National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

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

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

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

AVIS

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

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

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

Canada

Oat seed storage protein genes: promoter studies in transgenic tobacco plants.

Bernard Potier

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfilment of the requirements for the
Ph. D. degree in the
Ottawa-Carleton Institute of Biology

1993

ISBN 0-315-93605-3

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Acknowledgments

This work was conducted in the laboratory of Dr. Illimar Altosaar whom I thank for giving me such an opportunity.

I would also like to express my gratitude to Dr. Connie Nozzolillo for her continued and persistent support, especially during the writing part of this thesis where Dr. Nozzolillo was of great help.

The financial support was provided by a Government of Canada Scholarship, administered by the World University Service of Canada and later by the International Council for Canadian Studies. Many thanks also to the School of Graduate Studies for their awards.

To all the members of my advisory committee, I wish to express my sincere thanks: Dr. L. Boisen, Dr. V. Burrows, Dr. K. Joy, Dr. L. Robert and Dr. J. Vierula.

For their expert advice in tobacco transformation and in plasmid sequencing, I thank Mrs. P. Donaldson and Dr. D. Albani.

For her careful and prompt photographic work, I thank Mrs. E. Szabo.

For their constant availability, I thank Drs. G. Drouin and D. Johnson, as well as all members of the Biology and Biochemistry Departments.

For her patient and careful work on tRNAs, I express my gratitude to Ms. Amy Yamazaki who was also a very pleasant person to work with.

Many thanks to all those not mentioned here but who were also helpful in the accomplishment of this work.

Abstract

The seed storage proteins of oat are mainly represented by the globulins (75%) and the prolamins (10%). These proteins are only found in the endosperm and accumulate within protein bodies during the maturation of the oat seeds.

Three genomic sequences encoding globulin polypeptides and one avenin genomic sequence encoding an oat prolamin (avenin) have been isolated and characterized.

The respective promoters of these genomic clones were fused to the coding sequence of the ß-glucuronidase reporter gene (GUS). These constructs, together with the entire sequence of a globulin gene, were transferred into tobacco via *Agrobacterium tumefaciens*. Analysis of the transgenic tobacco seeds showed for the first time that the promoters of the oat globulin and avenin genes are able to regulate the expression of the GUS sequence in an endosperm-specific and developmentally controlled manner in a dicot plant, as in oat seeds with the original seed storage protein gene.

One of the globulin promoters was shown to be probably inactive, whereas two other promoters appear to direct strong expression in the seeds of transgenic tobacco plants. A deletion analysis on one of the functional promoters demonstrated that a portion of the promoter upstream of nucleotide -259 (relative to the start of transcription as determined in oat) was required for expression. Sequence analysis of the globulin promoters showed the lack of conserved elements which are found in other storage protein gene promoters, and believed to play an important role in the regulation of seed storage protein genes. It was nonetheless demonstrated in this study that the absence of such elements did not prevent a correct functionality of the oat globulin promoters in transgenic tobacco seeds.

The avenin promoter, when fused to the GUS sequence, also showed strong expression in the endosperm of transgenic tobacco seeds. Sequence analysis of the upstream region of the avenin gene confirmed the presence of a highly conserved 'prolamin box'. The possible role of this element was therefore further demonstrated in this work.

This work has also showed for the first time a difference in the choice of transcriptional start sites of two monocot (oat) seed storage protein genes after transfer into a dicot species (tobacco).

Résumé.

Les protéines de réserve des graines d'avoine sont principalement représentées par les globulines (75%) et les prolamines (10%). Ces protéines sont uniquement observées dans l'albumen et s'accumulent dans des corpuscules protéiques durant le développement des graines d'avoine.

Trois séquences génomiques de globulines et une de prolamine (avénine) ont été isolées et caractérisées.

Les promoteurs respectifs de ces clones génomiques ont été fusionnés à la séquence codante du gène 'reporter' ß-glucuronidase (GUS). Ces constructions, ainsi que la sequence complète d'un gène de globuline, ont été transférées dans le tabac par l'intermédiaire d'*Agrobacterium tumefaciens*. L'analyse des graines de tabac transgénique a montré pour la première fois que les promoteurs des gènes de globuline et d'avénine sont capables de réguler l'expression du gène GUS de manière spécifique, pendant la maturation des graines et dans l'albumen. Cette régulation est similaire à celle observée pour ces gènes dans l'avoine même.

Il a été révélé qu'un des promoteurs des gènes de globulines était probablement inactif, alors que les deux autres promoteurs paraissent diriger une forte expression dans les graines de tabac transgénique. Une analyse par délétion sur un des promoteurs fonctionnels a démontré qu'une partie en amont du nucléotide -259 (relativement à la position du site de début de transcription qui a été déterminé pour l'avoine) était requise pour diriger une expression. L'analyse des séquences de promoteurs de gènes de globulines a permis de démontrer l'absence d'éléments conservés identifiés dans des promoteurs d'autres gènes de protéines de réserve. On pense que ces éléments jouent un rôle important dans la régulation des gènes de protéines de réserve.

Il a néanmoins été démontré dans cette étude que l'absence de tels éléments conservés n'empêchait pas le fonctionnement correct des promoteurs des gènes de globulines d'avoine dans des graines de tabac transgénique.

Le promoteur du gène d'avoine a aussi montré une forte expression dans l'albumen des graines de tabac transgénique après avoir été fusionné à la séquence codante de GUS. L'analyse de la séquence promotrice du gène d'avénine a confirmé la présence d'un élément très conservé, la"boîte prolamine". Le fait que cet élément puisse jouer un rôle a ainsi été encore confirmé.

L'étude qui a été réalisée a aussi montré pour la première fois une différence dans le choix d'un site de début de transcription de deux gènes de monocotylédones (avoine) après avoir été transférés dans une dicotylédone (tabac).

Table of contents

List of figures.

List of tables.

List of abbreviations

AP	Alkaline phosphatase
BA	6-benzyladenine
BCIP	5-bromo-4-chloro-3-indolylphosphate p-toluidine salt
Bis-Acrylamide	N,N'-methylene bisacrylamide
bp	Base pair
CaMV	Cauliflower mosaic virus
CAT	Chloramphenicol acetyltransferase
cDNA	Complementary DNA
CIP	Calf intestine phosphatase
CDTA	trans-1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid
DAP	Day after pollination
DEPC	Diethylpyrocarbonate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
ER	Endoplasmic reticulum
GUS	ß-glucuronidase
HMW	High molecular weight
IEF	Isoelectric focusing
Kan^R	Kanamycin resistance
kDa	kilo Dalton
LB	Luria Bertani medium
LMW	Low molecular weight
mRNA	Messenger RNA
MS	Murashige-Skoog
MUG	4-methyl umbelliferyl ß-D-glucuronide
MW	Molecular weight
NAA	alpha-naphthalene acetic acid
NBT	Nitroblue tetrazolium chloride
NOSpro	Nopaline synthase promoter
NOSter	Nopaline synthase terminator
NPTII	Nopaline phosphotransferase II
nt	Nucleotide
PCR	Polymerase chain reaction
pI	Isoelectric point
PVPP	Polyvinyl-polypyrrolidone

rpm	Revolutions per minute
RT-PCR	Reverse transcriptase-polymerase chain reaction
SDS	Sodium dodecyl sulfate
SSC	Sodium chloride and sodium citrate solution
TAE	Tris-acetic acid-EDTA buffer
TBS	Tris-buffered saline
T-DNA	Transfer DNA
TEMED	N,N,N',N'-tetramethylethylenediamine
TRIS	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UTR	Untranslated region
X-Gluc	5-bromo-4-chloro-3-indolyl glucuronide
YEP	Yeast extract, peptone and NaCl medium

Chapter 1
Introduction

1.1. Oat seed storage proteins.

1.1.1. Oat plants.

Among the cereals, oats (family Poaceae or grass family) are considered a secondary crop. They are cultivated under temperate climates, cool and moist, particularly in Northern and Eastern Europe, and in northern North America. The most cultivated form, *Avena sativa*, accounts for about three quarters of the world production, which tends to stabilize around 45 to 50 million metric tons per year. Oats are used for pasture crop or livestock feed (78%), for human food (18%) and for industrial and seed purposes (4%) (Schrickel 1986).

1.1.2. Oat seed storage proteins.

1.1.2.1. Storage proteins.

By definition, grains are the fruits of the members of the grass family. In this study, the term seed will refer to a grain which has been stripped at maturity from its removable floral parts (glumes, palea and lemma, together termed as hulls). The seed (or groat in the case of oat) will therefore be composed of the embryo (or germ), the endosperm, the aleurone layer and the bran, which results from the fusion of the seed coat and the ovary wall.

The embryo and the endosperm are two distinct tissues which originate from the double fertilization occurring in angiosperm plants. In contrast to most legumes where the endosperm is rapidly absorbed by the cotyledons, the endosperm is the major storage tissue of cereal seeds. It contains large amounts of starch (in the starch granules), proteins (in the protein bodies) and also fat and ß-glucans (gums).

Storage proteins are defined by their seed-specificity, their temporal accumulation during seed development and their localization within protein bodies (in the endosperm and/or embryo) (Derbyshire *et al.* 1976). Their only known biological function during the plant life cycle is to provide the developing seedlings with nitrogen (also, to a lesser extent, with carbon and sulphur).

The most abundant storage proteins in cereals are represented by the prolamin class (alcohol-soluble proteins). This fraction accounts for about 50% of the total storage proteins in wheat, barley, rye and maize. Part of this class in wheat is composed of the glutenins which confer the bread-making quality to the wheat flour. The second important class of storage proteins is called globulin and this fraction, which constitutes about 4 to 20% of the total seed proteins in cereals other than oat and rice, is extracted with a salt solution. The remaining proteins are classified as albumins for the water-soluble proteins and as glutelins for the proteins not extracted with either water, salt solution or alcohol (Osborne 1924).

In contrast to of wheat, rye, barley and maize, the major class of seed storage proteins in oat and rice is represented by the globulins. Most of the cereal grains show a deficiency in some essential amino acids, mainly lysine, and also tryptophan (maize), threonine (barley) and phenylalanine and tyrosine (sorghum) (Pernollet and Mossé, 1983). Since globulins have a better amino acid balance, the nutritional value of oat and rice is therefore enhanced (Maruyama *et al.* 1975 and Wu *et al.* 1972). Oat globulins have been commonly estimated to make up about 50% of the total seed storage proteins. However, since it has been shown that the major part of glutelins is actually composed of globulins, the total percentage of globulins has been increased to 75% (Robert *et al.* 1985b).

1.1.2.2. Oat globulins.

The oat globulin fraction, usually extracted by a 1M NaCl solution (Galle *et al.* 1988) can be separated into 3S, 7S and 12S fractions by sucrose density gradient ultracentrifugation (Burgess *et al.* 1983). The 3S fraction (MW< 20 kDa) contains protease inhibitors and some unidentified proteins (Pernollet *et al.* 1989). The 7S fraction is composed of ≈55 kDa proteins that are not dissociated when a reducing agent is added. The 12S fraction, which is the major one, gives rise, upon SDS-PAGE, to proteins with MW ranging from 52 to 70 kDa. When a reducing agent is added, these proteins dissociate into two classes of polypeptides: the α polypeptides (acidic, pI between 5.0 and 7.0, MW ranging from 35 to 42 kDa) and the ß polypeptides (basic, pI between 8.0 and 9.0, MW ranging from 20 to 25 kDa) (Burgess *et al.* 1983). A model for the arrangement of these subunits has been proposed (Peterson 1978) where the native globulins are composed of six monomers, one unit being the result of a covalent linkage by disulfide bridges between the α and ß polypeptides, thus giving rise to a 320 to 360 kDa hexamer (12S protein). A more detailed analysis of the posttranslational events in glycinin assembly (glycinin being the 12S globulin in soybean) has been conducted (Dickinson *et al.* 1987 and 1989). It appears that the posttranslational cleavage of the proglycinin into acidic and basic polypeptide chains plays an important role in the assembly of glycinins first into trimers and then into hexamers.

Oat globulins begin to accumulate within the protein bodies of the endosperm about one week to 12 days after anthesis (Robert *et al.* 1983, Luthe 1987). They are synthesized as precursors (signal peptide, α and ß polypeptide) on membrane-bound polysomes, sequestered in the ER, transported into protein bodies where they are cleaved into α and ß polypeptides, and then associate to form the 12S oligomer (Adeli *et al.* 1984, Bednarek and Raikhel 1992). The exact mode of transportation of the precursors into the protein bodies is not

clear yet: the involvement of dictyosomes has not been demonstrated, and because 12S globulins are not glycosylated, there are no reasons to implicate them.

Oat globulins share structural features (Brinegar and Peterson 1982) as well as antigenic homologies (Fabijanski *et al.* 1985 and Robert et al. 1985) with other cereal globulins and dicot legumins. The presence of globulin-like polypeptides has been well documented in dicot as well as in monocot seeds (Derbyshire *et al.* 1976, Casey and Domoney 1987, Luthe 1992). The globulin seed storage proteins of angiosperms are thought to derive from two ancestral genes (Borroto and Dure 1987). The relatively recent availability of numerous additional nucleotide sequences (from cDNA and genomic clones) are confirming these findings.

1.1.2.3. Oat prolamins.

Oat prolamins, called avenins, represent 5 to 15% of the total seed storage proteins. The microheterogeneity of avenins has been well characterized (Kim *et al.* 1978, Robert *et al.* 1983) and the N-terminal protein sequences have also been established for several members of this class (Bietz 1982, Pernollet *et al.* 1987, Pernollet *et al.* 1989) together with the complete amino acid sequence of an avenin (Egorov 1988). The amino acid composition of the avenins shows a high proportion of proline and glutamine residues (hence their name of prolamins). Prolamins of different cereal species are homologous and it has been suggested that they originated from a common ancestor (Heidecker and Messing 1986, Casey and Domoney 1987, Kreis and Shewry 1989, Shewry and Tatham 1990). Prolamin-type proteins are found only in monocotyledonous species.

Accumulation of avenins in the endosperm begins about two weeks after anthesis and reaches a maximum at about four weeks (Robert *et al.* 1983).

1.1.3. Molecular biology aspects of oat seed storage protein genes.

As determined at the protein level, seed storage protein genes belong to large multigene families. Access to the primary structure of the proteins has been achieved by studying the genes that code for them. Numerous genes and cDNAs encoding dicot and monocot seed storage proteins have been isolated and characterized. The structures, compositions and homologies of these proteins have been widely confirmed at the DNA level. Thus, concerning oat globulin, an average of 40% and 70% similarity respectively was reported when DNA coding sequences were compared to those of dicot 11S legumins and rice glutelin (Shotwell *et al.* 1988, Nielsen *et al.* 1989, Wen and Luthe 1985). In addition, a hypervariable region at the COOH terminus of the acidic polypeptide was identified. The variability occurring in this region reflects the higher heterogeneity of the α subunit when compared to the ß subunit, as shown by IEF (Burgess *et al.* 1983). As for the avenins, the homology between various prolamins of different cereal species has been confirmed and attempts were made to explain the possible evolution of these genes from a common ancestor (Kreis *et al.* 1985).

The relatively similar levels of mRNAs encoding oat globulins (18S mRNA) and avenins (15S and 12S mRNA) do not reflect the proportions of synthesized proteins in mature seeds (Chesnut *et al.* 1989), nor do they correspond to the total copy number of their respective multigene families: about 300 copies for the globulin sequences and approximately 150 for the avenin sequences (Chesnut *et al.* 1989). Evidence for a translational control of storage proteins in oat seeds has been reported, suggesting that the final high proportion of globulins in oat seeds is mainly regulated by the translation elongation or termination reactions (not the initiation rates) occurring in the endosperm (Fabijanski and Altosaar 1985, Boyer *et al.* 1992).

The promoter region of an oat globulin genomic clone (OG1-E1, Shotwell *et al.* 1990) was shown not to possess any of the possible regulatory elements thought to be involved in the tissue-specificity and temporal storage protein gene expression. However, an avenin clone, AV45-X1 (Shotwell *et al.* 1990) exhibits the highly conserved prolamin box or '-300 element' (see section 1.2.2.). An oat globulin promoter fused to the GUS coding sequence has been reported to be expressed in the embryos of transgenic tobacco seeds (Heim *et al.* 1991). However, because of the nature of the report (poster abstract), very little information is available.

Because seed storage protein genes are highly regulated, spatially, temporally and translationally, they are one of the model systems for plant gene expression studies. The expression of seed storage protein genes, and other constructs using their promoters, in heterologous systems (i.e. transgenic plants) can act as a useful tool in determining the underlying mechanisms of developmental and spatial (seed-specific) gene regulation.

1.2. Molecular biology aspects of the expression of seed storage protein genes from other monocotyledonous and dicotyledonous plants.

Upstream DNA sequences of eukaryotic genes have been widely reported to play an important role in the regulation of gene expression. Recent developments in molecular biology (especially during the past ten years) have permitted the isolation and manipulation of specific DNA sequences. In addition, special vectors and methods have allowed the production of transgenic plants (Horsch *et al.* 1985, Bevan 1984, Murashige and Skoog 1962, Jefferson 1987) where the isolated genes can be further studied.

1.2.1. Seed-specific gene promoter in dicots.

Seed storage protein genes from dicot species were the first to be characterized and introduced in transgenic plants via *Agrobacterium tumefaciens*. Table 1 summarizes most of the genes and their promoters that have been transferred into tobacco or other plants[1]. It soon appeared that the regulation of the expression of dicot storage protein genes or reporter genes under the control of storage protein gene promoters in transgenic plants was very similar to that observed in the original plant. The tissue-specificity is maintained, together with the developmental regulation, in the embryos of transgenic plants. However, differences occurred, notably in the processing of the proteins when intact genes or 'minigenes' were used to produce transgenic plants. Examples were given where the size of the proteins found in transgenic seeds was smaller than the expected ones (i.e. the proteins found in the original host plant). This was the case with a pea convicilin (Newbigin *et al.* 1990), the ß-subunit of phaseolin (Sengupta-Gopolan *et al.* 1985), the α' subunit of ß-conglycinin (Beachy *et al.* 1985), a pea vicilin (Higgins *et al.* 1988), a ß-phaseolin (Bagga *et al.* 1992) and four phaseolin glycoforms (Bustos *et al.* 1991). In all these examples, the percentage of correctly processed proteins varies around 50%. Since all the processed proteins were found to be smaller in size to the expected one, it is admitted that the modifications occurred from a fully and correctly expressed protein. The changes are then thought to be post-translational and to reflect the differences present in an heterologous plant (cleavage sites, glycosylation events or protease sensitivity). However, other examples were reported where the protein precursor was correctly processed: pea legumin (Ellis *et al.* 1988), phaseolin 'minigene' (Chee *et al.* 1986). When the transcriptional start site of the introduced gene was

[1] Other seed-specific genes have been transferred into tobacco with the more specific purpose of studying the protein targeting mechanisms (Voelker et al. 1987, Sturm et al. 1988, Greenwood and Chrispeels 1985)

Table 1: Seed storage protein genes or constructs with their promoters that have been transferred into tobacco, petunia, alfalfa or sunflower.

	Type of construct	Reference
A: Dicot genes.		
Phaseolin (French bean)	minigene	Chee *et al.* 1986
ß-conglycinin (soybean)	promoter deletions-GUS	Chen *et al.* 1986
ß-conglycinin (soybean)	promoter sequence inserted in the 35S CaMV promoter-CAT	Chen *et al.* 1992
ß-conglycinin (soybean)	promoter-GUS	Chamberland *et al.* 1992
ß-conglycinin (soybean)	intact gene	Beachy *et al.* 1985
ß-subunit of ß-conglycinin (soybean)	intact gene	Bray *et al.* 1987
ß-phaseolin (French bean)	promoter-GUS	Bustos *et al.* 1989
ß-phaseolin (French bean)	minigenes	Bustos *et al.* 1991
ß-phaseolin (French bean)	intact gene + promoter deletions	Burow *et al.* 1992
ß-phaseolin (French bean)	promoter deletions-GUS	Sen *et al.* 1993
ß-phaseolin (French bean)	35S CaMV promoter-Phaseolin coding sequence	Bagga *et al.* 1992
ß-phaseolin (French bean)	modified coding sequence	Hoffman *et al.*1988
ß-phaseolin (French bean)	intact gene	Sengupta-Gopalan *et al.* 1985
Glycinin (french bean)	promoter deletions-GUS	Lelievre *et al.* 1992
Legumin (pea)	intact gene	Ellis *et al.* 1988 Shirsat *et al.* 1989
Vicilin (pea)	intact gene	Higgins *et al.*1988
Convicilin (pea)	intact gene	Newbigin *et al.*1990
Phytohemagglutinin and phaseolin (French bean)	promoters-GUS	Riggs *et al.* 1989
Helianthinin (sunflower)	promoter-GUS	Jordano *et al.* 1989
Legumin (Faba bean)	promoter deletions-NPTII	Bäumlein *et al.* 1991
Legumin (Faba bean)	promoter deletions-NPTII	Bäumlein *et al.* 1992
Legumin (pea)	promoter deletion	Reric *et al.* 1991
B: Monocot genes.		
Glutelin (rice)	promoter-CAT	Leisy *et al.* 1990
Glutelin (rice)	promoter deletions-CAT	Takaiwa *et al.* 1991
Zein (maize)	ß-phaseolin promoter-15kDa zein coding seq.	Hoffman *et al.* 1987
Zein (maize)	promoter-GUS (and CAT)	Schernthaner *et al.* 1988
Zein (maize)	promoter deletions-GUS	Matzke *et al.* 1990
Zein (maize)	intact gene	Ueng *et al.* 1988
HMW glutenin (wheat)	promoter deletions-CAT	Colot *et al.* 1987
HMW glutenin (wheat)	promoter deletions-GUS	Thomas and Flavell 1990
HMW glutenin (wheat)	promoter deletions-CAT and intact gene	Robert *et al.* 1989
HMW glutenin	promoter deletions-GUS	Halford *et al.* 1989
B hordein (barley)	promoter-CAT	Marris *et al.* 1988

determined in the transgenic tobacco seeds, it appears that it was not different from the one in the original plant (pea vicilin gene, Higgins *et al.* 1988 and phaseolin, Chee *et al.* 1986). From these experiments, it has been shown that expression of seed storage protein genes in a new host remains seed(embryo)-specific and developmentally regulated but that the post-translational processes vary depending on the expressed proteins.

The use of reporter genes (GUS, CAT, NPTII or luciferase) has permitted a deeper focus on the role of the promoter itself in the regulation of gene expression because of an easier detection and localization of the gene products. Promoter deletion analyses were also used in attempts to specifically characterize the minimum length of promoter region and/or the portion of the promoter responsible for the seed-specific expression of the storage protein genes. A second approach was also used for the isolation and characterization of these *cis*-acting elements: computer analyses of the available promoter regions of seed storage protein genes were performed in order to define conserved regions that might be involved in the specific regulation of these genes.

Such a conserved element has been found in the 7S-storage protein gene promoters and was called the "vicilin box" (Gatehouse *et al.* 1986 and Higgins *et al.* 1988) but proved to be insufficient to act as a seed-specific promoter (Newbigin and Higgins 1989) and required additional upstream sequences for correct regulation.

Another putative regulatory element was identified by Bäumlein *et al.* (1986) in the promoter regions of legumin genes. This conserved sequence of approximatively 30 bases is different from the "vicilin box" and was called the "legumin box". This box is approximately located in the -100 bp-region upstream of the transcriptional start point. Centrally located within the "legumin box" lies another very conserved element, the CATGCATG box, also called RY repeat (Dickinson *et al.* 1988). The RY repeat is widely present in numerous

legume seed gene promoters at different locations within the 300 bp upstream of the transcriptional start site. Experiments have shown that the RY repeat is necessary but not sufficient for proper tissue-specific regulation (Bäumlein *et al.* 1991 and 1992, Lelievre *et al.* 1992, Chamberland *et al.* 1992)). Additional upstream elements affecting the level of expression are required (Shirsat *et al.* 1989 and Rerie *et al.* 1991).

Other results obtained with promoter deletion analysis of an α' subunit of ß-conglycinin gene suggest a functional role played by repeated sequences in the proximal region of the promoter (from -250 to -200 bp upstream of the transcriptional start site) (Chen *et al.* 1986). The repeated sequences thought to regulate the embryo-specific expression are composed of 6 bp (AA/G/CCCCA) and have been shown to increase gene expression in embryos, when inserted within the 35S CaMV promoter controlling a reporter gene (CAT) (Chen *et al.* 1988).

A *cis*-acting element from a French bean phaseolin gene promoter was also reported to regulate the embryo-specific expression of GUS when this element (A/T-rich) was inserted in the 35S CaMV promoter (Bustos *et al.* 1989). This element (found between positions -682 and -628) also binds nuclear proteins. However, additional elements are required for expression only in the embryos: the A/T-rich region identified by Bustos *et al.* (1991) also increases the expression in roots. It has also been shown that a promoter from ß-phaseolin contains negative elements repressing the expression in stem and root, and two different portions of the promoter that temporally regulate the expression in embryos (Burow *et al.* 1992, Sen *et al.* 1993). By studying DNA/protein interactions, it has also been proposed that the motif CACGTG is a major *cis*-acting regulatory element controlling spatial and temporal expression of a ß-phaseolin gene (Kawagoe and Murai 1992).

1.2.2. Seed-specific gene promoters in monocots.

Identification of seed-specific sequence elements in monocot storage protein genes is more recent (Table 1). Initial studies have been principally conducted using DNA material from the most important cereals (wheat, rice, maize and barley). The methods of investigation are similar, however, to those used for the dicot seed storage protein genes: transformation of tobacco or petunia plants with the intact genes or with promoter constructs using a reporter gene, and computer analysis of different promoter regions in order to define conserved region(s) that could play a role in the regulation of these genes.

Zein, a maize prolamin, is specifically expressed in the endosperm of maize seeds. When transferred into tobacco, a zein gene encoding a 23-kDa protein was shown to be specifically expressed in the tobacco seeds (supported by the presence of zein transcripts), although the protein itself could not be detected. When fused to a reporter gene (GUS), the endosperm-specifity of the zein promoter ($Z4_{pro}$) was then demonstrated, but was found to be very weak (Schernthaner *et al.* 1988). Both the 23-kDa zein and a 19-kDa zein were stably expressed in the endosperm when their coding regions were fused to the 35S CaMV promoter, confirming the weakness of the Z4 promoter and the high sensitivity of the GUS reporter gene (Schernthaner *et al.* 1988). The stability of zein proteins (19 kDa and 15 kDa) was also demonstrated in the embryo tissues of petunia and tobacco transgenic plants by using a ß-phaseolin gene promoter (Williamson *et al.* 1988, Hoffman *et al.* 1987). Deletion analysis of the zein promoter ($Z4_{pro}$) in transgenic tobacco plants showed that 174 bp upstream of the transcriptional start site were sufficient to direct a developmental and endosperm-specific expression of the GUS reporter gene (Matzke *et al.* 1990). However, in transgenic petunia plants, expression of the 19-kDa zein gene was reported not to be tissue-specific (Ueng *et al.* 1988). A consensus sequence in the 5' flanking region of zein genes

was identified and shown to bind a nuclear factor (Maier *et al.* 1987). Surprisingly, the -174 deletion of the $Z4_{pro}$ (see above) only contained part of this sequence and was still able to specifically regulate the expression of the GUS gene. However, a portion of this consensus sequence (TGTAAAG) is still present and is also found in another consensus sequence called '-300 element' or 'prolamin box' (Forde *et al.* 1985).

The promoters from a low molecular weight (LMW) and a high molecular weight (HMW) glutenin gene were tested in transgenic tobacco plants (Colot *et al.* 1987). Plants transformed with these promoters, shortened by a series of deletions from their 5' end and fused to the reporter gene CAT, showed that CAT activity was exclusively confined to the endosperm of transformed seeds. In addition, a 326 bp-sequence upstream of the transcriptional start point was sufficient to direct endosperm-specific CAT activity, but a 160 bp-promoter did not show any CAT activity in the transformed plants. The TGTAAAG core motif is present twice within the 326 bp of the LMW glutenin promoter and can also be found in the promoter region of the HMW glutenin gene, but outside of the region shown to be essential for tissue-specific regulation (see below). This HMW glutenin gene promoter was similarly shown to direct CAT activity only in the seeds of transformed tobacco plants. Promoter deletion analysis of the HMW glutenin gene revealed that 433 bp upstream of the transcription start point were sufficient to direct an endosperm-specific expression of the CAT coding sequence (Robert *et al.* 1989). Further experiments to localize *cis*-acting sequences were performed on the HMW glutenin promoter (Thomas and Flavell 1990). A 40-bp region located 170 bp upstream of the transcription start point was shown to be involved in the specific expression pattern of the HMW glutenin. Only a variant of the '-300 element' core motif was found in this region (TGCAAAG), but it was suggested that a direct repeat of the pentanucleotide sequence TTGCT and a pair of GCTCC direct

repeats could play a role in the expression of the HMW glutenin. Functional analyses of the 5' flanking regions of the Glu-1Dx5 HMW glutenin gene showed that 277 bp upstream of the transcription start point are required for endosperm-specific expression of GUS (Halford *et al.* 1989). Comparison with a silent homologous promoter indicated that silencing is due to mutations within the 280-bp region. Two '-300 element' core motifs are present within the 280-bp region but no further evidence is given for their possible functional role.

Similarly, a 549 bp 5' flanking region of the B1 hordein gene (including the prolamin box) contains all the necessary elements to confer tissue (endosperm) and developmental control to the CAT reporter gene in transgenic tobacco plants (Marris *et al.* 1988).

Expression of a chimeric gene composed of the promoter of a rice glutelin (Gt3, homologous to the 11S globulins) fused to the CAT reporter gene showed that developmental and endosperm-specific regulation was conferred by a 980-bp promoter region upstream of the initiation codon (Leisy *et al.* 1990). Promoter deletion analysis of another rice glutelin gene promoter showed that a region between positions -441 and -237 was required for temporal and endosperm-specific expression of the GUS reporter gene in transgenic tobacco seeds (Takaiwa *et al.* 1991). An AACA motif found in the 5' flanking sequences of six glutelin genes was not required for the correct expression of the GUS gene. The core motif of the prolamin box can be found in four different locations within repeated sequences, one of them remains in a deletion construct that did not show any GUS activity. The possibility that this motif serves as an enhancer-element rather than as a tissue-specific box is suggested (Takaiwa *et al.* 1991 and Shirsat *et al.* 1989).

In summary, previous studies have shown that:

1. The tissue-specific (embryo for dicot genes and endosperm for monocot genes) and developmentally controlled expression of seed storage protein genes are maintained in a different system (i.e. transgenic plants),

2. The promoters of these genes are responsible for this regulation,

3. Conserved portions of the promoter sequences are likely to play an important role in the regulation mechanisms,

4. These conserved regions, or boxes, vary among the different protein gene families,

5. The correct expression (at the mRNA level) of the coding region of a seed storage protein gene in an heterologous system does not ensure the stable accumulation of the protein in the transgenic seed and a correct processing of the protein precursor,

5. The interaction of nuclear DNA-binding protein(s) with portions of the promoter region is an integral part of the regulation of the expression of these genes, and

6. The possibility that a tissue-specific box (for the quality of expression) requires enhancer-elements (for the quantitative level of expression) exists and further demonstrates the complexity of the seed storage protein gene expression.

1.3. Objectives of the study.

1.3.1. Why study oat seed storage protein genes.

Oat globulin genes are expressed in a legumin-like manner (in terms of quantities) in the endosperm of oat grains. Oat globulins also share many structural features that are found in the dicot legumins and also in the rice glutelins. However similar the coding regions of the genes might be, it seems that, from preliminary studies, the oat

globulin gene promoters are very different from those of other storage protein genes, both in dicot and in monocot species. It is therefore interesting to study a promoter that does not seem to be related to any other promoter but which can still regulate the expression of its coding sequence in a highly specific manner.

Considering the different levels of gene expression regulation (transcriptional, posttranscriptional, translational and posttranslational), it is important not to focus exclusively on one type of storage protein gene. The choice of an avenin gene, in addition to the globulin genes, for promoter study is consequently one of the best ways for direct comparison between two different promoters that are similarly regulated.

Besides the need of a better understanding of the regulation of plant gene expression, a relevant model of study is also required. As a cereal plant that produces grains of high nutritional quality, oats, despite being a secondary crop, are a suitable system for the analyses of seed storage proteins, which constitute the most important source of proteins, in terms of quantity, in the human diet.

1.3.2. Tobacco as a model system.

At the beginning of this work, the use of transgenic plants for the study of foreign DNA was essentially restricted to dicot species susceptible to *Agrobacterium tumefaciens*. Tobacco, petunia, tomato and *Arabidopsis thaliana* were widely used in plant transformation experiments (Klee *et al.* 1987). The relative ease with which tobacco plants can be transformed and regenerated (Horsch *et al.* 1985) makes it a suitable system for the analyses of transferred genes. In addition, the production of a good number of transgenic plants (about 10) corresponding to individual transformation events is routinely achieved, and reduces the risks of artefactual results due to position effect or possible rearrangement of the inserted sequences.

Another important point for the use of tobacco as a working system is the presence of the endosperm tissue which is not absorbed by the embryonic tissues during the maturation of the seed and is still present in mature tobacco seeds. From the different studies that were available when this study was started, it appeared that tobacco seeds retained the ability to specifically express seed storage protein genes in the endosperm or in the embryo, depending on the original source of the sequences (see 1.2.1. and 1.2.2.).

1.3.3. Use of a reporter gene.

When studying the promoter activity of a gene, it is important to reduce to a minimum the possible translational and posttranslational events that can also affect the quantity and quality of the final product. This was particularly evident in the case of storage proteins which underwent partial proteolytic events in tobacco seeds (Ellis *et al.* 1988, Chee *et al.* 1986). The use of the coding region of the original gene necessitates a detection method (*e.g.* immunological) that is always different from one gene to the other and makes comparisons very difficult among independent studies. In addition, it requires the production of antibodies, which in some cases were unable to detect a weak expression (Schernthaner *et al.* 1988). The use of a reporter gene increases the sensitivity of detection and is therefore particularly attractive in promoter studies.

Among the different available reporter genes (mainly NPTII, CAT, luciferase and GUS) (Walden and Schell 1990), the GUS system was chosen (Jefferson 1987). The *E. coli* ß-glucuronidase is a hydrolase that catalyses the cleavage of a wide variety of ß-glucuronides. There is little or no detectable ß-glucuronidase activity in tobacco tissues. This enzyme is very stable and different substrates can be used. The detection methods are also multiple (spectrophotometric, visual and the most sensitive one, spectrofluorimetric). Assaying the presence of the GUS enzyme is easily carried out, does not require manipulation of

radioactive material and can be standardized for easier comparisons with other studies. For all these reasons, the GUS gene fusion system was chosen in this study.

1.3.4. Specific objectives.

As shown in section 1.1.3., the regulation of plant gene expression is complex and involves multiple mechanisms at different levels. A step by step approach is therefore essential to better understand the tissue-specificity and developmental regulations of seed storage protein genes. The three initial steps of such a study were carried out in this work:

1. Isolation and identification of genomic sequences corresponding to the major seed storage proteins in oat[2].

2. Characterization of the gene structure by sequence analysis. Since seed storage proteins belong to a multigene family, an initial study of these genes has to be performed on the DNA sequence itself to detect possible pseudogenes that are present. Coding sequences that possess mutation(s) introducing an in-frame stop codon have already been reported in oat seed storage protein genes (Shotwell et al. 1990). Although an in-frame stop codon does not preclude a non-functional promoter, this is also a possibility.

3. Assay for promoter activity and functionality using the intact gene and GUS fusions in transgenic tobacco plants. This analysis will allow the identification of functional promoter(s) on which preliminary experiments will be done for a more detailed analysis of the functional parts of the promoter. In addition, such a study will confirm, raise new questions or invalidate the unpublished report (Heim *et al.* 1991) of an oat globulin promoter directing an embryo-specific expression of the GUS coding sequence.

---

[2]The focus of this work was primarily put on the globulin sequences. The availability of an avenin promoter in the course of this study was made possible (Dr. M. Giband) and its functional analysis was therefore included in this work.

Chapter 2
Materials and methods

2.1. Plant material.

2.1.1 Oat seeds and plants.

Oat seeds, *Avena sativa* L. cv. Hinoat, Ogle and Donald were obtained from Dr. V. Burrows at the Plant Research Center of Agriculture Canada (Ottawa). Harvests of oat seeds at different stages of maturation, for RNA isolations, were conducted in the field at Agriculture Canada, Ottawa, in the early summer of 1989 and 1990. These seeds were frozen on site into liquid nitrogen and mechanically dehulled as described by Garson *et al.* (1983). They were kept in cotton bags at -70°C.

Oat seeds corresponding to different levels of ploidy were provided by Plant Gene Resources of Canada, Agriculture Canada (Ottawa). Their genomic constitution and accession numbers were as follows: *Avena magna* Murphy et Terrel (CAV4078, AACC), *Avena murphyi* Ladiz. (CAV2832, AACC), *Avena longiglumis* Dur. (CAV2215, A_LA_L), *Avena prostrata* Ladiz. (CAV5263, A_PA_P), *Avena wiestii* Steud. (CAV5185, A_WA_W), *Avena clauda* Dur. (CAV4908, C_PC_P), *Avena ventricosa* Bal. (CAV2208, C_VC_V), *Avena pilosa* M. Bieb. (CAV4910, C_PC_P).

2.1.2. Tobacco plants.

Tobacco plants, *Nicotiana tabacum* cv. Delgold (Pandeya and White, 1984), used for the transformation experiments were provided by Dr. R. Pandeya at the Plant Research Center, Agriculture Canada, Ottawa.

2.2. Chemicals and enzymes.

Chemicals used for the preparation of the different solutions were from Sigma, BDH, ICN, Fisher Scientific and Gibco BRL.

The restriction and modifying enzymes were from BRL, Pharmacia, NEN Dupont, Amersham, Promega and Boehringer Mannheim.

The radionucleotides (^{32}PdATP, ^{32}PdCTP and ^{35}SdATP) were from NEN Dupont and Amersham.

The antibiotics (ampicillin, kanamycin and carbenicillin) were from Sigma.

The oligonucleotide primers were synthesized by Dr. G. Alvarado from Synthaid Biotechnologies Inc.

2.3. Bacterial strains, growing conditions and transformation procedures.

2.3.1. *Escherichia coli* DH5α competent cells were from BRL (Bethesda Research Laboratories) and were transformed as suggested by the manufacturer. The plasmids used for transformation were mixed on ice with an aliquot of these competent cells and kept on ice for 20min. Following a heat shock (20s at 37°C), 800µl of LB (Luria-Bertani) medium (10g/l of bactotryptone, 5g/l of yeast extract, 10g/l of NaCl, pH 7.0 with NaOH) were added and the mixture was incubated at 37°C with occasional mixing for one hour. The cell suspension was centrifuged for 1min at low speed (4,000 rpm). The supernatant was discarded. The pelleted cells were resuspended in about 100µl of LB medium and the cells were spread on LB medium agar plates (100mmX15mm, Baxter) containing 100µg/ml of ampicillin (when using the pGEM-4 plasmid, Promega) or 50µg/ml of kanamycin (when using the binary transformation vector pBin19 plasmid, Bevan 1984). The plates were incubated at 37°C for 12 to 15hrs. Individual colonies were transferred for growth in LB medium with addition of the appropriate antibiotic. The blue/white selection was used by adding X-Gal and IPTG to the medium (the ß-

galactosidase gene in the plasmid is disrupted by the insertion of a sequence).

2.3.2. *Agrobacterium tumefaciens* GV3101/pmp90 was used for tobacco transformations. About one µg of plasmid DNA (pBin19) was added to a 100µl-aliquot of *A.tumefaciens* competent cells (procedure from Clontech) and the mixture was incubated on ice for 30min. The cells were then frozen with liquid nitrogen and subsequently incubated at 37°C for 2min. One ml of YEP medium (10g/l of yeast extract, 10g/l of peptone, 5g/l of NaCl, pH 7.0) was added and the mixture was incubated for two hours at 28°C with gentle shaking. The cells were pelleted by centrifugation and resuspended in about 100µl of YEP medium. They were then spread on a YEP agar plate containing 100µg/ml of kanamycin. The plate was incubated at 28°C. The transformed colonies appeared in about 2 to 3 days.

2.4. Genomic cloning.

2.4.1. Oat globulin genomic clones.

Genomic DNA from oat was restricted by EcoRI[3] and size-fractionated on a sucrose gradient. The fraction that gave a good hybridization signal when probed with a partial globulin cDNA clone (ILL2) was subcloned into λgt10 (as an EcoRI-EcoRI insert). The primary screening and the library were prepared by E. Hansen. A secondary and a tertiary screening were performed (with Dr. M. Giband), resulting in the isolation of two clones, pGlav-1 and pGlav-3 (Glav for Globulin from *Avena*). Another screening of the library by J. Byrne produced a third clone, pGlav-12. The methods used were as described by Maniatis (1982). The three clones were subcloned into pGEM-4. The pGlav-1 and pGlav-3 clones were analysed by restriction enzymes to determine their respective restriction map. The localization of the coding region was estimated by digesting the clones

[3]The general methodology for the molecular biology work was taken from Maniatis *et al.* 1982 and Sambrook *et al.* 1989.

by different enzymes, separating the fragments on an agarose gel, transferring the DNA onto a nitrocellulose membrane and using the partial cDNA clone as a probe.

2.4.2. Oat avenin genomic clones (isolated and sequenced by Dr. M. Giband).

Two clones were obtained using the PCR technique. Two primers were designed from the published sequence of an avenin genomic clone (Shotwell *et al.* 1990) and were used to amplify specific avenin sequences from oat genomic DNA. The two primers were AvPC5 (5' CTCGAATTCGGTTTATGCCTAACAGCCTG 3') situated at the 5'end and AvPC3 (5' CTCAAGCTTAAGGTCACAACTCACAAGCC 3') situated at the 3'end of the genomic clone. Two restriction sites (EcoRI in AvPC5 and HindIII in AvPC3) were artificially introduced into the primers allowing the forced subcloning of the PCR products into pGEM-7 (Promega).

2.5. Plasmid DNA minipreparation (minipreps).

The method of Birnboim (1979) was used. Overnight-grown cells (1.5ml out of a 3ml-culture) were centrifuged in a microfuge tube and the pelleted cells were resuspended in 80µl of freshly prepared resuspension buffer (25mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0 and 15% sucrose). The tube was put on ice and 10µl of 20mg/ml lysozyme in 20mM Tris-HCl pH 8.0 were added. The mixture was incubated on ice for 10min. Two hundred µl of freshly prepared 0.2M NaOH, 1% SDS were then added and mixed by inverting the tube. The tube was incubated on ice for 5min. One hundred µl of 3M Na acetate pH 4.6 was added and the tube was gently mixed. The tube was incubated on ice for another 5min and then centrifuged for 8min in a bench-top centrifuge at top speed. The supernatant was transferred into a clean tube and extracted with an equal volume of phenol/chloroform (v/v). Two volumes of 100% ethanol were added and the tube was placed at

-20°C to precipitate the DNA. The DNA pellet was resuspended in 100µl of TE (10mM Tris-HCl pH 7.5 and 1mM EDTA pH 8.0) containing 20µg/ml boiled RNase A. For sequencing purposes, the plasmid DNA was purified with another phenol/chloroform extraction and reprecipitated. Plasmid DNA minipreparations from *A. tumefaciens* were performed on 6mls of culture by the same method.

2.6. Oat seed RNA extraction.

Five grams of frozen oat seeds (at different stages of maturation) were put in a coffee grinder and ground into a fine powder. This powder was transferred into a 50-ml polypropylene tube and 10mls of extraction buffer were added (200mM Tris-HCl pH 8.5, 50mM $MgCl_2$, 20mM KCl, 10mM DTT and 2% Triton X-100) (Fabijanski *et al.* 1985). The mixture was further ground in a polytron with addition of 10mls of guanidium chloride solution (6M guanidium chloride, 0.1M sodium acetate pH 6 and 1% ß-mercaptoethanol). The tube was then spun for 10min at 5,000 rpm and a 5ml-aliquot of the supernatant was deposited on top of 6mls of a cesium chloride solution (5.7M CsCl, 0.1M NaCl, 10mM Tris-HCl pH 7.5 and 1mM EDTA pH 8.0). Centrifugation was performed for 16hrs at 15°C at 100,000xg (28,600 rpm in a SW41 rotor, in a Beckman L8-55M ultracentrifuge). After centrifugation, the supernatant was removed by aspiration and discarded and the tube was placed upside down to let the last drops of liquid drain from the pellet. The RNA pellet was resuspended in about 400µl of resuspension solution (0.1M Na acetate, 0.01% SDS and 5mM EDTA pH 8.0). Two phenol/chloroform (v/v) and one chloroform extractions were then performed before precipitating the RNA by the addition of two volumes of ethanol in the cold. After recovering the RNA by centrifugation, its integrity was checked on an agarose gel and its concentration determined by spectrophotometry at 260nm.

