Order Number 9312222

Expression of bean arcelin-1 and β-phaseolin genes in seeds of common bean and transgenic tobacco

Wang, Ming-Li, Ph.D.

University of Hawaii, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

EXPRESSION OF BEAN ARCELIN-1 AND β-PHASEOLIN GENES IN SEEDS OF COMMON BEAN AND TRANSGENIC TOBACCO

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BOTANICAL SCIENCES
(PLANT PHYSIOLOGY)

DECEMBER 1992

By

Ming-Li Wang

Dissertation Committee:

Samuel S. M. Sun, Chairperson
H. Michael Harrington
Tom Humphreys
Morton Mandel
John I. Stiles

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor, Dr. Sam Sun, for his guidance and support during the development of this dissertation, and to the other members of my committee, Dr. Mike Harrington, Dr. Tom Humphreys, Dr. Morton Mandel, and Dr. John Stiles, for their valuable comments.

The emotioanl and material support of my parents, Tong-Hsiang and Dr. Hsi-Hua Wang, and the patience and understanding of my fiancee, Dave Robichaux, have been the greatest help to finish my graduate study.

Special thanks to my three BIG brothers, Heinrich Albert, Stefan Moisyadi, and Wayne Borth, for their academical and emotional support during the past four years, and to Dr. Piera Sun, for her kind advice in many aspects. Also thanks to all the other members of the lab and friends who deserve recognition of their support. Mahalo!

ABSTRACT

A gene, PHVARC1-1, encoding the insecticidal protein arcelin-1 has been isolated from the common bean (*Phaseolus vulgaris*, line SARC1-7). It contains no intron and is 2692 bp long, including 1167 bp in the 5' upstream region, 798 bp in the coding region, and 727 bp in the 3' downstream region. The coding region of PHVARC1-1 is identical to those of pARC1-11, (Osborn et al, 1988b) and Phvarc1a (Anthony et al, 1991).

Arcelin and phaseolin are the major seed storage protein in domestic beans. The expression of these genes has been compared at the mRNA and transcriptional levels in the common bean plants (*Phaseolus vulgaris*). Northern blot analysis revealed that the onset of arcelin mRNA accumulation coincides with a reduction in the phaseolin mRNA level. This reduction at least partly due to a decrease in the transcription rate of phaseolin genes, as shown by nuclear run-on assays. Potential gene interaction mechanisms are discussed.

The expression of these two genes was also examined in transgenic tobacco plants. The arcelin and/or phaseolin gene(s) under the control of their respective or alternative promoters were introduced into tobacco (*Nicotiana tabacum* cv Xanthi.nc) via the *Agrobacterium*/Ti plasmid vector system. All of these transgenes were

expressed in the transformed tobacco seeds. The accumulation of the phaseolin mRNA in the developing tobacco seeds peaked about 2 days earlier than the arcelin mRNA. However, the changes at the mRNA level observed for the two genes in the common beans were not observed in the transgenic tobacco plants. The levels of arcelin mRNA driven by either promoter were lower than those of the phaseolin mRNA. The possible causes of this finding are discussed.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ABBREVIATIONS xi

CHAPTER I. INTRODUCTION 1
Literature Review 2
Phaseolin 3
Phaseolin genes and their regulation 3
Lectin 11
Roles of lectins in plant/insect interaction 12
Lectin genes and their regulation 14
Biosynthesis, processing, transport, and localization of seed proteins 19
Arcelin 21
Insecticidal activity of arcelin 25
Arcelin genes 27
Phenomena of gene interaction 28
Regulatory elements of bean seed storage protein genes 29
Post-transcriptional regulation of seed storage protein gene expression . . . 32
Hypothesis 35

CHAPTER II. CLONING AND EXPRESSION OF A BEAN ARCELIN-1 GENE IN COMMON BEAN PLANTS 37
Introduction 37
Materials and Methods 38
Chemicals 38
Plant materials 38
Protein extraction and SDS-PAGE 38
Plasmids and Synthetic Oligonucleotide . . 38
Southern analysis 39
Construction and screening of the genomic library 39
Subcloning and sequencing of positive clones 40
RNA extraction 41
Primer extension 41
Northern analysis 42
In vitro translation 43
Nuclei isolation 43
RNA synthesis in isolated nuclei 44
Detection of arcelin-1 and phaseolin transcripts 45

Results and Discussion 46
Protein profiles of different bean lines. . 46
Genomic Southern analysis and gene copy number estimation 47
Construction and screening of genomic library 50
Sequence of arcelin-1 gene 52
Coding and 3' region 55
Transcription start site and TATA box 56
5' promoter region 58
Expression of arcelin-1 and phaseolin gene during seed development 59
Accumulation of mRNA 59
Relative transcription rate 62
Possible mechanisms of gene interaction 66

