Durham University

Durham E-Theses

Functional mapping of pea legumin upstream regulatory elements using TI plasmid vectors

Garrett, Christine

How to cite:

Garrett, Christine (1991) *Functional mapping of pea legumin upstream regulatory elements using TI plasmid vectors*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/6041/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Declaration

I declare that the work contained within this thesis submitted by me for the degree of Doctor of Philosophy is my original work and has not been submitted for a degree at this or any other University.

FUNCTIONAL MAPPING OF PEA LEGUMIN UPSTREAM REGULATORY ELEMENTS USING TI-PLASMID VECTORS

Christine Garrett, B.Sc.

A thesis submitted in accordance with
the requirements for the Degree of
Doctor of Philosophy in the University
of Durham

The copyright of this thesis rests with the author.
No quotation from it should be published without
his prior written consent and information derived
from it should be acknowledged.

October, 1991

Department of Biological Sciences.

18 AUG 1992

Abstract

The *leg A* gene from *Pisum sativum L.* has been extensively characterised and a distinct pattern of developmental and organ-specific gene expression demonstrated. Homology between legumin genes from other species has given some indication of those sequences which may be responsible for the regulation at the level of transcription. This study was designed to provide a functional analysis of the upstream sequences.

A number of plasmid vectors containing a maximum of 1.2 kb of upstream sequence from the *leg A* gene of *Pisum sativum L.*, ligated to the coding region of the nopaline synthase (*nos*) gene, were constructed. The use of smaller promoter fragments and the insertion of spacer DNA within the promoter region was employed in an effort to localise the regions of 5′ flanking sequence which may play a role in tissue specific expression.

In a minority of tumours derived from tissue transformed with the vector containing the ' full-length ' *leg A* promoter, low levels of nopaline were detected, but not with those containing a shorter promoter fragment. Results from the analysis of seed tissue indicatesthat 800 bp of the *leg A* promoter was insufficient to direct tissue-specific expression of the fused nopaline synthase gene in transgenic *Nicotiana tabacum*, although one individual plant showed a constitutive pattern of nopaline synthesis.

However, published results obtained with legumin and other storage protein gene promoters would suggest that this promoter fragment should have been sufficient to confer seed-specific expression. This suggests that there may have been undesirable secondary structures, or small undetected rearrangements, introduced during the construction of the transcriptional fusions between *leg A* and *nos*. Alternatively the marker gene may be inadequately sensitive to permit detection of low levels of expression.

Acknowledgements

I would like to acknowledge the outstanding feats of memory, support and cajoling my supervisor has provided throughout the course of this studentship.

I must acknowledge Dr Ron Croy and Professor Boulter for the use of the genomic clone of *leg A*; Dr Ray Ellis and Dr Andy Hepher for their advice on tissue culture techniques and the provision of a *Nicotiana* tissue line which constitutively expressed nopaline synthase; Dr Catherine O'Reilly and Dr Anderton Ryan for the provision of many molecular biology techniques, suggested approaches and valuable short cuts that are not within the pages of Maniatis; Christine Shaw for basic techniques used for *Agrobacterium* manipulations; and Dr Angray Kang for his help with the attempt to develop a nopaline ELISA technique.

I finally wish to thank all those who have provided the humour and a range of tips and advice over the years.

Abbreviations

All abbreviations used here are derived from the 'Instructions to Authors,' Biochemical Journal **249** (1988) 1-20, with the following exceptions:

A_{260}	=	absorbance at 260nm
Ap	=	ampicillin
ARE	=	anaerobic regulatory element
bp	=	base pairs
BSA	=	bovine serum albumin
C-terminal	=	carboxy terminus of a peptide
CAM	=	crassulacean acid metabolism
CaMV	=	cauliflower mosaic virus
CAT	=	chloramphenicol acetyl transferase
cDNA	=	complementary DNA
CIP	=	calf intestinal alkaline phosphatase
Cm	=	chloramphenicol
c.p.m.	=	counts per minute
d.a.f.	=	days after flowering
DEPC	=	diethylpyrocarbonate
DMSO	=	dimethylsulphoxide
DMF	=	dimethylformamide
DNAse	=	deoxyribonuclease
dNTP	=	deoxyribonucleoside triphosphate
DTT	=	dithiothreitol
ER	=	endoplasmic reticulum
Ery	=	erythromycin
EtBr	=	ethidium bromide
Gm	=	gentamycin
GUS	=	β-glucuronidase
HSE	=	heat shock element
kb	=	kilobase pairs

kD	=	kilodalton
Km	=	kanamycin
KTi I	=	kunitz trypsin inhibitor I
leg A	=	legumin A gene from *Pisum sativum*
LIH	=	limited internal homology
LRE	=	light responsive element
N-terminal	=	amino terminus of a peptide
Nm	=	neomycin
nos	=	nopaline synthase gene
npt*II*	=	neomycin phosphotransferase
ocs	=	octopine synthase gene
ORF	=	open reading frame
PEG	=	polyethylene glycol
LEG 0.7	=	legumin *A* promoter fragment from pDUB1301
LEG 1.2	=	legumin *A* promoter fragment from pDUB1300
poly A^+ RNA	=	polyadenylated RNA
Rif	=	rifampicin
RNAse	=	ribonuclease
SDS	=	sodium dodecyl sulphate
SDW	=	sterile distilled water
SEV	=	split end vectors
Sm	=	streptamycin
Sp	=	spectinomycin
SSC	=	saline sodium citrate
SV40	=	simian virus 40
Tc	=	tetracycline
T-DNA	=	transforming DNA of the Ti-plasmid
Ti-plasmid	=	tumour inducing plasmid
X-gal	=	5-dibromo-4-chloro-3-indoylgalactoside
5′	=	5′ terminal phosphate of a DNA or RNA molecule
3′	=	3′ terminal hydroxyl of a DNA or RNA molecule

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

Historically, plant seed proteins have been the focus of investigation because of their economic importance and their ease of study. The seed protein genes are only expressed in a specific organ at a predetermined stage of development, allowing the relatively simple identification of the seed-specific transcripts. The study of the controlling elements involved in seed protein expression is worthy of interest to shed light on the determination of developmental expression and to enable seed-specific expression of introduced protein genes using current genetic engineering techniques. Additionally, plants are a good system for studying the regulation of developmentally-regulated and tissue-specific genes because of the ease with which whole plants can be regenerated from undifferentiated, transformed tissue of an increasing number of plant species.

Major improvements in the yield of a number of crop plants have been made in recent decades, achieved by the use of fertilizers, pest control and by genetic improvements. These genetic improvements are the result of conventional plant breeding techniques, for example, exploiting hybrid vigour and selecting for high yielding and pest resistant strains. The introduction of a useful characteristic by sexual crosses requires numerous backcrossing experiments in order to produce a plant with just the desired characteristic without any detrimental effects on other aspects of the plant and is therefore time-consuming. However, the rate of these improvements has decreased in recent years and new methods of crop improvement have been investigated.

These techniques have included plant tissue culture, protoplast fusion and transformation of plant tissue with foreign or altered genes. The use of these techniques has circumvented certain limitations of conventional plant breeding by allowing useful characteristics to be introduced into the desired plant from a donor that is sexually incompatible. The use of plant tissue culture allows rapid propagation of improved varieties and the introduction of a single gene into a crop plant avoids the lengthy back-crossing experiments and opens up the possibility

of adding genes from distantly related organisms, even from prokaryotes, if the correct controlling sequences are introduced. Therefore, genetic engineering techniques may be employed in suitable cases to avoid some of the more protracted aspects of conventional plant breeding. Intensive farming methods used in the developed world have already provided high yield crop production and there may be scope for qualitative rather than quantitative improvements in crop plants, which are more amenable to genetic engineering techniques, being encoded by a single or a few genes. However, these new methods are likely to supplement rather than replace conventional plant breeding techniques, particularly for those aspects of plant physiology that have not been fully characterised at the gene level (Miflin and Lea, 1984).

A number of methods are currently available to introduce DNA into plant cells including transformation of protoplasts with DNA using electroporation, calcium chloride or PEG (Shillito and Saul, 1988) and more recently, microprojectiles (Klein *et al.* 1988). The regeneration of transformed, fertile maize plants has recently been reported using the last technique (Gordon-Kamm *et al.*, 1990). These methods are used especially for the Gramineae which are not easily transformed using Ti-plasmid vectors. Unfortunately, the regeneration of fertile plants from these transformed cell lines has often proved difficult. More success has been achieved with the dicotyledenous plants which are relatively easily transformed with Ti-plasmid vectors, using tissue pieces or protoplasts, and in an increasing number of species whole plants have been regenerated e.g. tobacco, petunia, tomato and potato.

A number of genes have been successfully expressed in transformed plants including various antibiotic resistance genes, pesticide resistance genes, viral proteins and seed storage proteins. Usually the plant genes are expressed in heterologous systems, particularly tobacco, because the majority of crop species are not readily regenerated. In the earliest work these genes were under the control of constitutive promoters such as the nopaline synthase promoter or promoters

from the 19S and 35S proteins of cauliflower mosaic virus. The possibility of introducing modified seed protein genes using these techniques will necessitate the expression of the foreign gene in the correct tissue in a developmentally-regulated manner.

The large protein-rich seeds of plants which belong to the legume family are an important source of both human and animal food in many areas of the world. The principal food legumes include peas, kidney beans, lima beans, lentils, chick peas, mung beans, cow peas, soy beans, peanuts, and broad beans. In many poor countries, the legumes are the most important high protein food (Chrispeels and Sadava, 1977). However, the use of legume crops as a sole protein source is limited by the amino acid composition of the seed proteins. In particular, the sulphur containing amino acids are poorly represented in most commercial legume varieties. Attempts to enhance the sulphur amino acid content of legumes by traditional plant breeding techniques have only been partially successful (Payne, 1983). Therefore, the improvement of the nutritional quality of legumes for both human and animal consumption by the genetic manipulation of the seed proteins would be an obvious candidate for these techniques (Shewry *et al.*, 1981; Croy and Gatehouse, 1985).

In order to express altered proteins in the seed, an understanding of the mechanisms which determine the expression of the storage proteins in the relevant tissue will be required. This study is intended as an initial exploration of the mechanism of the tissue-specific expression of one legume storage protein, legumin, from *Pisum sativum L.*

1.2 INTRODUCTION OF FOREIGN GENES INTO PLANTS

The study of plant gene regulation has required the introduction of modified genes into plant cells. A number of systems are available including transient expression in protoplasts, stable integration into the plant genome using direct transfer techniques or Ti-plasmids, to generate callus tissue or mature plants.

The technique chosen will depend on the type of gene to be expressed and the availability of reliable protocols for the plant species to be studied.

Transient expression of introduced genes in plant protoplasts allows rapid analysis of promoter deletions and has facilitated the investigation of certain inducible genes (Howard *et al.*, 1987). Some genes including ribulose 1,5-bisphospate carboxylase are switched off during protoplast formation (Fleck *et al.*, 1979; Vernet *et al.*, 1982) which limits the usefulness of this approach.

Foreign DNA has been introduced and integrated into the plant genome by direct transfer methods and transgenic plants regenerated from the treated protoplasts (Hain *et al.*, 1985; Paszkowski *et al.*, 1984; Potrykus *et al.*, 1985a). These techniques have been used to investigate the expression of introduced genes in a number of monocots e.g. *Triticum monococcum* (Lörz *et al.*, 1985; Werr and Lörz, 1986); *Oryza sativa* (Uchimiya *et al.*, 1986); *Lolium multiflorum* (Potrykus *et al.*, 1985b). Electroporation has been developed to transform plant cells (Fromm *et al.*, 1985) and the integration of the introduced DNA into a monocot genome has been reported using this technique (Fromm *et al.*, 1986). More recently, a report demonstrating expression of an introduced *npt*II gene in plants regenerated from electroporated maize protoplasts has been published (Rhodes *et al.*, 1988).

The ability of *Agrobacterium tumefaciens* to integrate stably part of a large tumour inducing plasmid into the plant genome during infection has resulted in the development of a number of vectors based on these plasmids. Transformation of plants using Ti-plasmid vectors is the simplest method currently available to introduce foreign genes into dicotyledonous plants. There are very few reports of successful transformation of monocotyledonous plants with *Agrobacterium tumefaciens*, although opine production has been detected following *A. tumefaciens* infection in certain species, including two members of the Lilaceae and Amaryllidacae families (Hooykaas-Van Slogteren *et al.*, 1984), *Asparagus* (Hernalsteens *et al.*, 1984), maize (Graves and Goldman, 1986a), *Gladiolus* (Graves and Goldman, 1986b) and sugar cane (Schäfer *et al.*, 1987).

Constitutive and light-regulated plant genes have been studied in callus tissue induced by oncogenic Ti-plasmid vectors, which rapidly produce assayable tumour mass. However, with more tightly regulated genes the results have sometimes been conflicting and organ-specific expression has necessitated the development of Ti-plasmids in which the genes coding for hormone biosynthesis, on the transferred portion of the Ti-plasmid (the T-DNA), have been deleted. Plant tissues transformed by these plasmids, known as disarmed vectors, have a normal hormone balance and allow the regeneration of phenotypically normal plants of some plant species. The most commonly used hosts include tobacco and petunia.

1.2.1 Development and biochemistry of Ti-plasmid vectors

The ability of *Agrobacterium tumefaciens* strains carrying a tumour inducing (Ti)-plasmid (Van Larebeke *et. al.*, 1974; Watson *et. al.*, 1975; reviewed by Bevan and Chilton, 1982) to transfer a specific portion of the plasmid, known as the T-DNA, to the plant genome (Chilton *et. al.*, 1977; Chilton *et. al.*, 1978; reviewed by Nester *et al.*, 1984), has resulted in the development of a range of vectors which utilise this phenomenon (Klee *et al.*, 1987).

The appearance of a tumour following infection with an *Agrobacterium tumefaciens* strain occurs because of the synthesis by the infected tissue of plant growth substances which lead to the disorganised tissue growth observed (Schröder *et al.*, 1983; Thomashow *et al.*, 1984; Thomashow *et al.*, 1986). At this stage the inciting bacteria are no longer required because of the stable transfer of the T-DNA to the plant genome which encodes the genes responsible for the auxin and cytokinin synthesis pathways under the control of sequences recognised by the plant RNA polymerases.

Additionally, the T-DNA encodes genes for the synthesis of opines, for example nopaline synthase or octopine synthase, which can be metabolised only by the inciting bacterium and other nopaline or octopine strains, respectively. The appearance of these opines has been used as a marker for the integration of the T-

DNA, especially where the genes encoding the phytohormones have been deleted in order to regenerate phenotypically normal plants.

In tumour tissues the T-DNA transcripts, which are not expressed in the bacterial cell, account for approximately 0.001% of the polyadenylated RNA (Willmitzer *et al.*, 1982). In most vectors the insertion of a gene coding for antibiotic resistance within the T-DNA allows the selection of transformed tissue in culture.

The near-perfect 25 bp repeats that flank the T-DNA (Zambryski *et al.*, 1982; Wang *et al.*, 1984) are recognised by a site-specific endonuclease encoded by the *virD* operon, which generates a single-stranded intermediate (Stachel and Zambryski, 1986; Yanofsky *et al.*, 1986). The orientation of the right border specifies the direction of the T-DNA transfer and deletion of these repeats abolishes transfer (Shaw *et al.*, 1984b; Wang *et al.*, 1984). The remainder of the T-DNA is not required, and vectors in which the T-DNA has been replaced by foreign genes, including a selectable marker gene, have been developed flanked by the essential border sequences.

The *vir* genes (Klee *et al.*, 1983; Horsch *et al.*, 1986), which are involved in this transfer event, are located in another portion of the Ti-plasmid. They are not cotransfered and are functional when present *in trans* on a separate replicon as part of a binary system (Hoekema *et al.*, 1983). The *vir* region comprises at least six operons which encode a number of functions of the T-DNA transfer process, including the original chemotactic response (Ashby *et al.*, 1987; Shaw *et al.*, 1988) and the induction of *vir* genes in response to plant exudates (Stachel *et al.*, 1985; Stachel and Zambryski, 1986).

Another locus, *overdrive*, located just outside of the flanking repeats of the T-DNA, has been identified; this increases the efficiency of T-DNA transfer (Peralta *et al.*, 1986) but fortuitously most vectors possess this fragment although transformation has been achieved when this region is absent.

1.2.2 Cointegrating vectors

This approach uses homology between sequences of a resident Ti-plasmid with an intermediate vector to introduce novel genes within the T-DNA region. The intermediate vector is capable of replication in *E.coli* and the original gene manipulations are carried out in this vector which, being much smaller than the Ti-plasmids, is more easily handled and may possess one or more unique restriction sites.

The intermediate vector may be unable to replicate in *A. tumefaciens*, which allows direct selection of integration by the ability of the Agrobacteria to grow on an antibiotic specified by the integrating vector, or alternatively may have a broad host range allowing it to replicate in both *E.coli* and *Agrobacterium tumefaciens*. The requirement for homology means that the vector will be capable of integrating into only one or a few specific Ti-plasmids which may limit host range (De Cleene and De Ley, 1976). In reports of successful transformation using these vectors, the majority of plants had only a single copy of the gene of interest (Spielmann and Simpson, 1987).

Zambryski and coworkers (1983) have developed a cointegrating vector based on the disarmed *Agrobacterium* Ti-plasmid pGV3850. The phytohormone genes have been replaced by pBR322 sequence, which allows any plasmid with homologous sequences to cointegrate. The Ti-plasmid provides the border sequences, a nopaline synthase gene and the *vir* functions. Cointegration with the pBR322 sequences places the integrated vector within the T-DNA region which will be cotransfered with the nopaline synthase gene, allowing easy identification of plant transformants.

Cointegrating vectors which also contain a selectable marker have been developed, for example the pLGVneo1103 vector which contains a single *Eco*RI site and a *npt*II gene fused to the *nos* promoter to code for kanamycin resistance in the transformed plant tissue (Herrera-Estrella *et al.*, 1984). An oncogenic variant,

pGV3851 has been constructed (Zambryski *et al.*, 1984) containing a smaller internal T-DNA deletion and retaining the *tmr* gene which generates a phytohormone independent shooty phenotype to enable easy identification of transformants.

The integrative vectors described by Rogers *et al.* (1986), such as pMON200, possesses an intact nopaline synthase gene, a multilinker for easy insertion of DNA fragments and an *npt*II gene under the control of *nos* flanking sequences. In addition, pMON200 contains a nopaline border sequence and a small portion of the octopine-type T-DNA, referred to as the Limited Internal Homology (LIH). The latter sequence allows recombination with octopine Ti-plasmids, such as pTiB6S3E in which the phytohormone genes have been removed. The Ti-plasmid vector contains the T_L left border and following the integration event, the left and right hand border sequences are correctly reconstructed flanking the genes of interest. This approach is known as the SEV (split end vectors) system (Fraley *et al.* 1985).

A transmission frequency for pBR322 based replicons from *E.coli* to *Agrobacterium tumefaciens* of $4.5 \times 10^{-}3$ has been reported by Van Haute *et al.* (1983) and a frequency of $2 \times 10^{-}2$ observed by Zambryski *et al.* (1983) for the recombination event with the homologous Ti-plasmid. The cointegrating vectors are small, easily manipulated and readily maintained in both the *E.coli* host and, once stably integrated in the Ti-plasmid, the *Agrobacterium* strain. There is also some evidence that efficiency of plant transformation may be higher with a cointegrating vector than a binary system (Zambryski *et al.*, 1984; McCormick *et al.*, 1986).

1.2.3 Binary vectors

Binary vectors exploit the observation that the *vir* region and the T-DNA can be physically separated on two plasmid replicons within an *Agrobacterium* host, without affecting the transfer of the T-region (de Frammond *et al.*, 1983; Hoekema *et al.*, 1983). This approach was first developed by Hoekema and co-workers (1983)

using pLA4404, a derivative of the octopine type Ti-plasmid pTiAch5, to provide the virulence functions in the transformation strategy. The binary vectors can be used with a disarmed Ti-plasmid or with a wild-type Ti-plasmid which may be useful in extending the host range of the vectors. The shuttle vectors tend to be less stable in the *Agrobacterium* host than the wild-type Ti-plasmids and require antibiotic selection for their maintenance.

An *et al.*, (1985) constructed a number of shuttle vectors with a range of useful characteristics including: i) a *nos-npt*II gene; ii) the ColE1 replicon; iii) the *cos* site of phage λ; iv) T-DNA border sequences; and v) a broad-host range replicon. The transformation of a range of plant species including tobacco, potato, tomato and *Arabidopsis thaliana* has been achieved using these vectors (An *et al.*, 1986b), although a wild-type Ti-plasmid was required for *Arabidopsis* transformation.

Similarly, the development of a number of binary vectors was reported by Bevan (1984), using the *vir* functions of pAL4404 to transfer the T-DNA from the shuttle vectors. One vector, pBIN6 contains a *nos-npt*II fusion as a selectable marker in plant tissues, a prokaryote kanamycin resistance gene and single restriction sites for *Sal*I and *Eco*RI. This vector also contains the left and right border sequences and a nopaline synthase gene. However, at 15 kb it is too large for efficient blunt end ligation and a derivative, pBIN19, of 10 kb was also constructed. To obtain the smaller vector the nopaline synthase gene was removed and unwanted T-DNA sequences were deleted with *Bal*31. In addition, a 440bp *Hae*II fragment from M13mp19, containing sites for *Eco*RI, *Bam*HI, *Hin*dIII, *Sst*I, *Kpn*I, *Sma*I, *Xba*I and *Sal*I, was inserted 80bp from the left border within the T-DNA. This also allows the presence of an insertion in this region to be distinguished using X-gal.

1.3 REPORTER GENES

Although, there have been numerous reports of the introduction of various plant genes under the control of their own regulatory sequences in heterologous

systems, the use of reporter genes allows the study of gene regulation in a homologous system. The enzyme activities encoded by these marker genes allows low levels of gene transcription to be detected where there is not an endogenous enzymatic activity. Although most plant species can be transformed using direct DNA transfer or Ti-plasmid vectors, the regeneration of fertile plants from transformed tissue of some commercial crop species has been less successful. The regeneration of transformed pea tissue has only very recently been achieved (Puonti-Kaerlas *et. al.*, 1990).

One of the earliest reporter genes used in transformed plants was octopine synthase (De Greve *et al.*, 1982). The opine product of octopine or nopaline synthase activity can be easily detected by the method of Otten and Schilperoort (1978). Originally used as the entire gene to register successful integration of the T-DNA, octopine synthase has also been used in chimaeric constructions with the coding region under the control of various putative promoter sequences in a number of expression cassettes (Herrera-Estrella *et al.*, 1983b; Jones *et al.*, 1985). Octopine synthase is not functional following amino terminal fusions, which limits its usefulness in some systems (Jones *et al.*, 1985).

Nopaline synthase activity has been successfully used as an indicator of gene expression by a number of workers (Depicker *et al.*, 1982; Bevan *et al.*, 1983a; Hepburn *et al.*, 1983; Willmitzer *et al.*, 1983; Shaw *et al.*, 1984a; Jones *et al.*, 1985). The assay is quick and does not employ expensive and potentially hazardous radiochemicals. The detection of the opine product in tumours transformed with vectors carrying various promoter deletions of the *nos* gene was used to map the nopaline synthase promoter (Shaw *et al.*, 1984a). It has been used, fused to a *rubisco* small subunit promoter fragment from soybean, to demonstrate light-inducible enzyme activity in a distantly related plant, *Kalanchöe daigremontiana* (Shaw *et al.*, 1986). These authors have also reported that the nopaline assay is 20-100 times more sensitive than dot-blot analysis of RNA (Shaw *et al.*, 1986). When this study was initiated, the nopaline synthase assay system was well char-

acterised, shown to be sensitive, easy to carry out and suitable vectors were available.

More recently a number of other reporter genes have been employed. The bacterial genes chloramphenicol acetyl transferase and neomycin phosphotransferase encode enzymes that are not normally found in plant tissues, although competing enzyme reactions from endogenous esterases, phosphatases and transferase can limit sensitivity. However, a number of plant promoters have been investigated using chimaeric fusions with these genes (see section 1.4; Bevan *et al.*, 1983b; Fraley *et al.*, 1983; Herrera-Estrella *et al.*, 1983a,b) despite the assays being relatively expensive and tedious to perform.

Chloramphenicol acetyltransferase has been widely used as a reporter of gene expression directed by fused prokaryote, animal and plant promoter sequences. A radiolabelled enzyme assay in the presence of plant extracts detects the conversion of [^{14}C]-chloramphenicol to its 1,3 and 1-3 acetylated derivatives. This results in altered mobility of the inactivated antibiotic, as detected by thin layer chromotography, when this gene is expressed. Certain plant extracts contain inhibitors of the enzyme which can be partially overcome by the addition of ascorbic acid and L-cysteine and others may have non-specific acetylases giving a high background signal in untransformed tissues.

The ability of the Tn5 gene, neomycin phosphotransferase, to phosphorylate specifically aminoglycoside antibiotics such as kanamycin and G 418 has allowed both direct selection of plant tissues transformed with this gene and the development of an assay to monitor its activity. The detection of this gene activity in transformed plant cells was first carried out by Herrera-Estrella *et al.* (1983b). The *nptII* gene retains enzyme activity when fused to an amino terminal peptide and has been used to study the effect of putative signal peptides on organelle transport (De Block *et al.*, 1985). An assay system has been developed using ^{32}P labelled ATP as a substrate which allows autoradiography to be used to detect the phosphorylation reaction (Reiss *et al.*, 1984). This assay is expensive but it

can detect as little as 1ng of active enzyme and demonstrate changes in the size of the enzymatically active proteins (Lichtenstein and Draper, 1984).

Other reporter genes have also been investigated. The luciferase reaction can generate interesting results in transgenic plants (Ow *et al.*, 1986) but the enzyme is not stable and difficult to assay accurately (DeLuca and McElroy, 1978). However, the development of an assay to detect enzyme activity from fusions with the coding region of the β- glucuronidase (GUS) gene was reported by Jefferson and co-workers (1987) and has been widely used (for recent examples see Benfey *et. al.*, 1990; Gordon-Kamm, *et. al.*, 1990; Ohl, *et. al.*, 1990; Schmid, *et. al.*, 1990; Szabados, *et. al.*, 1990; Thomas and Flavell, 1990). This does not appear to have the same problems as β- galactosidase, which suffers from high endogenous levels in some plant tissues (Helmer *et al.*, 1984) although this is not true for all tissues studied (Plegt and Bino, 1989). The enzyme is functional following amino terminal fusions, can be detected using a range of commercially available spectrophotometric, fluorimetric, and histochemical subtrates, and is relatively stable (Jefferson, *et.al.* 1987).

1.4 PLANT GENE EXPRESSION

1.4.1 Eukaryote gene expression

In 1984, when this investigation started, there was little information available on those sequences which were functionally significant for plant gene regulation. In other eukaryote systems a number of common motifs of similar sequence had been identified upstream of the start of transcription. One of these, the promoter sequence called the 'TATA' box (Breathnach and Chambon, 1981) appear to be essential for the precise localisation of the transcription initiation site. A conserved 'CAAT' box (Benoist *et al.*, 1980), required for the expression of many animal genes (Corden *et al.*, 1980; Hentschel and Birnstiel, 1981; Grosveld *et al.*, 1981; Tsai *et al.*, 1981; Tokunaga *et al.*, 1984), had also been identified.

Similar sequences have been identified in the 5′ region of plant genes that have

been sequenced (Messing *et al.*, 1983; Heidecker and Messing, 1987).

An element called an 'AGGA' box was identified (Messing *et al.*, 1983) in addition, or instead of, the CAAT box in some plant genes, although whether this is functionally significant is unclear.

Differences in codon usage between animal and plant genes have also been noted (Lycett *et al.*, 1983a, Lütcke *et al.*, 1987). There is more variability in the conserved sequences in the 3′ untranslated region (Lycett *et al.*, 1983a) compared to animal genes and experiments have shown that animal gene polyadenylation signals are not properly recognised by plant cells (Hunt *et al.*, 1987).

The diverse pattern of eukaryote gene expression, both spatial and temporal, requires that many genes possess additional regulatory elements which modulate the level, tissue-specificity or developmental stage of expression. Some authors have proposed a tripartite structure for a functional promoter (Grosschedl and Birnstiel, 1980), comprising an initiator, selector and modulator elements. An initiator sequence proximal to the start of transcription ('TATA' and 'CAAT' boxes) controls accurate initiation of transcripts but the tissue in which the gene is active may be controlled by an element upstream, the selector. Similar sequences may also respond to the environmental state in which expression of the gene occurs. The activity of the promoter may be modulated by an enhancer which influences the level of expression and may be tissue-specific (Voss *et al.*, 1986) or more general in action and largely independent of location or orientation on the same piece of DNA.

These *cis*-acting promoter elements are thought to act by interaction with *trans*-acting transcription factors (Dyan and Tjian, 1985), sequence-specific DNA-binding proteins such as the Sp1 factor, found in mammalian cells and required for SV40 transcription. Some experimental results suggest that these regulatory sequences act cooperatively to produce the particular pattern of expression of their associated gene, via interactions of their specific binding proteins (Garcia *et al.*, 1986; McKnight and Tjian, 1986; Schüle *et al.*, 1988). Although in some cases

the specific DNA binding proteins have been shown to be ubiquitous, being found in cell types or under environmental conditions in which the gene is inactive, post-translational modification of these factors may be involved in their regulatory role (Maniatis *et al.*, 1987).

The factor that binds the TATA sequence, designated TFIID, has been isolated from a number of eukaryote organisms and shown to possess conserved amino acids in the C-terminal region (reviewed by Latchman, 1990; Lewin, 1990; Ptashne and Gann, 1990). Binding of this factor to the TATA box causes a conformational change which facilitates the binding of other factors such as TFIIC, TFIIE and RNA polymerase to form a stable transcription complex. Two cDNA clones for TFIID have been identified in *Arabidopsis thaliana* (Gasch *et. al.*, 1990). In addition, the use of sequence-specific DNA affinity chromotography has been used to isolate DNA binding proteins which bind the CCAAT motif found in many animal genes (Cohen *et al.*, 1986; Jones *et al.*, 1987).

Many of these *trans* acting factors have now been shown to conform to three protein structural types (reviewed in Johnson and McKnight, 1989). The first of these is called the helix-turn-helix motif in which two α-helixes are separated by a sharp beta turn. These proteins bind as dimers to DNA sequences showing dyad symmetry, one α-helix of each protein interacting with the specific recognition sequence situated in the major groove of the DNA helix. Examples of this type of structure are found in the products of *Drosophila* homeotic genes and the mating type locus of *Saccharomyces cerevisiae.*

A second type of DNA-binding protein identified is the zinc finger motif which is characterised by conserved cysteine and histidine residues which generate a loop structure within the protein molecule via the binding of a zinc ion. The first example of this arrangement was observed in the transcription factor IIIa from *Xenopus laevis.* A constitutive element of a pea light-inducible promoter has been shown to bind a protein factor, termed 3AF1, in a metal dependent manner. A tetramer of this element was used to screen a cDNA library and the deduced

protein product of a positive clone showed a putative zinc finger structure (Lam *et. al.*, 1990).

Another conserved structure has been noted in the transforming proteins Fos, Myc, Jun and the yeast regulatory protein GCN4. These proteins share a heptad array of leucine residues which are involved in the dimerization of these molecules and termed a leucine zipper. This arrangement is thought to bring an adjacent basic region of each protein in the dimer into direct contact with the DNA. This feature has been found in some plant regulatory proteins such as EmBP-1 which binds the ABA response element (Guiltinan *et. al.*, 1990) and ASF-1 (TGA1a) which binds the motif TGACG found in the promoters of nopaline synthase, wheat histone H3 and the 35S promoter of cauliflower mosaic virus (Katagiri *et. al.*, 1989).

Sequencing of related genes from plants has focused attention on certain conserved elements in the 5′ region of the genes which may play a role in their regulation. Computer analysis of upstream sequences that are highly conserved between members of a gene family were implicated by Davidson and associates (1983) in the particular type of regulation of their associated genes, creating selection pressure against substitution or deletion of these nucleotides. Similarly, a more recent analysis of highly recurring sequence elements in eukaryotic DNAs has shown that these are often homologous to regulatory sequences or protein binding sites (Bodnar and Ward, 1987).

Specific, conserved elements have been noted in the histone, globin, actin and glucocorticoid-responsive gene families in animals (Grosschedl *et al.*, 1980; Dierks *et al.*, 1983; Groner *et al.*, 1984; Nudel *et al.*, 1985) and the light-inducible and seed-specific gene families in plants (Bäumlein *et al.*, 1986; Brown *et al.*, 1986; Coruzzi *et al.*, 1986; Kuhlemeier *et al.*, 1987b). Another example is the heat shock protein genes that are found in organisms as divergent as *Drosophila* (Pelham, 1982), man (Wu *et al.*, 1986) and soybean (Czarnecka *et al.*, 1985; Schöffl *et al.*, 1984) and shown to possess a distinctive sequence motif which is a protein

binding site in *Drosphila* (Wu, 1984). However, the binding of protein factors to specific DNA sequences may show a quite complex pattern. For example, Weinberger and associates (1988) demonstrated that distinct factors bound apparently homologous sequences in the immunoglobulin heavy chain enhancer. Therefore it is not sufficient to simply identify homologous sequences because the exact context may play a significant role.

An enhancer element was first identified in the DNA tumour virus SV40 (Banerji *et al.*, 1981) distinguished by the ability to stimulate expression of a gene independent of its position relative to the start of transcription in a variety of tissue types. Similarly, in the constitutively expressed octopine synthase gene, an element that can stimulate expression of heterologous plant promoters has been identified (Ellis *et al.*, 1987a, 1987b). Enhancers associated with developmentally regulated genes expressed in a specific cell-type have been identified in the 5′ flanking region and introns of the immunoglobulin gene families (Bergman *et al.*, 1984; Voss *et al.*, 1986) as well as within the structural genes of human α and β globin (Charnay *et al.*, 1984). A light-responsive element in plants has been isolated from the upstream region of the small-subunit of ribulose-1,5-bisphosphate carboxylase which shows enhancer-like properties (Kuhlemeier *et al.*, 1987b). However, the precise mechanism of action of enhancer elements is unclear but it has been proposed that there may be looping of the DNA (Ptashne, 1986) such that distant enhancer elements can interact with promoter sequences, presumably mediated by DNA-binding proteins.

A number of animal genes were used to transform plant tissue when suitable plant genes were not available but expression of these genes was rarely detected. The genes included developmentally regulated genes e.g. rabbit β-globin (Shaw *et al.*, 1983) or chicken ovalbumin (Koncz *et al.*, 1984), which are only expressed in one particular tissue-type in the animal system, so this is hardly surprising. However, these experiments did demonstrate the successful integration of foreign genes into the plant genome, usually employing *Agrobacterium tumefaciens* and

vectors derived from Ti-plasmids (see section 1.2).

As most plant genes have been isolated from commercial species which are not readily transformed (most monocotyledonous species) or regenerated from transformed tissue (some dicots e.g. pea), functional analysis of domains important in the regulated expression of these genes has involved the transformation of more amenable species such as tobacco or petunia. In most cases the introduced genes have been correctly expressed in the introduced tissue but there is some evidence that monocot pre-mRNAs are not efficiently processed in tobacco (Keith and Chua, 1986).

Chimaeric constructs of putative regulatory sequences with various marker genes, encoding a more readily detectable protein product have been employed (see section 1.3). These marker genes can increase the likelihood of detecting low levels of expression in a heterologous system and are particularly useful in homologous systems when reintroducing various promoter deletions of an endogenous gene product.

In plants, the earliest genes to be studied were those showing a constitutive pattern of expression, allowing relatively rapid and simple analysis in callus tissue. The functional mapping of nopaline synthase (*nos*) was one of the earliest studies undertaken (Shaw *et al.* 1984a, An *et al.* 1986a). Although the gene does not occur naturally in plants, being introduced into the genome during infection with virulent Agrobacteria, it shows eukaryotic characteristics at the 5′ end and is constitutively expressed in plant tissues. These constitutive promoters have been used to test implied enhancer action of isolated sequence elements(Simpson et. al. 1986b

1.4.2 Constitutive plant genes

The ability to express constitutively an antibiotic resistance gene in plant tissues has formed the basis of the selection of phenotypically normal transformed plant tissue. Previously, transformation had been identified by hormone-independent growth following infection with oncogenic Ti-plasmid vectors, but the abnormal hormone levels prevented regeneration of whole plants in most cases. The study of constitutive plant promoters has allowed the development of chimaeric constructs in which these promoters drive the expression of bacterial antibiotic resistance genes (Bevan *et al.*, 1983b).

1.4.2.1 The genes of the T-DNA from *Agrobacterium tumefaciens*

Koncz and co-workers (1983), showed that the genes responsible for the synthesis of opines carried by Ti-plasmids had all the signals necessary for expression when transferred to plants during *Agrobacterium* infection. The nopaline synthase gene is expressed in all tissues examined in transformed tobacco (De Block *et al.*, 1984; Horsch *et al.*, 1984). Similarly, the octopine synthase gene functions in tumours, leaves, roots, stem and seeds of tobacco (Otten *et al.*, 1981). Analysis of the 5′ flanking sequence of the nopaline synthase gene revealed sequences resembling the 'CAAT' and 'TATA' boxes (Depicker *et al.*, 1982, Bevan *et al.*, 1983a). In contrast, the octopine synthase gene does not possess a recognisable 'CAAT' box.

Initial experiments demonstrated that sequences upstream of -261, relative to the cap site, were not required for *nos* expression (Koncz *et al.*, 1983) and more detailed analysis by Shaw and co-workers (1984a) delineated a 88 bp fragment immediately upstream of the cap site, which includes the 'TATA' and 'CAAT' boxes, sufficient for wild-type levels of nopaline synthase expression in *Kalanchoë* callus tissue. Deletion of the 'CAAT' box from this sequence reduces the level of expression by an order of magnitude and removal of the 'TATA' box completely abolishes expression (Shaw *et al.*, 1984a).

However, other workers (An *et al.*, 1986a) using a bacterial gene, chloramphenicol acetyl transferase (CAT), to monitor expression from the nopaline synthase promoter have described an additional element beween -130 and -101 important for *nos* promoter activity and showed some activity after deletion of the 'TATA' box from the 3′ end. These authors carried out both 3′ and 5′ deletions on chimaeric *nos*-CAT fusions and used a binary Ti-plasmid vector to generate transformed tobacco calli. The differences in the marker gene employed and the plant system used may be responsible for the conflicting patterns of expression observed. These authors have reported similar findings in transgenic tobacco and a transient expression system (Ebert *et al.*, 1987) and demonstrated that duplication of the upstream element tripled the promoter activity. The importance of sequences upstream of -100 is also shown by the work of Lam and co-workers (1990) in which binding of the tobacco nuclear factor, ASF-1, to the region -138 to -103, was shown. A synthetic tetramer of the region -131 to -111, designated *nos*-1, conferred leaf and root expression when fused to a truncated CaMV 35S promoter but not in a similar construction in which nucleotides within the two TGACG motifs of *nos*-1 had been mutated.

The nopaline synthase gene does show some differences in organ-specificity and developmental stage (An *et al.*, 1988). Higher levels of expression are found in younger plants than older specimens. Recent work indicates that the nopaline synthase promoter can be induced by wounding, a response which is enhanced by auxin (An *et. al.*, 1990). Examination of promoter deletions suggest that the important sequences for this response reside in the same region as *nos*-1.

It has been shown that the octopine synthase gene requires sequences between -170 and -294 for expression in tobacco and sunflower (Konz *et. al.*, 1983). Later work showed that fusion of an *ocs* promoter element 5′ to a chimaeric *Adh-1*-CAT gene was essential for high levels of anaerobically-induced gene expression (Ellis *et al.*, 1987a) in transgenic tobacco. Further analysis of this segment delineated a 16 base pair palindrome, located at -193 to -178 , capable of enhancing expression

when placed either 3′ or 5′ of the *Adh-1*-CAT gene in a transient expression system (Ellis *et al.*, 1987b). There is a homologous sequence in the flanking region of nopaline synthase gene in which 12/16 bases correspond and although this sequence does not appear to act as enhancer (Ellis *et al.*, 1987b), it lies in a 29 bp region (-130 to -101) identified by An *et al.* (1986a) as essential for *nos* expression.

More recently it has been shown that the *ocs* enhancer can confer tissue-specific expression on a gene construction consisting of the GUS coding region and the CaMV 35S TATA box (Fromm *et. al.*, 1989). GUS gene expression was detected in the root tip and the shoot apex of transgenic tobacco. These workers also showed that root-specific expression could be confered on the *rbcS*-3A promoter when the *ocs* palindrome was inserted at -55 relative to the start of transcription. Using gel shift assays and competition experiments it was shown that the *ocs* enhancer binds activation sequence factor (ASF)-1, a factor from tobacco nuclear extracts that interacts with the *as*-1 element of the CaMV 35S promoter (Fromm *et. al.*, 1989). Specific protein binding to the *ocs* element has also been found with maize nuclear extracts (Tokuhisa *et. al.*, 1990).

In the Ti-plasmid-encoded 780 gene, which is transcribed in crown gall tissue, Bruce and Gurley (1987) have identified an activator element located -440 to -229 upstream of the transcription start site which leads to a 100-fold decrease in transcription when deleted. Removal of the TATA element reduced promoter activity to $< 0.1\%$ of a cointegrated wild-type 780 gene and additional elements upstream have been indicated by some of the internal deletion experiments performed.

1.4.2.2 The 19S and 35S proteins of the Cauliflower Mosaic Virus

The domains of the promoter regions important for the constitutive expression of the two major transcripts from cauliflower mosaic virus (CaMV) have been characterised. The 35S promoter was active in all tissues examined in transgenic tobacco (Odell *et al.*, 1985), shows no light-regulation (Fluhr and Chua, 1986) and is a stronger promoter than the nopaline synthase (Sanders *et al.*, 1987)

or the 19S promoter (Lawton *et al.*, 1987) in transgenic plants. High levels of expression are obtained with a 35S promoter fragment from -343 to -46 (Odell *et al.*, 1985). Mutants deleted to -46, which still possess the TATA motif, show low levels of correctly initiated transcripts in tobacco, although this is not the usual host for the virus (Odell *et al.*, 1985). These authors have also shown that a region from -46 to -105, containing the 'CAAT' box, an inverted repeat region and a sequence resembling the consensus core for enhancers in animal systems (GTGGA/TA/TA/TG), gave increased levels of transcription and appears to play an accessory role in increasing levels of transcription in conjunction with upstream sequences (Fang *et. al.*, 1989). This more recent work has used an internal reference gene to standardize expression levels from different transgenic plants. It was shown that a 35S fragment (-209 to -46) could act as an enhancer to activate transcription from a heterologous TATA box (Fang *et. al.*, 1989).

DNA-footprinting studies have identified a region from -58 to -90 (*as*-1) in the 35S promoter which is protected by a protein factor present in leaf and root extracts from tobacco and pea (Lam *et al.*, 1989). This factor, ASF-1, binds the sequence TGACG in *as*-1 (activation sequence-1) but not mutated sequences. Mutations in the TGACG motif of the promoter attenuates root and stem expression in transgenic tobacco. When the *as*-1 sequence is inserted into a light-regulated promoter expression was detected in roots with some increase in leaf expression. A second factor (ASF-2) which binds in the -100 region (*as*-2) of the 35S promoter has been identified in nuclear extracts from tobacco leaves but not roots and shown to confer leaf expression when a tetramer of this sequence was fused to a truncated 35S promoter (Lam and Chua, 1989). Although this sequence shows homology to a GATA repeat conserved among several *cab* gene promters, the tetramer does not confer light-inducibility.

A 0.4 kb DNA fragment containing the promoter domain of the 19S transcript is sufficient to promote constitutive expression of the *nptII* gene in transformed petunia cells using oncogenic and partially disarmed Ti-plasmids (Koziel *et al.*,

1984). Similarly, by direct transfer to tobacco, a construct containing the coding region of *nptII* with the 5′ and 3′ signals from the 19S gene was shown to be functional (Potrykus *et al.*, 1985a). An enhancer-like element at -100 to -40 relative to the transcription start site of the 19S gene has also been identified (Kuhlemeier *et al.*, 1987b) using homologous and heterologous promoter fusions.

1.4.3 Inducible plant genes

Inducible plant genes have also been studied. In addition to the canonical promoter sequences found in plant genes, namely the TATA box and in some cases the CAAT sequence, additional sequences must be responsible for the induction of gene expression in response to environmental stimuli. Such elements have been identified in light-inducible and heat-inducible genes. Other plant genes have been investigated which require wounding or anaerobic conditions for expression.

1.4.3.1 Light-inducible genes

Transcription of many genes involved in photosynthesis is controlled by light (Tobin and Silverthorne, 1985; Ellis, 1986). This response is mediated by photoreceptors, the most thoroughly characterised in higher plants being phytochrome. Phytochrome negatively regulates the expression of its own gene or genes (Colbert *et al.* 1983, 1985) and controls the genes encoding the small subunit (*rbcS*) of ribulose-1,5-bisphosphate carboxylase (Tobin, 1981; Thompson *et al.* 1983) and the chlorophyll a/b binding (*cab*) protein of the light-harvesting chlorophyll-protein complex (Apel, 1979; Tobin, 1981).

A number of phytochrome-regulated genes studied showed different kinetics of accumulation of mRNA in response to a pulse of red light (Kaufman *et al.*, 1987) which suggests that there must be other factors in the pathway from light pulse to gene expression which control the exact pattern of the response. The control of these light regulated genes is complex, for example, members of the pea *rbcS* gene family show different light responses, levels of accumulation and tissue specificity

(Kuhlemeier *et al.*, 1987b).

Less than 1 kb of 5′ flanking region of a number of ribulose-1,5-bisphosphate carboxylase genes has proved sufficient to direct light-regulated and organ-specific expression (Broglie *et al.*, 1984; Herrera-Estrella *et al.*, 1984; Morrelli *et al.*, 1985; Nagy *et al.*, 1985; Timko *et al.*, 1985; Fluhr *et al.*, 1986a; Shaw *et al.*, 1986). The major *rbcS* transcripts in mature green leaves of pea are modulated by both phytochrome and a blue-light receptor (Fluhr and Chua, 1986). This photoregulation is maintained, together with leaf-specificity, when the genes, *rbcS-3A* and *rbcS-3C*, were transfered to petunia with 0.4 kb and 2 kb of upstream sequences, respectively (Fluhr and Chua, 1986). Similar *rbcS-3A* promoter fragments are sufficient to confer light-inducibility and organ-specificity on a CAT fusion or when ligated to a truncated CaMV promoter (Fluhr *et al.*, 1986a). Negative regulatory elements have been found in the 5′ noncoding region (-50 to -169) of *rbcS-3A* which decrease the level of transcription in the dark (Kuhlemeier *et. al.*, 1987a). A 58bp sequence in this region contains two regulatory elements, one of which is highly homologous to the SV40 core enhancer, and the other to an adenovirus enhancer and the constitutive part of the human interferon-β gene enhancer.

Pea *rbcS-E9* and *rbcS*-SS3.6, together account for less than 7% of rubisco small-subunit transcripts in mature green leaves (Fluhr *et al.*, 1986b). Relatively low levels of expression were observed in petunia calli transformed with a chimaeric *rubisco*-CAT construct containing 1 kb of upstream sequences from *rbcS-E9* (Morelli *et al.*, 1985) but normal levels of expression were observed with a 352 bp fragment in transgenic plants (Nagy *et al.*, 1985). A smaller upstream fragment from *rbcS-E9* (-317 to -82) is sufficient to direct light-inducible and leaf-specific expression (Fluhr *et al.*, 1986a).

Similarly, light-induction was obtained with a 900 bp fragment (-973 to -90) from the upstream region of the pea *rbcS-SS3.6* gene in a callus system using *npt II* as a reporter gene (Timko *et al.*, 1985). This element shows enhancer-like properties in that it functions independent of orientation but only when placed

5′ of the reporter gene. Simpson *et al.*, (1986b) were able to demonstrate light-regulation with a similar 850 bp fragment from *rbcS-SS*3.6, but no phytochrome response. Specific protein binding has been demonstrated for a region spanning -574 to -433 to a factor from pea, designated AT-1 (Datta and Cashmore, 1989). These workers identified a 33bp sequence (-566 to -533) as the smallest fragment to form a protein-DNA complex with this factor. Inspection of this sequence has revealed two copies of an AT-rich element (the AT-1 box) which is also present in other light regulated genes, including pea *rbcS-3A*, tomato *rbcS-3A* and the tobacco *Cab* E gene. The AT-rich fragments from tomato *rbcS*-3A and *Cab* E from *Nicotiana plumbaginifolia* were able to compete with the pea *rbcS*-3.6 promoter fragment (-574 to -433) for binding of AT-1. Phosphorylation of the AT-1 protein causes almost complete loss of activity which is consistent with the mechanism of regulation of some other regulatory proteins (Datta and Cashmore, 1989).