2.7. Tobacco RNA extractions (from seed, stem, leaf and root).

Tobacco seed-pods were immediately frozen into liquid nitrogen upon harvest. The seeds were scratched off the pods under liquid nitrogen and ground in a mortar and pestle until a very fine powder was obtained (with addition of glass powder). The seeds from two seed-pods generally yielded about 300 to 400µl of powder. This powder was transferred into a microfuge tube filled with 400µl of extraction buffer (100mM Tris-HCl pH 9.0, 100mM NaCl, 20mM CDTA, 1% SDS, 10mM ß-mercaptoethanol), 40µl of a 10% (w/v) insoluble polyvinyl-polypyrrolidone (PVPP) solution and 20µl of a 10% (w/v) ascorbic acid solution. The mixture was thoroughly vortexed and centrifuged for 5 minutes at top speed. The supernatant was kept and the pellet was re-extracted as described above. Two phenol/chloroform (v/v) and one chloroform extractions were performed on the supernatants. The nucleic acids were precipitated by addition of 1/20 vol. of 3M sodium acetate pH 5.0 and 2.5 vol. of 100% ethanol at -20°C for one hour. After centrifugation of the tubes for 15 minutes at 4°C, the pellets were usually coloured (green, yellow, brown or even black). At this stage, a RQ1 DNase digest was performed using two units of DNase per sample for half an hour at 37°C. The RNA was precipitated by ethanol after a phenol/chloroform extraction. The pellets were resuspended in 200µl of DEPC-treated water and 40µl of a 10% (w/v) activated charcoal solution were added. After mixing well and letting stand at room temperature for one minute, the microfuge tube was centrifuged at top speed for five minutes. This procedure was repeated until total disappearance of colour was achieved. A phenol/chloroform extraction was performed to remove any traces of charcoal, and the RNA was precipitated by ethanol. The pellet was resuspended in 100µl of DEPC-treated water and 5µl were checked on an agarose gel.

The RNA extractions from leaves and roots were performed with the same procedure except that the activated charcoal treatment was omitted.

2.8. Oat and tobacco total DNA extraction.

Young green leaves from tobacco or oat seedlings germinated in the dark at 20°C (1 to 2g) were harvested and immediately frozen in liquid nitrogen. The frozen tissue was ground in a mortar and pestle after addition of a little glass powder. During the grinding, liquid nitrogen was added if necessary to keep the tissue frozen until a fine powder was obtained. This powder was scooped into a 50ml polypropylene centrifuge tube containing 5mls of buffer S at room temperature (100mM Tris-HCl pH 8.5, 100mM NaCl, 50mM EDTA pH 8.0, 2% SDS. Immediately before use 0.1mg/ml Proteinase K and 10mM DTT were added). The tube was mixed well but gently and let stand at 37°C with occasional swirling. After one hour, 5mls of phenol were added (the phenol had been equilibrated in 100mM Tris-HCl pH 8.5, 100mM NaCl and 50mM EDTA) and the mixture was gently shaken on a rotary shaker for 30 minutes. A centrifugation of 10 minutes at 3,500 rpm was performed to separate the different phases. The aqueous phase (top phase) was transferred to a clean tube. If this aqueous phase was clear, 0.5ml of 3M Na acetate was added and mixed, then 10mls of 100% ethanol were gently poured in at room temperature. The tube was gently rocked until there was a small clump of nucleic acids which sank to the bottom. The ethanol was poured off (if necessary, a short centrifugation was performed to make the nucleic acids stick to the bottom). The pellet was gently resuspended in 1ml of TE containing 100µg of boiled RNase A, and incubated at 37°C for 1 to 2 hours with occasional agitation to assist the resuspension. A phenol/chloroform extraction (phenol equilibrated in TE) was performed. The aqueous phase was recovered and the DNA was precipitated by addition of 1/10 vol. of 3M sodium

acetate and 2 vol. of ethanol, at -20°C for at least 2 hours. After centrifugation, the DNA pellet was resuspended in 500µl of dH_2O (the DNA was not allowed to dry out). It was sometimes necessary to incubate at 37°C to ensure a complete resuspension. The DNA concentration and integrity were finally visually estimated by loading 2µl on a 0.5% agarose gel along with known amounts of DNA (lambda DNA restricted by HindIII).

2.9. Restriction digests of total DNA from oat and tobacco.

Total plant DNA (5.5µg) was restricted in a final volume of 400µl using the buffer provided by the enzyme supplier. Addition of casein (Flash Kit, Stratagene) at a concentration of 100µg/ml greatly helped the digestion (Dreyer and Schulte-Holthausen 1991). The amount of enzymes added to the mixture was 30 units, and the reaction was allowed to proceed overnight at the required temperature. Ten more units of enzyme were added in the morning. After 2 to 3 hours of reaction, 20µl were run on an 0.8% agarose gel to verify the complete digestion of the DNA. The DNA was precipitated at -20°C overnight after addition of 1/20 vol. of 3M sodium acetate and 2 vol. of ethanol. After centrifugation of the tubes at 12,000 rpm (4°C), the DNA pellets were dried (Speed Vac, Savant), resuspended in 36µl of dH_2O and 4µl of 10X loading buffer (see 2.10.1. for composition).

2.10. DNA and RNA agarose gels.

2.10.1. DNA gels.

Agarose gels for DNA separations were run in TAE buffer (50X TAE buffer was prepared by mixing 242g of Tris base, 57.1mls of glacial acetic acid and 100mls of 0.5M EDTA pH 8.0). The 10X loading buffer was 40% sucrose in 10X TAE with bromophenol blue as a migration marker. The agarose gels were submerged in running buffer and were usually run at about 5 to 10V/cm (constant voltage). The concentration of agarose varied from 0.5% to 1.5% depending on the

size of the DNA to be separated. At the end of the run, the gel was immersed in an ethidium bromide solution (0.5µg/ml) for 15min and the DNA was viewed under UV light. When necessary, a photograph was taken with a Polaroid apparatus.

2.10.2. RNA gels.

The separation of RNAs was performed in 1.2% agarose gels. The 10X running buffer was 500mM boric acid, 50mM disodium tetraborate, 100mM sodium sulfate and 10mM EDTA. The working solution was adjusted to pH 8.3. A 4X loading buffer was prepared as follows: 800µl of 10X running buffer, 400µl of glycerol, 720µl of dH_2O, 80µl of methylmercuric hydroxide, 0.4% (w/v) of bromophenol blue and 0.4% (w/v) of xylene cyanol. Before being loaded into the wells, the samples were heated at 65°C for 5 to 10min. A constant voltage of approximately 5V/cm was applied. Examination of the RNAs was performed as described for the DNA gels except that a longer time of ethidium bromide staining was required (30min).

2.11. Plasmid constructs.

2.11.1. pGEM-4-GUS (Figure 1)[4].

The GUS coding sequence and the NOS terminator sequence were taken from pBI101.1 (Clontech) as a EcoRI-XbaI fragment and subcloned into pGEM-4.

2.11.2. Av-GUS (Figure 2).

The avenin promoter was amplified by PCR from a genomic clone inserted into pGEM-7, using the T7 primer and the Av-1 primer. A BamHI restriction site (bold in the sequence) was introduced at the 5' end of the Av-1 primer (5' CGGGATCCCATGGTGGATTTGTGTTGAC 3'). The resulting PCR product was digested by BamHI and SphI and inserted into pGEM-4-GUS (also cut by BamHI and SphI). The whole construct (avenin promoter-GUS-NOSter) was then taken out of pGEM-4 as an EcoRI-EcoRI fragment and inserted into pBin19 (cut by

[4]The drawings are not necessarily to scale when no scale is indicated in the figures.

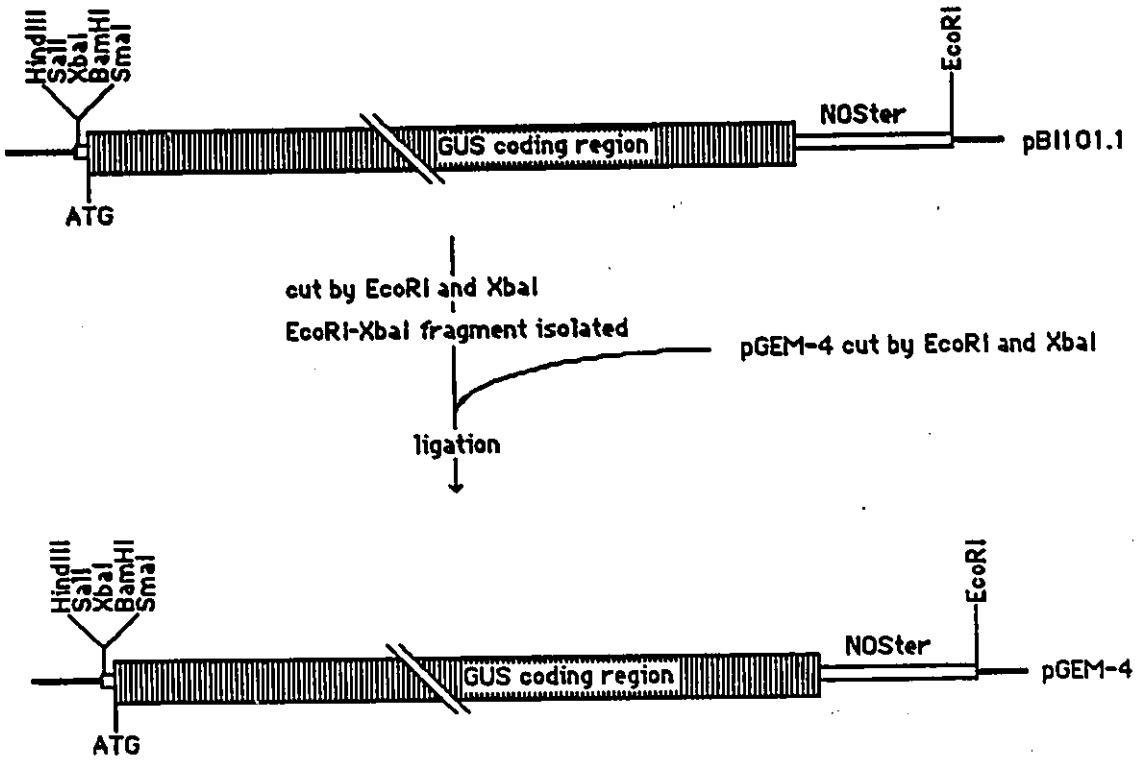


Figure 1 : pGEM-4-GUS.

Insertion of the GUS coding region and the NOSter into pGEM-4. The GUS and NOSter sequences were taken from pBI101.1 as an EcoRI-XbaI fragment and inserted in the EcoRI and XbaI sites of pGEM-4.

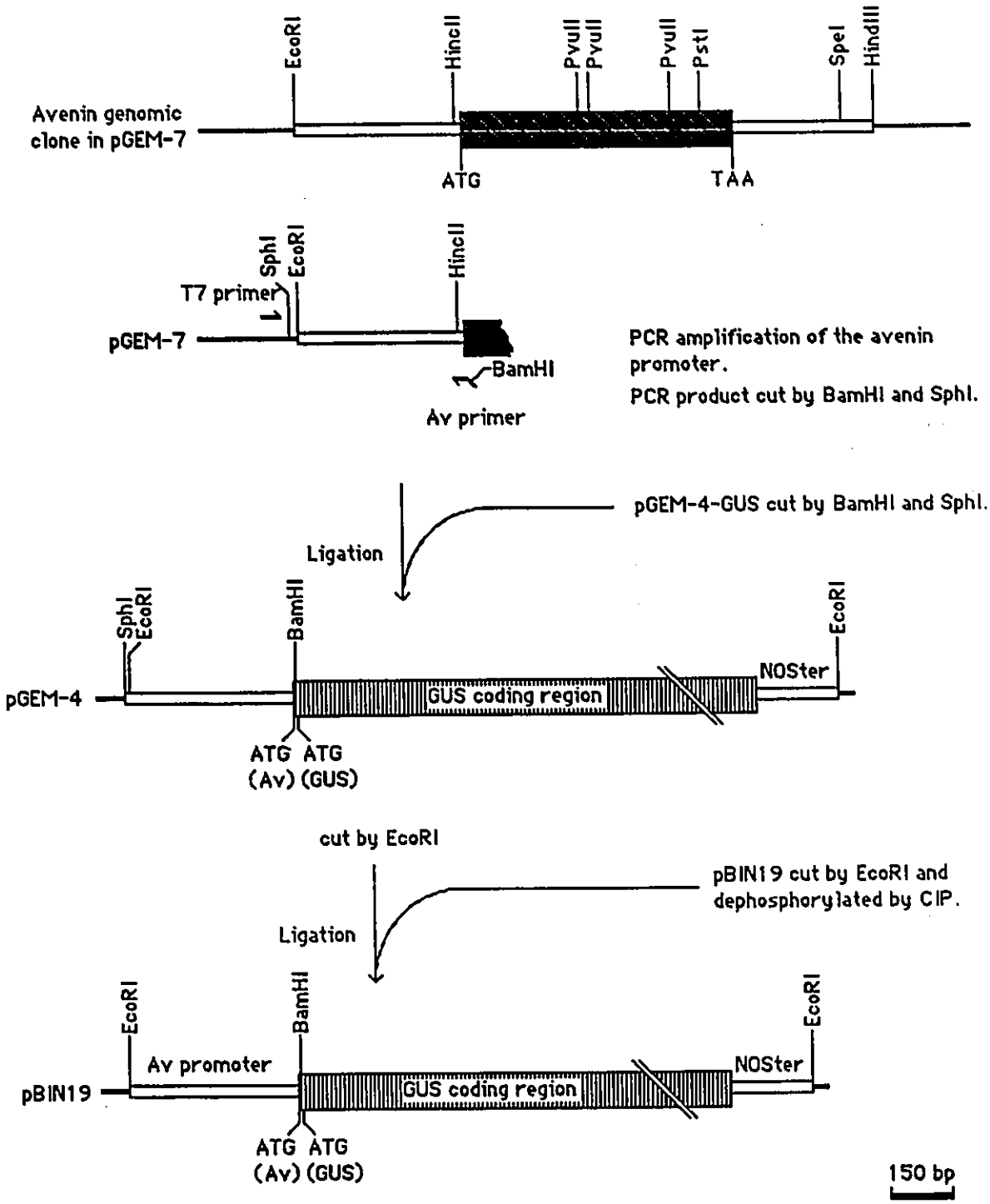


Figure 2: Construction of the Av-GUS plasmid.
The avenin promoter was PCR-amplified and inserted in front of the GUS coding sequence using restriction sites SphI and BamHI. The whole construct was then excised from pGEM-4 as an EcoRI fragment and cloned in the EcoRI site of pBin19.

EcoRI and dephosphorylated by CIP).

2.11.3. Globulin-GUS (Figure 3).

A deletion fragment obtained for the sequencing of the Glav-1 promoter was cut by HindIII and ligated to itself. The resulting plasmid was cut by SpeI and EcoRI. In between these sites was inserted the GUS and NOSter sequence obtained as a XbaI-EcoRI fragment from pGEM-4-GUS (see 2.11.1. and Figure 1). SpeI and XbaI have compatible ends that enabled the ligation. In the unique HindIII site that remained (Figure 4) was inserted the Glav-12 promoter (gel-purified as a 2.5kb HindIII-HindIII fragment from the genomic clone). The whole construct (Glav-12 promoter-GUS-NOSter) was taken out of pGEM-4 as an EcoRI-EcoRI fragment and inserted into pBin19 (cut by EcoRI and dephosphorylated by CIP).

2.11.4. G1-GUS, Δ-2364-GUS and Δ-1420-GUS.

The entire promoter of the Glav-1 genomic clone and two deletions of this promoter (Δ-2364 and Δ-1420) were also fused to the GUS sequence directly into pBI101.1. To insert the entire Glav-1 promoter in front of GUS (Figure 5), the plasmid G1SRL (see 2.15.) was first cut by EcoRI, blunt-ended with Klenow and then cut by SstI. The resulting fragment was subcloned into pGEM-4 cut by XbaI, blunt-ended with Klenow and cut by SstI. This subcloning allowed the fusion of the complete Glav-1 promoter as a SalI-BamHI fragment into the pBI101.1 cut by SalI and BamHI. The two deletion fragments were taken from the clones obtained for sequence determination of the Glav-1 promoter. They were inserted as BamHI-BamHI fragments into the BamHI site of pBI101.1.

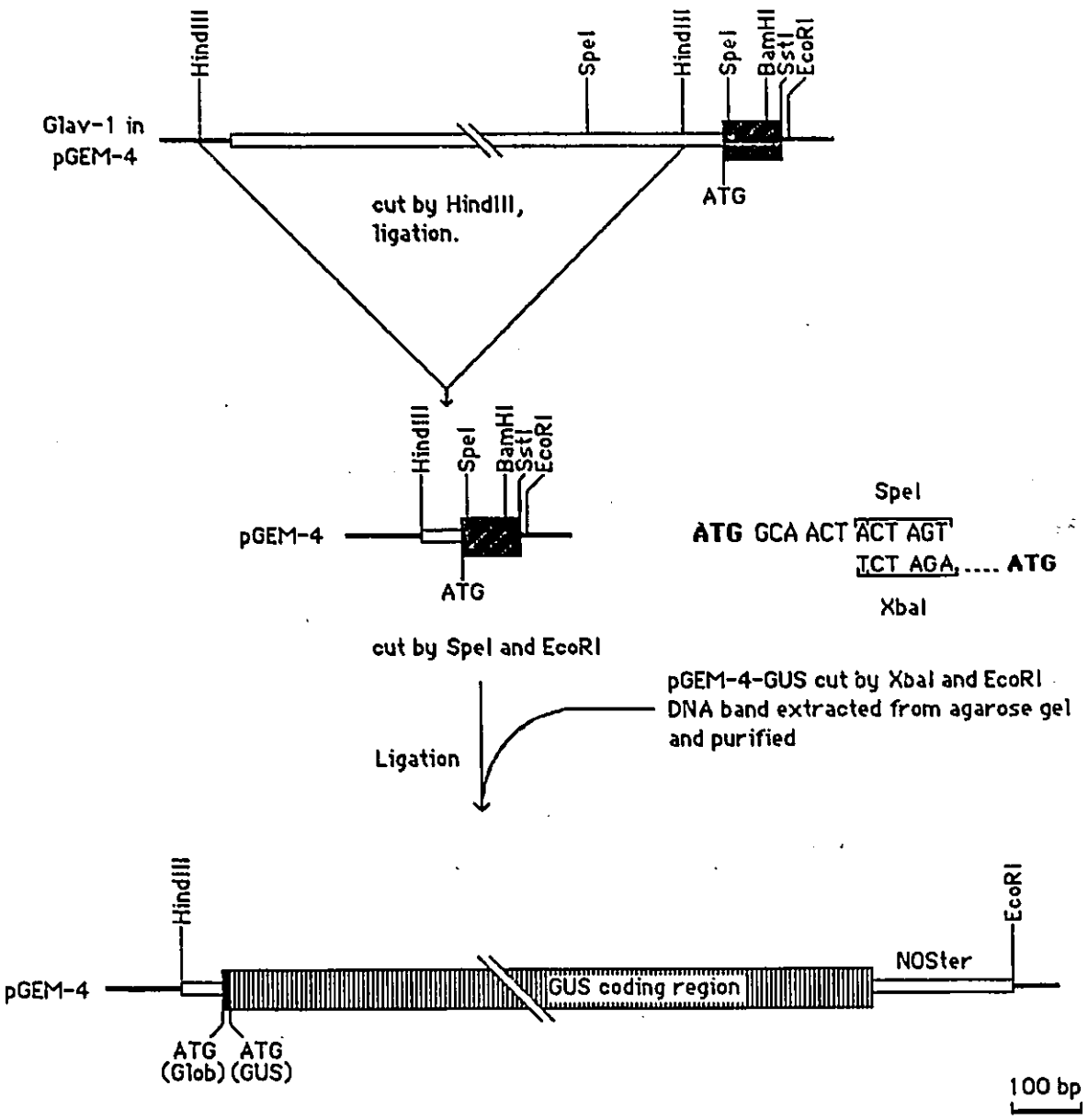


Figure 3 : Translational fusion of the beginning of the globulin promoter to the GUS coding sequence.
The promoter sequence of pGlav-1 upstream of the HindIII site was removed by cutting the plasmid by HindIII and performing a religation. The resulting plasmid was cut by SpeI and EcoRI. In these sites was inserted the GUS coding sequence and the NOSter as a XbaI-EcoRI fragment.

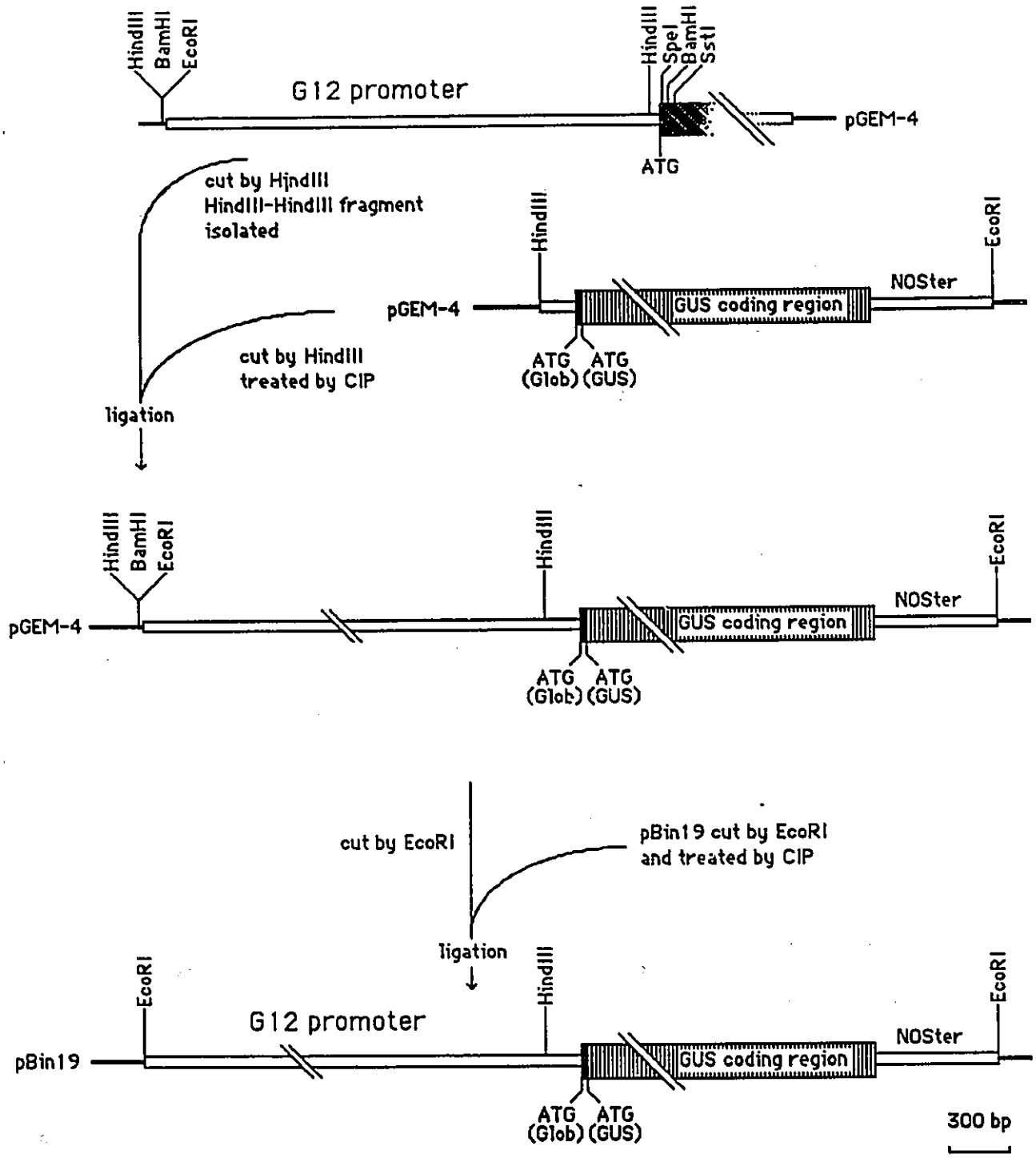


Figure 4 : Construction of the G12-GUS plasmid.
The entire promoter of Glav-12 was excised as a HindIII fragment and cloned in the HindIII site of the GUS translational fusion construct obtained in figure 3. The whole construct was then transferred in pBin19 as a EcoRI insert.

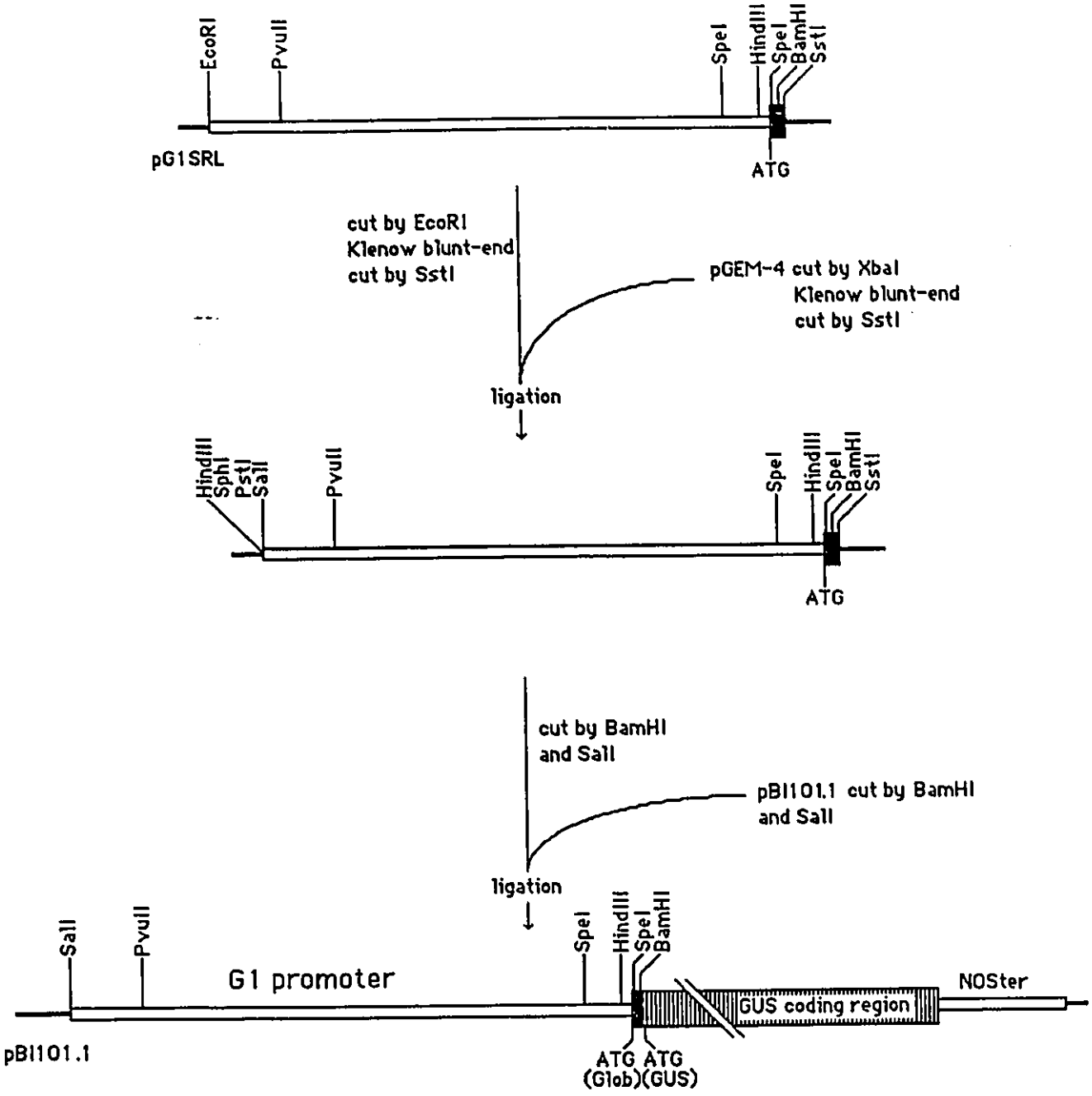


Figure 5 : Construction of the G1-GUS plasmid.
The plasmid G1SRL was resubcloned as an EcoRI-blunt end insert in pGEM-4 in order to have more suitable restriction sites available. The Glav-1 promoter was then translationally fused to the GUS sequence in pBI101.1 by using the BamHI and SalI restriction sites.

2.11.5. Precise deletions of the Glav-3 promoter fused to GUS (Figure 6, 2 pages).

Comparison of the 5' upstream sequences of the oat globulin genomic clones (Glav-1, Glav-3 and Glav-12) has revealed the presence of a duplicated element in Glav-1 that is only present once in Glav-3 and Glav-12. To assess the potential role of this sequence, a deletion analysis was performed which necessitated obtaining precisely located deletions (see section 3.1.5.).

A BamHI-BamHI fragment of pGlav-3 was subcloned into pGEM-4. This fragment included part of the promoter, the translational start codon and a portion of the coding region. Another portion of the promoter was removed by cutting the plasmid by XhoI and XbaI, creating blunt ends with Klenow and ligating the plasmid on itself. The resulting plasmid was cut by SphI and SalI. An Exonuclease III digestion was performed (6 time points at 20s intervals) and the DNA was treated by the nuclease S1. After having blunt-ended the plasmids with Klenow, the latter were ligated on themselves. The screening of the deletions was first done by size on agarose gel (HindIII-HindIII fragment) and then by sequence analysis using the primer G-1 (5' CTAGTAGTTGCCATG 3'). Three deletions were inserted in front of GUS in the same way as was the Glav-12 promoter. These three constructs were transferred into pBin19 (cut by EcoRI and SmaI) as PvuII-EcoRI fragments.

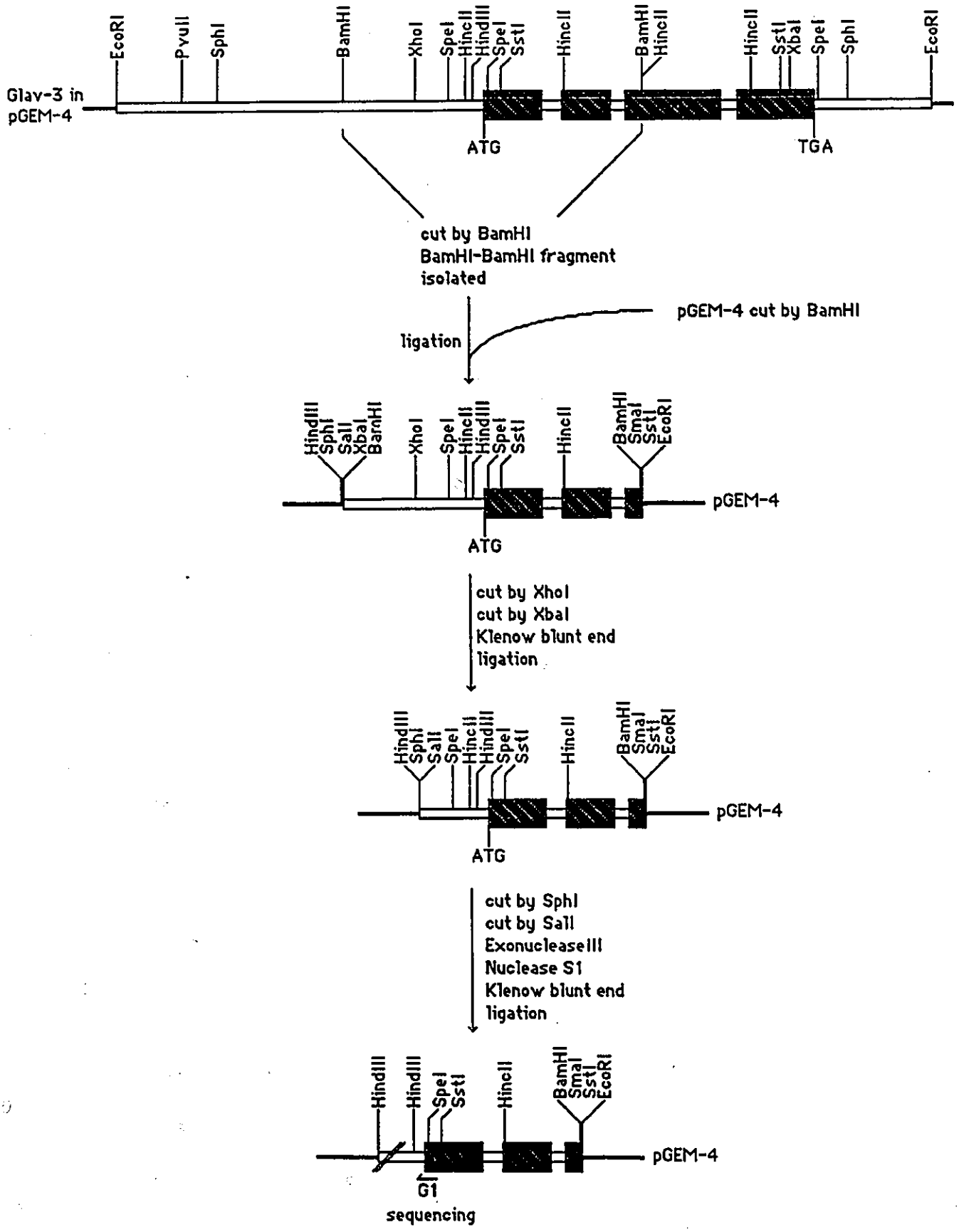
2.11.6. Other constructs.

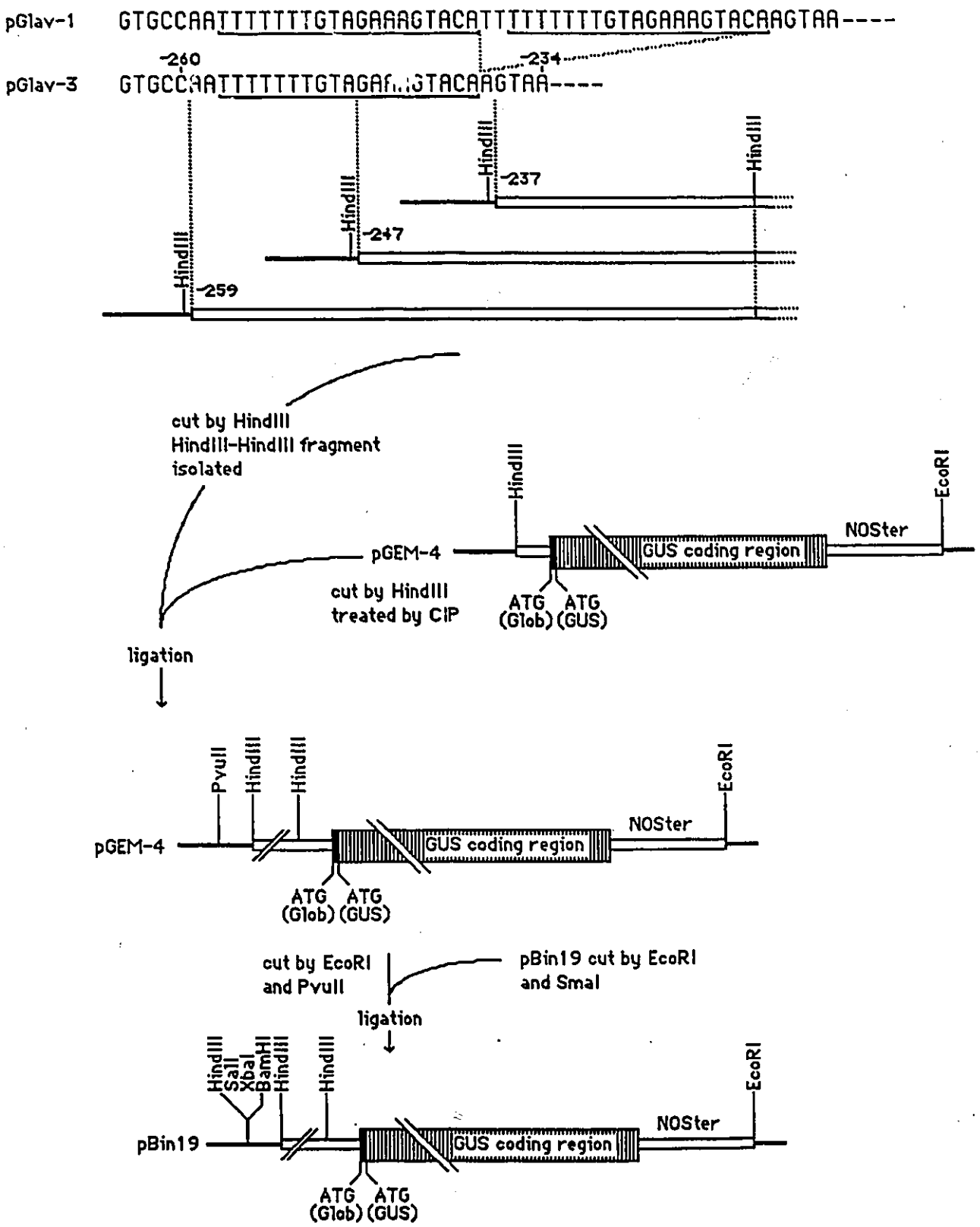
In addition, the complete genomic clones (Glav-1 and Glav-3) were cloned into pBin19 as EcoRI-EcoRI inserts. Glav-3 with a shorter promoter was also cloned into pBin19 (EcoRI-SmaI) by digesting pGlav-3 with PvuII and EcoRI (Figure 7).

A summary of all the constructs (T-DNA) is presented in Figure 8.

Figure 6 (2 pages): Production of precise deletions of the Glav-3 promoter and fusion to the GUS coding sequence.

A BamHI fragment of pGlav-3 was subcloned in pGEM-4. A portion of the promoter was removed (cut by XhoI and XbaI, then blunt-ended) and the plasmid was ligated on itself. The promoter was sequentially deleted by exonuclease III. Three deletions were selected after sequencing using the primer G-1. The corresponding sequence of the Glav-1 promoter is shown for comparison. These three deleted promoters, starting at position -259, -247 and -237 from the transcriptional start point, were inserted as a HindIII fragment in front of the translational GUS fusion construct (see Figure 3). These constructs were extracted from pGEM-4 as EcoRI-PvuII fragments and inserted in the SmaI and EcoRI sites of pBin19.


Glav-3 in pGEM-4
EcoRI
PvuII
SphI
BamHI
XhoI
SpeI
HincII
HindIII
SpeI
SstI
HincII
BamHI
HincII
HincII
SstI
XbaI
SpeI
SphI
EcoRI
ATG
TGA
cut by BamHI
BamHI-BamHI fragment
isolated
ligation
pGEM-4 cut by BamHI
HindIII
SphI
SalI
XbaI
BamHI
XhoI
SpeI
HincII
HindIII
SpeI
SstI
HincII
BamHI
SmaI
SstI
EcoRI
pGEM-4
ATG
cut by XhoI
cut by XbaI
Klenow blunt end
ligation
HindIII
SphI
SalI
SpeI
HincII
HindIII
SpeI
SstI
HincII
BamHI
SmaI
SstI
EcoRI
pGEM-4
ATG
cut by SphI
cut by SalI
ExonucleaseIII
Nuclease S1
Klenow blunt end
ligation
HindIII
HindIII
SpeI
SstI
HincII
BamHI
SmaI
SstI
EcoRI
pGEM-4
G1
sequencing


pGlav-1
GTGCCAATTTTTTTGTAGAAAGTACATTTTTTTTTGTAGAAAGTACAAGTAA----
pGlav-3
GTGCCAATTTTTTTGTAGAAAGTACAAGTAA----
-260
-234
HindIII
-237
-247
-259
cut by HindIII
HindIII-HindIII fragment
isolated
pGEM-4
cut by HindIII
treated by CIP
GUS coding region
NOSter
EcoRI
ATG (Glob)
ATG (GUS)
ligation
PvuII
pGEM-4
cut by EcoRI
and PvuII
pBin19 cut by EcoRI
and SmaI
SalI
XbaI
BamHI
pBin19

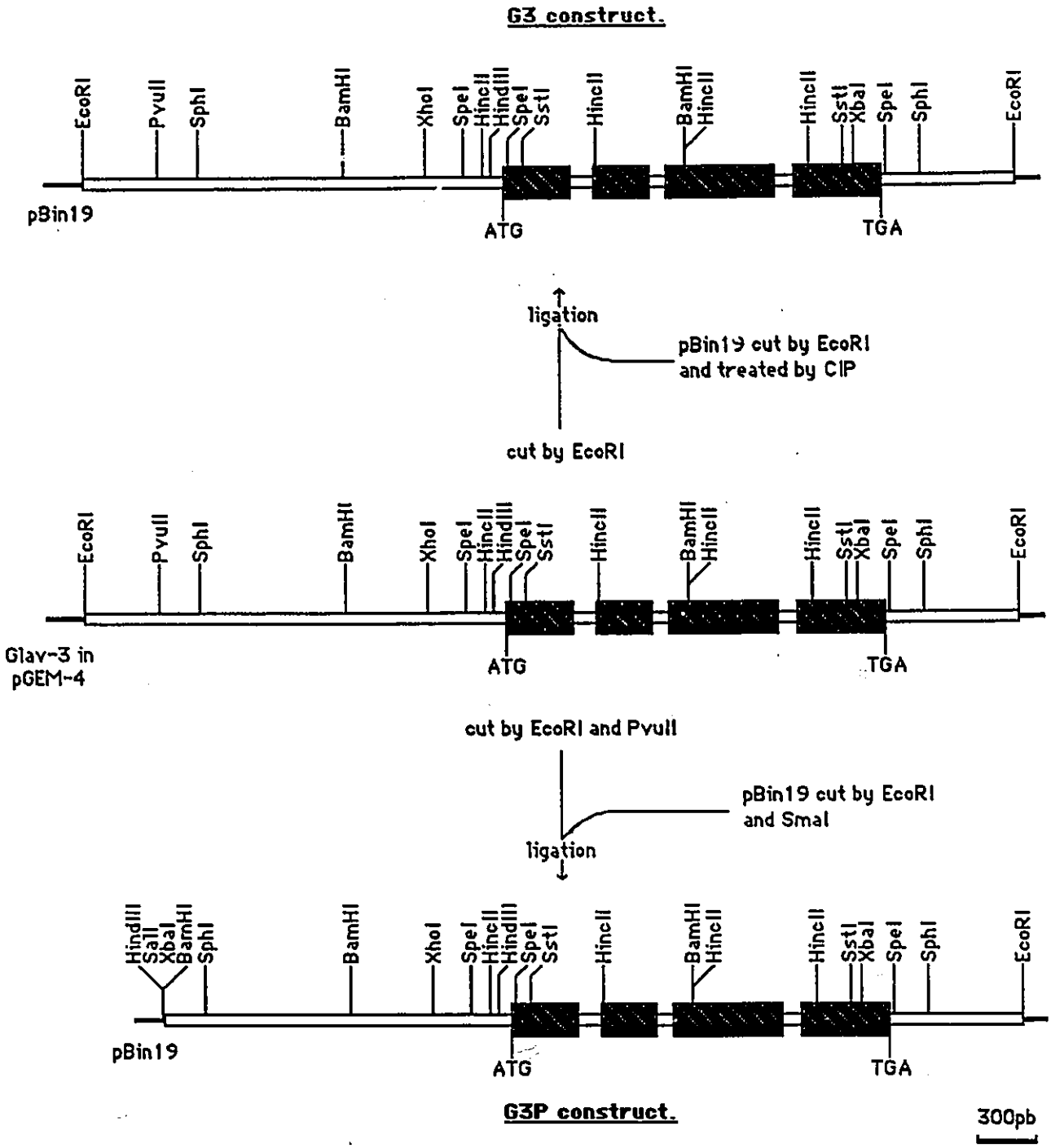


Figure 7 : Constructions of the G3 and G3P plasmids.
The entire Glav-3 sequence was transferred in the Bin19 plasmid as an EcoRI insert (G3 construct). The promoter of Glav-3 was shortened by a PvuII cut and the PvuII-EcoRI fragment inserted in the EcoRI and SmaI sites of the Bin19 plasmid (G3P construct).