CHAPTER III. EXPRESSION OF THE ARCELIN-1 AND THE β-PHASEOLIN GENES IN TRANSGENIC TOBACCO SEEDS . . 72
Introduction 72
Materials and Methods 73
Chemicals 73
Plant materials 73
Bacterial strains 74
Plasmid construction 74
Site-directed mutagenesis 75
Triparental mating between *Agrobacterium* and *E. coli* 79
Tobacco transformation and regeneration . . 81
GUS assay 82
DNA extraction 82
Southern analysis 83
RNA extraction 83
Northern and dot blot analysis 83
Results and Discussion 83
Tobacco transformation and regeneration . . 83
Southern analysis and gene copy number estimation 84
Expression of the arcelin-1 and β-phaseolin genes during seed development 90

CHAPTER IV. CONCLUSION 100

REFERENCES 103

LIST OF TABLES

Table		Page
1	Properties of phaseolin, lectin, and arcelin	23
2	Legume seed storage protein genes and their regulatory elements	31
3	Genomic library of common bean, *Phaseolus vulgaris*, line SARC1-7	50
4	Summary of nuclear run-on assays	63
5	Codon usage of phaseolin (P), arcelin (A), and lectin (PHA-L & PHA-E) genes	70
6	Summary of tobacco transformation - 1	86
7	Summary of tobacco transformation - 2	86

LIST OF FIGURES

Figure		Page
1	Amino acid sequence comparison of arcelin and lectin genes	22
2	SDS-PAGE of total seed storage proteins from common bean lines	47
3	Southern blot analysis of bean (SARC1-7) genomic DNA	48
4	Southern blot analysis of potential arcelin-1 clones	51
5	Complete nucleotide sequence of PHVARC1-1	53
6	Schematic comparison of three arcelin-1 gene family members	55
7	Determination of transcription start (CAP) site	57
8	PHVARC1-1 transcription start site vs. plant consensus	57
9	Northern blot analysis of developing bean seeds	60
10	*In vitro* translation of RNA from developing bean seeds	61
11	Nuclear run-on assays	64
12	Relative transcription rates of phaseolin and arcelin genes during development	65
13	Arcelin and phaseolin gene translation start site vs. plant consensus	68
14	Strategy to create HindIII sites at the 3' ends of the inserts containing the phaseolin and/or arcelin genes	76
15	Restriction maps of the inserts of A, pA3, pP7, & pPA3; B, pBI121; and C, pPA2.5-3, pAP4.0-6, & pT-2-3	77
16	Oligonucleotides for site-directed mutagenesis	78

Figure Page

17 GUS staining of wild-type and pA3 transformed tobacco leaves.....85

18 Southern blot analysis of tobacco genomic DNA digested with HindIII.....87

19 Southern blot analysis of pPA3 and pT-2-3 transformants.....88

20 Southern analysis of transgenic tobacco genomic DNA.....89

21 Northern blot analysis of RNA from 20-DAF tobacco seeds - 1.....91

22 Northern blot analysis of RNA from 20-DAF tobacco seeds - 2.....92

23 Comparison of Northern and dot blot analyses of developing seeds from P7-19.....93

24 Developmental expression of the arcelin or phaseolin gene in single-gene transformed plants.....94

25 Developmental expression of the arcelin and phaseolin genes in the 2-gene transformed plants.....95

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
cfu	colony forming unit
CIP	calf intestinal phosphatase
CPM	count per minute
CTAB	cetyltrimethylammonium bromide
DAF	day after flowering
DEPC	diethylpyrocarbonate
DMF	N,N-dimethyl formamide
DMSO	dimethylsulfoxide
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenedinitrilo-tetraacetic acid
ER	endoplasmic reticulum
GUS	xi-glucuronidase
IEF	isoelectric focusing
kb	kilobase
kD	kilodalton
LB	left border (of T-DNA in pBI121) or Luria-Bertani (medium)

M-MLV	Moloney murine leukemia virus
mRNA	messenger RNA
M_r	relative molecular weight
MS	Murashige and Skoog (medium)
M.W.	molecular weight
NOS	nopaline synthase
NPTII	neomycin phosphotransferase II
NRB	nuclear resuspension buffer
NRS	negative regulatory sequence
NTP	ribonucleotide triphosphate
$O.D._{660}$	optical density at 660 nm
PAGE	polyacrylamide gel electrophoresis
pfu	plaque forming unit
PHA	phytohemagglutinin
pI	isoelectric point
poly(A)	poly-adenylate
RB	right border of T-DNA
RNase	ribonuclease
SDS	sodium dodecyl sulfate
TCA	trichloroacetic acid
T-DNA	transfer DNA
Ti plasmid	Tumor-inducing plasmid
Tris	Tris[hydroxymethyl]aminomethane
UAS	upstream activating sequence
X-gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid

CHAPTER I. INTRODUCTION

Over millions of years of evolution under the threats of herbivorous predators, plants have developed some effective defense mechanisms to minimize the damages. These defense systems include physical protection, such as impenetrable surfaces and sharp spines, and chemical protection, such as certain proteins and secondary metabolites that are toxic to predators. Insects are one of the most important plant predators, especially to crop plants of economic importance. A wide variety of insecticides, most of them chemically synthesized, have been used to control insect damage in agriculture. The application of synthetic insecticides is often costly and causes environmental hazards. Biological controls, particularly those which use the natural defense systems of plants, may provide a viable alternative to chemical insecticides. Conventional breeding and genetic engineering allow the introduction of agronomically desirable traits, such as insect resistance, into crops. However, it is essential to first understand the basis of the traits of interest.