A conserved promoter structure, involved in the light regulation of rubisco genes from highly diverged plant species, is also implied by the observation by Shaw and associates (1986). A 900 bp promoter fragment from soybean, a C_3 plant, was sufficient to drive light regulated expression of a marker gene in transformed *Kalanchoë*, a CAM (Crussulacean Acid Metabolism) plant, despite the inherent differences in the tissue-specific pattern of rubisco expression between these species (Edwards and Huber, 1981). Therefore it is not surprising that a comparison of the upstream regions of a number of *rbcS* genes from higher plants has revealed a consensus sequence G/ATGTGG/TC/TCA/TATAT/AG in the -140 region (Coruzzi *et. al.*, 1984; Green *et. al.*, 1987) called the GT motif. The 13 bp motif in this region, found in a number of *rbcS* genes from pea, is completely conserved between *rbcS-E*9, *rbcS-3A*, *rbcS-3C* and *rbcS-SS*3.6. These sequences are highly similar to the enhancer motifs of SV40 and adenovirus (Kuhlemeier *et. al.*, 1987b).

Using the enhancer element from the CaMV 35S promoter and the CAT coding region (Kuhlemeier *et al.*, 1987b), a light regulatory element (LRE) was identi-

fied in a 58 bp fragment (from -168 to -110), which contains two sequence boxes conserved among members of the pea *rbcS* gene family. However, this activity could not be observed without the CaMV enhancer element, although a *rubisco* promoter fragment from -189 to -50 could mediate light induction of the CAT gene, suggesting there may be quantitative elements responsible for the level of expression in this larger fragment. Additional LREs have been indicated further upstream which allow light-induction in a tissue-specific manner when the conserved boxes in the -150 region are deleted (Kuhlemeier *et al.*, 1987a). These authors have also shown binding of the upstream region of the *rbc-3A* gene, but not the CaMV 35S promoter, to one or more protein factors present in nuclei from mature pea leaves (Green *et al.*, 1987). In DNA foot-printing experiments two protected sites have been identified at -140 and -220 which contain the GT motif (Coruzzi *et al.*, 1984). The protein factor which binds to these sites (also called boxes II and III) has been designated GT-1 (Green *et al.*, 1987). In experiments in which these boxes were mutated, mutation of a single element was insufficient to produce a profound effect on expression (Kuhlemeier *et al.*, 1988).

Another conserved element, termed the G-box, which shows dyad symmetry and an inverted repeat, has been identified in the 5′ region of 14 different *rbcS* genes from pea, tomato, tobacco, petunia, soybean and arabidopsis (Giuliano *et. al.*, 1988). This sequence is recognised by a nuclear factor obtained from tomato and *Arabidopsis thaliana* seedlings, designated the G box binding factor (GBF). The protein factor is absent in root extracts but is present in extracts from dark-adapted plants (Giuliano *et. al.*, 1988). The ABA response element in the promoter of the wheat Em gene also contains an element conforming to the G-box motif (Guiltinan *et. al.*, 1990).

The regulation of the chlorophyll a/b binding protein (*cab*) gene expression, which shows a similar pattern of regulation to the small subunit of *rubisco* has also been studied. Although, expression of a wheat *rbcS* gene could not be detected when transfered to tobacco (Keith and Chua, 1986), a wheat *cab* gene

showed light-regulated and organ-specific expression when used to generate transgenic tobacco and petunia (Lamppa *et al.*, 1985). An upstream fragment from -354 to -90 was shown to confer bi-directional enhancement, tissue-specificity and phytochrome control when ligated adjoining the 35S TATA box fused to the CAT gene (Kuhlemeier *et al.*, 1987b). Further analysis of promoter deletions indicated that a LRE is located between -180 and -90.

Simpson *et al.* (1986a) have shown that a 400 bp upstream fragment from the pea *cab* gene (*AB*80) was sufficient to promote light-inducible and tissue-specific expression of a fused *npt II* coding region. Sequences further upstream have been shown to increase the level of expression of homologous chimaeric constructs (Simpson *et al.*, 1985). These authors have identified a fragment from -100 to -347 which is responsible for light-inducibility when placed in both orientations upstream of the homologous promoter, or the constitutively expressed *nos* promoter, fused to the *npt II* coding sequence (Simpson *et al.*, 1986b). In the same constructs, this fragment was shown to have a silencer effect on the *nos* promoter, being sufficient to switch off expression of the formerly constitutive promoter in plant roots. Additionally, duplicate copies of this enhancer-like element were shown to have an additive effect on the level of induction by light.

Another light regulated gene that has been studied is phosphoenolpyruvate carboxylase from maize which accounts for about 10% of total soluble protein in green leaves. The expression of the gene is mediated by phytochrome. Sequencing of the 5′ flanking region by Matsuoka and Minami (1989) has revealed a number of putative regulatory sequences including a GC box (Dyan and Tjian, 1985), six long GC-rich directly repeated sequences and an LRE (Grob and Stuber, 1987). Later work has shown these GC-rich repeats to be a binding site for a protein factor found in nuclear extracts from green leaves but not etiolated leaves or roots (Kano-Murakami *et. al.*, 1991). This factor, designated PEP-I, also binds a synthetic oligonucleotide, based on the consensus sequence of these repeats (CCCTCTCCACATCC). Therefore it has been proposed that this factor

may be involved in the light regulation and tissue specificity.

1.4.3.2 Heat shock protein genes

Heat-shock proteins are a set of specific polypeptides, synthesized in response to environmental stress, which are highly conserved between animals and plants. These proteins are at least partly regulated at the transcriptional level and the characteristic heat-shock elements (HSE) found in the promoter region of animal heat-shock protein genes have been found in the equivalent plant genes by sequence comparison (Czarnecka *et al.*, 1985; Schöffl and Baumann, 1985).

A 457 bp upstream sequence from the *Drosophila hsp*70 gene was sufficient to direct heat-regulated expression of a fused bacterial reporter gene when introduced into tobacco calli (Spena *et al.*, 1985) and in transgenic plants (Spena and Schell, 1987). Similarly, a construct containing 1.1 kb of upstream sequences from a maize heat shock protein gene, which is highly homologous to the *Drosophila* gene, showed thermally induced expression in transgenic petunia (Rochester *et al.*, 1986).

Similar results have been obtained with a number of soybean heat shock protein genes (Gurley *et al.*, 1986). A soybean heat shock protein gene, *hsp*17.5*E*, with 3.25 kb of upstream flanking sequence, showed the expected heat induction of the mRNA in transgenic sunflower hypocotyls and tumour tissue at a level comparable to the homologous system. However in sunflower tumours a basal level of expression was also observed with this construct at low temperatures (Gurley *et al.*, 1986). Deletion downstream of −1175 bp increased the basal level observed in tumours and a further deletion to −95 bp reduced both the basal and induced level of expression.

In a similar fashion, these authors transfered the soybean *hs*6871 gene with 1 kb of 5′ flanking sequence, which contains several copies of an element homologous to that found in *Drosophila* heat shock genes, to sunflower and showed thermo-regulated transcription, although levels were significantly reduced in tu-

mour tissue. However, Schöffl and co-workers (1986) have shown that the levels of induced native heat shock proteins in sunflower tumour tissue were also significantly reduced compared to other tissues. It has been reported that heat shock elements from soybean are sufficient, when placed upstream of a truncated 35S promoter, to drive heat-induction in transgenic tobacco of a fused bacterial reporter gene (Kuhlemeier *et al.*, 1987b).

More recently heat shock protein genes from *Arabidopsis thaliana* were isolated and shown to possess multiple overlapping copies of HSE-like sequences (Takahashi and Komeda, 1989). These authors also demonstrated heat-inducible expression with 913 bp of 5′ flanking sequences from one of these genes fused to GUS when introduced into petunia.

1.4.3.3 Anaerobically-induced genes

Another stress response that has been studied in some detail is the alcohol dehydrogenase (*Adh*) genes which are induced under anaerobic conditions. Comparison of the flanking regions of *Adh* genes from maize and *Arabidopsis thaliana* has revealed that the only blocks of homology are located within the first 290 bp upstream of the transcription start site (Chang and Meyerowitz, 1986). A DNase-I hypersensitive site has also been detected at -40 to -100 under anaerobic conditions although another hypersensitive region, from -150 to -400, was present under non-induced conditions (Paul *et al.*, 1987).

In vivo genomic footprinting of *Adh*-1 from maize has revealed binding sites for possible regulatory molecules located at -100 to -108 and -186 to -190 in the induced state, and addition binding sites under normal conditions at -117 to -120 and -138 to -145 (Ferl and Nick, 1987). The factor binding at the latter site was found to alter its binding characteristics when the gene was induced. Differences were found between maize *Adh*-1 and *Adh*-2, which both possess anaerobic responsive elements (AREs). Later experiments using in vivo DMS footprinting showed that factor binding was constitutive for *Adh*-2 sequences (Paul and Ferl,

1991).

In transformed maize protoplasts, 247 bp of upstream sequence from a homologous *Adh*-1 gene is sufficient for anaerobic induction of a fused CAT gene (Howard *et al.*, 1987) and linker scanning mutants have identified a 40 bp region capable of promoting anaerobic regulation of a heterologous promoter (Walker *et al.*, 1987). However, a heterologous enhancer is required for induction of the *Adh*-1/CAT fusion in transgenic tobacco (Ellis *et al.*, 1987a).

1.4.3.4 Other inducible genes

The interaction of bacterial and plant genes in nodule formation has also been studied in depth and early- and late-induced *nodulin* genes identified in response to *Rhizobium* infection. Potential regulatory elements have been noted in the 5′ flanking region of three soybean genes, *nodulin*-23, *nodulin*-24 and leghaemoglobin (Lbc_3), induced early in infection (Mauro *et al.*, 1985). Three conserved regions were revealed by sequence comparison, one of which, an octanucleotide (GTTTCCT) was 100% homologous.

Deletion analysis of the 5′ and 3′ flanking regions of the soybean leghaemoglobin Lbc_3 gene fused to CAT in transgenic *Lotus corniculatus* plants has shown a positive regulator element at -1100 to -950 and a weaker element at -230 to -170 (Stougaard *et al.*, 1987). Nodule specific expression was confered by 37 bp located between -139 and -120 but deletions downstream of the gene did not identify any essential elements.

Jensen *et. al.* (1988) identified a nuclear factor (now designated NAT2) isolated from soybean nodules that recognizes two AT rich sequences in the promoter of leghaemoglobin Lbc_3 gene. Binding sites for this factor has also been found in the promoters from leghaemoglobin genes of the stem nodulating legume, *Sesbania rostrata* (Metz *et. al.*, 1988) and the soybean nodulin gene N23 (Jacobsen *et. al.*, 1990). This more recent work has identified additional factors that bind AT rich sequences in the promoter region of the soybean nodulin N23 gene, NAT2

from roots and nodules and LAT1 in young leaves (Jacobsen *et. al.*, 1990).

Another type of inducible plant gene that has been studied is those activated by wounding. Induced expression of a potato proteinase inhibitor II gene has been observed in transgenic tobacco plants, on wounding or following treatment of detached leaves with oligosaccharides (Sanchez-Serrano *et al.*, 1987). The induced gene expression was systemic, involving non-wounded leaves, stem and roots of the transgenic plants. This wound-inducible gene from potato, which is homologous to that found in tomato, was reported to require both 3′ and 5′ flanking sequences to confer wound induced expression of a fused CAT gene in transgenic tobacco (Thornburg *et al.*, 1987), but more recent work has excluded the involvement of 3′ sequences (Keil *et. al.*, 1990). Promoter deletions of this gene fused to CAT indicates that wound-inducibility requires sequences between -700 and -514 (Keil *et. al.*, 1990). High levels of wound-inducible expression were observed when additional upstream sequences were present. Orientation-independent expression on wounding was observed with the promoter fragment -1300 to -195 fused to a truncated CaMV promoter (Keil *et. al.*, 1990).

1.4.4 Tissue-specific plant genes

The sequences responsible for the tissue-specific expression of the small subunit of ribulose bisphosphate carboxylase and chlorophyll a/b binding protein genes have already been described (section 1.4.3). Leaf, stem and tuber-specific genes from potato have been studied in both heterologous and homologous tissues. Other obvious candidates for study are the storage proteins which are synthesized in the seed, in either the endosperm or embryo tissues. Tissue-specific expression of introduced genes was demonstrated by using a 17 kb fragment of soybean genomic DNA, containing a lectin gene and at least four nonseed genes, to transform tobacco, and subsequently detecting the various gene products in their expected tissues (Okamuro *et al.*, 1986).

In potato, the ability to regenerate transformed plants from this species, has

allowed the study of patatin and other potato genes in a homologous system. A leaf and stem specific gene, *ST-LS*1, requires only the sequences from -334 to +11 to direct leaf/stem specific and light-inducible expression of a fused CAT gene (Stockhaus *et al.*, 1987). Enhancer-like properties have been attributed to an upstream fragment from *ST-LS*1 (-98 to -675) which confers organ-specific expression in a head-to-head fusion with a minimal CaMV 35S promoter and the CAT coding region, which is not expressed in the absence of this fragment. Tuber specific expression of the patatin gene is confered by a 3.8 kb upstream fragment (Rosahl *et al.* 1987; Twell and Ooms, 1987) in both tobacco and potato.

The seed storage proteins show both temporal and organ-specific gene expression, accumulating in the developing seed at a predetermined time after pollination (Higgins, 1984). Although these genes are normally highly regulated, sunflower tumour tissues allowed expression of an introduced zein gene, resulting in detectable mRNA at low levels, but no observable protein accumulation (Matzke *et al.*, 1984). Similarly, a phaseolin gene was correctly expressed and the protein product detected in sunflower tumours (Murai *et al.*, 1983). These authors compared the levels of transcription of the introduced phaseolin gene with a chimaeric construct of the phaseolin coding region fused to the octopine synthase promoter. At 0.025% of the total polyA^+ RNA, the transcription level from the phaseolin promoter was very much lower than the level of 0.5% achieved with the octopine synthase promoter fused to the phaseolin coding region.

A number of storage protein genes have been used to generate transgenic plants. In the seeds of regenerated tobacco, an introduced phaseolin gene, containing 863 bp of 5′ and 1226 bp of 3′ flanking DNA, was correctly transcribed to give comparable protein levels to that found in bean (Sengupta-Gopalan *et al.*, 1985). Phaseolin starts to accumulate 16 days after anthesis in both bean and transgenic tobacco, although the native tobacco storage proteins first appeared 9 days after anthesis. Analysis of callus and seedlings showed significantly lower levels of phaseolin protein, approximately one thousandth of that found in seed.

Tobacco storage proteins are synthesized in both embryonic and endosperm tissues, but phaseolin was found only in the embryo, reflecting the localisation in bean seeds. However, some degradation of phaseolin into smaller peptides was observed.

This phaseolin promoter fragment contains a region (-628 to -682) that binds nuclear proteins from immature bean cotyledons (Bustos *et. al.*, 1989). The sequence of the protected region identified in foot-printing studies was found to contain two inverted A/T-rich motifs. Although fusion of 0.8 kb of the phaseolin promoter to a GUS reporter gene generated the correct spatial and temporal expression in transgenic tobacco, a different pattern was seen when the 55 bp fragment (-628 to -682) was fused to a minimal CaMV 35S promoter. This isolated fragment gave the strongest expression in roots and the timing of expression was altered. More recent work by these workers has revealed different regions involved in the spatial and temporal control of expression (Bustos *et. al.*, 1991). A number of constructions were tested in both a transient assay in bean cotyledon protoplasts and transgenic tobacco. These authors defined two upstream activating sequences (UAS1 and UAS2) which direct tissue specificity. UAS1 (-295 to -109) gave seed-specific expression, confined to the cotyledons and shoot meristem, when fused to both homologous and heterologous promoters. UAS2 (-468 to -391) extended the observed expression to the hypococotyl. Temporal control was shown to be generated by two negative regulatory sequences NRS1 (-391 to -295) and NRS2 (-518 to -418) in combination with UAS1.

In contrast, transgenic petunia plants transformed with the gene coding for the alpha′ subunit of β-conglycinin from soybean showed organ-specific expression of the introduced gene but the timing followed that of the endogenous petunia proteins (Beachy *et al.*, 1985). Analysis of 5′ deletions has implicated sequences between -131 and -257 in the high level of organ-specific expression observed (Chen *et al.*, 1986). A low level of correctly regulated expression with a -159 deletion, and complete loss of activity with a -69 deletion, were also shown. The tissue-

specific region has enhancer-like properties, indicated by experiments in which this fragment was cloned at position -90 of the 35S CaMV promoter or downstream of the 3′ non-coding region of a fused CAT gene, being sufficient to confer organ-specific CAT gene expression (Chen *et al.*, 1988). This conclusion is also suggested by experiments in which constructs, containing the α' and β subunit genes in direct or convergent orientations, were used to transform petunia (Naito *et. al.*, 1988). This comparison showed the greatest β subunit gene expression when the two promoters were in a divergent orientation, suggesting an influence of the α' gene promoter proximity on the other gene.

Interestingly, when the coding region of the α' subunit gene was fused to a constitutive promoter (either the 19S or 35S promoter from cauliflower mosaic virus), a higher level of immunodetectable protein was observed in the seeds of transformed petunia compared to leaf or callus tissue. The higher level of the α' subunit polypeptide seen in seed tissues probably reflect protein stability rather than transcriptional activity in this experiment (Lawton *et al.*, 1987).

In legumin genes a 28 bp sequence has been identified (section 1.5.3) at between 100 to 150 bp upstream of the mRNA cap site. In pea, the *leg A* gene with 1.2 kb of flanking sequence is sufficient to direct seed specific expression in transgenic tobacco (Ellis *et al.*, 1988). Analysis of promoter deletions in transgenic tobacco revealed that 549 bp of upstream sequence was required for seed-specificity and temporal regulation (Shirsat *et. al.*, 1989). Quantitative elements were identified in two additional upstream fragments from -549 to -833 and -833 to -1203. Subsequent experiments demonstrated pea seed nuclear proteins binding to the -549 promoter fragment but not to a smaller fragment which contained the 'legumin ' box (Shirsat *et. al.*, 1990).

Expression of a legumin gene in transgenic tobacco has been obtained by Baümlein and co-workers (1988) with 2.7 kb of the 5′ flanking region of the *LeB4* gene from *Vicia faba.* Similarly, a 400 bp fragment from a helianthin gene, situated 322 bp upstream of the transcription start site, when fused to a truncated CaMV 35S promoter-GUS construct, enhanced GUS expression in transformed tobacco embryos (Jordano *et. al.*, 1989).

Goldberg has identified a soybean embryo DNA binding protein of 60kd, which binds to the 5′ region of a lectin and the Kunitz trypsin inhibitor gene but not non-seed genes such as leghaemoglobin. (Jofuku *et al.*, 1987). Using gel retardation and DNase I footprinting, two binding regions for this protein were identified upstream of the lectin gene (within a region from -77 to -217) with a common seven-nucleotide core motif of 5′ ATTA/TAAT 3′, which is also found in the 5′ flanking region of the Kunitz trypsin inhibitors, KTi 1 and KTi 2. These authors also reported experiments which indicated that only 0.5 kb of lectin 5′ sequence and 0.4 kb of KTi 2 5′ flanking sequences were required to program gene expression in developing tobacco seeds.

Expression of zein proteins has been achieved at a low level in tumour tissues (Goldsbrough *et al.*, 1986; Matzke *et al.*, 1984; Kuhlemeier *et al.*, 1987b). Subsequent experiments using sunflower tumour tissues or carrot protoplasts have identified an upstream sequence of a 19 kd zein gene from -337 to -125, which contains five regions that share homology with the SV40 enhancer core sequence. These sequences are required for maximal expression of a fused CAT gene, with only very low levels of transcription observed with a minimal promoter fragment containing only the CAAT and TATA boxes (Roussell *et al.*, 1988).

The 5′ flanking sequence of a B-hordein gene (Marris *et al.*, 1988), and the upstream sequences of wheat LMW and HMW glutenin genes (Colot *et al.*, 1987), are sufficient to promote endosperm-specific expression of a fused CAT gene in transgenic tobacco. A deletion series of the LMW sequence identified a region between 326 bp and 160 bp upstream of the transcription start site essential for this pattern of activity (Colot *et al.*, 1987).

Conserved sequence motifs have been identified in the 5′ region of certain gene families of monocot storage proteins. A sequence that is conserved between the prolamin genes of barley, wheat and maize located at around 300 bp upstream of the ATG codon, which has not been found in other cereal genes (Forde *et al.*, 1985), has been described. A 15 nucleotide motif, which overlaps with the

sequence described by Forde *et al.*, (1985), present in all zein genes examined (Brown *et al.*, 1986), binds a nuclear protein factor from maize endosperm (Maier *et al.*, 1987). This sequence is also found in the maize sucrose synthase gene which displays a similar pattern of expression in the endosperm (Werr *et al.*, 1985) as zein genes. Additionally, there is some preliminary evidence that protein factors from barley nuclei bind with high specificity to the 5′ flanking region of a B hordein gene (Kreis *et al.*, 1987).

In rice, the promoter (-677 to -45) of a glutelin gene has been shown to possess multiple binding sites for nuclear factors from immature rice seeds (Kim and Wu, 1990; Takaiwa and Oono, 1990), using gel retardation and DNase-I footprinting. Two of the binding sites identified contain the motif TGAGTCA which is known to bind the transcription factors *jun* and GCN4 (Curran and Franza, 1988). This binding site is also found in the -300 element of various seed storage proteins of wheat and barley (Forde *et. al.*, 1985).

A different approach to the identification of elements essential for gene expression is the comparison of conventional mutants, which do not express the gene of interest, with the wild-type. The production of mutants which has generated information in bacterial systems is not usually applicable to plants because of their much larger and often polyploid genomes, multiple gene families and ' non-functional DNA.' However, comparisons of the 5′ region of alleles from the wild-type plant and a cultivar of *Phaseolus vulgaris* showing altered levels of phytohaemaglutinin genes has revealed deletions which appear to cause the reduced levels observed (Voelker *et al.*, 1986). The sequence of four phytohaemagglutinin genes were nearly identical from -400 to 100 bp downstream of the stop codon. In one of the genes, *dlec*2, a deletion just upstream of the TATA box (-40 to -200 upstream of the ATG codon) was identified but the gene was fully active suggesting that this region is not essential. In the *Pinto* mutant the *Pdlec*2 gene had an apparent deletion in the -250 to -360 region, which in the other alleles consisted of a direct repeat of 55 nucleotides with stretches of palindromic TGCA repeats, and which may ac-

count for the weak expression of this gene during seed development. It has been demonstrated that the differences in expression of these alleles are maintained in transgenic tobacco (Voelker *et al.*, 1987).

1.5 PEA STORAGE PROTEINS

The most common seed storage proteins found in dicotyledonous plants are the globulins, defined by their solubility in salt at neutral pH (Osborne, 1924). Although most monocotyledonous plants have prolamins as their predominant seed storage protein, many also contain a globulin fraction. The globulin component of proteins found in legumes are divided into two different size classes, 7S and 12S (Danielsson, 1949). Reflecting their function as a source of nitrogen for the developing seedling, plant storage proteins show relative abundance of the amino acids, asparagine, glutamine, and arginine or proline (Spencer, 1984). Some amino acids are poorly represented and the protein of pea seeds, in common with that of many legume grains, is limited in its nutritional suitability for man and other monogastric animals by its low content of the sulphur-containing essential amino acids, methionine and cysteine (Higgins, 1984). In cereal grains the content of lysine in the seed proteins is the limiting essential amino acid (Shewry *et al.*, 1981), although high lysine maize strains have been obtained using conventional plant breeding techniques (Mertz *et al.*, 1964; Nelson *et al.*, 1965).

The two major groups of storage protein in pea are the vicilins and legumins, distinguishable by differing salt-solubility, heat coagulation of the former and sedimentation coefficients of 7S and 12S, respectively (Derbyshire *et al.*, 1976). Additionally, legumin can be distinguished by the absence of any associated carbohydrate (Gatehouse *et al.*, 1980), in contrast to the glycoprotein, vicilin (Derbyshire *et al.*, 1976). Legumins are relatively sulphur-rich, containing more methionine and cysteine than vicilins. Considerable variation in the content of sulphur-containing amino acids from legumins from different pea strains has been reported (Casey and Short, 1981). Additionally, different pea cultivars have

been shown to possess considerable variation in the ratio of vicilin to the relatively sulphur-rich legumin component, of between 0.5 - 4.0 (Croy and Gatehouse, 1985). It has therefore been considered that an increase in the proportion of legumin to vicilin would be a desirable objective in the breeding of pea seeds with increased nutritional quality.

1.5.1 Structure and synthesis of legumin proteins.

Legumin occurs in the seeds of many of the Leguminosae (Derbyshire *et al.*, 1976) including peas, broad beans (Wright and Boulter, 1974; Croy *et al.*, 1979) and soybean (Neilsen, 1984). Homologous 12S proteins have also been found in pumpkin, sunflower rapeseed, rice and oats (Allen *et. al.*, 1985; Casey *et al.*, 1986). Legumins derived from a number of leguminous plants have been shown to have a common hexameric structure, each monomer comprising an acidic and basic subunit covalently joined by a disulphide bond (Croy *et al.*, 1979).

Legumin proteins exhibit a lot of heterogeneity in both molecular weight and charge (Casey, 1979; Gatehouse *et al.*, 1980; March *et al.*, 1987). The technique of two-dimensional non-reducing/reducing gel electrophoresis in conjunction with isofocusing was used by Matta *et al.*, (1981) to show the specificity of α and β-subunit pairing. These authors estimated that *Pisum* legumin comprised at least 22 different α- and 11 different β-polypeptides. The subunit pairs have been divided into 'major' and 'minor' legumin species (Casey *et al.*, 1981; Matta *et al.*, 1981) based on abundance. The 'minor' legumins have been further subdivided into 'big' (α-subunit 39-42,000 Kda) and 'small' (α-subunit $< 25,000$ Kda) legumins (Matta *et al.*, 1981).

The acidic and basic subunits are derived from a single translation product in peas as identified by *in vitro* translation studies (Croy *et al.*, 1980) and the predicted sequence from cDNA clones (Croy *et al.*, 1982; Domoney and Casey, 1984). Similarly, legumin precursor molecules have been identified in other species including *Vicia faba* and *Glycine max* as well as in non-legumes such as oats

and rice (Croy and Gatehouse, 1985). At least three different types of legumin precursor have been identified in pea (Croy *et al.,* 1980; Spencer and Higgins, 1980) with ^{M}r of between 60,000-65,000. A minor component of the legumin precursor molecules with a ^{M}r of 80,000 has also been identified (Domoney and Casey, 1984).

The storage proteins of pea are synthesized during the early phase of seed development and are sequestered in membrane bound structures called protein bodies (Pernollet, 1978). Hurkman and Beevers (1982) have shown that the storage proteins of pea cotyledons are synthesized exclusively by membrane-bound polysomes. Other workers (Chrispeels *et al.,* 1982a) demonstrated in pulse-chase experiments that the newly synthesized storage protein accumulated transiently in the rough endoplasmic reticulum (ER) prior to transport to the Golgi and protein bodies. The subcellular localisation of the legumin mRNA was confirmed by *in situ* hybridization with biotinylated cDNA probes (Harris and Croy, 1986; Harris *et. al.,* 1989). These protein bodies are thought to be of vacuolar origin in legumes (Craig *et al.,* 1979) in contrast to those of cereals that are derived from the ER. The legumin precursors are assembled into 8S oligomers in the ER but proteolytic cleavage and final assembly of legumin probably takes place in the protein bodies approximately 1-2 hrs after their synthesis (Chrispeels *et al.,* 1982b). In lupin the proteolytic cleavage of legumin-like molecules appears to be developmentally regulated as immature cotelydons are incapable of fully processing the precursor molecules (Johnson *et al.,* 1985).

1.5.2 Genetics and regulation of legumin synthesis.

The globulin storage proteins are probably derived from two ancestral genes (Borroto and Dure, 1987), giving rise to the 7S and 12S fractions. Hence, legumin-like molecules from different species show greater homology than the different globulin fractions from the same plant species. A number of related legumin proteins are produced in the developing seed, encoded by a small multigene family

of approximately ten members. Three genetic loci (Lg-1, Lg-2 and Lg-3) were proposed based on segregation of the subunits in crossing experiments carried out by Matta and Gatehouse (1982).

In agreement with the three types of legumin precursor peptides identified (Croy *et al.*, 1980), three classes of cDNA have been indicated by mRNA hybrid-selection/translation experiments (Domoney and Casey, 1984) which are expressed at different times in the developing seed (Domoney and Casey, 1987). The hybridization patterns of these cDNAs to pea genomic DNA (Domoney and Casey, 1985) has allowed the gene copy numbers of the three classes to be determined. Using plasmids pCD43, pCD40 and pCD32, gene copy numbers of 4-6, 2-3 and 1-2, respectively, were estimated for the haploid genome. One class of legumin genes which hybridized with pCD43 sequences mapped close to the *r* locus on chromosome 7, corresponding to Lg-1 (Domoney *et al.*, 1986). The *r* locus is associated with the characteristic of wrinkled or smooth seeds, increased levels of legumin being associated with the latter (Davies, 1980). The *r* locus is also associated with other aspects of metabolism e.g. starch biosynthesis. Another class (pCD40-related genes) mapped near to the *a* locus of chromosome 1 and a third class (pCD32-related genes) also showed linkage to this locus in one cross analysed (Domoney *et al.*, 1986). The organisation of the legumin gene families in *Pisum sativum L.* is illustrated in Fig 1.1.

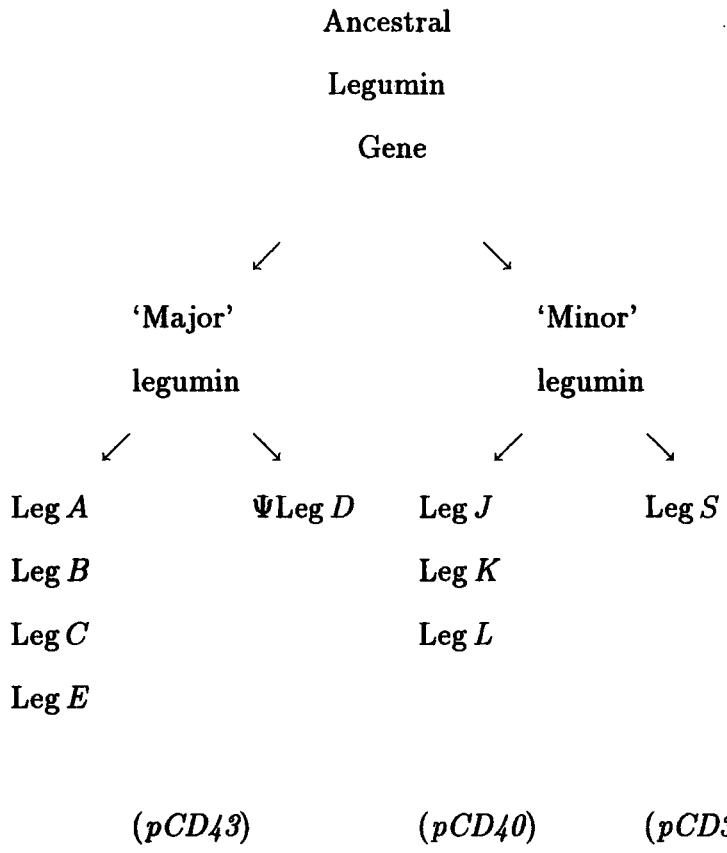
The synthesis of storage protein in pea commences 9 days after flowering (d.a.f.) for vicilin and 10 d.a.f for legumin (Millerd and Spencer, 1974; Boulter, 1981). The accumulation of these storage proteins is essentially complete 22 d.a.f. (Boulter, 1981; Gatehouse *et al.*, 1982). This synthetic activity occurs after cell division during a phase of cell expansion (Dure, 1975) coincident with endoreduplication of the DNA within the developing cotyledon (Millerd and Spencer, 1974). At this time there is an increase in the amount of mRNA transcription, but a reduction in the variety of sequences produced, resulting in a few highly expressed mRNA species (Morten *et al.*, 1983). The additional genomic DNA above

the 2C level does not appear to act as a template for RNA synthesis (Millerd and Spencer, 1974), possibly due to limitations on the available RNA polymerase II.

By Northern blot analysis of pea seed polyA^+-RNA, Gatehouse *et al.* (1982) showed that the levels of legumin and vicilin mRNA species increased and decreased in agreement with the estimated rates of synthesis of the respective polypeptides. These authors could not detect any legumin mRNA in leaf polyA^+-RNA, the limits of detection being one thousandth that found in cotyledon polyA^+-RNA, demonstrating tissue-specific expression. Evans and associates (1984a) showed by 'run-off' transcription from isolated cotyledon nuclei at various times after flowering, that the pattern of transcripts obtained reflected the synthesis of polypeptides during development. Together, these results suggest that regulation of storage protein synthesis in pea is primarily at the level of transcription. Some post-transcriptional regulation is indicated in work by Thompson *et. al.*, (1989) in which steady state mRNA levels and transcription rates of *LegA*, *LegJ* and *LegS* during cotyledon development showed little correlation. Also at later stages of development, mRNA stability may play a regulatory role because continued protein synthesis occurs despite declining mRNA transcription (Morton *et al.*, 1983). These authors presented evidence for a relatively long half-life, of >10 hrs, for some seed mRNAs (Morton *et al.*, 1983).

The levels of the various storage proteins of pea are modulated by various environmental conditions. Under conditions of sulphur deficiency the relative levels of legumin are decreased but sulphur-poor vicilin is increased (Randall *et al.*, 1979; Evans *et al.*, 1985b). This regulation of storage protein production in response to environmental sulphur levels is thought to occur at both the transcriptional and post-transcriptional (i.e. stability of mRNA) level since legumin mRNA transcription is decreased by a smaller factor than the decrease in mRNA level (Beach *et al.*, 1985; Evans *et al.*, 1985b). When levels of potassium or phosphorus are limited the levels of legumin are increased (Randall *et al.*, 1979).

Fig 1.1 Legumin gene families of *Pisum sativum L.*



1.5.3 Legumin gene structure

A number of cDNAs for legumin genes have been isolated (Croy *et al.*, 1982; Lycett *et al.*, 1984b; Casey *et al.*, 1986). Comparison of published protein sequences with a cDNA isolated from developing pea cotyledons showed that the basic subunit was located at the 3′ end of the legumin messenger RNA (Croy *et al.*, 1982; Lycett *et al.*, 1984b). These cDNAs have been used to isolate genomic clones from a λ genomic library derived from *Pisum sativum* cv. 'Feltham First' (Lycett *et al.*, 1984a).

One genomic clone λLeg 1 contains two genes; one is *leg A* which appears to be functional and another in the same orientation, approximately 1.3 kb 3′ from *leg A*, denoted Ψ*leg D*, which is presumed to be a pseudogene (Lycett *et al.*, 1984a; Bown *et al.*, 1985). The gene sequence of *leg A* shows strong homology to several previously sequenced cDNAs (Croy *et al.*, 1982; Lycett *et al.*, 1984b)

and complete homology with one particular cDNA clone, pDUB8. The predicted amino acid sequence of *leg A* agrees closely with the L3 subunit pair (Matta *et al.*, 1981) and also with the α^M-subunit described by Casey *et al.* (1981) in a different pea strain. This indicates that the gene is transcriptionally active *in vivo* as well as *in vitro*, as demonstrated by Evans *et al.* (1985a) in a heterologous expression system.

The *leg A* gene has been completely sequenced (Lycett *et al.*, 1984a) and shown to contain three introns, two within the α-subunit coding region and one within the β-subunit sequence. The boundary sequences obey the GT/AG rules (Breathnach and Chambon, 1981) and are in agreement with the plant concensus sequence identifed by Slightom *et al.* (1983). The intervening sequences are relatively short in length (88-99 bp), AT rich and at least two have been found in the *leg C* gene from pea which belongs to the same legumin gene family. Both the class I and class II glycinin genes contain three introns in analogous positions to those in *leg A*. (Fischer and Goldberg, 1982; Nielsen, 1984). The *leg J* gene sequence (Gatehouse *et al.*, 1988) has only two introns but belongs to a different legumin sub-family (Casey *et al.*, 1986). The same two-intron structure has been described by Wobus *et al.* (1986) for the type-B legumin from *Vicia faba.*

Comparison of the predicted amino acid sequence of *leg A* (Lycett *et al.*, 1984a) with the amino acid sequence determined by Casey and associates (1981) for an α-subunit indicates that a short hydrophobic region extends beyond the N-terminus of the mature protein. This, together with evidence presented by *in vitro* translation studies (Spencer and Higgins, 1980), suggests that the 21 amino acid peptide functions as a signal peptide (Blobel and Dobberstein, 1975). Similarly, *leg J* possesses a 22 amino acid leader peptide (Gatehouse *et al.*, 1988). A number of other legumin genes have been shown to possess a leader sequence: e.g. those from *Vicia faba* (Bäumlein *et al.*, 1986) and *Glycine max* (Tumer *et al.*, 1982). Leader sequences have been identified in other storage proteins of pea, one vicilin gene encodes a 15 amino acid hydrophobic signal sequence (Lycett *et*

al., 1983b) and another has a 27 amino acid leader sequence (Spencer, 1984).

The *leg A* coding region gives a protein containing 5 cysteine residues and 4 methionine residues (Lycett *et al.*, 1984a) which places *leg A* towards the top of the range for sulphur amino acid content compared to other legumins from various pea strains (Casey and Short, 1981). A higher sulphur-amino acid content is desirable, so the *leg A* gene is a good candidate for genetic manipulation. The *leg A* gene contains three direct repeats near the C-terminal region of the acidic polypeptide (Lycett *et al.*, 1984a; Evans *et al.*, 1984b). These repeated sequences have also been found in *leg C* but not in *leg J* (Casey *et al.*, 1986). The glycinin $A_5A_4B_3$ gene, which is a class II gene showing homology with *leg J*, also has three copies of a repeated sequence in this region (Momma *et al.*, 1985). There is significant homology between *leg A* and the pseudogene, Ψ*leg D*, but the latter contains in-frame stop codons, deletions and frame-shift errors and is not thought to be transcriptionally active (Bown *et al.*, 1985).

The transcription start site of *leg A*, determined by S1 nuclease mapping (Lycett *et al.*, 1984a), occurs 25 bp downstream of the TATA box within a sequence (C/TATC/A) identified at the start of other plant mRNAs (Vodkin *et al.*, 1983). The 5′ region of the *leg A* gene contains promoter elements observed in other plant genes (Messing *et al.*, 1983). Sequences showing homology with the 'TATA' box are found at position -66 and a 'CAAT' box at -126 (Lycett *et al.*, 1984a) relative to the ATG codon. Immediately upstream of the 'CAAT' box, a motif, showing partial homology with the 'AGGA' box proposed for plant genes by Messing *et al.* (1983), has been identified (Lycett *et al.*, 1984a). This overlaps a sequence conserved between all the legumin genes currently sequenced (Bäumlein *et al.*, 1986).

The 5′ sequences of a number of pea legumin genes have been compared and considerable homology observed in the region immediately upstream of the transcription start site (Lycett *et al.*, 1985). The 5′ region of *leg A*, *leg B* and *leg C* are identical for 300 bp upstream of the start codon and the pseudogene Ψ*leg D* shows

homology up to -145, although the 'CAAT' box is significantly diverged (Bown *et al.*, 1985). There is no recognisable 'CAAT' box in *leg J* although homology with a cDNA clone suggests that the gene is expressed (Gatehouse *et al.*, 1988). In *leg B* and *leg C*, an additional 550 bp upstream shows homology, including an additional 'TATA' and 'CAAT' box in this region. The sequence upstream of -900 in *leg B* and *leg C* diverges, the later contains a sequence resembling a plant insertion sequence and designated *Pis*I (Shirsat, 1988).

The genes of one legumin gene family, namely *leg A*, *leg B* and *leg C* contain a sequence 90% homologous to the SV40 enhancer core sequence (Weiher *et al.*, 1983) at -160 to -167 and another sequence 80% homologous to the adenovirus core enhancer (Hearing and Shenk, 1983) at -181. Analysis of the promotor region of a number of legumin genes from various sources has revealed a highly conserved sequence of at least 28 bp located approximately 100 bp upstream of each cap site (Bäumlein *et al.* 1986). This sequence (Fig 1.2) has been named the 'legumin' box by these authors. The 'legumin' box has been found in the region of the 'AGGA/CAAT' box in all the pea legumin genes currently sequenced, including the pseudogene Ψ*leg D*, as well as in the homologous region of the 11S genes of

Fig 1.2 Sequence conservation of legumin genes - the 'legumin' box

LeB4 -110 TCCATAGCCATGCATGCTGAAGAATGTC -80

Leg *A* GCTTCCATAGCCATGCA*A*GCTG*C*AGAATGTC

G1 AGGCTTCCATAGCCATGCAT*A*CTGAAGAATGTC

LeB4 is a functional B-type legumin gene from field bean (*Vicia faba*). Leg *A* and G1 (clone DA28-30, Goldberg, R. and Sim, T.) are A-type legumin genes. Adapted from Bäumlein *et al.*, 1986.

soybean and field bean (Bäumlein *et al.*, 1986). Homologous sequences have not been found in other plant storage protein genes including representatives of the prolamin class (e.g. zein, gliadin) or in the 7S globulins (e.g. vicilin, phaseolin). This sequence is highly specific for one class of storage proteins from divergent sources but no functional significance has been demonstrated, although a number of workers are currently investigating the problem by deletion analysis in transgenic plants.

Multiple polyadenylation signals have been identified in the 3′ region of the *leg A*, *Ψleg D* and *leg J* genes (Bown *et al.*, 1985; Casey *et al.*, 1986; Lycett *et al.*, 1984b; Gatehouse *et al.*, 1988) but these are not homologous. The *leg A* gene sequence has three sequences at the 3′ end resembling polyadenylation signals, the second of which is followed by the sequence ATTTCATGC. This is similar to the sequence found at the 3′ end of many eukaryote messages (Benoist *et al.*, 1980), and close to a site of potential secondary structure (Lycett *et al.*, 1984a). Indeed, polyadenylation is thought to occur approximately 19-20 bp downstream of the second sequence in agreement with the polyadenylated cDNAs sequenced which do not possess the third polyadenylation signal (Lycett *et al.*, 1984a).

1.6 AIMS OF THE PROJECT

The main aim of this project was the delineation of 5′ flanking sequences responsible for the tissue and stage-specific pattern of expression of the legumin *A* gene from *Pisum sativum L.*. The modular nature of eukaryote promoters may also allow the modification of constitutive promoters by isolated elements to give a new pattern of expression. Identification of a seed-specific enhancer or a non-seed silencer sequence could potentially be used to perform this function.

The *leg A* gene has been sequenced and shown to be transcriptionally active. Related legumin genes have also been sequenced allowing the identification of conserved elements which may play a role in tissue-specific expression. Additionally, the *leg A* gene belongs to the more abundant 'major ' class of legumin genes which

may indicate that the promoter is relatively more active than those of some other legumin seed protein genes.

Improvement of nutritional quality of seed proteins could be achieved by the introduction of modified or foreign seed proteins under the control of a promoter shown to be active in the seed. Altenbach *et. al.* (1989) used a methionine rich 2S seed protein gene from Brazil nut to enhance the methionine levels of transgenic tobacco. An additional copy of the 2S albumin gene (*at2S1*) of *Arabidopsis thaliana* with modifications in the large subunit was introduced and gave apparently increased levels of this 2S protein (determined by hybridization to the 3′ unmodified end) without obvious effects on other 2S proteins (Guerche *et. al.* 1990). Other workers (Dickinson *et. al.,* 1990) have increased the methionine content of the 11S protein glycinin without disrupting subunit assembly. *Leg A* has a higher methionine content than some other legume seed proteins and would be a good candidate for a similar approach.

The different *leg A* promoter fragments used in this study were selected to test the functional significance of various elements that had been identified by sequence analysis. The smallest promoter fragment (0.1 kb) tested contained the TATA and CAAT boxes and 12 bp of the ' legumin box '. Larger promoter fragments of 0.7 kb and 0.8 kb were used which, in addition to the elements of the smallest promoter fragment, contained an intact ' legumin box ' and two of three copies of an element closely related to the consensus sequence TGHAAARK found in the glutenin genes. The longest promoter fragment (1.2 kb) contained an additional glutenin-like element.

In this study a number of promoter deletions were placed upstream of the coding region of the nopaline synthase gene in pDUB1111 (Shaw *et al.,* 1986), a vector derived from pASK1029 (Shaw *et al.,* 1983), which contains pBR322 sequences and the *Hin*dIII fragment 23 of pTiC58. This vector was a product of a deletion strategy used to study the functional sequences of the nopaline synthase promoter and effects on oncogenicity (Shaw *et al.,* 1984a, b). This vector contains

the right border of the T-DNA and the coding region of the nopaline synthase gene. A unique *Eco* RI site in pDUB1111 immediately upstream of the translation start codon of *nos* allows simple insertion of *leg A* promoter fragments.

After integration of each these plasmids into a suitable *Agrobacterium* host, *Kalanchoë* plants were inoculated and the tumour tissue harvested. Analysis of this tissue for nopaline production was carried out to test for the absence of any upstream silencer sequences within the individual promoter deletions. In addition, the plasmid pBIN19 was used to construct a number of similar *leg A* promoter fusions with the nopaline synthase coding region, in order that whole plants could be regenerated and seed harvested to test the effect of the deletions on tissue-specific expression.

Nopaline synthase activity has been used as a reporter gene by a number of workers (Section 1.3). The detection of the opine product in tumours transformed with vectors carrying various promoter deletions of the *nos* gene was used to map the nopaline synthase promoter (Shaw *et al.*, 1984a). A deletion of the entire promoter region of this gene was a convenient starting point for a number of *leg A* promoter fusions in this work. When this study was initiated, the nopaline synthase assay system was well characterised, shown to be sensitive, easy to carry out and suitable vectors were available.

After the experimental portion of this study was completed similar work with the *legA* gene was published by Ellis *et. al.*, (1988). These workers transferred a 3.4 kb genomic fragment containing the *legA* gene to *Nicotiana plumbaginifolia* using the Bin19 binary vector system. Using messenger affinity paper dot blots strong hybridization to a labelled *legA* fragment was demonstrated with RNA obtained from transgenic seeds but not leaves. More detailed analysis of one transformed plant line allowed the demonstration of the introduced legumin gene protein predominantly in the embryonic tissue of the seeds. Analysis of the legumin protein by immunoblotting suggested that the protein was correctly processed in tobacco seeds. The total amount of *legA* protein produced in transgenic

tobacco seeds was only 4% of the level found in pea seeds.

This work was extended to examine the pattern of expression of a number of deletions of the *legA* promoter in transgenic tobacco (Shirsat *et. al.*, 1989). Analysis of transgenic *Nicotiana plumbaginifolia* showed that 97bp of the *legA* promoter, including the CAAT and TATA boxes, was insufficient for expression. A low level of temporally and tissue-specific expression was obtained with 549bp of 5′ flanking sequence from *legA*. Longer promoter fragments of 833 and 1203 bp gave increasing levels of expression suggesting quantitative elements within these sequences.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemical and biological reagents

All chemical reagents, with the exception of those listed below, were obtained from BDH Chemicals Ltd., Poole, Dorset, U.K., and were Analar or the best grade obtainable.

All antibiotics, adenosine 5′ -triphosphate (ATP), 6-benzyl amino purine, bovine serum albumin (BSA), deoxynucloeside 5′ -triphosphates, dithiothreitol (DTT), ethidium bromide (EtBr), lysozyme, α-naphthaleneacetic acid, nopaline, octopine, ornithine, pronase E, RNase A, salmon sperm DNA and spermidine were obtained from Sigma Chemical Co., Poole, Dorset, U.K.

Restriction endonucleases were from Bethesda Research Laboratories (U.K.) Ltd., (BRL), Cambridge, U.K., The Boehringer Corporation (London) Ltd., Lewes, East Sussex, U.K. or Northumbria Biologicals Ltd., Cramlington, Northumberland, U.K.

T4 DNA ligase, Klenow fragment (DNA Pol I), 5-dibromo-4-chloro-3- indoyl-galactoside (X-gal) were from The Boehringer Corporation (London) Ltd., Lewes, East Sussex, U.K.

[^{32}-P] α-dCTP and nick translation kit N.5000 were from Amersham International p.l.c., Amersham, Bucks, U.K.

Sephadex G-50 and Ficoll 400 were from Pharmacia Fine Chemicals, Uppsala, Sweden.

Caesium chloride, sodium chloride and 9,10-phenanthraquinone were from Koch-Light Ltd., Haverhill, Suffolk, U.K.

Nitrocellulose filters (BA85, 0.45μm) were from Scheicher and Schüll, Anderman and Co. Ltd., Kingston-upon-Thames, Surrey, U.K.

BBL trypticase peptone was from Becton Dickinson and Co., Cockeyville, M.D., U.S.A.

Bacto-agar and noble agar were from Difco Laboratories, Detroit, Michigan, U.S.A.