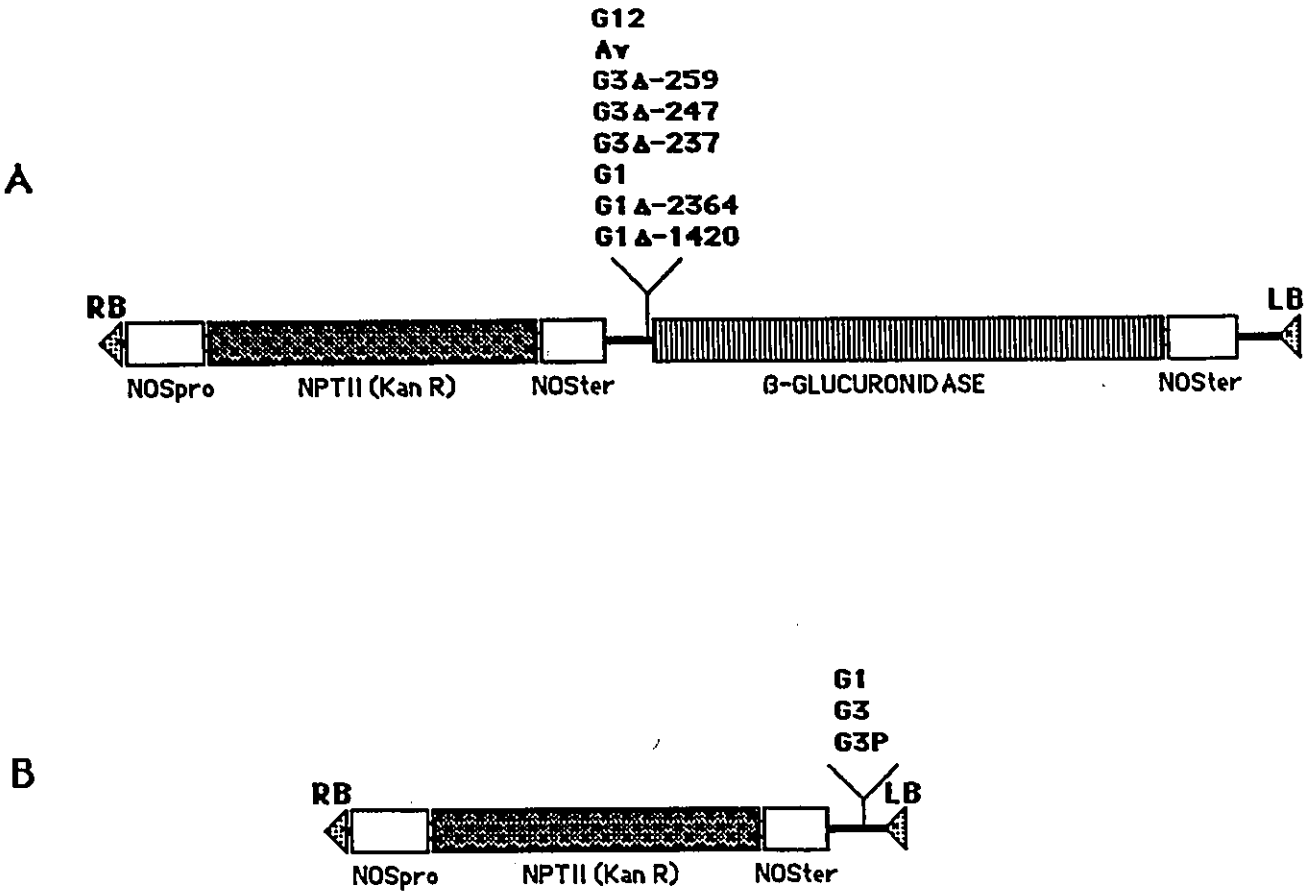


Figure 8: Summary of the constructs used for tobacco transformations presented as T-DNA inserts.
A: GUS fusions
G12, Av, G3Δ-259, G3Δ-247 and G3Δ-237 are in the Bin19 plasmid. G1, G1Δ-2364 and G1Δ-1420 are in pBI101.1.
B: Globulin sequences (promoter, coding region and terminator) in pBin19.
RB: Right border LB: Left border.

2.12. Polymerase chain reaction (PCR) and reverse transcriptase-PCR (RT-PCR).

2.12.1. PCR reaction.

The PCR reaction was carried out in a total volume of 100µl which included the buffer (50mM KCl, 20mM Tris-HCl pH 8.4, 2.5mM MgCl2), the dNTPs (2µl of a 2mM solution of each dNTP), the two primers (20ng each), the DNA (approximately 20ng) and one unit of Taq DNA polymerase (BRL or Biolabs) (Innis *et al.* 1990). The reaction mixture was placed in a 0.5ml microfuge tube and covered with a drop (100µl) of light mineral oil (Sigma). Thirty five cycles (45s at 95°C, 45s at 55°C, 45s at 72°C) were performed in a Microcycler (Eppendorf) after an initial denaturation at 95°C for 3min. At the end of the program, an extension step at 72°C for 3min was performed. Ten µl of the reaction were analysed on a 1.5% agarose gel.

2.12.2. RT-PCR.

The reverse transcriptase reaction was performed in a total volume of 20µl. The total RNA (5µg) was mixed with the reaction buffer (the same as for the PCR reaction), 1µl of dNTPs, 10ng of the primer, 1µl of RNase inhibitor (RNasin, Promega) and 200 units of Moloney Murine Reverse Transcriptase (BRL). The reaction was carried out at 42°C for 1 hour. The amplification of the cDNAs was realized as described in section 2.12.1. with the PCR method.

2.13. Plasmid sequencing.

The sequencing was done using the dideoxy chain termination method (Sanger *et al.* 1977) following the manufacturer's instructions (T7 sequencing kit, Pharmacia). About 2µg of plasmid DNA were denatured by boiling the sample for 3min. The primers were annealed to the DNA template in a buffered solution containing DTT and $MgCl_2$ (provided by the kit supplier) at 37°C for 20min. The tube was then placed at room temperature for 10min. The labelling reaction was

carried out by adding dCTP, dGTP, dTTP, the T7 DNA polymerase (3 units) and the ^{35}SdATP. The reaction mixture was kept at room temperature for five minutes, then aliquots were dispensed into four tubes each containing the four dNTPs plus one different ddNTP. The termination reaction was allowed to proceed for 5min at 37°C and was stopped by addition of an acrylamide gel loading solution (deionised formamide with EDTA, xylene cyanol 0.2% and bromophenol blue 0.2%).

2.14. Polyacrylamide gel electrophoresis under denaturing conditions for sequence analysis.

The glass plates were thoroughly washed with soap, rinsed with distilled water and wiped with ethanol. The inside of the short plate was Silane-treated. Uniform spacers were inserted and the plates were clamped together. The bottom of the gel assembly was taped. The sample application comb was inserted upside-down at the top of the plates. The 6% acrylamide solution was carefully poured in between the two plates. The gel was allowed to polymerize for at least one hour. The composition of the gel was 5.7% acrylamide, 0.3% bis-acrylamide, 46% urea in TBE (45mM Tris base, 45mM boric acid and 2mM EDTA pH 8.0). Just before pouring the gel, 100µl of a 10% ammonium persulfate solution and 70µl of TEMED were added to 100mls of gel solution. After complete polymerization of the gel, the tape and the comb were removed. The top of the assembly was washed with distilled water and the comb was inserted with the teeth slightly entering the gel. The gel was placed in the electrophoresis apparatus (Biorad), the running buffer (TBE) was added and the gel was pre-run for half an hour. The samples were denatured by boiling them for 3min in a water-bath and loaded into the wells. The gel was run for 2.5hrs (short run) or 5hrs (long run) at a constant power of 40W. After completion of the electrophoresis, the gel was removed from the assembly, transferred to a supporting sheet of filter paper,

covered with a plastic wrap and dried in a gel dryer (connected to a vacuum trap) for 1h at 80°C. The plastic wrap was then removed and the gel exposed overnight to an X-ray film.

2.15. Sequencing of pGlav-1.

The sequencing of pGlav-1 was done using two subclones of the original clone. The first subclone, pG1SRL (**pG**lav-**1** **S**stI **R**e**L**igation), was obtained by cutting pGlav-1 with SstI and then doing a ligation of the plasmid to itself (Figure 9). The internal SstI-SstI fragment was subcloned into pGEM-4 cut open with SstI and dephosphorylated using C.I.P (Figure 10). This second subclone was called pG1SF (**pG**lav-**1** **S**stI **F**ragment).

The plasmid pG1SRL was cut by EcoRI, a phenol/chloroform extraction was performed and the DNA was precipitated by ethanol (Figure 9). The DNA was resuspended in dH_2O and an equal volume of 2X exonuclease III buffer was added (100mM Tris-HCl pH 8, 10mM $MgCl_2$ and 2mM DTT). An excess of exonuclease III (Henikoff 1984) was then added (200 to 300 units) and the reaction mixture placed in a water-bath set at 37°C. Aliquots of 5µl were removed every minute and immediately transferred into a tube containing the same volume of TE 10:10 (10mM Tris-HCl pH 7.5 and 10mM EDTA pH 8) and placed on ice. Twelve time points were taken. The exonuclease III was heat-inactivated for 10min at 70°C. A volume of 200µl of nuclease S1 solution was added in each tube (30mM Na acetate pH 4.8, 60mM NaCl, 1mM $ZnSO_4$, 5% glycerol and 50 U/ml of nuclease S1). The reaction was allowed to proceed for 30min at 37°C. A phenol/chloroform (v/v) extraction was performed, followed by the precipitation of the DNA by ethanol. The DNA was blunt-ended by the Klenow DNA polymerase (large fragment) for 10min at room temperature (the reaction buffer was 25mM Tris-HCl pH 8 and 10mM $MgCl_2$). The DNA was restricted by PstI and SstI. The deletion fragments (5' blunt-ended and 3' SstI) were ligated into pGEM-4 cut

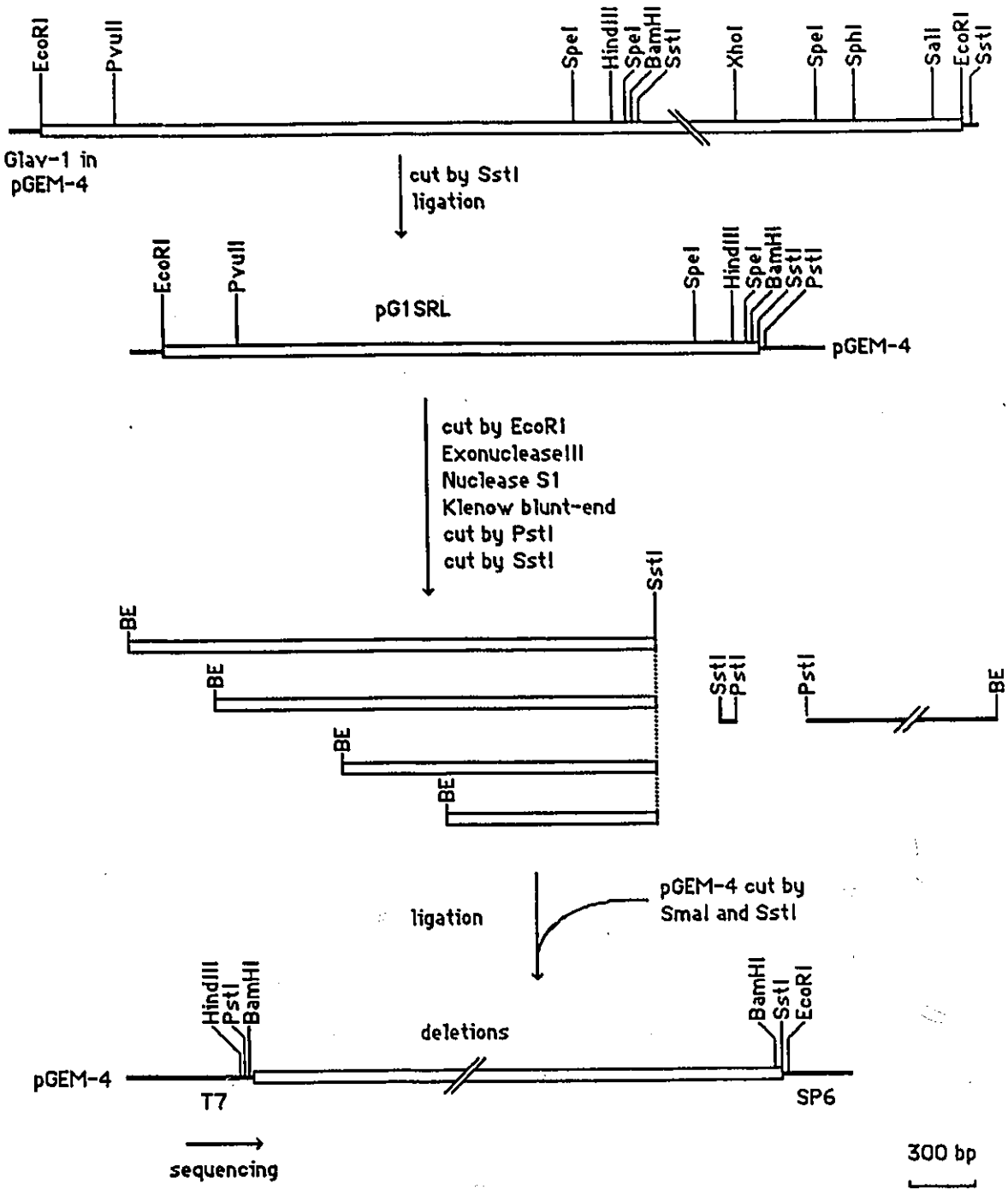


Figure 9: Sequencing of pG1SRL.
The plasmid G1SRL was obtained after a SstI cut of the Glav-1 sequence followed by a ligation. An exonuclease treatment resulted in the obtaining of progressively deleted fragments which were then subcloned in pGEM-4 and sequenced using the T7 primer.

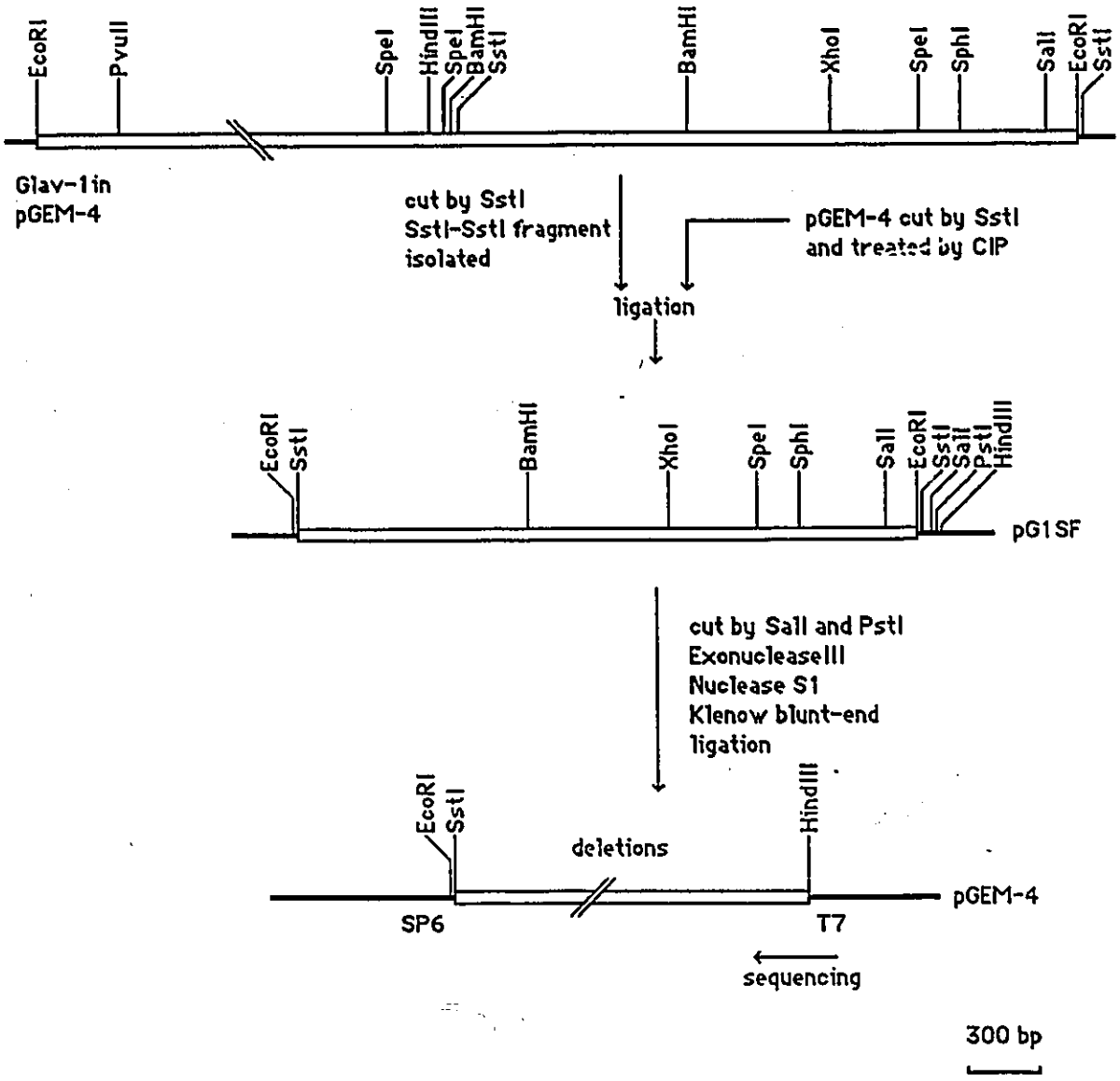


Figure 10 : Sequencing of pG1SF.
The plasmid G1SF was obtained by subcloning the SstI fragment of pGlav-1 in the SstI site of pGEM-4. This fragment was subjected to an exonuclease III treatment resulting in the obtaining of clones of various lengths. These clones were then sequenced using the T7 primer.

by SmaI (blunt-end) and SstI. One third of the ligation mixture was used to transform *E. coli* DH5α competent cells. Plasmid DNA minipreparations were performed on colonies grown overnight in 3mls of LB containing 50µg/ml of ampicillin. The size of the inserted fragments was determined by agarose gel electrophoresis after digestion of the plasmids by BamHI and 8 plasmids were selected for sequencing. Verification that the different plasmids belonged to the Glav-1 sequence was done by a short sequencing of the plasmids using the SP6 primer. Each subclone was sequenced at least twice using the T7 primer.

The plasmid pG1SF (Figure 10) was double cut by SalI and PstI. Deletions were obtained by using exonuclease III followed by a nuclease S1 treatment (as described above). The DNA was blunt-ended by Klenow and the plasmids religated on themselves. The deletions were screened by digesting the plasmids with HindIII and SstI, and the selection of the deletions was based on the size of the fragment. The plasmid sequencing was performed using the T7 primer.

2.16. Sequencing of the Glav-12 promoter (Figure 11).

The Glav-12 promoter and a short portion of the beginning of the coding region were cloned as an EcoRI-SstI fragment into pGEM-4 (pG4-G12ES). The plasmid was cut by PstI and SmaI. An exonuclease III digestion was performed as described previously (10 time-points at 60s intervals) followed by a nuclease S1 treatment. The resulting plasmids were blunt-ended by Klenow and ligated to themselves. The EcoRI-HindIII fragments were screened for the correct size. Twelve deletions were selected for sequence determination which was done using the T7 primer. In addition, a subclone was obtained by cutting pG4-G12ES with HincII and ligating the resulting plasmid on itself. The T7 primer was used for the sequence determination.

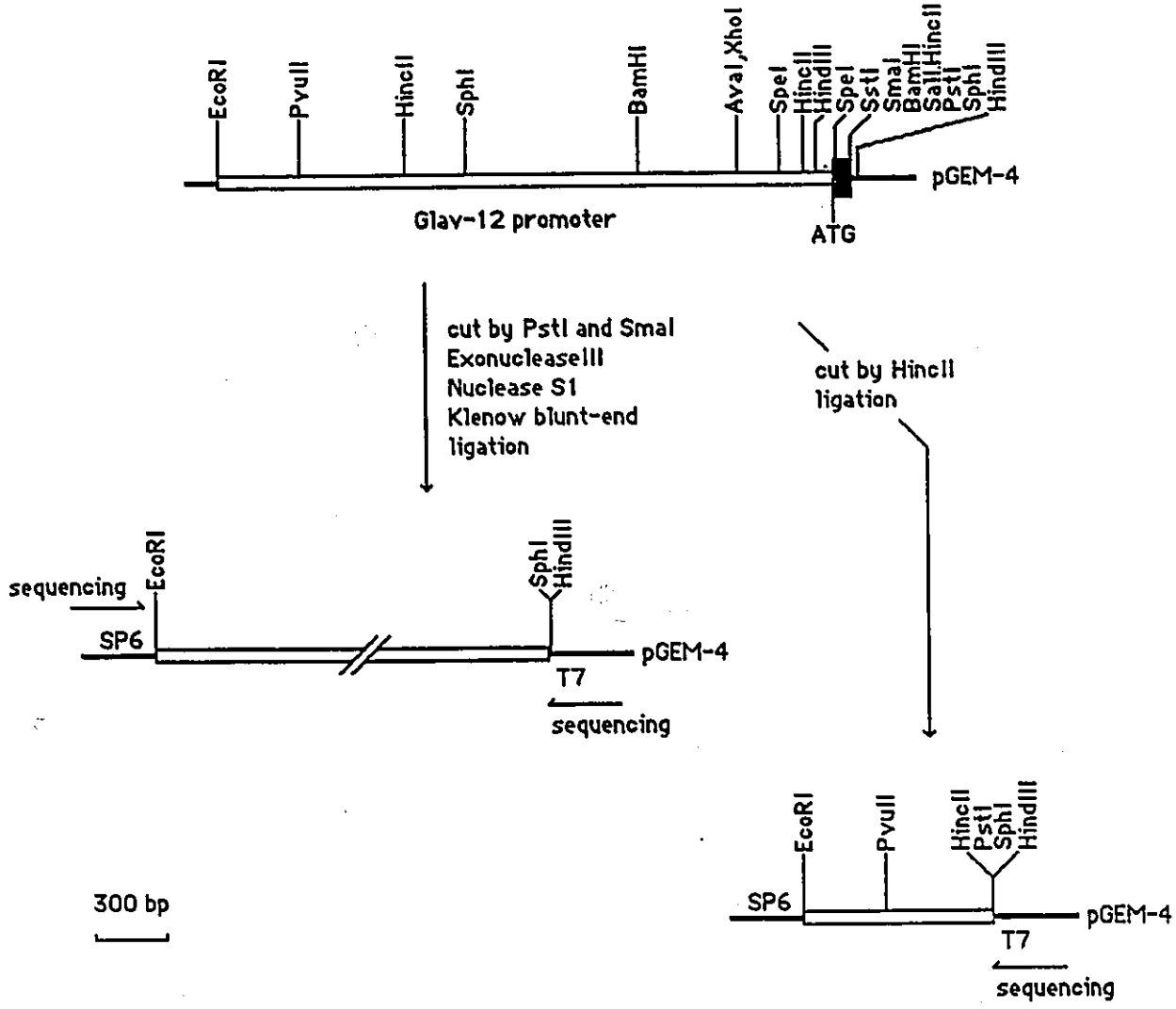


Figure 11 : Sequencing of the Glav-12 promoter.
The promoter of Glav-12 was cloned as an EcoRI-SstI fragment in pGEM-4. This clone was subjected to an exonuclease III treatment which resulted in the formation of clones of different lengths. These clones were sequenced using the T7 primer. A subclone of the 5' end of the promoter was also obtained by cutting the Glav-12 promoter by HincII and ligating the plasmid on itself. The sequence was obtained using the T7 primer.

2.17. Primer extension on RNA from tobacco or oat seeds.

2.17.1. Oligonucleotide end labeling.

Three different primers were end labelled. GUS-p was specific to the GUS coding region downstream from the AUG. Av-1 was specific to the avenin sequences around the AUG (complementary sequence in bold) and G-1 was specific to the globulin sequences also around the AUG (complementary sequence in bold).

GUS-p 5' GCCCACAGGCCGTCGAG 3'

Av-1 5' CGGGATCCCATGGTGGATTTGTGTTGAC 3'

G-1 5' CTAGTAGTTGCCATG 3'

The end labelling was carried out at 37°C for one hour in the following buffer: 60mM Tris-HCl pH7.8, 10mM $MgCl_2$ and 15mM ß-mercaptoethanol. The quantities of oligonucleotides were 100ng of GUS-p and 30ng each of Av-1 and G-1. Ten units of T4 polynucleotide kinase were used in each reaction. The amounts of $^{32}P\gamma ATP$ were 60µCi for GUS-p and 20µCi each for Av-1 and G-1. The total volume of each reaction was 50µl. At the end of the reactions, the volume was brought to 250µl with distilled water and a phenol/chloroform extraction was performed. The labelled oligonucleotides were precipitated by addition of 1/10 vol. of Na acetate and 2 vol. of 100% ethanol overnight at -20°C. After centrifugation, the pellets were resuspended in 5µl (Av-1 and G-1) or 10µl (GUS-p) of dH_2O.

2.17.2. Primer annealing to the RNA.

The end labelled primers were annealed to the mRNAs in 100mM NaCl and 10mM Tris-HCl. Av-1 and G-1 were annealed to oat seed mRNA (2.5µg) and GUS-p was annealed to tobacco seed total RNA (10µg). The total volume was 20µl. The mixtures were heated at 90°C for one minute and then placed at 42°C for 30min after addition of 5 units of RNase inhibitor (RNasin, Promega).

2.17.3. Extension.

Prior to the addition of the different components of the reaction, 2.5µl of a solution of actinomycin D (1mg/ml ethanol) was placed in the microfuge tube and the ethanol was evaporated (Speed Vac, Savant). The primer extension was performed at 37°C for one hour using 150 units/reaction of the SuperScript Reverse Transcriptase (BRL). The reaction buffer was 50mM Tris-HCl pH 8.3, 75mM KCl, 3mM $MgCl_2$ and 1mM DTT. Ten µl of the annealing mix were used, 2µl of dNTPs (2mM of each dNTP) and 10 units of RNase inhibitor were added to the reaction. The final volume was 50µl. At the end of the reaction, a phenol/chloroform extraction was performed, followed by the precipitation of the nucleic acids after addition of 5µg of *E. coli* tRNAs as a carrier. The pellet was resuspended in 5µl of dH_2O and 3µl of loading buffer (formamide, 10mM EDTA pH 8.0, 0.001% xylene cyanol and 0.001% bromophenol blue) were added. The samples were boiled for 2min and 2µl were loaded on a polyacrylamide gel along with the sequencing reactions of the plasmid DNA (using the same primer) corresponding to the GUS construct transferred into the plants or the genomic clones (avenin or globulin).

2.18. Southern blots.

The restricted DNA was separated on an agarose gel (0.8 or 1%) at about 5V/cm, in TAE buffer. The DNA was stained with ethidium bromide and a Polaroid photograph (Polaroid MP.4 LandCamera) was taken under UV light (Fotodyne). The gel was soaked in a denaturing solution (1.5M NaCl, 0.5M NaOH) for 1 hour, followed by a 1 hour soaking in a neutralizing buffer (1M Tris-HCl pH 7.5, 1.5M NaCl). The DNA was transferred to a nitrocellulose membrane (Amersham) as described by Maniatis (1982): the gel was placed on a filter paper with both ends in contact with the transfer solution (10X SSC: 1.5M NaCl, 150mM sodium citrate, pH 7.0). The previously wetted nitrocellulose membrane (in dH_2O) was placed on top of the gel, then

six layers of 3MM filter paper (the first two wetted in 2X SSC), and finally a stack of paper towels were added. A weight of approximately 500g was applied uniformly on top of the assemblage. The transfer was allowed to proceed for at least 12 hours. The nitrocellulose membrane was removed, gently washed in 2X SSC, allowed to dry and finally baked for two hours at 80°C under vacuum to fix the DNA. Verification that the transfer was complete was done by restaining the gel with ethidium bromide.

2.19. Northern blots.

The RNA was directly transferred to a nylon membrane using the same technique as described for the Southern transfer, but without the denaturing and neutralizing steps. After completion of the transfer (12 to 15hrs), the locations of the wells were marked with a pen and the membrane was gently washed in 2X SSC and allowed to air-dry. The RNA was fixed to the membrane by first UV cross-link (3min illumination on each side of the membrane) and then by baking at 80°C under vacuum for 2hrs.

2.20. Radioactive DNA probes (random primer labeling).

2.20.1. Avenin probe (Figure 12.A).

The genomic clone was cut by HincII and PstI, and the DNA fragments were separated on an agarose gel. A light ethidium bromide staining was performed to locate the DNA band which was then excised from the gel. The agarose plug was centrifuged in a SpinX tube and the recovered DNA in solution was further purified on a silica matrix (Glass Milk, Bio101). The DNA quantity was estimated on an agarose gel. Fifty ng of DNA were denatured by heating the sample in boiling water for 3min. The random primer extension buffer was added to the denatured DNA (the random hexanucleotides were in a buffer containing Hepes pH 6.6, magnesium chloride, bovine serum albumin and ß-mercaptoethanol, Random Primer Extension Labeling

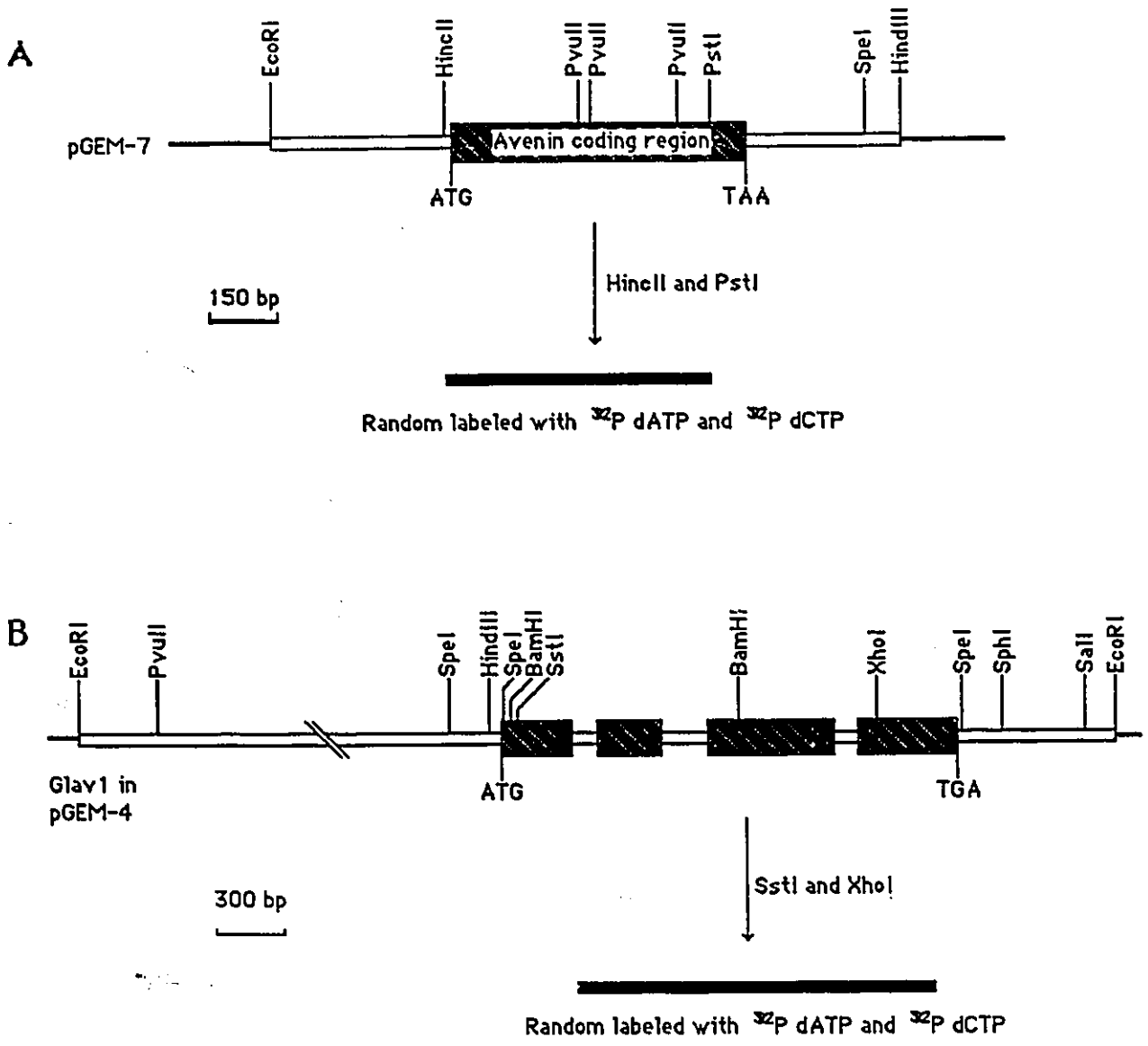


Figure 12 : Radioactive DNA probes.

A: Avenin probe. The avenin genomic clone was cut by HincII and PstI and the resulting fragment isolated. Incorporation of radionucleotides was done by the random labelling method.

B: Globulin probe. The Glav-1 sequence was cut by SstI and XhoI and the resulting fragment isolated. Incorporation of radionucleotides was done by the random labelling method.

System, NEN Dupont). Two µl of a 2mM solution of dGTP and dTTP were added. The ^{32}P-labeled dATP and dCTP (40µCi each) were mixed in, and 2 units of Klenow fragment (large fragment DNA polymerase I) were finally added. The reaction was allowed to proceed for 3hrs at room temperature. The labeled DNA fragments were purified on a Nensorb™20 cartridge following instructions by the manufacturer (NEN Dupont).

2.20.2. Globulin probe (Figure 12.B).

The same method as described above for the avenin probe was used to prepare a radioactive DNA probe specific for globulin sequences. The DNA fragment used in the random primer extension reaction was obtained by cutting the genomic clone pGlav-1 by SstI and XhoI.

2.20.3. GUS probe.

The GUS DNA fragment was obtained by cutting the plasmid pGEM-4-GUS by EcoRI and BamHI. The purifications and labeling of the DNA were done as described in 2.20.1.

2.21. Prehybridization and hybridization (Northern and Southern blots).

The nitrocellulose or nylon membranes were wetted in a 5X SSC solution for 5min and transferred to an hybridization tray. The prehybridization of the membranes was carried out in 50% deionised formamide, 5X SSC, 5X Denhardt's solution (50X Denhardt's solution was 1% ficoll, 1% PVPP and 1% BSA), 200µg/ml salmon sperm DNA and 0.1% SDS. The membrane was gently shaken (60 rpm) at 42°C for at least 5hrs. The radioactive probe was directly added to the prehybridization solution after having been boiled for 5min. The hybridization tray was carefully sealed and incubated overnight (12 to 15hrs) at 42°C on a rotating shaker (60 rpm). The DNA blots were

washed twice for 15min in a 2X SSC and 0.1% SDS solution at 45°C. They were thereafter washed twice for 10min in a 0.1X SSC and 0.1% SDS solution at 65°C. The RNA blots were washed twice for 10min in a 2X SSC and 0.1% SDS solution at 20°C, and then twice for 10min in a 0.2X SSC and 0.1% SDS at 45°C. The membrane, while still damp, was wrapped in plastic food wrap and placed in a cassette with an X-ray film and an intensifying screen.

2.22. Tobacco transformations.

2.22.1. Source of young leaves.

Tobacco seeds (*Nicotiana tabacum* cv. Delgold) were surface-sterilised in 70% ethanol for 30s, then in 1% Javex (6% sodium hypochlorite) for no more than 10 min and washed three times in sterile dH_2O. The seeds were germinated on MS 2% sucrose medium (Murashige and Skoog 1962, see appendix for compositions) solidified with 0.5% (w/v) agar. After two weeks, the seedlings were individually transferred to Magenta boxes containing 100 ml of sterile MS 2% sucrose medium solidified with 1% agar. These aseptically-grown young tobacco plants were kept at room temperature with a photoperiod of 16h of light and 8h of darkness and served as a source of young leaves for transformation.

2.22.2. Preparation of the inoculum.

A 50-ml YEP medium was inoculated with *A. tumefaciens* containing the Bin19 plasmid of interest. The medium contained 100µg/ml of kanamycin and was incubated at 28°C with shaking (250 rpm) for 24h. One ml of this culture was used to re-inoculate a fresh 50-ml culture. After 4 hours of incubation in the same conditions as above, the culture was poured into a sterile centrifuge bottle and centrifuged at 5,000 rpm for 5 minutes at 4°C. In the sterile environment of a laminar flow hood, the supernatant was poured out and the bottle inverted for a few minutes to remove all

the medium. The pellet was then resuspended in 5 ml of YEP (without kanamycin).

2.22.3. Transformation (Horsch *et al.* 1985).

Young tobacco leaves were cut from the aseptically-grown plants and placed in a Petri dish. One ml of *A. tumefaciens* suspension was added and diluted with 20 ml of YEP medium. The leaves were cut into small pieces (about 0.5 cm X 0.5 cm) while immersed in the bacterial suspension. These pieces were placed on a sterile filter paper covering a solidified medium (MS 3% sucrose, 1mg/l 6-benzyladenine (BA), 0.1mg/l alpha-naphthalene acetic acid (NAA), pH 5.7-5.8, agar 0.8%) in a small Petri dish (60mmX20mm, Nunc) which was then closed and sealed with parafilm. About 10 leaf "disks" were placed with *A. tumefaciens* in each Petri dish and about 12 such dishes were prepared for each construct. The light conditions were 16h/8h light/darkness at 20°C. After 48h, the leaf pieces were transferred onto the same solidified medium as described above but containing 500µg/ml of carbenicillin and without the sterile filter paper. The carbenicillin was added to kill the *A. tumefaciens* which had, by this time, transferred their T-DNA(s).

2.22.4. Regeneration and propagation of transformed plants.

Two days later, a transfer was made to MS 3% sucrose, 1mg/l BA, 0.1mg/l NAA, pH 5.7-5.8, agar 0.8%, 500µg/ml carbenicillin and 300µg/ml kanamycin, the latter being the selective agent. After a variable time (usually 2 to 4 weeks), small calli appeared, from which shoots developed (Figure 13. A and B). When these shoots were well developed, they were cut and placed into a rooting medium (MS 2% sucrose, pH 5.7-5.8, agar 0.8%, 500µg/ml carbenicillin and 300µg/ml kanamycin). After the roots had reached a length of at least 1cm (Figure 13. C), the plantlets were transferred into potting soil (Jiffy-7 pots) and kept in a high humidity environment by covering them with

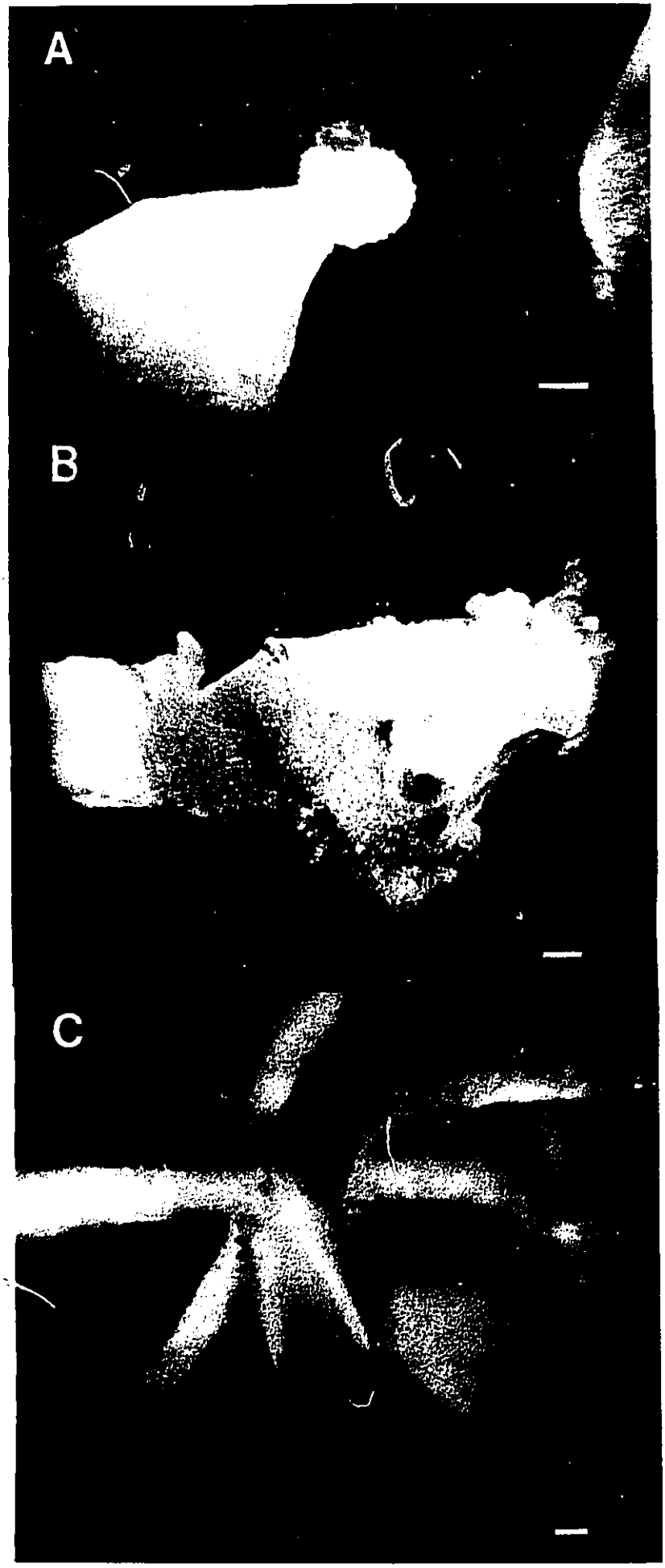

A
B
C

Figure 13 : Regeneration and propagation of transformed tobacco plants.

A: Tobacco leaf disk with a callus on which a shoot is forming. The agar medium is MS 3% sucrose with BA and NAA, carbenicillin and kanamycin.

B: Formation of multiple shoots on the edge of the tobacco leaf disk. The medium is the same as in A.

C: Formation of roots after the transfer of shoots on rooting medium. The agar medium is MS 2% sucrose with carbenicillin and kanamycin.

Bar represents 1mm.

a transparent plastic wrapping. When the roots visibly penetrated the sides of the pots, the Jiffy pots were placed into a bigger pot (20cm in diameter, 25cm high) filled with a mixture of potting soil (Premier), sphagnum peat moss (Premier), vermiculite (Peters) and turface (Aimcor) (5/5/2/2). Slow-release fertilizer (Nutricoat 14/14/14 type 100, Chisso-Asahi Fertilizer Co.) was directly added to the mixture. The plants were regularly watered and light conditions were the same as described above.

The total procedure, from transformation to seed maturation, took approximately 9 months.

2.23. Protein extractions.

2.23.1. Globulin extractions from oat and tobacco seeds.

Mature tobacco seeds (50mg) were frozen in liquid nitrogen and ground to a fine powder in a mortar and pestle. The powder was then transferred into a microfuge tube containing 500µl of extraction solution (1M NaCl). The mixture was thoroughly vortexed and let stand for one hour at 5°C with occasional vortexing. The cellular debris was removed by centrifugation (bench-top microfuge, top speed for 20 minutes). The supernatant was dialysed overnight (dialysis tubing n°3, Spectra/Por) against distilled water in a cold room with constant magnetic stirring of the water. The precipitated proteins were recovered and resuspended in a buffer (62.5mM Tris-HCl pH 6.8, 2% SDS). The protein concentration was estimated by the Bio-Rad Protein Assay Kit using casein as a reference (the presence of SDS required a dilution of the sample to reduce its interference with the kit reagents). The same procedure was used for oat seeds, except that 0.5g was ground and 5mls of extraction solution were used.

2.23.2. GUS protein extraction.

The GUS proteins were extracted as described by Jefferson (1987). The tobacco seeds (50mg) were ground in a mortar and pestle with

600µl of extraction buffer (50 mM $NaPO_4$ pH 7.0, 10 mM DTT, 1 mM EDTA pH 8.0, 0.1% sodium lauryl sarcosine, 0.1% Triton X-100). After centrifugation, the supernatant was kept and assayed to estimate the protein concentration.

All the different protein samples were kept at -70°C.