The common bean (*Phaseolus vulgaris*) is an important dietary protein source for humans and livestock throught the world. In the tropics, 50 to 70% of the beans are

often damaged by bean weevils during storage (Taylor, 1981). Domestic bean cultivars have very low levels of resistance to the Mexican bean weevil (*Zabrotes subfasciatus*), the most important bean pest (Schoonhoven and Cardona, 1982), whereas high levels of resistance have been observed in the wild accessions (Schoonhoven et al, 1983). This resistance is associated with the presence of a novel seed storage protein, arcelin (Osborn et al, 1988a). The arcelin genes have been introduced into domestic bean lines by breeding (Romero Andreas et al, 1986). In all arcelin containing beans, the content of phaseolin, the major seed storage protein in domestic beans, was greatly reduced.

I am interested in the interrelation between the expression of the arcelin and phaseolin genes, and propose to analyze the expression of these storage protein genes at the molecular level.

Literature Review

A seed storage protein may be defined as any protein accumulated in significant quantities in developing seeds that is rapidly hydrolyzed during germination to provide nitrogen for the early stages of growth (Higgins, 1984). These proteins are tissue specific and developmentally regulated. Three main types of seed storage proteins have

been identified and characterized in the common bean, phaseolin, lectin, and arcelin.

Phaseolin. Phaseolin is the major seed storage protein of cultivated beans. It represents 40 to 60% of the total protein (Osborn, 1988). Phaseolin is a globulin which is completely soluble in 0.5 M NaCl at all pHs. Several names have been assigned to this protein, including vicilin, glycoprotein II, globulin-I, euphaseolin, and α-globulin, but they are all identical. The most extensively studied phaseolin variant (T-type) consists of 3 subunits, α, β, and γ (γa and γb). The corresponding molecular weights of the subunits are 51, 48, and 45.5 kD.

The accumulation of phaseolin during French bean embryo development was analyzed by electrophoresis, rocket immunoelectrophoresis, *in vitro* translation of polysomes, and Western blots (Sun et al, 1978). No phaseolin was detected before the seeds reached 7 mm in length (cotyledon, 3 mm) or before 12 days after flowering (DAF). However, phaseolin was synthesized and accumulated during the following 14 days of development.

Phaseolin genes and their regulation. Phaseolin genes were found to be preferentially sensitive to DNase I in cotyledons relative to leaves (Murray and Kennard, 1984). However, no DNase hypersensitive sites could be identified in the flanking regions of chromatin. The nucleosomal repeat length is shorter (177 bp) in the developing

cotyledons than in the expanding leaves (191 bp). It was debated that the difference in spacing was due to DNA duplication rather than active transcription. Polyploidy is common in plant storage tissues. Since storage protein synthesis often follows the onset of polyploidy, it was argued that the change in chromatin structure resulting from DNA synthesis could, in part, be involved in the regulation of storage protein gene expression.

Phaseolin mRNA accumulation during embryo development was shown by Northern analysis (Chappell and Chrispeels, 1986). The transcription rate, determined by nuclear run-on experiments, was high in the early stages of development and declined to very low rates at the later stages. In the cultivar Greensleeves, phaseolin genes have higher trasncription rates than lectin genes (PHA-L). By comparing the transcription rates and the relative amounts of mRNA, the authors concluded that the storage protein mRNA is very stable, with a half life of several days.

A phaseolin cDNA clone was isolated by hybridization with a cDNA library prepared from 16S mRNA, which is essentially phaseolin mRNA in developing cotyledons, and confirmed by hybrid-arrested translation (HART) (Hall et al, 1980). This clone was later used to screen a genomic library (Sun et al, 1981).

A β-phaseolin gene containing 80 bp of 5' untranslated region, 1,263 bp of coding sequence (interrupted by 5

introns ranging from 72 bp to 128 bp in length), and 135 bp of 3' untranslated region was isolated (Sun et al, 1981 and Slightom et al, 1983). Three TATA boxes are located at positions -28, -37, and -39 (Hall et al, 1983). A CCAAAT box is present at position -74. A poly(A) addition signal, AATAAA, is also present in the 3' region, 16 bp before the first nucleotide of poly(A) tail. The deduced amino acid composition is similar to that of the phaseolin β subunit. The hydrophobic region at the N-terminus (about 21 to 26 amino acids in length) probably serves as a signal peptide for co-translational insertion into the ER. Two possible glycosylation sites, Asn-Leu-Thr (amino acids 251-253 or nucleotides 1,115-1,123) and Asn-Phe-Thr (amino acids 340-342 or nucleotides 1,510-1,518) were identified (Hall et al, 1983).