Yeast extract was from Sterilin Ltd., Teddington, U.K.

Agarose was from Bethesda Research Laboratories (U.K.) Ltd.,

3MM paper was from Whatman Ltd., Maidstone, Kent, U.K.

Linkers were from Bethesda Research Laboratories (U.K.) Ltd.

2.1.2 Glassware and plasticware

All glassware and plasticware was sterilized by autoclaving. Glassware used for the preparation of small amounts of DNA was siliconised by first degreasing with chloroform and then immersing in Repelcote, a solution of dichlorodimethylsilane, followed by baking at 100°C for two hours and autoclaving.

2.1.3 Bacterial strains and plasmids

Bacterial strains and plasmids are listed in Table 2.1. The following convention has been used to assign numbers to plasmids constructed by the *Durham Agrobacterium Research Group*:

pDUB1000-1049 are nopaline Ti-plasmid derivatives in *A. tumefaciens.*

pDUB1100-1149 are subcloned Ti-plasmid fragments in pBR322 in *E.coli.*

pDUB1200-1249 are Ti-plasmid subclones containing mini-Sa (pGV1106).

pBUB1300-1303 are legumin promoter fragment subclones in pUC vectors.

pDUB1500-1503 are *leg-nos* fusions in pBIN19.

Table 2.1 Bacterial strains and plasmids

Bacterial strains or plasmids	Markers	Specifications	Reference or source
E. coli			
K514	-	*thr leu thi hsd* R	Colson *et al.*, 1965
JM83	-	*ara*Δ (*lac-pro*AB) *rps*L (= *str*A) ϕ80, *lac*Z Δ M15	Bethesda Research Laboratories (BRL)
DH5α	-	F$^-$, *end*A1, λ^-, *rec* A1, *hsd* R17(r_k^-, m_k^+), *sup* E44, *thi*-1, *gyr*A496, *rel*A1, *Rec*$^+$, ϕ 80d*lac*Z Δ M15	Bethesda Research Laboratories (BRL)
A.tumefaciens			
C58C1Rif^r	Rif	a derivative of C58 cured of its pTiC58 plasmid	Van Larebeke *et al.*, 1974
C58C1Sm^{r}Sp^r	Sm/Sp	"	
C58C1Ery^{r}Cm^r	Ery, Cm	"	Holsters *et al.* 1980
LBA4404	Rif, Sm	A136 harbouring pAL4404	Hoekema *et al.* 1983

Table 2.1 Bacterial strains and plasmids (continued)

Bacterial strains or plasmids	Markers	Specifications	Reference or source
Plasmids			
pAL4404	-	A derivative of the octopine plasmid, pTiAch5, with deletion of the T-DNA region	Hoekema *et al.*, 1983
pAS11	Ap	*nos* gene in pUC8	Shirsat (Department of Botany, Durham)
pBIN19	Km	Disarmed Ti vector containing the right border, the *npt* II gene and a multiple cloning site	Bevan 1984
pDUB24	Ap	*leg A* gene in pUC8	Lycett *et al.* 1984a
pDUB1003Δ31	Km/Nm, Vir^+, Nos^+, Onc^+, Tra^+, Occ^+	Small *nos* promoter deletion in pTiC58	Shaw *et al.* 1984a
pDUB1006	Km/Nm, Vir^+, Nos^-, Onc^+ Tra^+, Occ^+	*Nos* promoter deletion in pTiC58	Shaw *et al.* 1986
pDUB1016	Vir^+, Onc^+, Tra^+, Occ^+	Gm^s Nm^s recombinant between pDUB1006 and pDUB1207	This work
pDUB1017	Vir^+, Onc^+, Tra^+, Occ^+	Nm^s recombinant between pDUB1006 and pDUB1215	This work

Table 2.1 Bacterial strains and plasmids (continued)

Plasmids	Markers	Specifications	Reference or source
pDUB1018	Vir^+, Onc^+, Tra^+, Occ^+	Gm^sNm^s recombinant between pDUB1006 and pDUB1216	This work
pDUB1019	Vir^+, Onc^+, Tra^+, Occ^+	Nm^s recombinant between pDUB1006 and pDUB1218	This work
pDUB1111	Ap	*Nos*-promoter deletion in pBR322	Shaw *et al.* 1986
pDUB1122	Ap	1.2 kb *leg A* promoter fragment from pDUB1300 in pDUB1111	This work
pDUB1128	Ap	0.7 kb *leg A* promoter fragment from pDUB1301 in pDUB1111	This work
pDUB1207	Ap Km/Gm	*Hin* dIII cut pDUB1122 ligated to *Hin* dIII cut pGV1106	This work
pDUB1215	Ap, Km/Gm, Sm/Sp	*Pst* I partial cut pDUB1122 ligated to *Pst* I cut pGV1106	This work
pDUB1216	Ap Km/Gm	*Hin* dIII cut pDUB1128 ligated to *Hin* dIII cut pGV1106	This work
pDUB1217	Ap, Km/Gm	As pDUB1215 with *Eco* RI fragment containing promoter sequences upstream of the *Pst* I site excised	This work

Table 2.1 Bacterial strains and plasmids (continued)

Plasmids	Markers	Specifications	Reference or source
pDUB1300	Ap	*Bam* HI/*Dde* I *leg A* promoter fragment from pDUB24 subcloned in pUC9 via *Eco* RI linkers	This work
pDUB1301	Ap	*Hae* III/*Eco* RI leg *A* promoter fragment from pDUB1300 subcloned in pUC9 via *Eco* RI linkers	This work
pDUB1302	Ap	pDUB1122 *Pst* I fragment containing *leg A-nos* sequences in pUC9	This work
pDUB1303	Ap	*Bgl* II/*Bam* HI *leg A* promoter fragment subcloned in pUC19	This work
pDUB1500	Km	*Bgl* II/*Stu* I fragment from pDUB1122 in pBIN19	This work
pDUB1501	Km	*Hin* dIII/*Stu* I fragment from pDUB1302 in pBIN19	This work
pDUB1502	Km	0.4 kb *Hin* dIII fragment from pDUB1303 in pDUB1500 in correct orientation	This work
pDUB1503	Km	As pDUB1502 with the *Hin*dIII fragment in the reverse orientation	This work

Table 2.1 Bacterial strains and plasmids (continued)

Plasmids	Markers	Specifications	Reference or source
pGV0601	Ap	*Hin*dIII fragment 23 of pTiC58 in a pBR322 derivative with a deleted *Bam*HI site	Shaw *et al.* 1983
pGV1106	Km, Sm/Sp	*Bgl* II-deletion mutant of Sa (MiniSa)	Leemans *et al.* 1981
pRK2013	Km	Km^r ColE1 derivative with the *tra* genes of RK2	Figurski and Helinski 1979
pRN3	Sm/Sp Tc Su	N-type plasmid	Datta and Hedges 1971
pUC9	Ap	*lac* Z	Vieira and Messing 1982
pUC19	Ap	*lac* Z	Vieira and Messing, 1982

2.2 BIOCHEMICAL TECHNIQUES

2.2.1 Purification of nucleic acids

Samples of DNA were dissolved in sterile distilled water or TE (10 mM Tris-HCl, 1mM EDTA, pH 8.0). In order to remove residual proteins from DNA preparations, or following enzymatic manipulations, the DNA sample was 'phenol-extracted' as follows. The sample was shaken with an equal volume of phenol/chloroform (phenol : chloroform : isoamyl alcohol 25:24:1, equilibrated with TE) to form an emulsion. This was centrifuged, the upper phase removed and mixed with a second volume of phenol/chloroform. After centrifugation, the upper phase was mixed with chloroform (chloroform : isoamyl alcohol 24:1) and

recentrifuged to remove traces of phenol. Alternatively, five ether extractions were carried out on the upper phase, particularly with high molecular weight DNA. The aqueous phase was recovered and ethanol precipitated. When purifying small amounts of DNA the organic phases were re-extracted with a 0.5 vol of TE.

2.2.2 Alcohol precipitation of DNA

The volume of the sample was estimated and 3M sodium acetate, pH 4.8 (0.1 vol) added, followed by cold 100% ethanol (2.5 vols). The solution was mixed and stored at $-80°C$ for 20 mins or $-20°C$ overnight. The DNA was pelleted by centrifugation at full speed for 5 mins in a microcentrifuge. Larger samples were spun in 30ml corex tubes at 10,000 rpm for 20 mins at $4°C$ using an 8 × 50 rotor in the M.S.E 18. centrifuge. The pellet was gently washed with 70% (v/v) ethanol and respun. The DNA was vacuum dried and resuspended in TE or sterile distilled water. When large volumes were to be precipitated, isopropanol (0.6 vol) was used instead of ethanol.

2.2.3 Quantitation of DNA

The concentration of small amounts of DNA was estimated by running a known volume on an agarose gel against a precise amount of a predetermined DNA sample. Larger amounts of DNA were estimated by absorbance at 260nm as measured on a Pye Unicam SP8-150 spectrophotometer, usually employed in a scanning mode from 200–300nm. A_{260} of 1 corresponds to 50μg/ml of double-stranded DNA, 40μg/ml of single-stranded DNA and RNA, and 20μg/ml for oligonucleotides; so the quantity of DNA present could be calculated. The purity of the sample could be determined by the ratio of A_{260}/A_{280} which should be 1.8 and 2.0 in an uncontaminated sample of DNA and RNA, respectively. A lower ratio might indicate contamination by protein or phenol.

2.2.4 Preparation of dialysis tubing

Dialysis tubing was cut into suitable lengths and boiled for 10 minutes in a large volume of 2% (w/v) sodium bicarbonate and 1mM EDTA. The dialysis tubing was washed thoroughly in distilled water, boiled in distilled water for 10 mins, allowed to cool and stored at $4°C$.

2.2.5 Preparation of denatured salmon sperm DNA

Denatured salmon DNA was incorporated into prehybridisation and hybridisation buffers when probing Southern blots of genomic DNA, to prevent non-specific hybridisation of labelled DNA. Salmon sperm DNA (10mg/ml) was dissolved in sterile distilled water by stirring overnight at room temperature. The DNA was sheared by extruding three times through an 18 gauge needle and denatured by heating to $100°C$ for 10 mins. The salmon sperm DNA was allowed to cool and stored in 0.5ml aliquots at $-20°C$. The aliquots of denatured salmon sperm DNA were heated in a boiling water bath for 10 mins and placed on ice immediately before use.

2.3 ENZYMATIC MODIFICATION OF DNA

2.3.1 Restriction endonuclease digestion

Plasmid DNA was cleaved with type-II restriction endonucleases in one of the four buffers recommended in Maniatis *et al.* (1982), outlined in Table 2.2. A number of enzymes were found to function at a fairly broad range of NaCl concentration, so that multiple digestions could be carried out in the same restriction buffer simultaneously. The restriction enzyme buffers were prepared at $10\times$ concentration and stored at $-20°C$.

Typically reactions consisted of $1\mu g$ of DNA digested with 5U of enzyme in a final volume of $25\mu l$. Digestions were carried out at the temperature recommended by the manufacturer, usually $37°C$ for 1–2hrs. When digesting plasmid DNA from crude minipreps or genomic DNA, 0.1M spermidine (0.1 vol) was added and the

incubation time increased. The reaction was stopped by phenol extraction or by the addition of a 0.2 vol of gel loading buffer (section 2.4).

Table 2.2. Buffers for restriction endonuclease digestion

Buffer	KCl (mM)	NaCl (mM)	Tris-HCl, pH 7.5 (mM)	$MgCl_2$ (mM)	DTT (mM)
Low	0	0	10	10	1
Medium	0	50	10	10	1
High	0	100	50	10	1
SmaI	20	0	10	10	1

2.3.2 Nick translation

Typically $0.1 - 1.0\mu g$ of DNA was labelled to $10^8 - 10^9$ cpm/μg in a reaction volume of $50\mu l$ using a method based on that described by Rigby *et al.*(1977). This was carried out using the Amersham nick-translation kit as described in its instructions. The reagents were added in this order:

Probe DNA	up to $30\mu l$
5× nick translation buffer, ($100\mu M$ dNTPs)	$10\mu l$
α-dCTP[^{32}P] ($50\mu Ci$; 125pmol)	$5\mu l$
Sterile distilled water	made up to $45\mu l$
Enzyme solution (2.5U DNA polymerase I, 50pg DNase I)	$5\mu l$

The reaction mixture was incubated at $16°C$ for 2 hrs. The ^{32}P-labelled DNA was separated from unincorporated nucleotides by chromatography using sephadex G50, packed into a 5ml disposable pipette, equilibrated and eluted with TENS (10mM Tris-HCl pH 8.0, 1mM EDTA, 100mM NaCl, 0.1% SDS). Aliquots (0.5ml) were collected and $1\mu l$ of each dispersed in scintillation fluid (3.37g PPO, 667ml toluene and 333ml Triton X-100 per litre) and the radioactivity counted using a Packard (PL Tri-Carb) liquid scintillation counter. The relevant fractions were pooled, used immediately in a hybridisation reaction, or stored at $-20°C$ for up to one week.

2.3.3 Phosphorylation of linkers

Linkers ($2-5\mu g$) were phosphorylated in kinase buffer (66mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 10mM dithiothreitol, 2mM ATP) with 10U T4 polynucleotide kinase in a total volume of $20\mu l$. The reaction mixture was incubated for 2hrs at $37°C$ and used immediately or stored at $-20°C$.

2.3.4 Filling recessed 3′ ends with Klenow fragment

In order to blunt-end ligate isolated DNA restriction fragments to linker oligonucleotides, treatment with the Klenow fragment of *E.coli* DNA polymerase was employed. The ethanol precipitated DNA was resuspended in $85\mu l$ sterile distilled water and $10\mu l$ of $10\times$ ligase buffer (660mM Tris-HCl pH7.5, 10mM EDTA, 100mM $MgCl_2$, 100mM dithiothreitol, 1mg/ml BSA and 5mM ATP). A solution containing dATP, dCTP, dTTP and dGTP was added to give a final concentration of $250\mu M$ of each nucleoside triphosphate. Finally, Klenow (9U) was added and incubated at $15-20°C$ for up to 1hr. At the end of the incubation period the reaction mixture was extracted first with phenol/chloroform, then chloroform/isoamyl alcohol and finally ethanol precipitated. The DNA pellet was recovered by centrifugation and resuspended in SDW.

2.3.5 Addition of linkers

In order to make the cohesive ends of restriction fragments compatible with the vectors used, linker adaption was employed. The linker molecules ($2\text{-}5\mu g$) were phosphorylated (section 2.3.3) and ligated to the isolated fragment or a total digest ($1\text{-}10\mu g$) which had been treated with Klenow (section 2.3.4) as described in section 2.3.8. The ligation mixture was phenol extracted, ethanol precipitated and the linkers digested with a vast excess of enzyme, usually in excess of 100U. The fragment of interest was gel purified in order to remove unincorporated linkers. Typically, the fragment was cloned into the relevant site in pUC9 to select for those molecules containing linkers. These were easily identified by the colour selection on L-agar containing the chromogenic substrate, 5-bromo-4-chloro-3-indolyl-β-D-

galactoside (X-gal) and ampicillin. The linker-adapted fragment could be easily excised and inserted into the vector of choice.

2.3.6 Phosphatase treatment of vectors

The terminal 5′ phosphates can be removed from DNA by treatment with calf intestinal alkaline phosphatase (CIP). This prevents recircularisation of the vector in the absence of insert DNA. The enzyme is inactivated by nitrilotriacetic acid (NTA), heating to $70°C$ and phenol extraction. The NTA (0.1M) solution was made by the addition of 0.19g NTA solid to 9.4ml SDW and 0.6ml of 5M NaOH to approximately give a final pH of 8.9.

The plasmid DNA (up to $15\mu g$) was digested with the appropriate restriction enzyme, phenol/chloroform extracted, ethanol precipitated and resuspended in $40\mu l$ of TE. The reaction was carried out in 50mM Tris-HCl, 0.1mM EDTA, pH 8.0 at $37°C$ with 22U CIP. After 30 mins the reaction was stopped by the addition of $4.8\mu l$ 0.1M NTA and heated to $70°C$ for 15 mins. The solution was extracted once with phenol/chloroform and the aqueous phase ethanol precipitated. The DNA was recovered and dissolved in TE to give a final concentration of approximately $0.1\mu g/ml$.

2.3.7 Ribonuclease treatment

In small scale preparations of plasmid DNA, contaminating RNA was removed by resuspending the precipitated nucleic acids in TE containing $50\mu g/ml$ RNase A. The contaminating RNA was digested during the incubation with restriction enzymes and allowed the visualization of small DNA fragments on agarose gels, which would otherwise have co-migrated with the contaminating RNA.

In order to remove contaminating DNase activity before use, RNase A was dissolved in 15mM NaCl, 10mM Tris-HCl (pH 7.5) at a concentration of 10mg/ml and heated to $100°C$ for 15 minutes. This was allowed to cool and dispensed into aliquots prior to storage at $-20°C$.

2.3.8 Ligation conditions

Typically, approximately 200ng of vector DNA, digested with an appropriate restriction enzyme, was ligated with a 10× molar excess of a DNA fragment with compatible cohesive ends. The reaction was carried out in 1× ligase buffer (section 2.3.4) with 1–2U T4 DNA ligase in up to 50μl. The reaction mixture was incubated at 4°C or 16°C for at least 4 hrs.

When ligating DNA with blunt ends, the vector and insert DNA concentration was increased by performing the ligation in a smaller volume, the final ATP concentration increased to 2mM and the reaction carried out at 4°C for at least 24hrs. Similar reactions conditions were employed to ligate kinased linkers (section 2.3.3) to DNA fragments with the exception that a vast molar excess of linker DNA (2μg) was used.

2.4 AGAROSE GEL ELECTROPHORESIS

Agarose gel electrophoresis was employed in order to identify and purify DNA fragments. For most applications 0.7% (w/v) agarose gels were used which are recommended for the separation of DNA fragments in the range 0.8–10kb (Maniatis *et al.* 1982).

The agarose was dissolved in an appropriate buffer by heating in a microwave oven for 5–8mins. Typically the agarose was dissolved in a Tris-acetate buffer (40mM Tris-acetate, 1mM EDTA pH 7.7) and cast in a 200ml perspex mould. For minigels (50ml), Tris-borate buffer (89mM Tris-borate, 89mM boric acid, 2mM EDTA) was used for its greater buffering capacity. When casting low melting point agarose gels, the mould was maintained at 4°C. Ethidium bromide was added to both buffer and gel at a concentration of 0.5μg/ml in order to visualise the DNA bands under ultraviolet light (Sharp *et al.* 1973).

Samples were mixed with a 0.2 vol of gel loading buffer (25mM Tris-HCl pH8.0, 50mM EDTA, 1% SDS, 7.5% ficoll, 0.25% bromophenol blue) and loaded into the gel slots. The gels were run at 20–30V overnight or at 100V for 4–6hrs. Minigels

were run at 100V for 1hr and were employed as a quick check for the success of digestions or ligations, before proceeding with the sample. After electrophoresis, the gels were photographed using transmitted UV light at 302nm, through a red filter, for 1sec at F8 on Polaroid film ASA 3000.

2.4.1 Recovery of DNA from agarose gels

Three methods were sucessfully employed to isolate DNA fragments from agarose gels. Typically 10μg of digested DNA was loaded into a wide gel slot and run on an agarose gel overnight. Electrophoresis at 100V for 4hrs was prefered when isolating DNA fragments smaller than 1kb.

2.4.1.1 DEAE cellulose method

Isolation of DNA fragments for nick translation were carried out using a method based on that described by Dretzen *et al.* (1981). The DNA obtained was also sucessfully ligated.

DEAE cellulose (Whatman DE-81) paper was soaked in 2.5M NaCl for several hours, rinsed several times in distilled water and finally in 1mM EDTA. The paper was dried in a vacuum oven and stored for future use.

The band of interest was located by UV transillumination and a slit cut each side. NaCl-treated DEAE cellulose paper was placed in the incisions and the gel returned to the gel tank. Electrophoresis was continued until the band had entered the paper. The paper was blotted dry on filter paper, rolled into a tube and placed in a 0.5ml eppendorf with a siliconised glass wool plug and a hole pierced in the top and bottom. Ice cold sterile distilled water (150μl) was added and spun out into a 1.5ml eppendorf. This step was repeated and then 150μl of extraction buffer (1.5M NaCl, 1mM EDTA, 20mM Tris-HCl pH 7.5) added to the paper tube. The small eppendorf was wrapped in nescofilm and incubated at 37°C for at least 2hrs. The extraction buffer was spun into a large eppendorf tube in a microcentrifuge, a second aliquot of extraction buffer added and spun

immediately into a second large eppendorf tube. The two aliquots of extraction buffer were combined and ethanol precipitated. The pellet was resuspended in 100μl TE buffer and phenol extracted, the aqueous phase ethanol precipitated and the phenol extraction repeated. Following ethanol precipitation, the DNA pellet was resuspended in 25μl TE buffer.

2.4.1.2 Low melting point agarose gels

After electrophoresis the desired fragment was excised, blotted dry on Whatman 3MM paper, and 20mM Tris HCl, 1mM EDTA, pH 8.0 (5 vols) added. This was heated to 65°C for 5 mins to melt the gel. The melted gel slice was allowed to cool to room temperature and extracted with an equal volume of phenol. The aqueous phase was recovered by centrifugation and extracted with phenol/chloroform. After centrifugation, the aqueous phase was recovered and extracted with chloroform. The DNA was recovered from the aqueous phase by ethanol precipitation.

2.4.1.3 Electroelution into dialysis tubing

This method is based on that described by McDonnell *et al.* (1977). The DNA band was located, excised and placed in a piece of dialysis tubing sealed at one end with a Medi-clip. This was filled with Tris-acetate buffer and sealed at the other end with a second Medi-clip, excluding any excess buffer.

The dialysis tubing was immersed in a shallow layer of Tris-acetate buffer and electrophoresis continued at 60–100V for about 45 mins. The polarity of the current was reversed for two minutes to release the DNA from the walls of the dialysis bag. The buffer was recovered and the dialysis tubing washed out with a small volume of Tris-acetate buffer. These were combined and the volume reduced by sequential extractions with butanol, which forms the upper phase after centrifugation. When a final volume of 100μl was obtained, the lower aqueous phase was ethanol precipitated. Two rounds of phenol extraction followed by ethanol precipitation were carried out and the DNA pellet finally redissolved in a

suitable volume of TE.

2.5 SOUTHERN BLOTTING

The verification of homologous recombination, and the subsequent resolution of introduced plasmids with the endogeous Ti-plasmid pDUB1006, was achieved by the method described by Southern (1977), in which restricted DNA is transfered to nitrocellulose filters and hybridised to ^{32}P-labelled DNA probes. The presence of introduced chimaeric *leg-nos* genes in regenerated *Nicotiana tabacum* plants was similarly identified by transfer of a digest of isolated genomic DNA to nitrocellulose filters and subsequent hybridisation with relevent gene probes (Botchan *et al.* 1976; Jeffreys and Flavell 1977).

2.5.1 Transfer of DNA to nitrocellulose filters

Agarose gels were run with digests of total bacterial or eukaryote DNA (10μg) and photographed. These were then soaked in 1% HCl for 15 mins to partially hydrolyse the DNA, in order to aid the transfer of large DNA fragments. The gel was rinsed twice in distilled water and soaked in denaturation buffer (1.5M NaCl, 0.5M NaOH) with occassional gentle agitation. After 45 mins the gel was rinsed twice in distilled water and soaked in three changes of neutralisation buffer (3M NaCl, 0.5M Tris-HCl, pH 7.0) for a total of 60 mins. The gel was then rinsed in 20× SSC (3M NaCl, 0.3M Sodium citrate) and transfered to the moistened filter paper of the blotting apparatus.

The blotting apparatus consists of a glass plate, the same width as the gel, suspended above a reservoir of 20×SSC, with a wick of Whatmans 3MM paper over the glass plate and in contact with the 20×SSC. The nitrocellulose filter, cut to the same size as the gel, was prewetted in distilled water followed by 20× SSC and placed over the gel, taking care to exclude air bubbles. One piece of Whatman 3MM paper moistened in 20× SSC, and three dry pieces, were placed on top. Absorbent cotton wool pads were placed above the filter papers and a heavy weight placed on top of the assembly.

The transfer was carried out overnight, the nitrocellulose filter then rinsed in 3× SSC and baked in a vacuum oven at 80°*C* for two hours and stored dry.

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

Labelled probes were used to locate specific sequences in filter-bound DNA from *in situ* hybridisation of bacterial colonies (section 2.9.2), total DNA from *Agrobacterium tumefaciens* (section 2.11) and total genomic DNA from transformed *Nicotiana tabacum* plants (section 2.13).

Prehybridisation and hybridisation of the nitrocellulose filter was carried out at 65°*C* in heat-sealed plastic bags, in a shaking water bath, under the conditions given in table 2.3. The filters were prehybridised for 4–16hrs and hybridised for 16–48 hrs with the labelled DNA (10^7 – 10^8 cpm). The longer incubation times were favoured when using filter-bound plant genomic DNA. The temperature and solutions used for the subsequent washing steps are also given in table 2.3. After washing the filters were dried for 2 hrs in a vacuum oven and autoradiographed.

Table 2.3 Conditions for hybridisation of ^{32}-P labelled probes

Type of filter-bound DNA	Prehybridisation buffer	Hybridisation buffer	Washing conditions
Colony hybridisation & Total *Agrobacterium* DNA	1× Denhardts 3× SSC	1× Denhardts 3× SSC 1mM EDTA 0.5% SDS	3× SSC, 0.1% SDS 4×15 mins., 65°*C* 3× SSC 4×15 mins., 65°*C*
Plant genomic DNA	Heparin sulphate (0.5g/l) Denatured salmon sperm DNA (50mg/l) Sodium pyro-phosphate (1g/l) 0.5%SDS 4×SSC	Same as prehybrid--isation buffer	0.1× SSC, 0.1% SDS 1 hr, 65°*C* Repeated at RT 0.1× SSC 30 mins, RT.

Denhardts was made as a 50× solution (10g/l BSA, 10g/l ficoll 400, 10g/l polyvinylpyrollidine) and stored at −20°*C*.

2.5.3 Autoradiography

Autoradiography was used to detect ^{32}P-labelled nucleic acids on nitrocellulose filters. The nitrocellulose filter was mounted on a glass plate, covered with clingfilm and the following procedure carried out in a dark room under a red safe-light.

A piece of X-ray film (Fuji RX) was preflashed and placed " flashed "-side down on the filter. An intensifier screen was placed on top of the film and another glass plate placed on top to complete the sandwich. This assembly was secured by elastic bands, placed inside a number of black plastic bags, within a light-tight box.

The autoradiograph was stored at $-70°C$ for at least an hour, and up to a month if there were no detectable counts on the filter. The X-ray film was developed in Kodak X-Omat developer for 5mins, rinsed in water and immersed in Kodak fixer for 5mins, rinsed and air-dried.

2.6 GROWTH AND MAINTENANCE OF BACTERIAL STRAINS

2.6.1 Media and antibiotics

For most applications, bacteria were grown in Luria broth containing suitable antibiotics as described by Miller (1972). For conjugations between strains of *Agrobacterium tumefaciens*, the bacteria were grown on minimal media (Miller, 1972). Selection of opine catabolism, of Ti-plasmid containing Agrobacteria, was achieved using supplemented nitrogen free media (Bomhoff *et al.* 1976). The media were prepared as outlined in table 2.4.

Solid media was prepared by adding Bacto agar (15g/l) before autoclaving the solutions, with the exception of NO.O media in which noble agar was used. To prevent precipitation of salts, the constituents of Minimal media and NO.O media were prepared as separate concentrated sterile solutions and added to sterile distilled water or a molten agar solution immediately before use.

Table 2.4 Bacterial Media

Media	Constituents
Luria Broth (L-broth)	10g Trypticase Peptone 5g Yeast Extract 5g NaCl made up to 1 litre and adjusted to pH 7.0
Minimal media (Min A)	1 ml 20% glucose 0.1 ml 1M $MgSO_4.7H_2O$ 20 mls 5× Min *A* salts (52.5g K_2HPO_4, 22.5g KH_2PO_4, 1g $(NH_4)_2\ SO_4$, 2.5g Na Citrate.$2H_2O$ per litre) 78 mls sterile distilled water
Nitrogen free media (NO.O)	0.28 ml 1M $MgSO_4.7H_2O$ 0.4 ml 22mM $CaCl_2.H_2O$ 5 mls 5× NO.O salts (51.3g K_2HPO_4, 61.75g KH_2PO_4, 0.75g NaCl per litre, adjusted to pH 7.5.) 95 ml sterile distilled water

Table 2.5 Antibiotic concentrations

Antibiotic	Concentration (μg/ml) in *E. Coli*	in *A. tumefaciens*
Ampicillin (Ap)[1]	50	–
Carbenicillin (Cb)	–	100
Chloramphenicol (Cm)	25	25
Erythromycin (Ery)	100	100
Gentamycin (Gm)	25	100
Kanamycin (Km)	50	50
Neomycin (Nm)	–	100
Rifampicin (Rif)	–	100
Spectinomycin (Sp)[2]	50	100
Streptomycin (Sm)[2]	20	300
Tetracycline (Tc)	10	2.5

1. X-gal was added dissolved in DMF, to a final concentration of 40mg/l

2. Higher concentrations of spectinomycin (200μg/ml) and streptomycin (1mg/ml) were used with minimal media.

Antibiotics were added after autoclaving when the media had cooled to 55°C, at the concentrations indicated in table 2.5. The chromogenic substrate, X-gal was added to the media to allow the detection of plasmids such as the pUC vectors (Messing *et al.*, 1983), which contains the amino terminal portion of the β-galactosidase gene, in host strains containing a defective β-galactosidase gene (eg. JM83, DH5α).

To select for the presence of a Ti-plasmid encoding for octopine catabolism, NO.O media was supplemented with octopine (100mg/l) and ornithine (2g/l).

2.6.2 Storage of bacterial strains

Bacterial strains were stored at 4°C on L-agar plates, containing relevant antibiotics and wrapped in nescofilm, for up to 4 weeks. For prolonged storage an overnight culture in L broth was mixed with an equal volume of 80% glycerol (v/v) and stored at −80°C.

2.7 BACTERIAL CONJUGATIONS

2.7.1 Conjugations between *E. coli* strains

Plasmids containing sequences from pGV1106 (e.g. pDUB1215), a broad host range vector, were mobilised from *E. coli* to *A. tumefaciens* using the *Inc*N plasmid pRN3. The helper plasmid, pRN3, cannot stably replicate in *A. tumefaciens* (Leemans *et al.*, 1981; Shaw *et al.*, 1983).

The *E.coli* strain K514, containing the pGV1106:pDUB1111 derivative, and K514 (pRN3) were each grown overnight at 37°C in L broth. An equal volume of the cultures were then mixed and an aliquot (0.1ml) spread on L agar plates. Aliquots (0.1ml) of each culture were also spread separately on L agar plates as controls.

After incubating for 4 hrs at 37°C, the bacterial lawn was removed by scraping the plates with 10mM $MgSO_4$ (3×1ml). The bacterial suspensions were diluted to 10^{-7} in 10mM $MgSO_4$ and aliquots (0.1ml) of the dilution series plated on L-agar

containing kanamycin (Km), streptomycin (Sm) and spectomycin (Sp), to select for transconjugants.

The numbers of the donor strain in the mixture were determined by plating the dilutions ($10^{-6} - 10^{-7}$) on L-Sm, Sp. The numbers of the recipient strain in the mixture were estimated by plating the dilutions ($10^{-6} - 10^{-7}$) on L-Km. The antibiotic selection was checked by plating dilutions of the strains incubated separately on L-Km, Sm, Sp plates.

Transconjugants were observed after 24 hrs at $37°C$ and checked by streaking on L agar plates containing the non-selected markers (Ap, Tc). An efficiency of transfer of pRN3 to the recipient strain, harbouring the pGV1106 derivative, of approximately 40% was observed.

A single transconjugant was purified by streaking for single colonies and used in the subsequent conjugation with *A. tumefaciens* harbouring the Ti-plasmid pDUB1006.

2.7.2 Conjugations between *E. coli* and *A. tumefaciens*

The pGV1106 derivatives were mobilised from *E. coli* (K514) containing pRN3 to the *Agrobacterium tumefaciens* strain C58C1Ery^{r}Cm^r [pDUB1006]. This allowed the incorporation of the pGV1106:pDUB1111 derivative into the T-DNA by homologous recombination with sequences derived from pDUB1111.

Overnight cultures of each strain were grown in L broth at $27°C$, equal volumes mixed and 0.1ml spread on an L agar plate. Aliquots (0.1ml) of the cultures were also spread on L-agar plates as controls for the subsequent selection on antibiotics. After 24hrs at $27°C$ the bacterial lawn was removed with 10mM $MgSO_4$ (3mls). The cell suspension was serially diluted in 10mM $MgSO_4$ and dilutions ($10° - 10^{-2}$) plated on L-Ery, Cm, Nm, Gm plates to select for the transcojugant e.g. C58C1Ery^{r}Cm^r[pDUB1006:1215]. The numbers of donor and recipient strains were determined by plating suitable dilutions on L-Km, Sm, Sp and L-Ery, Cm, Nm, respectively.

Colonies were observed after incubation at 27°C for 3-5 days. Transconjugates were isolated at an average frequency of 1.9×10^{-6}.

2.7.3 Conjugations between *A. tumefaciens* strains

Selection for homologous recombination between the pGV1106 derivative and pDUB1006 in C58C1Ery^{r}Cm^r was achieved by mobilisation to a cured strain of *Agrobacterium tumefaciens*, to distinguish between bacteria harbouring the two plasmids as separate entities. In order to induce the *tra* functions, which are negatively regulated, the strains were grown on minimal media. The cured strain used was typically C58C1Rif^r and the cointegrated form was selected by growth on L-agar containing rifampicin, gentamycin and neomycin.

A single colony from each strain was used to inoculate Min A (10ml) and incubated for up to 48 hrs at 27°C. The cultures were mixed and aliquots (0.1ml) plated on Min A and incubated at 27°C. The individual cultures were plated on Min A as controls for the antibiotic selection. After 24 hrs the bacterial lawn was removed with 10mM $MgSO_4$ and serially diluted. The cured strain, C58C1Rif^r, harbouring the cointegrate between the pGV1106 derivative and pDUB1006 was selected by plating on L-Rif, Gm, Nm. Colonies harbouring the cointegrate appeared after 3-5days at 27°C.

The efficiency of transfer of all Ti plasmids, determined by selection on L-Rif, Nm was approximately 4%. Of these 1.9×10^{-4} were the cointegrate formed by a single crossover event between the two plasmids. The presence of the cointegrate was verified by Southern blotting (section 2.5) of total DNA (section 2.11) from Rif^{r}Gm^{r}Nm^r colonies.

Selection for a second crossover event was achieved by transmission of the cointegrate to the cured strain C58C1Sm^{r}Sp^r and selection for spectinomycin/streptomycin-resistant, octopine catabolising (Occ^+) transconjugants which displayed the genotype, Nm^sGm^s (Shaw *et al.*, 1983). The spectinomycin/streptomycin resistance encoded by pGV1106 had previously been eliminated by the

ligation of the vector containing the chimaeric *leg-nos* gene within the *Hind III* site.

The conjugation was carried out as described previously but the selection on solid media was carried out after two or three enrichment steps in liquid culture containing NO.O media, supplemented with octopine, ornithine, spectinomycin and streptomycin. Each enrichment step was carried out for 48 hrs at $27°C$ and finally aliquots of a dilution series plated on NO.O agar containing octopine, ornithine, spectinomycin and streptomycin. The colonies obtained were screened for neomycin and gentamycin sensitivity.

2.7.4 Triparental mating

Bin 19 derivatives (Bevan, 1984), containing a variety of chimaeric *leg-nos* constructs, were mobilised from *E. coli* JM83 to the *Agrobacterium tumefaciens* strain LBA4404 (pAL4404) (Hoekema *et al.* 1983) using HB101 (pRK2013). This was achieved during a triparental mating of the strains, as described by Ditta *et al.* (1980). Although pRK2013 confers kanamycin resistance, it is not stably maintained in *A. tumefaciens.* Selection for kanamycin resistant *Agrobacteria* on minimal media as described by Bevan (1984) proved unpredictable, so selection on L-Rif, Km, Sm was employed.

Each strain was grown overnight in L-broth, mixed, plated on L-agar and incubated for 24 hrs at $27°C$. The bacteria were resuspended in 10 mM $MgSO_4$, diluted and LBA4404 (pAL4404) harbouring the Bin 19 derivative selected by plating on L-Rif, Km, Sm. Single colonies were streaked on L-agar or Min A using Rif ($50\mu g/ml$), Km ($50\mu g/ml$) and Sm (1mg/ml) before use in subsequent plant transformation steps. The prescence of the Bin 19 derivative was verified by probing southern blots of total bacterial DNA.

2.8 TRANSFORMATION OF *E. COLI* BY PLASMID DNA

2.8.1 Preparation of competent cells

Competent cells were prepared using calcium chloride as described by Dagert and Ehrlich (1979). A washing step in NaCl, and a relatively low calcium chloride concentration, were employed to enable prolonged storage of the competent cells at $-80°C$.

Briefly, *E.coli* were grown in L-broth (50ml) to an OD_{650} of 0.2, the cells spun down at 6,000g for 5 mins at $4°C$, resuspended in ice-cold 10mM NaCl (25 mls) and respun. The pellet was resuspended in ice-cold 30mM $CaCl_2$ (25 mls) and recentrifuged. The bacteria were resuspended in a second aliquot of 30mM $CaCl_2$ and left on ice for 20 mins before the cells were spun down. Finally, the bacterial pellet was resuspended in 4 mls of 30mM $CaCl_2$, 15% glycerol (v/v), and stored in aliquots ($200\mu l$) at $-80°C$ until required.

2.8.2 Transformation

Competent cells were thawed slowly on ice and 100ng of ligated DNA, diluted at least 1:2 in 30mM $CaCl_2$, added. These were left on ice for at least 30 mins, then heat-shocked at $42°C$ for 2 mins, placed briefly on ice and L-broth (1 ml) added. This was incubated, shaking at $37°C$ for one hour to allow the transformed cells to express the relevant antibiotic resistance and then aliquots ($100\mu l$) plated on L-agar containing suitable antibiotics. The plates were incubated at $37°C$ for 16–20 hrs.

2.9 IDENTIFICATION OF RECOMBINANT CLONES

When the pUC vectors were used the recombinant plasmids were identified by the appearance of white colonies when plated on L-agar plates containing X-gal. The phosphatase treatment of the vector molecules prevented recircularisation of the vector without the presence of the $5'$ phosphate groups on the insert DNA, so most colonies resulting from the transformation were recombinant plasmids and

were verified by small-scale isolation of the plasmid DNA. With these exceptions the recombinant plasmids were identified by *in situ* hybridisation of the plasmid DNA to nitrocellulose filters, followed by hybridisation of a ^{32}P labelled probe, complementary to the insert DNA.

2.9.1 Small scale isolation of plasmid DNA

Plasmid DNA was obtained from transformed *E. coli* by alkali lysis as described by D. Ish-Horowitz (Maniatis *et al.*, 1982). This is a modification of the method of Birnboim and Doly (1979). Briefly a 1.5ml aliquot of an overnight culture was spun down in an eppendorf and the pellet vortexed with $100\mu l$ of an ice-cold solution of 50mM glucose, 10mM EDTA, 25mM Tris-HCl, pH 8.0. After 5 mins at room temperature a freshly prepared solution of 0.2M NaOH, 1%SDS ($200\mu l$) was added, mixed by inverting gently and stored on ice. After 5 mins, an ice-cold solution of 5M potassium acetate, pH 4.8 ($150\mu l$) was added, vortexed vigorously and stored on ice for 5 mins. The precipitated proteins were removed by centrifugation for 5 mins and then the supernatant extracted with an equal volume of phenol/chloroform. The plasmid DNA was precipitated with ethanol (2 vols) at room temperature for 2 mins and pelleted in a microcentrifuge. The pellet was drained on tissue, washed with 70% ethanol (v/v) and dried under vacuum. The pellet was resuspended in TE ($50\mu l$) containing DNase-free pancreatic RNase. A yield of at least $3\mu g$ plasmid DNA was routinely achieved.

2.9.2 *In situ* hybridization of bacterial colonies

After transformation the resulting colonies were picked onto fresh plates in a regular grid formation for easy subsequent identification. Nitrocellulose filters (Schleicher & Schüll BA85/3, 85mm diameter) were placed gently on the agar surface, the orientation of the filter marked and left for 5 mins. The nitrocellulose filters were removed and the agar plates returned to the incubator in order that the colonies could regrow.

The filters were placed with the adhering colonies uppermost onto a succession

of pre-wetted Whatman 3MM filter papers. The filters were moistened with the following solutions and left for the times indicated:

1.)	0.5N NaOH, 1.5M NaCl	7 mins.
2.)	1.0M Tris-HCl pH 7.6	2 mins.
3.)	1.0M Tris-HCl pH 7.6	2 mins.
4.)	1.5M NaCl, 0.5M Tris-HCl pH 7.6	4 mins.

The filters were then placed on a piece of dry Whatman 3MM paper and allowed to dry at room temperature. After approximately 15 mins the filters were placed sequentially onto Whatman 3MM soaked in 20%, 40%, and 70% (v/v) ethanol and then air-dried on a dry filter paper. Finally the nitrocellulose filters were baked in a vacuum oven at 80°C for two hours. The filters were probed with ^{32}P-labelled probes complementary to the insert DNA (section 2.5.2.).

2.10 LARGE-SCALE ISOLATION OF PLASMID DNA

The large-scale isolation of plasmid DNA was carried out by essentially the same method described in section 2.9.1, except the initial buffer was supplemented with lysozyme.

A single colony was used to inoculate 5ml of L broth and grown at 37°C, shaking overnight. The overnight culture (1ml) was used to inoculate 500ml of L broth and incubated for 6hrs shaking at 37°C. Suitable plasmids were amplified overnight by the addition of chloroamphenicol (170μg/l) to the culture. The cells were harvested by centrifugation at 6,000rpm in a 6 × 500 rotor in the Sorvall RC-5B at 4°C.

The bacterial pellet was resuspended in 10ml of cold 50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA, 5mg/ml lysozyme and stored at room temperature for 10 mins. The suspension was transfered to 50ml Sorvall tubes and 20ml of 0.2N NaOH, 1%SDS added, mixed gently and stored on ice for 5–10 mins. An ice-cold solution of 5M potassium acetate, pH 4.8 (15ml) was added, vortexed and

the cell debris spun down, after 10 mins on ice, by centrifugation at 10,000 rpm for 20 mins. The supernatant was carefully decanted into two 30 ml corex tubes, a 0.6 vol ($\approx$ 12ml) of isopropanol added, mixed and left for at least 15 mins. at room temperature. The DNA was recovered by centrifugation at 10,000 rpm for 30 mins.in a Sorvall HB4 rotor at room temperature. The pellet was allowed to drain, washed in 70% ethanol (v/v) and resuspended in TE.

In order to set up caesium chloride gradients, the volume of the DNA solution was measured. Solid CsCl (1g/ml) and ethidium bromide (600μg/ml) were added and the density checked, and if necessary, modified, to give a final density of 1.55g/ml. Occassionally, a clearing spin was carried out, in order to remove precipitated proteins, by centrifugation at 10,000 rpm for 20 mins at room temperature. The DNA solution was transfered to quick-seal Beckman tubes, placed in a Vti 50 vertical Beckman rotor and centrifuged for 24–36 hrs in a Sorvall OTD 65 Ultracentrifuge at 44,000 rpm at 15°C.

The DNA bands were visualised under UV light and the lower supercoiled band recovered. The ethidium bromide was removed by sequential extractions with CsCl-saturated isopropanol. The plasmid DNA was then dialysed against three changes of TE at 4°C to remove the CsCl. The DNA was either stored directly or concentrated by ethanol precipitation.

2.11 PREPARATION OF TOTAL BACTERIAL DNA

Putative recombinant Ti-plasmids were screened by Southern blotting of total *A. tumefaciens* DNA as described by Dhaese *et al.* 1979. An overnight culture (1.5ml) was centrifuged, the supernatant discarded and the pellet resuspended in 400μl of 50mM Tris-HCl (pH 8.0), 20mM EDTA, 0.8% sodium lauroyl sarcosinate, 1mg/ml pronase. This was incubated at 37°C for at least 1 hr and then sheared by pipetting through a glass pasteur three times. The proteinaceous material was removed by sequential phenol extraction and ethanol precipitation. The nucleic acids were finally resuspended in TE.

2.12 TRANSFORMATION OF PLANT TISSUE

2.12.1 Transformation of Plants with oncogenic Ti-plasmids

Tumours were induced on stems and leaves of *Kalanchoëdaigremontiana* by wounding with a sterile syringe needle, immediately followed by inoculation of the wound site with a fresh overnight culture of *A. tumefaciens* in L broth. Tumours were assayed for nopaline after 4–6 weeks.

Tumour tissue was excised from axenically grown plants and transfered to hormone-free Murashige and Skoog medium (Murashige and Skoog, 1962) containing carbenicillin (1mg/ml) and sucrose (30g/l). The tumour tissue was transfered every three weeks.

2.12.2 Transformation of *Nicotiana tabacum* with Bin 19 vectors

In order to maintain sterility all manipulations described were carried out in a laminar flow hood.

The presence of a Bin 19 derivative in LBA 4404 was confirmed by streaking the cells on L-Rif, Km, Sm. Single colonies were used to inoculate L broth containing $50\mu g/ml$ kanamycin and grown shaking at $27°C$ for 48 hrs. The overnight culture (1ml) was used as the inoculant for 50mls of L broth($50\mu g/ml$ kanamycin) which was grown at $27°C$ for 24hrs and harvested by centrifugation. The pellet was washed by resuspending in an equal volume of 2mM $MgSO_4$ and respun. This step was repeated twice and the bacteria finally resuspended in twice the volume of Murashige and Skoog medium (Murashige and Skoog, 1962) containing 10g/l sucrose.

Tissue from *Nicotiana tabacum* SR1 leaves was surface sterilised in the following way. The cut petiole from a fully expanded leaf was sealed with wax, immersed in 70% (v/v) ethanol for 30 secs, washed in sterile distilled water and immersed in 5% (v/v) chloros solution with 2-3 drops of Tween 20 for at least 15 mins. The leaf was then washed three times in a large volume of sterile distilled water and

cut into pieces approximately 1 cm square.

The leaf pieces were mixed with the bacterial suspension and left for 10 mins, placed on 0.8% (w/v) agar plates containing shooting medium (Murashige and Skoog medium, 20g/l sucrose, 1mg/ml 6-benzyl aminopurine, 0.1mg/ml α-naphthaleneacetic acid, pH 5.8) and maintained at $25°C$ in a plant growth room.

After 48hrs the leaf pieces were removed and placed in liquid shooting medium with carbenicillin (1mg/ml). These were incubated, shaking, overnight, then washed briefly in shooting medium, blotted and placed on shooting medium plates containing carbenicillin ($500\mu g/ml$) and kanamycin ($200\mu g/ml$) in a plant growth room until green shootlets appeared.

The shootlets were excised and transfered to rooting medium (Murashige and Skoog medium, 20mg/ml sucrose, 0.8% agar, pH 5.8) in 60ml sterilin containers. Roots appeared after 7–10 days and after three weeks when the roots were established the plants were transfered to pots containing 50:50 perlite:soil. The plants were covered in plastic bags and allowed to stand in water for 1 week after which the plastic covering was slit. After three days the bags were removed completely and watered as required. The plants flowered after three months.

2.13 ISOLATION OF PLANT DNA

2.13.1 Large-scale isolation of plant DNA

Genomic DNA from putative transformed *Nicotiana tabacum* was isolated by a method based on that described by Graham (1977). Leaf tissue (approx 4g) was frozen in liquid nitrogen and ground to fine powder in a precooled pestle and mortar. The ground tissue was transfered to a 100ml flask and the following added:-

4ml	Homogenising buffer (100mM NaCl, 25mM EDTA, 2% SDS, 0.1% DEPC)
1ml	5M sodium perchlorate
4ml	phenol
4ml	chloroform/octanol (99:1)

This was shaken gently on ice for at least 50mins. The mixture was then transfered to a 30ml corex tube and spun at 8,000g for 5mins at $10°C$ in a 8×50 rotor in an MSE 18 centrifuge. The supernatant was removed to a fresh corex tube and shaken with an equal volume of chloroform/octanol (99:1). The tube was spun for 1 min at 8,000g, the aqueous phase removed to a fresh tube and the DNA precipitated with cold ethanol (2 vols). The DNA was recovered by spinning at 8,000g for 20 mins at $4°C$.