2.24. Western blots.

2.24.1. SDS-PAGE.

Sample buffer was added to the proteins (1 vol. of loading buffer for 1 vol. of protein sample). The composition of the loading buffer was adjusted according to Laemmli (1970): 62.5mM Tris-HCl pH 6.8, 20% glycerol, 2% SDS, 0.00125% (w/v) bromophenol blue and 10% 2-ß-mercaptoethanol (in reducing conditions). The separating gels and stacking gels were prepared as in Table 2.

The preparation of the acrylamide/Bis 30% stock was made using 29.2% acrylamide and 0.8% Bis-acrylamide (Sigma).

The dimensions of the gels were approximately 10x7x0.75 (cmXcmXmm). The running buffer was composed of 25mM TrisHCl pH8.3, 192mM glycine and 0.1% SDS. The apparatus was a Mini-V 8.10 electrophoresis apparatus (BRL) and the gels were run at 180V (constant voltage) for 50 min. Staining was performed with 0.1% Coomassie blue R-250 in fixative (40% methanol, 10% acetic acid). Destaining was done in 40% methanol, 10% acetic acid.

2.24.2. Protein transfer.

Transfer of the proteins was achieved using the same apparatus as for gel electrophoresis (see 2.24.1.), but with a different set-up. The polyacrylamide gel was equilibrated for 15min in the transfer buffer (25mM Tris-HCl, 150mM glycine, 20% methanol, pH 8.3). The nitrocellulose membrane (Amersham) was wetted in deionised dH_2O then placed in the transfer buffer for a few minutes. Assemblage was as described by the apparatus manufacturer: from bottom to top,

Table 2: Preparation of the stacking and separating gels for the SDS polyacrylamide gel electrophoresis.

	Separating gels (0.375M Tris, pH 8.8)		4% stacking gel (0.125M Tris, pH 6.8)
	12%	15%	
Distilled water	3.35ml	2.35ml	3.05ml
1.5M TrisHCl, pH 8.8	2.5ml	2.5ml	-
0.5M TrisHCl, pH 6.8	-	-	1.25ml
10% (w/v) SDS	100µl	100µl	50µl
Acrylamide/Bis (30% stock)	4.0ml	5.0ml	0.65ml
10% ammonium persulfate	50µl	50µl	25µl
TEMED	5µl	5µl	5µl
Final volume	10ml	10ml	5ml

pressure pad, chromatography paper, polyacrylamide gel, nitrocellulose membrane, chromatography paper and pressure pad. The transfer was done at constant voltage (150V) for one hour. The nitrocellulose membrane was air-dried for 30min and the gel was Coomassie stained (although the prestained molecular weight markers (Pharmacia) testified that the transfer had been complete).

2.24.3. Probing and detection (Amersham guidelines).

The membrane was then placed in a blocking buffer for one hour at room temperature with constant gentle rocking (blocking buffer: TBS-5% non-fat dry milk, TBS is 20mM Tris-HCl, 135mM NaCl, 2.7mM KCl, pH 7.8). Two different rabbit sera have been obtained: the first one was prepared by Dr. J. Schernthaner and the second one by Dr. G. Matlashewski (1983). The binding of the primary antibodies was carried out in TBS-1% non-fat dry milk overnight at room temperature with a rabbit serum dilution of 1/750. The membrane was washed thrice successively for 10min with TBS-Tween20 (0.05%). The binding of the secondary antibodies was performed in TBS-1% non-fat dry milk for two hours at room temperature. The goat anti-rabbit serum was at a dilution of 1/7500 (alkaline phosphatase was linked to the antibodies). The membrane was washed three times as described above. Detection of the alkaline phosphatase (AP) was achieved by placing the membrane in the reaction buffer in the presence of the substrate (the AP buffer was 100mM Tris-HCl pH 9.5, 100mM NaCl, 5mM $MgCl_2$, the substrates were NBT and BCIP, respectively 66µl and 33µl in 10mls of AP buffer). The reaction was stopped by several washes in dH_2O.

2.25. GUS fluorimetric assays.

The determination of GUS activities was performed as described by Jefferson (1987). The assay buffer contained 1 mM MUG (4-methyl umbelliferyl ß-D-glucuronide) in extraction buffer (see above). The

reaction volume was 500 µl which included 10µl of seed extract (for the Av-GUS and G12-GUS seeds) or 50µl of seed extract (for the Δ-259, Δ-247, Δ-237 and untransformed seeds). The reactions involving the Av-GUS and G12-GUS protein extracts were performed at room temperature in assay buffer pre-heated to 37°C. After 5, 10 and 30 minutes, 100µl of the reaction were removed and immediately mixed with 900µl of stop solution (0.2M Na_2CO_3). For the reactions with the deletions and untransformed seed protein extracts, 50µl of seed protein extract were used and the reactions were performed at 37°C. The different time points were 30 min, 1.5hr and 18h. The fluorimetric readings were taken on a Hitachi F-2000 Fluorescence Spectrophotometer calibrated with 0, 100nM and 1µM methylumbelliferone in stop solution. The average value of two fluorimetric readings per sample was established. A linear regression was calculated for each sample and the GUS activity was adjusted to the protein concentration and the initial volume assayed.

2.26. GUS histochemical staining (Jefferson 1987).

Hand-cut sections of fresh mid-maturation to late maturation whole tobacco seed pods (2mm thick) were incubated in phosphate buffer (50mM $NaPO_4$ pH 7.0) containing 0.5mg/ml of X-Gluc (5-bromo-4-chloro-3-indolyl glucuronide). The ß-glucuronidase substrate was infiltrated into the tissues by placing the preparation under vacuum for 5min. The incubation was then carried out at 37°C.

Removal of the embryo from the tobacco seeds was performed under a stereoscope. Each part of the seed (embryo and seed coat/endosperm) was individually placed into a well of a micro-titration plate. These tissues were submerged with the phosphate buffer containing X-Gluc, placed under vacuum for 5min and incubated at 37°C.

2.27. Computer software.

DNAstar was used for the analysis of the DNA and protein sequences.

Figures were drawn using Superpaint version 1.0 from Silicon Beach Software, Inc.

This text was written using Microsoft Word version 4.0 from Microsoft Corporation.

Graphs were drawn using Cricket Graph version 1.0 from Cricket Software.

The bibliography was arranged using EndNote from Niles & Associates, Inc.

The computer, hard disk and printer were Macintosh.

Chapter 3
Results

3.1. Oat globulin genes and promoters.

3.1.1. Isolation of oat globulin genes.

Two different globulin sequences were isolated from the λgt10 library using a partial globulin cDNA clone as a probe. They were named Glav-1 and Glav-3 (for Globulin of *Avena*). They were cloned into pGEM-4 for analysis and a restriction map was established. At a later time, an additional clone (Glav-12) was made available. The approximate location of the coding region was determined by Southern analysis: the two clones were double digested by different sets of restriction enzymes and the DNA fragments were probed with the partial cDNA clone. On both clones (pGlav-1 and pGlav-3), the coding region was located on the 3' end of the inserts. The orientations of the sequences were also established by a short sequencing on both ends of the inserts of some subclones, which also confirmed that the clones were actual globulin clones. From these initial analyses, it appeared that the promoter of the Glav-1 sequence was of slightly greater length than the promoter of Glav-3; pGlav-1 was therefore primarily chosen for further analyses.

3.1.2. Glav-1.

3.1.2.1. Sequence of pGlav-1.

The Glav-1 sequence was divided into two subclones of similar length, pG1SF and pG1SRL (see section 2.15). These two subclones were progressively shortened by an exonuclease III treatment (Henikoff 1984). Nine subclones for pG1SRL and fourteen for pG1SF were thus generated. All these clones were independently sequenced at least twice and the overlap between two adjacent sequences was at least 30 nucleotides. In most of the cases, overlapping was in the order of 50 to 100 nucleotides, which, with the estimated size of the deleted fragments, should have prevented the omission of repeated sequences. The complete sequence of pGlav-1, which is 5273 bp long, is presented in Figure 14.

3.1.2.2. Start of transcription.

The start of transcription of the globulin mRNAs was determined by primer extension analysis on oat RNA enriched in polyA$^+$ RNA by one round on oligo-dT cellulose. One major signal was obtained (Figure 21.A) which corresponds to an adenosine residue on the DNA sequence at position 2583. The 5' untranslated region (5'UTR) of the globulin mRNA is therefore 34 nucleotides long. Although the 5'UTR of Glav-1 is very similar to a previously published oat globulin genomic sequence (OG1-E1, Shotwell *et al.* 1990), the site of the start of transcription is different. The 5' end of the mRNA had been assigned to a T residue, making the 5'UTR 38 nucleotides long. Another oat globulin genomic sequence (ASglo5, Schubert *et al.* 1990) has a transcriptional start site at the same position as in Glav-1.

Comparison of the three UTRs (below) shows a very high degree of similarity. The start of transcription is indicated by an asterisk.

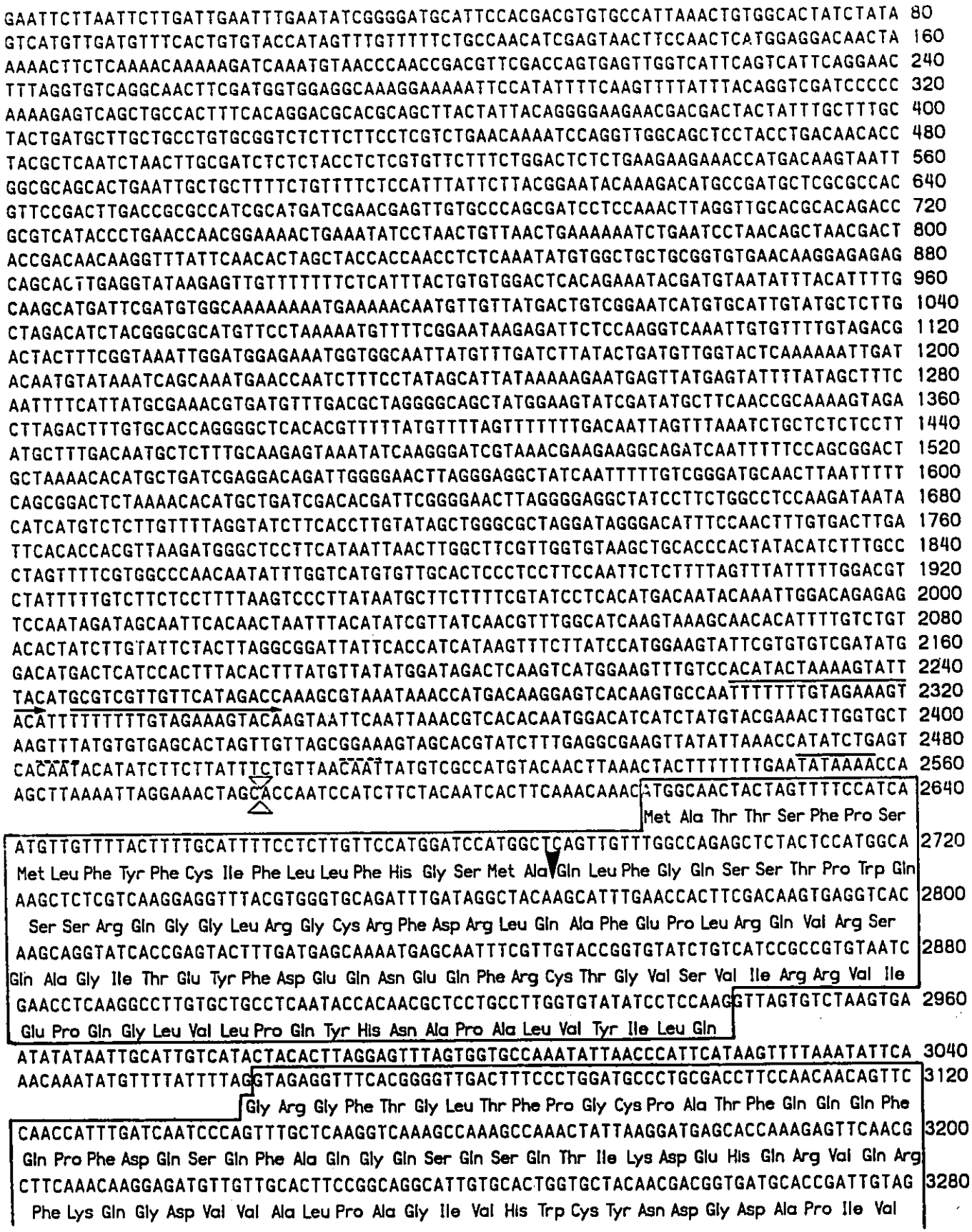
```
OG1-E1    *TAGCACCAATCCACCTTCTACAATCTCTTCAAACAATCATG
Glav-1    TAGC*ACCAATCCATCTTCTACAATCACTTCAAACAAACATG
Asglo5    TAGC*ACCAATCCACCTTCTACAATCTCTTCAAACAATCATG
```

Figure 14 : Nucleotide and deduced amino acid sequences of the oat genomic clone Glav-1.

Promoter: The start of transcription, as determined by primer extension analysis, is indicated by two open arrowheads. The possible TATA box is overlined. Two possible CAAT boxes are overlined by a dotted line. The two arrows indicate a repeated sequence.

Coding sequence: The coding nucleotide sequence and the deduced amino acid sequence are boxed. The first solid arrowhead points to the cleavage site of the signal peptide. The second solid arrowhead indicates the cleavage site between the α and ß polypeptide. The arrows overline the five octapeptide repeats.

3' end: The potential polyadenylation sequences are underlined.




CAATCTATGTCTTTGATGTAAACAACAACGCTAATCAGCTTGAACCTAGACAAAAGGTAATTATATAAATTAATCCACAT 3360
Ala Ile Tyr Val Phe Asp Val Asn Asn Asn Ala Asn Gln Leu Glu Pro Arg Gln Lys
AACAAATATATAAATGTTTAGAAGTTCTTGTTGGCTGGTAACAACAAGTTCTTGTTGGCTGGTAACAACGCTAATCAGCT 3440
TGAACCTAGACAAAAGGTAATTATACAAAAGTATATATTATTGGGGTATTAATGTGAACTTTGTTTTACTCTATCCATAT 3520
AAATTGTCAGGAGTTCTTGTTGGCTGGTAACAACAAGAGAGAGCAACAGTCTGGAAATAACATATTCAGTGGATTAAGTG 3600
Glu Phe Leu Leu Ala Gly Asn Asn Lys Arg Glu Gln Gln Ser Gly Asn Asn Ile Phe Ser Gly Leu Ser
TCCAACTTCTTAGTGAGGCCCTTGGTATAAGTCAACAAGCAGCACAAGGATCCAAGTCAAATGACCAAAGAGGTAGAGTA 3680
Val Gln Leu Leu Ser Glu Ala Leu Gly Ile Ser Gln Gln Ala Ala Gln Gly Ser Lys Ser Asn Asp Gln Arg Gly Arg Val
ATTCGTGTGAGTCAAGGCCTTCAATTCTTGAAGCCCATTGTGTCCCAACAAGTACCAGTAGAGCAGCAAGTCTACCAACC 3760
Ile Arg Val Ser Gln Gly Leu Gln Phe Leu Lys Pro Ile Val Ser Gln Gln Val Pro Val Glu Gln Gln Val Tyr Gln Pro
AATTCAAACTCAAGATGTACAAGCAACCCAATACCAGGTAGGGCAATCAACCCAATACCAGGTAGGGAAATCAACCCCAT 3840
Ile Gln Thr Gln Asp Val Gln Ala Thr Gln Tyr Gln Val Gly Gln Ser Thr Gln Tyr Gln Val Gly Lys Ser Thr Pro
ATCAAGGAGGACAATCAAGTCAATACCAGGCAGGACAGTCATGGGACCAAAGTTTCAACGGTTTGGAGGAAAACTTTTGT 3920
Tyr Gln Gly Gly Gln Ser Ser Gln Tyr Gln Ala Gly Gln Ser Trp Asp Gln Ser Phe Asn Gly Leu Glu Glu Asn Phe Cys
TCATTGGAGGCAAGGAAAAACATTGAAAACCCCCAACATGCCGACACATACAACCCACGTGCTGGCAGGATAACACGTCT 4000
Ser Leu Glu Ala Arg Lys Asn Ile Glu Asn Pro Gln His Ala Asp Thr Tyr Asn Pro Arg Ala Gly Arg Ile Thr Arg Leu
CAATAGCAAGAATTTCCCCATCCTTAACATTGTGCAAATGAGTGCTACAAGAGTAAATCTATACCAGGTATATATGATAC 4080
Asn Ser Lys Asn Phe Pro Ile Leu Asn Ile Val Gln Met Ser Ala Thr Arg Val Asn Leu Tyr Gln
TACACTCTATAGACTCTCTTATTTTTAGGTAGTATAAGATTCATAACAATCGGTTAATAATATGATCTTATAAATTACTA 4160
TTGCAGAATGCCATTCTTTCACCATTCTGGAACATTAATGCGCACAGTGTCATCTACATGATCCAAGGACATGCTCGAGT 4240
Asn Ala Ile Leu Ser Pro Phe Trp Asn Ile Asn Ala His Ser Val Ile Tyr Met Ile Gln Gly His Ala Arg Val
TCAAGTCGTCAACAACAATGGCCAGACTGTATTCAGTGATATTCTTCATCGAGGACAACTTCTAATCGTACCACAACACT 4320
Gln Val Val Asn Asn Asn Gly Gln Thr Val Phe Ser Asp Ile Leu His Arg Gly Gln Leu Leu Ile Val Pro Gln His
TTGTTGTTCTCAAGAATGCAGAGCGTGAAGGATGCCAATACATTTCATTCAAGACTAACCCAAACTCCATGGTTAGTCAC 4400
Phe Val Val Leu Lys Asn Ala Glu Arg Glu Gly Cys Gln Tyr Ile Ser Phe Lys Thr Asn Pro Asn Ser Met Val Ser His
ATCGCAGGAAAGACCTCCATCCTACGTGCCTTACCTATCGATGTCCTCGCCAATGCATACCGCATTTCTAGACAGGAAGC 4480
Ile Ala Gly Lys Thr Ser Ile Leu Arg Ala Leu Pro Ile Asp Val Leu Ala Asn Ala Tyr Arg Ile Ser Arg Gln Glu Ala
CCGAAACCTTAAAAACAACCGAGGAGAAGAGTTTGGTGCATTCACACCTAAACTTACCCAAACAGGCTTCCAGAGTTATC 4560
Arg Asn Leu Lys Asn Asn Arg Gly Glu Glu Phe Gly Ala Phe Thr Pro Lys Leu Thr Gln Thr Gly Phe Gln Ser Tyr
AAGACATCGAGGAGGCGTCATCTTCGGCTGTTAGAGCATCCGAGTGAATGAGTGTAATGGGAACTAGTATAGTAAAATAA 4640
Gln Asp Ile Glu Glu Ala Ser Ser Ser Ala Val Arg Ala Ser Glu
AGGCATTGCAAGTGTGCAACCTAGTGTCATATAAGTGCTTACCTGAATAAAAACTTTACCATGTTATATCACTTGTTTGT 4720
GTCTTGTACCTTTCTTAAATTTATCTCTTCTGAATAATCTCCTTCCCCCTCCCTCGATTTTCTGCTTGTTTTTACCTCAT 4800
GTGCATGCATATCGACGAGGTATATGGGCCAGTTGGACAGGATTTTTTTTTCAAAGTACATCTTTTTTGGTCCTTCCTAT 4880
AACTACACTTTTTATTTAGTTCTTGGTTGCTAGATTTTTTCATAAATTCGAACCACCTAAAATTATGAATCACATAGGTT 4960
AAAAGTTCTATGTCAAAGTTGTGCAACTTTAATAAATAAAACCCTAATTTTTGGTTTATTGATTATGATACCTTTTATAC 5040
AGGACATTGCACACTGCAACTCCATTAAGCATCTCTTATCACATATTATTGACATACCGAGCACACATTGAAGATAATTA 5120
ATCTTGACTACATCCTTGAGTACACACACGTCGACAATGTTGTGTTGTTTGTTGGCAAGGTGAGGATTTTTCCTGTAGCT 5200
GAAGGAGCTACATAATTTATCTGTACATAATTCATTTATCAAGATTTATGATGTGTCTTCAGTTCGAGAATTC 5273

As for the CAAT box, multiple potential copies can be found in the three sequences at different places.

3.1.2.3. Coding regions (exons), introns and possible termination signals.

The coding sequences and the potential cleavage sites of Glav-1 were determined by comparison with the published sequences of oat globulin cDNAs (Shotwell *et al.* 1988, Walburg and Larkins 1986) and oat globulin genomic clone OG1-E1 (Shotwell *et al.* 1990) and ASglo5 (Schubert *et al.* 1990). The coding regions (four exons) encode a 526 amino acid polypeptide. The signal peptide is 24 amino acids long, the cleavage site being between an alanine and a glutamine residue. The acidic polypeptide corresponds to the two first exons and part of the third. This acidic polypeptide is 300 amino acids long. The basic polypeptide is 202 amino acids long and corresponds to part of the third exon and the fourth exon. The cleavage site between the α (acidic) and ß (basic) polypeptide lies between an asparagine and a glycine residue, as previously determined by amino acid sequencing of the N terminus of the basic polypeptide (Shotwell *et al.* 1988). Five octapeptide repeats are present in the variable region of the acidic polypeptide; only four are present in OG1-E1.

The A+T content of the three introns are respectively 73%, 71% and 74%, and the A+T content of the four exons are respectively 54%, 54%, 56% and 55%. This high AT content is required for the efficient splicing out of the pre-mRNAs in plants (Goodall and Filipowicz 1989). The splice sites of these introns conform to the splice junctions of plant genes (Brown 1986).

A possible coding region is present in the second intron (Figure 15). This extra exon would be in-frame with the nucleotide sequence of the second exon and would add 26 amino acids to the acidic polypeptide. The splice sites and the A+T content of the introns (84%

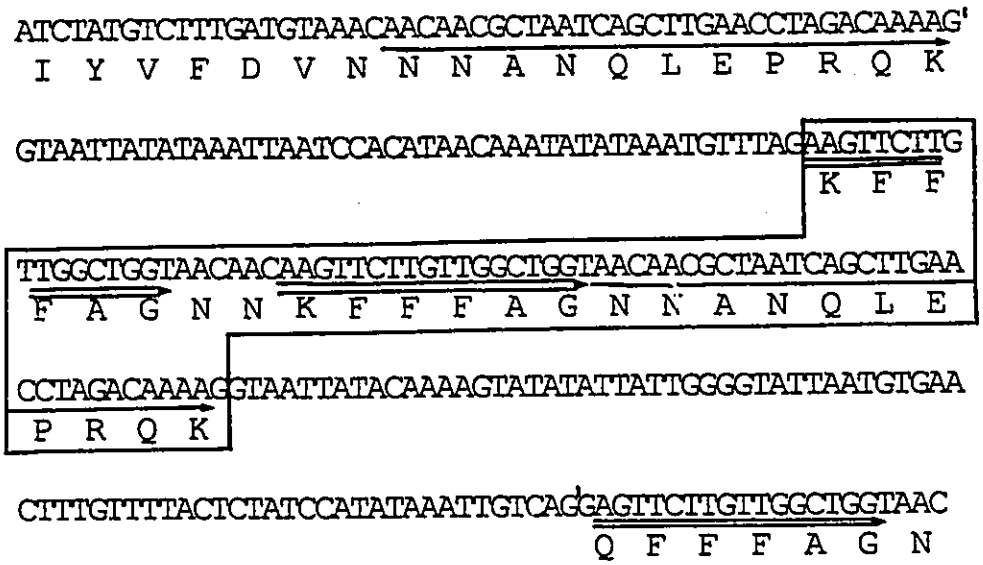


Figure 15 : Second intron of Glav-1

Shown are the end of the second exon and the beginning of the third exon with their respective amino acid sequence. A potential coding sequence in this second intron is boxed. The arrows indicate the presences of duplicated sequences.

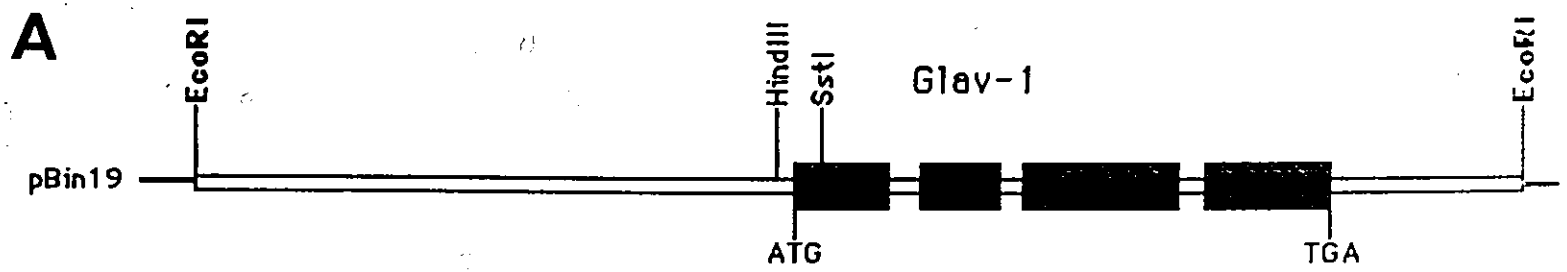
and 75%) and the exon (58%) are valid to support the actual presence of this coding sequence. However, the length of the first intron (45 nt) seems to be too short to be efficiently spliced out, but the second intron (74 nt) would probably be correctly spliced out (Goodall and Filipowicz 1990). This possible extra region of coding sequence is composed of two small amino acid repeats (FFFAG) probably duplicated from the 5' end of the third exon, followed by a longer duplicated sequence which is present at the end of the Glav-1 second exon. This latter duplicated sequence also marks the end of this possible extra exon (Figure 15).

Two possible polyadenylation sequences are present in the 3' end of Glav-1. They are located 29 and 79 nucleotides downstream of the stop codon and their positions are very similar to the ones found in OG1-E1 (33 and 83 nt downstream of the stop codon).

3.1.2.4. Tobacco transformation experiments with the complete Glav-1 sequence.

a. Southern analysis of the regenerated plants.

The complete sequence of Glav-1 was. inserted in pBin19 and tobacco plants were transformed via *A. tumefaciens*. Thirteen plants were regenerated and analysed by Southern blots (Figure 16). All the plants tested were transformed, as shown by the presence of the expected 5.2 kb hybridizing fragment. Plant 6 does not show a signal because the DNA escaped from the well during the loading but showed a hybridization signal in another blotting experiment (data not shown). Rearrangements occurred in two of the plants (2 and 3). The copy number of the insert was not determined but greatly varies from plant to plant.


A
EcoRI
HindIII
SstI
Glav-1
EcoRI
pBin19
ATG
TGA

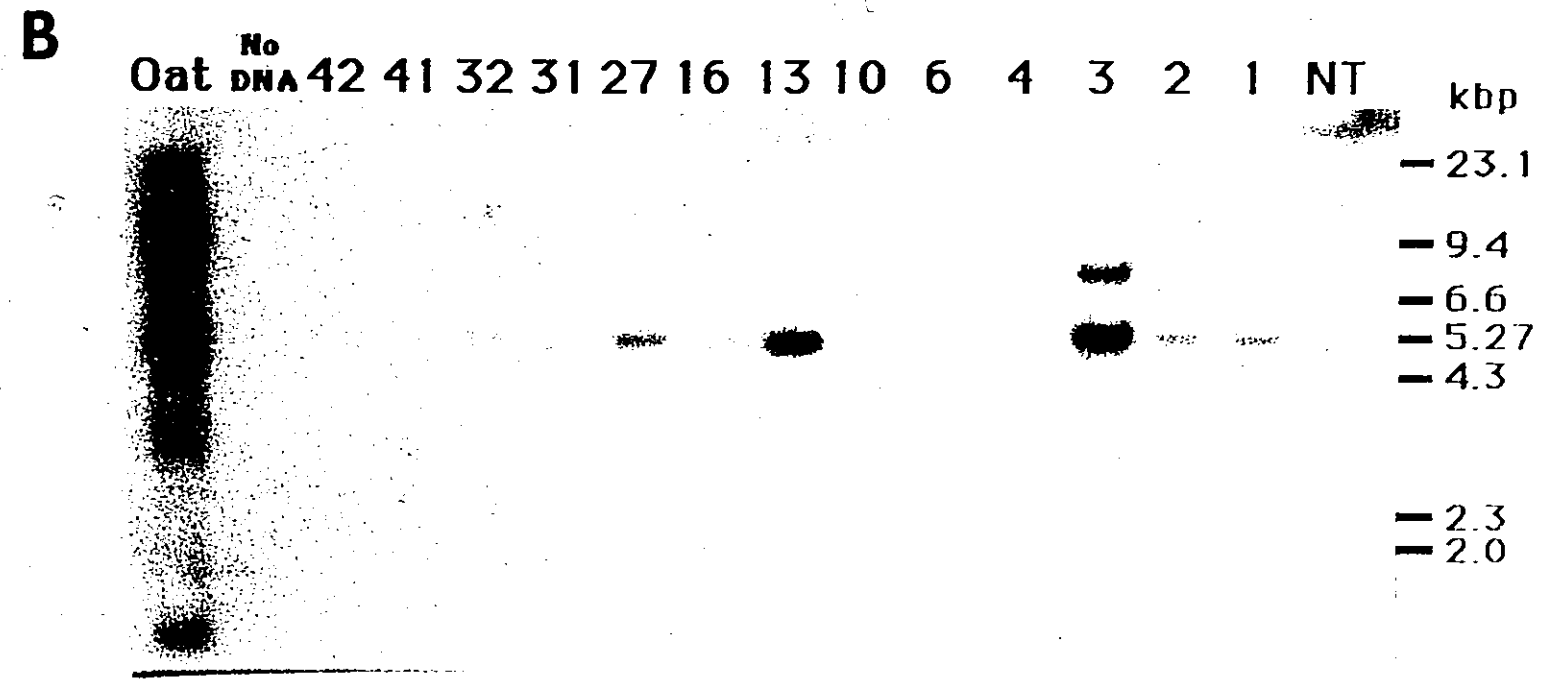

B
Oat
No DNA
42
41
32
31
27
16
13
10
6
4
3
2
1
NT
kbp
23.1
9.4
6.6
5.27
4.3
2.3
2.0

Figure 16: Southern blot analysis of tobacco plants transformed with the entire Glav-1 sequence.

A: Construct containing the Glav-1 sequence used for the transformation of tobacco plants.

B: Leaf DNA (5µg) extracted from different plants was digested by EcoRI and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose filter and hybridized to a random labeled globulin probe. The lane numbers indicate the number of the plants. NT indicates an untransformed plant. The size markers are from lambda DNA digested by HindIII.

b. RT-PCR analysis of the transformed plants.

Northern analyses of transformed tobacco seeds did not show any hybridization signal. At this point of the study, three hypotheses could be proposed:

1) the promoter is not functional,

2) Glav-1 has a very low degree of expression,

3) the second intron is not correctly spliced out (see 3.1.2.3.) and/or the mRNA becomes highly unstable.

The Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) was used to test these hypotheses. Two oligonucleotide primers were designed, one at the end of the first exon and the other at the beginning of the second, which would allow testing the three hypotheses at once. RT-PCR products run on an agarose gel showed multiple bands when performed on oat seed RNA but nothing when used with mid-maturation tobacco seed RNA. However, when tobacco seed RNA was mixed with oat RNA, no PCR products could be obtained (result not shown). By using ^{32}PdCTP in the reverse transcriptase step and running the products on a denaturing agarose gel, it was then shown that there was no incorporation of radioactive dCTP when tobacco seed RNA was mixed with oat seed RNA, whereas there was incorporation into cDNAs of different sizes when oat seed RNA only was used. It was then concluded that the tobacco seed RNA extracts contained co-extracted compounds that inhibit the reverse transcriptase (different types of reverse transcriptase were used and they were all inhibited). Attempts to further purify the tobacco seed RNA were made so that it could be used in RT-PCR experiments. Several methods were used to try to remove the coloured contaminations which seemed to be phenolic and polyphenolic compounds. These attempts included the use of different RNA extraction kits supplied by different manufacturers, the use of an oligo(dT) cellulose column and anti-oxidants during the extraction procedure. Although addition of ascorbic acid in the extraction buffer

considerably reduced the colouring of the RNA extract, it was still not enough to prevent the inhibition of the reverse transcriptase. Finally, addition of activated charcoal to the RNA resuspended in water (no salt present) completely removed the colouring and the RNA extract no longer inhibited the reverse transcriptase. However, the analysis of transformed tobacco seeds from three different plants by the RT-PCR procedure showed the absence of Glav-1 transcripts. Even a Southern analysis of the RT-PCR products failed to detect any amplified sequence (data not shown).

3.1.2.5. Tobacco transformations with the Glav-1 promoter fused to the GUS coding region.

In parallel to the transformation experiments with the entire Glav-1 genomic sequence, 2680 bp of 5' upstream sequence of Glav-1 and two derivative deletions (Δ-2364 and Δ-1420 from the transcription start point) were fused to the GUS coding region from pBI101.1 and these constructs were used for tobacco transformations. Self-pollinated seeds from the regenerated plants were directly analysed by GUS histochemical staining. No GUS activity was detectable in the mature tobacco seeds, even after prolonged incubation with the X-Gluc substrate (data not shown).

3.1.2.6. Summary.

From the results of the above experiments using the Glav-1 promoter, the first hypothesis (promoter not functional) appears likely. Considering the high sensitivity of the RT-PCR, the second hypothesis (low degree of expression by the Glav-1 promoter) could probably be ruled out. The third hypothesis (instability or low efficiency of the pre-mRNA maturation) remains difficult to assess because of the non-functionality of the promoter. Since there is no in-frame stop codon in the coding regions, Glav-1 cannot be defined as a pseudogene. Glav-1 can therefore be called an inactive gene. The

term silent gene would refer to a non-functional promoter fused to a functional coding region (i.e. no in-frame stop codon and a relatively good efficiency of the splicing mechanisms) and would require additional investigations, especially in the second intron of Glav-1 (see 3.1.2.3.)

3.1.3. Glav-12 promoter.

3.1.3.1. Sequence of the promoter.

After the exonuclease III treatment of pG4-G12ES (see 2.16.), twelve deletion subclones and one subclone obtained with the use of restriction sites were selected and sequenced. The complete sequence of the Glav-12 5' genomic end is presented in Figure 17. This sequence (2702 bp long) includes the 5' UTR, the translational start codon and a small portion of the first exon. At least the first sixty nucleotides upstream of the ATG and also the sequence complementary to the G-1 primer are identical. The primer extension signal obtained in section 3.1.2.2. is therefore valid for the positioning of the start of transcription on the Glav-12 promoter. The possible TATA and CAAT boxes are also situated at the same positions with regard to the ATG. A more detailed analysis of the promoter sequence is presented in section 3.1.5.1.

3.1.3.2. Tobacco transformation with the Glav-12 promoter fused to the GUS coding region.

The complete 5' upstream region of the Glav-12 genomic clone was fused to the GUS coding sequence (translational fusion) and transferred into tobacco plants via *A. tumefaciens*. The translational fusion changed the NH2-terminal region of the GUS protein as shown below:

GUS fusion sequence: **ATG** GCA ACT ACT AGA GGA TCC CCG GGT GGT CAG TCC CTT **ATG** ...

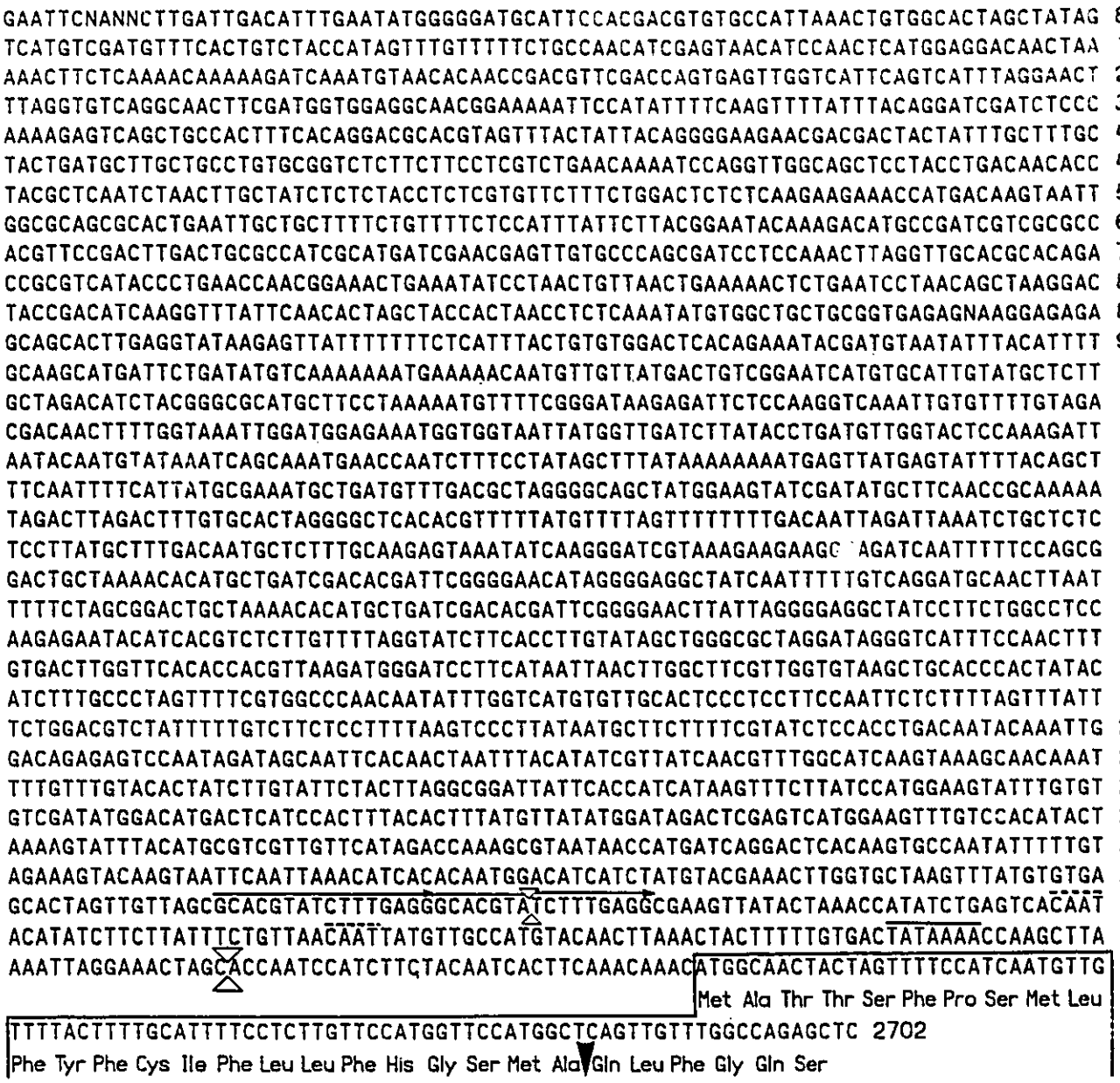


Figure 17 : Nucleotide sequence of the Glav-12 promoter.

The beginning of the coding sequence and its deduced amino acid sequence is partially boxed. The solid arrowhead indicates the cleavage site of the signal peptide.
The two big open arrowheads point to the start of transcription determined by primer extension on oat mRNAs. The two little open arrowheads point to the start of transcription determined by primer extension on tobacco RNA after fusion of the promoter to the GUS reporter gene. The two arrows overline a repeated sequence. Possible TATA and CAAT boxes are overlined respectively by a solid line and a dotted line.

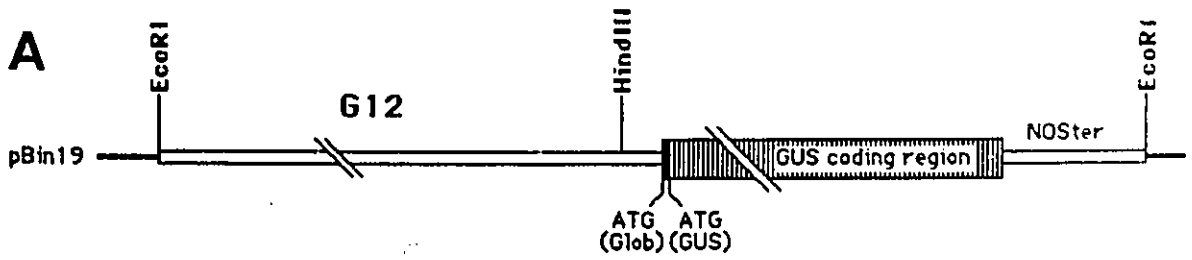
The first ATG corresponds to the ATG of the globulin sequence whereas the second ATG, which remains in frame with the first one, belongs to the GUS coding sequence. The nucleotides in between these two ATGs, except for the six nucleotides downstream of the first ATG, came from the polylinker region of pBI101.1.

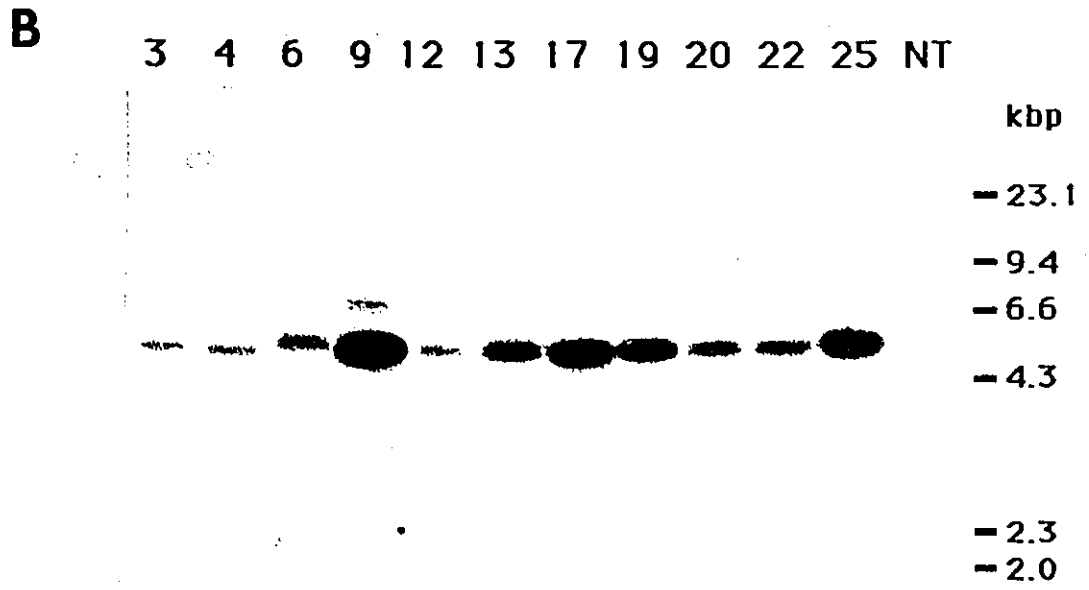
a. Southern analysis of the regenerated plants.

Southern analysis was performed on the first eleven regenerated plants. The integrity and presence of the gene fusion in the tobacco genome was determined by digesting the tobacco DNA with EcoRI and using the GUS coding sequence as a probe. Figure 18.B shows that all the tested plants were transformed and possess the expected ≈5 kb hybridizing fragment. Rearrangements occurred in four of the plants (6, 9, 17 and 25) resulting in the loss of at least one EcoRI site. In order to estimate the number of copies inserted in the tobacco genome, the genomic DNA was also restricted with HindIII which cuts only once within the T-DNA. Figure 18.C shows that only one copy was present in plants 3, 6, 12 and 20. Two copies were inserted in plant 4, three in plant 13 and 22, four copies in plant 9, seven in plant 17, five in plant 19 and about five in plant 25. The more intense bands present in the hybridization patterns of plants 17, 19 and 25 are counted as two inserts. Whether these two inserts correspond to tandem repeats or were only random integration events at the same distance from a HindIII site cannot be determined.

b. Northern analysis of some transformed plants.

Two attempts to analyse by Northern hybridizations the RNA extracted from leaf, stem, root and seeds at different stages of maturation from two transformed plants were made. Unfortunately, the experiments were not successful. The hybridization signals were smears and it was not possible to determine the technical reasons for the inability to detect the GUS transcripts.