Sequence data from 8 partial and five complete phaseolin cDNAs revealed that phaseolin polypeptides are encoded by two unique gene families, which differ in their coding regions by the presence or absence of two direct repeats (Slightom et al, 1985). The α-phaseolin genes contain direct repeats that encode 14 extra amino acids. Otherwise, there is a 98% homology at the nucleotide level between the α- and β-type genes. The difference in the sizes of the polypeptides is mainly due to post-translational processing, namely glycosylation. Within a family, there are more base substitutions in the α-type

genes. No nucleotide substitutions are found in the 5' untranslated region, signal peptide, or the first 59 amino acids. The nucleotide changes among intra- and intergenic family members appear to favor amino acid replacement rather than silent substitution. This suggests that there are parts of the polypeptide that can be engineered without adversely effecting its function. Hydropathy analysis indicated that the phaseolin polypeptides are predominately hydrophilic, and that the two potential N-glycosylation sites are located in different hydropathic environments.

The β-phaseolin gene has been expressed in transgenic yeast cells, sunflower tissues, and tobacco seeds. In sunflower tissue culture cells, phaseolin was expressed under the control of its own promoter or octopine synthase gene promoter (Murai et al, 1983). The transcripts were correctly spliced and the protein was detected by phaseolin antiserum.

The obligatory sequence, TACTAAC, for proper intron splicing in yeast is not present in the phaseolin gene. Consequently, when the entire phaseolin genomic sequence (3.8 kb, including 863 bp 5'-, 1700 bp coding, and 1226 bp 3'- region) was transformed into yeast cells, the resulting transcripts were too small and no phaseolin could be detected by immunoprecipitation (Hall et al, 1984 and Cramer et al, 1985 and 1986). When the introns were eliminated by substituting the coding region with a cDNA

sequence, a full length β-phaseolin was synthesized, and the protein underwent glycosylation and probable signal peptide cleavage.

The same phaseolin "minigene" (construct without introns) was also transferred into tobacco via Ti-plasmid/*Agrobacterium tumefaciens* (Chee et al, 1986 and Chee et al, 1991). The results of Northern analyses and S1 nuclease mappings provided evidence that the phaseolin introns and intron splicing are not necessary for correct spatial and temporal expression, and that no alternative splicing site was introduced by the removal of the five introns.

The glycosylation sites (Asn^{252} and Asn^{341}) of the phaseolin "minigene" were modified (to Gly^{252} and Ser^{341}, respectively) and the expression of these constructs was determined in transgenic tobacco plants (Bustos et al, 1991a). These modifications did not change the phaseolin mRNA level. However, reduced levels (41 and 73% of what were expressed in the wild-type gene transformants) of full-length protein were detected in the protein bodies of the plants transformed with these 2 mutant constructs. Thus, glycosylation is apparently involved in maintaining the stability of the full-length protein but not the targeting. The 3.8-kb fragment used for the yeast transformation was inserted into the A66 Ti plasmid of *A. tumefaciens* and used to transform tobacco (Sengupta-Gopalan

et al, 1985a and b). The seeds of the transformed plants contained phaseolin mRNA of the expected size, indicating correct excision of introns and addition of poly(A). Phaseolin was synthesized, properly processed, and accumulated in the protein bodies of embryonic cells, the storage tissue of beans, and in some endosperm cells, which is the storage tissue for tobacco (Greenwood and Chrispeels, 1985b). The timing of accumulation was also similar to that for beans. However, a significant portion of the phaseolin was degraded into smaller peptides, that are usually formed during germination of bean seeds.

In an attempt to increase the methionine codons (from 3 to 9), a 45-bp synthetic duplex containing 6 methionine codons was inserted into the *Xba*I site in the third exon of the phaseolin gene (Hoffman et al, 1988). The modified gene was introduced into tobacco via the Ti-plasmid/*Agrobacterium* system. Although a similar levels of transcripts was detected in the plants transformed with either the normal gene or the modified gene, the level of modified protein was only 0.2% of the normal protein. As the normal phaseolin, the modified protein was glycosylated, assembled into a trimer, and expressed in the correct temporal and organ specific fashion. However, it was localized in the ER and Golgi apparatus, but not in the protein bodies. These results indicated that the modified protein was degraded in the Golgi complex or protein

bodies. The met-rich peptide was inserted into a conserved region among phaseolin genes (Slightom et al, 1985). This may have disrupted the proper conformation and function of phaseolin. This argument was supported by the studies of the 3-dimensional structure of phaseolin (Lawrence et al, 1990): the *Xba*I site is located within a helix motif (the first helix in the helix-turn-helix motif that is similar to that of a DNA binding domain). An insertion in this region may distort the tertiary and/or quaternary structure of the protein.

By injecting wild-type and C-terminal truncated phaseolin RNA's into *Xenopus* oocytes, Ceriotti and co-workers (1991) showed that phaseolin trimerization is dependent on the level of protein synthesis and is necessary for intracellular transport. The C-terminal, which interacts with the N-proximal α-helix of the adjacent subunitdomain, is crucial for assembly (Lawrence et al, 1990).

Reporter genes, such as luciferase (Riggs et al, 1989) and β-glucuronidase (GUS) (Bustos et al, 1989), have been used to analyze the phaseolin promoter. In both studies, engineered Ti plasmids containing the chimeric reporter genes were transferred into *A. tumefaciens* and transformed into tobacco leaf cells. Correct temporal and spatial expression of the reporter genes were observed in the regenerated plants.