The pellet was resuspended by shaking with 2 mls of resuspension buffer (50mM Tris-HCl, 10mM EDTA, pH 8.0) on ice overnight. Pronase was added to a final concentration of $500\mu g/ml$ and incubated at $37°C$ for 3 hrs.The nucleic acid solution was made up to 6.8ml with resuspension buffer and 6.8g of caesium chloride added followed by ethidium bromide to a final concentration of $200\mu g/ml$. The solution was mixed thoroughly and centrifuged in a 10×10 rotor in an MSE Prepspin 65 centrifuge overnight at 44,000rpm at $15°C$. The DNA band was collected, the EthBr extracted with CsCl saturated isopropanol. The CsCl present was removed by dialysis against three changes of TE. The DNA was ethanol precipitated and resuspended in 1 ml of TE. The concentration of DNA and purity was estimated by scanning from 200–300nm.

2.13.2 Miniprep method of plant DNA isolation

The Dellaporta plant DNA miniprep technique (Dellaporta *et al.* 1983) was used to initally screen large numbers of putatively transformed tobacco plants. This method had the advantage of requiring less than a gram of tissue and allowed the DNA to be isolated in one day. The speed of isolation allowed relatively undegraded material to be retrieved but considerable RNA contamination was usual.

At least 0.5 - 0.75g of leaf tissue was weighed, transfered to a pestle and mortar, frozen in liquid nitrogen and ground to a fine powder. This was transfered to 30ml corex tube on ice and 15ml extraction buffer (100mM Tris-HCl, 50mM

EDTA, 500mM NaCl, 10mM mercapatoethanol, pH 8.0) and 1ml 20% (w/v) SDS added. The mixture was mixed thoroughly and incubated at 65°C. After 10 mins 5M potassium acetate (15ml) was added, shaken thoroughly and placed on ice for 20 mins. The mixture was spun at 8,000 rpm for 30mins in a HB-4 rotor in the Sorvall at 4°C. The supernatant was poured through three layers of muslim cloth into a 30 ml corex tube containing 10ml isopropanol. This was mixed, incubated at $-20°C$ for at least 30 mins and spun at 8,000 rpm for 20 mins. The supernatant was discarded and the pellet allowed to drain on tissue. The pellets were resuspended in 0.7 ml resuspension buffer I (50 mM Tris-HCl, 10 mM EDTA, pH 8.0), transfered to an eppendorf tube and spun in a microcentrifuge at full speed for 10 mins. The supernatant was transfered to a clean eppendorf and the DNA precipitated with 3M sodium acetate, pH 4.8 ($75\mu l$) and isopropanol(0.5 ml). The DNA was pelleted by centrifugation for 30 secs, the pellet washed in 70% (v/v) ethanol, dried and redissolved in 100–200μl resuspension buffer II (10 mM Tris-HCl, 1mM EDTA, pH8.0). Usually at least 40μg of DNA was isolated by this method.

2.14 NOPALINE ASSAYS

The presence of nopaline in a variety of transformed plant tissues was detected using a simple paper electrophoresis method based on that described by Otten and Schilperoort (1978). This method is sufficient to distinguish between nopaline and octopine and a variety of related products depending on the relative migration of the substance and the colour of the stained product. However, attempts were also made to develop more sensitive methods of nopaline detection and provide a more readily quantitative assay. These included HPLC which can detect a related product, nopalinic acid, and an ELISA technique. Unfortunately an assay based on these methods was not successfully developed in the time available.

The paper electrophoresis was carried out as follows. Approximately 100mg of plant tissue was placed in a small eppendorf tube and then crushed with a

metal rod. These were spun down for 5 mins and the supernatant removed. The supernatant (20μl) from each sample was spotted, at 1 cm intervals, onto a 30×21 cm piece of Whatman 3MM paper, 4.5 cm from the anodal side. Nopaline, octopine and arginine standards were also loaded onto the paper, together with colour markers (methyl green, orange G and xylene cyanol) to monitor the progress of the samples. The samples were allowed to dry completely. The Whatman 3MM paper was placed in a Shandon 600×100 electrophoresis apparatus, wetted gently with the running buffer (5:15:80 formic acid : acetic acid : water, pH 1.8) and run for 1–2 hrs at 400V. The paper was dried and then sprayed with a 50:50 mixture of 9,10-phenanthraquinone (20mg/ml) in ethanol and 10% (w/v) NaOH in 60% (v/v) ethanol, mixed immediately before use. The paper was dried in a fume hood and the opines visualised under a UV lamp.

CHAPTER 3

RESULTS

3.1 INTRODUCTION

A series of chimaeric *leg-nos* genes were constructed using two different vector systems. These complementary sets of vectors were designed to study promoter deletions in both callus tissue and regenerated plants.

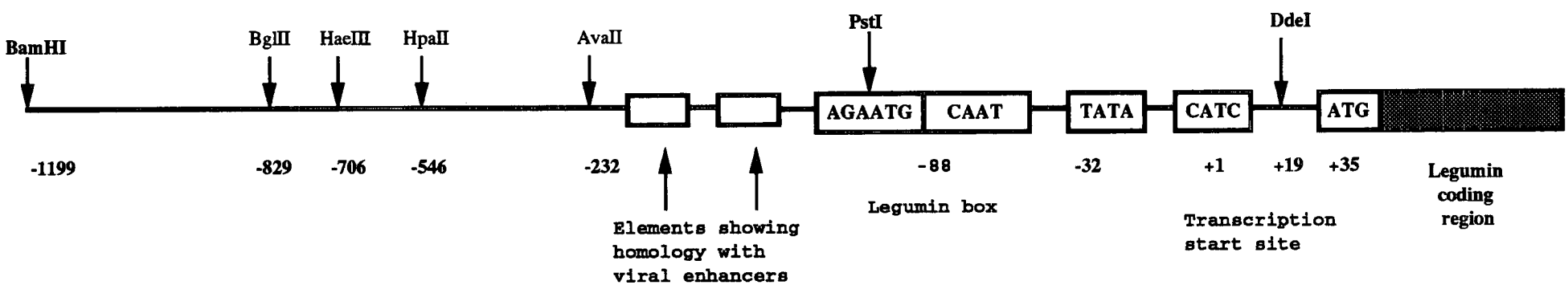
Initially sequences from the upstream region of *leg A* were combined with the nopaline synthase coding region in pDUB1111. These *leg-nos* fusions allowed the construction of a number of vectors based on the oncogenic Ti-plasmid pDUB1006. This Ti-plasmid vector has no intrinsic *nos* activity but allows the homologous recombination of the different chimaeric *leg-nos* genes, constructed in pDUB1111. These constructs were intended to address the question of potential silencer sequences in the *leg A* promoter region. It was hoped that analysis of different deletion mutants would reveal differences in levels of expression in callus tissue and therefore the sites of any silencer sequences that may be present. Although the integration of the chimaeric gene sequences required laborious conjugations and selection of the desired crossing-over event, the callus tissue could be quickly assayed for any nopaline synthase activity.

The more time-consuming regeneration of mature, flowering plants, required the synthesis of a parallel set of *leg-nos* genes in a disarmed vector. These vectors were required in order to dissect the minimum length of upstream sequence required for tissue and stage-specific expression of the *leg A* gene. The undifferentiated state of callus tissue is therefore unsuitable. The vector used was pBIN19 (see section 1.2.3), which possesses a polylinker sequence for simple insertion of the *leg-nos* fusions. Once the constructs were inserted into pBIN19 the plasmids could be rapidly transfered from the *E.coli* host to the *Agrobacterium tumefaciens* strain LBA4404 during a triparental mating. Mature plants were obtained and various tissues tested for the presence of nopaline, which would indicate nopaline synthase activity. Where time permitted, the plants were allowed to self-pollinate and seed collected for analysis.

3.1 Restriction sites upstream of the Legumin *A* gene

Restriction sites that have been used to generate suitable promoter fragments are indicated (not drawn to scale). These have been used to create a set of chimaeric genes, when fused to the nopaline synthase coding region. The location of the canonical promoter sequences of eukaryote genes that have been identified in the 5′ flanking region of the legumin *A* gene from *Pisum sativum L.* are shown, together with sequences showing homology to the adenovirus and SV40 enhancers.

FIG 3.1 Restriction sites upstream of the Legumin A gene



3.2 ANALYSIS OF PROMOTER DELETIONS IN CALLUS TISSUE

3.2.1 Strategy for construction of oncogenic Ti-plasmid vectors.

The legumin promoter sequences were contained within pDUB24, a genomic clone of *legA* in pUC8 (Lycett *et al.* 1984a), which is described in section 1.5.3. Analysis of the *leg A* sequence revealed a *Dde*I site located between the transcription start site and the coding region. This has allowed the relatively simple construction of a number of chimaeric genes, containing upstream fragments of the legumin *A* gene from *Pisum sativum L.*, ligated to the coding region of the nopaline synthase gene.

The legumin promoter sequences were inserted into pDUB1111 which contains the 25 bp repeat essential for T-DNA transfer. This vector was derived from a pBR322 derivative containing *Hin*dIII fragment 23 of pTiC58 in which sequences between -1 and -287 of the *nos* promoter had been replaced by a unique *Eco* RI site (Shaw *et. al.*, 1986). Different restriction fragments from within the 5′ flanking region of *leg A* starting from the *Dde* I site upstream to suitable restriction sites (shown in fig 3.1) were adapted with *Eco* RI linkers to allow insertion into the unique *Eco* RI site in pDUB1111. To check for the addition of the linkers the fragments were first subcloned into pUC9, which also allowed large-scale preparation of these fragments for ^{32}P-labelled probes.

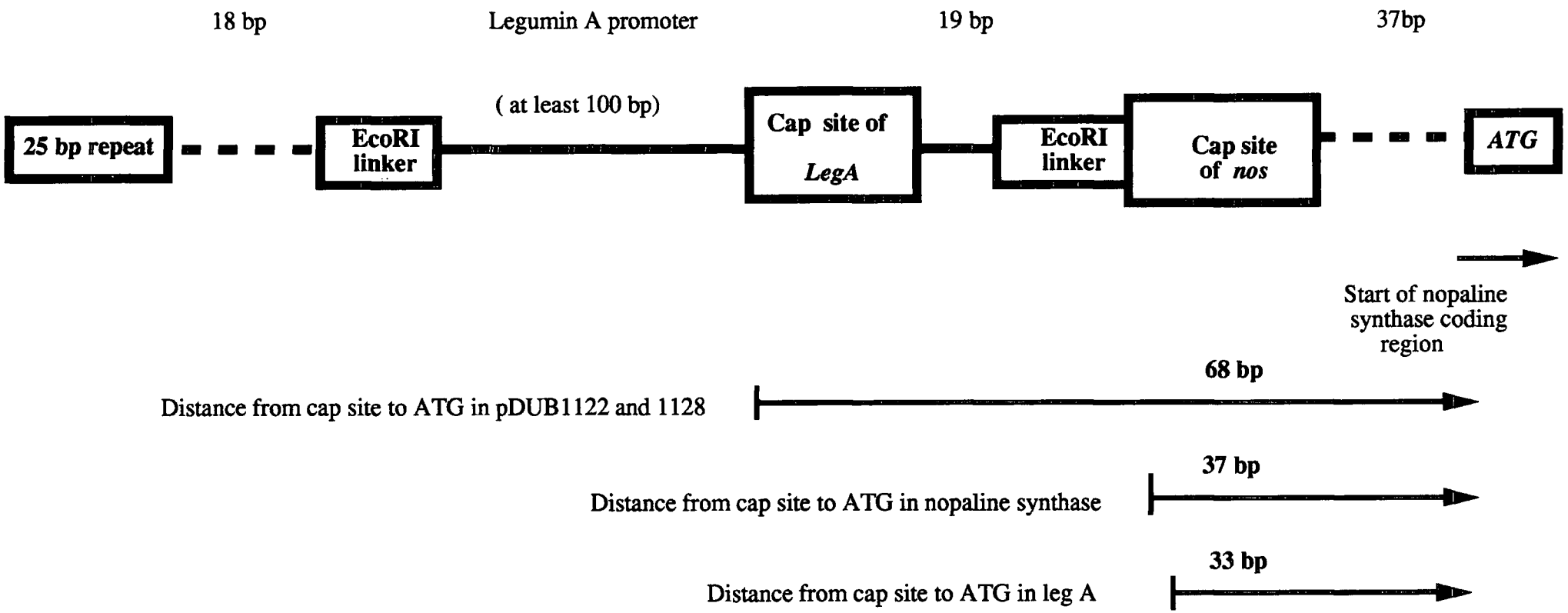
Following the insertion of the pSa derivative, pGV1106, to enable replication of these constructs in both *Agrobacterium* and *E.coli*, the plasmids were mobilised to the strain, C58C1Ery^{r}Cm^r[pDUB1006], mediated by pRN3. Therefore, the chimaeric genes could be integrated into the endogenous Ti-plasmid, pDUB1006, by homologous recombination between the *nos* coding region or right border sequences. Tumours were then induced on the leaves of *Kalanchoë daigremontiana.*

Four chimaeric *leg-nos* genes with differing lengths of 5′ legumin *A* sequences were constructed. The details of synthesis of these constructs are outlined in the following sections. The *leg-nos* fusions contain the start of transcription from

3.2 Structure of the chimaeric *Leg-nos* genes at the site of fusion

The genetic origin of the transcription and translation start signals, of the gene fusions between legumin *A* and nopaline synthase sequences, are shown. Sequences derived from pDUB1111, flanking the *leg A* promoter fragment, are indicated by a broken line. Sequences from the legumin promoter (0.1-1.2 kb) are indicated by a solid line.
Not drawn to scale.

Fig 3.2 Structure of the chimaeric leg-nos genes at the site of fusion



leg A, separated by 68 bp, and the *Eco*RI linker, from the nopaline synthase translation start codon. This results in tandem cap sites, which appears to have little effect on gene expression (Shaw *et. al.*, 1986). The arrangement of potential transcription and translation start points following fusion of sequences from the two constituent genes in pDUB1122 and pDUB1128 is illustrated in fig 3.2.

Initially, the "full-length" (1.2 kb) legumin promoter fragment was inserted into the plasmid pDUB1111 to give pDUB1122. This plasmid has formed the starting point for a number of additional *leg-nos* chimaeric genes (fig 3.3) as well as a range of homologous disarmed Ti-plasmid vectors (described in section 3.3.1). A homologous construct with a smaller, 0.7 kb promoter fragment was constructed and designated pDUB1128.

The two other plasmids were constructed to test the functional relevance of the legumin box and other sequences showing homology with the SV40 and adenovirus enhancers. A promoter fragment including the legumin transcription start site up to the *Pst* I restriction site at -95 (relative to the start of transcription) ligated to the *nos* coding region was constructed. This 0.1 kb promoter fragment retains the TATA and CAAT boxes (Fig 3.1) and the first twelve nucleotides of the conserved sequence known as the 'legumin box' (see section 1.5.3). However, a promoter fragment of < 0.1 kb was difficult to isolate from agarose gels so alternatively pGV1106 was inserted into the *Pst* I site at -95 on the *leg A* promoter fragment in pDUB1122, to give pDUB1215. This spacially separated the promoter sequences upstream of the *Pst* I site by at least 8 kb. An *Eco* RI fragment containing the sequences upstream of -95 was excised from pDUB1215 to create an additional construct, pDUB1217, in order to exclude the possibility of any enhancer function in this upstream fragment.

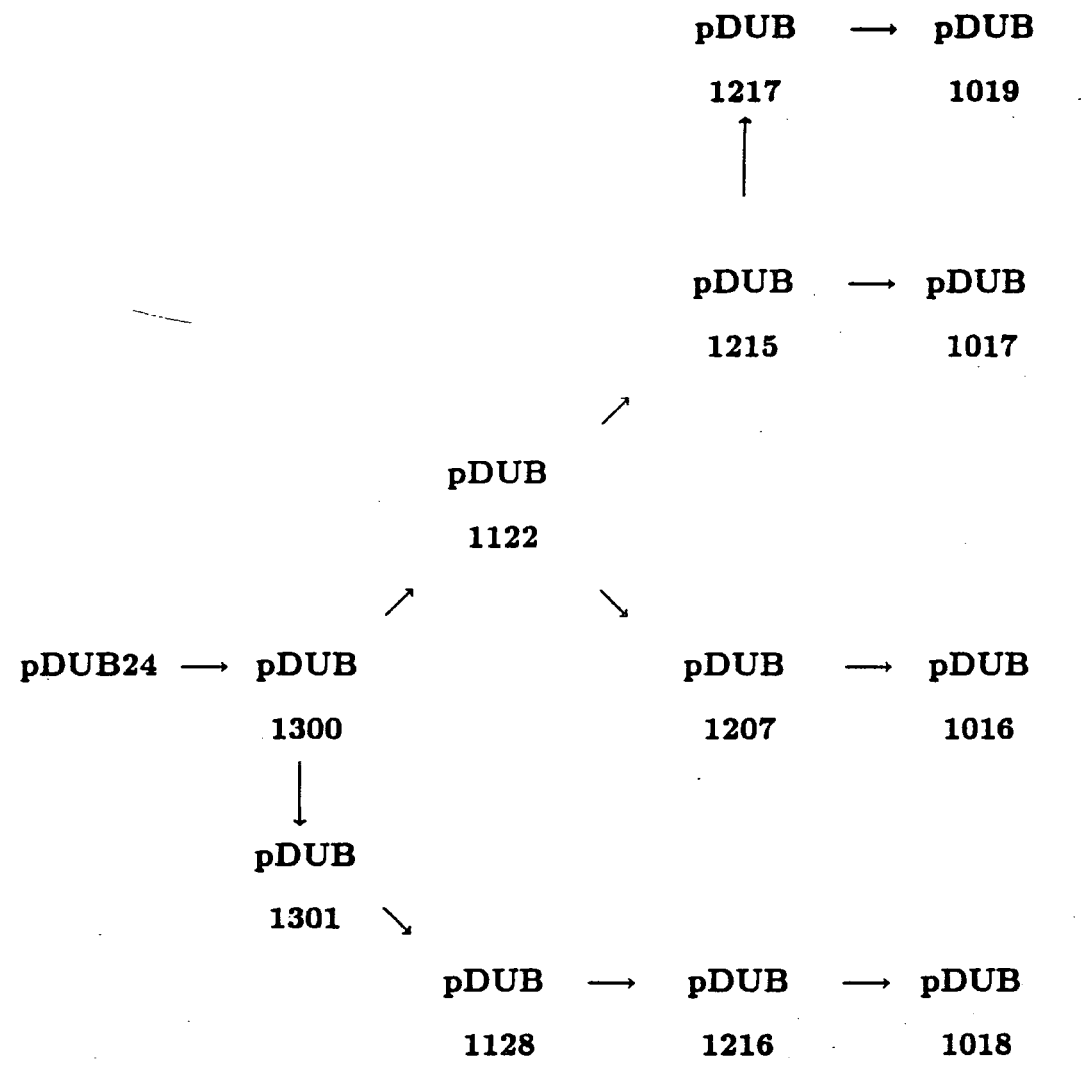
3.2.2 Construction of *legA* promoter subclones

Promoter fragments from the *leg A* gene which extend 1.2 kb (*Bam* HI site) and 0.7 kb (*Hae* III site) upstream from the *Dde* I restriction site were subcloned

3.3 Genealogy of oncogenic chimaeric *Leg-nos* constructs

The construction of oncogenic Ti-plasmid vectors containing chimaeric *Leg-nos* genes has been carried out in four steps:
i) Construction of promoter fragment subclones in pUC9; ii) insertion into the *Eco* RI site in pDUB1111; iii) insertion of pGV1106 into the pDUB1111 derivatives to create intermediate vectors; and finally, iv) recombination and resolution of cointegrates with the Ti-plasmid pDUB1006.
Plasmids pDUB1016, pDUB1122, pDUB1207 and pDUB1300 contain the full-length legumin promoter. Plasmids pDUB1018, pDUB1128, pDUB1216 and pDUB1301 contain 0.7 kb of the legumin promoter. Plasmids pDUB1019 and pDUB1217 contain 0.1 kb of the legumin promoter. Plasmids pDUB1017 and pDUB1215 contain the full-length promoter with pGV1106 inserted at -95.

Fig 3.3 Genealogy of oncogenic chimaeric leg-*nos* constructs



into pUC9 using *Eco* RI linkers to give pDUB1300 and pDUB1301, respectively (Fig 3.4). Briefly, the plasmid, pDUB24, which contains 1.2 kb of 5′ flanking sequence from the *leg A* gene was digested with *Bam* HI and *Dde* I, the ends rendered flush with Klenow and *Eco* RI linkers added (see section 2.2). Following *Eco* RI digestion, the 1.2 kb legumin promoter fragment, which spans a region from +19 to -1200 upstream of the legumin *A* gene, with its associated *Eco* RI linkers, was isolated. The isolation of the promoter fragment and removal of unbound linkers was achieved in a single step, by purification of the 1.2 kb fragment from an agarose gel (section 2.4.1), the linkers and other restriction fragments from the digested plasmid migrating much further down the gel. The 1.2 kb fragment was ligated to an *Eco* RI digest of pUC9 and colonies containing a recombinant plasmid identified on X-gal/Ap plates. The resulting plasmid was designated pDUB1300 (fig 3.4).

Similarly, pDUB1301 was constructed by the addition of *Eco* RI linkers to a *Hae* III digest of pDUB1300, followed by an *Eco* RI digestion. The 0.7 kb fragment was gel purified and inserted into the *Eco* RI site of pUC9 (fig 3.4).

3.2.3 Construction of chimaeric *leg-nos* genes in pDUB1111

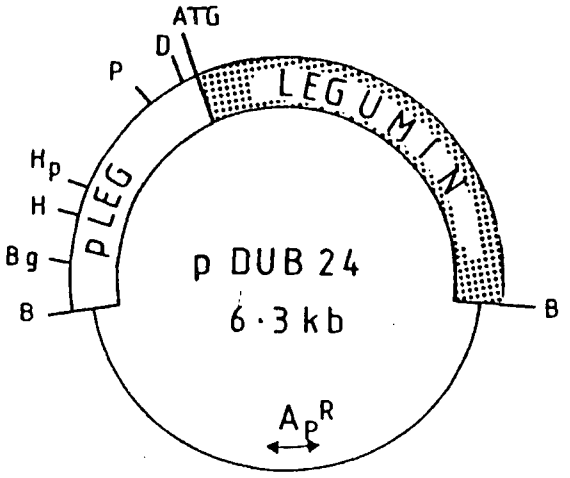
The promoter fragments from the pUC9 subclones were isolated and inserted into pDUB1111 (fig 3.5). A chimaeric *leg-nos* gene construct, pDUB1122, containing the entire length of the legumin *A* promoter sequences from pDUB24 was prepared by the insertion of the 1.2 kb *Eco* RI fragment from pDUB1300 into the single *Eco* RI site of pDUB1111. Following transformation and ampicillin selection, colonies containing the insert were identified by colony hybridisation (section 2.9.2) with a labelled 1.2 kb *Eco* RI fragment (section 2.3.2) from pDUB1300 (LEG 1.2). Of the eight positive colonies isolated, the clone containing the promoter sequences in the correct orientation relative to the nopaline synthase coding region was identified by restriction analysis (fig 3.6). Five isolates showing the correct insert and vector size by *Eco* RI digestion were digested with *Pst* I. The

3.4 Construction of *leg A* promoter subclones in pUC9

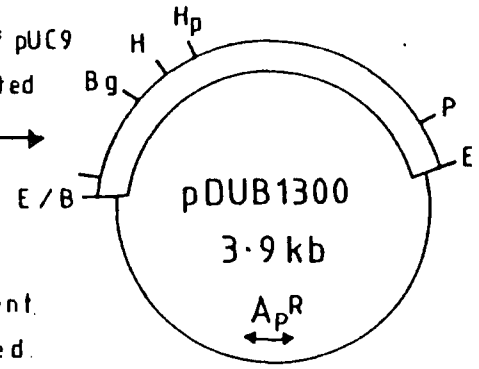
Legumin *A* promoter fragments (represented by open boxes) spanning 1.2 kb and 0.7 kb of upstream sequences were modified with *Eco RI* linkers and inserted into pUC9. The restriction sites are indicated as follows: B=*Bam* HI, Bg=*Bgl*II, D=*Dde*I, E=*Eco* RI, H=*Hae*III, Hp=*Hpa*II, P=*Pst* I. ATG signifies the position of the initiation codon of the legumin coding region. Only the HaeIII site used for the promoter manipulations is indicated.
(Not drawn to scale.)

Fig 3·4

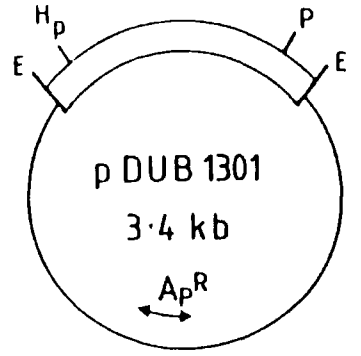
Manipulations in E. Coli



1. DdeI / Bam HI digest
2. Addition of EcoRI linkers.
3. EcoRI digest.
4. Gel purification of 1·2 kb fragment.
5. Ligation to pUC9 EcoRI digested.



1. Hae III
2. Addition of EcoRI linkers
3. EcoRI digest
4. Gel purification of 0·7kb fragment
5. Ligation of pUC9 EcoRI digested



Legumin A gene subcloned in pUC 9

Legumin A promoter fragments subcloned in pUC 9

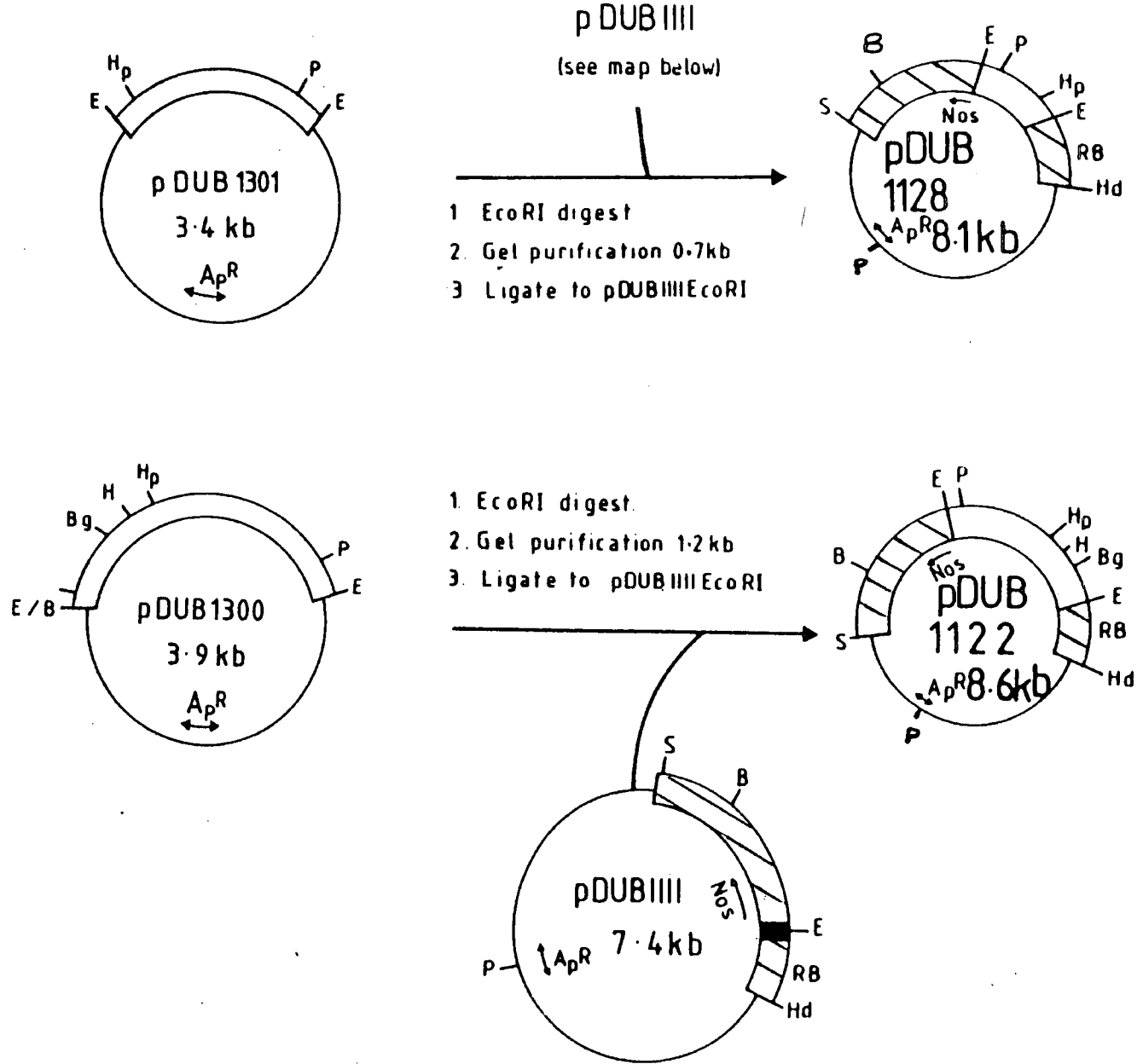
3.5 Construction of chimaeric *Leg-nos* genes in pDUB1111

The promoter fragments subcloned into pUC9 were reisolated and inserted into the *Eco* RI site immediately upstream of the nopaline synthase translation start codon, to generate pDUB1122 (LEG 1.2) and pDUB1128 (LEG 0.7). Sequences homologous to the Ti-plasmid vector, pDUB1006, namely the right border repeat and the nopaline synthase coding region (*Hin*dIII fragment 23 from pTiC58), are indicated by hatched boxes. Hd=*Hin*dIII and RB=right border sequences. Other symbols used are the same as Fig 3.4.

Only the HaeIII site used for the promoter manipulations is shown, similarly other restriction sites shown are limited to those used in the plasmid constructions.

(Not drawn to scale).

Fig 3·5



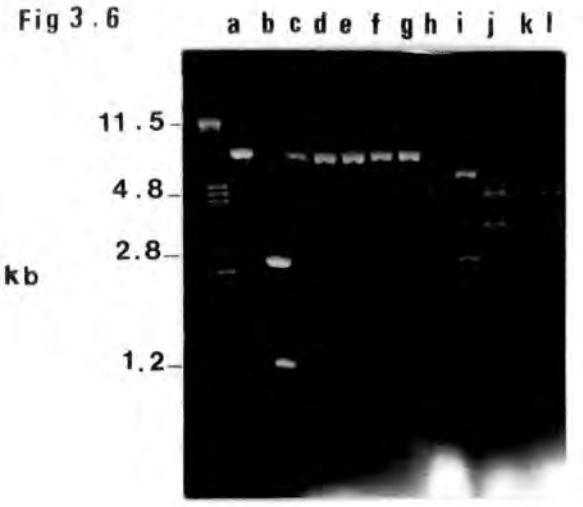
Legumin A promoter ligated to nopaline synthase coding region in pDUB IIII.

3.6 Restriction analysis of pDUB1122

Isolates identified by colony hybridisation were cultured and plasmid DNA digested with restriction enzymes. The DNA fragments from these digestions were sized on agarose gels against a *Pst* I digest of λ DNA. Isolate 2 (lanes d and i) contains the correct fusion between *nos* and the *leg A* promoter.

Key: lane (a) pDUB1111 *Eco* RI; (b) pDUB1300 *Eco* RI;
(c) - (g) are isolates 1-5 digested with *Eco* RI;
(h) - (l) are isolates 1-5 digested with *Pst* I.

Fig 3.6



presence of asymmetric *Pst* I sites in the construct, generating fragments of 5.8 kb and 2.8 kb, allowed the correct orientation to be distinguished.

Similarly, the 0.7 kb promoter fragment from pDUB1301 was isolated and ligated with *Eco* RI cleaved pDUB1111 to give pDUB1128 (fig 3.5). By colony hybridisation with LEG 1.2 it was shown that 20% of the colonies obtained had the insert DNA. Again, the clone with the promoter sequences in the correct orientation was identified by restriction analysis (data not shown).

Two additional vectors with only 95 bp of promoter sequences immediately upstream of the legumin *A* gene proximal to the *nos* coding region were prepared by the insertion of pGV1106 into pDUB1122 (section 3.2.4).

3.2.4 Construction of intermediate vectors

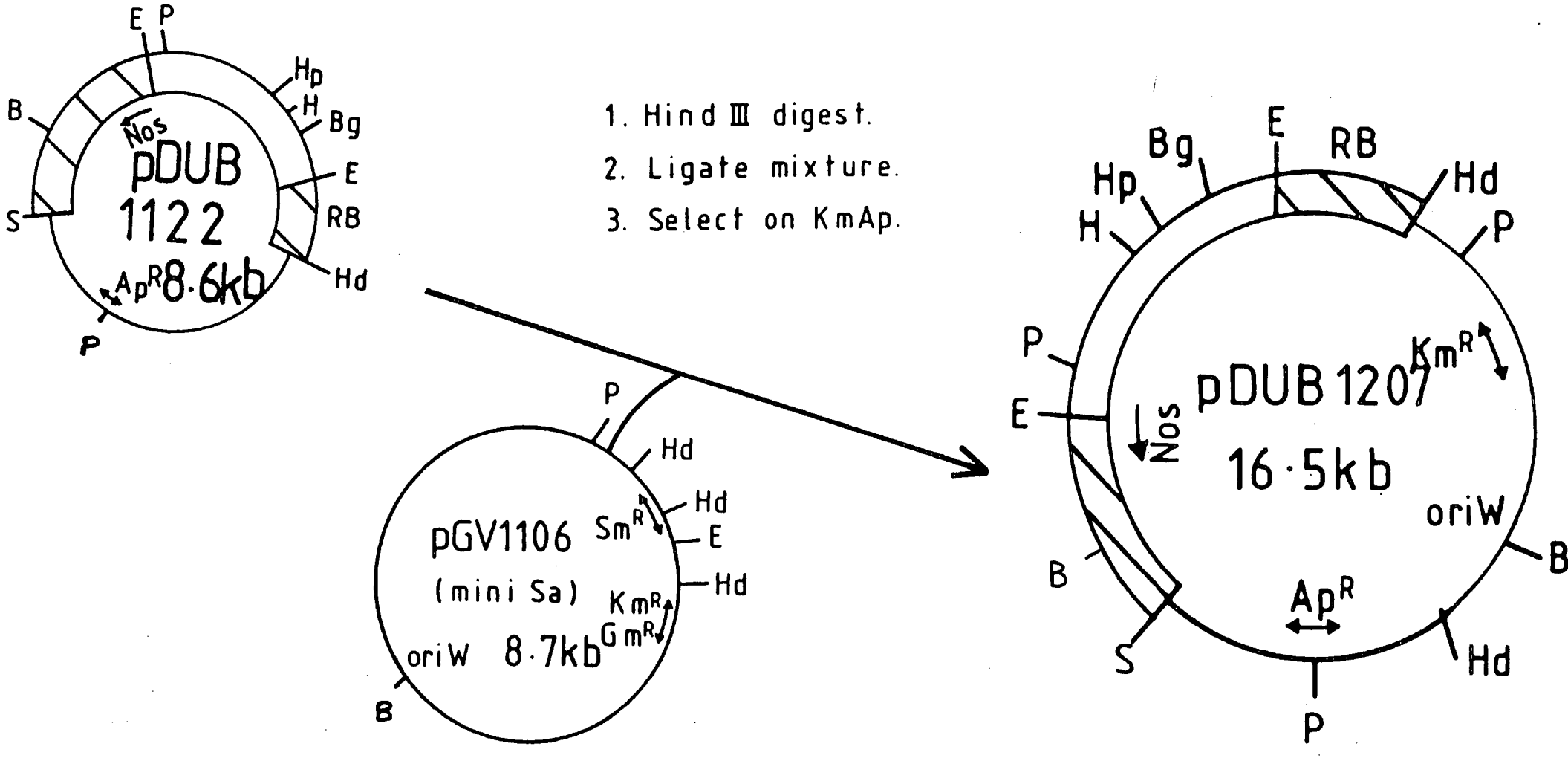
Leg-nos constructs capable of replication in *Agrobacterium tumefaciens* were made using the following approach. Sequences from pGV1106 (Mini Sa), which can replicate in both *E. coli* and *A. tumefaciens*, were inserted into pDUB1122 and pDUB1128, and designated pDUB1207 and pDUB1216, respectively. An additional oncogenic construct, with the first 95 bp of the legumin promoter spatially separated from upstream promoter sequences, has been made by the insertion of pGV1106 within the *Pst* I site of the legumin promoter of pDUB1122 to give pDUB1215. A construct with only the 95 bp promoter fragment, pDUB1217, was obtained by the excision of an *Eco* RI fragment from pDUB1215. The DNA manipulations and selection steps required are described in detail below.

The intermediate vector, pDUB1207, was constructed by mixing equimolar quantities of pGV1106 and pDUB1122, *Hin* dIII digesting and religating (fig 3.7). The recombinant plasmid was selected by ampicillin and kanamycin resistance of transformed *E.coli* [K514]. The integration and orientation of the isolates obtained was verified by *Pst* I digestion (fig 3.8a and b). In one orientation fragments of 2.7, 4.0 and 9.5 kb were obtained (isolates 2, 4 and 5) and in the other a doublet of 2.7 kb and an 11 kb band (isolates 1 and 3). The 2.1 kb fragment seen in the

3.7 Construction of intermediate vector pDUB1207

The intermediate vector pDUB1207, constructed by religating a *Hin*dIII digest of the two constituent plasmids, pDUB1122 and pGV1106, was identified by selection with ampicillin and kanamycin. Legumin *A* promoter sequences are represented by open boxes; sequences homologous to *Hin*dIII fragment 23 from pTiC58 are indicated by hatched boxes. The restriction sites are indicated as follows: B=*Bam* HI, Bg=*Bgl*II, D=*Dde*I, E=*Eco* RI, H=*Hae*III, Hd=*Hin*dIII, Hp=*Hpa*II, P=*Pst* I. (Not drawn to scale). Only the restriction sites relevant to the plasmid constructions are shown.

Fig 3.7


pDUB 1122
Nos
Ap^R 8·6kb
E
P
Hp
H
Bg
E
RB
Hd
B
S
P
1. Hind III digest.
2. Ligate mixture.
3. Select on KmAp.
pGV1106
(mini Sa)
oriW 8·7kb
Sm^R
Km^R
Gm^R
P
Hd
Hd
E
Hd
B
pDUB 1207
16·5kb
Nos
Km^R
oriW
Ap^R
Bg
E
RB
Hd
Hp
H
P
P
E
B
S
P
Hd
B

3.8 Restriction analysis of pDUB1207 and pDUB1216

(a) Minilysate DNA from the $Km^r Ap^r$ clones (putative pDUB1207 isolates) was digested with *Pst* I to establish the correct relative orientation of the two constituent plasmids. The bacterial isolate 1 generated a suitable restriction pattern (11.0 and 2.7 kb fragments) and was used for the subsequent conjugation with pDUB1006.
Key: lanes (a) - (e) isolates 1-5 digested with *Pst* I; lane (f) λ *Pst* I.

(b) The restriction pattern of pDUB1207 and pDUB1216 was checked by digestion with *Eco* RI (E), *Bam* HI (B) and *Pst* I (P). The 1.2 kb promoter fragment (LEG 1.2) derived from pDUB1300 and used as a probe in subsequent experiments is shown.

Fig 3·8a

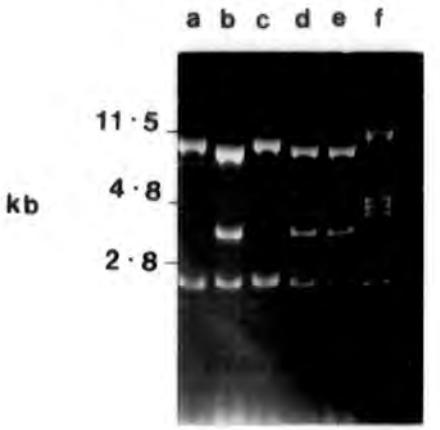
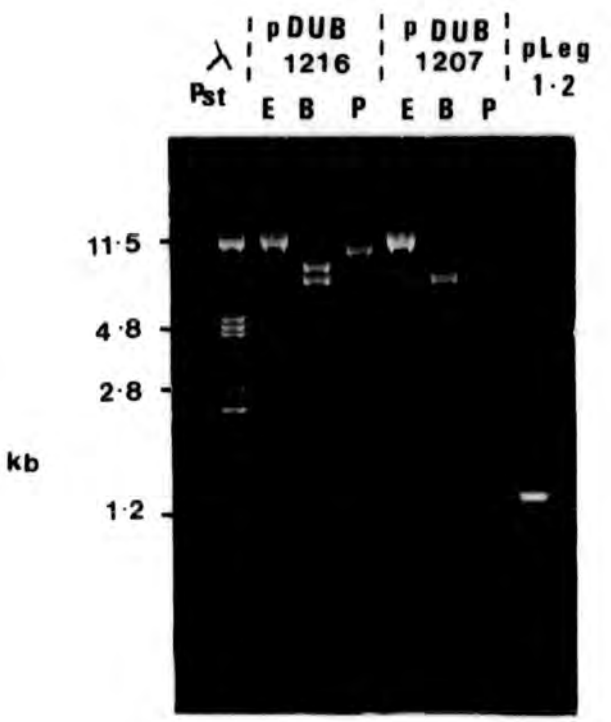


Fig 3·8 b



Bam HI digest of pDUB1207 (fig 3.8b) indicates that *Bam* HI site of the original *leg A* promoter was unexpectedly maintained after Klenow treatment and the addition of *Eco* RI linkers.

The orientation of the two constituent plasmids, seen in isolates 1 and 3, ensured that the possibility of activation of the deleted tetracycline gene by adjacent fragments was unlikely, which is important for the subsequent conjugation with a pRN3-containing strain which is selected by tetracycline resistance. The desired antibiotic sensitivity was confirmed by streaking on L-agar containing tetracycline and only those isolates which failed to grow selected for subsequent conjugations.

Similarly, pDUB1216 was constructed by the ligation of *Hind* III digests of pDUB1128 and pGV1106, followed by Ap/Km selection of transformed K514 cells (fig 3.9). This event was confirmed and the desired orientation of the constituent plasmids selected using restriction analysis (fig 3.8b). The correct orientation was indicated by 2.3, 2.7 and 11 kb fragments on *Pst* I digestion and two large fragments with *Bam* HI.

The construct, pDUB1215, containing a *leg-nos* gene with 95 bp of legumin sequences proximal to the the *nos* coding region and 1.1 kb of these sequences spacially separated by the insertion of pGV1106, was derived from pDUB1122. The plasmid, pDUB1122 contains two *Pst* I sites, one in the legumin promoter fragment and the other in the ampicillin resistance gene. The insertion of pGV1106 into the first of these two sites could be easily detected by the retention of a functional gene for ampicillin resistance, in addition to the kanamycin resistance encoded by the inserted sequences. A partial *Pst* I digest of pDUB1122 was mixed with a complete *Pst* I digest of pGV1106 and ligated (fig 3.10). The recombinant plasmid, pDUB1215, was identified by ampicillin and kanamycin resistance.

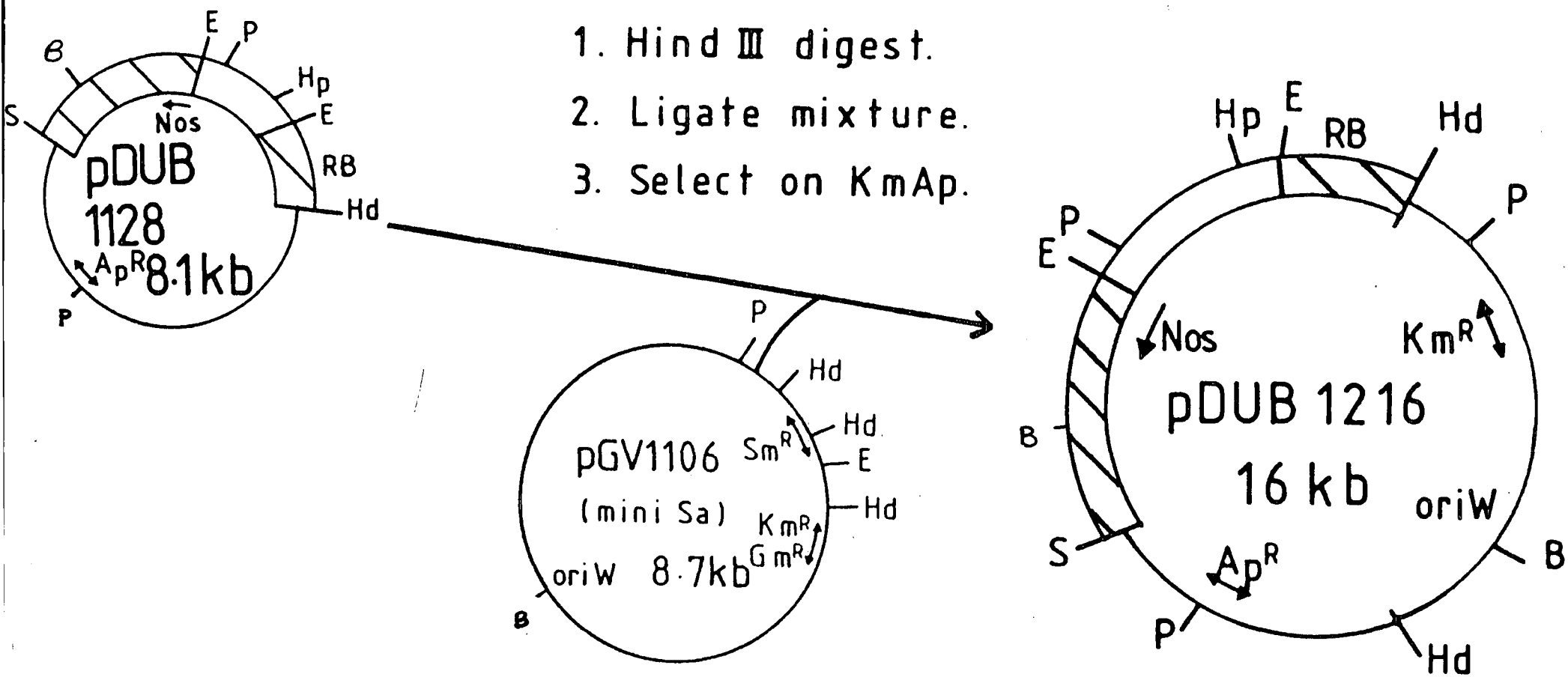
The orientation of the two plasmid components in pDUB1215 was determined and the isolate containing the single *Eco* RI site of pGV1106 proximal to the larger *leg A* promoter fragment was selected for further manipulations. This ori-

3.9 Construction of intermediate vector pDUB1216

The intermediate vector pDUB1216, constructed by religating a *Hin*dIII digest of the two constituent plasmids, pDUB1128 and pGV1106, was identified by selection with ampicillin and kanamycin. Legumin *A* promoter sequences are represented by open boxes; sequences homologous to the Ti-plasmid pDUB1006 are indicated by hatched boxes. The restriction sites are indicated as follows: B=*Bam* HI, Bg=*Bgl*II, D=*Dde*I, E=*Eco* RI, H=*Hae*III, Hd=*Hin*dIII, Hp=*Hpa*II, P=*Pst* I, RB=right border sequences.

(Not drawn to scale). Only restriction sites relevant to the plasmid construction are shown.