A
EcoRI
G12
HindIII
EcoRI
pBin19
GUS coding region
NOSter
ATG (Glob)
ATG (GUS)


B
3 4 6 9 12 13 17 19 20 22 25 NT
kbp
23.1
9.4
6.6
4.3
2.3
2.0

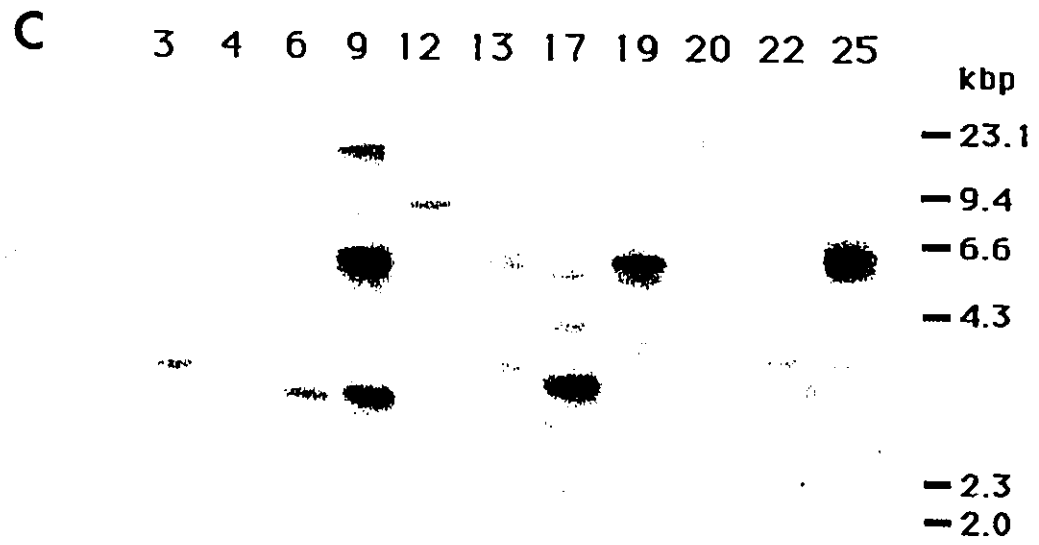

C
3 4 6 9 12 13 17 19 20 22 25
kbp
23.1
9.4
6.6
4.3
2.3
2.0

Figure 18 : Southern blot analysis of tobacco plants transformed with the G12-GUS construct.

A: G12-GUS construct used for tobacco transformation.

B: Leaf DNA (5µg) extracted from different plants was digested by EcoRI and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose membrane and hybridized to a random labeled GUS probe. The lane numbers refer to the number of the plants. NT indicates an untransformed plant. The size markers are from lambda DNA cut by HindIII.

C: Same as in B except that the DNA was digested by HindIII.

c. Histochemical analysis of transformed tobacco seed pods.

Hand-cut cross-sections of transformed tobacco seed pods were subjected to histochemical staining. The maturation stage of the seed pods was about three weeks after hand pollination. GUS activity was made visible by the formation of a blue precipitate at the site of enzyme activity. After about half an hour of incubation at 37°C, a blue precipitate started to appear in the endosperm of the tobacco seeds. After one hour of incubation, photographs were taken and are presented in figure 19 A to D. The presence of the blue precipitate was restricted to the endosperm tissue of the transformed seeds (a histochemical assay on untransformed tobacco seed is shown in figure 32.D). Neither the seed coat nor the embryo or other maternal tissues forming the tobacco seed pod were stained. However, when the incubation was allowed to proceed for a longer period of time, the blue precipitate also appeared in the embryos of the seeds. To verify the endosperm specificity of the GUS activity, individual seeds were dissected: the embryo was separated from the rest of the seed (seed coat and endosperm) and the different parts were individually incubated with the GUS substrate. Figure 19.E shows that the GUS activity is present only in the endosperm. The embryos, even after an overnight incubation with X-Gluc, did not show any GUS activities. There is a possibility that the primary product of GUS (soluble and colourless indoxyl derivative) diffuses from the endosperm to the embryo and be oxidized in the embryo to form the blue precipitate (by oxidative dimerization). This would account for the blue staining of the embryo when it is left in contact with the endosperm. Although not completely ruled out, diffusion of the GUS enzyme itself is unlikely because of its size (around 70 kDa). The blue precipitate occurs only when the endosperm is made directly available to the substrate by cutting the seed and exposing the endosperm.

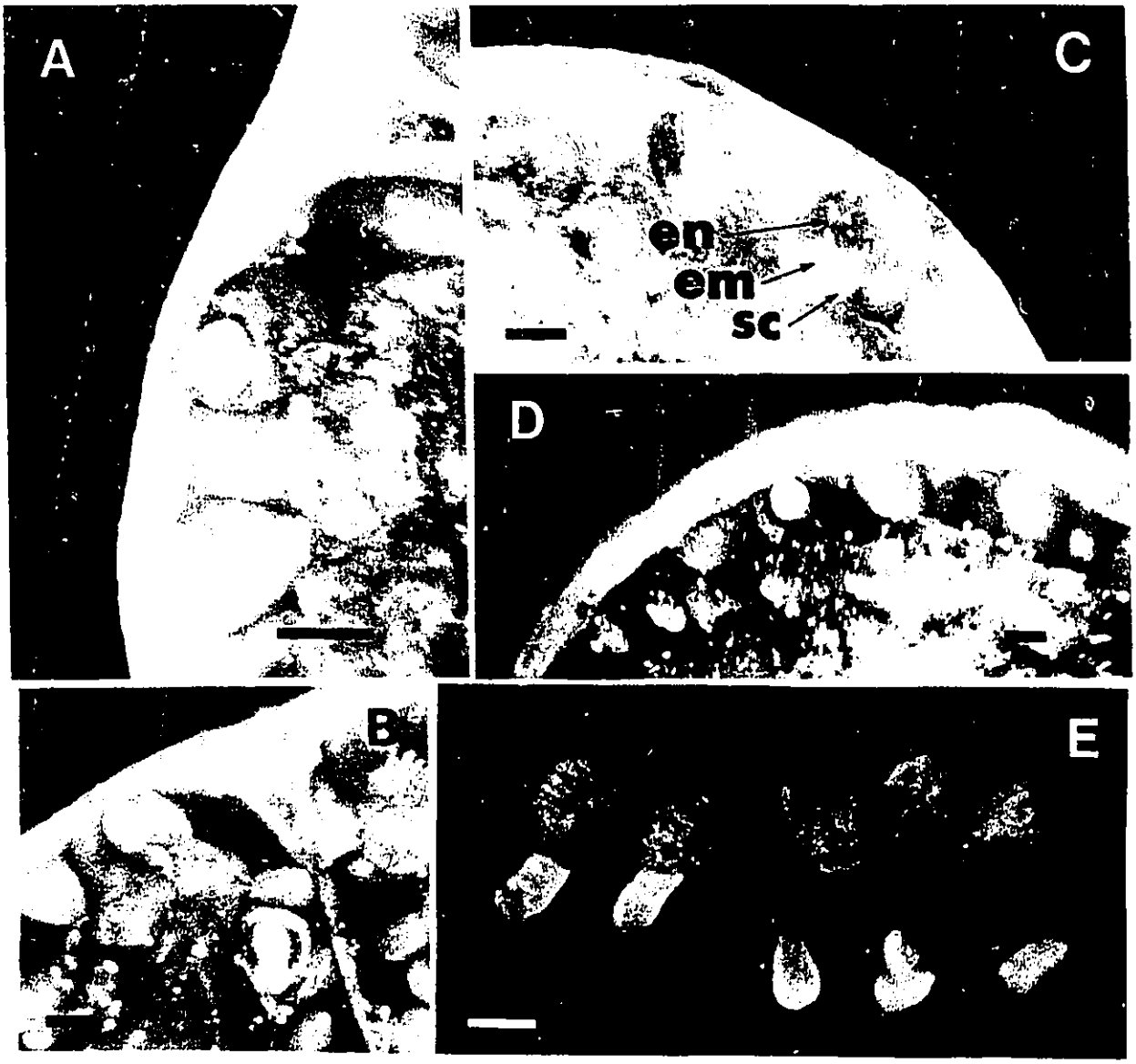

A
B
C
en
em
sc
D
E

Figure 19 : Histochemical GUS staining of cross-sections of seed pods from tobacco plants transformed with the G12-GUS construct.

A is the same as **B** but at a higher magnification. The seed pod is from plant 7. The blue staining indicates the presence of the GUS which is restricted to the endosperm.

C: Cross-section of a tobacco seed pod from plant 22. The blue staining is restricted to the endosperm (en). The maternal tissues, the embryo (em) and the seed coat (sc) do not show any GUS activities.

D: Cross-section of a tobacco seed pod from plant 7.

For **A, B, C** and **D**, the staining was performed for one hour.

E: Embryos from tobacco seeds (plant 19) were dissected and separated from the endosperm and seed coat. The staining reaction was done separately for each part of the seed for three hours at 37°C. Only the endosperm shows GUS activity. Each vertical set of embryo/endosperm-seed coat comes from the same seed.

Bar represents 0.2mm.

When the tobacco seeds are not cut open, GUS activities cannot be detected by histochemical staining (Figure 19).

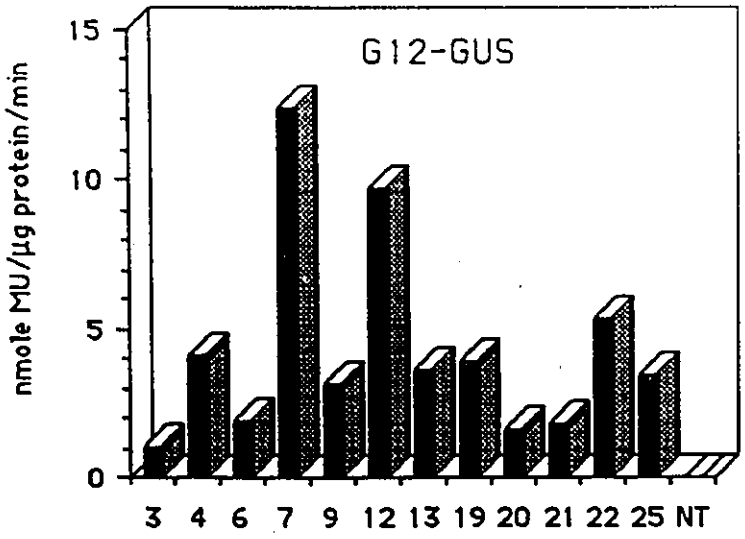
GUS enzyme activity was not detected at early stages of seed maturation. There was no formation of blue precipitate when seeds at stage 7 days after pollination or earlier were used for histochemical staining (results not shown). This is partial evidence that the GUS is developmentally expressed during seed maturation when fused to the Glav-12 promoter, with the GUS proteins starting to accumulate in the endosperm one week after pollination.

d. GUS activities in mature seeds.

Mature seeds collected from the different plants (including those that had not previously been analysed by Southern hybridization) were assayed for GUS activity. These seeds were positive for GUS activity on the basis of histochemical staining. Figure 20 shows the GUS activities of the different seed extracts. There is no evidence of correlation between the copy number of integrated GUS genes and the level of GUS activities in the different plants. It is particularly visible in plant 12 where a unique copy was integrated and which shows high GUS activity. In that special case, the site of integration of the T-DNA is presumably located at a transcriptionally active DNA region. It is therefore very likely that the observed variability of GUS expression is determined by the position effect of the introduced genes.

e. Primer extension analysis.

RNA extracted from mid-maturation tobacco seeds from two transformed plants was analysed by the primer extension method to determine the 5' end of the GUS transcripts. An oligonucleotide primer specific to the beginning of the GUS coding sequence was used for that purpose. The extension step was carried out twice in two independent


G12-GUS
nmole MU/µg protein/min
15
10
5
0
3 4 6 7 9 12 13 19 20 21 22 25 NT

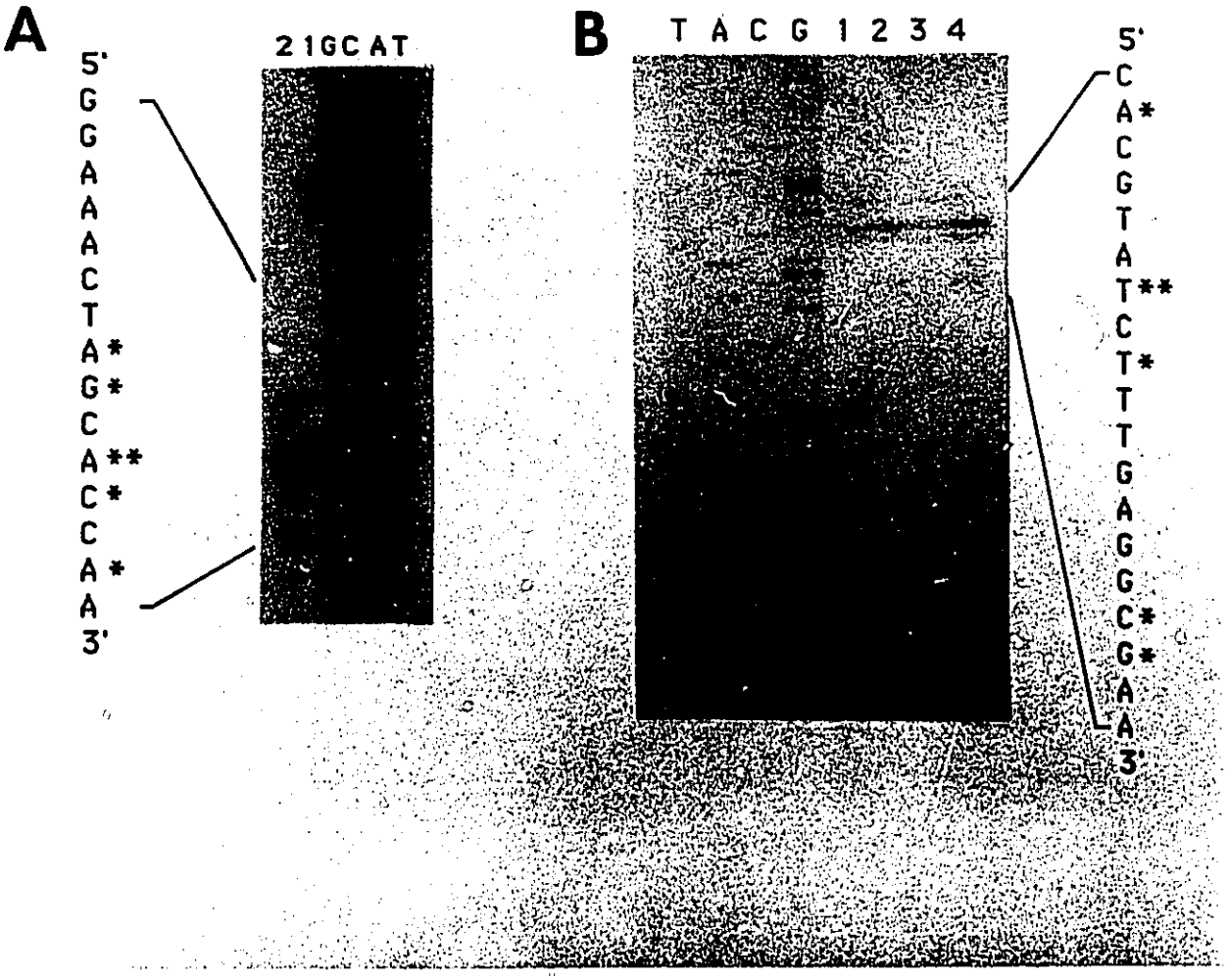

A
21GCAT
5'
GGAAACTA*G*CA**C*CA*A
3'
B
T A C G 1 2 3 4
5'
CA*CGTAT**CT*TTGAGGC*G*AA
3'

Figure 20 : GUS activities in mature seeds from tobacco plants transformed with the G12-GUS construct.

GUS activities were measured fluorometrically and expressed in nmole 4-MU per µg protein per min. The plant numbers are indicated below the bars. NT represents an untransformed plant.

Figure 21 : Determination of the 5' terminus of the globulin mRNA in oat seed and of the GUS mRNA in transgenic tobacco seeds by primer extension analysis.

A: The G-1 primer was ^{32}P end-labeled and annealed to 2.5µg of oat seed mRNA. Two extension reactions were realized and are indicated by lane numbers 1 and 2. The same primer was used to sequence the Glav-1 genomic clone by the dideoxy chain termination method (lanes G, C, A and T). The two asterisks indicate the major product obtained and a single asterisk shows the position of the minor bands.

B: The GUS-p primer was ^{32}P end-labeled and annealed to 10µg of total tobacco seed RNA. Two extension reactions were performed on RNA extracted from two plants transformed with the G12-GUS construct: lanes 1 and 2 correspond to plant 7, and lanes 3 and 4 correspond to plant 9. The same oligonucleotide primer was used to sequence the plasmid pBin19-G12-GUS by the dideoxy chain termination method (lanes T, A, C and G). The two asterisks indicate the major product obtained and a single asterisk shows the position of the minor bands.

reactions with the same annealing mixture. Figure 21.B shows that one major signal is obtained together with other less intense bands.

The band patterns obtained from the two different plants (plants 7 and 9) are identical, suggesting that the DNA rearrangement observed in one copy of the inserted GUS constructs in plant 9 did not affect qualitatively the transcription. The 5' end of the mRNA is assigned to a T residue, making the 5'UTR 138 nucleotides long. This transcriptional site is located in approximately the middle of the second direct repeated sequence of 16 nucleotides present in the Glav-12 promoter between positions 2416 and 2447 in Figure 17.

f. Summary.

From the different experiments conducted with the Glav-12 promoter, it has been shown that the expression of a reporter gene under the control of an oat globulin promoter is restricted to the endosperm of transgenic tobacco seeds. This spatial expression is accompanied by a developmental regulation of the accumulation of the GUS protein which is stably stored within the endosperm. However, as seen in the primer extension analysis, there is a difference in the transcriptional mechanism of tobacco versus oat.

3.1.4. Glav-3.

3.1.4.1. Sequence of pGlav-3.

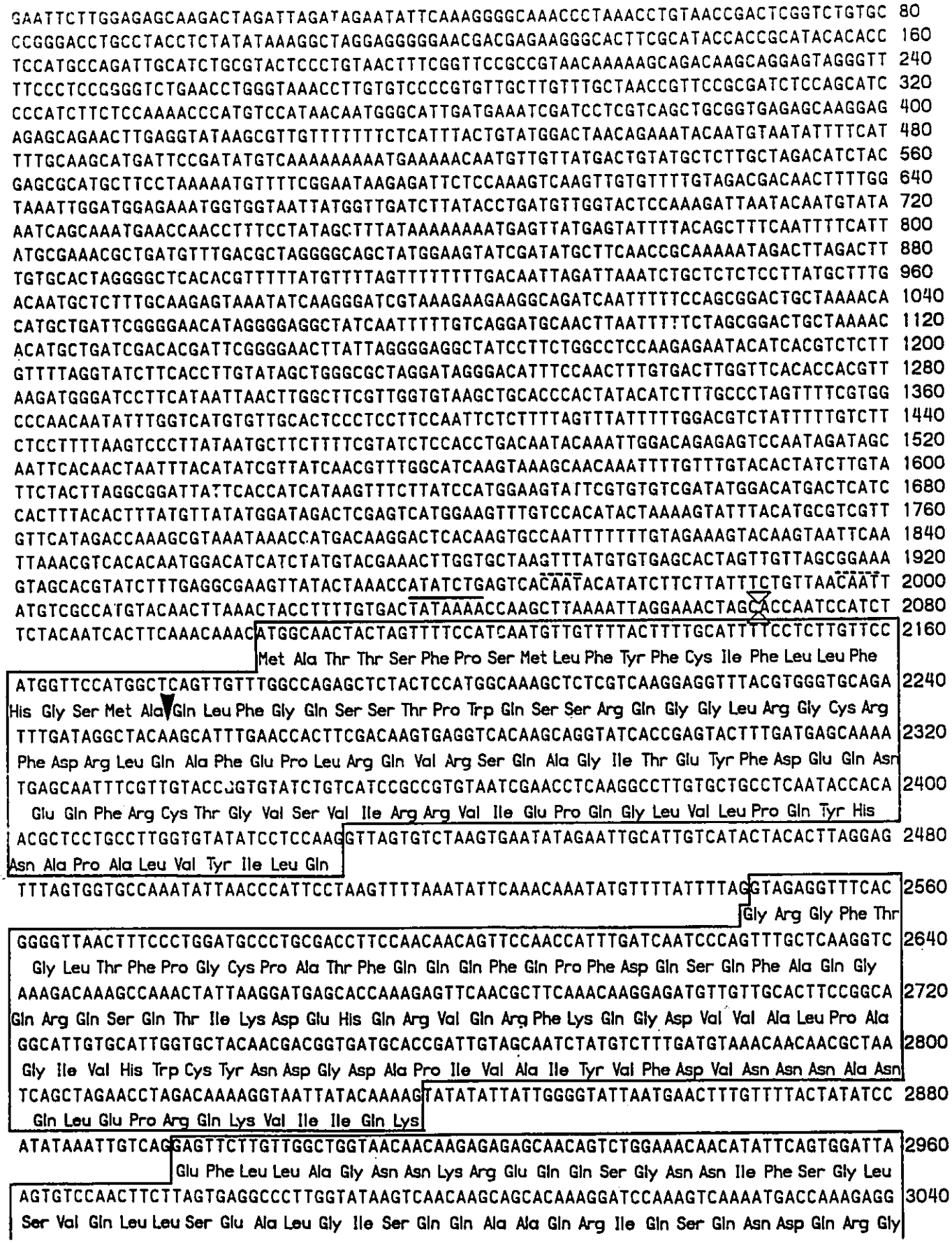
The complete sequence of Glav-3, sequenced by Dr. M. Tanchak, is presented for comparison in Figure 22. The two genomic clones, pGlav-1 and pGlav-3, are homologous. The nucleotide similarity between the two clones is very high (over 95%), as well as the deduced amino acid sequences of the exons. The introns are positioned at the same place in both Glav-1 and Glav-3. The predicted length of the Glav-3 protein is 527 amino acids, including a signal peptide of 24 amino acids. The α polypeptide consists of 301 amino

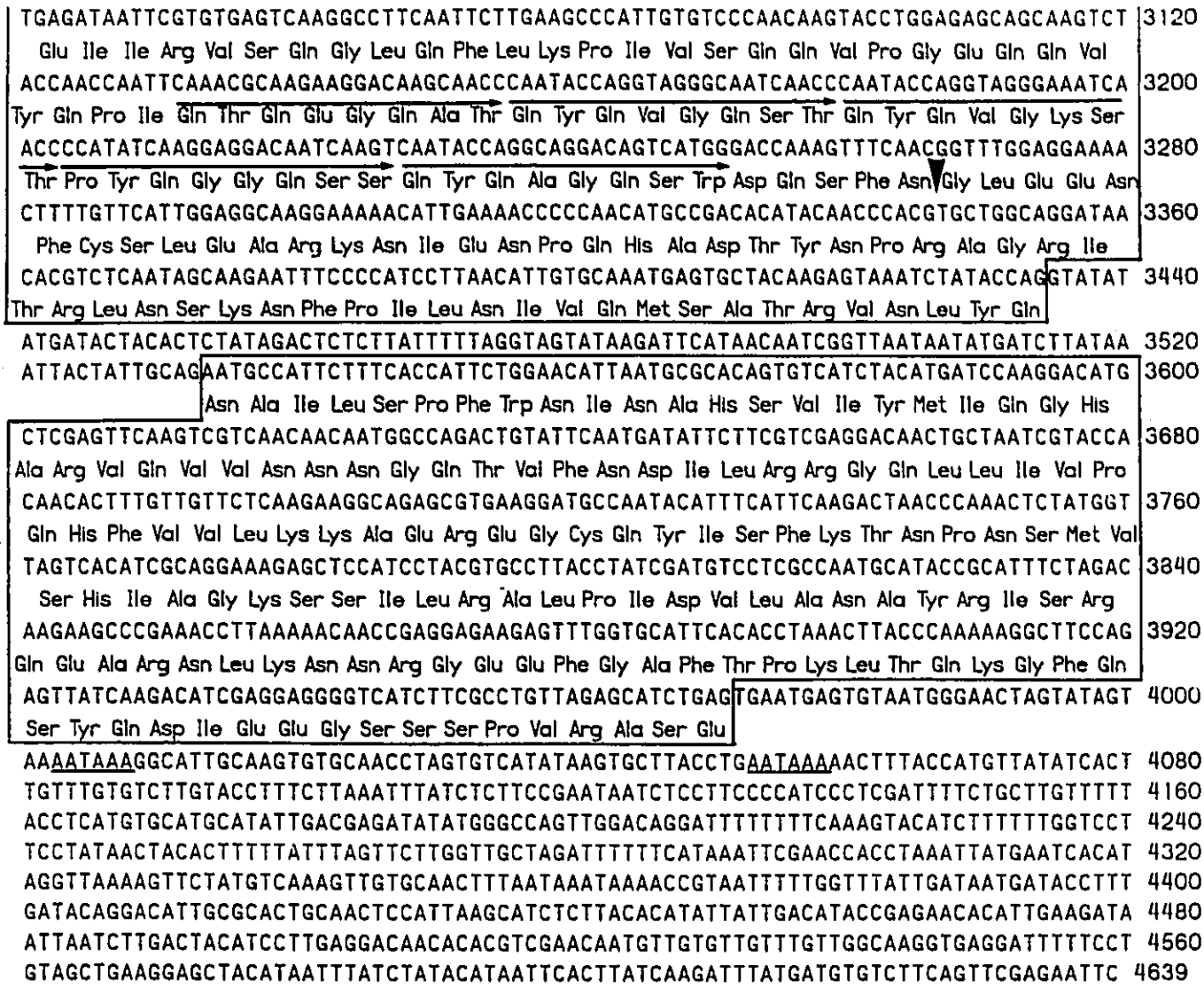
Figure 22 : Nucleotide and deduced amino acid sequences of the oat genomic clone Glav-3 (sequenced by Dr. M. Tanchak).

Promoter: The start of transcription, as determined by primer extension analysis, is indicated by two open arrowheads. The possible TATA box is overlined. Two possible CAAT boxes are overlined by a dotted line.

Coding sequence: The coding nucleotide sequence and the deduced amino acid sequence are boxed. The first solid arrowhead points to the cleavage site of the signal peptide. The second solid arrowhead indicates the cleavage site between the α and ß polypeptide. The arrows overline the five octapeptide repeats.

3' end: The potential polyadenylation sequences are underlined.




TGAGATAATTCGTGTGAGTCAAGGCCTTCAATTCTTGAAGCCCATTGTGTCCCAACAAGTACCTGGAGAGCAGCAAGTCT 3120
Glu Ile Ile Arg Val Ser Gln Gly Leu Gln Phe Leu Lys Pro Ile Val Ser Gln Gln Val Pro Gly Glu Gln Gln Val
ACCAACCAATTCAAACGCAAGAAGGACAAGCAACCCAATACCAGGTAGGGCAATCAACCCAATACCAGGTAGGGAAATCA 3200
Tyr Gln Pro Ile Gln Thr Gln Glu Gly Gln Ala Thr Gln Tyr Gln Val Gly Gln Ser Thr Gln Tyr Gln Val Gly Lys Ser
ACCCCATATCAAGGAGGACAATCAAGTCAATACCAGGCAGGACAGTCATGGGACCAAAGTTTCAACGGTTTGGAGGAAAA 3280
Thr Pro Tyr Gln Gly Gly Gln Ser Ser Gln Tyr Gln Ala Gly Gln Ser Trp Asp Gln Ser Phe Asn Gly Leu Glu Glu Asn
CTTTTGTTCATTGGAGGCAAGGAAAAACATTGAAAACCCCCAACATGCCGACACATACAACCCACGTGCTGGCAGGATAA 3360
Phe Cys Ser Leu Glu Ala Arg Lys Asn Ile Glu Asn Pro Gln His Ala Asp Thr Tyr Asn Pro Arg Ala Gly Arg Ile
CACGTCTCAATAGCAAGAATTTCCCCATCCTTAACATTGTGCAAATGAGTGCTACAAGAGTAAATCTATACCAGGTATAT 3440
Thr Arg Leu Asn Ser Lys Asn Phe Pro Ile Leu Asn Ile Val Gln Met Ser Ala Thr Arg Val Asn Leu Tyr Gln
ATGATACTACACTCTATAGACTCTCTTATTTTTAGGTAGTATAAGATTCATAACAATCGGTTAATAATATGATCTTATAA 3520
ATTACTATTGCAGAATGCCATTCTTTCACCATTCTGGAACATTAATGCGCACAGTGTCATCTACATGATCCAAGGACATG 3600
Asn Ala Ile Leu Ser Pro Phe Trp Asn Ile Asn Ala His Ser Val Ile Tyr Met Ile Gln Gly His
CTCGAGTTCAAGTCGTCAACAACAATGGCCAGACTGTATTCAATGATATTCTTCGTCGAGGACAACTGCTAATCGTACCA 3680
Ala Arg Val Gln Val Val Asn Asn Asn Gly Gln Thr Val Phe Asn Asp Ile Leu Arg Arg Gly Gln Leu Leu Ile Val Pro
CAACACTTTGTTGTTCTCAAGAAGGCAGAGCGTGAAGGATGCCAATACATTTCATTCAAGACTAACCCAAACTCTATGGT 3760
Gln His Phe Val Val Leu Lys Lys Ala Glu Arg Glu Gly Cys Gln Tyr Ile Ser Phe Lys Thr Asn Pro Asn Ser Met Val
TAGTCACATCGCAGGAAAGAGCTCCATCCTACGTGCCTTACCTATCGATGTCCTCGCCAATGCATACCGCATTTCTAGAC 3840
Ser His Ile Ala Gly Lys Ser Ser Ile Leu Arg Ala Leu Pro Ile Asp Val Leu Ala Asn Ala Tyr Arg Ile Ser Arg
AAGAAGCCCGAAACCTTAAAAACAACCGAGGAGAAGAGTTTGGTGCATTCACACCTAAACTTACCCAAAAAGGCTTCCAG 3920
Gln Glu Ala Arg Asn Leu Lys Asn Asn Arg Gly Glu Glu Phe Gly Ala Phe Thr Pro Lys Leu Thr Gln Lys Gly Phe Gln
AGTTATCAAGACATCGAGGAGGGGTCATCTTCGCCTGTTAGAGCATCTGAGTGAATGAGTGTAATGGGAACTAGTATAGT 4000
Ser Tyr Gln Asp Ile Glu Glu Gly Ser Ser Ser Pro Val Arg Ala Ser Glu
AAAATAAAGGCATTGCAAGTGTGCAACCTAGTGTCATATAAGTGCTTACCTGAATAAAAACTTTACCATGTTATATCACT 4080
TGTTTGTGTCTTGTACCTTTCTTAAATTTATCTCTTCCGAATAATCTCCTTCCCCATCCCTCGATTTTCTGCTTGTTTTT 4160
ACCTCATGTGCATGCATATTGACGAGATATATGGGCCAGTTGGACAGGATTTTTTTTCAAAGTACATCTTTTTTGGTCCT 4240
TCCTATAACTACACTTTTTATTTAGTTCTTGGTTGCTAGATTTTTTCATAAATTCGAACCACCTAAATTATGAATCACAT 4320
AGGTTAAAAGTTCTATGTCAAAGTTGTGCAACTTTAATAAATAAAACCGTAATTTTTGGTTTATTGATAATGATACCTTT 4400
GATACAGGACATTGCGCACTGCAACTCCATTAAGCATCTCTTACACATATTATTGACATACCGAGAACACATTGAAGATA 4480
ATTAATCTTGACTACATCCTTGAGGACAACACACGTCGAACAATGTTGTGTTGTTTGTTGGCAAGGTGAGGATTTTTCCT 4560
GTAGCTGAAGGAGCTACATAATTTATCTATACATAATTCACTTATCAAGATTTATGATGTGTCTTCAGTTCGAGAATTC 4639

acids with four octapeptide repeats (a fifth could be included although it is imperfect) at the COOH terminus and the estimated molecular weight is 34 kDa. The ß polypeptide is 202 amino acids long with a molecular weight of 23 kDa. The start of transcription, the possible TATA and CAAT boxes and the possible polyadenylation signals are located at the same positions as in the Glav-1 sequence.

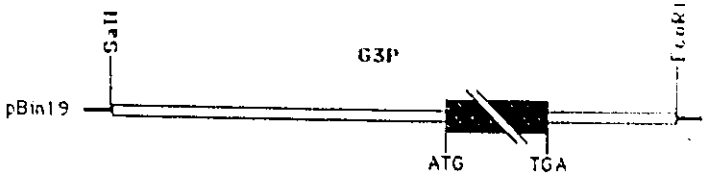
3.1.4.2. Tobacco transformation with the Glav-3 sequence.

The entire sequence of Glav-3 was inserted in the T-DNA of pBin19. Two independent tobacco transformations were attempted but neither of them produced any calli. It seemed that the T-DNA would not transfer or insert itself properly. Both transformation attempts were done using the Glav-3 sequence inserted in the same direction. DNA conformation of the T-DNA could be a possible reason for the failure to obtain transformed calli. It was then decided to remove about 300 bp of the Glav-3 sequence (at the 5' end) so as to slightly modify the construct. Transformation experiments made with that new construct (G3P) enabled the production of calli from which plants were regenerated.

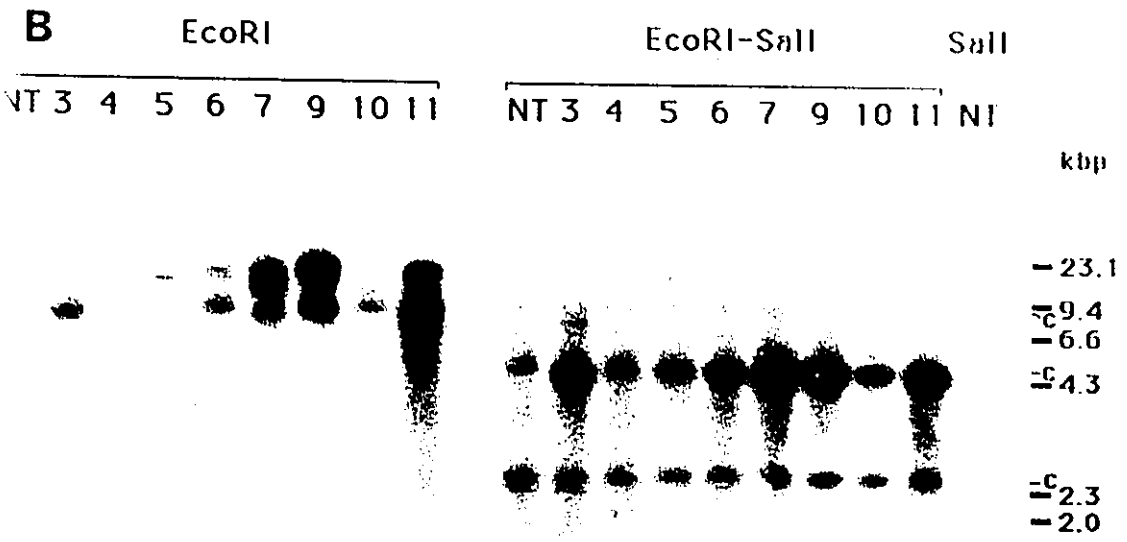
a. Southern analysis of the regenerated plants.

The first eight regenerated plants were assayed for the presence of the Glav-3 sequence by Southern analysis. Figure 23.B shows that all the plants carried at least two copies of the insert (EcoRI digest). Three of the plants show a rearrangement of the insert (plants 3, 7 and 9) but still carry multiple copies of intact inserts as shown by the hybridization pattern when the DNA is double cut with EcoRI and SalI (Figure 23.B).

A


SalI
G3P
EcoRI
pBin19
ATG
TGA

B


EcoRI
EcoRI-SalI
SalI
NT 3 4 5 6 7 9 10 11
NT 3 4 5 6 7 9 10 11 NT
kbp
23.1
9.4
6.6
4.3
2.3
2.0

C

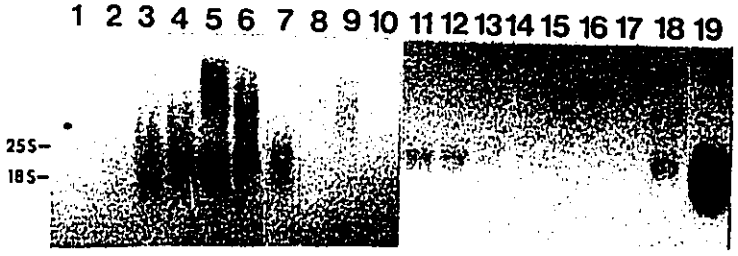

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
25S–
18S–

Figure 23 : Southern and Northern blot analyses of tobacco plants transformed with the Glav-3 genomic sequence (G3P construct).

A: Construct G3P used for the tobacco transformations.

B: Southern blot analysis of regenerated tobacco plants. DNA (5µg) was digested by EcoRI and EcoRI-SalI and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose membrane and hybridized to a random labeled globulin probe. The lane numbers refer to the number of the transgenic plants. NT indicates an untransformed plant. The size markers are from lambda DNA cut by HindIII. A contamination (probably in the SalI reaction buffer) is indicated by the letter c.

C: Northern blot analysis of transformed tobacco seeds. RNA was extracted from seeds at different stages of maturation and from leaf, stem and root. Five µg of total RNA were separated in a 1.2% agarose gel and blotted on a nylon membrane. A random labeled globulin probe was used for the detection of globulin transcripts.
Lanes **1** to **6**: Tobacco seed RNA from plant 3 at stages 5, 10, 15, 20, 25 and 30 DAP. Lane **7**: Tobacco seed RNA from plant 20 at stage 20 DAP. Lanes **8, 9** and **10**: Tobacco RNA from plant 4 extracted respectively from leaf, stem and root. Lanes **11, 12** and **13**: Tobacco seed RNA from plant 4 extracted respectively at stages 20, 25 and 30 DAP. Lane **14**: Tobacco seed RNA (mid-maturation) from an untransformed plant. Lanes **15** and **16**: Tobacco RNA from leaf and stem of plant 7. Lane **17**: Tobacco seeds RNA (mid-maturation) from plant 12 transformed with the G12-GUS construct. Lane **18**: Tobacco seed RNA (mid-maturation) from plant 7. Lane **19**: 350ng of oat seed mRNA (mid-maturation).

b. Northern analysis of some transformed plants.

Plants transformed with the oat globulin Glav-3 sequence were manually self-pollinated and the tobacco seed pods were collected at different time points. RNA was extracted from the seeds and analysed by Northern blots. Figure 23.C shows the presence of the globulin transcripts in seeds from plants 3, 4, 7 and 20. The expression of the globulin gene is restricted to the seeds: there is no detectable globulin transcripts in the leaves, roots or stems of the transformed plants (lanes 8, 9, 10, 15 and 16). There is also no expression of the Glav-3 gene in untransformed tobacco seeds (lane 14) or in the seeds of a plant transformed with a GUS construct (lane 17). Lanes 1 to 6 respectively correspond to RNA extracted during the maturation of the seeds at 5, 10, 15, 20, 25 and 30 days after pollination (DAP). The increase of the hybridization signal shows that the expression of the Glav-3 gene is developmentally regulated. There is no detectable expression at day 5 or before (lane 1), and the expression starts at or around 10 DAP (lane 2). The maximum expression seems to be at 25 DAP (lane 5) for plant 3, and between 20 and 25 DAP for plant 4 (lanes 11 and 12). The size of the transcripts, compared to the oat RNA (lane 19), seems to be longer than the majority of the transcripts obtained in oat seeds. This could be due to the longer 5' UTR of the globulin transcript when expressed in tobacco seeds, assuming that the Glav-12 and Glav-3 promoters initiate transcription at the same site in tobacco seeds and that the poly(A) tails are also of similar length.

c. Western analysis of some transformed plants.

Globulins were extracted from mature tobacco seeds and subjected to Western analysis under reduced or unreduced conditions. Two batches of anti-oat globulin rabbit sera were used. A specific hybridization signal (Figure 24.D), although weak with reduced

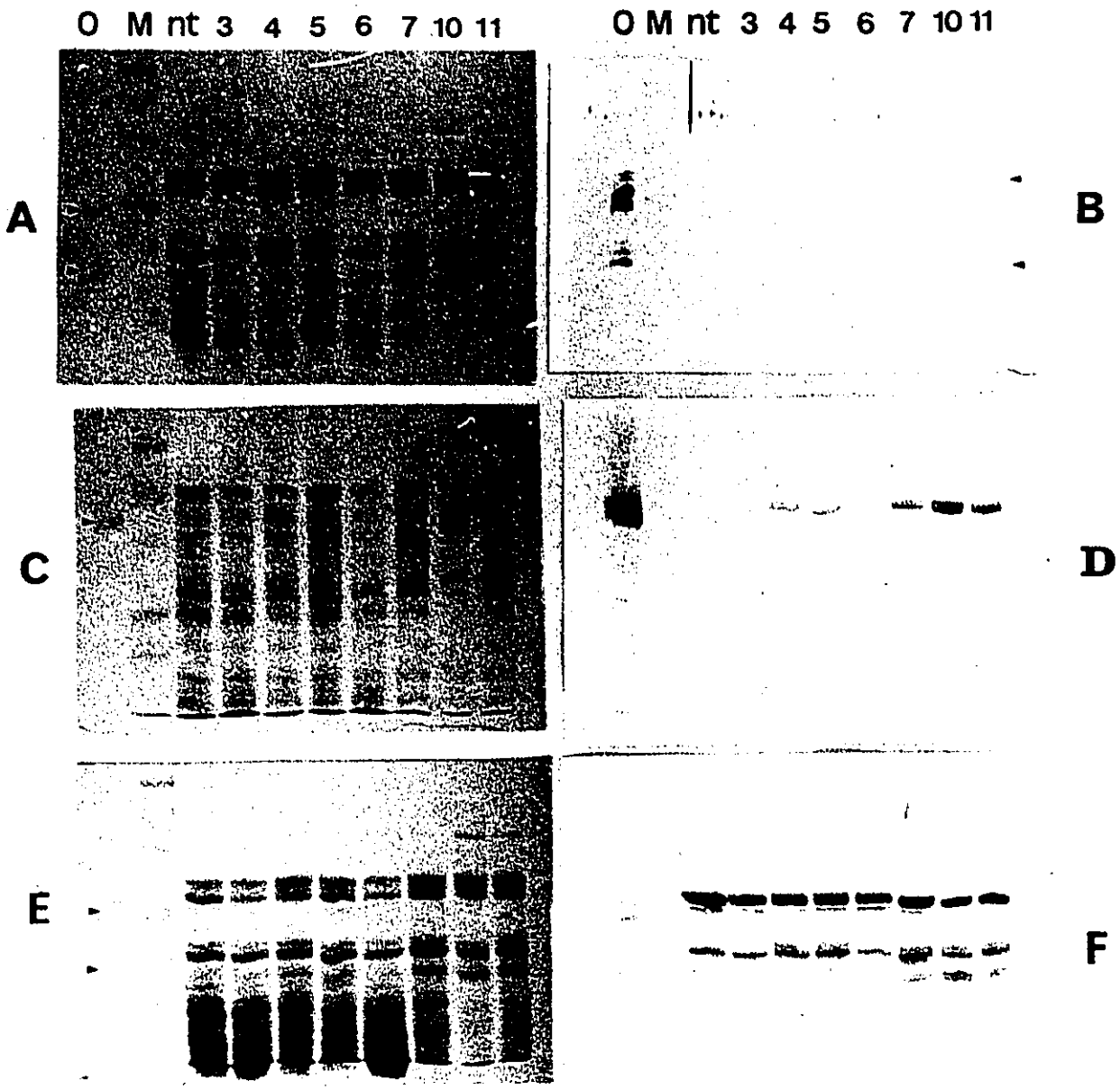

O M nt 3 4 5 6 7 10 11
O M nt 3 4 5 6 7 10 11
A
B
C
D
E
F

Figure 24 : Western blot analysis of tobacco plants transformed with the G3P construct.

A: Coomassie stained polyacrylamide gel (15%) of reduced globulins (5µg) extracted from seeds of transformed tobacco plants. The lane numbers indicate the transgenic plant numbers. The two arrowheads point to the two groups of reduced oat globulins (300ng) in lane O. The prestained molecular weight markers (lane M) are respectively from top to bottom: bovine albumin (66kDa), egg albumin (45kDa), G3P dehydrogenase (36kDa), carbonic anhydrase (29kDa), trypsinogen (24kDa), trypsin inhibitor (20.1kDa) and a lactalbumin (14.2kDa). nt: reduced tobacco seed globulins from an untransformed plant.