The binding of a protein factor present in the nuclear extract of immature embryos to a phaseolin gene upstream region (-628 to -682), containing two inverted A/T rich motifs, has been shown by gel retardation and foot printing experiments (Bustos et al, 1989). Fusion of a 103-bp fragment or a 55-bp synthetic oligonucleotide, containing these motifs in either direction, to a minimal 35S promoter and GUS gene resulted in GUS expression in seeds, roots, hypocotyls, and cotyledons of seedlings. These results suggest that the upstream A/T rich region that binds nuclear protein can activate transcription *in general*. Additional elements may be required for embryo-specific gene expression. The same binding activity has been demonstrated by using nuclear extract from somatic carrot embryos (Guiltinan et al, 1989). In this case, however, the active protein factor was found only in the embryonic tissues. In other words, the binding activity is modulated by embryogenesis. These results indicate that the cis- and trans-acting elements controlling gene expression have been conserved during evolution but that the function of these elements may not be completely the same in different plants. The phaseolin promoter was further dissected by transient and stable transformation assays using GUS as the reporter gene (Bustos et al, 1991b). Two upstream activating sequences (UAS1 and UAS2) and two negative regulatory sequences (NRS1 and NRS2) were identified. UAS1

(-295 to -109) was sufficient for seed-specific expression and required for temporal control. UAS2 (-468 to -391) expanded the expression to hypocotyl. Deletion of either NRS's (-391 to -295, and -518 to -418) resulted in early GUS expression, suggesting that the 2 NRS's are involved in the temporal control of gene expression during seed development.

Lectin. Lectins, or phytohemagglutinins (PHA), are proteins of plant origin which have the ability to agglutinate red blood cells (Lierner, 1976). Although some bean lines lack seed lectins, they constitute 6 to 12% of the total seed protein in many varieties (Osborn, 1988). The major bean lectin is comprised of five glycoproteins that are isomeric tetramers with varying proportions of two different subunits (L_4, L_3E_1, L_2E_2, L_1E_3, and E_4, Miller et al, 1975). The L subunit is a potent leucoagglutinin, whereas E is an erythroagglutinin. There are striking similarities between these two peptides. They have similar molecular weights (34 and 36 kd), and similar amino acid and sugar compositions. Both lack methionine and cysteine (Leavitt et al, 1977). However, they are only partially identical by immunochemical criteria. The isolectins were isolated from the albumin fraction with pIs ranging from 4.6 to 5.2, while the globulin isolectins contain several more lectins with higher pIs (Pusztai and Stewart. 1978). Two-dimensional gel electrophoresis (Osborn et al, 1983)

revealed that there is a greater complexity for the lectin polypeptide composition than the previous model (Miller et al, 1975). A total of five instead of two polypeptides were detected. However, the agglutinating activities are identical. Lectins are encoded by a family of genes that are tightly linked (Brown et al, 1981).

Roles of lectins in plant/insect interaction. Lectins are widely distributed in the plant kingdom. Their nutritional value is important because they constitute a large portion of the proteins in seeds of legume, that are a major dietary protein source. Lectins probably play a role in plant defense (Chrispeels and Raikhel, 1991). There are several ways these vacuolar proteins can get in contact with molecules from the attacker: they are released during imbibition as well as after cellular structures are disrupted by predators or fugal hyphae.

The lectins of *Phaseolus vulgaris* have been shown to be toxic to animals, such as Japanese quails (Jayne-Williams and Burgess, 1974), rats (Evans et al, 1973, King et al, 1980 and 1982, and Wilson et al, 1980), and bruchid beetles (*Callosobruchus maculatus*) (Janzen and Juster, 1976, and Gatehouse et al, 1984).

It is well known that lectins agglutinate red blood cells and initiate mitosis of human leukocytes (Liener, 1976). The inclusion of high-lectin 'Processor' beans in the diet for rats causes severe disruption and abnormal

development of intestinal microvilli and a dramatic overgrowth of bacteria, thus interfering with the digestive and absorptive processes (King et al, 1980 and Wilson et al, 1980). Ultrastructural observation revealed the intracellular changes associated with microvilli disruption, including disorganization of the terminal web, swelling of the apical cytoplasm, and an increase in the number of lysosomes. This suggests a disturbance in membrane-associated transport processes (King et al, 1982). None of the above phenomena was observed in rats fed with diets containing Pinto III beans, which contains no lectin.

In a feeding trial comparing the effect of bean lectins and cowpea trypsin inhibitor on a bruchid beetle (*C. maculatus*), the latter was found to have virtually no effect on the bruchid larvae (Janzen and Juster, 1976). However, reduced number of the larvae survived on the diets containing 0.1% (dry weight) of bean lectin. No larvae survived to adulthood on the diets with 5% lectin. Bruchid beetles or weevils (Bruchidae), the major insect pests of beans, are extremely prey-specific. The authors hypothesized that the bean lectins are of adaptive significance in being one of the main reasons why these insects cannot prey on every species of beans. Lectins from some other plants, including peanut agglutinin, osage orange lectin, potato lectin, jimson weed lectin, and wheat germ agglutinin, have also been demonstrated to have an

anti-metabolic effect on *C. maculatus* (Murdock et al, 1990). These lectins causes significant delay in larval development. No effect was observed among the other 12 lectins tested. The results indicate that the physiological/biochemical systems in *C. maculatus* are vulnerable to selected plant lectins. Feeding trials using different bean lectins suggest that the antimetabolite is the E-type phytohemagglutinin (PHA-E, Gatehouse et al, 1984). Immunofluorescence investigations show that the lectin binds to midgut epithelial cells. It was suggested that the mechanism of lectin toxicity is analogous to what occurs in the rats, namely that lectin causes disruption of epithelial cells of the larval midgut leading to interference with nutrient absorption and transport.