Fig 3.9

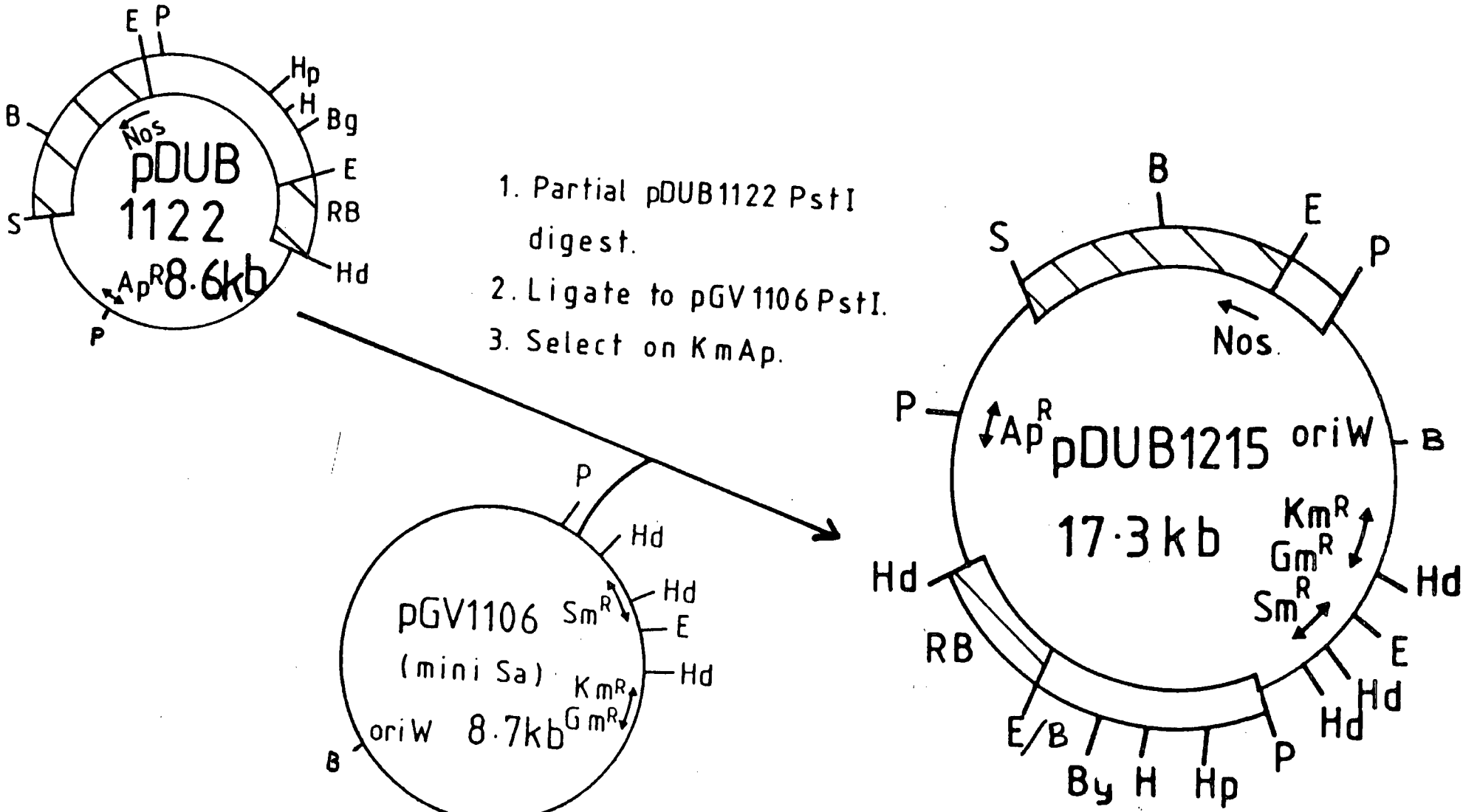

1. Hind III digest.
2. Ligate mixture.
3. Select on KmAp.
pDUB 1128 ApR 8.1kb
Nos
E
P
Hp
E
RB
Hd
S
B
P
pGV1106 (mini Sa) 8.7kb
SmR
KmR
GmR
oriW
Hd
Hd
E
Hd
P
B
pDUB 1216 16 kb
Nos
KmR
oriW
ApR
Hp
E
RB
Hd
P
P
E
B
S
P
Hd

3.10 Construction of intermediate vector pDUB1215

The intermediate vector, pDUB1215 was constructed by religating a partial *Pst* I digest of pDUB1122 to a *Pst* I digest of pGV1106. Recombinant plasmids were identified by selection with ampicillin and kanamycin. Legumin *A* promoter sequences are represented by open boxes; sequences homologous to the Ti-plasmid pDUB1006 are indicated by hatched boxes. The restriction sites are indicated as follows: B=*Bam* HI, Bg=*Bgl*II, D=*Dde*I, E=*Eco* RI, H=*Hae*III, Hd=*Hin*dIII, Hp=*Hpa*II, P=*Pst* I, RB=right border sequences.

(Not drawn to scale). Only restriction sites relevant to the plasmid constructions are shown.

Fig 3.10


E
P
Hp
H
Bg
E
RB
Hd
B
S
Nos
pDUB
1122
ApR 8.6kb
P
1. Partial pDUB1122 PstI digest.
2. Ligate to pGV1106 PstI.
3. Select on KmAp.
P
Hd
Hd
E
Hd
pGV1106
(mini Sa)
SmR
KmR
GmR
oriW 8.7kb
B
B
E
S
P
Nos.
P
ApR pDUB1215 oriW
B
17·3kb
KmR
GmR
SmR
Hd
Hd
RB
E
Hd
Hd
E/B
P
By
H
Hp

entation of the *Eco* RI sites allowed the simple excision of a 2.3 kb *Eco* RI fragment containing the larger *leg A* promoter fragment and the Sm/Sp resistance gene of pGV1106. The deleted variant, pDUB1217, was obtained by a partial *Eco* RI digest of pDUB1215 which was religated and Ap^r, Km^r colonies selected. Of 400 transformants, six of the colonies were also Sm/Sp sensitive, indicating the loss of the *Eco* RI fragment. This was confirmed by digestion of the plasmid DNA which shows the absence of the smallest *Eco* RI fragment and the loss of a *Pst* I site within this fragment (fig 3.11).

3.2.5 Insertion of chimaeric *leg-nos* genes into the T-DNA

The insertion of the chimaeric *leg-nos* genes into the T-DNA has been achieved by the mobilisation of the intermediate vectors to the *Agrobacterium tumefaciens* strain, C58C1EryCm[pDUB1006] and homologous recombination. The deleted *nos* gene and the 25 bp repeat in pDUB1006 are homologous to sequences in the intermediate vectors derived from pDUB1111 and allow a crossover event at these two locations, as indicated in fig 3.12. The resultant recombinant Ti-plasmid, following a single crossover event, can have one of the two configurations shown in figure 3.13. The two different configurations can be distinguished by probing southern blots of restricted bacterial DNA with homologous fragments. A double crossover event was detected by antibiotic selection and hybridisation data indicating the replacement of the Km/Nm-resistance gene in the Tn903 fragment by *leg-nos* sequences.

The mobilisation of the intermediate vectors to C58C1EryCm[pDUB1006] was mediated by pRN3 (Datta and Hedges, 1971). The Inc*N* plasmid, pRN3, which contains the two genes coding for tetracycline and streptomycin/spectinomycin resistance, was transfered to *E. coli* [K514] containing the intermediate vectors by conjugation (section 2.7.1). With the exception of conjugations involving pDUB1215, which retains a functional gene encoding Sm/Sp resistance, the transfer was detected by selection of transconjugants on L-Km/Sm/Sp and verified by

3.11 Restriction analysis of pDUB1215 and pDUB1217

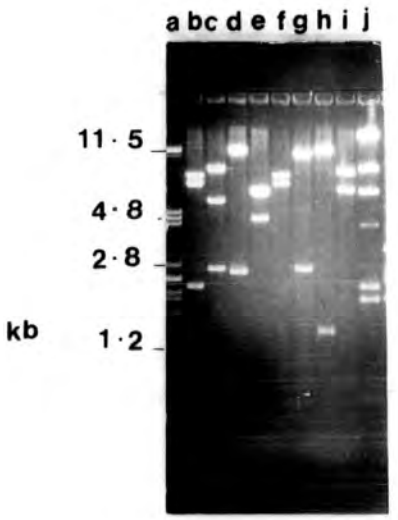
The ligation of the two plasmids pDUB1122 and pGV1106 (MiniSa) in the kanamycin and ampicillin resistant clones isolated, was confirmed by digestion with a number of restriction enzymes. The derivative plasmid, pDUB1217 was obtained from pDUB1215 by the excision of a 2.3 kb *Eco* RI fragment.

Key: lane (a) λ *Pst* I

(b) pDUB1215 *Eco* RI	(f) pDUB1217 *Eco* RI
(c) pDUB1215 *Pst* I	(g) pDUB1217 *Pst* I
(d) pDUB1215 *Hin*dIII	(h) pDUB1217 *Hin*dIII
(e) pDUB1215 *Bam* HI	(i) pDUB1217 *Bam* HI

and (j) λ *Hin*dIII

Fig 3·11

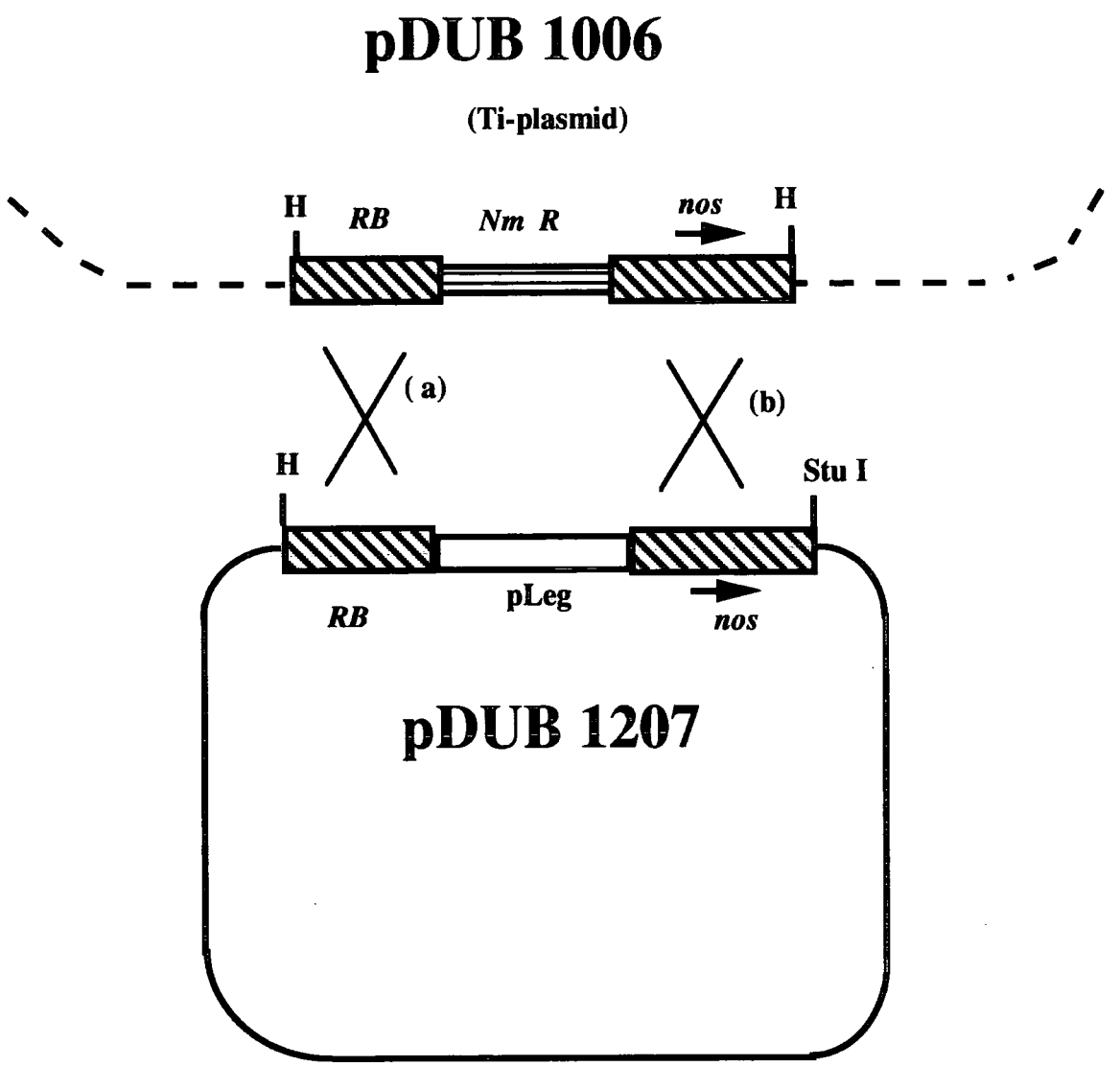


3.12 Cointegration of pDUB1207 and pDUB1006

A conjugation was carried out between *E.coli*, harbouring pDUB1207, and the *Agrobacterium tumefaciens* strain C58C1 $Ery^r Cm^r$ [pDUB1006]. A cointegrate plasmid is formed by a single crossover event through either *Hin*dIII fragment 23 a or b, as indicated. The plasmids, not to scale, show the sequences derived from *Hin*dIII fragment 23 as hatched boxes, legumin sequences as open boxes, the fragment containing the neomycin gene from pDUB1006 as a box with straight lines, the Ti-plasmid sequences as broken lines and pBR322 and pGV1106 as solid lines. A double crossover event results in the direct replacement of the neomycin resistance gene by the legumin promoter fragment.

UNIV. DUBLIN

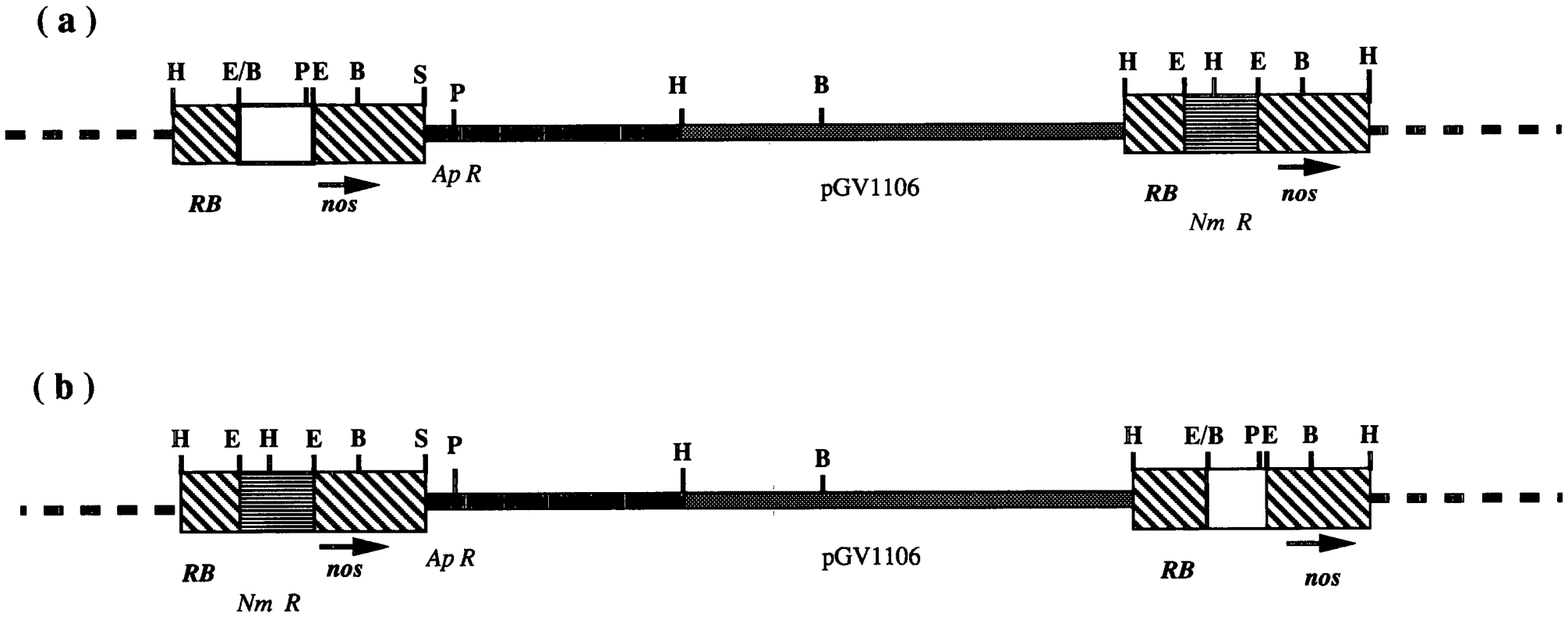
Fig 3.12


pDUB 1006
(Ti-plasmid)
H
RB
Nm R
nos
H
(a)
(b)
H
Stu I
RB
pLeg
nos
pDUB 1207

3.13 The two possible results of a single crossover between pDUB1207 and pDUB1006

The cointegrate plasmid formed by a single crossover event generates two different plasmid conformations depending on the site of the crossover event, indicated in this figure as (a) the right border or (b) the nopaline synthase coding region.
The plasmids show the sequences derived from *Hin*dIII fragment 23 as hatched boxes, legumin sequences as open boxes and pBR322 as a black box and sequences from pGV1106 as a stippled box. The fragment containing the neomycin resistance gene is indicated by a box with straight lines and the remainder of the Ti-plasmid sequences are indicated by broken lines. Restriction sites are indicated as follows: B=*Bam* HI, E=*Eco* RI, H=*Hin*dIII , P=*Pst* I, RB=right border sequences.

Fig 3.13


(a)
H
E/B
PE
B
S
P
H
B
H
E
H
E
B
H
RB
nos
Ap R
pGV1106
RB
nos
Nm R
(b)
H
E
H
E
B
S
P
H
B
H
E/B
PE
B
H
RB
nos
Nm R
Ap R
pGV1106
RB
nos

streaking on media containing the antibiotics, tetracycline and ampicillin. However, with an efficiency of transfer of up to 70%, most isolated colonies contained both plasmids.

The four different intermediate vectors were each transmitted from *E.coli* to the *Agrobacterium tumefaciens* strain C58C1EryCm [pDUB1006] during conjugation (section 2.7.2). These transconjugants were identified by selection with erythromycin, chloramphenicol, neomycin and gentamycin at a frequency of between 10^{-5}-10^{-8}. However, this selection does not distinguish between bacteria harbouring the two plasmids separately and those containing the desired cointegrate. To select for the cointegrate, a purified transconjugant was transmitted to a cured strain, C58C1Rif^r (section 2.7.3). An unintegrated pGV1106 derived plasmid cannot be transfered and the transconjugant can be distinguished from the original Ti-plasmid by the antibiotic markers transferred to the cured strain. The Ti-plasmids were transmitted to the recipient strain during conjugation with an efficiency of 4%.

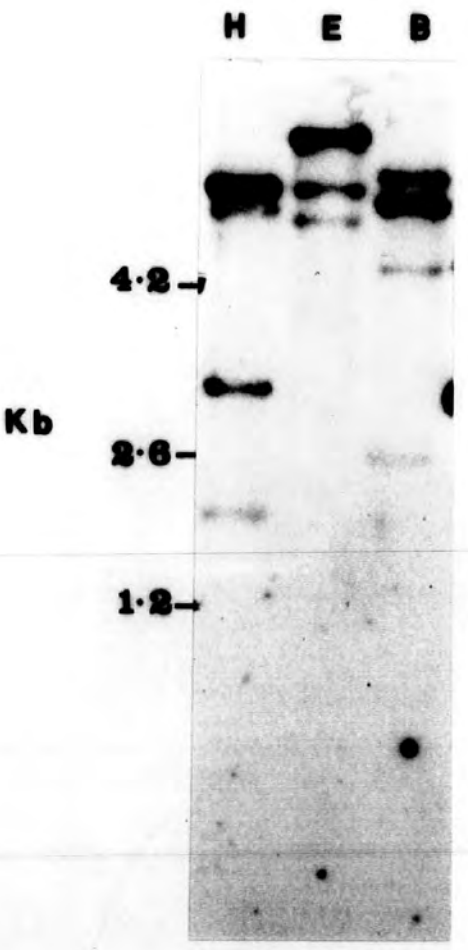
In this way the cointegrates, C58C1Rif[1207::1006] and C58C1Rif[1216::1006] were isolated at a frequency of $10^{-5} - 10^{-7}$. These cointegrates were selected by plating on rifampicin, neomycin and gentamycin and checked for Sm^s before the next conjugation step. The successful formation and orientation of the cointegrate with the intermediate vector was confirmed by Southern blotting of total *Agrobacterium* DNA (section 2.5) using pGV0601 (fig 3.14) and LEG 1.2 (fig 3.15 and 3.16) as probes.

The probe, pGV0601, contains the *Hin*dIII 23 fragment from pTiC58 and is homologous to sequences in both pDUB1006 and the intermediate vector, pDUB1207 (fig 3.14). Hybridisation with this probe to a *Hin*dIII digest of pDUB1207::1006 would give radioactive bands at 7.1, 4.3 and 1.5 kb if the crossover event had occurred through the nopaline synthase coding region. In fig 3.14 the hybridising *Hin*dIII fragments of 8.6, 2.8 and 1.5 are diagnostic of crossover be-

3.14 Southern blot of total DNA from a cointegrate plasmid between pDUB1207 and pDUB1006, probed with pGV0601

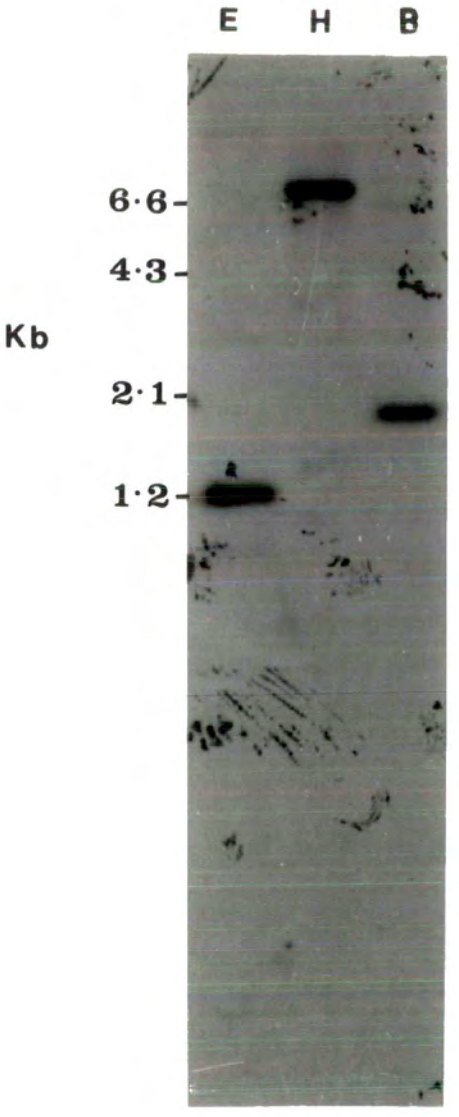
An *Agrobacterium* isolate containing a putative cointegrate between plasmids pDUB1207 and pDUB1006 was confirmed by probing total DNA (section 2.11) with ^{32}-P labelled pGV0601. 10μg of isolated DNA was digested with *Eco* RI (E), *Hin*dIII (H) or *Bam* HI (B). These digests were run on agarose gels, Southern blotted onto nitrocellulose and probed with labelled DNA (section 2.5)

Fig 3·14



3.15 Southern blot of total DNA from a cointegrate plasmid between pDUB1207 and pDUB1006, probed with LEG 1.2

The presence of the legumin promoter fragment in the putative cointegrate was confirmed by probing total DNA from lysed cells with a ^{32}-P labelled 1.2 kb fragment isolated from pDUB1300. 10μg of total isolated DNA from individual isolates was digested with *Eco* RI (E), *Hin*dIII (H) or *Bam* HI (B). These digests were run on agarose gels, Southern blotted onto nitrocellulose and probed with labelled DNA.

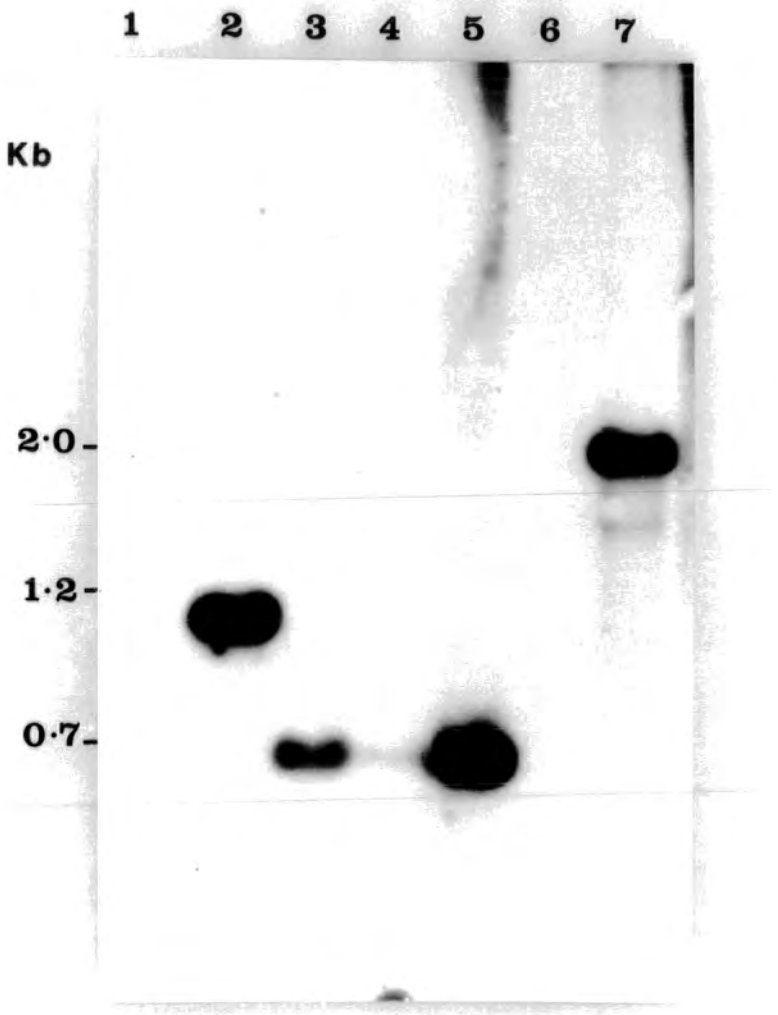

E
H
B
6·6
4·3
Kb
2·1
1·2

3.16 Southern blot of total *Agrobacterium tumefaciens* DNA from a cointegrate plasmid between pDUB1006 & pDUB1216 probed with LEG 1.2

The presence of the legumin promoter fragment in the putative cointegrate, pDUB1216::1006, was confirmed by probing *Eco* RI digested total DNA from lysed cells (section 2.11) with a ^{32}-P labelled 1.2 kb fragment (section 2.5) isolated from pDUB1300.

Key: Lane (1) pDUB1006 *Eco* RI; (2) pDUB1207::1006 *Eco* RI;
(3)-(5) Three isolates of pDUB1216::1006 *Eco* RI digested;
(6) pDUB1006 *Bam* HI; (7) pDUB1207::1006 *Bam* HI.

Fig 3·16



tween the right border seqences. An *Eco*RI digest of either crossover product would give almost identical hybridising fragments because of the similarity in size of the legumin promoter fragment and the neomycin resistance gene of pDUB1006.

The legumin promoter probe (LEG 1.2) hybridises to a single *Eco*RI band of 1.2 kb in the plasmid cointegrate pDUB1207::1006, confirming the presence of the promoter sequences (fig 3.15). The LEG 1.2 probe also hybridises to a single *Hin*dIII fragment of approximately 7 kb and a 2.1 kb *Bam*HI fragment in this plasmid (fig 3.15). Using the same methodology, the putative *Agrobacterium* isolates of the cointegrate between pDUB1216 and pDUB1006 were verified (fig 3.16).

The second crossover event within the cointegrate plasmids, pDUB1207::1006 and pDUB1216::1006, was identified following a conjugation to a second cured strain, C58C1Sm^{r}Sp^r, and the plasmids finally obtained designated pDUB1016 and pDUB1018. This crossover event results in the loss of genes encoding for both gentamycin (encoded by pGV1106) and neomycin (encoded by Tn903) resistance so the transmitted plasmids were selected by octopine catabolism. This involved subculturing twice in liquid media before plating the bacteria on octopine and ornithine supplemented NO.O media, containing streptomycin and spectinomycin (section 2.7.3). The colonies obtained were replica plated onto L-Nm and L-Gm to isolate bacteria that had lost both resistance genes. Plasmids derived from the second crossover event were isolated at a frequency of between $10^{-2} - 10^{-3}$.

Initial experiments had used the cured strain C58C1Ery^{r}Cm^r in which to select the double crossover, but the observed sensitivity of isolated colonies to neomycin and gentamycin was often the result of the loss of the entire Ti-plasmid. Subsequent conjugations were carried out with the C58C1Sm^{r}Sp^r strain and DNA from a number of opine catabolising isolates digested with *Eco*RI to screen for the presence of the legumin promoter, as shown in figures 3.17 and 3.18.

The majority of putative isolates of the plasmids pDUB1016 and pDUB1018 hybridised to the Leg 1.2 probe giving a characteristic, 1.2 kb (fig 3.17) or 0.7 kb

3.17 Southern blot of total *Agrobacterium tumefaciens* DNA from putative pDUB1016 isolates probed with LEG1.2

The resolved plasmid pDUB1016 results in the loss of the genes encoding neomycin and gentamycin resistance and can be selected (after enrichment) on solid media containing opines as the sole nitrogen source. However, to exclude the possibility of the phenotype being a result of the loss of the Ti-plasmid, total DNA from individual isolates was cleaved with *Eco* RI and probed with LEG 1.2. The ten putative isolates analysed all possessed the 1.2 kb promoter fragment. The four isolates used for subsequent tumour induction are indicated.

Fig 3·17

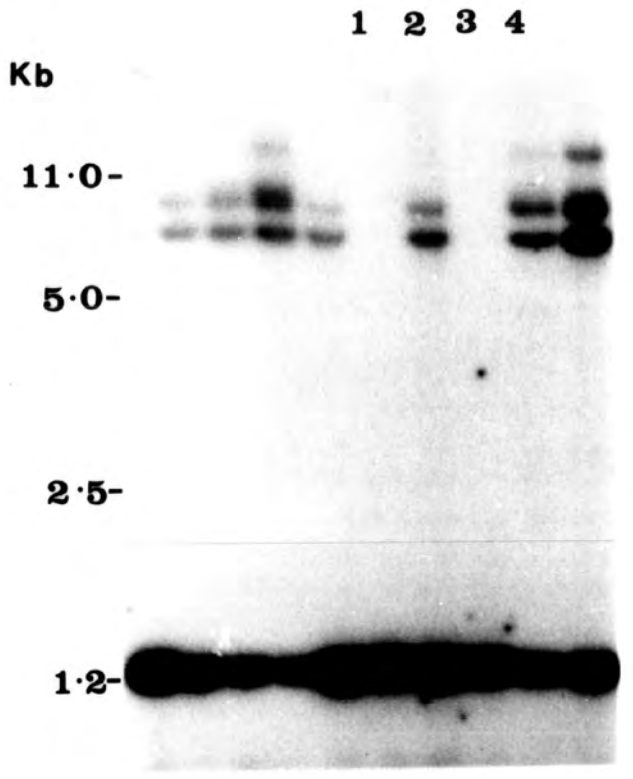
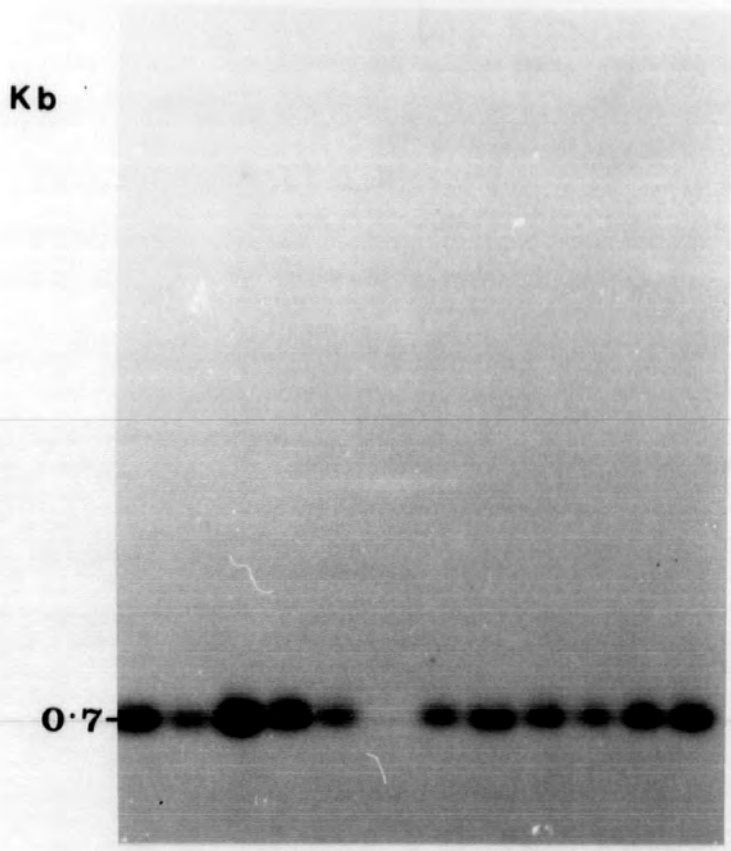


Fig 3·18



3.18 Southern blot of total *Agrobacterium tumefaciens* DNA from putative pDUB1018 isolates probed with LEG 1.2

The resolved plasmid pDUB1018 results in the loss of the genes encoding neomycin and gentamycin resistance and can be selected (after enrichment) on solid media containing opines as the sole nitrogen source. However, to exclude the possibility of the phenotype being a result of the loss of the Ti-plasmid, total DNA from twelve individual isolates was cleaved with *Eco* RI and probed with LEG1.2. The majority of the isolates tested had the 0.7 kb promoter fragment.

band (fig 3.18), respectively. In fig 3.17 the additional hybridising bands observed are probably caused by partially digested products which would be 7.4, 9.2 and 12.2 kb in length. Additionally, to verify these crossover events total DNA from two different isolates was cleaved with *Bam* HI, *Eco* RI, and *Hin*dIII and Southern blots probed with the labelled legumin promoter (fig 3.19). The hybridising bands in each case showed the expected pattern, generating a *Bam* HI band of 2.1 kb and a 4.3 kb *Hin*dIII band. The additional partial *Eco* RI digestion products were not observed in this experiment.

Similarly, the cointegrates between pDUB1006 and the other pGV1106 derivatives, pDUB1215 and pDUB1217, were selected on rifampicin and gentamycin at a frequency of 10^{-7}. The loss of the Tn903 Km/Nm fragment from these plasmids by a double crossover event was observed at a frequency of $> 10^{-2}$ and could be screened directly by replica-plating of the colonies obtained onto neomycin plates and sensitive colonies determined. The $Rif^r Gm^r Nm^s$ cointegrates of pDUB1215 and pDUB1217 with pDUB1006 were designated pDUB1017 and pDUB1019, respectively.

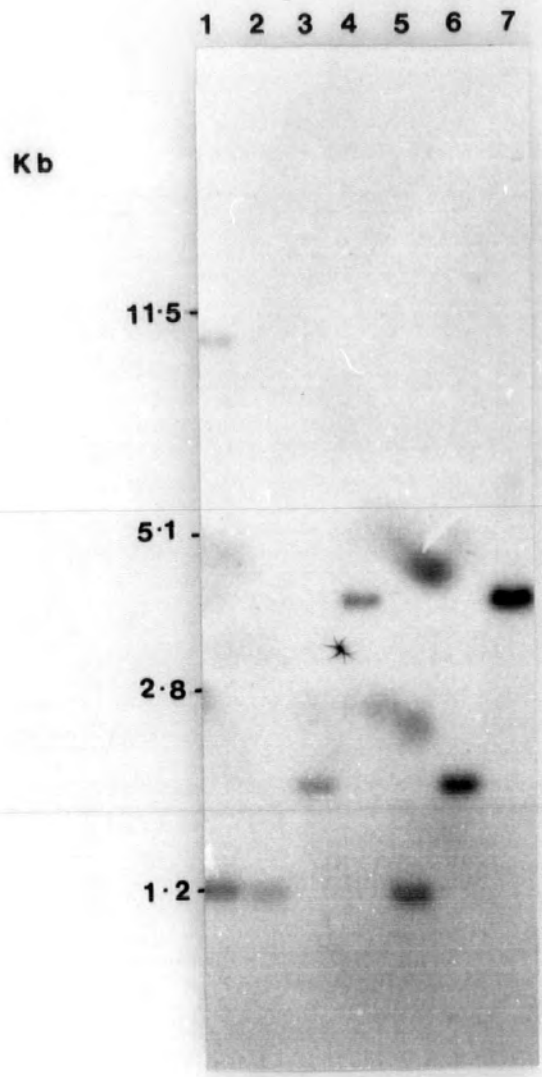
The crossover event was confirmed by probing Southern blots of total *Agrobacterium* DNA digested with *Eco* RI, *Bam* HI and *Pst* I (fig 3.20) with the legumin promoter fragment. In fig 3.20b it can be seen that the legumin *A* promoter probe would be expected to hybridise to two bands with digests of pDUB1017 but only one band with pDUB1019. In *Eco* RI digests of pDUB1017 two hybridising bands are seen (fig 3.20) but the more intense signal of the 1.1 kb legumin fragment is absent in the corresponding pDUB1019 digests. The larger promoter fragment is found on a 2.3 kb *Eco* RI fragment in pDUB1017 and the small promoter fragment is found on a 7.6 kb *Eco* RI fragment. Some larger hybridising *Eco* RI bands are probably caused by partial digestion products.

3.19 Southern blot of total *Agrobacterium tumefaciens* DNA from pDUB1016 isolates probed with LEG1.2

The plasmid pDUB1016 isolates identified in figure 3.17 were confirmed by probing 10μg of isolated DNA, digested with *Eco* RI, *Bam* HI, or *Hin*dIII, with LEG 1.2.

Key:
Lane (1) pDUB1207::1006 *Eco* RI; (2) pDUB1016 (isolate 1) *Eco* RI ; (3) pDUB1016 (isolate 1) *Bam* HI ; (4) pDUB1016 (isolate 1) *Hin*dIII; (5) pDUB1016 (isolate 4) *Eco* RI ; (6) pDUB1016 (isolate 4) *Bam* HI; (7) pDUB1016 (isolate 4) *Hin*dIII.

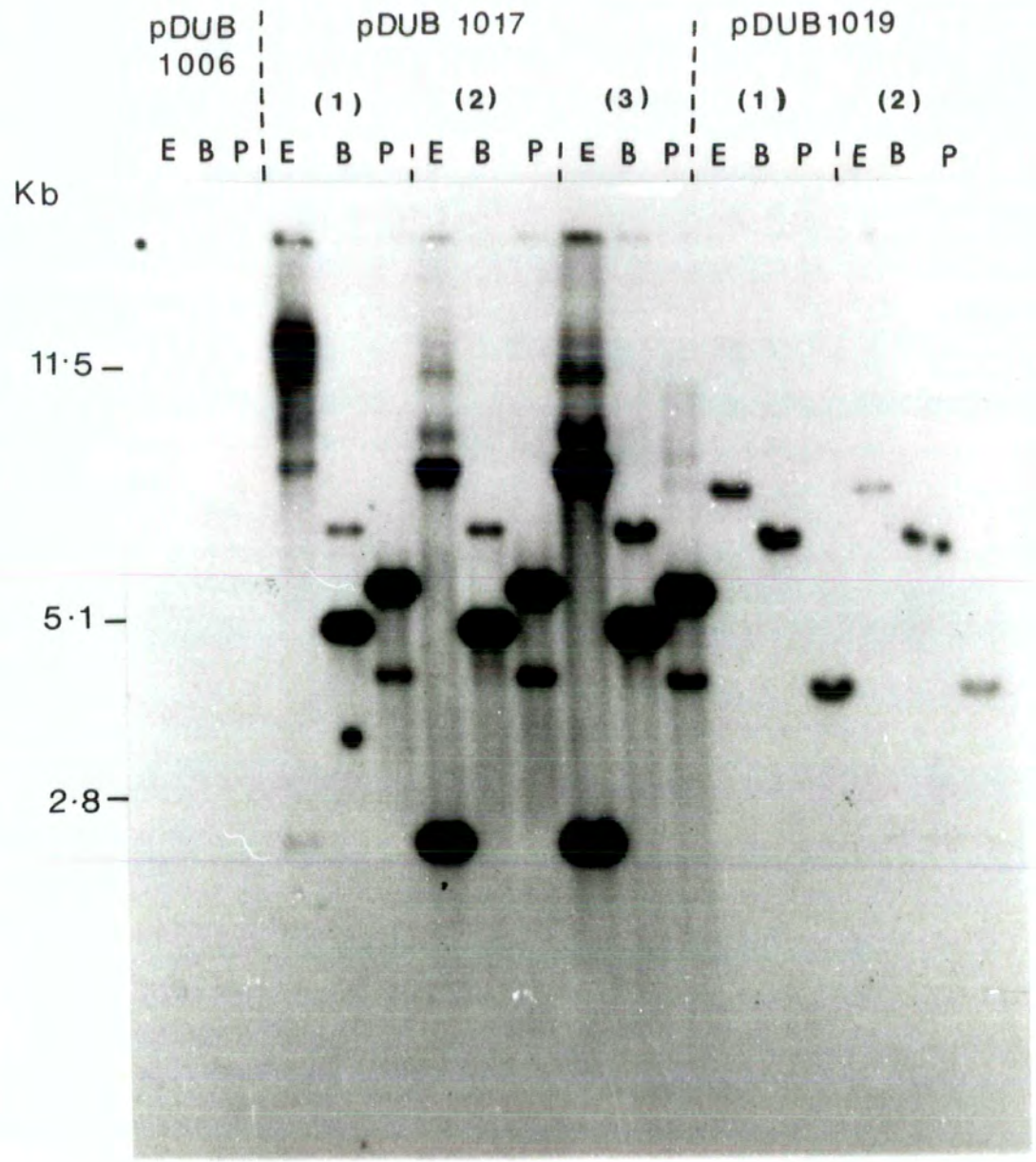
Fig 3·19



3.20 Southern blot of total *Agrobacterium tumefaciens* DNA from putative pDUB1017 and pDUB1019 isolates probed with LEG 1.2

These plasmids were identified by the loss of the gene encoding neomycin resistance. Total DNA from three putative isolates of pDUB1017 and two isolates of pDUB1019 were cleaved with *Eco* RI (E), *Bam* HI (B) or *Pst* I (P) and probed with LEG 1.2.

Fig 3·20


pDUB 1006
pDUB 1017
pDUB1019
(1)
(2)
(3)
(1)
(2)
E B P E B P E B P E B P E B P E B P
Kb
11·5
5·1
2·8

3.2.6 Analysis of callus tissue

The four Ti-plasmid vectors finally constructed, containing the different *leg-nos* fusions shown in fig 3.21, were screened for nopaline production in callus tissue obtained from *Kalanchoë*. Cultures of *Agrobacteria* harbouring these Ti-plasmid vectors were used to inoculate multiple wound sites on *Kalanchoë* leaves. Additionally the same plants were inoculated with *Agrobacteria* containing pDUB1006 and pDUB1003Δ31, as negative and positive controls, respectively. The proportion of inoculation sites which produced visible tumours appeared to depend on the physiological state of the individual plant rather than differences in the inducing strain. However differing responses were noted between tumours incited by plasmids, pDUB1016 and pDUB1018, as compared to those initiated by plasmids which contained sequences derived from pGV1106, namely, pDUB1017 and pDUB1019. Plasmids pDUB1016 and pDUB1018 appeared to be relatively more oncogenic, producing larger and faster growing tumours.

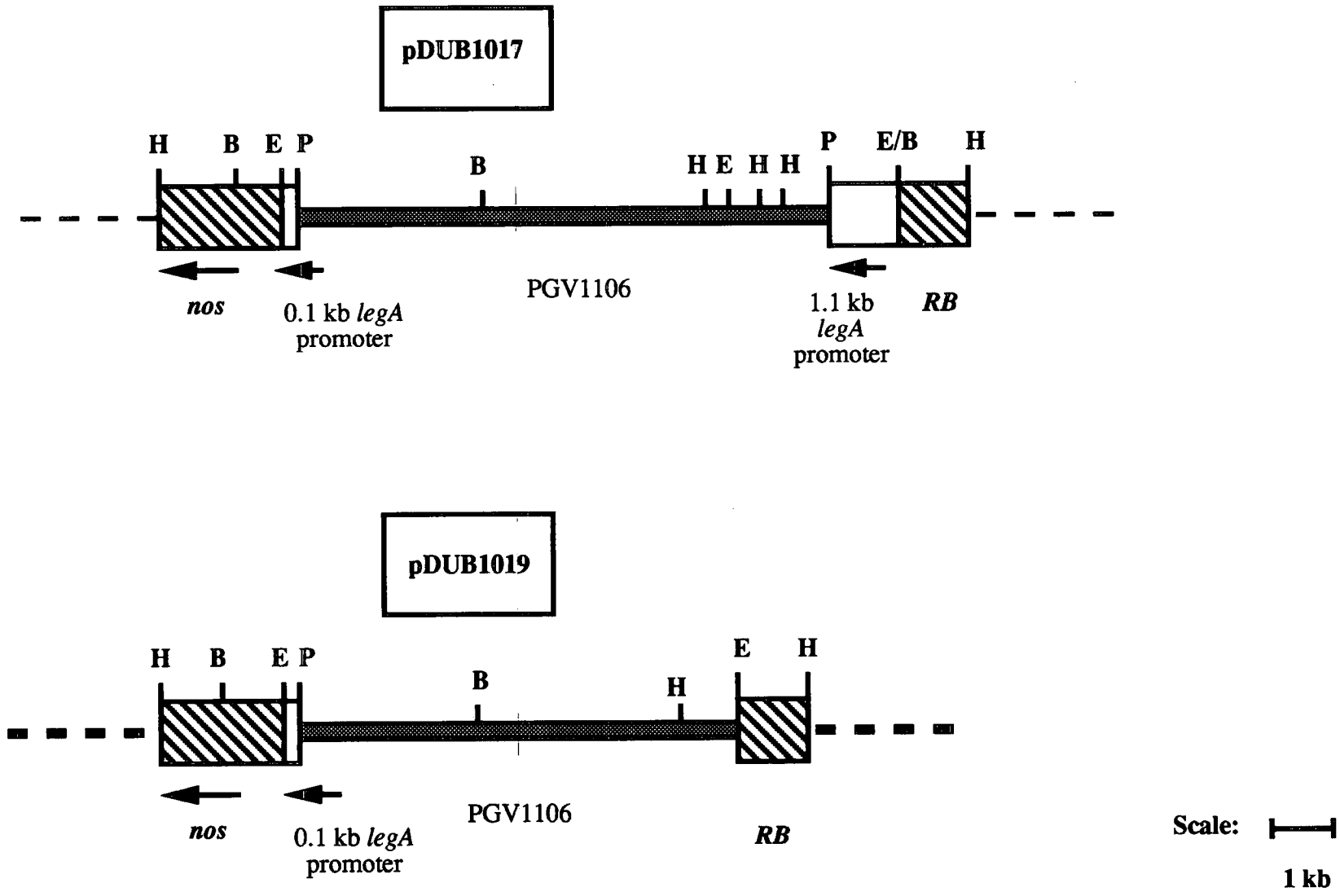
The tumours induced on *Kalanchoë* leaves infected with *Agrobacterium tumefaciens* harbouring pDUB1017 or pDUB1019, which contain just 0.1 kb of the *leg A* promoter proximal to the *nos* coding region, were slow growing and fewer were obtained. Analysis was hindered by the small amount of tumour tissue obtained but no detectable nopaline was observed in tumour extracts (fig 3.22). UV fluorescent material was observed in the extracts of tumours derived from pDUB1019 (fig 3.22b), and to a lesser extent in other inducing strains, but this did not co-migrate with the nopaline standards. Similarly, no nopaline production was detected in tumour extracts derived from agrobacteria harbouring pDUB1018, the 0.7 kb promoter construct (data not shown).

Chronologically pDUB1016, the full-length leg *A* promoter fusion, was the first oncogenic vector of the series to be constructed and four separate isolates were used to infect *Kalanchoë* plants. Infection with agrobacteria harbouring pDUB1016, the ' full-length' promoter construct, gave rise to tumours, a few of which produced low levels of nopaline, as shown in figure 3.23. All these nos^+

3.20b Structure of pDUB1017 and pDUB1019

The product of a double crossing over event between (1) pDUB1215 and pDUB1006; and (2) pDUB1217 and pDUB1006 are shown. These products were identified by a Rif^r Gm^r Nm^s phenotype.
The plasmids show the sequences derived from *Hin*dIII fragment 23 as hatched boxes, legumin sequences as open boxes, sequences from pGV1106 as a stippled box and the Ti-plasmid sequences are indicated by broken lines. Restriction sites are indicated as follows: B=*Bam* HI, E=*Eco* RI, H=*Hin*dIII , P=*Pst* I, RB=right border sequences.