B: Western blot of a duplicated gel run in parallel with the gel presented in A and containing the same samples. The proteins were transferred to a nitrocellulose membrane and probed with a rabbit anti-oat globulin serum. The two arrowheads point to the weak signals obtained in lanes 3, 4, 5, 7, 10 and 11.

C: Coomassie stained polyacrylamide gel (12%) of unreduced globulins (6µg) extracted from seeds of transformed tobacco plants. Lane denominations are the same as in A. The arrowhead points to the group of unreduced oat globulins.

D: Western blot of a duplicated gel run in parallel with the gel presented in C and containing the same samples. The proteins were transferred to a nitrocellulose membrane and probed with a rabbit anti-oat globulin serum (same serum as in B).

E: Same as in A.

F: Same as in B except that a different anti-oat globulin serum was used. This serum is less specific than the one used in B and D. An extra hybridization band can be seen corresponding to the ß polypeptide in lanes 4, 5, 7, 10 and 11.

proteins (Figure 24.B), is obtained with the more recent serum (prepared by Dr. J. Schernthaner in 1990). The second rabbit serum (more than 10 years old) cross-hybridizes with most of the globulin polypeptides from tobacco (Figure 24.F). This is not very surprising considering the wide presence of globulin-like polypeptides in dicot seeds (Luthe 1992), even in gymnosperm seeds (Jensen and Berthold, 1989) and the immunological similarities that can be encountered between different seed storage proteins, both in dicots and in cereals (Robert et al., 1985a and 1985b). However, in the experimental conditions which were used, it has been possible to identify the presence of a globulin polypeptide that is not found in untransformed tobacco seeds and corresponds to the size of the oat globulins, both in reduced and unreduced conditions. Plant 6 does not show any hybridization signal although the Southern analysis testifies that the plant has been transformed. This is probably due to the position effect of the inserts in the tobacco genome.

d. Summary.

The insertion of an oat globulin genomic sequence at a favorable site in the tobacco genome results in proper spatial and temporal expression of the globulin transcript and polypeptide in tobacco seeds. This oat globulin protein is correctly processed (i.e. cleaved into an acidic and a basic polypeptide) and stably stored within the tobacco seed. However, with the type of experiments that have been conducted, it is not possible to verify the proper cleavage of the signal peptide, nor to determine the subcellular localization of the Glav-3 protein. Since the Glav-12 and Glav-3 promoters are very similar, it is very likely that the expression of the G3P construct, like the expression of the Glav-12 GUS fusion, is restricted to the endosperm of transgenic tobacco seeds.

3.1.5. Analysis of the globulin promoter.

3.1.5.1. Comparison of the different globulin promoters.

The 1000 nucleotides upstream of the ATG from Glav-1, Glav-3 and Glav-12 and the 985 nucleotides upstream of the ATG from OG1-E1 (Shotwell *et al.* 1990) were aligned by computer (Figure 25). The degree of similarity between the three Glav sequences remains very high throughout the 1000 nucleotides but the OG1-E1 sequence diverges rapidly from them (around position 470) although it is still possible to find similar stretches of nucleotides further upstream[5].

As already mentioned by Shotwell *et al.* (1990) for the OG1-E1 clone, the promoters of oat globulin genes do not show any of the conserved elements described in other seed storage protein gene promoters. These elements are the RY repeat (5' CATGCATG 3') of legume seed protein genes (Dickinson *et al.* 1988), the '-300 element' of prolamin genes (Forde *et al.* 1985), the 'legumin box' in pea-legumin genes (Baumlein *et al.* 1986) and the embryo-specific element AA/GCCCA (Chen *et al.* 1986). A possible 'CACA' element (which interacts with soybean embryo nuclear proteins) is described in OG1-E1 but is not found in any of the Glav sequences.

Since most of the regulatory sequences are thought to be in the vicinity of the transcriptional start site (around -300), a more thorough investigation of the available portions of promoter sequences has not been conducted in the more distal regions.

The finding that the globulin 5' UTR is different in oat seeds and in tobacco seeds brings the problem of identifying possible CAAT and TATA sequences upstream for the transcriptional start site used in transgenic tobacco seeds. Two possible CAAT sequences starting at positions 761 and 776 are underlined in Figure 25. A possible TATA sequence is also present at position 789, this sequence is an imperfect

[5]It is not known whether the promoter of OG1-E1 is functional or not.

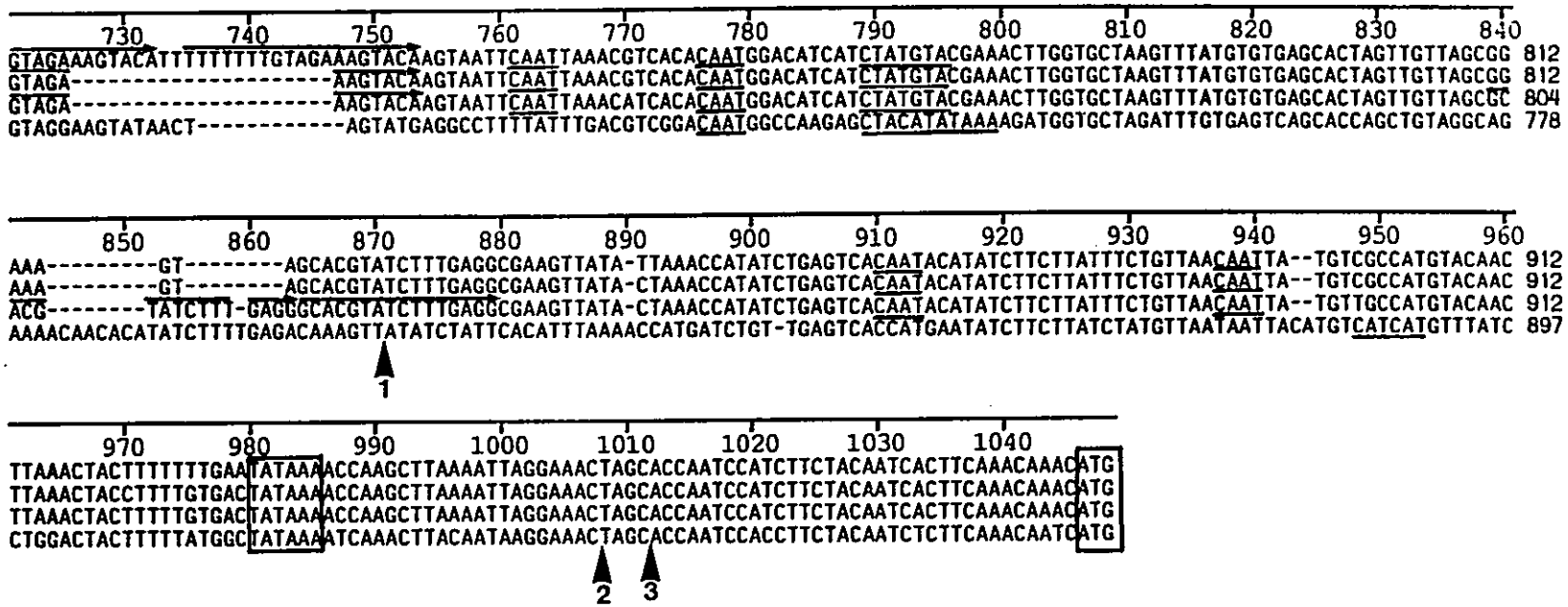
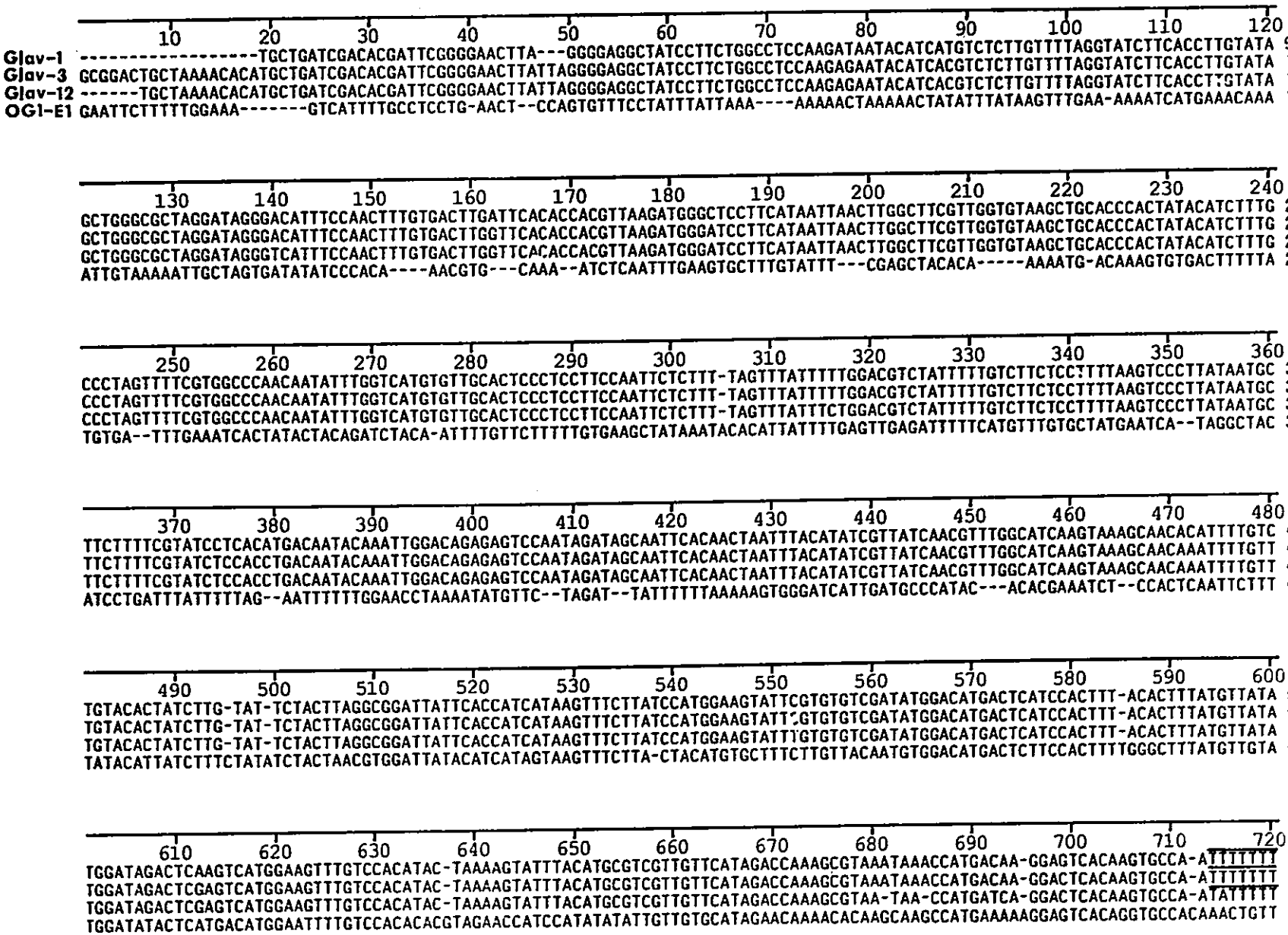


Figure 25 : Multiple alignment of oat globulin promoters.

The 1000 nucleotides (Glav-1, Glav-3 and Glav-12) and the 982 nucleotides (OG1-E1, Shotwell *et al.* 1990) upstream of the start codon (ATG) were aligned using the clustal method of the DNAstar program.

Boxed at the end of the alignment are the start codons of the 4 sequences. The possible TATA boxes (position 980) are also boxed. The possible CAAT boxes are underlined (positions 761, 776, 910, 937 and 948). Other possible TATA boxes (position 789) are underlined. The repeated sequence in Glav-1 (position 714-753) is indicated by two arrows, the corresponding sequence in Glav-3 and Glav-12 is also indicated by an arrow. A repeated sequence in Glav-12 (position 839-879) is indicated by an arrow.

Arrowhead 1 points to the transcriptional start point of Glav-12 as determined in transgenic tobacco seeds. Arrowheads 2 (for OG1-E1) and 3 (for Glav-1, Glav-3 and Glav-12) point to the transcriptional start points as determined in oat seeds.


10 20 30 40 50 60 70 80 90 100 110 120
Glav-1 ----------TGCTGATCGACACGATTCGGGGAACTTA---GGGGAGGCTATCCTTCTGGCCTCCAAGATAATACATCATGTCTCTTGTTTTAGGTATCTTCACCTTGTATA 99
Glav-3 GCGGACTGCTAAAACACATGCTGATCGACACGATTCGGGGAACTTATTAGGGGAGGCTATCCTTCTGGCCTCCAAGAGAATACATCACGTCTCTTGTTTTAGGTATCTTCACCTTGTATA 120
Glav-12 ------TGCTAAAACACATGCTGATCGACACGATTCGGGGAACTTATTAGGGGAGGCTATCCTTCTGGCCTCCAAGAGAATACATCACGTCTCTTGTTTTAGGTATCTTCACCTTGTATA 114
OGI-E1 GAATTCTTTTTGGAAA-------GTCATTTTGCCTCCTG-AACT--CCAGTGTTTCCTATTTATTAAA----AAAAACTAAAAACTATATTTATAAGTTTGAA-AAAATCATGAAACAAA 105
130 140 150 160 170 180 190 200 210 220 230 240
GCTGGGCGCTAGGATAGGGACATTTCCAACTTTGTGACTTGATTCACACCACGTTAAGATGGGCTCCTTCATAATTAACTTGGCTTCGTTGGTGTAAGCTGCACCCACTATACATCTTTG 219
GCTGGGCGCTAGGATAGGGACATTTCCAACTTTGTGACTTGGTTCACACCACGTTAAGATGGGATCCTTCATAATTAACTTGGCTTCGTTGGTGTAAGCTGCACCCACTATACATCTTTG 240
GCTGGGCGCTAGGATAGGGTCATTTCCAACTTTGTGACTTGGTTCACACCACGTTAAGATGGGATCCTTCATAATTAACTTGGCTTCGTTGGTGTAAGCTGCACCCACTATACATCTTTG 234
ATTGTAAAAATTGCTAGTGATATATCCCACA----AACGTG---CAAA--ATCTCAATTTGAAGTGCTTTGTATTT---CGAGCTACACA-----AAAATG-ACAAAGTGTGACTTTTTA 207
250 260 270 280 290 300 310 320 330 340 350 360
CCCTAGTTTTCGTGGCCCAACAATATTTGGTCATGTGTTGCACTCCCTCCTTCCAATTCTCTTT-TAGTTTATTTTTGGACGTCTATTTTTGTCTTCTCCTTTTAAGTCCCTTATAATGC 338
CCCTAGTTTTCGTGGCCCAACAATATTTGGTCATGTGTTGCACTCCCTCCTTCCAATTCTCTTT-TAGTTTATTTTTGGACGTCTATTTTTGTCTTCTCCTTTTAAGTCCCTTATAATGC 359
CCCTAGTTTTCGTGGCCCAACAATATTTGGTCATGTGTTGCACTCCCTCCTTCCAATTCTCTTT-TAGTTTATTTCTGGACGTCTATTTTTGTCTTCTCCTTTTAAGTCCCTTATAATGC 353
TGTGA--TTTGAAATCACTATACTACAGATCTACA-ATTTTGTTCTTTTTGTGAAGCTATAAATACACATTATTTTGAGTTGAGATTTTTCATGTTTGTGCTATGAATCA--TAGGCTAC 322
370 380 390 400 410 420 430 440 450 460 470 480
TTCTTTTCGTATCCTCACATGACAATACAAATTGGACAGAGAGTCCAATAGATAGCAATTCACAACTAATTTACATATCGTTATCAACGTTTGGCATCAAGTAAAGCAACACATTTTGTC 458
TTCTTTTCGTATCTCCACCTGACAATACAAATTGGACAGAGAGTCCAATAGATAGCAATTCACAACTAATTTACATATCGTTATCAACGTTTGGCATCAAGTAAAGCAACAAATTTTGTT 479
TTCTTTTCGTATCTCCACCTGACAATACAAATTGGACAGAGAGTCCAATAGATAGCAATTCACAACTAATTTACATATCGTTATCAACGTTTGGCATCAAGTAAAGCAACAAATTTTGTT 473
ATCCTGATTTATTTTTAG--AATTTTTTGGAACCTAAAATATGTTC--TAGAT--TATTTTTTAAAAAGTGGGATCATTGATGCCCATAC---ACACGAAATCT--CCACTCAATTCTTT 431
490 500 510 520 530 540 550 560 570 580 590 600
TGTACACTATCTTG-TAT-TCTACTTAGGCGGATTATTCACCATCATAAGTTTCTTATCCATGGAAGTATTCGTGTGTCGATATGGACATGACTCATCCACTTT-ACACTTTATGTTATA 575
TGTACACTATCTTG-TAT-TCTACTTAGGCGGATTATTCACCATCATAAGTTTCTTATCCATGGAAGTATTCGTGTGTCGATATGGACATGACTCATCCACTTT-ACACTTTATGTTATA 596
TGTACACTATCTTG-TAT-TCTACTTAGGCGGATTATTCACCATCATAAGTTTCTTATCCATGGAAGTATTTGTGTGTCGATATGGACATGACTCATCCACTTT-ACACTTTATGTTATA 590
TATACATTATCTTTCTATATCTACTAACGTGGATTATACATCATAGTAAGTTTCTTA-CTACATGTGCTTTCTTGTTACAATGTGGACATGACTCTTCCACTTTTGGGCTTTATGTTGTA 550
610 620 630 640 650 660 670 680 690 700 710 720
TGGATAGACTCAAGTCATGGAAGTTTGTCCACATAC-TAAAAGTATTTACATGCGTCGTTGTTCATAGACCAAAGCGTAAATAAACCATGACAA-GGAGTCACAAGTGCCA-ATTTTTTT 692
TGGATAGACTCGAGTCATGGAAGTTTGTCCACATAC-TAAAAGTATTTACATGCGTCGTTGTTCATAGACCAAAGCGTAAATAAACCATGACAA-GGACTCACAAGTGCCA-ATTTTTTT 713
TGGATAGACTCGAGTCATGGAAGTTTGTCCACATAC-TAAAAGTATTTACATGCGTCGTTGTTCATAGACCAAAGCGTAA-TAA-CCATGATCA-GGACTCACAAGTGCCA-ATATTTTT 705
TGGATATACTCATGACATGGAATTTTGTCCACACACGTAGAACCATCCATATATATTGTTGTGCATAGAACAAAACACAAGCAAGCCATGAAAAAGGAGTCACAGGTGCCACAAACTGTT 670

match of the consensus sequence of the TATA box as described by Forde *et al.* (1985) and presented below.

```
Globulin:      CTATGTACG
               '''' ''
Consensus:     CTATATATA
                   A A
```

The OG1-E1 sequence possesses only one possible CAAT box but the TATA box more closely matches the TATA consensus sequence. An interesting point is that all these sequences are aligned and present in a relatively well conserved region of the different promoters.

Although a complete prolamin box could not be identified, an imperfect central element (TGTAGAAAGT) can be found in the Glav promoters. A comparison of the prolamin box and the sequence found in the globulin promoters is shown here (W=G or T):

```
Prolamin box:   ANNTGT..AAAGWWAATNNGATGAWWCATG
                 ''''' ''' ' ' '' ''
Glav sequence:  TTTTGTAGAAAGTACAAGTAAT
```

This central element is present only once in Glav-3 and Glav-12, but is repeated once in Glav-1 (between positions 713 and 753 in Figure 25). Because this DNA element is present around position -250 and because of its partial similarity to the prolamin box, experiments were performed to test whether that portion of the promoter played a role in the tissue-specific expression of oat globulin genes and whether the doubling of this sequence is responsible for the non-functionality of the Glav-1 promoter. For that purpose, precise deletions of the Glav-3 promoter were performed[6].

[6]The design of these precise deletions was done before knowing that the transcriptional start site in tobacco seeds was different than the one in oat seeds.

3.1.5.2. Obtaining of fine deletions of the upstream region of Glav-3.

Using the exonuclease III method, three deletions of the Glav-3 promoter were obtained and fused to the GUS sequence. These three deletions correspond to positions -259, -247 and -237 of the transcriptional start point of the globulin promoter in oat seeds. The G3Δ-259 occurs two nucleotides before the beginning of the DNA sequence that is repeated in the Glav-1 promoter. The G3Δ-247 occurs in the middle of this sequence and the G3Δ-237 starts one nucleotide after the end of the DNA stretch (see Figure 6).

3.1.5.3. Tobacco transformations with the deleted promoters fused to the GUS coding region.

The three constructs (G3Δ-259-GUS, G3Δ-247-GUS and G3Δ-237-GUS) were used to transform tobacco plants and about twelve plants per construct were regenerated.

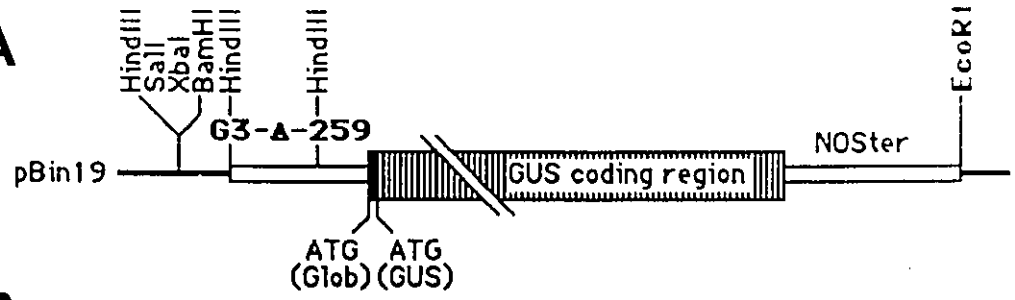
a. Southern analysis of the regenerated plants.

Only one set of plants corresponding to the longer promoter (G3Δ-259-GUS) was analysed by Southern blots. Figure 26 shows that all the plants tested were transformed. The copy number varied from one (plant 3, 7, 8 and 10) to about 5 (plant 1). It is assumed that the plants transformed with the two other constructs are also carrying the inserts since in all the transformation experiments conducted so far, no escapees were found.

b. GUS activities in mature transformed seeds.

Self-pollinated mature seeds from the three sets of transformed plants were collected and spectrofluorimetrically assayed for GUS activity. The results are presented in Figure 27. Although the GUS

A


HindIII
SalI
XbaI
BamHI
HindIII
HindIII
EcoRI
G3-A-259
pBin19
GUS coding region
NOSter
ATG (Glob)
ATG (GUS)

B

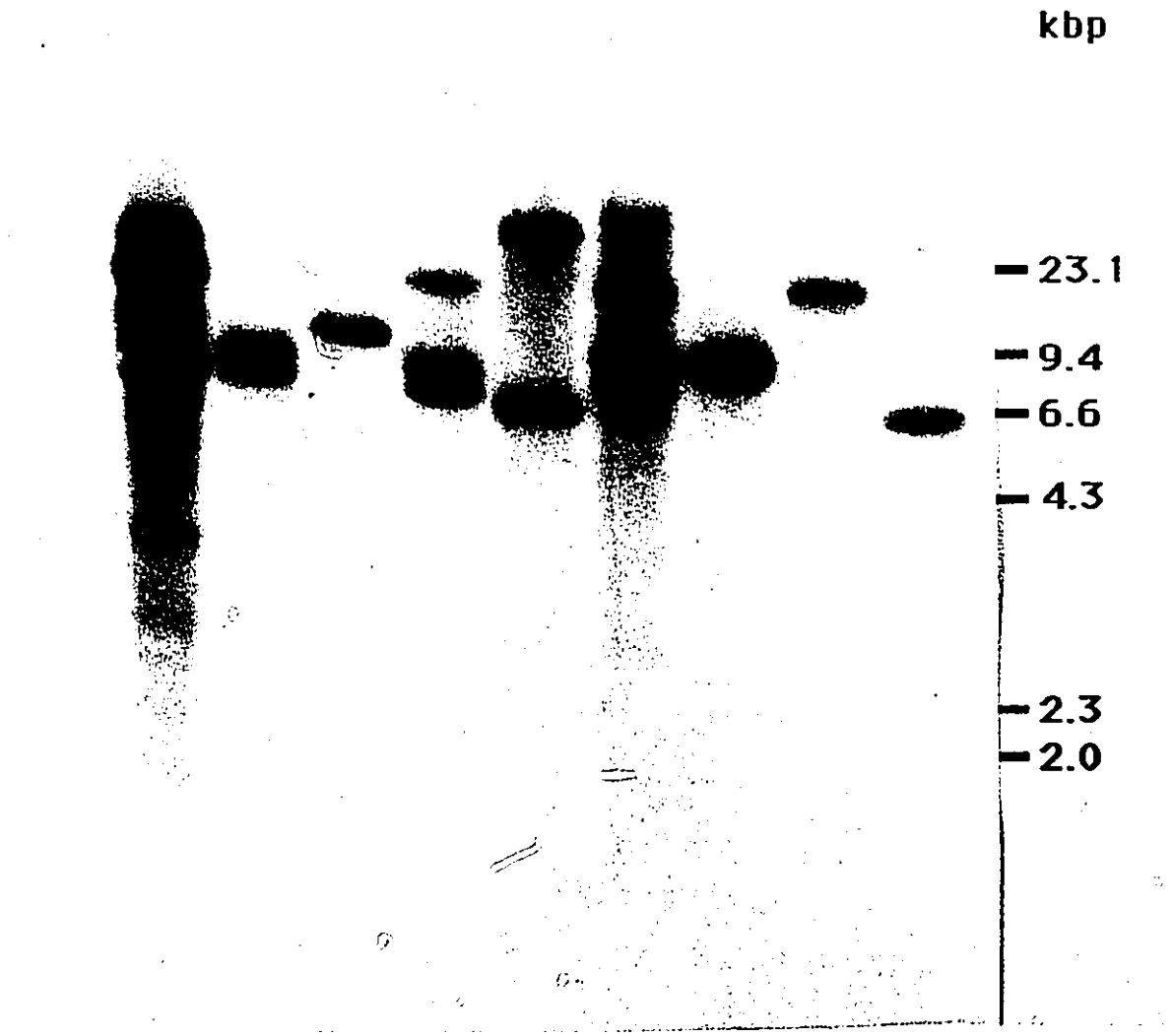

NT 1 2 3 4 5 6 7 8 10
kbp
23.1
9.4
6.6
4.3
2.3
2.0

Figure 26 : Southern blot analysis of tobacco plants transformed with the G3Δ-259-GUS construct.

A: Construct used for the tobacco transformations.

B: Southern blot analysis of regenerated tobacco plants. DNA (5µg) was digested by EcoRI and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose membrane and hybridized to a random-labeled GUS probe. The lane numbers refer to the number of the plants. NT indicates an untransformed plant. The size markers are from lambda DNA cut by HindIII.

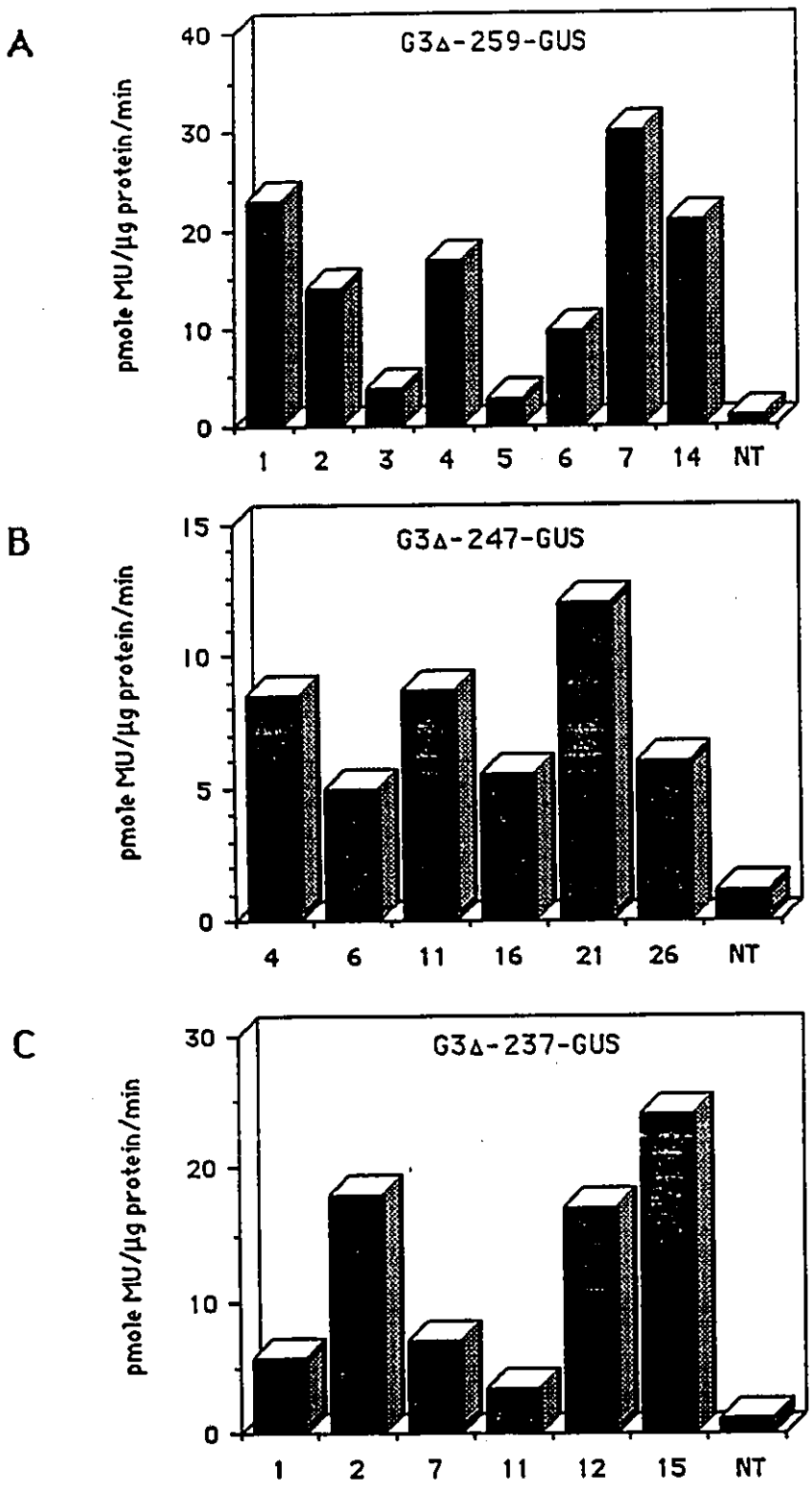


Figure27: GUS activities in mature seeds:
A: Plants transformed with the G3Δ-259-GUS construct
B: Plants transformed with the G3Δ-247-GUS construct
C: Plants transformed with the G3Δ-237-GUS construct
The numbers represent the different plants from which the seeds were collected. NT: not transformed.

activities obtained with the transformed seeds are higher than the GUS activity obtained with the untransformed seeds, there are no consistent differences between the three sets of plants. Another observation is that the values (pmoles) are about 1000-fold lower than those obtained with the G12-GUS construct (nmoles). This very low level of GUS activity is probably due to some residual level of promoter expression. The variations between the plants could be attributed to the position effect of the insertions.

c. Summary.

Oat globulin promoter sequences do not share any conserved elements (other than the CAAT and TATA boxes) with other seed-specific storage protein gene promoters. If we assume that a *cis*-acting element is present in the promoter of oat globulin genes, it is probably lying close upstream of the region chosen for promoter deletion analysis since the nucleotide similarity between globulin promoters from the two different oat varieties rapidly fades away further upstream. However, two members of an abundant gene (≈ 300 copies in total) family have been shown to be functional and regulated, both developmentally and spatially, in similar ways in tobacco seeds and in oat seeds.

3.1.6. Globulin genes in different oat species.

The DNA from oat species representing different levels of ploidy was restricted with EcoRI, HindIII and BamHI. After electrophoresis, the restriction fragments were hybridized to a random labeled globulin probe (see 2.20.2.). Figure 28 shows the autoradiograms obtained. As mentioned by Chesnut *et al.* (1989), the HindIII restriction fragments that hybridize to globulin sequences are well conserved among the oat species (Figure 28.B). However, when oat genomic DNA is restricted with EcoRI or BamHI, a larger degree of polymorphism is observed (Figure 28 A and C).

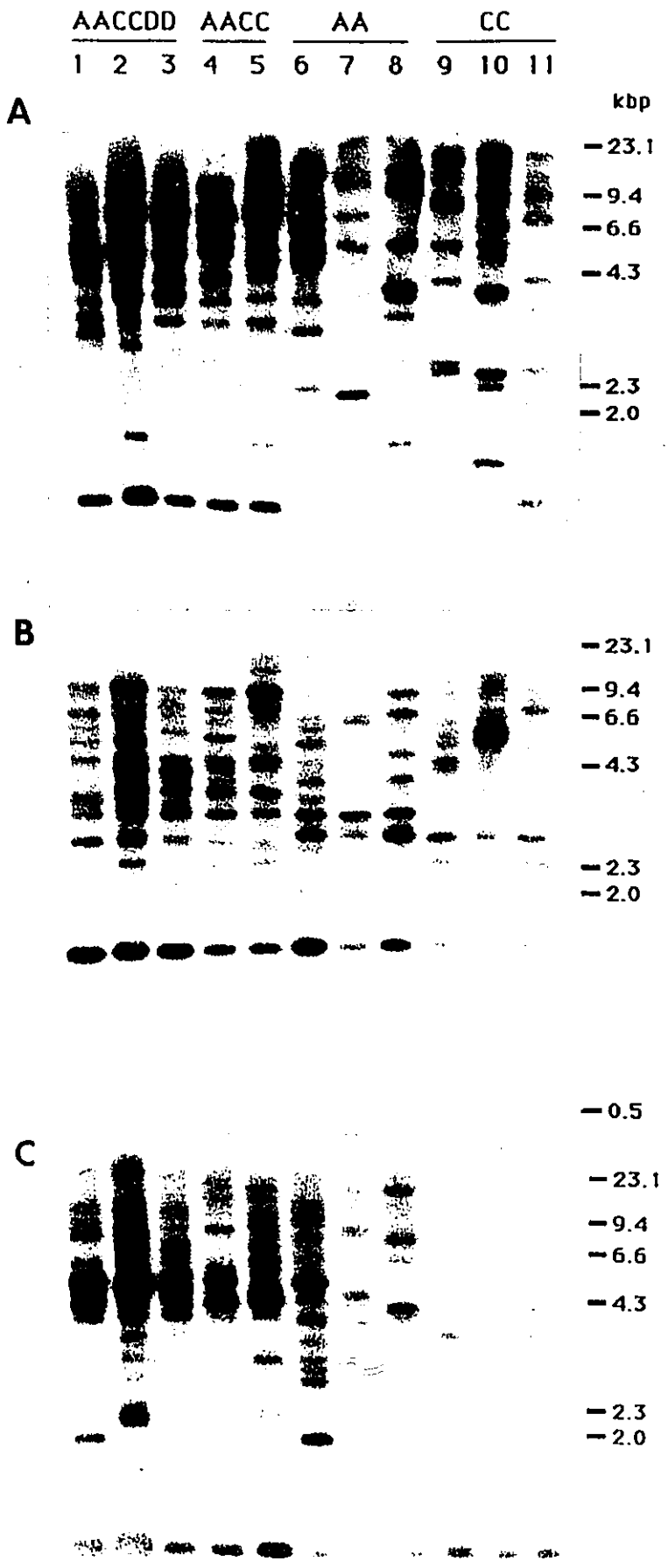

AACCDD
AACC
AA
CC
1 2 3 4 5 6 7 8 9 10 11
A
kbp
23.1
9.4
6.6
4.3
2.3
2.0
B
23.1
9.4
6.6
4.3
2.3
2.0
C
0.5
23.1
9.4
6.6
4.3
2.3
2.0

Figure 28 : Analysis of globulin genes in different oat species by Southern hybridization.

DNA (5µg) from oat species representing different levels of ploidy was digested with EcoRI (**A**), HindIII (**B**) or BamHI (**C**) and electrophoresed in a 0.8% agarose gel. The gels were blotted on nitrocellulose membranes and hybridized to a random-labeled globulin probe.

The hexaploid oat species (genome AACCDD) are *A. sativa* cv. Hinoat (lane 1), *A. sativa* cv. Donald (lane 2) and *A. sativa* cv. Ogle (lane 3).
The tetraploid oat species (genome AACC) are *A. magna* (lane 4) and *A. murphyi* (lane 5).
The diploid oat species (genome AA) are *A. longiglumis* (lane 6), *A. prostrata* (lane 7) and *A. wiestii* (lane 8). The genome CC is represented by *A. clauda* (lane 9), *A. ventricosa* (lane 10) and *A. pilosa* (lane 11).

The molecular weight markers are from lambda DNA digested with HindIII.

In the diploid genomes (CC), we can see that *A. clauda* (lane 9) and *A. pilosa* (lane 11) are more closely related to each other than to *A. ventricosa* (lane 10). This confirms their genome denominations: C_VC_V for *A. ventricosa* and C_PC_P for *A. pilosa* and *A. clauda*. As for the AA genomes (lanes 6, 7 and 8), they have been assigned different denominations (A_LA_L for *A. longiglumis*, A_PA_P for *A. prostrata* and A_WA_W for *A. wiestii*), although the globulin hybridization patterns are very similar when the DNA is restricted with HindIII. The differences become notable when the DNA is restricted with EcoRI, and to a lesser extent with BamHI.

In the tetraploid genomes (AACC), the two species that have been analysed (*A. magna* and *A. murphyi*) share remarkable similarities (lanes 4 and 5 in panels A, B and C). It is difficult, however, to determine the respective progenitors of their AA and CC genomes, especially since not all the different possible sources of the diploid genomes were analysed in this study.

The *A. sativa* cultivars (Hinoat, Donald and Ogle), with a AACCDD genome, are very similar on the basis of their hybridization patterns (lanes 1, 2 and 3). These hexaploid cultivars are probably derived from a AACC tetraploid species (Chesnut *et al.* 1989), but to determine which species is their progenitor would require more analyses.

The globulin gene copy number has been estimated to be about 50 copies per haploid genome (Chesnut *et al.* 1989) in *A. sativa*. Judging by the intensities of the hybridization signals, it seems that duplication of globulin genes also occurred after the fusions of the different genomes which produced the hexaploid species. Out of these estimated 50 copies, there are also probably a certain number of pseudogenes.

3.2. Oat prolamin genes.

3.2.1. Sequence of a prolamin gene.

Two avenin genomic sequences were isolated and sequenced by Dr. M. Giband. These avenin genes were obtained by PCR using oligonucleotide primers designed from the genomic clone Av45-X1 (Shotwell *et al.* 1990). The sequence of one of the genomic clones is shown in Figure 29.

The coding region of the avenin gene and its structure are very similar to previously described oat coding sequences (Shotwell *et al.* 1990, Chesnut *et al.* 1989, Pernollet *et al.* 1987 and 1989, Egorov 1988): a 19-amino-acid signal peptide, a NH_2-terminal region, a first region of repeats, a conserved central region, a second region of repeats and the COOH-terminal region.

The start of transcription of the avenin mRNA was mapped by primer extension analysis at the A residue 71 nucleotides upstream of the ATG (Figure 35.A). The 5' end of the Av45-X1 has been mapped to the T residue 73 bp upstream of the start codon (Shotwell *et al.* 1990). A possible TATA box begins 34 nucleotides upstream of the translational start site. A possible CAAT box is also present, starting 31 nucleotides upstream of the possible TATA box. A prolamin '-300 element' (Forde *et al.* 1985) is located 261 bp upstream of the transcriptional start point. At the 3' end, we can find two possible polyadenylation signal sequences (the first one as a doublet) positioned 75 and 138 nt dowstream of the stop codon.

3.2.2. Promoter comparison of some prolamin-type gene promoters.

The complete genomic sequence of the avenin gene was used to find similar sequences in EMBL+GenBank Release 73. The sub-database thus created was screened again only with the promoter of the avenin gene (about 400 bp). The different sequences generated by

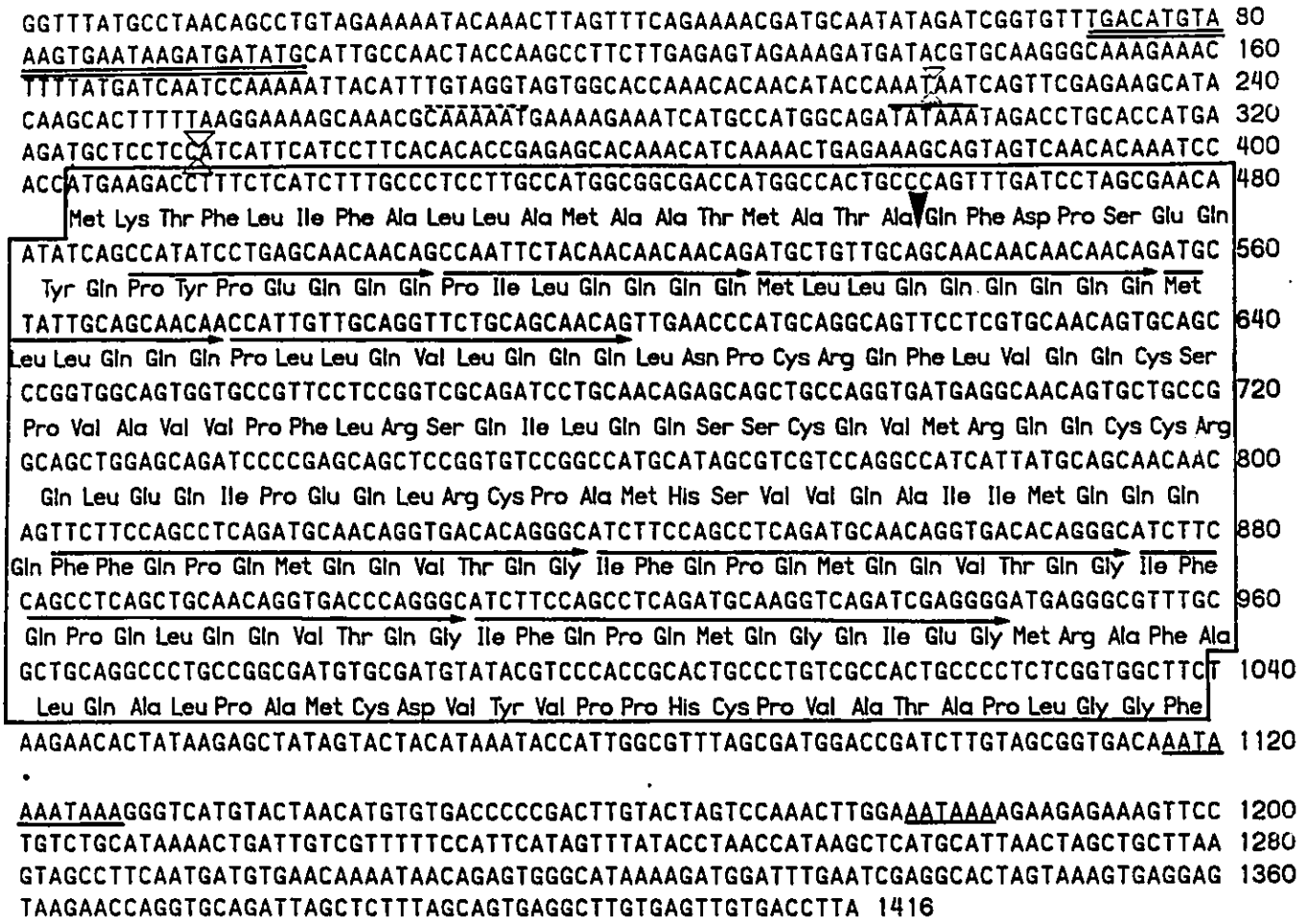


Figure 29 : Nucleotide and deduced amino acid sequences of an avenin gene (isolated and sequenced by Dr. M. Giband).

Promoter: The start of transcription, as determined by primer extension analysis on oat seed mRNA, is indicated by the two big open arrowheads. The start of transcription in tobacco seeds is indicated by the two small open arrowheads. The possible TATA box is overlined. A possible CAAT box is overlined by a dotted line. The prolamin box is double underlined.

Coding sequence: The coding nucleotide sequence and the deduced amino acid sequence are boxed. The solid arrowhead points to the cleavage site of the signal peptide. The arrows overline two different series of peptide repeats.

3' end: The potential polyadenylation sequences are underlined.

this search were aligned (only the 400 bp upstream of the start codon) and are presented in Figure 30. Some duplicated sequences were omitted. It appears that the promoter sequences from storage protein genes of different cereal species share a high degree of similarity. This is particularly visible with the TATA box and the '-300 element', which was identified by alignment of only a few promoter sequences (Forde *et al.* 1985). It is interesting to see that, as more genomic sequences are available, this prolamin box can be identified in a wide variety of storage protein genes that are related to the prolamin gene family, thus confirming the potentially important role played by this element in the regulation of gene expression. For this reason, it was decided to fuse the promoter region of the avenin gene to the GUS coding sequence.