Lectins may also act as digestive enzyme inhibitors. At least one lectin has been identified as an α-amylase inhibitor for insect and mammalian α-amylases but not plant α-amylase (Moreno and Chrispeels, 1989). This inhibitor, and not PHA, is the major source of resistance in common beans to cowpea weevils (*C. maculatus*, Huesing et al, 1991). Since the precursor of this α-amylase did not show inhibitory activity, proteolytic processing and/or glycan modification in the Golgi are probably necessary for the acquisition of the inhibitory activity.

Lectin genes and their regulation. A lectin cDNA clone was isolated from *P. vulgaris* cv. Tendergreen and the

complete sequence was determined (Hoffman et al, 1982). It contains 4 AUG codons in the 5' region, but only the most 3' one falls into the favored initiator sequence, A/CXXAUGG. the first 3 are within the CATGAATG (RY) repeat. There are 5 possible N-glycosylation sites (Asp-X-Ser or Thr) in the coding region and 2 overlapping polyadenylation signals (AATAAA) in the 3' region. The newly synthesized protein consists of a signal sequence of about 22 amino acid residues in length and a mature protein of 223 residues. The lectin transcript (1.1 kb) starts to accumulate in cotyledons by the 9-mm stage. The amount of this transcript increases until about the 16-mm stage, and dramatically decreases during further maturation.

This cDNA clone was used to screen a λ1059 genomic library of the same bean line (Hoffman and Donaldson, 1985). One of the λ clones isolated had 2 complete lectin genes in the same orientation about 4 kb apart. According to the deduced amino acid sequences, these two genes, *dlec1* and *dlec2*, encode erythro- and leucoagglutinating phytohemagglutinins (PHA-E and PHA-L), respectively. The coding regions of these 2 genes are 90% homologous, suggesting that they may have evolved by tandem duplication of a single ancestral gene. Both genes are intronless and have a short 5' untranslated sequence (11-14 bp), as estimated by S1 mapping. Most of the 5' untranslated regions contain alternating pyrimidines and purines, or RY

repeats (CATGCATG), which is also present in other legume seed protein genes (Dickinson et al, 1988). The sequences further upstream are also homologous and include more RY repeats. Both genes have overlapping double (*dlec1*) or triple (*dlec2*) polyadenylation signals and share 84% homology in the 3' untranslated regions. Both genes are expressed in bean cotyledons, as indicated by hybridization experiments using gene specific oligonucleotide probes.

Seeds of the bean cultivar Pinto III and Pinto U1 III contain very low levels of PHA-L and no PHA-E can be detected (Vitale et al, 1985, Staswick et al, 1986). The levels of PHA mRNAs and the transcription rates in Pinto III were significantly lower than those in Greensleeves, a normal bean line (Chappell and Chrispeels, 1986). In contrast, the expression of phaseolin genes is similar in both lines. A low level of PHA transcription in Pinto III was measurable but no mRNA was detected, indicating that the PHA deficiency is due to a reduced transcription rate and possibly, to instability of the mRNA.

Pinto and normal cultivars contain about 5 lectin genes (Staswick et al, 1986). The deficiency of PHA may be due to a mutation that has a coordinate negative effect on the small multigene family. In soybean, the insertion of a transposable element (TGM1) into the coding region of a lectin gene resulted in inactivation of gene expression (Rhodes and Vodkin, 1985). Two PHA genes from Pinto

phenotype (U1 III), *Pdlec1* and *Pdlec2*, were cloned and sequenced (Voelker et al, 1986). *Pdlec1* contains a 1-bp frameshift mutation close to the 5' end of the coding region, resulting in no PHA-E expression. *Pdlec2* codes for a polypeptide very similar to PHA-L, encoded by *dlec2*. There is a 114-bp deletion in the upstream non-coding region of *Pdlec2* compared with *Pdlec1*, *dlec1*, and *dlec2*. This region contains 2 almost perfect repeats of 56 and 63 bp with each copy carrying an inverted RY repeat and has enhancer-like activity. It is not known whether the same mutations are also present in the other copies of lectin genes.

The differences in expression between *dlec2* and *Pdlec2* are maintained in transgenic tobacco (Voelker et al, 1987). In *dlec2* transformed plants, PHA-L is mainly in the embryos and much less in the endosperm at the same stage of seed protein accumulation. The protein is correctly glycosylated, processed in the Golgi apparatus, and targeted to the protein bodies (Sturm et al, 1988).