Fig 3.20 b

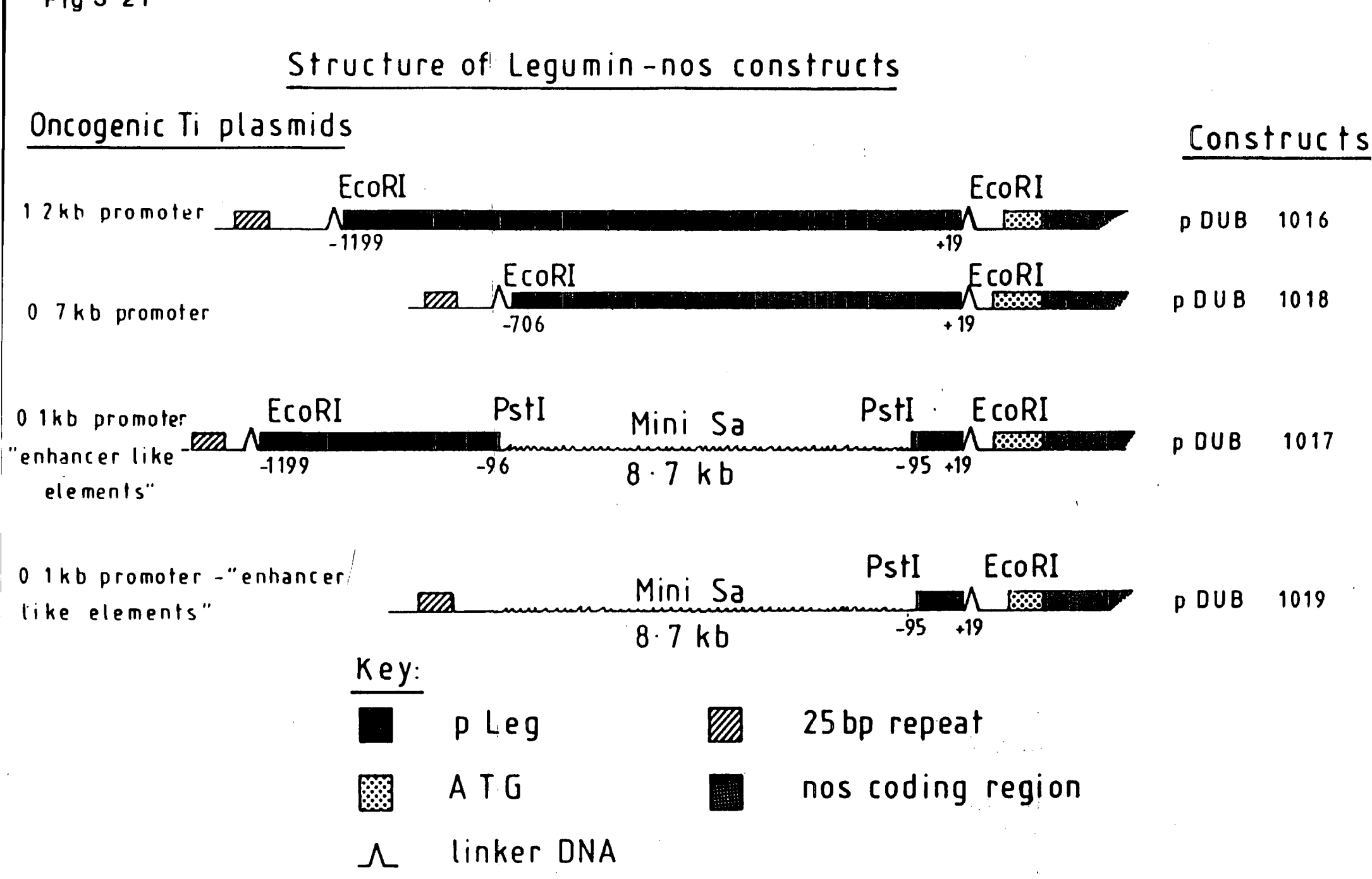

pDUB1017
H B E P
B
H E H H
P E/B H
nos
0.1 kb legA promoter
PGV1106
1.1 kb legA promoter
RB
pDUB1019
H B E P
B
H
E H
nos
0.1 kb legA promoter
PGV1106
RB
Scale:
1 kb

3.21 Structure of the oncogenic *leg-nos* constructs

The four different promoter deletions of *leg*A fused to nopaline synthase are shown. The *Eco* RI linkers, the signal sequences required for transfer to the plant genome and the nopaline synthase translational start site are indicated. The plasmid numbering system is outlined in Table 2.1.

Not drawn to scale .

Structure of Legumin-nos constructs



3.22 Opine assays on tumours induced with pDUB1017 and pDUB1019

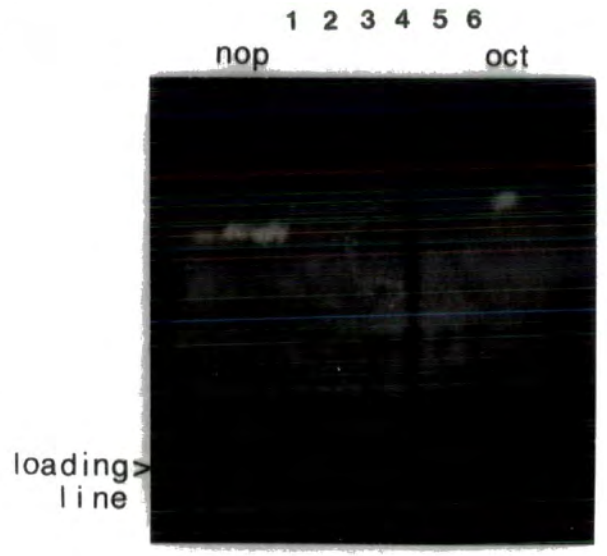
Approximately 50μg of tumour tissue, derived from separate wound sites, was crushed in an eppendorf with a glass rod and the supernatant loaded onto 3MM paper. After electrophoresis and staining the UV fluorescent spots were visualised and photographed. Nopaline (2.5, 12.5 and 25 μg) and octopine standards (12 μg) are indicated. The tumours were induced with C58C1 containing the following:

Key: **(a)** 1. pDUB1006;
2. pDUB1017;
3. pDUB1017;
4. pDUB1017;
5. pDUB1017;
6. pDUB1017.

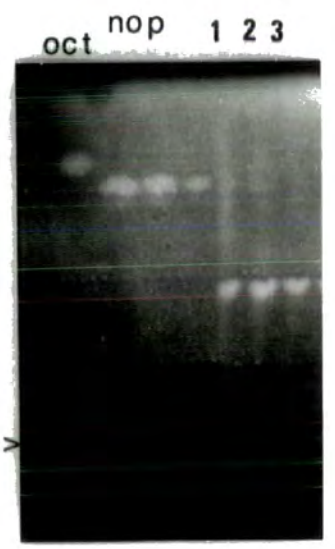
(b) 1. pDUB1006;
2. pDUB1019
3. pDUB1019.

Fig 3・22

(a)



(b)



3.23 Opine assays on tumours induced with pDUB1016

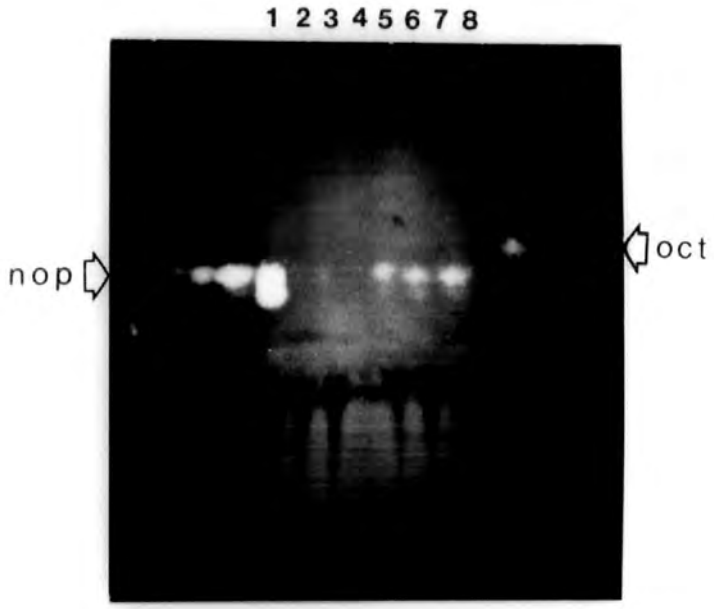
Approximately 50μg of tumour tissue, derived from separate wound sites, was crushed in an eppendorf with a glass rod and typically 20μl of supernatant was loaded onto 3MM paper. After electrophoresis and staining the UV fluorescent spots were visualised and photographed. Nopaline (6.25, 12.5 and 25 μg) and octopine standards (12 μg) are indicated. The tumour were induced with the following strains:

Key: **(a)** 1. pDUB1003Δ31;
2. pDUB1006;
3. pDUB1006;
4. pDUB1017;
5. pDUB1016 isolate 4 with 20μl extract;
6. pDUB1016 isolate 4 with 30μl extract;
7. pDUB1016 isolate 4 with 40μl extract;
8. pDUB1016 isolate 3.

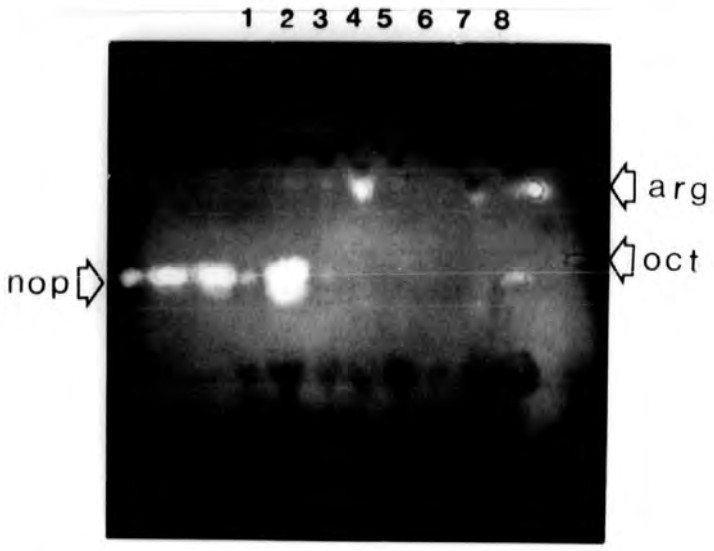
(b) 1. pDUB1006;
2. pDUB1003Δ31;
3. pDUB1016 isolate 1;
4. pDUB1016 isolate 1;
5. pDUB1016 isolate 3;
6. pDUB1016 isolate 3;
7. pDUB1016 isolate 4;
8. pDUB1016 isolate 4;

Fig 3·23

(a)



(b)



tumours were derived from one particular isolate, indistinguishable from those strains which did not produce nopaline, on the Southern blots probed with the legumin promoter fragment (fig 3.17 and fig 3.19). At least half of the tumours obtained with pDUB1016 (isolate 4) showed a fluorescent spot that co-migrated with the nopaline standard (fig 3.23b), corresponding to an average of 0.1mg/mg of tumour tissue. This level of nopaline production is only 5-10% of that observed with the positive control, pDUB1003Δ31.

These results were consistently obtained when these isolates were used to inoculate an additional set of *Kalanchoë* plants and the resulting tumours assayed for nopaline production.

3.3 ANALYSIS OF PROMOTER DELETIONS IN REGENERATED TOBACCO PLANTS

3.3.1 Strategy for the construction of disarmed Ti-plasmid vectors

A series of disarmed vectors containing the promoter fragments from *leg*A ligated to the *nos* coding region, were prepared in parallel using pBin19 (Bevan, 1984). The disarmed Ti-plasmid vector,

pBIN19, contains the polylinker from M13 mp19, the *npt* II gene regulated by the nopaline synthase promoter and a kanamycin resistance gene for selection in bacteria. Therefore the insertion of chimaeric *leg-nos* fusions, derived from pDUB1122, within the polylinker of pBIN19, was simply achieved. An example of the construction of one of these vectors, pDUB1500, is shown in figure 3.24. The other disarmed (*leg-nos*) Ti-plasmid vectors were made in a similar fashion, as outlined in fig 3.25a and fig 3.25b.

Four different transcriptional fusions, derived from pDUB1122, were finally synthesised and ligated into pBIN19, spanning 0.1, 0.8 and 1.2 kb of the *leg A* promoter and were designated pDUB1501, pDUB1500 and pDUB1502. An additional construct, pDUB1503, containing the "full-length" legumin promoter, but possessing an inverted upstream promoter fragment relative to the remain-

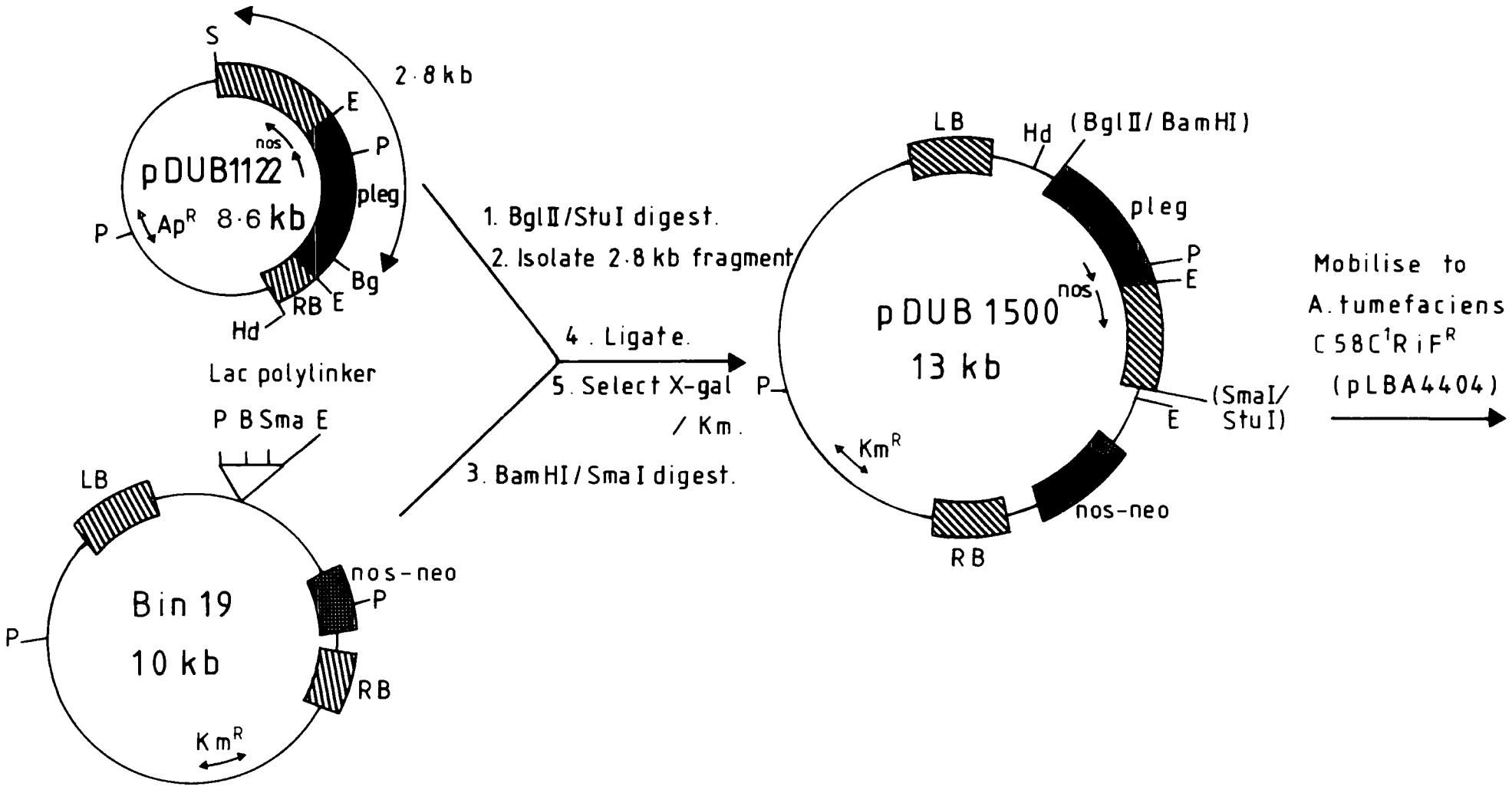
3.24 Construction of chimaeric *leg-nos* genes in pBIN19

The legumin-nopaline synthase fusion in pDUB1122 was isolated by a *Bgl*II/*Stu*I digest. The chimaeric gene was isolated from an agarose gel as a 2.8 kb fragment and ligated to a *Bam* HI/*Sma*I digest of pBIN19. The insertion of the fragment into the polylinker of pBIN19 allows the simple detection of recombinant plasmids by selection on X-gal plates containing kanamycin.

Sequences derived from the Ti-plasmid, namely the right (RB) and left border (LB) repeats and the nopaline synthase coding region (*Hin*dIII fragment 23 from pTiC58), are indicated by hatched boxes. The neomycin phosphotransferase gene under the control of the *nos* promoter is indicated by a cross-hatched box and legumin promoter sequences by a black box. B=*Bam* HI; Bg=*Bgl*II; E=*Eco* RI; Hd=*Hin*dIII; P=*Pst* I; S=*Stu*I.

Fig 3·24

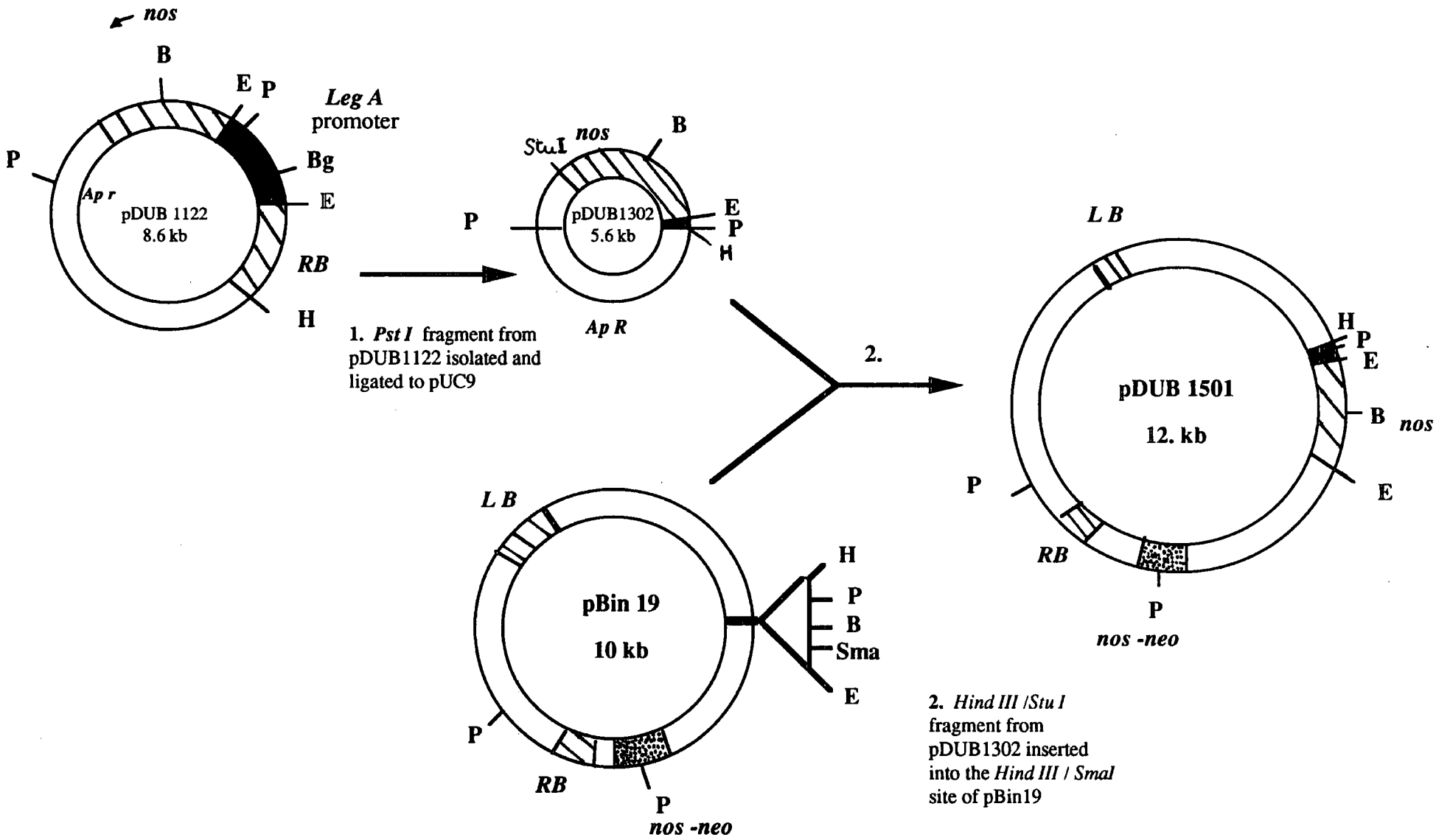
Legumin - nos gene in Bin 19



3.25a Construction of pDUB1501

The 0.1 kb *leg-nos* construct was derived from pDUB1122 by excision with *Pst* I. This 2.8 kb fragment was subcloned into pUC9 to give pDUB1302, which allows the isolation of a 2 kb fragment using *Hin*dIII and *Stu*I, which can be ligated into the polylinker region of pBIN19. Sequences derived from the Ti-plasmid, namely the right (RB) and left border (LB) repeats and the nopaline synthase coding region (*Hin*dIII fragment 23 from pTiC58), are indicated by hatched boxes. The neomycin phosphotransferase gene under the control of the *nos* promoter is indicated by a stippled box and legumin promoter sequences by a black box. B=*Bam* HI; Bg=*Bgl*II; E=*Eco* RI; H=*Hin*dIII; P=*Pst* I.

Fig 3.25 a

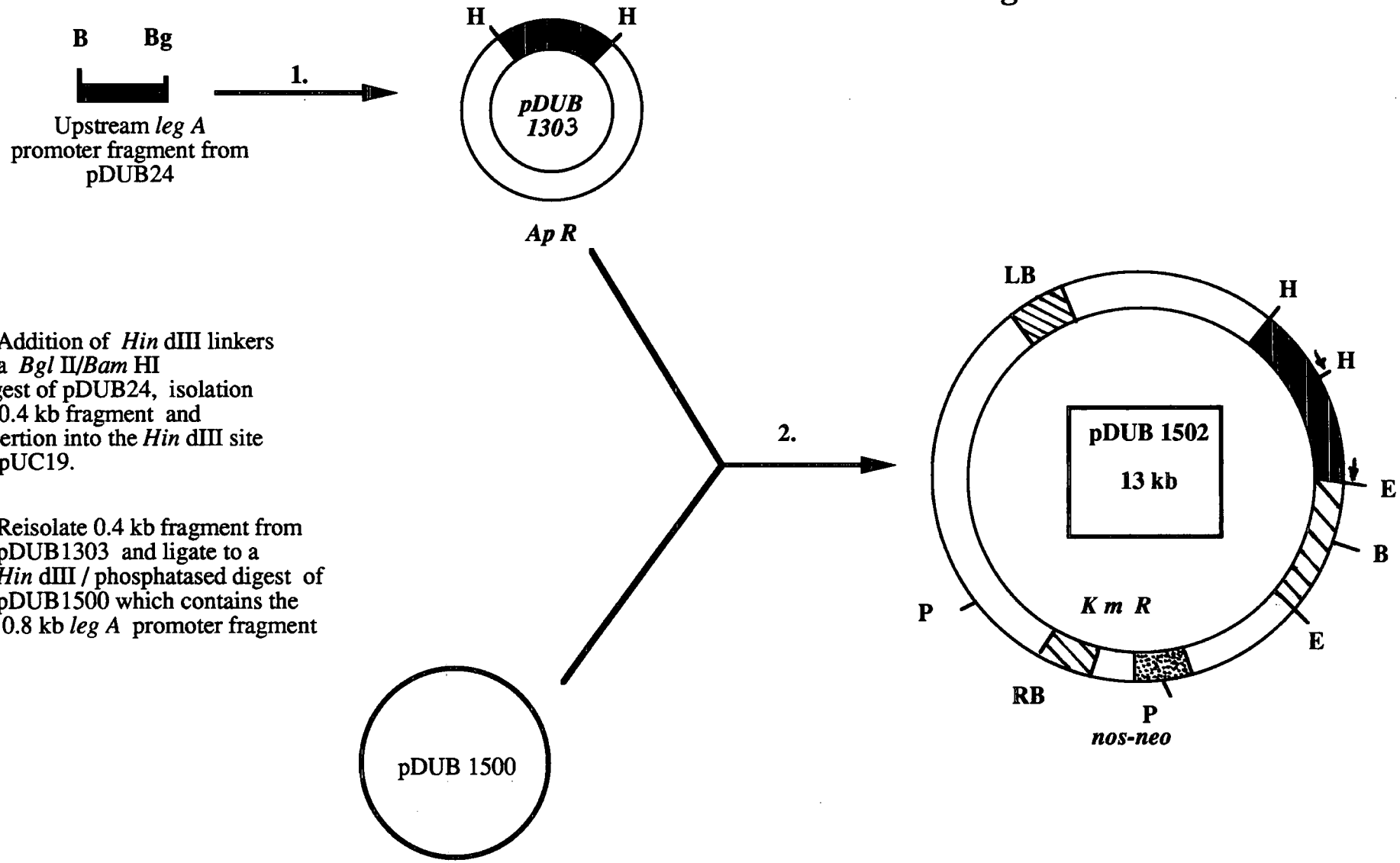

nos
B
E P
Leg A
promoter
P
Bg
E
Ap r
pDUB 1122
8.6 kb
RB
H
1. Pst I fragment from pDUB1122 isolated and ligated to pUC9
StuI
nos
B
P
pDUB1302
5.6 kb
E
P
H
Ap R
2.
L B
H
P
E
pDUB 1501
12. kb
B
nos
P
E
RB
P
nos -neo
L B
pBin 19
10 kb
H
P
B
Sma
E
P
RB
P
nos -neo
2. Hind III /Stu I fragment from pDUB1302 inserted into the Hind III / Smal site of pBin19

3.25b Construction of pDUB1502

The 'full length' promoter fragment was reconstructed by the addition of *Hin*dIII linkers to a *Bgl*II/*Bam* HI fragment from pDUB24 and the insertion of this fragment into the *Hin*dIII site of the polylinker region of pDUB1500, which lies upstream of the 0.8 kb promoter fragment of this vector.

Sequences derived from the Ti-plasmid, namely the right (RB) and left border (LB) repeats and the nopaline synthase coding region (*Hin*dIII fragment 23 from pTiC58), are indicated by hatched boxes. The neomycin phosphotransferase gene under the control of the *nos* promoter is indicated by a stippled box and legumin promoter sequences by a black box. B=*Bam* HI; Bg=*Bgl*II; E=*Eco* RI; H=*Hin*dIII; P=*Pst* I.

Fig 3.25b


B
Bg
Upstream *leg A* promoter fragment from pDUB24
1.
H
H
pDUB 1303
Ap R
1. Addition of *Hin* dIII linkers to a *Bgl* II/*Bam* HI digest of pDUB24, isolation of 0.4 kb fragment and insertion into the *Hin* dIII site of pUC19.
2. Reisolate 0.4 kb fragment from pDUB1303 and ligate to a *Hin* dIII / phosphatased digest of pDUB1500 which contains the 0.8 kb *leg A* promoter fragment
pDUB 1500
2.
LB
H
H
pDUB 1502
13 kb
E
B
P
K m R
E
RB
P
nos-neo

der of the legumin promoter, was also synthesised (section 3.3.3) The pBin19 derivatives were mobilised from their *E. coli* host to the *Agrobacterium tumefaciens* strain, LBA4404(pAL4404) as described in section 2.7.4. Total DNA from bacteria showing the Rif/Sm/Km-resistant phenotype were Southern blotted and hybridised with suitable probes to confirm the presence of the pBin19 derivative in LBA4404.

These constructs were used to infect leaf discs of *Nicotiana tabacum* (section 2.12.2) and plants regenerated from Km-resistant shootlets. A variety of plant tissues, including callus, leaf and seed, were assayed for nopaline production (section 2.14). Genomic DNA was extracted from the transformed plant tissue as described in section 2.13, and analysed for unrearranged insertion of the chimaeric *leg-nos* genes by probing Southern blots (section 2.5).

3.3.2 Construction of subclones in pUC9 and pUC19

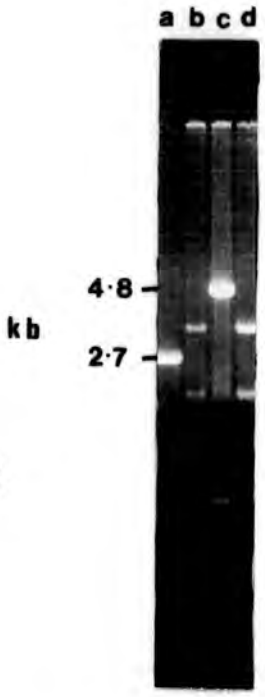
In order to construct pDUB1501, a subcloning step in pUC9 was used. A *Pst* I fragment from pDUB1122, which encompasses pBR322 sequences from the Ap^r gene, the entire *nos* coding region and *leg A* sequences up to the *Pst* I site at -95, was subcloned into the *Pst* I site in pUC9. The resulting plasmid, pDUB1302, was maintained in the rec^- *E.coli* strain DH5α to prevent homologous recombination between the sequences derived from pBR322 and the vector, pUC9. The *Pst* I fragment from pDUB1122 could combine with pUC9 in two orientations so a number of clones were screened by a *Hin*dIII/*Stu*I digestion of miniprep DNA. One recombinant plasmid was selected in which the orientation of the vector and insert were such that a fragment containing the legumin and nopaline synthase sequences could be excised with *Hin*dIII and *Stu* I, generating a fragment of approximately 2 kb (fig 3.26). The purpose of the subcloning step was essentially to convert the *Pst* I site at -95 in the *leg A* promoter to a *Hin*dIII site using the polylinker from pUC9 There are three *Pst* I sites in pBIN19 but *Hin*dIII is a unique site in the disarmed vector.

3.26 Restriction analysis of putative pDUB1302 isolates

A *Pst* I fragment from pDUB1122 was subcloned into pUC9. The orientation of insertion was determined such that a *leg-nos* fragment spanning 2 kb could be excised by *Hin*dIII/*Stu*I digestion. Miniprep DNA was digested with *Hin*dIII and *Stu*I. Lanes b and d show the correct orientation, giving characteristic bands of 2.0 and 3.6 kb. The opposite orientation produces fragments of 0.8 and 4.8 kb

Key: lane (a) pUC9 *Pst* I; (b) - (d) pUC9 subclones of the pDUB1122 *Pst* I fragment digested with *Hin*dIII and *Stu*I.

FIG 3·26



b & d = pDUB 1302

3.27 Restriction analysis of putative pDUB1303 isolates

*Hin*dIII linkers were ligated to the 0.4 kb fragment from pDUB24, *Hin*dIII digested and gel purified. Following ligation to *Hin*dIII cleaved pUC19 and transformation, miniprep DNA obtained from a number of white colonies was digested and run on an agarose gel. Isolates 1, 3 and 4 appear to have the correct insert.

Key: lanes 1-5. pUC19 subclones digested with *Hin*dIII; 6. λ *Pst* I.

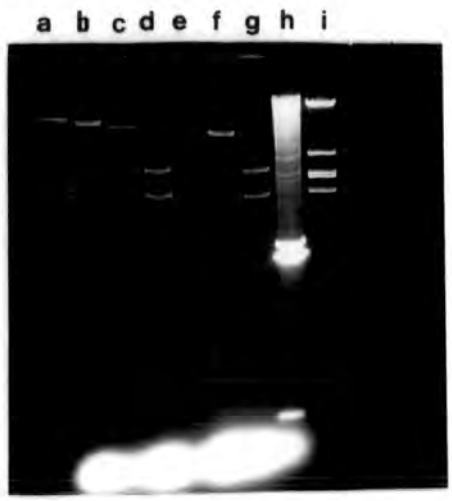
Fig 3·27

Fig 3.28 Restriction analysis of pDUB1500 and pDUB1501

Plasmids pDUB1500 and pDUB1501 were digested with a range of restriction enzymes. pDUB1500 contains a 0.6 kb *Pst* I fragment absent in pDUB1501. An *Eco* RI fragment of 2 kb which corresponds to the *nos* coding region is common to both vectors.

Key: lane (a) λ *Pst* I; (b) pDUB1500 *Bam* HI; (c) pDUB1500 *Eco* RI; (d) pDUB1500 *Pst* I; (e) pDUB1501 *Bam* HI; (f) pDUB1501 *Eco* RI; (g) pDUB1501 *Pst* I; (h) pDUB1303 *Hin*dIII; (i) λ *Eco* RI.

Fig 3 · 28



The constructs containing the full-length legumin promoter were synthesised in two steps. First a disarmed vector containing 0.8 kb of the legumin promoter (pDUB1500) was constructed (see section 3.3.3) and then a 0.4 kb sequence was inserted immediately upstream. This 0.4 kb sequence was derived from a *Bgl* II/*Bam* HI fragment of pDUB24 which contains promoter sequences upstream from the *Bgl* II site at -829. This was subcloned into pUC19 after the addition of *Hin*dIII linkers to give pDUB1303 (fig 3.27).

3.3.3 Insertion of *leg-nos* sequences into pBIN19

The disarmed vector, pDUB1501 (Leg 0.1) was constructed by insertion of the *Hin*dIII/*Stu I* fragment from pDUB1302, which contains 0.1 kb of the legumin promoter fused to the nos coding region, into the *Hin*dIII/*Sma* I sites within the polylinker of pBIN19 (fig 3.25). Another vector, pDUB1500 (Leg 0.8) was constructed by the insertion of the 2.8 kb *Bgl* II/*Stu* I fragment from pDUB1122 into the *Bam* HI and *Sma* I sites of the pBin 19 polylinker (fig 3.24). These constructs were checked by restriction mapping, the larger of these constructs gives an additional *Pst* I fragment corresponding to the legumin promoter upstream of the *Pst* I site at -95 (fig 3.28).

The vectors, pDUB1502 and pDUB1503, were obtained by the insertion of the 0.4 kb *Hin*dIII fragment from pDUB1303 into the polylinker immediately upstream of the promoter fragment of pDUB1500, utilising the *Hin*dIII site adjacent to the former *Bam* HI site. The *Hin*dIII digest of pDUB1500 was phosphatased (section 2.3.6) prior to ligation to ensure efficient selection of recombinant plasmids. An assymmetric *Nco* I site within the 0.4 kb insert allowed the orientation of insertion to be deduced (fig 3.29). The plasmid with the two pLEG fragments in phase was designated, pDUB1502 and the opposite orientation was designated, pDUB1503.

3.3.4 Integration of deleted *leg-nos* genes into the tobacco genome.

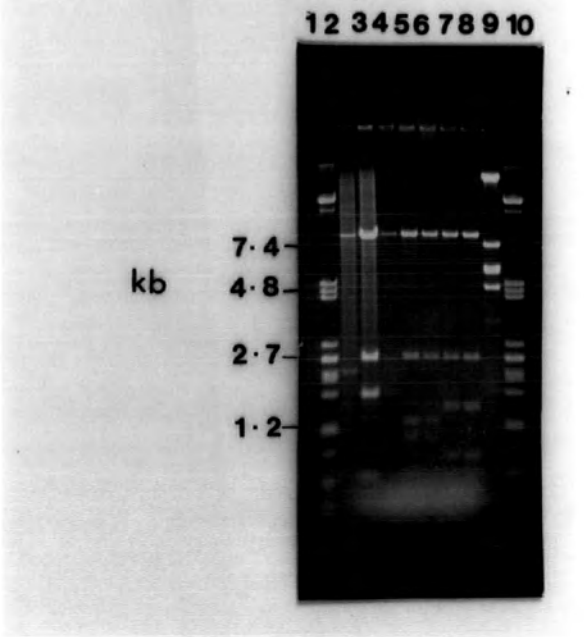
The structures of the four chimaeric *leg-nos* genes in pBIN19 are given in

3.29 Restriction analysis of pBIN19 derivatives

The different legumin promoter fragments in the four disarmed vectors constructed could be demonstrated by *Nco*I digestion. Plasmids resulting from the ligation of the *Hin*dIIIfragment of pDUB1303 to pDUB1500 were further characterised by *Nco*I digestion so that the correct orientation of the promoter fragment relative to the *nos* coding region could be identified. The potentially functional orientation was designated pDUB1502 and the reverse, PDUB1503.

Key: Lane 1. λ *Pst* I; 2. pBIN19 *Nco*I; 3. pDUB1501 *Nco*I; 4. pDUB1500 *Nco*I; 5. pDUB1502 *Nco*I; 6. pDUB1502 *Nco*I; 7. pDUB1503 *Nco*I; 8. pDUB1503 *Nco*I; 9. λ *Eco* RI; 10. λ *Pst* I.

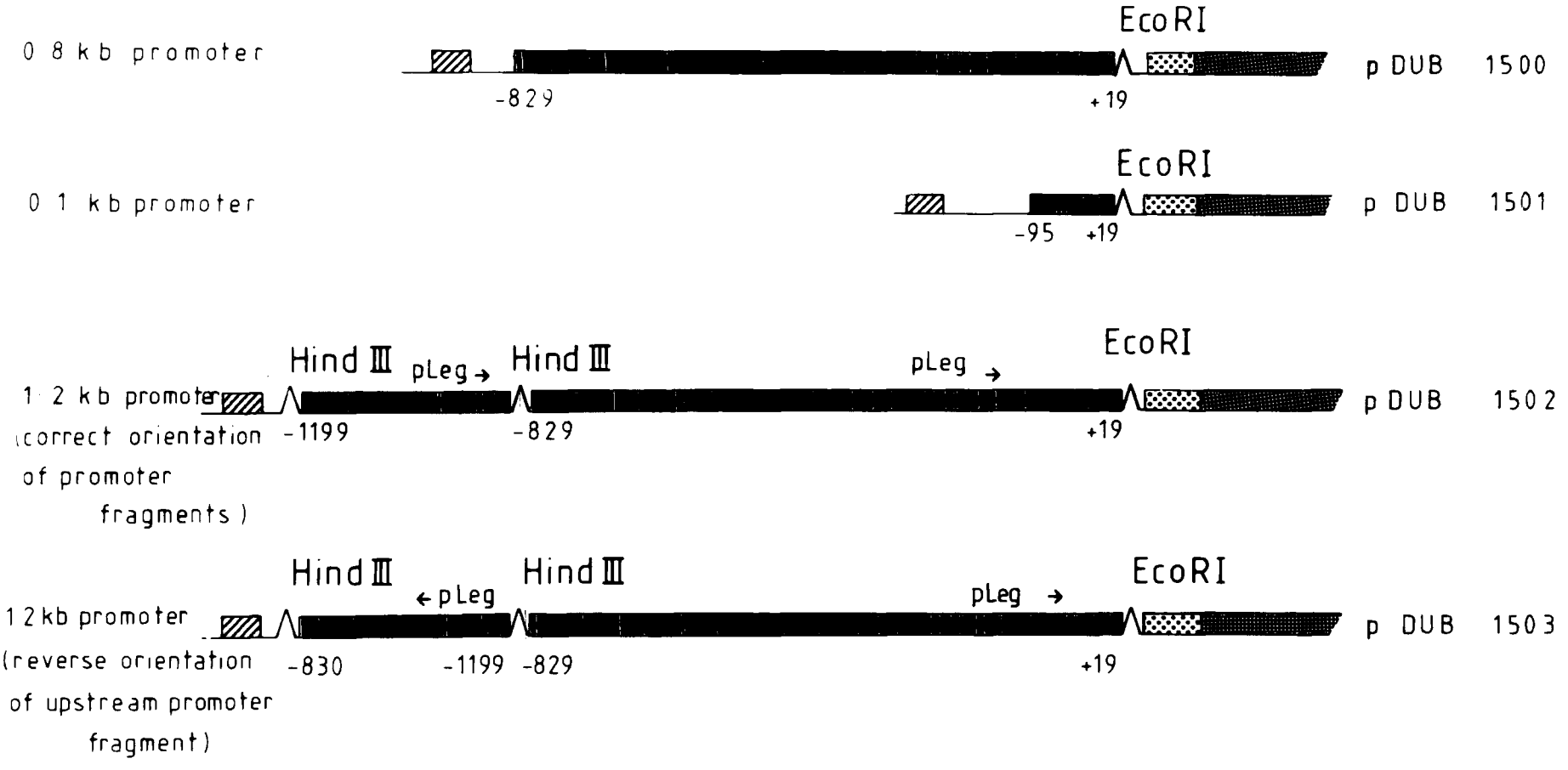
Fig 3·29



3.30 Structure of disarmed *leg-nos* constructs

The four disarmed *leg-nos* constructs are shown, indicating the location of the *Eco* RI and *Hin*dIII linkers, the right border sequences and the translation start site for nopaline synthase. The key is given previously in fig 3.21 and the plasmid numbering is outlined in table 2.1.

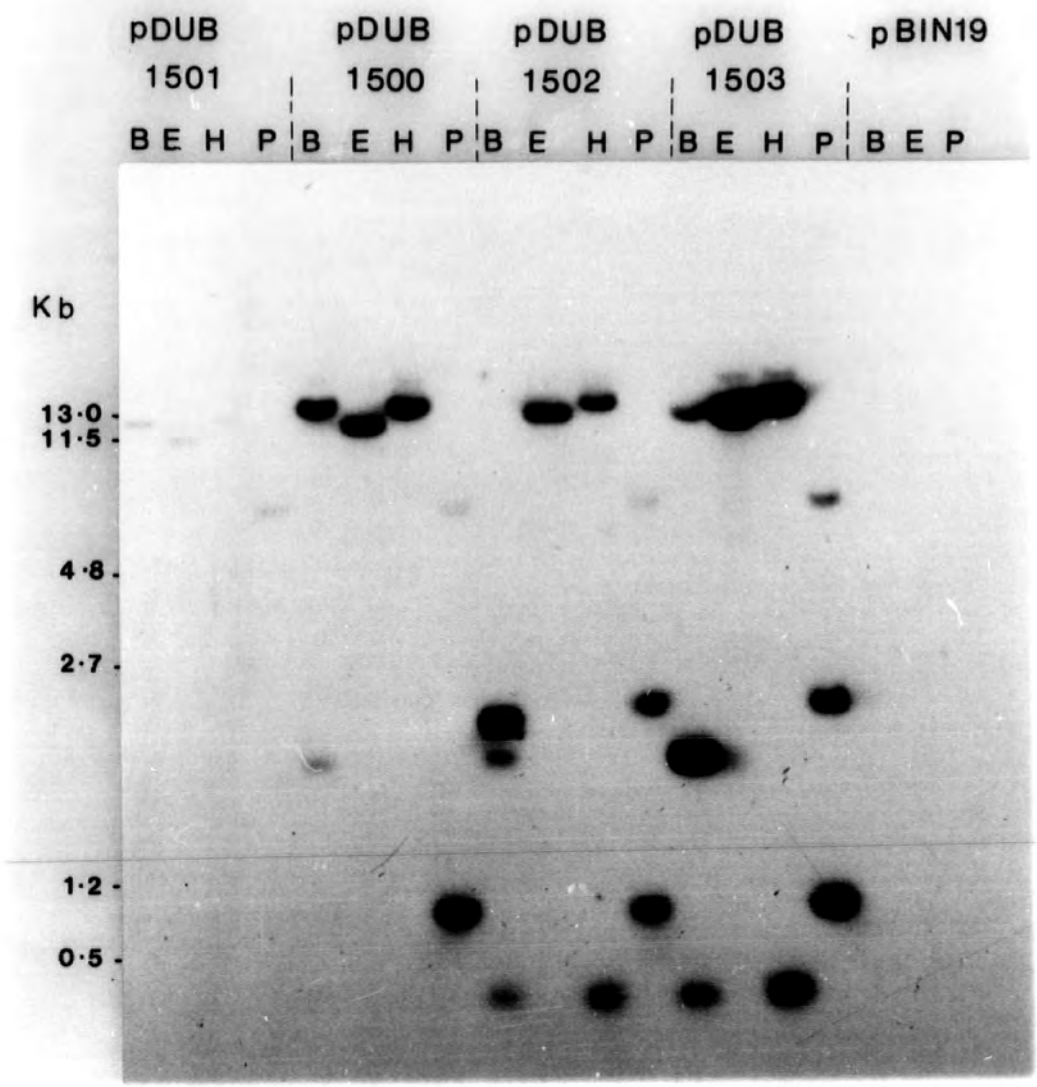
Fig 3·30 Structure of disarmed leg—nos constructs



3.31 Southern blot of total *Agrobacterium* DNA from disarmed vectors probed with LEG 1.2

Agrobacterium total DNA from lysed cells was probed with a ^{32}-P labelled promoter fragment from *leg*A. 10μg of total DNA from individual isolates was digested with *Bam* HI (B), *Eco* RI (E), *Hin*dIII (H) or *Pst* I(P). These digests were run on agarose gels, Southern blotted onto nitrocellulose and probed with labelled DNA. As a negative control, total DNA from pBIN19 was also run on the gel. The plasmid DNA gave the expected hybridisation pattern, pDUB1501 with only 0.1 kb of *leg*A sequences gives only faint hybridisation.

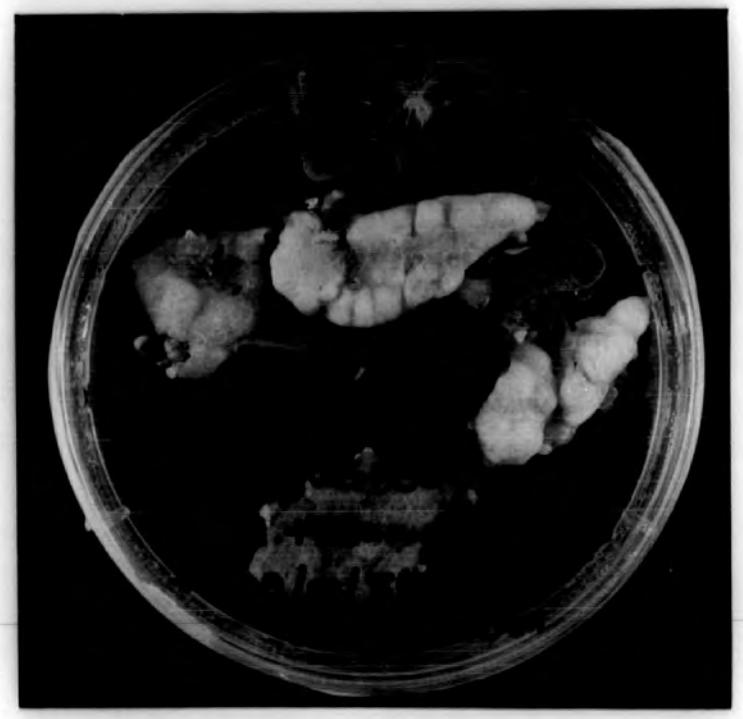
Fig 3·31


pDUB 1501
pDUB 1500
pDUB 1502
pDUB 1503
pBIN19
B E H P B E H P B E H P B E H P B E P
Kb
13·0
11·5
4·8
2·7
1·2
0·5

3.32 Kanamycin resistant shootlets derived from transformed tobacco tissue

Shootlets were obtained after 4-6 weeks on MS medium supplemented with sucrose (5g/l), carbenicillin (500mg/l) and kanamycin (200mg/l).

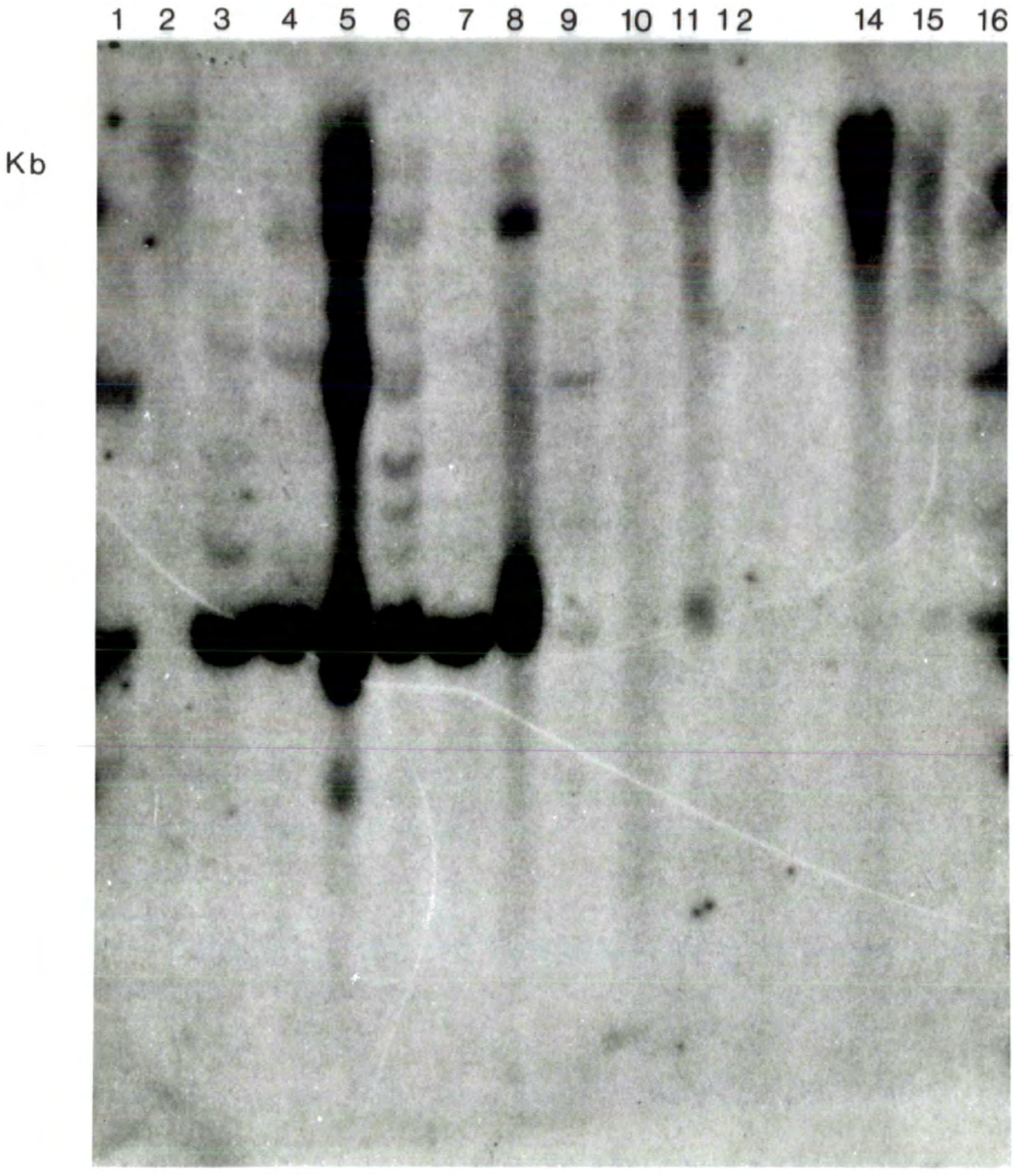
Fig 3·32



3.33 Southern blot of DNA from transgenic plants probed with *nos*

DNA from 15 transgenic plants, transformed with pDUB1500, was prepared by the Dellaporta method (section 2.13) and digested with *Eco* RI. These digests were run on an agarose gel to check for complete digestion and to allow transfer of the DNA to nitrocellulose. The Southern blot was probed with a labelled *Eco* RI/*Stu*I fragment from pDUB1111 which corresponds to the *nos* coding region. The strongly hybridising band seen with plants 1, 3,4,5,6,7 and 8 corresponds to this 2.0 kb fragment. Fainter hybridisation is also seen in some of the other tracks.