3.2.3. Tobacco transformation with the avenin promoter fused to the GUS coding region.

The complete available promoter sequence of the avenin gene (406 bp to the ATG start codon) was translationally fused to the GUS coding region, transferred into pBin19 and used for tobacco transformation via *A. tumefaciens*. As for the G12-GUS fusion, the making of the construct resulted in the presence of two ATGs in frame. Shown below is the beginning of the coding region of the Av-GUS construct:

ATG AGA TTC CCG GGT GGT CAG TCC CTT **ATG** ...

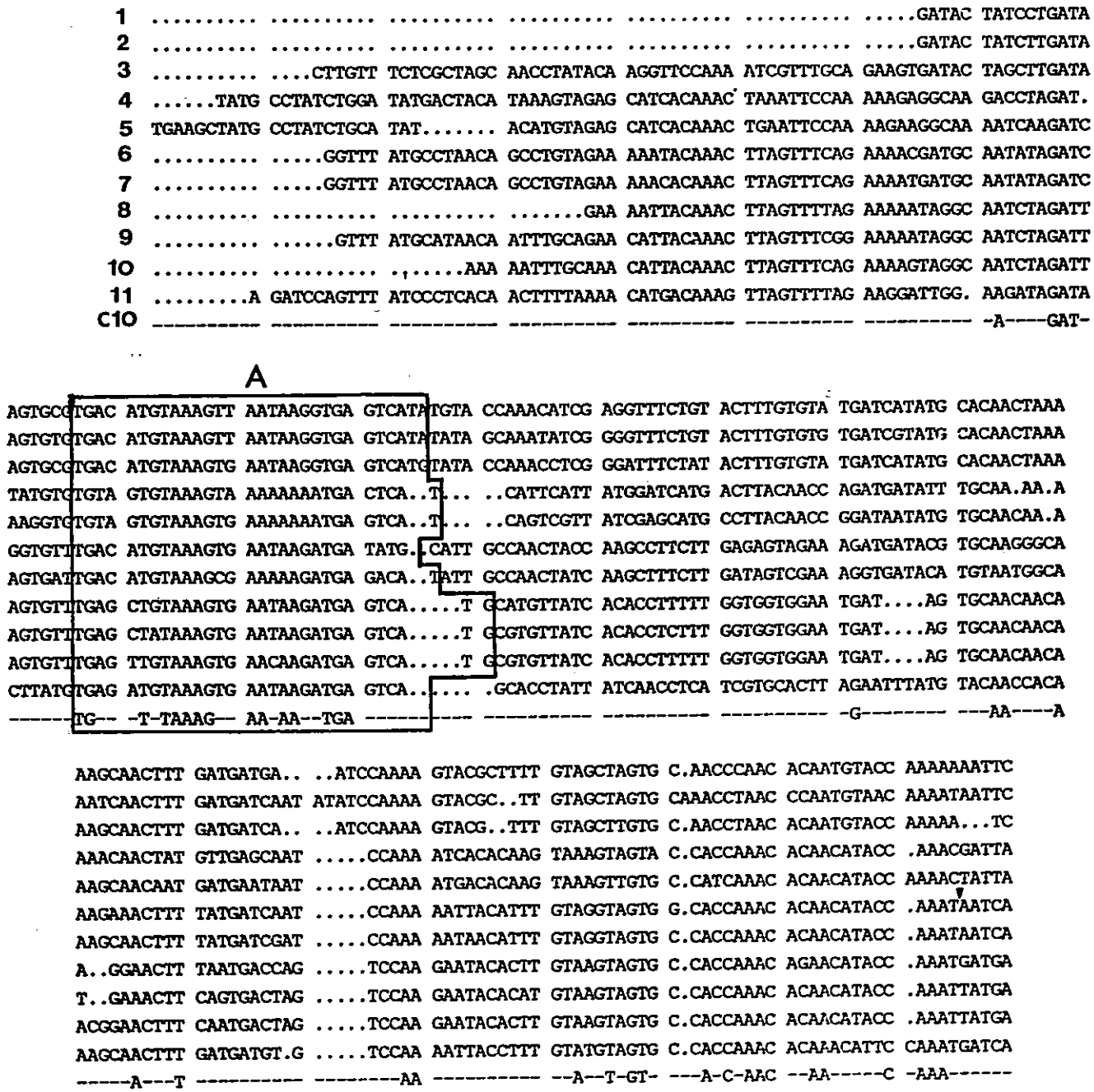
The first ATG corresponds to the avenin start codon and the second ATG to the GUS coding region. Nucleotides in between are from the pBI101.1 multiple cloning site.

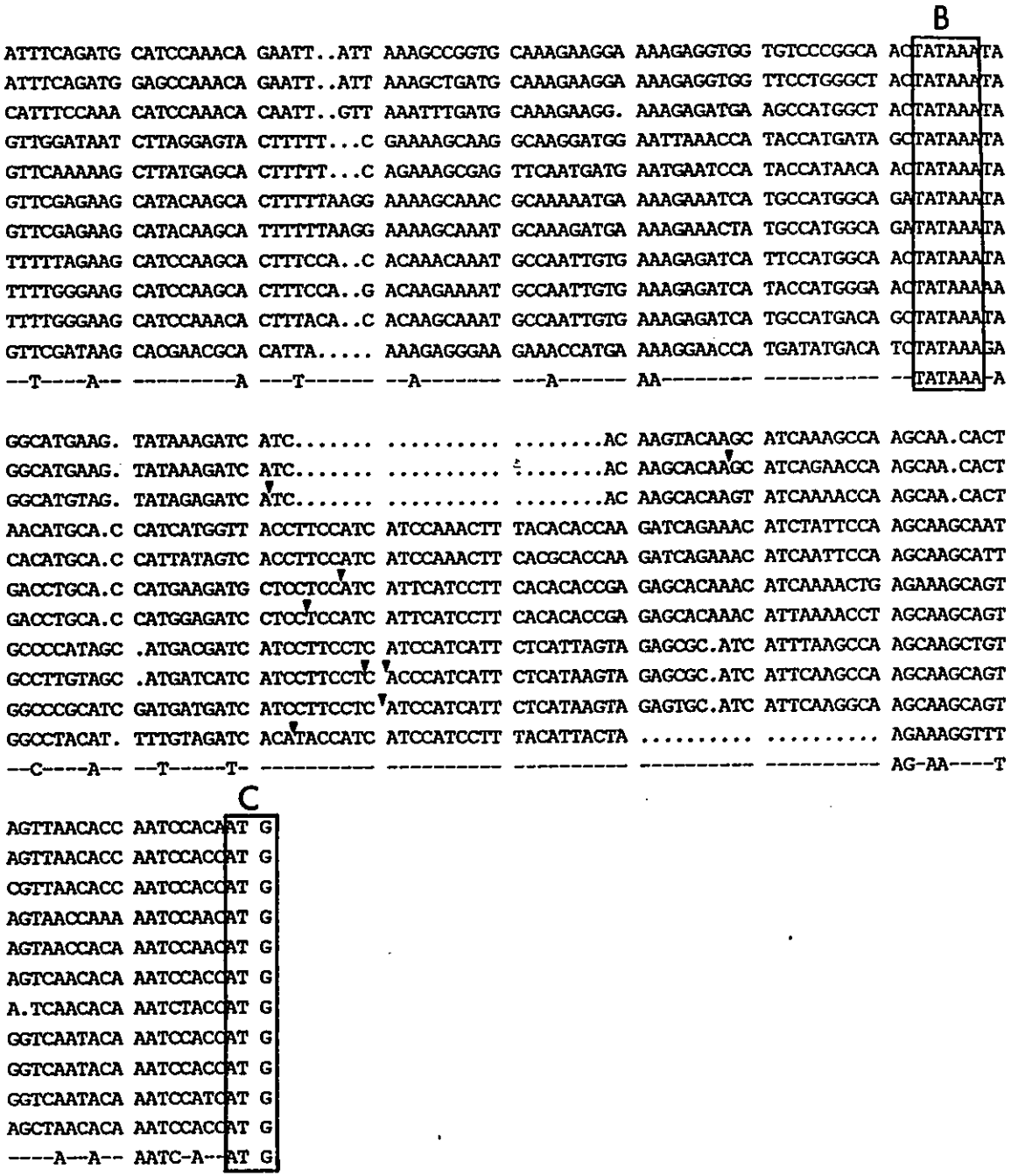
Figure 30 : Multiple alignment of promoter sequences of different storage protein genes.

The avenin genomic sequence was used to screen the EMBL+GenBank Release 73 database. The resulting sub-database was then searched for sequence similarities with the avenin promoter. A multiple alignment was performed on sequences representing a variety of different plants and storage proteins. Approximately 400 nucleotides upstream of the ATG were selected from each sequence.

1-WHTGLUT: wheat LMW glutenin. Pitts *et al.* 1988.
2-WHTGLU1D1: wheat LMW glutenin. Colot *et al.* 1989.
3-BLYHORB: barley B1 hordein. Forde *et al.* 1985.
4-BLYHORDCA: barley C-hordein. Entwistle, 1988.
5-RYESEC1B: rye omega secalin. Hull *et al.* 1991.
6-AVENIN: oat avenin (from this study).
7-ASTSSPA: oat avenin. Shotwell *et al.* 1990.
8-WHTGLIC: wheat gliadin. Rafalski *et al.* 1984.
9-WHTGLNA: wheat alpha/beta-type gliadin. Reeves and Okita, 1987.
10-WHTGLIABD: wheat alpha/beta gliadin. Sumner-Smith *et al.* 1983.
11-BLYG1HORDA: barley gamma hordein. Cameron-Mills and Brandt, 1988.
C10-Consensus: a similar nt which appears in 10 out of the 11 sequences is reported.

The prolamin box (-300 element) is boxed (A). The possible TATA box is boxed (B). The translational start codon (ATG) is boxed at the end of the sequences (C). Arrowheads point to the transcriptional start sites, when available. Multiple possible CAAT boxes can be found in each sequence and are therefore not indicated. The second most upstream arrowhead in sequence 6 points to the transcriptional start site in transgenic tobacco seeds.


A


B
C

a. Southern analysis of the regenerated plants.

Sixteen plants were regenerated, the DNA was extracted from young leaves and analysed by Southern hybridization using the GUS coding sequence as a probe. The results are shown in Figure 31. All the plants were transformed (Figure 31.B) and the number of inserted copies varies from 1 (plants A14 and A17) to more than 6 (plant A5). Figure 31.C shows that most insertions are not rearranged and an expected 2.6 kb fragment is observed, but a relatively high proportion of the plants (8 out of 16) exhibits rearrangements of the construct. It was then verified that the hybridization fragments were not due to partial digests of the tobacco DNA. For that purpose, different amounts of EcoRI enzymes were used to digest total DNA from plants A1, A3, A4 and A18. Figure 34 shows that the extra hybridization band(s) are not due to partial digests since the intensities of these signals do not decrease when more enzymes are used.

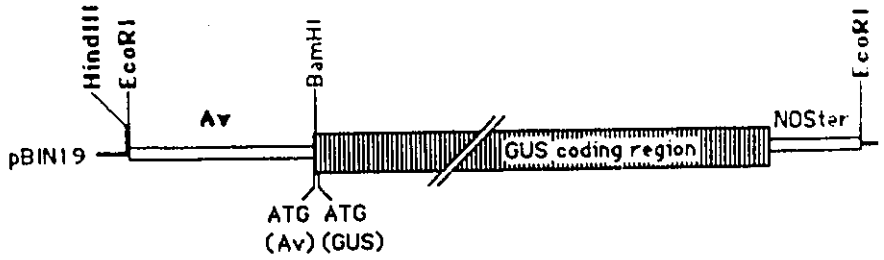
b. Northern analysis of some transformed plants.

Attempts to detect the GUS transcripts at different stages of seed maturation, and also from leaf, stem and root were unsuccessful. The experiments were conducted together with the analysis of the plants transformed with the G12-GUS construct (see 3.1.3.2.b).

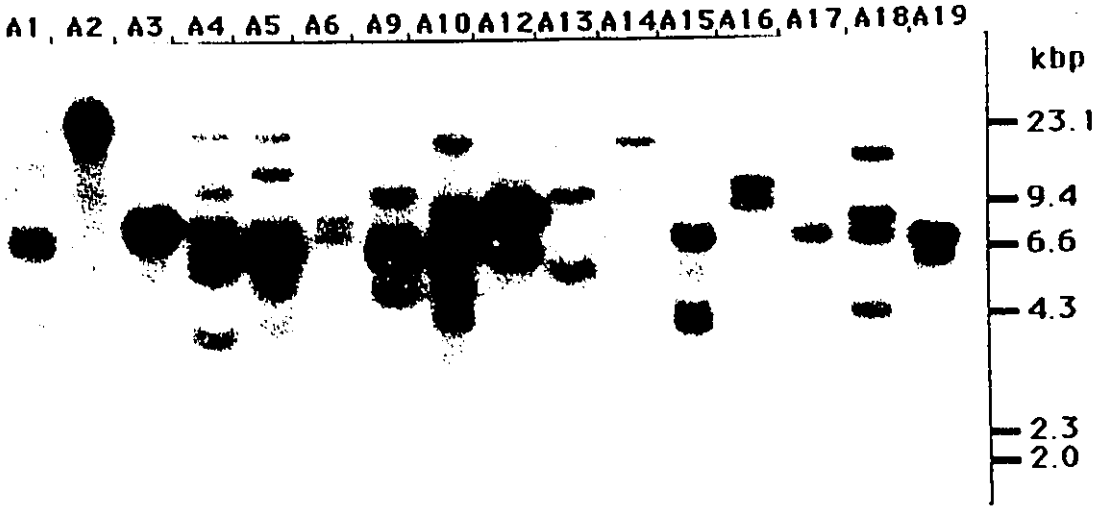
c. Histochemical analysis of transformed tobacco seed pods.

Hand-cut cross-sections of transformed tobacco seed pods were analysed by histochemical staining. The presence of the GUS enzymes was detected by the production of a blue precipitate. As was the case with the globulin promoter fused to GUS, the GUS activity is restricted to the endosperm of transformed seeds (Figure 32 A, B and C). The blue precipitate was visible very rapidly after the addition of the substrate (about half an hour) and did not appear in any other seed-pod tissues. However, if incubation time was prolonged, a blue

A


HindIII
EcoRI
BamHI
EcoRI
Av
NOSter
pBIN19
GUS coding region
ATG ATG
(Av) (GUS)

B


NT, A1, A2, A3, A4, A5, A6, A9, A10, A12, A13, A14, A15, A16, A17, A18, A19
kbp
23.1
9.4
6.6
4.3
2.3
2.0

C

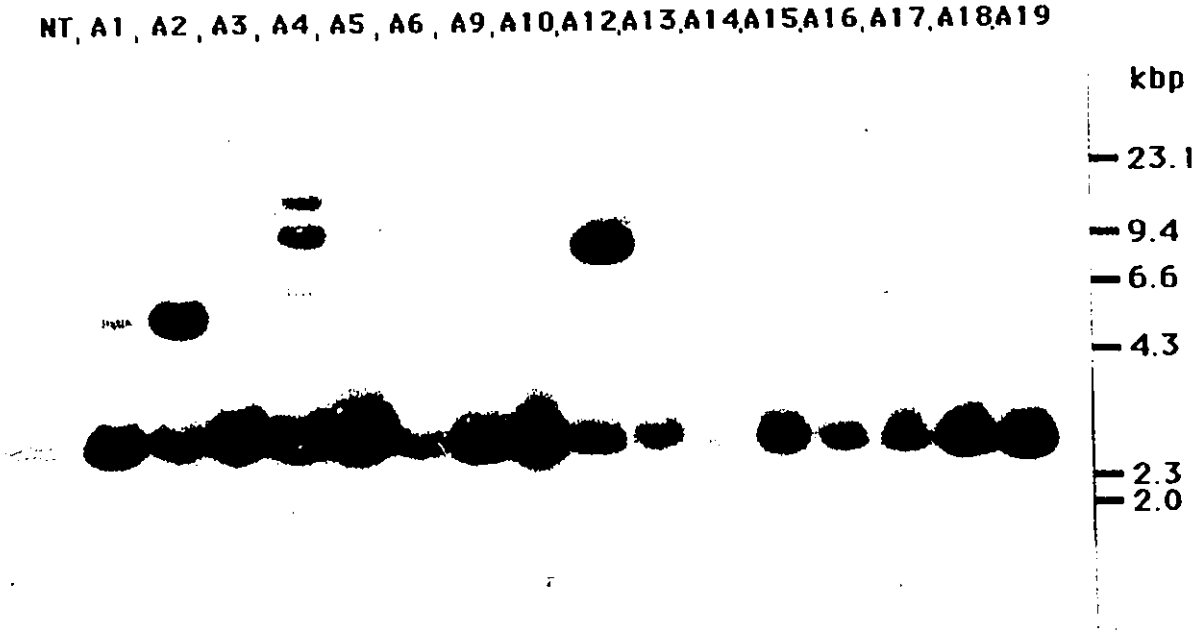

NT, A1, A2, A3, A4, A5, A6, A9, A10, A12, A13, A14, A15, A16, A17, A18, A19
kbp
23.1
9.4
6.6
4.3
2.3
2.0

Figure 31 : Southern blot analysis of tobacco plants transformed with the Av-GUS construct.

A: Construct Av-GUS used for the tobacco transformations.

B: Southern blot analysis of the regenerated tobacco plants. DNA (5µg) was digested by HindIII and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose membrane and hybridized to a random labeled GUS probe. The lane numbers refer to the number of the plants. NT indicates an untransformed plant. The size markers are from lambda DNA cut by HindIII.

C: Same as **B** except that the DNA was digested with EcoRI.

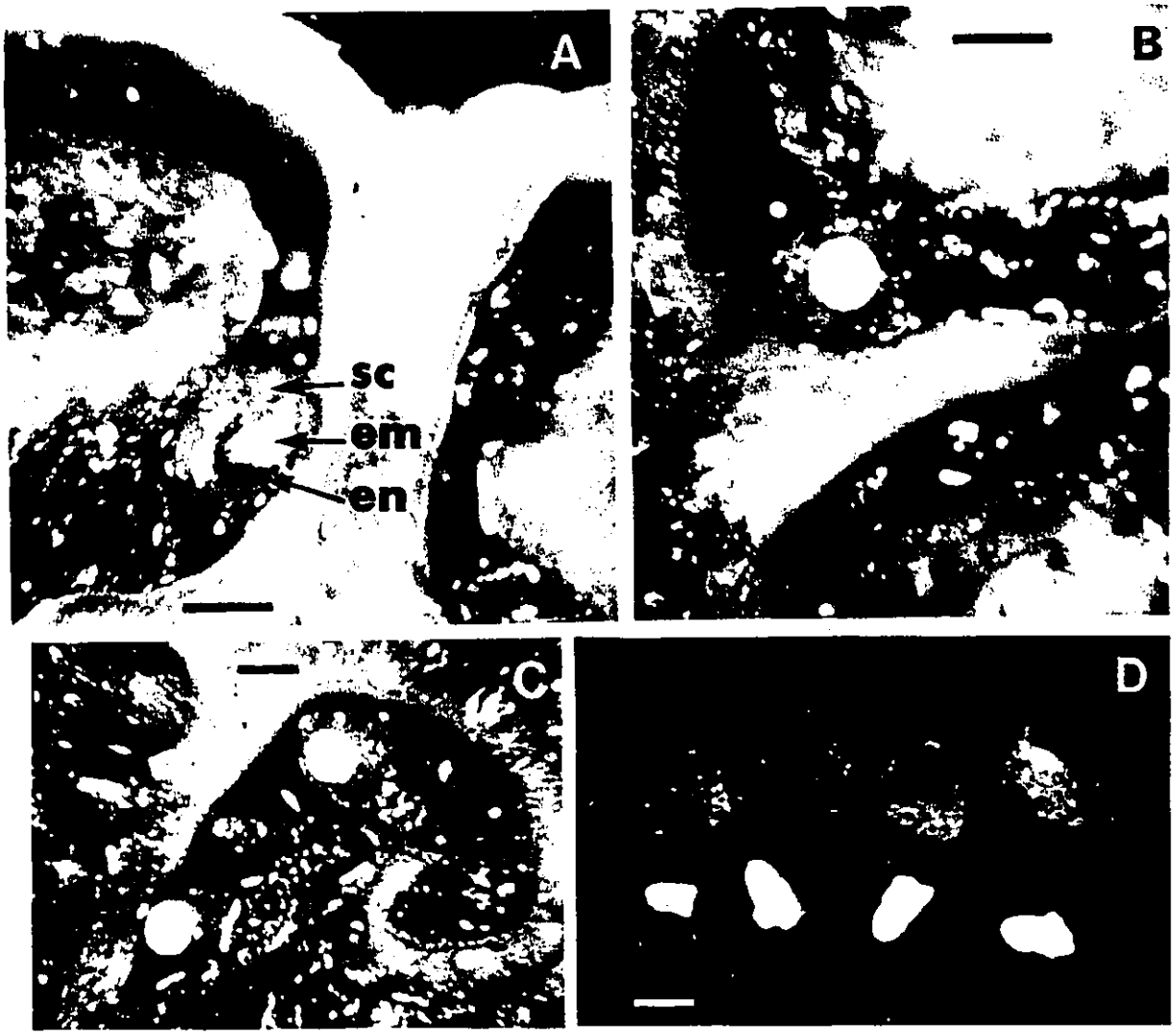

A
sc
em
en
B
C
D

Figure 32 : Histochemical GUS staining of cross-sections of seed pods from tobacco plants transformed with the Av-GUS construct.

A and B: Cross-sections of tobacco seed pods from plant A15. The blue staining, indicating the presence of the GUS protein, is restricted to the endosperm (en) of the seed. The maternal tissues, as well as the embryo (em) and the seed coat (sc), do not show any GUS activities.

C: Cross-section of a tobacco seed pod from plant A10

For **A, B** and **C**, the staining was performed for one hour.

D: Embryos from tobacco seeds (plant A13) were dissected and separated from the endosperm and seed coat. The staining reaction was done separately for each part of the seed for three hours at 37°C. Only the endosperm shows GUS activity. Each vertical set of embryo/endosperm-seed coat comes from the same seed. The set at the right of the photograph comes from an untransformed plant.

Bar represents 0.2mm.

precipitate formed in the embryo. To verify that the GUS activity was endosperm specific, embryos were dissected from the rest of the seeds and incubated separately in the GUS staining buffer. Figure 32.D shows that the GUS protein is absent from the embryo. A possible reason for the staining of the embryo when left to incubate for a longer period of time is diffusion of the primary product of the GUS enzyme reaction. This product is colorless and soluble. Because of the stage of the seed pods when assayed (between two and three weeks after pollination), the seed tissues are still relatively soft and not yet dehydrated, and would allow diffusion of small soluble molecules. This artifactual staining is greatly reduced when mature seeds are used for histochemical staining. The same phenomenon appeared with the seeds from the tobacco plants transformed with the G12-GUS construct (see 3.1.3.2.c). Young tobacco seeds, less than 10 DAP, did not show any GUS activities, indicating that the GUS protein expression is developmentally regulated.

d. GUS activities in mature seeds.

Mature seeds were collected and GUS activities were fluorimetrically determined (Figure 33). The levels of GUS activity are independent of the GUS construct copy number. The A5 plant (with about 6 GUS inserts) shows a low GUS activity (1.3 nmole 4-MU/µg protein/min) whereas the plants A14 and A17 (with single copy insertion) show GUS activities 4 to 5 times higher (5.8 and 6.3 nmoles 4-MU/µg protein/min). As was the case with the globulin promoter-GUS construct (G12-GUS, see 3.1.3.2.d), there is no direct correlation between the copy number of the GUS genes and the GUS activities in mature seeds. This would again confirm the important role of the position effect on gene expression.

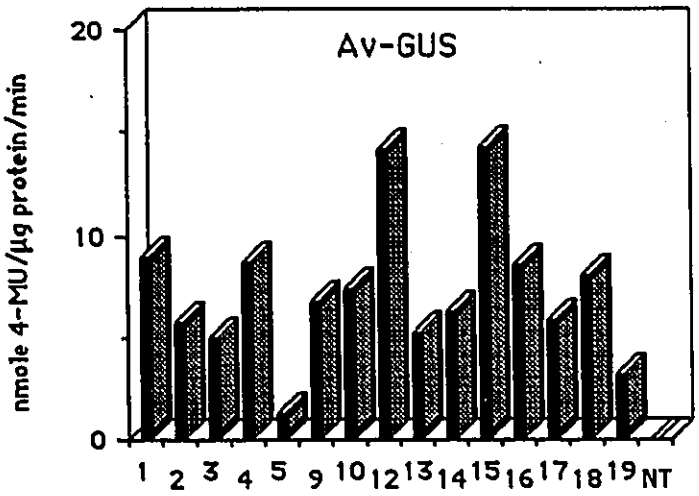


Figure 33: GUS activities in mature seeds from tobacco plants transformed with the Av-GUS construct.

GUS activities were measured fluorometrically and expressed in nmole 4-MU per µg protein per min. The plant numbers are indicated below the bars. NT represents an untransformed plant.

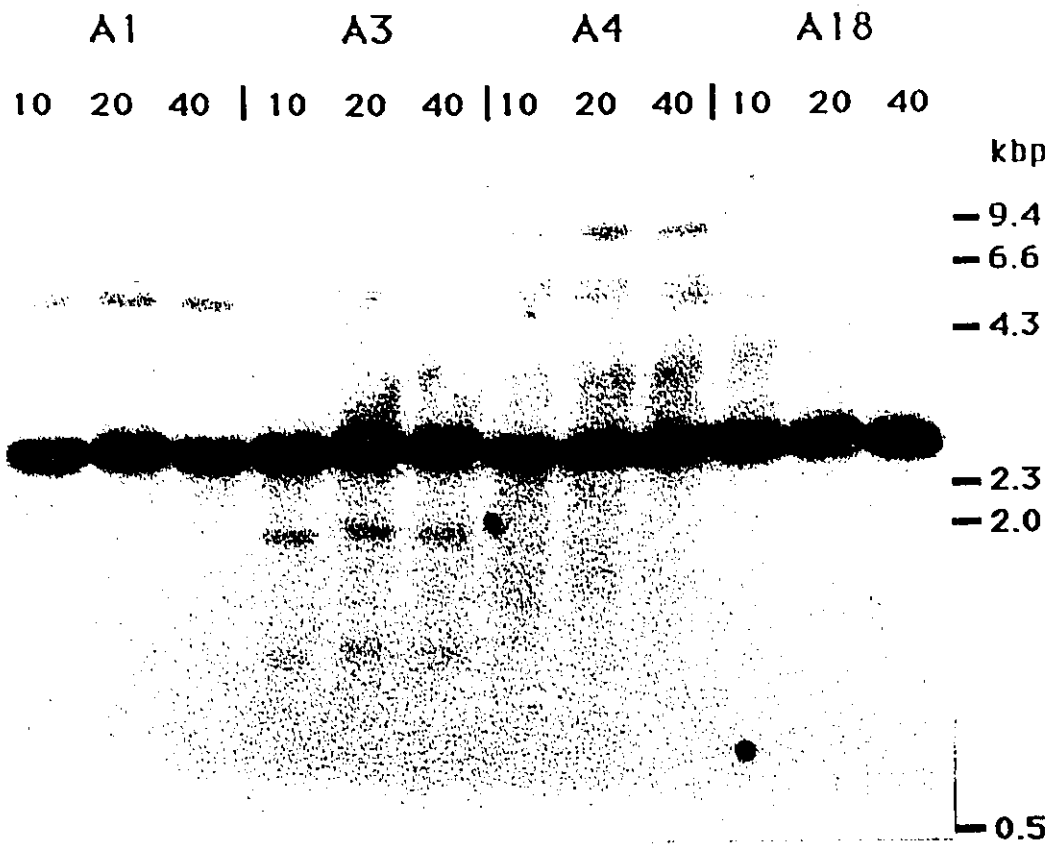
e. Primer extension analysis.

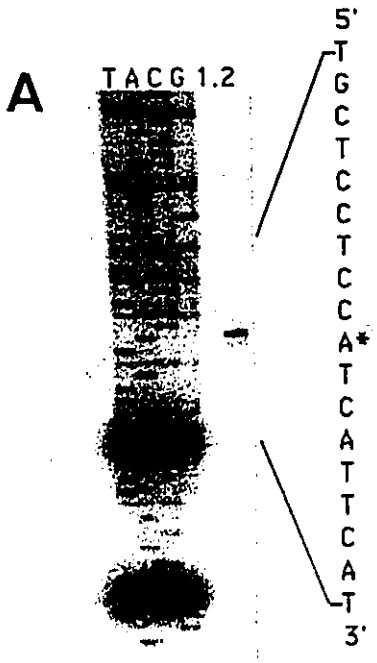
RNA from mid-maturation transformed tobacco seeds (plant 19) was analysed by primer extension using an oligonucleotide primer specific to the GUS coding region. Two independent extension reactions were carried out and loaded on a polyacrylamide sequencing gel. The signals obtained (Figure 35.B) enable mapping of the 5' end of the GUS transcripts at the A residue 182 nucleotides upstream of the avenin ATG (for the major signal). The GUS transcript, compared to the avenin mRNA, is 111 nucleotides longer and is also indicated in Figure 29. Although possible CAAT boxes can be found in the upstream region of this new transcriptional start site, a manual survey of the avenin promoter did not identify any putative TATA box that would play a role in the transcription mechanism in transgenic tobacco seeds.

The possibility that the primer extension products (both for the Av-GUS and the G12-GUS constructs) is artefactual (i.e. is not derived from the inserted constructs) is unlikely since the calculated length of the two 5' UTRs differs by 22 nucleotides.

f. Summary.

An avenin gene promoter was shown to direct a tissue-specific and developmentally controlled expression of the GUS reporter gene in transgenic tobacco seeds. This promoter includes in its sequence a very conserved region known as the '-300 element' or prolamin box which may play a role in the specific regulation of storage protein genes which harbor this element. As was the case with the G12-GUS construct, the transcriptional mechanisms in the tobacco endosperm tissues are different from those occurring in oat endosperms, resulting in a different transcriptional start site. However, this difference does not alter the spatial and temporal expression of the GUS gene.


A1
A3
A4
A18
10 20 40 | 10 20 40 | 10 20 40 | 10 20 40
kbp
9.4
6.6
4.3
2.3
2.0
0.5


A
T A C G 1.2
5'
T G C T C C T C C A*
T C A T T C A T
3'

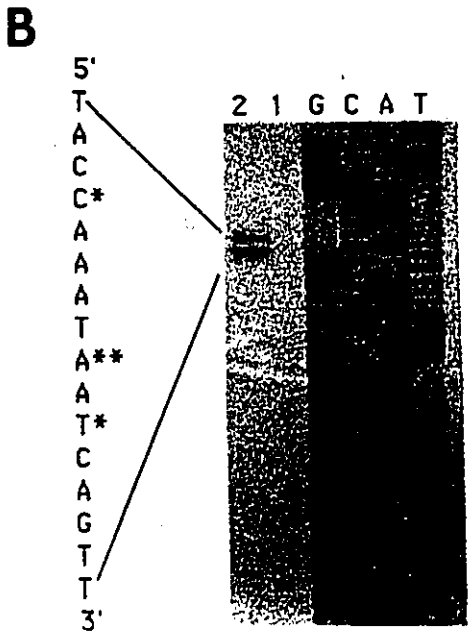

B
5'
T A C C* A A A T A** A T* C A G T T
3'
2 1 G C A T

Figure 34 : Southern blot analysis of 4 tobacco plants transformed with the Av-GUS construct using increasing amounts of enzymes.

DNA (5µg) from plant A1, A3, A4 and A18 was digested with 10, 20 or 40 units of EcoRI and electrophoresed in a 0.8% agarose gel. The gel was blotted on a nitrocellulose membrane and hybridized to a random labeled GUS probe. The photograph shows that the extra bands are due to real rearrangements and not to partial digestions of the DNA.

The molecular weight markers are from lambda DNA digested with HindIII.

Figure 35 : Determination of the 5' terminus of the avenin mRNA in oat seeds and of the GUS mRNA in transgenic tobacco seeds by primer extension analysis.

A: The Av-1 primer was ^{32}P end-labeled and annealed to 2.5µg of oat seed mRNA. Two extension reactions were realized and are indicated by lane numbers 1 and 2. The same primer was used to sequence the avenin genomic clone by the dideoxy chain termination method (lanes T, A, C and G). The asterisk indicates the position of the obtained product.

B: The GUS-p primer was ^{32}P end-labeled and annealed to 10µg of total tobacco seed RNA. Two extension reactions were performed on RNA extracted from plant A19 transformed with the Av-GUS construct. The same oligonucleotide primer was used to sequence the plasmid Av-GUS by the dideoxy chain termination method (lanes G, C, A and T). The two asterisks indicate the major product obtained and a single asterisk shows the position of the minor bands.

3.2.4. Avenin genes in different oat species.

DNA from different oat species was restricted with HindIII, EcoRI and BamHI. The resulting fragments were hybridized to an avenin probe (see 2.20.1.) and the autoradiograms are shown in Figure 36. By comparing the hybridization patterns obtained by Chesnut *et al.* (1989) on DNA restricted with HindIII with those in Figure 36.A, it is possible to confirm the presence of at least two classes of avenin genes which did not cross-hybridize at the washing stringencies which were used: some hybridization signals obtained in this study (the washing stringency was 0.1XSSC at 65°C) are not present in the autoradiogram presented by Chesnut *et al.* (1989), even though the washing stringency was lower (1XSSC at 60°C). It seems therefore possible that the copy number of the avenin genes in *A. sativa*, determined by Chesnut *et al.* (1989) at about 25 per haploid genome by using one probe, are underestimated.

Avenin genes have been shown to be organized in clusters, at least for some of them (Shotwell *et al.* 1990). This is, to some extent, verified by the Southern analysis in Figure 36.C where the majority of the restriction fragments that hybridize to the avenin probe are found in the high molecular weight DNA region, after restriction with BamHI. This is not due to partial digestion of the total DNA since the DNA samples were digested in duplicated reactions at the same time, one that served for globulin hybridization and the second for avenin hybridization. Figure 28.C shows that the DNA was completely digested by BamHI.

As determined in section 3.1.6., *A. clauda* and *A. pilosa* (both with a CC genome, lanes 9 and 11) are closely related. In the AA genomes (lanes 6, 7 and 8), small differences appear in each species, which make their relationship and phylogeny difficult to assess. In the hexaploid species, it seems that the cultivars Donald and Ogle (lanes 2 and 3) are more closely related to each other than to the cultivar Hinoat (lane 1).

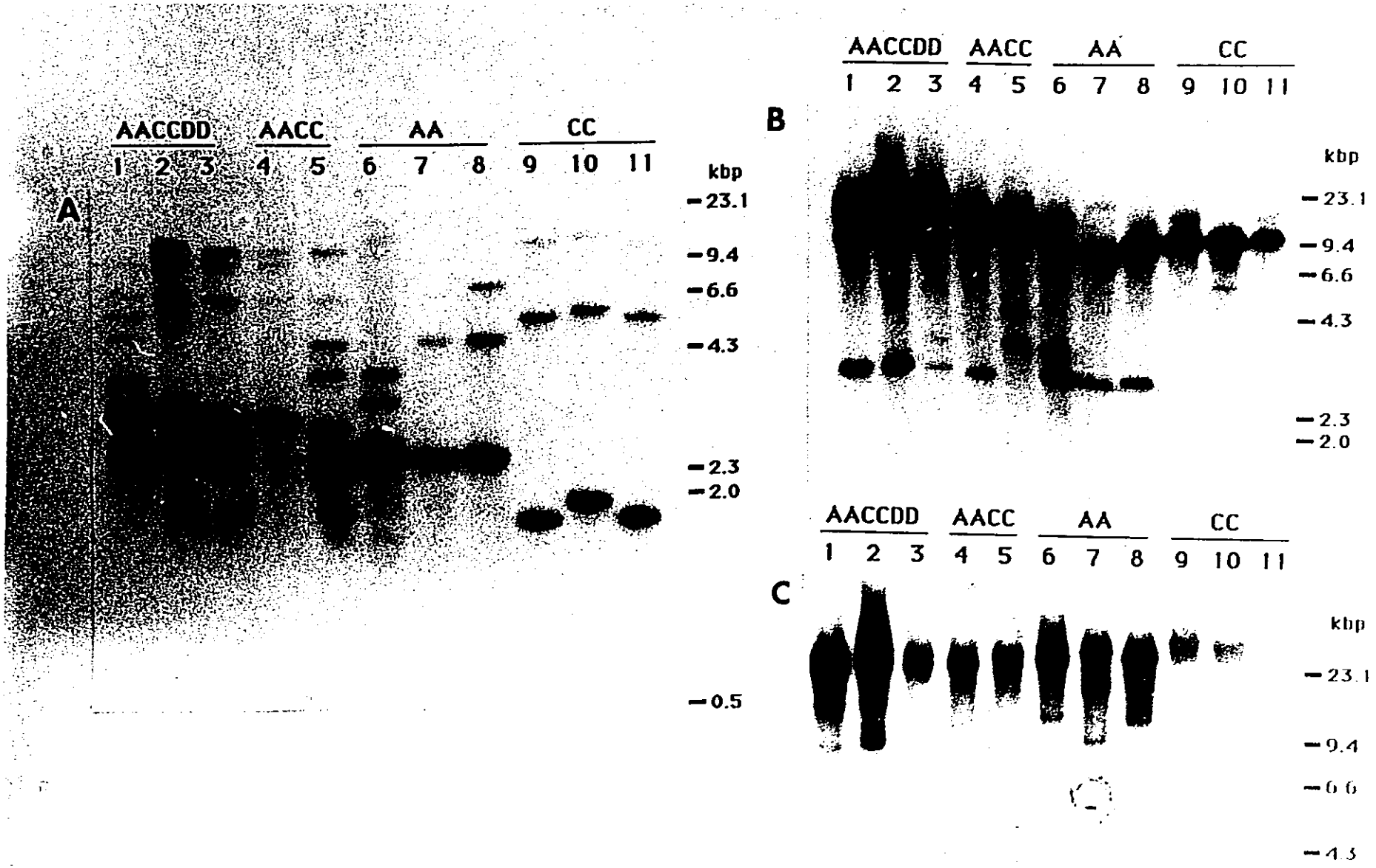

A
AACCDD
AACC
AA
CC
1 2 3 4 5 6 7 8 9 10 11
kbp
23.1
9.4
6.6
4.3
2.3
2.0
0.5
B
AACCDD
AACC
AA
CC
1 2 3 4 5 6 7 8 9 10 11
kbp
23.1
9.4
6.6
4.3
2.3
2.0
C
AACCDD
AACC
AA
CC
1 2 3 4 5 6 7 8 9 10 11
kbp
23.1
9.4
6.6
4.3

Figure 36 : Analysis of avenin genes in different oat species by Southern hybridization.

DNA (5µg) from oat species representing different levels of ploidy was digested with HindIII (A), EcoRI (B) or BamHI (C) and electrophoresed in a 0.8% agarose gel. The gels were blotted on nitrocellulose membranes and hybridized to a random labeled avenin probe.

The hexaploid oat species (genome AACCDD) are *A. sativa* cv. Hinoat (lane 1), *A. sativa* cv. Donald (lane 2) and *A. sativa* cv. Ogle (lane 3).
The tetraploid oat species (genome AACC) are *A. magna* (lane 4) and *A. murphyi* (lane 5).
The diploid oat species (genome AA) are *A. longiglumis* (lane 6), *A. prostrata* (lane 7) and *A. wiestii* (lane 8). The genome CC is represented by *A. clauda* (lane 9), *A. ventricosa* (lane 10) and *A. pilosa* (lane 11).

The molecular weight markers are from lambda DNA digested with HindIII.

Chapter 4.
General discussion.

4.1. Functional oat promoters in transgenic tobacco seeds.

4.1.1. Globulin promoters.

From three oat globulin genomic clones, it has been demonstrated that two of them (Glav-3 and Glav-12) possess functional promoters and that the promoter of Glav-1 is likely to be inactive. The 2610 bp of the Glav-12 promoter region (including the 5' UTR) are able to direct the endosperm-specific and developmentally regulated expression of the GUS reporter gene. Furthermore, 1721 bp of the promoter region of the Glav-3 globulin clone (5' UTR included) are sufficient for the seed-specific and temporal expression of the globulin protein. The expression of an oat globulin promoter translationally fused to the GUS coding sequence (G12-GUS construct) was shown by histochemical staining to be specifically restricted to the endosperm of transgenic tobacco seeds. The analysis of transgenic plants transformed with the genomic sequence of an oat globulin clone (G3P construct) showed that the developmental and spatial regulatory controls were maintained in the transgenic dicot plant. This is the first example of cereal globulin promoters functioning properly in a dicot plant (tobacco).

The lack of conserved elements found in other promoters (from either monocot or dicot genes) in the promoter region of oat globulin genes is demonstrated in this study by the sequence analysis of the three oat globulin gene promoters and raises interesting questions

about the regulation controls of the expression of the oat globulin multigene family members.

An imperfect core motif of the prolamin box was identified and shown not to direct any expression (construct G3Δ-259-GUS). A similar finding was reported with the promoter of a rice glutelin gene (Takaiwa *et al.* 1991) where a core motif of the prolamin box was present in a promoter deletion construct that did not direct any expression. In this latter study, the promoter region responsible for the endosperm-specific expression of the GUS sequence has been shown to lie between positions -441 and -237 from the transcriptional start site. Accordingly, this present study has shown that additional sequences upstream of the nucleotide at position -259 are required for the expression of the oat globulin promoter fused to the GUS coding region. However, as suggested earlier (Shirsat *et al.* 1989), involvement of the prolamin box core motif is not ruled out if we suppose that it plays a role of enhancement (quantitative) rather than a developmentally-specific function (qualitative).

4.1.2. Avenin promoter.

An avenin promoter fragment (335 bp upstream of the transcriptional start site) was shown for the first time to be sufficient to properly regulate the expression of the GUS reporter gene in transgenic tobacco seeds. Both endosperm-specificity and temporal regulation were maintained in the tobacco heterologous system. This promoter region includes the well conserved prolamin box and a 71-bp stretch of sequence upstream of this '-300 element'.

It is not known whether a lengthier promoter would direct a stronger expression. Comparisons with GUS activities obtained with other seed-specific promoters are difficult because of differences in experimental conditions. However, the values obtained with the avenin promoter expressed in tobacco seeds (around 5nmoles 4-MU/min/µg prot.) are higher than those reported in the literature

with other cereal seed storage gene promoters (0.004 to 0.018nmoles 4-MU/min/µg prot. with a maize zein promoter, Schernthaner *et al.* 1988, around 0.05 nmoles 4-MU/min/µg prot. with a rice glutelin promoter, Takaiwa *et al.* 1991 and around 2 pmoles 4-MU/min/µg prot. with a wheat HMW glutenin promoter, Thomas and Flavell 1990).

The strong expression in tobacco of an avenin gene promoter fused to GUS is further demonstrated by the time-course appearance of the blue precipitate in the GUS histochemical staining of the transgenic seeds: between 1/2 and 1 hour with the avenin promoter, 24 hours with the zein promoter and overnight with the rice glutelin promoter.

It seems therefore that the avenin promoter, like the Glav-12 promoter, is not only recognized by the transcriptional machinery of a dicot plant such as tobacco, but is able to direct a high level of expression in a plant where prolamin-type proteins are known to be absent.

4.1.3. Start of transcription.

It has been shown in this study that the transcriptional start site of the avenin and globulin promoters is different in tobacco from the one determined in the oat plant. However, the tissue-specificity was not altered in either case. This is the first report of the mapping of the transcriptional start site of a monocot gene promoter after transfer into a dicot plant.

Although possible additional CAAT and TATA boxes can be identified in the Glav-12 promoter sequence, no obvious TATA box which would be related to the new transcriptional start site in tobacco can be found in the avenin promoter. The TATA box has been defined as a consensus sequence, and to what extent another sequence can act as a binding site for the RNA polymerase complex remains to be clarified. Very little is known about plant TATA boxes, or other functionally related sequences (Kuhlemeier 1992). Two reports have

appeared where the transcriptional start site of dicot embryo-specific genes expressed in tobacco was similar to the one in the original plant: a pea vicilin gene (Higgins *et al.* 1988) and a phaseolin minigene (Chee *et al.* 1986).