Very little PHA-L accumulats in the *Pdlec2* transformed tobacco, probably due to the absence of the cis-element present in the 114-bp deletion (Voelker et al, 1987). Furthermore, Southern analysis of *Bam*HI/*Sst*I digested DNA from *Pdlec2* transformed plants showed that in addition to the correct size (2.7kb) fragment, a high molecular weight band also hybridizes to the probe. This might be due to

gene rearrangement or to partial digestion resulting from the sensitivity of these restriction enzymes to methylation. The latter case may suggest a role of DNA methylation in *Pdlec2* regulation. The methylation status of DNA from bean cotyledons, leaves, stems, and roots, was examined by Southern analysis, using the PHA gene, a "lectin-like" protein gene, and the phaseolin gene as probes and by methylation sensitive restriction enzymes (Riggs and Chrispeels, 1988). DNA fragments containing the PHA genes are undermethylated in cotyledons, where the genes are expressed, but the same sites are methylated in other tissues.

The effect of deletions of the *dlec2* upstream sequence on the level of PHA-L expression was examined in transgenic tobacco (Riggs et al, 1989). A negative regulatory element is located between -1000 and -675 and two positive elements lie between -550 and -345, and -345 and -125. The two positive elements increase the expression by 200 fold. No binding activity was detected for the positve element between -345 and -125, a RY repeat. However, the other two regions bind cotyledon nuclear proteins from mid-mature seeds, as shown by gel retardation assays. The binding activity of an AT-rich region between -550 and -345 is competed by upstream sequences from phaseolin, soybean lectin, and Kunitz trypsin inhibitor genes, suggesting that these legume seed protein genes may be regulated by

evolutionally conserved protein/DNA interactions. However, the binding of activator(s) at this site is not sufficient for high-level expression. *Pdlec2* from the PHA deficient line also contains this AT-rich sequence but the RY repeats are deleted (Voelker et al, 1987), indicating that both elements are required. The putative PHA silencer is also AT rich (Riggs et al, 1989). Binding of this element was localized within a 269-bp *Rsa*I fragment, which includes a 70-nucleotide region having the potential to form a stem and loop structure.

Biosynthesis, processing, transport, and localization of seed proteins. Techniques such as wheat germ (Sun et al, 1978 and Hall et al, 1977 and 1978) and *Xenopus* oocyte (Matthews et al, 1981) cell-free systems for *in vitro* translation of poly(A) RNA, and free and membrane-bound polysomes (Bollini and Chrispeels, 1979), *in vivo* labelling (Chrispeels, 1983), polysome run-off (Bollini et al, 1983 and Vitale et al, 1984), immunodetection, and electronmicroscopy (Greenwood and Chrispeels, 1985a, Greenwood et al, 1984, and Manen and Pusztai, 1982) have been used to investigate the synthesis, processing, transport, and localization of phaseolin and lectins. In summary, the polypeptides are synthesized on rough ER. Signal peptide cleavage and glycosylation are both co-translational events. Transport is mediated by the Golgi apparatus. Protein bodies, ER or tonoplast derived

membrane bound organelles, are the deposition sites (Chrispeels, 1984). There are common steps in the transport of proteins in plant and animal cells prior to the Golgi apparatus, although the targeting mechanisms are apparently different (Bustos et al, 1988, Bassüner et al, 1983, and Voelker et al, 1986). The mechanisms are, however, conserved between yeast and plant (Hall et al, 1984 and Tague and Chrispeels, 1987).

The signal peptide of PHA-L is not sufficient to direct protein to the protein bodies in transgenic tobacco (Dorel et al, 1989 and Hunt and Chrispeels, 1991). When glycosylation was inhibited by tunicamycin in beans (Bollini et al, 1985) or by the mutations of the glycosylation sites in transgenic tobacco (Voelker et al, 1989, and Herman et al, 1989), PHA-L was still transported to the protein bodies. Glycans probably function to prevent proteolysis and to maintain protein stability (Bustos et al, 1991a). The N-terminal region of the mature PHA-L is sufficient to direct invertase to the vacuole (Tague et al, 1991). The C-terminal propeptide is necessary and adequate to sort barley lectin to the vacuole (Bednarek and Raikhel, 1991). However, multiple protein sequence domains are probably required for the sorting of legumin in pea (Saalbach et al, 1991). In different plant species, the sequences recognized by the sorting

mechanism(s) may be present in different parts of the proteins.

Arcelin. A novel seed storage protein in the common bean, arcelin, was revealed by SDS-PAGE of proteins from the seeds of a wild, bruchid-resistant bean line (Romero Andreas et al, 1986). It was named after Arcelia, the town in the Guerrero, Mexico near where the wild bean (PI 325690) was collected (Gentry, 1969). This protein has toxic effects on one of the most important bean bruchid pest, *Zabrotes subfasciatus* (Osborn et al, 1988a).

When subjected to two-dimensional gel electrophoresis, four arcelin variants (35 to 42 kD), arcelin-1, 2, 3, and 4, with different electrophoretical mobilities from phaseolin and lectin were observed (Osborn et al, 1986). These variants were originally isolated from 4 different wild bean accessions. These polypeptides had more basic pIs than phaseolin. The alleles for the four variants are co-dominant.