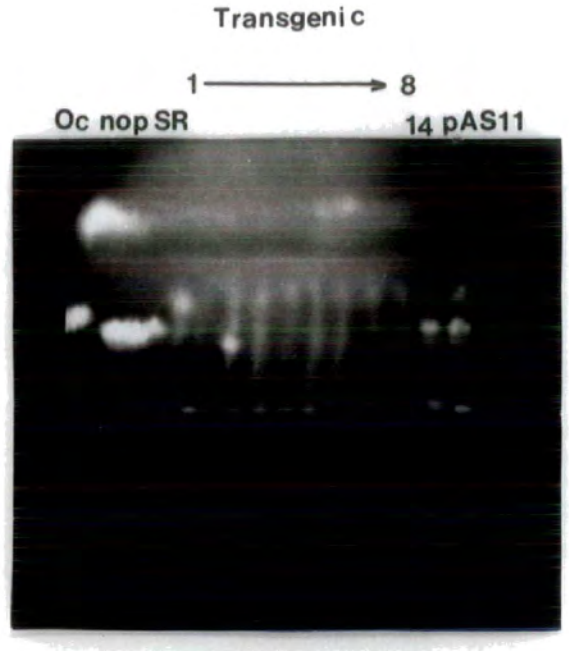
Fig 3·33 Transgenic plants



3.34 Opine assays on leaf tissue from transformed SR1

The presence of nopaline in tissues from transgenic SR1 transformed with pDUB1500 was sreened. As controls, tissue from untransformed SR1 and a cell line producing *nos* constitutively were used. Extracts from sixteen regenerated plants were made and run beside nopaline (nop) and octopine (oc) standards.

Fig 3·34


Transgenic
1
8
Oc nop SR
14 pAS11

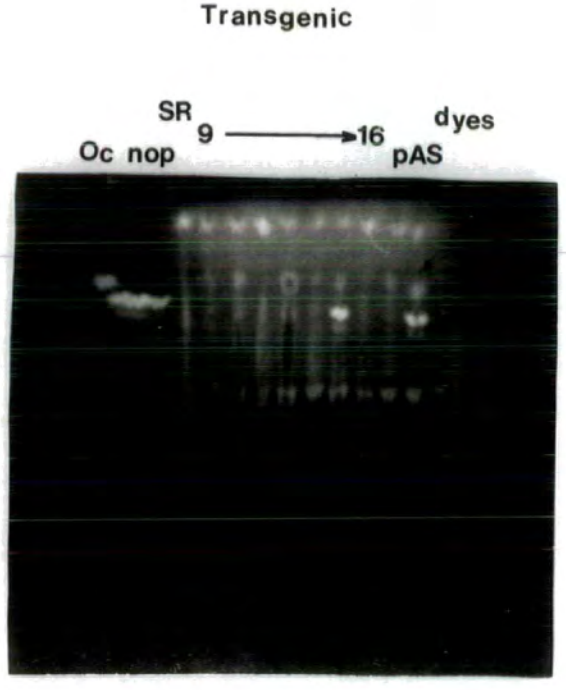

Transgenic
SR
9
16
dyes
Oc nop
pAS

fig 3.30. The disarmed vectors constructed were mobilised from *E.coli* to *Agrobacterium tumefaciens* in a triparental mating (Section 2.7.4). LBA4404, harbouring these plasmids were isolated on L-agar containing rifampicin, streptomycin and kanamycin and verified by Southern blotting (fig 3.31). The resulting strains were used to transform *Nicotiana tabacum* leaf tissue and the kanamycin resistant shootlets obtained (fig 3.32) were regenerated into mature plants as described (section 2.12.12) Sixteen tobacco plants showing kanamycin resistance were obtained following infection of leaf discs with pDUB1500, which contains 0.8 kb of the leg *A* promoter.

Initially six plants were selected and genomic DNA isolated by the method of Graham (section 2.13.1). In order to screen rapidly the remaining plants, a miniprep method for preparing genomic DNA was used (section 2.13.2) and the DNA cleaved with *Eco*RI. Southern blots were probed with a 2.0 kb *Eco*RI/*Stu*I fragment which encompasses the entire *nos* coding region. Of the fifteen plants analysed in this way, DNA from eleven plants showed the expected band pattern after hybridisation to a probe for the *nos* coding region (fig 3.33). The regeneration of transformed tobacco plants from the other disarmed constructs generated only a few specimens which failed to flower and analysis of the integration of the binary vector was not carried out.

3.3.5 Analysis of transgenic tobacco

(donated by A. Hepher)

A callus line of *Nicotiana tabacum*, which had been transformed with pAS11, a plasmid containing a nopaline synthase gene homologous to pDUB1003Δ31, was used to regenerate fertile plants to function as a positive control for nopaline production. High levels of opine production were observed in the leaves of regenerated plants (fig 3.34). Similar results were observed with seeds except that this material produced a high level of background fluoresence which made interpretation difficult. The tissues from untransformed SR1 was used as a negative control.

Callus, leaf and seed tissues from plants transformed with pDUB1500 were

assayed for nopaline production. No nopaline was detected in leaf tissue (fig 3.34) from these plants with the exception of one plant (14). Analysis of seed harvested after self-pollination was carried out and only pAS11 and plant 14 showed any nopaline production (data not shown).

3.4 SUMMARY

Nopaline production was not detected in callus tissue with the exception of a number of individual transformants with 1.2 kb of *leg A* upstream sequences. There was no evidence to suggest that seed-specific expression can be successfully achieved with 0.8 kb of the *leg A* promoter region, when fused to the *nos* gene, in transgenic tobacco.

CHAPTER 4

DISCUSSION

4.1 DISCUSSION OF RESULTS

4.1.1 Design and construction of vectors

The oncogenic Ti-plasmids were used to elucidate possible silencer sequences in the *leg A* promoter and required numerous sub-cloning and cointegration steps. Negative controlling elements have been identified in the 5′ upstream region of the *cab* gene from *Pisum sativum* (Simpson *et al.*, 1986b) and *Nicotiana plumbaginifolia* (Castresana *et al.*, 1988). In contrast the pBIN19-derived vectors involved fairly simple ligation steps which were easily verified by restriction analysis. This set of vectors was used to test for tissue-specific expression in regenerated *Nicotiana.*

The initial plasmid manipulations were carried out in *E.coli* and verified by restriction mapping. The construction of the oncogenic vectors involved cointegration and resolution of homologous sequences in pDUB1006 and the pGV1106-derivatives (pDUB1207,1215,1216,1217). The retention of the introduced sequences from the cointegrating vectors was verified by probing Southern blots of digested DNA from bacterial isolates demonstrating the desired pattern of antibiotic resistance and sensitivity.

The loss of the antibiotic markers, involving a second crossover event, was distinguished from the loss of the Ti-plasmid by selecting for opine catabolism. However there is some evidence that the cured strain was able to grow on the selective media despite the absence of a supplied nitrogen source. Therefore the crossover event was verified by Southern blotting of a number of independent colonies obtained on opine supplemented nitrogen-free media, after a number of enrichment steps in the liquid media.

The ability of the resulting plasmid constructs obtained to produce tumours provided further evidence that the restriction fragments corresponding to the chimaeric gene constructs were integrated into pDUB1006 rather than present as an autonomous replicating units. However, although no gross rearrangements were

observed, small changes in individual bases would not be detected. It would have been feasible to sequence the original plasmid constructions in pDUB1111, in particular the crucial region of the legumin and nopaline synthase gene fusion, but mutations occuring at a later stage could not be ruled out.

4.1.2 Nopaline synthase as a reporter gene

The various chimaeric *leg-nos* constructs failed to synthesise sufficient levels of the enzyme, nopaline synthase, to produce quantities of nopaline comparable to the levels observed with the positive controls, pDUB1003Δ31 in callus tissue, or the homologous construct pAS11 in regenerated tobacco tissues.

In *Kalanchoë* tumour tissue, low levels of nopaline were observed with one particular isolate of pDUB1016, containing the full-length legumin promoter, but similar results were not obtained with three apparently identical strains. The other oncogenic vectors containing deleted *leg-nos* genes did not produce nopaline in the tumours examined. Similarly, analysis of callus tissue obtained during the transformation of *Nicotiana tabacum* using the homologous binary vectors, failed to demonstrate opine production.

In transgenic *Nicotiana tabacum* SR1 plants, regenerated from tissue transformed with pDUB1500, which contains 800 bp of 5′ flanking sequences from the *leg A* gene, no evidence for opine production in any plant tissue was normally obtained. The only regenerated plant which produced nopaline, apart from the positive controls, gave an abberrant pattern on Southern blots and showed a constitutive type of gene expression. Unfortunately, there was not time to regenerate flowering plants from tissue transformed with the full-length *leg A* promoter so the possibility that factors upstream of the *Bgl* II site at -829 are required for high levels of gene expression cannot be excluded.

Normally opines are produced in response to *Agrobacterium tumefaciens* infection and have been used as an indicator of integration of the T-DNA into the plant genome. These opines have generally not been detected in wild-type tis-

sues (Biemann *et al.*, 1960; Braun *et al.*, 1962; Holderbach and Biederbeck, 1976; Lioret, 1956; McKee, 1962). However, there have been a few isolated reports of the detection of opines in normal plant tissues (Johnson *et al.*, 1974; Seitz and Hochster, 1964; Wend-Gallitelli and Dobrigkeit, 1973) although one report was subsequently retracted (Montoya *et al.*, 1977).

These opines have generally been detected by the paper electrophoretic technique of Otten and Schilperoort (1978) which involves pre-incubation of the tissue with arginine. In a more recent publication by Christou *et al.* (1986), significant levels of nopaline was detected in soybean leaf and callus and small amounts in tissue from cotton. A structurally similar guanido-compound, acetopine which is an acetylated derivative of arginine, was detected in tobacco, soybean and cotton tissue. During electrophoresis, acetopine migrates between the spots corresponding to nopaline and arginine, so do not pose a problem in identification. If the tissues were not treated with arginine these compounds were barely detectable (Christou *et al.*, 1986.). The nopaline assays described in this work were also carried out without pre-incubation with arginine. However in order to detect low levels of expression in seed tissues, an extract from 40 seeds was loaded. Following electrophoresis and staining a background smear stretching from the origin was observed in both transformed and untransformed seed tissue which may have masked very low levels of nopaline production.

The nopaline synthase marker gene system may have been insuffuciently sensitive to detect low levels of expression. The detection of nopaline in seed from *Nicotiana tabacum* transformed with pAS11$_A$ by Dr A. Hepher indicates that it is possible to detect *nos* gene expression using the enzyme product as an indicator in this plant tissue. Nopaline was detected in seed extracts without preincubation with arginine, which suggests that although the seed is actively synthesising storage proteins from the amino acid pool in this tissue, the levels of arginine were not limiting for nopaline synthase activity.

Although the nopaline assay has the advantage of being quick, and inexpensive

to perform, it is probably not as sensitive as some other reporter genes currently in use. Different results were obtained by authors mapping the nopaline synthase promoter using *nos*as a marker gene (Shaw *et al.*, 1984a) compared to those using CAT (An *et al.*, 1986a). It is also difficult to quantitate the results. However, nopaline production was detected in leaf and seed of the transgenic tobacco plants which served as a positive control, so the failure to detect nopaline production may not be a fault of the marker gene used. Indeed one particular plant transformed with pDUB1500 produced nopaline in all tissues analysed which suggests that the nopaline synthase gene was potentially functional, although it may be only after some, as yet undefined rearrangement, or possibly the fortuitous proximity of an enhancer element in the plant genome.

When promoters from monocot storage protein genes have been used to direct expression in dicot tissues, the often low levels of expression observed can make the marker gene system employed crucial to subsequent detection. Schernthaner and associates (1988) failed to detect expression of a 23kd zein gene (Z4) in transgenic tobacco, but were able to demonstrate low levels of expression with some plants transformed with a Z4-promoter-CAT fusion gene and in all transformed tobacco with a Z4-GUS fusion. Therefore, GUS gene fusions appear to provide sensitivity without using radioactive products. However, in a recent paper (Plegt and Bino, 1989), tissue-specific differences in endogenous glucuronidase activity during flower development were observed.

4.2 COMPARISON OF RESULTS WITH OTHER STORAGE PROTEIN GENES

Ellis *et al.*, (1988), using a 3.4 kb genomic fragment containing the *leg A* gene, including 1.2 kb of 5′ flanking sequences, have demonstrated accumulation of the legumin protein in the seeds of transgenic *Nicotiana plumbaginifolia* using an ELISA detection system. The vector used was pBIN19 containing both the intact *leg A* gene and the marker gene, nopaline synthase. Of the 50 kanamycin plants

assayed for nopaline and immuno-detectable legumin, only three appeared to possess functional copies of both genes. About half of these plants gave a positive result using an ELISA for pea legumin but only ten plants gave a signal on the nopaline assay. The non-functional genes were associated with rearrangement but in at least one plant line, a rearranged legumin gene produced detectable legumin protein. These authors reported a failure to correlate gene copy number with the levels of expression obtained.

In subsequent work (Shirsat *et al.*, 1989) with a series of promoter deletions these workers identified 549 bp of flanking sequences required for tissue-specificity and temporal regulation, although elements further upstream were required for high level expression. Considerable variation between individual transformants was observed. Binding of nuclear proteins from pea seed has been correlated with these findings (Shirsat *et al.*, 1990) but no binding to the legumin box was demonstrated.

The expression of phaseolin and zein genes in callus tissue has been observed using transformed sunflower, tobacco and petunia. (Matzke et al. 1984; Murai et al 1983) Although one isolate of pDUB1016 showed reproducible levels of nopaline, albeit at a lower level than the controls, a number of other similar isolates failed to elicite nopaline production. This may be explained be some alteration in the chimaeric gene that was not detected by the Southern blots carried out. The other oncogenic constructs, containing deleted *leg-nos* genes failed to produce nopaline in the tumour tissues examined.

The detection of seed-specific expression has been achieved in seed promoter fusions with a number of marker genes, including octopine synthase, CAT and *npt II*. Only 863 bp of flanking sequences were required for the organ-specific expression of the phaseolin gene in transgenic tobacco plants (Sengupta-Gopalan *et al.*, 1985) although there was evidence for degradation of the protein product. Sequences between -131 and -257 of the soybean protein β-conglycinin have been implicated in the high level of organ-specific expression and correct tissue-

specificity is maintained on further deletion to -159, albeit at a low level (Chen *et al.*, 1986).

Seed proteins from non-legumes, such as barley and wheat have been succesfully expressed in a heterologous system using a dicotyledonous host. This suggests that there are similar regulatory signals even in such evolutionally diverged species. A region of the 5′ flanking sequence, from -337 to -125, of a 19kd zein protein is sufficient for maximal, tissue-specific expression of a fused CAT gene in carrot and sunflower but not in tobacco (Roussell *et al.*, 1988). The 5′ flanking sequences from a B-hordein gene are able to confer tissue-specific expression of a fused CAT gene in transgenic tobacco (Marris *et al.*, 1988). Similarly a promoter fragment from a LMW glutenin gene of wheat, spanning 160 to 326 bp upstream of the transcription start site will direct tissue-specific expression of a fused CAT gene in transgenic tobacco (Colot *et al.*, 1987).

These regions of seed-specific promoters are very close to the essential promoter motifs, namely the CAAT and TATA boxes and dissecting tissue-specific elements from those sequences required for transcriptional initiation would appear to be problematical. The upstream elements that have been fused to constitutive genes has generally resulted in a tissue-specific pattern of regulation. However, these results suggest that 800 bp of the legumin A flanking sequence should have been sufficient to direct seed-specific expression.

4.3 FACTORS INFLUENCING GENE EXPRESSION IN TRANSFORMED PLANTS

Many factors control the levels of expression of introduced genes in the plant genome, including position effects, methylation and the unrearranged insertion of the foreign DNA. The precise fusion of the legumin and nopaline synthase sequences may have introduced secondary structures that interfere with the initiation of transcription. Results from a wide range of chimaeric gene fusions in transgenic plants indicate that lower levels of expression are observed compared

to the native genes, particularly in a heterologous system. Small deletions or rearrangements in the coding or promoter components of the chimaeric genes may have occurred during the various subcloning steps, which were not distinguishable by restriction analysis. The copy number may influence the level of expression although in many cases this correlation was not observed, suggesting position effects.

4.3.1 Unrearranged insertion of introduced genes

In a comparison of binary and cointegrate Ti-plasmid vectors by Spielmann and Simpson (1987), many reports of rearrangements of the T-DNA structures were indicated. The workers studying legumin expression in transgenic tobacco using pBIN19 reported a high incidence of rearrangements (Ellis *et al.*, 1988) which may have been partially caused by having two copies of the *nos* promoter in these constructions.

Point mutations which would be difficult to spot, except by laborious sequencing of all constructions, may have a profound effect if they occur in the coding region of the marker gene and produce a truncated or modified protein which is no longer functional. Similarly a mutation in an element of the promoter region could also have profound effects.

4.3.2 Transcriptional *versus* translational fusions

The use of a marker gene can increase the sensitivity of detection of gene expression from a plant promoter, which may be particularly important if the promoter deletions result in very low levels of expression. The detection of the protein product of an introduced storage protein gene in transgenic plants can be problematical if there are similar endogenous products in the host tissue, which makes the use of a reporter gene desirable. Similarly, if reintroducing promoter deletions into a homologous system, to distinguish between the endogenous gene and an introduced gene, fusions of promoter sequences with a reporter gene have been employed. However, the fusion of a plant promoter with a reporter gene

may introduce novel sequences or potentially secondary structures which interfer with normal transcription and translation. Additionally, the use of linkers and the restriction sites used in the chimaeric gene fusion often alters the spacing of the cap site relative to the AUG codon. In different plant genes a considerable range in this distance had been observed (Joshi, 1987). The precise context of the AUG codon is important for translation (Kozak, 1987).

Jones *et al.* (1985), using a fusion between the petunia *cab* promoter and the octopine synthase coding region, presented evidence that a translational fusion consistently gave a higher level of gene expression than a comparable transcriptional fusion. These authors suggested that in order to achieve high levels of expression of introduced genes, the sequence of the 5′-untranslated region should be as close to the original plant gene sequence as possible. Alterations in this region may exert deleterious effects on expression by increasing turnover of the messenger RNA by the destabilising effects of GC-rich regions. Alternatively, there may be regions downstream from the transcription start site that are important for the regulation of transcription initiation as has been found for the globin genes (Charnay *et al.*, 1984).

In this laboratory, Shaw and co-workers have generated a number of transcriptional fusions between various promoters and the nopaline synthase coding region and demonstrated expression. The *leg*-nos fusions reported in this work and other fusions using the nopaline synthase gene (Shaw *et al.*, 1986) have altered the spacing between the cap site and the AUG codon as well as possessing two potential transcriptional initiation sites. However, in most cases the associated nopaline synthase activity has been demonstrated although the prefered transcriptional initiation site has not been determined by these workers.

4.3.3 Position effects of introduced genes

The occurrence of position effects on foreign genes has been reported in most transgenic organisms. The site of integration greatly influenced the level of ex-

pression of introduced genes in transgenic mice (Palmiter and Brinster, 1985) and in *Drosophila* (Bourouis and Richards, 1985). There is also one report of an introduced rabbit β-globin gene being expressed in an unexpected tissue in transgenic mice (Lacey *et al.*, 1983).

Some authors have reported differences in the level of expression of introduced genes which cannot be explained by the number of copies detected by Southern hybridisation (An, 1986; Czernilofsky *et al.*, 1986; Jones *et al.*, 1985). This has been observed with both calli and regenerated plants, up to a 200 fold difference has been reported between individual calli and even two distinct genes on the same transfered T-DNA may show little relationship in their levels of expression (An, 1986).

Analysis of a number of transgenic petunia plants carrying one or two copies of the pea *rbc-E9* gene revealed differences of between 25 and 50-fold in the ratio of levels of mRNAs for this gene compared to the contransfered *nos-npt*II gene (Nagy *et al.*, 1985). Jones *et al.* (1985) carried out a number of experiments to compare the level of expression of a number of transcriptional and translational fusions of the petunia *cab* promoter and the *ocs* gene using a cointegrated Ti-plasmid vector to transform both petunia and tobacco. Most transformed plants had between 1-5 copies of the gene, but one transformant which relative low levels of *ocs* mRNA had greater than ten copies in the genome. Individual transformed plants showed a 200-fold variation in their accumulated levels of chimaeric mRNA (Jones *et al.*, 1985). In these experiments a *nos* gene was also cotransfered and independent variation observed in the level of expression of these two genes, about 10% of the Ocs^+ plants showed no nopaline production and more rarely the ocs^- nos^+ phenotype (2/100).

Beachy *et al.* (1985) did not report similar effects when the α subunit of β-conglycinin was used to transform petunia. Four independent transformants showed nearly identical amounts of an accumulated bean protein.

4.3.4 Methylation of genes introduced into the plant genome

Plant DNA contains a high content of methylated cytosine residues, preferentially at sites containing the nucleotides CG or C-X-G (Gruenbaum *et al.*, 1981). Unusual phenotypes following transformation with wild-type Ti-plasmids can sometimes be attributed to the loss of expression of a particular T-DNA transcript caused by methylation (Hepburn *et al.*, 1983). These authors reported a flax tumour line containing 22-24 copies of the T-DNA but only showing very low levels of nopaline synthase expression, which correlated with the methylation level of the DNA. This occurs indepedently of the methylation state of the surrounding plant DNA.

4.3.5 Effects of the host tissues

Although, Ellis *et al.* (1988) reported that the legumin protein deposited in seeds of transformed *N. plumbaginifolia* was correctly processed and undegraded, other workers have reported partial hydrolysis of the protein products of introduced genes. The vicilin-like proteins, phaseolin and β-conglycinin showed some degradation to smaller peptides in transgenic plants (Beachy *et al.*, 1985; Sengupta-Gopalan *et al.*, 1985) However, the levels of nopaline synthesised in the tissues of transgenic plants containing an intact *nos* gene would indicate that this was not a factor, although a lower level of gene expression may have produced insufficient enzyme to be detected.

The host plant can affect the level of expression. Nagy *et al.* (1985) reported that the level of expression of the pea *rbcS*-E9 gene is greater in transgenic petunia compared to tobacco (relative to an internal control). In transgenic petunia, the *rbcS*-E9 gene is expressed at only 0.2-10% of the level found in pea seedlings.

4.4 CONCLUSION

In order for an introduced gene to be expressed in a transgenic plant, a number of conditions have to be satisfied. When chimaeric gene fusions are used, the

precise site of fusion may generate novel sequences that are incompatible with efficient transcription/translation. The introduced gene has to be integrated, without rearrangement or mutation into a portion of the genome where the gene is likely to be actively transcribed (position effects). The site of integration may also influence the susceptibility to methylation of the integrated gene sequences. The gene messenger RNA has to be stable in the tissue in which it is expressed, efficiently translated and the protein product undegraded if the enzyme activity is to be detected. The failure to achieve each of these conditions has been widely reported.

The advances in our knowledge of the control of gene expression in plants has focussed attention on DNA-binding proteins in the promoter domain. This approach is attractive because the experiments are quick to perform and do not require the time-consuming regeneration of fertile plants, the results obtained give a functional role for the identified regions and can pinpoint sequences that deserve further characterisation by techniques such as scanning linker deletions. In addition new reporter genes have been developed whose assays are relatively inexpensive, quick and safe to perform and have been shown to be more sensitive than the more traditional marker genes in some systems.

The dissection of controlling elements for expression in seed tissues have a number of potential applications. The expression of modified seed proteins or other proteins with a desirable nutritional status is an obvious application but the presence of multi-gene families would entail multiple insertions of the foreign gene sequences that were highly expressed in order to have any appreciable effect. Additionally legumes seeds can be viewed as a green fermenter (Croy and Gatehouse, 1985) producing high levels of proteins in the seed and requiring little in the way of exogenous energy inputs. Some seed proteins are undesirable and deletions might be affected by linking a strong seed-specific promoter to the anti-sense coding region of the deleterious protein.

REFERENCES

Allen, R. D., Nessler, C. L. and Thomas, T. L. (1985). Developmental expression of sunflower 11S storage protein genes. Plant Mol. Biol. **5**, 165-173.

Altenbach, S. B., Pearson, K. W., Meeker, G., Staraci, L. C. and Sun, S. S. M. (1989). Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol. Biol. **13**, 513-522.

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 tumor line. Mol. Gen. Genet. **197**, 437-446.

An, G. (1986). Development of promoter expression vectors and their use for analysis of differential activity of nopaline synthase promoter in transformed tobacco cells. Plant Physiol. **81**, 86-91.

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

An, G., Ebert, P. R., Yi, B-Y., and Choi, C-H. (1986a). Both TATA box and upstream regions are required for the nopaline synthase promoter activity in transformed tobacco cells. Mol. Gen. Genet. **203**, 245-250.

An, G., Watson, B. D. and Chiang, C. C. (1986b). Transformation of tobacco, tomato, potato, and *Arabidopsis thaliana* using a binary Ti vector system. Plant Physiol. **81**, 301-305.

An, G., Costa, M. A., Mitra, A., Ha, S-B. and Marton, L., (1988). Organ-specific and developmental regulation of nopaline synthase promoter in transgenic tobacco plants. Plant Physiol. **88**, 547-552.

An, G., Costa, M. A. and Ha, S-B. (1990) Nopaline synthase promoter is wound inducible and auxin inducible. Plant Cell **2**, 225-233.

Apel, K. (1979). Phytochrome-induced appearance of mRNA activity for the apoprotein of the light-harvesting chlorophyll a/b-protein of barley (*Hordeum vulgare*). Eur. J. Biochem.**97**, 183-188.

Aryan, A. P., An. G. and Okita, T. W. (1991). Structural and functional analysis of promoter from gliadin, an endosperm-specific gene of *Triticum aestivum L.* Mol. Gen. Genet. **225**, 65-71

Ashby, A. M., Watson, M. D. and Shaw, C. H. (1987). A Ti-plasmid determined function is responsible for chemotaxis towards the plant wound product acetosyringone. FEMS. Micro. Letts., **41**, 189-192.

Banerji, J., Rusconi, S. and Schaffer, W. (1981). Expression of the rabbit β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**, 299-308.

Bäumlein, H., Müller, A. J., Schiemann, J., Helbing, D., Manteuffel, R. and Wobus, U. (1988). Expression of a *Vicia faba* legumin B gene in transgenic tobacco plants: Gene dosage-dependent protein accumulation. Biochem. Physiol. Pflanzen. **183**, 205-210.

Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F. C. (1986). The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res. **14**, 2707-2720.

Beach, L. R., Spencer, D., Randall, P. J. and Higgins, T. J. V. (1985). Transcriptional and post transcriptional regulation of storage protein gene expression in sulphur-deficient pea seeds. Nucleic Acids Res. **13**, 999-1013.

Beachy, R. N., Chen, Z.-L., Horsch, R. B., Rogers, S. G., Hoffmann, N. J. and Fraley, R. T. (1985). Accumulation and assembly of soybean beta-conglycinin in seeds of transformed petunia plants. EMBO J. **4**, 3047-3053.

Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B. and Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. Gene **19**, 327-336.

Benfey P. N., Takatsuji, H., Ren, L., Shah, D. M. and Chua, N. -H. (1990). Sequence requirements of the 5-enolpyruvylshikimate-3-phosphate synthase 5′- upstream region for tissue-specific expression in flowers and seedlings. Plant Cell. **2**, 849-856.

Benoist, C., O' Hare, K., Breathnach, R., Chambon, P. (1980). The ovalbumin gene - sequence of putative control regions. Nucleic Acids Res. **8**, 127-142.

Bergman, Y., Rice, D., Grosschedl, R. and Baltimore, D. (1984). Two regulatory elements for immunoglobulin κ light chain gene expression. Proc. Natl. Acad. Sci. USA **81**, 7041-7045.

Bevan, M. W. (1984). Binary *Agrobacterium* vectors for plant transformation. Nucleic Acid Res. **12**, 8711-21.

Bevan, M. W. and Chilton, M.-D. (1982). T-DNA of *Agrobacterium* Ti and Ri plasmids. Ann. Rev. Genet. **16**, 357-384.

Bevan, M. W., Flavell, R. B. and Chilton, M.-D. (1983a). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. Nature **304**, 184-187.

Bevan, M. W., Barnes, W. M. and Chilton, M.-D. (1983b). Structure and transcription of the nopaline synthase gene region of T-DNA. Nucleic Acids Res. **11**, 369-386.

Biemann, K., Lioret, C., Asselineau, J., Lederer, E. and Polonsky, J. (1960). On the structure of lysopine, a new aminoacid isolated from crown gall tissue. Biochim. Biophys. Acta **40**, 369-370.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**, 1513-23.

Blobel, G. and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol.

67, 835-851.

Bodnar, J. W. and Ward, D. C. (1987). Highly recurring sequence elements identified in eukaryotic DNAs by computer analysis are often homologous to regulatory sequences or protein binding sites. Nucleic Acids Res. **15**, 1835-1851.

Bomhoff, G. Klaplwijk, P. M., Kester, H. C. M., Schilperoort, R. A., Hernalsteens, J. P. and Schell, J. (1978). Octopine and nopaline synthesis and breakdown genetically controlled by a plasmid of *Agrobacterium tumefaciens.* Mol. Gen. Genet. **145**, 177-182.

Borroto, K. and Dure, L. (1987). The globulin seed storage proteins of flowering plants are derived from two ancestral genes. Plant Mol. Biol **8**, 113-131.

Botchan, M., Topp, W. and Sambrook, J. (1976). The arrangement of simian virus 40 sequences in the DNA of transformed cells. Cell **9**, 269-273.

Boulter, D. (1981). Biochemistry of storage protein synthesis and deposition in the developing legume seed. Adv. Bot. Res. **9**, 1-31.

Boulter, D., Evans, I. M., Ellis, J. R., Gatehouse, J. A. and Croy, R. R. D. (1987). Differential gene expression in the development of *Pisum sativum.* Plant Physiol Biochem **25**, 283-289.

Bourouis, M. and Richards, G. (1985). Remote regulatory sequences of the Drosophila glue gene *sgs*3 as revealed by P-element transformation. Cell **40**, 349-357.

Bown, D., Levasseur, M., Croy, R. R. D., Boulter, D. and Gatehouse, J. A. (1985). Sequence of a pseudogene in the legumin gene family of pea (*Pisum sativum L.*). Nucleic Acids Res. **13**, 4527-4538.

Braun, A. C. (1962). Tumor inception and development in the crown gall disease. Ann. Rev. Plant Physiol. **13**, 533-558.

Breathnach, R. and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. **50**, 349-383.

Broglie, R., Corrozzi, G., Fraley, R. T., Rogers, S. G., Horsch, R. B., Neidermeyer, J. G., Fink, C. L., Flick, J. S. and Chua, N. -H. (1984). Light-regulated expression of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene in transformed plant cells. Science **224**, 838-843.

Brown, J. W. S., Wandelt, C., Feix, G., Neuhaus, G. and Schweiger, H.-G. (1986). The upstream regions of the zein genes: Sequence analysis and expression in the unicellular green alga *Acetabularia*. Eur. J. Cell Biol. **42**, 161-170.

Bruce, W. B. and Gurley, W. B. (1987). Functional domains of a T-DNA promoter active in crown gall tumours. Mol. Cell. Biol.**7**, 59-67.

Bustos, M. M., Begum, D., Kalkan, F. A., Battraw, M. J. and Hall, T. C. (1991). Positive and negative cis-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J. **10**, 1469-1479.

Bustos, M. M., Guiltinan, M. J., Jordano, J., Begum, D., Kalkan, F. A. and Hall, T. C. (1989). Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a french bean β-phaseolin gene. The Plant Cell. **1**, 839-853.

Bytebier, B., Deboeck, F., De Greve, H., Van Montagu, M. and Hernalsteens, J. -P. (1987). T-DNA organization in tumor cultures and transgenic plants of the monocotyledon *Asparagus officinalis*. Proc. Natl. Acad. Sci. USA **84**, 5345-5349.

Casey, R. (1979). Genetic variability in the structure of the α-subunits of legumin from *Pisum* - a two dimensional gel electrophoresis study. Heredity **43**, 265-272.

Casey, R. and Short, M. N. (1981). Variation of amino acid composition of legumin from *Pisum*. Phytochemistry **20**, 21-23.

Casey, R., March, J. F., Sharman, J. E. and Short, M. N. (1981). The purification, N-terminal amino acid sequence and other properties of an α^M-subunit

of legumin from the pea (*Pisum sativum L.*). Biochim. Biophys. Acta **670**, 428-432.

Casey, R., Domoney, C. and Ellis, N. (1986). Legumin storage proteins and their genes. Oxford Surveys Plant Mol. Cell. Biol. **3**, 1-95.

Castresana, C., Garcia-Luque, I., Alonso, E., Malik, V. S. and Cashmore, A. R. (1988). Both positive and negative regulatory elements mediate expression of a photoregulated CAB gene from *Nicotiana plumbaginifolia.* EMBO J. **7**, 1929-1936.

Chandler. P. M., Higgins, T. J. V., Randall, P. J. and Spencer, D. (1983). Regulation of legumin levels in developing pea seeds under conditions of sulfur deficiency. Plant Physiol. **71**, 47-54.

Chang, C. and Meyerowitz, E. M. (1986). Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA **83**, 1408-1412.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. and Maniatis, T. (1984). Differences in human α- and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**, 251-263.

Chen, Z.-L., Schuler, M. and Beachy, R. N. (1986). Functional analysis of regulatory elements of a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA **83**, 8560-8564.

Chen, Z.-L., Pan, N.-S. and Beachy, R. N. (1988). A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. EMBO J. **7**, 297-302.

Cheung, S., Arndt, K. and Lu, P. (1984). Correlation of *lac* operator DNA imino proton exchange kinetics with its function. Proc. Natl. Acad. Sci. USA **81**, 3665-3669.

Chilton, M. -D., Drummond, M. D., Merlo, D. J. and Sciaky, D. (1978) Highly

conserved DNA of Ti-plasmids overlaps T-DNA maintained in plant tumours. Nature **275**, 147-149.

Chilton, M. -D., Drummond, M. D., Merlo, D. J., Sciaky, D., Montoya, A. L., Gordon, M. P. and Nester, E. W. (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell. **11**, 263-272.

Chrispeels, M. J. and Sadava, D. (1977). Plants, food and people. Freeman, W. H. and Co. (San Fransisco, USA).

Chrispeels, M. J., Higgins. T. J. V., Craig, S. and Spencer, D. (1982a). Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. J. Cell Biol. **93**, 5-14.

Chrispeels, M. J., Higgins. T. J. V. and Spencer, D. (1982b). Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. J. Cell Biol. **93**, 306-313.

Christou, P., Platt, S. G., and Ackerman, M. C. (1986). Opine synthesis in wild-type plant tissue. Plant Physiol. **82**, 218-221.

Citovsky, V., De Vos, G. and Zambryski, P. (1988). Single-stranded DNA binding protein encoded by the *vir A* locus of *Agrobacterium tumefaciens*. Science **240**, 501-503.

Cohen, R. B., Sheffrey, M. and Kim, C. G. (1986). Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α-1 globin gene. Mol. Cell. Biol. **6**, 821-832.

Colbert, J. T., Hershey, H. P. and Quail, P. H. (1983). Autoregulatory control of translatable phytochrome mRNA levels. Proc. Natl. Acad. Sci. USA **80**, 2248-52.

Colbert, J. T., Hershey, H. P. and Quail, P. H. (1985). Phytochrome regulation of phytochrome mRNA abundance. Plant Mol. Biol. **5**, 91-101.

Colot, V., Robert, L. S., Kavanagh, T. A., Bevan, M. W. and Thompson, R. D. (1987). Localization of sequence in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J. **6**, 3559-3564.

Colsen, C., Glover, S. W., Symonds, N. and Stacey, K. A. (1965). The location of the genes for host controlled modification and restriction in *Escherichia coli* K-12. Genetics **52**, 1043-1450.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Promoter sequences of eukaryotic protein coding genes. Science **209**, 1406-14.

Cordes, S., Deikman, J., Margossian, L. J. and Fischer, R. L. (1989). Interaction of a developmentally regulated DNA-binding factor with sites flanking two different fruit-ripening genes from tomato. Plant Cell **1**, 1025-1034.

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

Craig, S., Goodchild, D. J. and Hardham, A. R. (1979). Structural aspects of protein accumulation in developing pea cotyledons. I. Qualitative and quantitative changes in parenchyma cell vacuoles. Aust. J. Plant Physiol. **6**, 81-98.

Croy, R. R. D., Krishna, T. G. and Boulter, D. (1979). Legumin of *Pisum sativum* and *Vicia faba*. New Phytol. **83**, 29-35.

Croy, R. R. D. and Gatehouse, J. A. (1985). Genetic engineering of seed proteins - current and potential applications. In Plant genetic engineering (Dodds, J. A., ed.), Cambridge University Press, pp.143-268.

Croy, R. R. D., Gatehouse, J. A., Evans, M. I. and Boulter, D. (1980). Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and

RNA from developing pea (*Pisum sativum L.*) I. Legumin. Planta **148**, 49-56.

Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N. and Boulter, D. (1982). Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature **295**, 76-79.

Czasrnecka, E., Gurley, W. B., Nagao, R. T., Mosquera, L. A. and Key, J. L. (1985). DNA sequence and transcript mapping of a soybean gene encoding a small heat shock protein. Proc. Natl. Acad. Sci. USA **82**, 3726-3730.

Czernilofsky, A. P., Hain, R., Herrera-Estrella, L., Lörz, H., Goyvaerts, E., Baker, B. J. and Schell, J. (1986). Fate of selectable marker DNA integrated into the genome of *Nicotiana tabacum.* DNA **5**, 101-113.

Dagert, M. and Ehrlich, N. D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene **6**, 23-27.

Danielsson, C. E. (1949). Seed globulins of the Gramineae and Leguminosae. Biochem. J. **44**, 387-400.

Datta, N., and Cashmore, A. S. (1989). Binding of a pea nuclear protein to promoters of certain photoregulated genes is modulated by phosphorylation. Plant Cell **1**, 1069-1077.

Datta, H. and Hedges, R. W. (1971). Compatibility groups among fi^- R factors. Nature **234**, 222-226.

Davidson, E. H., Jacobs, H. T. and Britten, R. J. (1983). Very short repeats and coordinate induction of genes. Nature **301**, 468-470.

Davies, R. D. (1980). The r_a locus and legumin synthesis in *Pisum sativum.* Biochem. Genet. **18**, 1207-1219.

Dean, C., Favreau, M., Tamaki, S., Bond-Nutter, D., Dunsmuir, P. and Bedbrook, J (1988a). Expression of tandem gene fusions in transgenic tobacco plants. Nucleic Acids Res. **16**, 7601-7617.

Dean, C., Jones, J., Favreau, M., Dunsmuir, P. and Bedbrook, J (1988b). Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. Nucleic Acids Res. **16**, 9267-9283.

Dean, C., Favreau, M., Bedbrook, J. and Dunsmuir, P. (1989a). Sequences 5′ to translation start regulate expression of petunia *rbcS* genes. Plant Cell **1**, 209-215.

Dean, C., Favreau, M., Bond-Nutter, D., Bedbrook, J. and Dunsmuir, P. (1989b). Sequences downstream of translation start regulate quantitative expression of two petunia *rbcS* genes. Plant Cell **1**, 201-208.

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

De Cleene, M. and De Ley, J. (1976). The host range of crown gall. Bot. Rev. **42**, 389-466.

de Frammond, A., Barton, K. A. and Chilton, M.-D. (1983). Mini-Ti; a new vector strategy for plant genetic engineering. Biotechnology **1**, 262-269.

De Greve, H., Dhaese, P., Seurinck, J., Lemmers, M., Van Montagu, M. and Schell, J. (1982). Nucleotide sequence and transcript map of the *Agrobacterium tumefaciens* Ti plasmid-encoded octopine synthase gene. J. Mol. Appl. Genet. **1**, 499-512.

DeLuca, M. and McElroy, W. D. (1978). Purification and properties of firefly luciferase. Methods Enzymol. **57**, 3-15.

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

Dellaporta, S. L., Wood, J. and Hicks, J. B. (1983). A plant DNA minipreparation version II. Plant Mol. Biol. Reporter **1**, 4, 19-21.

Derbyshire, E., Wright, D.J. and Boulter, D. (1976). Legumin and vicilin, storage proteins of legume seeds. Phytochemistry **15**, 3-15.

Dhaese, P., De Greve, H., Decraemer, H., Schell, J. and Van Montagu, M. (1979). Rapid mapping of transposon insertion and deletion mutations in the large Ti-plasmids of *Agrobacterium tumefaciens*. Nucleic Acids Res. **7**, 1837-1841.

Dickinson, C. D., Evans, P. R. and Neilson, N. C. (1988). RY repeats are conserved in the 5′ flanking regions of legume seed protein genes. Nucleic Acids Res. **16**, 371.

Dickinson, C. D., Scott, P. M. Hussein, E. H. A., Argos, P. and Neilson, N. C. (1990). Effect of structural modifications on the assembly of a glycinin subunit. The Plant Cell, **2**, 403-413.

Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**, 695-706.

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

Domoney, C. and Casey, R. (1984). Storage protein precursor polypeptides in cotyledons of *Pisum sativum L.* Identification of, and isolation of a cDNA clone for, an 80,000 Mr legumin-related polypeptide. Eur. J. Biochem.**139**, 321-327.

Domoney, C. and Casey, R. (1985). Measurement of gene number for seed storage proteins in *Pisum*. Nucleic Acids Res. **13**, 687-699.

Domoney, C. and Casey, R. (1987). Changes in legumin messenger RNA throughout seed development in *Pisum sativum L.* Planta **170**, 562-566.

Domoney, C., Ellis, N. T. H. and Davies, R. D. (1986). Organization and mapping of legumin genes in *Pisum*. Mol. Gen. Genet. **202**, 280-285.

Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981). A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112**, 295-298.

Dure, L. S. (1975). Seed formation. Ann. Rev. Plant Physiol.**26**, 259-278.

Dyan, W. S. and Tjian, R. (1985). Control of eukaryote messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature **316**, 774-778.

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.

Edwards, G. E. and Huber, S. C. (1981). in The Biochemistry of Plants. A Comprehensive Treatise, Stumpf, P. K. and Conn, E. E. Eds., **8**, 237-281, Academic Press, New York.

Ellis, R. J. (1986). Photoregulation of gene expression. Biosci. Rep. **6**, 127-136.

Ellis, J, G., Llewellyn, D. J., Dennis, E. S. and Peacock, W. J. (1987a). Maize *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.

Ellis, J, G., Llewellyn, D. J., Walker, J. C., Dennis, E. S. and Peacock, W. J. (1987b). The *ocs* element: A 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J. **6**, 3203-3208.

Ellis, J. R., Shirsat, A. H., Hepher, A., Yarwood, J. N., Gatehouse, J. A., Croy, R. R. D. and Boulter, D. (1988). Tissue-specific expression of a pea legumin gene in seeds of *Nicotiana plumbaginifolia.* Plant Mol. Biol. **10**, 203-214.

Evans, I. M., Gatehouse, J. A., Croy, R. D. D. and Boulter, D. (1984a). Regulation of the transcription of storage-protein mRNA in nuclei isolated from developing pea (*Pisum sativum L.*) cotyledons. Planta **160**, 559-568.

Evans, I. M., Gatehouse, J. A., Lycett, G. W. and Boulter, D. (1984b). Two cDNA clones for the legumin protein of *Pisum sativum L.* contain sequence repeats. Plant Mol. Biol. **3**, 91-96.

Evans, I. M., Bown, D., Lycett, G. W., Croy, R. R. D., Boulter, D. and Gatehouse, J. A. (1985a). Transcription of a legumin gene from pea (*Pisum sativum L.*) in vitro. Planta **165**, 554-560.

Evans, I. M., Gatehouse, J. A. and Boulter, D. (1985b). Regulation of storage-protein synthesis in pea (*Pisum sativum L.*) cotyledons under conditions of sulphur deficiency. Biochem. J. **232**, 261-265.

Fang, R. -X., Nagy, F., Sivasubtamaniam, 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.

Ferl, R. J. and Nick, H. S. (1987). *In vivo* detection of regulatory factor binding sites in the 5′ flanking region of maize *Adh*1. J. Biol. Chem. **262**, 7947-7950.

Figurski, D. H. and Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans.* Proc. Natl. Acad. Sci. USA **76**, 1648-1652.

Fischer, R. L. and Goldberg, R. B. (1982). Structure and flanking regions of soybean seed protein genes. Cell **29**, 651-660.

Fleck, J., Durr, A., Lett, M. C. and Hirth, L. (1979). Changes in protein synthesis during the initial stage of life of tobacco protoplasts. Planta **145**, 279-285.

Fluhr, R. and Chua. N.-H. (1986). Developmental regulation of two genes encoding ribulose-bisphosphate carboxylase small subunit in pea and transgenic petunia plants: phytochrome responses and blue-light induction. Proc. Natl. Acad. Sci. USA **83**, 2358-2362.

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

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. and Chua. N.-H. (1986b). Expression dynamics of the pea rbcS multigene family and organ distribution of the transcripts. EMBO. J. **5**, 2063-2071.

Forde, B. G., Heyworth, A., Pywell, J. and Kreis, M. (1985). Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res. **13**, 7327-7339.

Fraley, R. T., Rogers, S. G., Horsh, R. B., Sanders, P. S., Flick, J. S., Adams, S. P., Bittner, M. L., Brand, L. A., Fink, C. L., Fry, J. S., Galluppi, G. R., Goldberg, R. B., Hoffman, N. L. and Woo, S. C. (1983). Expression of bacterial genes in plant cells. Proc. Natl. Acad. Sci. USA **80**, 4803-4807.

Fraley, R. T., Rogers, S. G., Horsh, R. B., Eichholtz, D. A., Flick, J. S., Fink, C. L., Hoffman, N. L. and Sanders, P. S. (1985). The SEV system: a new disarmed Ti plasmid vector for plant transformation. Biotechnology **3**, 629-635.

Fraley, R. T., Rogers, S. G. and Horsh, R. B. (1986). Genetic transformation in higher plants. CRC Crit. Rev. Plant Sci. **4**, 1-46.

Fromm, H., Katagiri, F. and Chua, N. -H. (1989). An octopine synthase enhancer element directs tissue-specific expression and binds ASF-1, a factor from tobacco nuclear extracts. Plant Cell **1**, 977-984.

Fromm, M. E., Taylor, L. P. and Walbot, V. (1985). Expression of genes transferred into monocot and dicot plant cells by electroporation. Proc. Natl. Acad. Sci. USA **82**, 5824-5828.

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

Gallie, D. R., Sleat, D. E., Watts, J. W., Turner, P. C., and Wilson, T. M. A. (1987). The 5′ leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. Nucleic Acids Res. **15**, 3257-3273.

Garcia, J. V., Le thi, B.-T., Stafford, J. and Queen, C. (1986). Synergism between immunoglobulin enhancers and promoters. Nature **322**, 383-385.

Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R. G. (1990). *Arabidopsis thaliana* contains two genes for TFIID. Nature **346**, 390-394.

Gatehouse, J. A., Croy, R. R. D. and Boulter, D. (1980). Iso-electric-focusing properties and carbohydrate content of pea (*Pisum sativum*) legumin. Biochem. J. **185**, 497-503.

Gatehouse, J. A., Evans, M. I., Bown, D., Croy, R. R. D. and Boulter, D. (1982). Control of storage-protein synthesis during seed development in pea (*Pisum sativum L.*). Biochem. J. **208**, 119-127.