Mapping of the 5' end of the globulin and the avenin mRNAs showed a difference of respectively 4 nt and 2 nt from the determination done by Shotwell *et al.* (1990). The additional length of the 5' UTR reported by these authors can be accounted for by the different method that was used. In the present study, mapping was made by primer extension analysis, whereas Shotwell *et al.* used the mung bean nuclease protection assay. It is then possible that the cap structure of the mRNA can protect a small portion of the hybridized genomic sequence, the length of this additional protected portion probably depending on the degree of methylation of the cap structure (Sonenberg 1988). A similar difference was also reported by Reeves and Okita (1987), who showed the transcriptional start site of a wheat alpha/beta gliadin transcript in wheat seeds was one nt longer when the determination was performed by nuclease S1 mapping as opposed to primer extension analysis (see also Figure 30).

4.1.4. Different factors are involved in the transcriptional mechanisms.

This study and other similar previous reports (*e.g.* Katagiri and Chua 1992) have shown the importance of the primary structure of the promoter: the DNA sequence itself. Within this sequence, the necessity for the presence of *cis*-acting element(s) has been indicated. The fact that such *cis*-acting elements are part of the transcriptional mechanisms has been documented in two different ways[7].

[7]CAAT and TATA boxes are part of the *cis*-acting elements and are required for maximal gene expression. They are however not sufficient to direct any specific expression. Similarly, the RNA polymerase complex will not be considered here, despite its obvious requirement for transcription.

1. Binding of specific nuclear factors to the promoter region,

2. Decrease or disappearance of the promoter functionality after removal or alteration of the element.

The specific binding of a nuclear factor to the consensus sequence of maize zeins was first observed by Maier *et al.* (1987). It is interesting to see that the core motif of the prolamin box is also present in this sequence. The consequences of the binding of this *trans*-acting factor can be multiple: turning-on of the genes, transcription rate enhancement or tissue-specific expression. A "virtual identity to the viral/animal enhancer core sequence" was also reported (Maier *et al.* 1987) concerning the core motif of the prolamin box. Other studies have since shown that multiple nuclear proteins can interact with the promoter region of different storage protein genes, among them Vellanoweth and Okita (1993) with the α/ß-gliadin gene, Kawagoe and Murai (1992) with the ß-phaseolin gene, Takaiwa and Oono (1990) with a rice glutelin gene, Lessard *et al.* (1991) with ß-conglycinin genes and Jofuku *et al.* (1987) with a soybean lectin gene. It is apparent from these studies that the presence of the DNA-binding proteins fluctuates in parallel with the seed protein gene transcriptional activity. A well-studied example of defective transcription of an endosperm-specific storage protein gene has been provided by the *Opaque-2* mutant line in maize. It has been shown that *O2* is a *trans*-acting transcriptional activator of the 22 kDa class of zein genes (Hartings *et al.* 1989). The *O2* protein was also shown to belong to a family of ubiquitous transcription factors, confirming that the specificity of gene expression is brought about by the combined presence of the *cis*- and *trans*-acting factors.

Modification of the promoter region, and specifically of the possible *cis*-acting elements can also lead to a better evaluation of the role of these elements. The first approach to detect functional regions in a promoter involves the progressive deletion of promoter portions (see section 1.2.). In the case of the oat globulin promoters, a major

difference in the proximal region between a potentially inactive promoter (Glav-1) and two functional promoters (Glav-3 and Glav-12), determined the locations of the deletions that were made. The second approach is then to alter the identified possible regulatory element by either removing it from the sequence, introducing mutational changes or inserting it in another sequence. The site-specific mutational analysis is believed to be the most appropriate method since deletion can affect gene expression indirectly by changing the conformational structure of the DNA in the region of interest (Lelievre *et al.* 1992).

Concerning the oat globulin promoters, this present work has determined that the regulatory element(s) lies upstream of the -259 nt. Additional work should enable the precise localization of the functional portion of the promoter. As for the avenin promoter, an analysis by site-specific mutations might be envisaged at this stage, if we consider the relative shortness of the functional promoter (221 bp long in tobacco or 332 bp long in oat) and the presence of the highly conserved prolamin box.

4.2. Other aspects of the regulation of seed storage protein gene expression.

4.2.1. DNA methylation.

Shortly after replication, DNA undergoes minor modifications by methylation, the most common being the methylation of the cytosine base. The presence of methylcytosine has a stabilizing effect on the double helix structure of the DNA. When present in the sites of early unwinding (initiation sites of both replication and transcription), the methylcytosines may prevent the entry of an RNA polymerase at an adjacent TATA box. In contrast, the methylation of an adenine is thought to play a positive role in the unwinding of DNA. However, adenine methylation occurs less frequently (Adams 1990). An

example of a DNA methylation study of maize seed storage protein genes has been provided by Bianchi and Viotti (1988). Although a common methylated pattern was reported in different tissues, it was shown that the glutelin and about 65% of the zein genes were undermethylated in the endosperm tissue. This undermethylated state is acquired or already present at an early stage of endosperm development and is likely to play a role in the activation and/or expression of seed storage protein genes. It would then be interesting, for comparison, to assess the methylated state of the oat storage protein genes in different tissues and see whether methylation is part of the possible regulatory factors for the expression of oat seed storage proteins.

4.2.2. Ploidy level of the endosperm tissue.

The double fertilization event occurring in angiosperm plants results in the formation of two separate tissues: the embryo and the endosperm. The ploidy level of each tissue is different, the endosperm resulting from the fusion of two maternal nuclei and one pollen nucleus, whereas the embryo is formed by the fusion of one maternal nucleus and one pollen nucleus. The consequence of the different ploidy levels is not clear. However, in relation to the methylated state of the storage protein genes, it has been hypothesized that "undermethylation could occur before fertilization in the two maternal polar nuclei, while the pollen genome would remain methylated and therefore account for most of the fraction of endosperm DNA (≈35%) which is methylated at zein and glutelin sequences" (Bianchi and Viotti 1988). The presence of other regulating processes is then obvious since no activity of zein or glutelin can be detected in the embryo. The question remains to be answered whether this genetic imprinting theory also applies to hexaploid plants such as *A. sativa*.

4.2.3. Physiological factors.

Examples of the effects of abscisic acid (ABA), a plant growth regulator, on the levels of seed storage protein gene mRNAs and proteins have been reported (DeLisle and Crouch 1989, among others). An increase in the mRNAs for napin and cruciferin (*Brassica napus* seed storage proteins) was reported in cultured embryos when ABA was applied. Similar results were obtained with embryos from other plants. It has been shown that during recovery from a sulfur deficiency stress, the amount of legumin mRNAs in pea seeds increased 20-fold, whereas that of vicilin mRNAs gradually decreased (Beach *et al.* 1985). It has been shown that the regulation had more to do with mRNA stability than with the rate of transcription itself and it was therefore pointed out that the possible key sequences could be located in the 5' or 3' UTRs (Spencer *et al.* 1990). Another report (Marcotte *et al.* 1988) demonstrated a rapid action of ABA on the initiation of gene transcription using the promoter region of a wheat embryo protein (Em). Are the transcription mechanisms of oat seed storage protein genes affected by ABA or other growth regulators? Is there a different response in the endosperm and in the embryo? Is the mRNA stability altered during the recovery from stress? The use of transgenic plants could bring some additional light in these processes, especially when a GUS translational fusion is used (*i.e.* the 5'UTR is exclusively derived from the original genomic sequence).

4.2.4. Post-transcriptional mechanisms.

The processing of the pre-mRNAs (splicing out of the introns and addition of poly(A)-tail) plays an important role in the regulation of gene expression. Demonstration has been made that the fidelity with which RNA processing signals are recognized in different plants (usually the original host plant *versus* tobacco) can vary (Keith and Chua 1986). Although the present study did not directly assess the correct splicing out of the introns, it is assumed that was the case

since the positioning of the introns in Glav-3 was similar to the ones observed in dicot globulin-like genes and that there was expression at the protein level of the introduced gene. It would certainly be interesting to verify whether the splicing-out of the introns was correctly performed in transgenic tobacco seeds on the Glav-3 transcripts, especially since the novel RNA purification method described in this study allows the use of the RT-PCR.

The addition of an intron within the coding sequence of a reporter gene has been shown to result in an increase of gene expression (intron 1 of the maize *Shrunken-1* gene, Maas *et al.* 1991, intron 1 of the maize alcohol dehydrogenase-1 gene, *Adh-1*, Callis *et al.* 1987, a modified intron of a castor bean catalase gene, Ohta *et al.* 1990, intron 2 or 6 of the maize *Adh-1* gene, Mascarenhas *et al.* 1990, intron 2 of the potato light-inducible ST-LS1 gene, Vancanneyt *et al.* 1990, intron 1 of a rice actin gene, McElroy *et al.* 1990). Explanations for these increases are numerous and often contradictory. It remains clear that there are other genes that do not contain introns and are still capable of high levels of expression, one case being the prolamin genes in monocot seeds.

It is known that mRNA stability and translatability (Vancanneyt *et al.* 1990) are part of the regulation of gene expression. The secondary and tertiary structure can increase the half-life time of the transcripts by shielding the RNA from the action of nucleases and favour the initiation of protein synthesis at privileged sites. It has been proposed that the intron splicing mechanisms themselves could provide such a protection to the mRNAs (Mascarenhas *et al.* 1990). Although less studied in plant genes, the 5' and 3' UTR are also part of the regulation mechanisms. In the case of the GUS reporter gene fused to the avenin and globulin promoters, the NOSter was used but additional constructs would be needed to assess its role in the endosperm expression of the GUS-fusion constructs used in this study.

4.2.5. Translational and posttranslational controls.

Initiation rate, elongation and termination phases are the different events taking place during the translation of a mRNA molecule.

In both the Av-GUS and the G12-GUS constructs, two possible translation initiation codons (AUG) are available (see 3.1.3.2. and 3.2.3.), the first derived from the original genomic sequences and the second corresponding to the GUS coding sequence. Although determination of which AUG is chosen for the translation initiation has not been carried out in this study, it is probable that translation of the GUS product initiated from the first AUG. Such a choice, when multiple AUGs are present, has been reported (Putterill and Gardner 1989) and has been shown to fit the modified Kozak's scanning model for translation (Kozak 1989).

In the case of storage protein synthesis in oat, Boyer *et al.* (1992) have proposed that the translation elongation or the termination reactions are likely to be the regulatory steps. Fidelity and rate of translation can be determining factors in the regulation of protein synthesis (see appendix 2).

When correctly expressed as a pro-protein, the seed storage proteins are subjected to various modifications and constraints that, if not properly processed, can substantially affect their accumulation in endosperm or embryo tissues. These numerous constraints are placed upon the targeting of the protein through the proper sub-cellular compartments, the cleavage, assembly and folding of the protein, its packaging into protein bodies and its susceptibility to special endoproteases. In the latter case, it has been shown that some storage proteins are degraded when they are expressed in plants which are not their original host (see section 1.2.3.). The high level of constraints during the posttranslational events has also been demonstrated when a modified storage protein (insertion of a 45-bp sequence containing a high proportion of methionine codons) was not able to be stably incorporated in the protein bodies of transgenic seeds (Hoffman *et al.*

1988), even though this protein was correctly expressed and processed.

Whether the oat globulin Glav-3, expressed in transgenic tobacco seeds, is correctly processed remains difficult to assess. Although the Glav-3 protein, immunologically detected, corresponds to the size of globulins synthesized in oat seeds, the signals were weak. A monoclonal antibody would be ideally suited for a deeper analysis of the targeting and fate of the Glav-3 protein in transgenic tobacco seeds.

4.3. Conclusion and future prospects.

It has been shown in this study that two promoters from oat globulin genes and one promoter from an avenin gene are similarly regulated in transgenic tobacco plants as they are in oat. This is the first time that the promoters of oat seed storage protein genes have been reported as endosperm specific when transferred into a dicot species.

Examination of the oat globulin promoters did not reveal any conserved elements that could be involved in the control of the regulation of these genes, except for an imperfect core motif of the '-300 element'. Deletion of the region upstream of this putative box showed that an additional portion of the promoter is required for correct expression. Concerning the avenin promoter, its structural homology with members of other prolamin gene families was confirmed. The presence of the 'prolamin box' within a functional promoter further demonstrates the possible role of this element in the control of the regulation of prolamin gene expression.

In a molecular biology context, the identification of functional promoters can be of great use in the understanding of the molecular events regulating plant gene expression. The possibilities of future

investigations are numerous and can involve many areas of research, from basic plant molecular biology to the more applied field of genetic engineering of plants.

The newly developed methods of gene transfer to cereals (Potrykus 1990) is giving more impact to the identification of functional seed-specific promoters. A recent report of the regeneration of fertile transgenic oat plants (Somers *et al.* 1992) is raising exciting hopes and prospects for the future. The potential agronomic and economic utility of obtaining transgenic plants with altered characteristics is undeniable (Gasser and Fraley 1989), even if many problems still have to be overcome before we can see a safe and routinely possible application of these discoveries. It is hoped that the work that was presented here will contribute to a better knowledge of plant biology.

Appendix 1: Preparation and composition of the media used for tobacco transformation and regeneration.

5X MS base (giving a 2% sucrose 1X MS solution):

	1 litre
KNO_3	9.5g
NH_4NO_3	8.25g
$MgSO_4$-$7H_2O$	1.85g
$CaCl_2$-$2H_2O$	2.2g
KH_2PO_4	0.85g
Sucrose	100g
KI stock	5mls
B5 vit.100X	50mls
MS µnutrient100X	50mls
FeNaEDTA	0.2g

1000X MS µnutrient

	1litre
$MnSO_4$-$4H_2O$	23.3g
H_3BO_3	6.2g
$ZnSO_4$-$7H_2O$	8.6g
Na_2MoO_4-$2H_2O$	0.25g
$CuSO_4$-$5H_2O$	0.025g
$CaCl_2$-$6H_2O$	0.025g

100X B5 vitamins

	1litre
Myo-inositol	10g
Nicotinic acid	0.1g
Pyridoxine-HCl	0.1g
Thiamine-HCl	1g

KI stock	0.83g/l

Appendix 2

tRNA extraction from mid-maturation seeds of different plants.

An increasing number of sequenced coding genes has now shown that synonymous codons are used preferentially over others and that a codon usage bias exists in most genes (Murray et al. 1989, Wain-Hobson et al. 1981). The translation mechanisms involve many different factors such as RNA secondary and tertiary structures, RNA stability, energy of interaction between codon and anticodon and tRNA availability. The fidelity and the rate of translation (efficiency of translation) take into account some or all of these components.

One of the first hypotheses for the bias in genetic code usage has been the tRNA adaptation theory (Ikemura 1985).The tRNA content in *E. coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae* have been determined and the frequencies of use of each isoacceptor have been plotted against its usage, showing a strong correlation between tRNA abundance and the choice of synonymous codons (Ikemura 1985). The correlation is even stronger when a distinction is made between highly and weakly expressed genes. Codons corresponding to minor tRNAs are avoided in these efficiently expressed genes (Grosjean and Fiers 1982).

The efficiency of translation is reflected by both the fidelity and the rate of translation. The rate-limiting step in the elongation cycle of polypeptide is tRNA selection (Varenne et al. 1984). When codon usage matches tRNA availability, it then becomes clear that the probability of correct codon-anticodon interactions is increased, thus affecting the translation rate. However, the main factor determining protein yield is the

initiation rate, but elongation and initiation can interact to modulate protein yield (Gouy and Gautier 1982).

It has been shown that the final high proportion of globulins in oat seeds is mainly regulated by the translation elongation or terminal reactions, not the initiation rates (Boyer et al. 1992). It then becomes clear that the relationship between tRNA abundance and high level of globulin-like proteins in oat, rice and dicot seeds will have to be investigated.

A preliminary study was conducted by isolating the tRNAs from mid-maturation seeds. Figure 37 shows the tRNA banding patterns of the different tRNA pools. No direct conclusion can be drawn, except that the tRNA population from oat, rice and soybean do not share specific banding patterns that would directly bring the Ikemura's adaptation theory into prospect for seed storage protein regulation. A more detailed analysis is therefore needed, especially in the determination and frequency of use of each isoacceptor.

Total tRNA isolation method (modified from Steinmetz and Weil 1986 and Cornelis 1978):

Ten grams of dehulled mid-maturation seeds were ground and homogenized with 1.5 vol. of buffer 1 (see below) in a 100ml beaker. The homogenate was filtered through a 50µm nylon cloth into a 250ml erlenmeyer containing 1vol of phenol saturated with buffer 1 and SDS (1% final concentration in the aqueous phase). The mixture was vigorously stirred for 1h at 4°C, reducing the speed of the magnetic stirring after 15min. The mixture was then divided into two disposable sterile screw cap tubes and centrifuged at 3,000 rpm for 15min. The aqueous phase was recovered and re-extracted with phenol. After centrifugation, the aqueous phase was divided into three 50ml

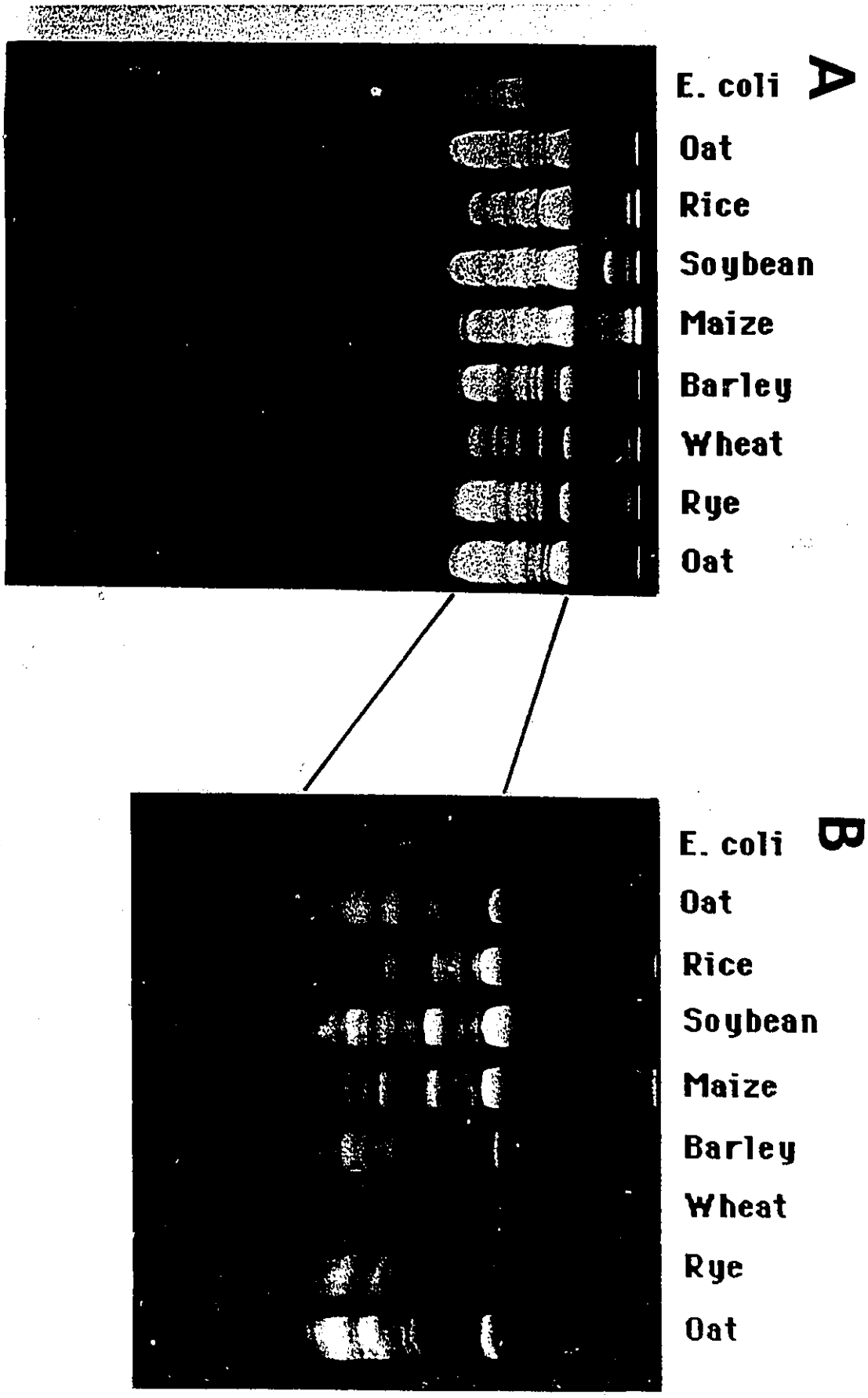

A
E. coli
Oat
Rice
Soybean
Maize
Barley
Wheat
Rye
Oat
B
E. coli
Oat
Rice
Soybean
Maize
Barley
Wheat
Rye
Oat

Figure 37 : tRNAs from mid-maturation seeds of different plants.

A: Each sample of 20µg of tRNAs was loaded on a 15% denaturing acrylamide gel (with a 6% stacking gel). The gel was run for 22 hours at 300V in a cold room. The tRNA bands were visualized by ethydium bromide staining.

B: Same as in A except that 5µg of each sample were loaded and the gel was run for 30 hours.

tubes and 0.1vol of 20% K acetate pH 5.0 and 2vol of cold 99% ethanol were added. The RNAs were precipitated at -20°C for 1h and recovered by centrifugation (10,000 rpm for 15min). The pellets were each resuspended in 700µl of DEPC-treated dH_2O and transferred to a 1.5ml centrifuge tube. The HMW RNAs were selectively precipitated by addition of 1vol of 4M ammonium sulfate (30min on ice). After centrifugation, the supernatant was loaded on a short column (about 1/3 cc) of Sepharose 4B equilibrated with the buffer 2. The column was washed with the same buffer and the tRNAs were eluted with DEPC-treated dH_2O (4ml). The eluate was collected in a 15ml Corex tube and 2.5vol of ethanol were added to precipitate the tRNAs (-20°C). After centrifugation (10,000 rpm, 15min), the tRNA pellets were resuspended in 1ml of DEPC-treated dH_2O and dialysed (No. 3 dialysis tubing, Spectra/Por) overnight against distilled water. After dialysis, 400µl-aliquots were withdrawn and transferred into 1.5ml centrifuge tubes. The tRNAs were then precipitated at -20°C after addition of 1/15vol of Na acetate and 2.5vol of ethanol. After centrifugation, the samples were resuspended in a total volume of 200µl. The deaminoacylation was carried out at 37°C for 1h in a small volume of buffer 3. The tRNAs were again precipitated and the concentration was spectrophotometrically determined.

Buffer 1: 10mM Tris-HCl pH 7.4
10mM $MgCl_2$

Buffer 2: 2mM ammonium sulfate
10mM Na acetate pH 4.5
10mM $MgCl_2$
1mM EDTA

Buffer 3: 1.8M Tris-HCl pH 8.0
1M NaCl

References.

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

Adeli K., Allan-Wojtas P. and Altosaar I. 1984. Intracellular transport and posttranslational cleavage of oat globulin precursors. Plant Physiol. **76**: 16-20.

Bagga S., Sutton D., Kemp J.D. and Sengupta-Gopolan C. 1992. Constitutive expression of the ß-phaseolin gene in different tissues of transgenic alfalfa does not ensure phaseolin accumulation in non-seed tissue. Plant Mol. Biol. **19**: 951-958.

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

Bäumlein H., Boerjan W., Nagy I., Panitz R., Inzé D. and Wobus U. 1991. Upstream sequences regulating legumin gene expression in heterologous transgenic plants. Mol. Gen. Genet. **225**: 121-128.

Bäumlein H., Nagy I., Villarroel R., Inzé D. and Wobus U. 1992. *Cis*-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. The Plant J. **2**: 233-239.

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 sulfur-deficient pea seeds. Nucl. Acids Res. **13**: 999-1013.

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

Bednarek S.Y. and Raikhel N.V. 1992. Intracellular trafficking of secretory proteins. Plant Mol. Biol. **20**: 133-150.

Bevan M. 1984. Binary *Agrobacterium* vectors for plant transformation. Nucl. Acids Res. **12**: 8711-8721.

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

Bietz J.A. 1982. Cereal prolamin evolution and homology revealed by sequence analysis. Biochem. Genet. **20**: 1039-1053.

Birnboim H.C. and Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**: 1513-1518.

Boyer S.K., Shotwell M.A. and Larkins B.A. 1992. Evidence for the translational control of storage protein gene expression in oat seeds. J. Biol. Chem. **267**: 17449-17457.

Brinegar A.C. and Peterson D.M. 1982. Separation and characterization of oat globulin polypeptides. Arch. Biochem. Biophys. **219**: 71-79.

Brown J.W.S. 1986. A catalogue of splice junction and putative branch point sequences from plant introns. Nucl. Acids Res. **14**: 9549-9559.

Burgess S.R., Shewry P.R., Matlashewski G.J., Altosaar I. and Miflin B.J. 1983. Characteristics of oat (*Avena sativa* L.) seed globulins. J. Exp. Bot. **34**: 1320-1332.

Burow M.D., Sen P., Chlan C.A. and Murai N. 1992. Developmental control of the ß-phaseolin gene requires positive, negative, and temporal seed-specific transcriptional regulatory elements and a negative element for stem and root expression. Plant J. **2**: 537-548.

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

Bustos M.M., Kalkan F.A., VandenBosch K.A. and Hall T.C. 1991. Differential accumulation of four phaseolin glycoforms in transgenic tobacco. Plant Mol. Biol. **16**: 381-395.

Callis J., Fromm M. and Walbot V. 1987. Introns increase gene expression in cultured maize cells. Genes and Development **1**: 1183-1200.

Cameron-Mills V. and Brandt A. 1988. A γ-hordein gene. Plant Mol. Biol. **11**: 449-461.

Casey R. and Domoney C. 1987. The structure of plant storage protein genes. Plant Mol. Biol. Reporter **5**: 261-281.

Chamberland S., Daigle N. and Bernier F. 1992. The legumin boxes and the 3' part of a soybean ß-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. Plant Mol. Biol. **19**: 937-949.

Chee P.P., Klassy R.C. and Slightom J.L. 1986. Expression of a bean storage protein 'phaseolin minigene' in foreign plant tissues. Gene **41**: 47-57.

Chen Z.L., Schuler M.A. and Beachy R.N. 1986. Functional analysis of regulatory elements in 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.

Chesnut R.S., Shotwell M.A., Boyer S.K. and Larkins B.A. 1989. Analysis of avenin proteins and the expression of their mRNAs in developing oat seeds. Plant Cell **1**: 913-924.

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

Colot V., Bartels D., Thompson R. and Flavell R. 1989. Molecular characterization of an active LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol. Gen. Genet. **216**: 81-90.

Cornelis P. 1978. Use of high salt sepharose 4B chromatography for the extraction of tRNA from plant tissues. Plant Sci. Lett. **11**: 3-10.

DeLisle A.J. and Crouch M.L. 1989. Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to exogenous abscisic acid. Plant Physiol. **91**: 617-623.

Derbyshire E., Wright D.J. and Boulter D. 1976. Legumin and vicilin, storage proteins of legume seeds. Phytochemistry **15**: 3-24.

Dickinson C.D., Floener L.A., Lilley G.G. and Nielsen N.C. 1987. Self-assembly of proglycinin and hybrid proglycinin synthesized *in vitro* from cDNA. Proc. Natl. Acad. Sci. USA **84**: 5525-5529.

Dickinson C.D., Evans R.P. and Nielsen N.C. 1988. RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. Nucl. Acids Res. **16**: 371.

Dickinson C.D., Hussein E.H.A. and Nielsen N.C. 1989. Role of posttranslational cleavage in glycinin assembly. Plant Cell **1**: 459-469.

Dreyer K. and Schulte-Holthausen H. 1991. Casein is a potent enhancer for restriction enzyme activity. Nucl. Acids Res. **19**: 4295.

Egorov T.A. 1988. The amino acid sequence of the "fast" avenin component (*Avena sativa* L.). J. Cereal Sci. **8**: 289-292.

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 in seeds of *Nicotiana plumbaginifolia*. Plant Mol. Biol. **10**: 203-214.

Entwistle J. 1988. Primary structure of a C-hordein gene from barley. Carlsberg Res. Commun. **53**: 247-258.

Fabijanski S. and Altosaar I. 1985. Evidence for translational control of storage protein biosynthesis during embryogenesis of *Avena sativa* L. (oat endosperm). Plant Mol. Biol. **4**: 211-218.

Fabijanski S., Altosaar I., Laurière M., Pernollet J.-C. and Mossé J. 1985a. Antigenic homologies between oat and wheat globulins. FEBS Lett. **182**: 465-468.

Fabijanski S., Matlashewski G.J. and Altosaar I. 1985b. Characterization of developing oat seed mRNA: evidence for many globulin mRNAs. Plant Mol. Biol. **4**: 205-210.

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. Nucl. Acids Res. **13**: 7327-7339.

Galle A.-M., Sallantin M. and Pernollet J.-C. 1988. Influence of ionic strength on the extraction of oat seed storage proteins. Plant Physiol. Biochem. **26**: 733-738.

Garson K., Matlashewski G.J., Adeli K., Robert L.S. and Altosaar I. 1983. Isolation of intact polysomes from mechanically dehulled developing grain. Anal. Biochem. **134**: 512-516.

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

Gatehouse J.A., Evans I.M., Croy R.R.D. and Boulter D. 1986. Differential expression of genes during legume seed development. Philos. Trans. R. Soc. Lond. (Biol.) **314**: 367-384.

Goodall G.J. and Filipowicz W. 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58**: 473-483.

Goodall G.J. and Filipowicz W. 1990. The minimum functional length of pre-mRNA introns in monocots and dicots. Plant Mol. Biol. **14**: 727-733.

Gouy M. and Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucl. Acids Res. **10**: 7055-7074.

Greenwood J.S. and Chrispeels M.J. 1985. Correct targeting of the bean storage protein phaseolin in the seeds of transgenic tobacco. Plant Physiol. **79**: 65-71.

Grosjean H. and Fiers W. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene. **18**: 199-209.

Halford N.G., Forde J., Shewry P.R. and Kreis M. 1989. Functional analysis of the upstream regions of a silent and an expressed member of a family of wheat seed protein genes in transgenic tobacco. Plant Sci. **62**: 207-216.

Hartings H., Maddaloni M., Lazzaroni N., Di Fonzo N., Motto M., Salamini F. and Thompson R. 1989. The O2 gene which regulates zein deposition in maize endosperm encodes a protein with stuctural homologies to transcriptional activators. EMBO J. 8: 2795-2801.

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

Heim U., Schubert R., Weschke W. and Wobus U. 1991. Seed protein gene promoters from faba bean and oat: specificities of expression in space and time in transgenic tobacco seeds. *in* Abstracts: Third International Congress of Plant Molecular Biology: Molecular biology of plant growth and development, Tucson, Arizona, USA.

Henikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351-359.

Higgins T.J.V., Newbigin E.J., Spencer D., Llewellyn D.J. and Craig S. 1988. The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. Plant Mol. Biol. 11: 683-695.

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: 3213-3221.

Horsch R.B., Fry J.E., Hoffmann N.L., Eichholtz D., Rogers S.G. and Fraley R.T. 1985. A simple and general method for transferring genes into plants. Science 277: 1229-1231.

Hull G.A., Halford N.G., Kreis M. and Shewry P.R. 1991. Isolation and characterization of genes encoding rye prolamins containing a highly repetitive sequence motif. Plant Mol. Biol. 17: 1111-1115.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13-34.

Innis M.A., Gelfand D.H., Sninsky J.J. and White T.J. 1990. PCR protocols. A guide to methods and applications. Academic Press, San Diego.

Jefferson R.A. 1987. Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Rptr. 5: 387-405.

Jofuku K.D., Okamura 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.

Katagiri F. and Chua N.-H. 1992. Plant transcription factors: present knowledge and future challenges. TIG **8**: 22-27.

Kawagoe Y. and Murai N. 1992. Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein *ß-phaseolin* gene conferring spatial and temporal control. Plant J. **2**: 927-936.

Keith B. and Chua N.-H. 1986. Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J. 5: 2419-2425.

Kim S.I., Charbonnier L. and Mosse J. 1978. Heterogeneity of avenin, the oat prolamin. Fractionation, molecular weight and amino acid composition. Biochim. Biophys. Acta **537**: 22-30.

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

Kozak M. 1989. The scanning model for translation: an update. J. Cell Biol. **108**: 229-241.

Kreis M., Shewry P.R., Forde B.G., Forde J. and Miflin B.J. 1985. Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surveys of Plant Mol. and Cell Biol. **2**: 253-317.

Kuhlemeier C. 1992. Transcriptional and post-transcriptional regulation of gene expression in plants. Plant Mol. Biol. **19**: 1-14.

Laemmli U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227**: 680-685.

Leisy D.J., Hnilo J., Zhao Y. and Okita T.W. 1990. Expression of a rice glutelin promoter in transgenic tobacco. Plant Mol. Biol. **14**: 41-50.

Lelièvre J.-M., Oliveira L.O. and Nielsen N.C. 1992. 5'-CATGCAT-3' elements modulate the expression of glycinin genes. Plant Physiol. **98**: 387-391.

Lessard P.A., Allen R.D., Bernier F., Crispino J.D., Fujiwara T. and Beachy R.N. 1991. Multiple nuclear factors interact with upstream sequences of differentially regulated β-conglycinin genes. Plant Mol. Biol. **16**: 397-413.

Luthe D.S. 1987. Storage protein synthesis during oat (*Avena sativa* L.) seed development. Plant Physiol. **84**: 337-340.

Luthe D.S. 1992. Electrophoretic analysis of seed proteins in the dicotyledoneae. Plant Mol. Biol. Rptr. **10**: 254-262.

Maas C., Laufs J., Grant S., Korfhage C. and Werr W. 1991. The combination of a novel stimulatory element in the first exon of the maize *Shrunken-1* gene with the following intron 1 enhances reporter gene expression up to 1000-fold. Plant Mol. Biol. **16**: 199-207.

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 J. 6: 17-22.

Maniatis T., Fritsch E.F. and Sambrook J. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Marcotte Jr W.R., Bayley C.C. and Quatrano R.S. 1988. Regulation of a wheat promoter by abscisic acid in rice protoplasts. Nature **335**: 454-457.

Marris C., Gallois P., Copley J. 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**: 359-366.

Maruyama K., Shands H.L., Harper A.E. and Sunde M.L. 1975. An evaluation of the nutritive value of new high protein oat varieties (cultivars). J. Nutr. **105**: 1048-1054.

Mascarenhas D., Mettler I.J., Pierce D.A. and Lowe H.W. 1990. Intron-mediated enhancement of heterologous gene expression in maize. Plant Mol. Biol. **15**: 913-920.

Matlashewski G.J. 1983. Studies on oat, Avena sativa L.: Part A. Partial characterization of oat lipase. Part B. Structure and biosynthesis of globulin. Ph.D. dissertation. University of Ottawa.

Matzke A.J.M., Stoger E.M., Schernthaner J.P. and Matzke M.A. 1990. Deletion analysis of a zein gene promoter in transgenic tobacco plants. Plant Mol. Biol. **14**: 323-332.

McElroy D., Zhang W., Cao J. and Wu R. 1990. Isolation of an efficient actin promoter for use in rice transformation. Plant Cell 2: 163-171.

Murashige T. and Skoog F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. **15**: 473-497.

Murray E.E., Lotzer J. and Eberle M. 1989. Codon usage in plant genes. Nucl. Acids Res. **17**: 477-498.

Newbigin E.J. and Higgins T.J.V. 1989. Regulation of expression of two 7S storage protein genes of pea (*Pisum sativum* L.) in a transgenic host. Proc. 8th Aust. Biotech. Conf. 104-109.

Newbigin E.J., de Lumen B.O., Chandler P.M., Gould A., Blagrove R.J., March J.F., Kortt A.A. and Higgins T.J.V. 1990. Pea convicilin: structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. Planta **180**: 461-470.

Nielsen N.C., Dickinson C.D., Cho T.-J., Thanh V.H., Scallon B.J., Fischer R.L., Sims T.L., Drews G.N. and Goldberg R.B. 1989. Characterization of the glycinin gene family in soybean. Plant Cell **1**: 313-328.

Ohta S., Mita S., Hattori T. and Nakamura K. 1990. Construction and expression in tobacco of a ß-glucuronidase (GUS) reporter gene containing an intron within the coding sequence. Plant Cell Physiol. **31**: 805-813.

Osborne T.B. 1924. Classification of vegetable proteins. *in* The vegetable proteins, Longmans, Green and Co, New York. pp16-56.

Pandeya R.S. and White F.H. 1984. Delgold: a new flue-cured tobacco. Can. J. Plant Sci. **64**: 233-236.

Pernollet J.-C. and Mossé J. 1983. *in* Seed proteins. Daussant, J., Mossé, J. and Vaughan, J. eds. Academic Press, London. pp155-191.

Pernollet J.-C., Huet J.-C., Galle A.-M. and Sallantin M. 1987. N-terminal sequences of oat avenins compared to other cereal prolamins. Biochimie **69**: 683-689.

Pernollet J.-C., Potier B., Galle A.-M., Huet J.-C., Beauvais F. and Sallantin M. 1989. 2D-HPLC separation, electrophoretic characterization and N-terminal sequences of oat seed prolamins. Phytochemistry **28**: 2565-2570.

Peterson D.M. 1978. Subunit structure and composition of oat seed globulins. Plant Physiol. **62**: 506-509.

Pitts E.G., Rafalski J.A. and Hedgcoth C. 1988. Nucleotide sequence and encoded amino acid sequence of a genomic gene region for a low molecular weight glutenin. Nucl. Acids Res. **16**: 11376.

Potrykus I. 1990. Gene transfer to cereals: an assessment. Bio/Tech. **8**: 535-542.

Putterill J.J. and Gardner R.C. 1989. Initiation of translation of the β-glucuronidase reporter gene at internal AUG codons in plant cells. Plant Science **62**: 199-205.

Rafalski J.A., Scheets K., Metzler M., Peterson D.M. and Hedgcoth C. 1984. Developmentally regulated plant genes: the nucleotide sequence of a wheat gliadin genomic clone. EMBO J. **3**: 1409-1415.

Reeves C.D. and Okita T.W. 1987. Analyses of alpha/beta-type gliadin genes from diploid and hexaploid wheats. Gene **52**: 257-266.

Rerie W.G., Whitecross M. and Higgins T.J.V. 1991. Developmental and environmental regulation of pea legumin genes in transgenic tobacco. Mol. Gen. Genet. **225**: 148-157.

Robert L.S., Matlashewski G.J., Adeli K., Nozzolillo C. and Altosaar I. 1983a. Electrophoretic and developmental characterization of oat (*Avena sativa* L.) globulins in cultivars of different protein content. Cereal Chem. **60**: 231-234.

Robert L.S., Nozzolillo C. and Altosaar I. 1983b. Molecular weight and charge heterogeneity of prolamins (avenins) from nine oat (Avena sativa L.) cultivars of different protein content from developing seeds. Cereal Chem. **60**: 438-442.

Robert L.S., Adeli K. and Altosaar I. 1985a. Homology among 3S and 7S globulins from cereals and pea. Plant Physiol. **78**: 812-816.

Robert L.S., Nozzolillo C. and Altosaar I. 1985b. Characterization of oat (*Avena sativa* L.) residual proteins. Cereal Chem. **62**: 276-279.

Robert L.S., Nozzolillo C. and Altosaar I. 1985c. Homology between legumin-like polypeptides from cereals and pea. Biochem. J. **226**: 847-852.

Robert L.S., Thompson R.D. and Flavell R.B. 1989. Tissue-specific expression of a wheat high molecular weight gene in transgenic tobacco. Plant Cell **1**: 569-578.

Sambrook J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual. second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger F., Nicklen S. and Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463-5467.

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.

Schrickel D.J. 1986. Oats production, value and use. *in* Oats: chemistry and technology. F.H. Webster, ed. AACC St. Paul, Minnesota, USA. pp1-11.

Schubert R., Baeumlein H., Czihal A. and Wobus U. 1990. Genomic sequence of a 12S seed storage protein gene from oat (*Avena sativa* L. cv. Solidor. Nucl. Acids Res. **18**: 377.

Sen P., Chlan C.A., Burow M.D., Lee W.S. and Murai N. 1993. Apical and lateral shoot apex-specific expression is conferred by a promoter of the seed storage protein B-phaseolin gene. Transgenic Res. **2**: 21-28.

Sengupta-Gopalan C., Reichert N.A., Barker R.F., Hall T.C. and Kemp J.D. 1985. Developmentally regulated expression of the bean ß-phaseolin gene in tobacco seed. Proc. Natl. Acad. Sci. USA **82**: 3320-3324.

Shewry P.R. and Tatham A.S. 1990. The prolamin storage proteins of cereal seeds: structure and evolution. Biochem. J. **267**: 1-12.

Shirsat A.H., Wilford N., Croy R.R.D. and Boulter D. 1989. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol. Gen. Genet. **15**: 326-331.

Shotwell M.A., Afonso C., Davies E., Chesnut R.S. and Larkins B.A. 1988. Molecular characterization of oat seed globulins. Plant Physiol. **87**: 698-704.

Shotwell M.A., Boyer S.K., Chesnut R.S. and Larkins B.A. 1990. Analysis of seed storage protein genes of oats. J. Biol. Chem. **265**: 9652-9658.

Somers D.A., Rines H.W., Gu W., Kaeppler H.F. and Bushnell W.R. 1992. Fertile, transgenic oat plants. Bio/Tech. **10**: 1589-1594.

Sonenberg N. 1988. Cap-binding proteins of eukaryoyic messenger RNA: functions in initiation and control of translation. Progress in Nucleic Acid Research and Molecular Biology **35**: 173-207.

Spencer D., Rerie W.G., Randall P.J. and Higgins T.J.V. 1990. The regulation of pea seed storage protein genes by sulfur stress. Aust. J. Plant Physiol. **17**: 355-363.

Steinmetz A. and Weil J.H. 1986. Isolation and characterization of chloroplast and cytoplasmic transfer RNAs. Methods Enzymol. **118**: 212-231.

Sturm A.J., Voelker T.A., Herman E.M. and Chrispeels M.J. 1988. Correct glycosylation, Golgi-processing, and targeting to protein bodies of the vacuolar protein phytohemagglutinin in transgenic tobacco. Planta **175**: 170-183.

Sumner-Smith M., Rafalski J.A., Sugiyama T., Stoll M. and Soell D. 1983. Conservation and variability of wheat alpha/beta-gliadin genes. Nucl. Acids Res. **11**: 3905-3916.

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.

Takaiwa F., Oono K. and Kato A. 1991. Analysis of the 5' flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco. Plant Mol. Biol. **16**: 49-58.

Thomas M.S. and Flavell R.B. 1990. Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell **2**: 1171-1180.

Ueng P., Galili G., Sapanara V., Goldsbrough P.B., Dube P. and Larkins B.A. 1988. Expression of a maize storage protein gene in petunia plants is not restricted to seeds. Plant Physiol. **86**: 1281-1285.

Vancanneyt G., Rosahl S. and Willmitzer L. 1990. Translatability of a plant-mRNA strongly influences its accumulation in transgenic plants. Nucl. Acids Res. **18**: 2917-2921.

Varenne S., Buc J., Lloubes R. and Lazdunski C. 1984. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. **180**: 549-576.

Vellenoweth R.L. and Okita T.W. 1993. Analysis of nuclear proteins interacting with a wheat a/ß-gliadin seed storage protein gene. Plant Mol. Biol. **22**: 25-41.

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.

Wain-Hobson S., Nussinov R., Brown R.J. and Sussman J.L. 1981. Preferential codon usage in genes. Gene. **13**: 355-364.

Walburg G. and Larkins B.A. 1986. Isolation and characterization of cDNAs encoding oat 12S globulin mRNAs. Plant Mol.Biol. **6**: 161-169.

Walden R. and Schell J. 1990. Techniques in plant molecular biology - progress and problems. Eur. J. Biochem. **192**: 563-576.

Wen T.-N. and Luthe D.S. 1985. Biochemical characterization of rice glutelin. Plant Physiol. **78**: 172-177.

Williamson J.D., Galili G., Larkins B.A. and Gelvin S.B. 1988. The synthesis of a 19 kilodalton zein protein in transgenic *Petunia* plants. Plant Physiol. **88**: 1002-1007.

Wu Y.V., Sexson K.R., Cavins J.F. and Inglett G.E. 1972. Oats and their dry-milled fractions: protein isolation and properties of four varieties. J. Agric. Food Chem. **20**: 757-761.