Gatehouse, J. A., Bown, D., Gilroy, J., Levasseur, M., Castleton, J. and Ellis, N. T. H. (1988). Two genes encoding 'minor'legumin polypeptides in pea (*Pisum sativum L.*). Biochem. J. **250**, 15-24.

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

Gilmartin, P. M., and Chua, N. -H. (1990). Spacing between GT-1 binding sites within a light-responsive element is critical for transcriptional activity. The Plant Cell, **2**, 447-455.

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

Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A. and Cashmore, A. R. (1988). An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. Proc. Natl. Acad. Sci. USA **85**, 7089-7093.

Goldsbrough, P. B., Gelvin, S. B. and Larkins, B. A. (1986). Expression of maize zein genes in transformed sunflower cells. Mol. Gen. Genet. **202**, 374-381.

Gordon-Kamm, W. J., Spencer, T. M., Mangano, M. L., Adams, T. R., Daines, R. J., Start, W. G., O'Brien, J. V., Chambers, S. A., Adams, W. R., Willets, N. G., Rice, T. B., Mackey, C. J., Krueger , R. W., Kausch, A. P. and Lemaux, P. G. (1990). Transformation of maize cells and regeneration of fertile transgenic plants. Plant Cell. **2**, 603-618.

Gorman, C. M., Moffatt, L. F. and Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells Mol. Cell. Biol. **2**, 1044-1051.

Graham, D. E. (1978). The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal. Biochem. **85**, 609-613.

Graves, A. C. F. and Goldman, S. L. (1986a). The transformation of *Zea mays* seedlings with *Agrobacterium tumefaciens*: detection of T-DNA specific enzyme activities. Plant Mol. Biol. **7**, 43-50.

Graves, A. C. F. and Goldman, S. L. (1986b). *Agrobacterium tumefaciens*- mediated transformation of the monocot genus *Gladiolus*: detection of expression of T-DNA genes. J. Bacteriol. **169**, 1745-1746.

Green, P. J., Kay, S. A. 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**, 2543-2550.

Groner, B., Kennedy, N., Skroch, P., Hynes, N. E. and Ponta, H. (1984). DNA sequences involved in the regulation of gene expression by glucocorticoid hormones. Biochim. Biophys. Acta **781**, 1-6.

Grosschedl, R. and Birnstiel, M. L. (1980). Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription *in vivo*. Proc. Natl. Acad. Sci. USA **77**, 7102-7106.

Grosveld, G. C., Shewmaker, C. K., Jat, P., Flavell, R. A. (1981). Localisation of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*.

Cell. **25**, 215-26.

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

Guiltinan, M. J., Marcotte, W. R. and Quatrano, R. S. (1990). A plant leucine zipper protein that recognizes an abscisic-acid response element. Science. **250**, 267-271.

Gurley, W. B., Czarnecka, E., Nagao, R. T., and Key, J. L. (1986). Upstream sequences required for efficient expression of a soybean heat shock gene. Mol. Cell. Biol. **6**, 559-565.

Hain, R., Stabel, P., Czernilofsky, A. P., Steinbiss, H. H., Herrera-Estrella, L. and Schell, J. (1985). Uptake, integration, expression and genetic transmission of a selectable chimaeric gene by plant protoplasts. Mol. Gen. Genet. **199**, 161-168.

Harris, N. and Croy, R. R. D. (1986). Localization of mRNA for pea legumin: *in situ* hybridization using biotinylated cDNA probes. Protoplasma, **130**, 57-67.

Harrison, M. J., Lawton, M. A., Lamb, C. J. and DIxon, R. A. (1991). Characterization of a nuclear protein that binds to three elements within the silencer region of a bean chalcone synthase gene promoter. Proc. Natl. Acad. Sci. USA **88**, 2515-2519.

Hearing, P. and Shenk, T. (1983). The Adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell **33**, 695-703.

Heidecker, G. and Messing, J. (1986). Structural analysis of plant genes. Ann. Rev. Plant Physiol. **37**, 439-466.

Hentschel, C. C. and Birnstiel. (1981). The organisation and expression of the histone gene families. Cell. **25**, 301-13.

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.

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

Hernalsteens, J.-P., Thia-Toong, L., Schell, J. and Van Montagu, M. (1984). An *Agrobacterium*-transformed cell culture from the monocot *Asparagus officinalis*. EMBO J. **3**, 3039-3041.

Herrera-Estrella, L., De Block, M. N., Messens, E., Hernalsteens, J.-P., Van Montagu, M., and Schell, J. (1983a). Chimeric genes as dominant selectable markers in plant cells. EMBO J. **2**, 987-995.

Herrera-Estrella, L., Depicker, A., Van Montagu, M. and Schell, J. (1983b). Expression of chimaeric genes transferred into plant cells using a Ti plasmid derived vector Nature **303**, 209-213.

Herrera-Estrella, L., Van den Broeck, G., Maenhaut, R., Van Montagu, M., Schell, J., Timko, M. and Cashmore, A. (1984). Light-inducible and chloroplast-associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a Ti plasmid vector. Nature **310**, 115-120.

Higgins, T. J. V. (1984). Synthesis and regulation of major seed proteins in seeds. Ann. Rev. Plant Physiol. **35**, 191-221.

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J. and Schilperoort, R. A. (1983). A binary plant vector strategy based on separation of *vir* and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature **303**, 179-181.

Hoffman, L. M., Donaldson, D. D., Bookland, R., Rashka, K. and Herman, E. M. (1987). Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seeds. EMBO J. **6**, 3213-3222.

Holderbach, E. and Biedrebeck, R. (1976). Octopingehalt in normal-und Tumorgeweben eineger höherer Pflanzen. Phytochemistry **15**, 955-956.

Holsters, M., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., Depicker, A., Inze, D., Engler, G., Villarroel, R., Van Montagu, M.and Schell, J. (1980). The functional organisation of the nopaline *A. tumefaciens* plasmid pTiC58. Plasmid **3**, 212-218.

Hooykaas-Van Slogteren, G. M. S., Hooykaas, P. J. J. and Schilperoort, R. A. (1984). Expression of Ti-plasmidgenes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. Nature **311**, 763-764.

Horsch, R. B. and Klee, H. J. (1986). Rapid assay of foreign gene expression in leaf discs transformed by *Agrobacterium tumefaciens*: Role of T-DNA borders in the transfer process. Proc. Natl. Acad. Sci. USA **83**, 4428-4432.

Horsch, R. B. Fraley, R. T., Rogers, S. G., Sanders, P. R., Lloyd, A. and Hoffman, N. (1984). Inheritance of functional foreign genes in plants. Science **223**, 496-498.

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

Horsch, R. B. Klee., H. J., Stachel, S., Winans, S. C., Nester, E. W., Rogers, S. G., and Fraley, R. T. (1986). Analysis of *Agrobacterium tumefaciens* virulence mutants in leaf discs Proc. Natl. Acad. Sci. USA **83**, 2571-2575.

Howard, E. A., Walker, J. L., Dennis, E. S. and Peacock, W. J. (1987). Regulated expression of an alcohol dehydrogenase 1 chimeric gene introduced into maize protoplasts. Planta **170**, 535-540.

Hunt, A. G., Chu, N. M., Odell, J. T., Nagy, F. and Chua, N.-H. (1987). Plant cells do not properly rcognize animal gene polyadenylation signals Plant Mol. Biol. **8**, 23-36.

Hurkman, W. J. and Beevers, L. (1982). Sequestration of pea reserve proteins by rough microsomes. Plant Physiol. **69**, 1414-17.

Jacobsen, K., Laursen, N. B., Jensen, E. O., Marcker, A., Poulsen, C. and Mar-

cker, K. A. (1990). HMG 1-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nodulin promoters. The Plant Cell **2**, 85-94.

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

Jeffreys, A. J. and Flavell, R. A. (1977). A physical map of the DNA regions flanking the rabbit β-globin gene. Cell **12**, 429-434.

Jensen, E. O., Marcker, K. A., Schell, J. and de Bruijn, F. J. (1988). Interaction of a nodule-specific, *trans*-acting factor with distinct DNA elements in the soybean leghaemoglobin lbc_3 5′ upstream region. EMBO J. **7**, 1265-1271.

Johnson, R., Guderian, R. H., Eden, F., Chilton, M.-D., Gordon, M. P. and Nester, E. W. (1974). Detection and quantitation of octopine in normal plant tissue and in crown gall tumors. Proc. Natl. Acad. Sci. USA **71**, 536-539.

Johnson, E. D., Knight, J. and Gayler, K. R. (1985). Biosynthesis and processing of legumin-like storage proteins in *Lupinus augustifolius* (Lupin). Biochem. J. **232**, 673-679.

Jofuku, K. D., Okamuro, J. K. and Goldberg, R. B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature **328**, 734-737.

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

Jones, K. A., Kadonaga, J. T. Rosenfeld, P. J., Kelly, T. J. and Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**, 79-89.

Jordano, J., Almoguera, C. and Thomas, T. L. (1989). A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. The Plant Cell. **1**, 855-866.

Jorgensen, R., Snyder, C. and Jones, J. D. G. (1987). T-DNA is organized predominantly in inverted repeats in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. Mol. Gen. Genet. **207**, 471-478.

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

Kaufman, L. S., Briggs, W. R., Thompson, W. F. (1985). Phytochrome control of specific mRNA levels in developing pea buds. Plant Physiol. **78**, 388-93.

Kay, S. A., Keith, B., Shinozaki, K., Chye, M. -L. and Chua, N.- H. (1989). The rice phytochrome gene: structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5′ upstream region. Plant Cell **1**, 351-360.

Keil, M., Sanchez-Serrano, J., Schell, J. and Willmitzer, L. (1990). Localization of elements important for the wound-inducible expression of a chimeric potato proteinase inhibitor II-CAT gene in transgenic tobacco plants. The Plant Cell, **2**, 61-70.

Keith, B. and Chua. N.-H. (1986). Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco cells. EMBO J. **5**, 2419-2426.

Kim, S. Y. and Wu, R. (1990). Multiple protein factors bind to rice glutelin promoter region. Nucleic Acids Res. **18**, 6845-6852.

Klee, H. J., White, F. F., Iyer, V. N., Gordon, M. P., Nester, E. W. (1983). Mutational analysis of the virulence region of an *Agrobacterium tumefaciens* Ti plasmid. J. Bacteriol. **153**, 878-883.

Klee, H., Horsch, R. and Rogers, S. (1987). *Agrobacterium*-mediated plant transformation and its further applications. Ann. Rev. Plant Physiol. **38**, 467-486.

Klein,T. M., Fromm, M., Weissinger, A., Tomes, D., Schaaf, S., Sletten, M. and Sanford, J. C. (1988) Transfer of foreign genes into intact maize cells with high-velocity microprojectiles. Proc. Natl. Acad. Sci. USA **85**, 4305-4309.

Koncz, C., De Greve, H., Andre, D., Deboeck, F., Van Montagu, M. and Schell, J. (1983). The opine synthase genes carried by Ti-plasmids contain all signals necessary for expression in plants. EMBO J. **2**, 1597-1603.

Koncz, C., Kreuzaler, F., Kalman, Z. S. and Schell, J. (1984). A simple method to transfer, integrate and study expression of foreign genes, such as chicken ovalbumin and α-actin in plant tumors. EMBO J. **3**, 1029-1037.

Kozak, M. (1987). An analysis of 5′-non coding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**, 8125-8148.

Koziel, M. G., Adams, T. L., Hazlet, M. A., Damm, D., Miller, J., Dahlbeck, D., Jayne, S. and Staskawicz, B. J. (1984). A cauliflower mosaic virus promoter directs expression of kanamycin resistance in morphogenic transformed plant cells. J. Mol. Appl. Genet. **2**, 549-62.

Kreis, M., Williamson, M. S, Forde, J., Clarke, J., Buxton, B., Pywell, J., Marris, C., Gallois, P. and Shrewry, P. R. (1987). Structure and expression of seed protein genes in barley. Plant Physiol. Biochem. **25**, 291-302.

Kuhlemeier, C., Cuozzo, M., Green, P. J., Goyvaerts, E., Ward, K. and Chua, N-H. (1988). Localization and conditional redundancy of regulatory elements in *RbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA **85**, 4662-4666.

Kuhlemeier, C., Fluhr, R., Green, P. J. and Chua, N-H. (1987a). Sequences in the pea *rbcS*-3A gene have homology to constitutive mammalian enhancers but function as negative regulatory elements. Genes Dev. **1**, 247-255.

Kuhlemeier, C., Green, P. J. and Chua, N-H. (1987b). Regulation of gene expression in higher plants. Ann. Rev. Plant Physiol. **38**, 221-57.

Kuhlemeier, C., Strittmatter, G., Ward, K. and Chua, N-H. (1989). The pea *rbcS*-3A promoter mediates light responsiveness but not organ specificity. Plant Cell **1**, 471-478.

Lacey, E., Roberts, S., Evans, E. P., Burtenshaw, M. D. and Constantini, F. D. (1983). A foreign β-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. Cell, **34**, 343-358.

Lam, E., Benfey, P. N., Gilmartin, P. M., Fang, R.-X. and Chua, N.-H. (1989). Site-specific mutations alter *invitro* factor binding and change promoter expression pattern in transgenic plants. Proc. Natl. Acad. Sci. USA **86**, 7890-7894.

Lam, E. and Chua, N.- H. (1990a). GT-1 binding site confers light responsive expression in transgenic tobacco. Science **248**, 471-474.

Lam, E., Kanomurakami, Y., Gilmartin, P. M., Niner, B. and Chua, N.-H. (1990b). A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter. Plant Cell **2**, 857-866.

Lam, E., Katagiri, F. and Chua, N.-H. (1990c). Plant nuclear factor *asf*-1 binds to an essential region of the nopaline synthase promoter. J. Biol. Chem. **265**, 9909-9913.

Lamppa, G., Nagy, F and Chua. N.-H. (1985). Light-regulated and organ-specific expression of a wheat Cab gene in transgenic tobacco. Nature **316**, 750-752.

Langdale, J. A., Taylor, W. C. and Nelson, T. (1991). Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site >3kb upstream of the gene. Mol. Gen. Genet. **225**, 49-55.

Latchman, D. S. (1990). Eukaryotic transcription factors. Biochem. J. **270**, 281-289.

Lawton, M. A., Tierney, M. A., Nakamura, I., Anderson, E., Komeda, Y., Dube, P., Hoffman, N., Fraley, R. T. and Beachy, R. N. (1987). Expression of a soybean β-conglycinin gene under the control of the cauliflower mosaic 35S and 19S promoters in transformed petunia tissues. Plant Mol. Biol. **9**, 315-324.

Leemans, J., Shaw, C. H., Deblaere, R., De Greve, H., Hernalsteens, J-P., Van

Montagu, M. and Schell, J. (1981). Site-specific mutagenesis of *Agrobacterium* Ti-plasmidand transfer of genes to plant cells. J. Mol. Appl. Genet. **1**, 149-164.

Lewin, B. (1990). Commitment and activation of PolII promoters: a tail of protein-protein interactions. Cell, **61**, 1161-1164.

Lichtenstein, C. and Draper, J. (1985). in *DNA cloning - A Practical Approach Vol II*, Glover, D. M. ed. pp.67-119.

Liu, X. J., Prat, S., Willmitzer, L. and Frommer, W. B. (1990). Cis regulatory elements direct tuber-specific and sucrose-inducible expression of a chimeric class-I patatin promoter GUS-gene fusion. Mol. Gen. Genet. **223**, 401-406.

Lloyd, J. C., Raines, C. A., John, U. P. and Dyer, T. A. (1991). The chloroplast FBPase gene of wheat: structure and expression of the promoter in photosynthetic and meristematic cells of transgenic tobacco plants. Mol. Gen. Genet. **225**, 209-216.

Logemann, J., Lipphardt, S., Lörz, H., Häuser, I., Willmitzer, L., and Schell, J. (1989). 5′ upstream sequences from the *wun*1 gene are responsible for gene activation by wounding in transgenic plants. The Plant Cell, **1**, 151-158.

Lorz, H., Baker, B. and Schell, J. (1985). Gene transfer to cereal cells mediated by protoplast transformation. Mol. Gen. Genet. **199**, 178-182.

Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. and Scheele, G. A. (1987). Selection of AUG initiation codons differs in plants and animals. EMBO J. **6**, 43-48.

Lycett, G.W., Delauney, A. J. and Croy, R.R.D. (1983a). Are plant genes different? FEBS Lett. **153**, 43-46.

Lycett, G.W., Delauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R.R.D., and Boulter, D. (1983b). The vicilin gene family of pea (*Pisum sativum L.*): a complete cDNA coding sequence for preprovicilin. Nucleic Acids Res. **11**, 2367-2380.

Lycett, G.W., Croy, R.D.D., Shirsat, A.H. and Boulter, D. (1984a). The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum L.*) Nucleic Acid Res. **12**, 4493-4506.

Lycett, G.W., Delauney, A. J., Zhao, W., Gatehouse, J. A., Croy, R.R.D., and Boulter, D. (1984b). Two cDNA clones coding for the legumin protein of *Pisum sativum L.* contain sequence repeats. Plant Mol. Biol. **3**, 91-96.

Lycett, G.W., Croy, R.D.D., Shirsat, A.H., Richards, D. M. and Boulter, D. (1985). The 5′-flanking regions of three pea legumin genes: comparison of the DNA sequences. Nucleic Acids Res. **13**, 6733-6743.

Maier, U.-G., Brown, J. W. S., Toloczyki, C. and Feix, G., (1987). Binding of a nuclear factor to a consensus sequence in the 5′ flanking region of zein genes from maize. EMBO **6**, 17-21.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (eds.) (1982). Molecular cloning: A laboratory Manual, published by Cold Spring Harbor Laboratory Press, New York.

Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. Science **236**, 1237-1245.

March, J. F., Domoney, C. and Casey, R. (1987). Legumin heterogeneity in *Pisum.* Biochem. Genet. **25**, 449-458.

Marocco, A., Wissenbach, M., Becker, D., Paz-Ares, J., Saedler, H., Salamini, F. and Rohde, W. (1989) Multiple genes are transcribed in *Hordeum vulgare* and *Zea mays* that carry the DNA binding domain of the *myb* oncoproteins. Mol. Gen. Genet. **216**, 183-187.

Marris, C., Gallois, P. and Kreis, m. (1988). The 5′-flanking region of a barley B hordein gene controls tissue and developmental specific CAT expression in tobacco plants. Plant Mol. Biol. **10**, 359-365.

Matta, N. K. and Gatehouse, J. A. (1982). Inheritance and mapping of stirage

protein genes in *Pisum sativum L.* Heridity **48**, 383-392.

Matta, N. K., Gatehouse, J. A. and Boulter, D. (1981). Molecular and subunit heterogeneity of legumin of *Pisum sativum L.* (Garden pea) - a multidimensional gel electrophoresis study. J. Exp. Bot. **32**, 1295-1307.

Matthysse, A. G. (1987). Effect of plasmid pSa and of auxin on attachment of *Agrobacterium tumefaciens*to carrot cells. Appl. Environ. Microbiol. **53**, 2574-2582.

Matzke, A. and Matzke, M. (1986). A set of novel Ti plasmid-derived vectors for production of transgenic plants. Plant Mol. Biol. **7**, 357-365.

Matzke, M. A., Susani, M., Binns, A. N., Lewis, E. D., Rubenstein, I. and Matzke, A. J. M. (1984). Transcription of a zein gene introduced into sunflower using a Ti-plasmid vector. EMBO J. **3**, 1525-1531.

Mauro, V. P., Nguyen, T., Katinakis, P. and Verma, D. P. S. (1985). Primary structure of the soybean nodulin-23 gene and potential regulatory elements in the 5′-flanking regions of nodulin and leghemoglobin genes. Nucleic Acids Res. **13**, 239-249.

McCormick, S., Niedermeyer, J., Fry, J., Barnason, A., Horsch, R., Fraley, R. (1986). Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens.* Plant Cell Rep. **5**, 81-84.

McDonnell, M. W., Simon, M. N. and Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**, 119-125.

McKee, H. S. (1962). Nitrogen metabolism in plants. Clarendon Press, Oxford.

McKnight, S. and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell **46**, 795-805.

McPherson, J. C. (1984). DNA sequence analysis of crown gall T-DNA encoding

the 0.7 kb transcript. Nucleic Acids Res. **12**, 2317-2326.

Meakin, P. J., Shirsat, A. H. and Gatehouse, J. A. (1990) Pea storage protein genes: promoter analysis and DNA-binding assays. J. Exp. Bot. [P5.24].

Mertz, E. T., Bates, L. S. and Nelson, O. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science **145**, 279-280.

Messing, J., Geraghty, D., Heidecker, G., Hu, N., Kendl, J. and Rubenstein, I. (1983). Genetic Engineering of Plants, Kosuge, T., Meredith, C.P. and Hollaender, A., Eds, pp.211-227.

Metz, B. A., Welters, P., Hoffman, H. J., Jensen, E. O., Schell, J. and de Bruijn, F. J. (1988). Primary structure and promoter analysis of leghemoglobin genes of the stem-nodulated tropical legume *Sesbania rostrata*: Conserved coding sequences, *cis*-elements and *trans*-acting factors. Mol. Gen. Genet. **214**, 181-191.

Miflin, B. and Lea, P. (1984). The genetic manipulation of crop plants. Nature **308**, 498-499.

Miller, J. (1972). Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York p. 466.

Millerd, A. and Spencer, D. (1974). Changes in RNA-synthesizing activity and template activity in nuclei from cotyledons of developing pea seeds. Aust. J. Plant Physiol. **1**, 331-341.

Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K. and Fukazawa, C. (1985). Glycinin $A_5A_4B_3$ mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean. Eur. J. Biochem. **149**, 491-496.

Montoya, A. C., Chilton, M.-D., Gordon, M. P., Sciaky, D. and Nester, E. W. (1977). Octopine and nopaline metabolism in *Agrobacterium tumefaciens* and crown gall tumor cells: Role of plasmid genes. J. Bacteriol. **129**, 101-107.

Morelli, G., Nagy, F., Fraley, R. T., Rogers, S. G. and CHua, N. -H. A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose 1,5-bisphosphate carboxylase small subunit of pea. Nature **315**, 200-204.

Morton, H., Evans, I. M., Gatehouse, J. A. and Boulter, D. (1983). Sequence complexity and abundance classes in messenger RNA from cotyledons of developing pea (*Pisum sativum L.*) seeds. Phytochemistry **22**, 807-817.

Murai, N., Sutton, D. W., Murray, M. G., Slightom, J. L., Merlo, D. J., Reichert, N. A., Sengupta-Gopalan, C., Stock, C. A., Barker, R. F., Kemp, J. D. and Hall, T. C. (1983). Phaseolin gene from bean is expressed after transfer to sunflower via tumour-inducing plasmid vectors. Science **222**, 476-482.

Murashige, T. and Skoog, S. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plantarum **15**, 473-497.

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. **4**, 3063-3068.

Nagy, F., Kay, S. A., Boutry, M., Hsu, M.-Y. and Chua. N. -H. (1986). Phytochrome-controlled expression of a wheat *cab* gene in transgenic tobacco seedlings. EMBO J. **5**, 1119-1128.

Nagy, F., Kay, S. A., and Chua. N.-H. (1988). Gene regulation by phytochrome. Trends. Genet. **4**, 37-41.

Naito, S., Dube, P. H. and Beachy, R. N. (1988). Differential expression of conglycinin α' and β subunit genes in transgenic plants. Plant Mol. Biol. **11**, 109-123.

Nelson, O. E., Mertz, E. T. and Bates, L. S. (1965). Second mutation gene affecting the amino acid pattern of maize endosperm proteins. Science **150**, 1469-1470.

Nester, E. W., Gordon, M. P., Amasino, R. M. and Yanofsky, M. F. (1984). Crown

gall: a molecular and physiological analysis. Ann. Rev. Plant Physiol.**35**, 387-413.

Nielsen, N. C. (1984). The chemistry of legume storage proteins. Phil. Trans. Roy. Soc. Lond. B. **304**, 287-296.

Nudel, U., Greenburg, D., Ordahl, C. P., Saxel, O., Neuman, S. and Yaffe, D. (1985) Developmentally regulated expression of a chicken muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA **82**, 3106-3109.

Odell, J. T., Nagy, F. and Chua, N -H. (1985). Identification of DNA sequences required for activity of the CaMV 35S promoter. Nature **313**, 810-12.

Odell, J. T., Knowlton, S., Lin, W. and Mauvais, C. J. (1988). Properties of an isolated transcription-stimulating sequence derived from the cauliflower mosaic virus 35S promoter. Plant Mol. Biol. **10**, 263-272.

Odl, S., Hedrick, S. A., Chory, J. and Lamb, C. J. (1990). Functional properties of a phenylalanine ammonia-lyase promoter from *Arabidopsis*. Plant Cell **2**, 837l-848.

Okamuro, J. K., Jofuku, K. D. and Goldberg, R. B. (1986). The soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. Proc. Natl. Acad. Sci. USA **83**, 8240-8244.

Opik, H. (1968). Development of cotyledon cell structure in ripening *Phaseolus vulgaris* seeds. J. Exp. Bot. **19**, 64-76.

Osborne, T. B. (1924). The vegetable proteins, 2nd edn., Longmans, Green and Co. N.Y.

Otten, L. and Schilperoort, R.A. (1978). A rapid microscale method for the detection of lysopine and nopaline dehydrogenase activities. Biochim. Biophys. Acta. **527**, 497-500.

Otten, L., De Greve, H., Hernalsteens, J.-P., Van Montagu, M., Schieder, O.,

Straub, J. and Schell, J. (1981). Mendelian transmission of genes introduced into plants by the Ti-plasmidsof *Agrobacterium tumefaciens*. Mol. Gen. Genet. **183**, 209-213.

Ow, D. L., Wood, L. V., DeLuca, M., DeWet, J. R., Helinski, D. R. and Howell, S. H. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. Science , **234**, 856-859.

Palmiter, R. D. and Brinster, R. L. (1985). Transgenic mice. Cell, **41**, 343-345.

Paszkowski, J., Shillito, R. D., Saul, M., Mandak, V., Hohn, T., Hohn, B. and Potrykus, I. (1984). Direct gene transfer to plants. EMBO J. **3**, 2717-2722.

Paul, A.-L. and Ferl, R. J. (1991). *Invitro* footprinting reveals unique cis-elements and different modes of hypoxic induction in maize *Adh*-1 and *Adh*-2. The Plant Cell **3**, 159-168.

Paul, A.-L., Vasil, I. K., Vasil, V., and Ferl, R. J. (1987). Constitutive and anaerobically induced DNase-I-hypersensitive sites in the 5′ region of the maize *Adh*1 gene. Proc. Natl. Acad. Sci. USA **84**, 799-804.

Payne, P. I. (1983). Breeding for protein quantity and protein quality in seed crops. In: Daussant J., Mosse, J., Vaughan, J. (eds.) Seed Proteins. Academic Press Inc., London, pp. 223-253.

Peerbolte, R., Ruigrok, P., Wullems, G. and Schilperoort, R (1987). T-DNA rearrangements due to tissue culture: somaclonal variation in crown gall tissues. Plant Mol. Biol. **9**, 51-57.

Pelham, H. R. B. (1982). A regulatory upstream promoter element in the Drosophila *Hsp* 70 heat-shock gene. Cell **30**, 517-528.

Peralta, E. G., Hellmiss, R. and Reamm, W. (1986) Overdrive, a T-DNA transmission enhancer on the *A. tumefaciens* tumor-inducing plasmid. EMBO J. **5**, 1137-1142.

Pernollet, J.-C. (1978). Protein bodies of seeds: ultrastructure, biochemistry, biosynthesis and degradation. Phytochemistry **17**, 1473-80.

Petit, A., Berkaloff, A. and Tempe, J. (1986) Multiple transformation of plant cells by *Agrobacterium* may be responsible for the complex organization of T-DNA in crown gall and hairy root. Mol. Gen. Genet. **202**, 388-393.

Plegt, L. and Bino, R. J. (1989). β-Glucuronidase activity during development of the male gametophyte from transgenic and non-transgenic plants. Mol. Gen. Genet. **216**, 321-327.

Potrykus, I., Paszkowski, J., Saul, M. W., Petruska, J. and Shillito, R. D. (1985a) Molecular and general genetics of a hybrid foreign gene introduced into tobacco by direct gene transfer. Mol. Gen. Genet. **199**, 169-77.

Potrykus, I., Saul, M. W., Petruska, J., Paszkowski, J., and Shillito, R. D. (1985b). Direct gene transfer to cells of a graminaceous monocot. Mol. Gen. Genet. **199**, 183-188.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature **322**, 697-701.

Ptashne, M., and Gann, A. F. (1990). Activators and targets. Nature **34622**, 329-331.

Puonti-Kaerlas, J., Eriksson, T. and Engstrom, P. (1990). Production of transgenic pea (*Pisum sativum L.*) plants by *Agrobacterium tumefaciens* - mediated gene transfer. Theor. Appl. Genet. **80**, 246-252.

Randall, P. J., Thomson, J. A. and Shroeder, H. E. (1979). Cotyledonary storage proteins in *Pisum sativum*. IV. Effects of sulfur, phosphorus, potassium and magnesium deficiencies. Aust. J. Plant Physiol. **6**, 11-24.

Reiss, B., Sprengel, R., Will, H. and Schaller, H. (1984). A new sensitive method for qualitative and quantitative assay of neomycin phosphotransferase in crude cell extracts. Gene **30**, 211-218.

Rhodes, C. A., Pierce, D. A., Mettler, I. J., Mescarenhas, D. and Detmer, J. J. (1988). Genetically transformed maize plants from protoplasts. Science , **240**, 204-210.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, B. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**, 237-245.

Robert, L. S., Nozzolillo, C. and Altosaar, I. (1985). Homology between leguminlike polypeptides from cereals and pea. Biochem. J. **226**, 847-852.

Rochester, D., Winer, J. A. and Shah, D. M. (1986). The structure and expression of maize genes encoding the major heat shock protein, *hsp*70. EMBO J. **5**, 451-458.

Rogers, S. G., Horsch, R. B. and Fraley, R. T. (1986). Gene transfer in plants Production of transformed plants using Ti-plasmid vectors. Methods Enzymol. **118**, 627-640.

Rosahl, S., Schell, J. and Willmitzer, L. (1987). Expression of a tuber-specific storage protein in transgenic tobacco plants: Demonstration of an esterase activity. EMBO J. **6**, 1155-1160.

Roussell, D. L., Boston, R. S., Goldsbrough, P. B. and Larkins, B. A. (1988). Deletion of DNA sequences flanking an M_r 19 000 zein gene reduces its transcriptional activity in heterologous plant tissues. Mol. Gen. Genet. **211**, 202-209.

Sanchez-Serrano, J. J., Keil, M., O'Conner, A., Schell, J. and Willmitzer, L. (1987). Wound-induced expression of a potato proteinase inhibitor II gene in transgenic tobacco plants. EMBO J. **6**, 303-306.

Sanders, P. R., Winter, J. A., Barnason, A. R., Rogers, S. G. and Fraley, R. T. (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. Nucleic Acids Res. **15**, 1543-1558.

Schäfer, W., Gorz, A. and Kahl, G. (1987). T-DNA integration and expression in

a monocot crop plant after induction of *Agrobacteriun.* Nature **327**, 529-532.

Schernthaner, J. P., Matzke, M. A. and Matzke, A. J. M. (1988). Endosperm-specific activity of a zein gene promoter in transgenic tobacco plants. EMBO J. **7**, 1249-1255.

Schmid, J., Doerner, P. W., Clonse, S. D., Dixon, R. A. and Lamb, C. J. (1990). Developmental and environmental regulation of a bean chalcone synthase promoter in transgenic tobacco. Plant Cell **2**, 619-631.

Schöffl, F., and Baumann, G. (1985). Thermo-induced transcripts of a soybean heat shock gene after transfer into sunflower using a Ti-plasmid vector. EMBO J. **4**, 1119-1124.

Schöffl, F., Baumann, G., Raschke, E. and Bevan, M. (1986). The expression of heat shock genes in higher plants. Philos. Trans. R. Soc. London Ser. B. **314**, 453-468.

Schöffl, F., Raschke, E. and Nagao, R. T. (1984). The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. EMBO J. **3**, 91-97.

Schröder, G., Waffenschmidt, S., Weiler, E. W. and Schröder, J. (1983). The T-region of Ti-plasmids codes for an enzyme indole-3-acetic acid. Eur. J. Biochem. **138**, 387-391.

Schüle, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988). Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. Nature **332**, 87-90.

Seitz, E. W. and Hochster, R. M. (1964). Lysopine in normal and in crown gall tissue of tomato and tobacco. Can. J. Bot. **42**, 999-1004.

Sengupta-Gopalan, C., Reichert, N. A., Barker, R. F., and Hall, T. C. (1985). Developmentally regulated expression of the bean β-phaseolin gene in tobacco seed. Proc. Natl. Acad. Sci. USA **82**, 3320-3324.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973). Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose. Biochemistry **12**, 3055-3063.

Shaw, C.H., Ashby, A. M., Brown, A., Royal, C., Loake, G. J. and Shaw, C.H. (1988). *virA* and *virG* are the Ti-plasmid functions required for chemotaxis of *Agrobacterium tumefaciens* towards acetosyringone. Molec. Microbiol. **2**, 413-417.

Shaw, C.H., Carter, G.H., Watson, M.D. and Shaw, C.H. (1984a). A functional map of the nopaline synthase promoter. Nucleic Acid Res. **12**, 7831-7846.

Shaw, C.H., Leemans, J., Shaw, C.H., Van Montagu, M. and Schell, J. (1983). A general method for the transfer of cloned genes to plant cells. Gene **23**, 315-330.

Shaw, C.H., Sanders, D.H., Bates, M.R and Shaw, C.H. (1986). Light regulation of a ssRubisco-nos chimaeric gene: photoregulatory control sequences from a C_3 plant function in cells of a CAM plant. Nucleic Acid Res. **14**, 6603-6612.

Shaw, C.H., Watson, M.D., Carter, G.H. and Shaw, C.H. (1984b). The right hand copy of the nopaline Ti-plasmid 25 bp repeat is required for tumour formation. Nucleic Acids Res. **12**, 6031-6041.

Sheikholeslam, S. N. and Weeks, D. P. (1987). Acetosyringone promotes high efficiency transformation of *Arabidopsis thaliana* explants by *Agrobacterium tumefaciens*. Plant Mol. Biol. **8**, 291

Shewry, P. R., Miflin, B. J., Forde, B. J. and Bright, S. W. J. (1981). Conventional and novel approaches to the improvement of the nutritional quality of cereal and legume seeds. Sci. Prog. (Oxford) **65**, 575-600.

Shillito, R. D. and Saul, M. W. (1988). Protoplast isolation and transformation. In Plant Molecular Biology (Shaw, C. H. ed.), IRL Press Ltd, Oxford, pp. 161-186.

Shirsat, A. (1988). A transposon like structure in the 5′ flanking sequence of a legumin gene from *Pisum sativum*. Mol. Gen. Genet. **212**, 129-136.

Shirsat, A., Meakin, P. J. and Gatehouse, J. A. (1990). Sequences 5′ to the conserved 28 bp Leg box element regulate the expression of pea seed storage protein gene *legA*. Plant Mol. Biol. **15**, 685-693.

Shirsat, A., Wilford, N., Croy, R. and Boulter, D. (1989). Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol. Gen. Genet. **215**, 326-331.

Siebertz, B., Logemann, J., Willmitzer, L. and Schell, J. (1989). *cis*-analysis of the wound-inducible promoter *wun*1 in transgenic tobacco plants and histochemical localization of its expression. Plant Cell **1**, 961-968.

Simpson, J., Timko, M. J., Cashmore, A. R., Schell, J., Van Montagu, M. and Herrera-Estrella, L. (1985). Light-inducible and tissue specific expression of a chimaeric gene under control of the 5′ flanking sequence of a pea chlorophyll a/b binding protein gene. EMBO J. **4**, 2723-2729.

Simpson, J., Schell, J., Van Montagu, M. and Herrera-Estrella, L. (1986a) Light-inducible and tissue-specific pea *lhcp* gene expression involves an upstream element combining enhancer- and silencer-like properties. Nature **323**, 551-554.

Simpson, J., Van Montagu, M. and Herrera-Estrella, L. (1986b) Photosynthesis-associated gene families: differences in response to tissue-specific and environmental factors. Science **233**, 34-38.

Slightom, J. L., Sun, S. M. and Hall, T. C. (1983). Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc. Natl. Acad. Sci. USA **80**, 1897-1901.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-512.

Spielmann, A. and Simpson, R. B. (1987). Physical structure and genetics of the T-DNA in plants transformed by *Agrobacterium tumefaciens*. In Plant-Microbial Interactions , Verma eds. p.27-31.

Spena, A. and Schell, J. (1987). The expression of a heat-inducible chimeric gene in transgenic tobacco plants. Mol. Gen. Genet. **206**, 436-444.

Spena, A., Hain, R., Ziervogel, U., Saedler, H. and Schell, J. (1985) Construction of a heat-inducible gene for plants. Demonstration of heat-inducible activity of the *Drosophila melanogaster hsp* 70 promoter in plants. EMBO J. **4**, 2739-2743.

Spencer, D. (1984). The physiological role of storage proteins in seeds. Phil. Trans. Roy. Soc. Lond. B. **304**, 275-285.

Spencer, D. and Higgins, T. J. V. (1980) The biosynthesis of legumin in maturing pea seeds. Biochem. Int. **1**, 502-509.

Stachel, S. E. and Zambryski, P. (1986) *VirA* and *virG* control the plant-induced activation of the T-DNA transfer process of *Agrobacterium tumefaciens.* Cell bf 46, 325-333.

Stachel, S. E., Messens, E., Van Montagu, M. and Zambryski, P. (1985) Identification of the signal molecules produced by wounded plants that activate T-DNA transfer in *Agrobacterium tumefaciens.* Nature **318**, 624-629.

Stockhaus, J., Eckes, P., Rocha-Sosa, M., Schell, J. and Willmitzer, L. (1987) Analysis of cis-active sequences involved in the leaf specific expression of a potato gene in transgenic plants. Proc. Natl. Acad. Sci. USA **84**, 7943-7949.

Stougaard, J., Sandal, N. N., Gron, A., Kuhle, A. and Marker, K. A. (1987). 5′ analysis of the soybean leghemoglobin lbc_3 gene: Regulatory elements required for promoter activity and organ specificity. EMBO J. **6**, 3565-3570.

Strittimatter, G. and Chua, N.-H. (1987). Artificial combination of two cis-regulatory elements generates a unique pattern of expression in transgenic plants. Proc. Natl. Acad. Sci. USA **84**, 8986-8990.

Sumrada, R. A. and Cooper, T. G. (1985). Point mutation generates constitutive expression of an inducible eukaryote gene. Proc. Natl. Acad. Sci. USA **82**, 643-647.

Szabados, L., Ratel, P., Grunenberg, B. and de Bruijn, F. J. (1990). Functional analysis of the *Sesbania rostrata* leghemoglobin *glb*3 gene 5′ -upstream region in transgenic *Lotus corniculatus* and *Nicotiana tabacum* plants. Plant Cell **2**, 973-986.

Tabata, T., Nakayama, T., Mikami, K. and Iwabuchi, M. (1991). HBP-1a and HBP-1b: leucine zipper type transcription factors of wheat. EMBO J. **10**, 1459-1467.

Takahashi, T. and Komeda, Y. (1989). Characterization of two genes encoding small heat-shock proteins in *Arabidopsis thaliana*. Mol. Gen. Genet. **219**, 365-372.

Takaiwa, F. and Oono, K. (1990). Interaction of an immature seed-specific *trans*-acting factor with the 5′ upstream region of a rice glutelin gene. Mol. Gen. Genet. **224**, 289-293.

Taylor, J. L., Jones, J. D. G., Sandler, S., Mueller, G. M., Bedbrook, J. and Dunsmuir, P. (1987). Optimizing the expression of chimeric genes in plant cells. Mol. Gen. Genet. **210**, 572-577.

Thomas, M. S. and Flavell, R. B. (1990). Identification of an enhancer element of the endosperm specific expression of high molecular weight glutenin. Plant Cell **2**, 1171-1180.

Thomashow, L. S., Reeves, S. and Thomashow, M. F. (1984). Crown gall oncogenesis: evidence that a T-DNA gene from the *Agrobacterium* Ti-plasmid pTiA6 encodes an enzyme that catalyzes synthesis of indoleacetic acid Proc. Natl. Acad. Sci. USA **81**, 5071-5075.

Thomashow, M. F., Hughly, S., Buchholz, W. and Thomashow, L. S. (1986) Molecular basis for the auxin independent phenotype of crown gall tumor tissues. Science **231**, 616-618.

Thompson, W. F., Everett, M., Polans, N. O., Jorgensen, R. A., Palmer, J. D.

(1983). Phytochrome control of RNA levels in developing pea and mung-bean. Planta **158**, 487-500.

Thornburg, R. W., An, G., Cleveland, T. E., Johnson, R. and Ryan, C. A. (1987) Wound inducible expression of a potato inhibitor II-chloramphenicol gene fusion in transgenic tobacco plants. Proc. Natl. Acad. Sci. USA **84**, 744-748.

Timko, M. P, Kausch, A. P., Castresana, C., Fassler, J., Herrera-Estrella, L., Van den Broeck, G., Van Montagu, M., Schell, J. and Cashmore, A. (1985). Light regulation of plant gene expression by an upstream enhancer-like element. Nature **318**, 579-582.

Tobin, E. M. (1981). Phytochrome-mediated regulation of messenger RNAs for the small subunit of ribulose-1,5-bisphosphate carboxylase and the light-harvesting chlorophyll a/b-protein in *Lemna gibba* Plant Mol. Biol. **1**, 35-51.

Tobin, E. M. and Silverthorne, J. (1985). Light regulation of gene expression in higher plants. Ann. Rev. Plant Physiol.**36**, 569-93.

Tokuhisa, J. G., Singh, K., Dennis, E. S. and Peacock, W.J. (1990). A DNA-binding protein factor recognizes 2 binding domains within the octopine synthase enhancer element. Plant Cell **2**, 215-224.

Tokunaga, K., Hirose, S. and Susuki, K. (1984). In monkey cos cells the TATA box and the cap site region are required for faithful and efficient initiation of the fibroin gene transcription. Nucleic Acids Res. **12**, 1543-58.

Tsai, S. Y., Tsai, M-J. and O' Malley, B. (1981). Specific 5′ flanking sequences are required for faithful initiation of *in vitro* transcription of the ovalbumin gene. Proc. Natl. Acad. Sci. USA **78**, 879-83.

Tumer, N. E., Richter, J. D. and Nielsen, N. C. (1982). Structural characterization of the glycinin precursors. J. Biol. Chem. **257**, 4016-4018.

Twell, D. and Ooms, G. (1987). The 5′ flanking DNA of a patatin gene directs tuber-specific expression of a chimeric gene in potato. Plant Mol. Biol. **9**, 365-

376.

Uchimiya, H., Fushimi, T., Hashimoto, H., Harada, H., Syono, K. and Sugawara, Y. (1986) Expression of a foreign gene in callus derived from DNA-treated protoplasts of rice (*Oryza sativa L.*). Mol. Gen. Genet. **204**, 204-207.

Van Haute, E., Joos, H., Maes, M., Warren, G., Van Montagu, M. and Schell, J. (1983). Intergenic transfer and exchange recombination of restriction fragments cloned into pBR322: a novel strategy for the reversed genetics of Ti plasmids of *Agrobacterium tumefaciens*. EMBO J. **2**, 411-418.

Van Lurebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Zaenen, I., Schilperoort, R. A., and Schell, J. (1974). Large plasmid in *Agrobacterium tumefaciens*essential for crown gall inducing ability. Nature **252**, 169-170.

Vernet, T., Fleck, J., Durr, A., Fritsch, C., Pinck, M. and Hirth, L. (1982). Expression of the gene coding for the small subunit of ribulose bisphosphate carboxylase during differentiation of tobacco protoplasts. Eur. J. Biochem. **126**, 489-495.

Vierra, J. and Messing, J. (1982). The pUC plasmids, ane M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**, 259-268.

Vodkin, L. O., Rhodes, P. R. and Golbdberg, R. B. (1983). A lectin gene insertion has the structural feature of a transposable element. Cell **32**, 1023-1031.

Voelker, T. A., Staswick, P. and Chrispeels, M. J. (1986) Molecular analysis of two phytohaemagluttunin genes and their expression in *Phaseolus vulgaris* cultivar pinto, a lectin deficient aultivar of the bean. EMBO J. **5**, 3075-3082.

Voelker, T. A., Sturm, A. and Chrispeels, M. J. (1987). Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. EMBO J. **6**, 3571-3577.

Voss, S. D., Schlokat, U. and Gruss, P. (1986) The role of enhancers in the regulation of cell-type-specific transcriptional control. Trends in Biological Sci. **11**, 287-

289.

Walburg, G. and Larkins, B. A. (1983). Oat seed globulin: subunit characterization and demonstration of its synthesis as a precursor. Plant Physiol. **72**, 161-165.

Walker, J. C., Howard, E. A., Dennis, E. S. and Peacock, W. J. (1987). DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase 1 gene. Proc. Natl. Acad. Sci. USA **84**, 6624-6628.

Wang, K., Herrera-Estrella, L., Van Montagu, M. and Zambryski, P. C. (1984) Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer to the plant genome. Cell **38**, 455-462.

Watson, B., Currier, T. C., Gordon, M. P., Chilton, M.-D. and Nester, E. W. (1975) Plasmid required for virulence of *Agrobacterium tumefaciens.* J. Bacteriol. **123**, 255-264.

Wang, K., Genetello, C., Van Montagu, M. and Zambryski, P. C. (1987). Sequence context of the T-DNA border repeat element determines its relative activity during T-DNA transfer to plant cells. Mol. Gen. Genet. **210**, 338-346.

Weiher, H., König, M. and Gruss, P. (1983). Multiple point mutations affecting the Simian virus 40 enhancer. Science **219**, 626-631.

Weinberger, J., Baltimore, D. and Sharp, P. A. (1986) Distinct factors bind apparently homologous sequences in the immunoglobulin heavy-chain enhancer. Nature **322**, 846-848.

Wend-Gallitelli, M. F. and Dobrigkeit, I. (1973). Investigations implying the invalidity of octopine as a marker for transformations by *Agrobacterium tumefaciens.* Z. Naturforsch **28c**, 768-771.

Werr, W. and Lörz, H. (1986) Transient gene expression in a gramineae cell line. Mol. Gen. Genet. **202**, 471-475.

Werr, W., Frommer, W.-B., Maas, C. and Starlinger, P. (1985). Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. EMBO J. **4**, 1373-1380.

Willmitzer, L., Simons, G. and Schell, J. (1982). TL-DNA in octopine crown gall tumors codes for 7 well-defined polyadenylated transcripts. EMBO J. **1**, 139-146.

Willmitzer, L., Dhaese, P., Schreier, P., Schmalenbach, W., Van Montagu, M. and Schell, J. (1983). Size, location and polarity of T-DNA-encoded transcripts in nopaline crown gall tumors, common transcripts in octopine and nopaline tumors. Cell **32**, 1045-1056.

Wingender, E. (1988). Compilation of transcription regulating proteins. Nucleic Acids Res. **16**, 1879-1902.

Wobus, U., Bäumlein, H., Bassüner, R., Heim, U., Jung, R., Müntz, K., Saalbach, G. and Weschke, W. (1986). Characterization of two types of legumin gene in the field bean (*Vicia faba* L. var. minor.) genome as revealed by cDNA analysis. FEBS Lett. **201**, 74-80.

Wright, D. J. and Boulter, D. (1974). Purification and subunit structure of legumin of *Vicia faba* L. (Broad bean). Biochem. J. **141**, 413-418.

Wu, B. J., Kingston, R. E. and Morimoto, R. I. (1986) Human *Hsp*70 promoter contains at least two distinct regulatory domains. Proc. Natl. Acad. Sci. USA **83**, 629-633.

Wu C. (1984) Activating protein factor binds *in vitro* to upstream control sequences in heat shock gene chromatin. Nature **311**, 81-84.

Yamazaki, K., Katagiri, F., Imaseki, H. and Chua, N-. H. (1990). *Tga*1A, a tobacco DNA-binding protein, increases the rate of initiation in a plant *invitro* transcription system. Proc. Natl. Acad. Sci. USA **87**, 7035-7039.

Yanofsky, M., Porter, S., Young, C., Albright, L., Gordon, M. and Nester, E. (1986) The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific

endonuclease. Cell **47**, 471-477.

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

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 altering their regeneration capacity. EMBO J. **2**, 2143-2150.

Zambryski, P., Herrera-Estrella, L., De Block, M., Van Montagu, M. and Schell, J. (1984). in Genetic Engineering, Principles and Methods, Vol 6. Setlow, J., and Hollaender, A., (eds.), Plenum Press, New York, p. 253.

UNIV DUNELM