Arcelins share many common properties with lectins. These include: similar subunit molecular weight, deglycosylated subunit molecular weight, amino acid composition, and timing of accumulation during development (Osborn et al, 1988b). They are also related antigenically. High degrees of sequence homology between arcelin and lectin also occur at both nucleotide (80%) and amino acid levels (60%, Figure 1). However, arcelins have

```
                1                    25                    50
                .                    .                     .
Arcelin-1       MASSNLLTLALFLVLLTHANSSNDASFNVETFNKTNLILQGDATVSSEGH
Arcelin-2       ..................................................
Lectin-Like     ....K..S.....A..S....ATET..IIDA.................--
PHA-L           .......S............ASQT...SFQR..E......R......K.Q
PHA-E           ....-KFFTV..............IY..FQR..E......R.S...S.Q

                51                   75                    100
                .                    .                     .
Arcelin-1       LLLTNVKGN---EEDSMGRAFYSAPIQINDRTIDNLASFSTNFTFRINAK
Arcelin-2       ..................................................
Lectin-Like     ---NGNLQLSYNSY...S..........R.S.TG.V...D....MN.RTH
PHA-L           .R....ND.GEPTLS.L...........W.N.TGAV.ASP.S...N.DVP
PHA-E           .R....N..GEPRLS.L...........W.N.TGAV...A.S...N.QVP

                101                  125                   150
                .                    .                     .
Arcelin-1       NIENSAYGLAFALVPVGSRPKLKGRYLGLFNTTNYDRDAHTVAVVFDT--
Arcelin-2       .N.............................A.................
Lectin-Like     RQA...V..D.V....--Q.ES..--------------D..T.E...--
PHA-L           .NSGP.D....V.L....Q..D..GL.....NYK..SN......E...LY
PHA-E           .NAGP.D...........Q..D..GF....--DGSNSNF.....E...LY

                151                  175                   200
                .                    .                     .
Arcelin-1       ------VSNRIEIDVNSIRPIATESCNFGHNNGEKAEVRITYDSPKNDLR
Arcelin-2       ------.....................................Y.......
Lectin-Like     ------FLS..S....-NND.KSVPWDVHDYD.QN.......N.STKVFS
PHA-L           NVHWDPKPRH.G......--KSIKTTTWDFVK..N...L.....STKL.V
PHA-E           NKDWDPTERH.G.......--SIKTTRWDFV...N...L.....STKL.V

                201                  225                   250
                .                    .                     .
Arcelin-1       VSLLYPSSEEKCHVSATVPLEKEVEDWVSVGFSATSGSKKETTETHNVLS
Arcelin-2       ..................................................
Lectin-Like     ...SNPSTGKSNN..T..E.....Y............AYQWSY...D...
PHA-L           A..V...LKTSFI..D..D.KSVLPE..I...T..T.IT.GNV..NDI..
PHA-E           A..V...QKTSFI..D..D.KSVLPE.........T.IT.GNV..ND...

                251                  275
                .                    .
Arcelin-1       WSF-SSNFINFKGKKSERSNILLNKIL
Arcelin-2       ...........L................
Lectin-Like     ...-......L.DQ.......V.....
PHA-L           ...A.KLSDGTTSEALNLA.FA..Q..
PHA-E           ...A.KLSDGTTSEGLNLA.LV..Q..
```

Figure 1. Amino acid sequence comparison of the arcelin and lectin genes. Dots indicate identical amino acids.

Table 1. Properties of phaseolin, lectin, and arcelin.

	Phaseolin	Lectin	Arcelin
M.W. (kD)			
Native	654 (tetramer of trimer)	146.1 (tetramer)	80.9 (dimer) 159.6 (tetramer)
Subunit	51, 48, 45.5		35 to 42
Deglycosylated	47, 43	29.9; 30.8	29.9
pI	4.4-5.6	5.2-5.4	6.7; 6.8
% of Seed Total Protein			
Cultivars(*)	60	6-12	0
Wild beans	14		10% seed wt
Accumulation			
Starting Time (DAF)	13	15	16
Cotyledon Length (mm)	>3		
Antigenically Related Protein	none	Arcelin	Lectin
Agglutinating Activity	none	yes	w/ pronase-treated blood cells

*: do not include those bred for arcelins.

some unique characteristics, such as having greater numbers of basic amino acid residues (more basic pI's), more cysteine residues, and a methionine residue (which PHA lacks). Arcelins only agglutinate pronase-treated erythrocytes and are present mostly as dimers instead of tetramers like native PHA (Table 1). Furthermore, despite the similarity, lectins do not show insecticidal activity against the most important bruchid pest.

Arcelin-1 is composed of both dimers and tetramers as determined by examining the native and denatured relative molecular wights (M_rs) of the protein (Hartweck et al, 1991). Arcelin-2 is composed of dimers only, whereas arcelin-3 and 4 exist as tetramers. These variants have immunological cross-reactivities and 77 to 100% identity of their first 37 N-terminal amino acids. Arcelin-1 and 2 genes also share one common DNA restriction fragment.

Arcelin genes from the wild beans were crossed into a domestic bean line, Sanilac. The analysis of F_2 and F_3 seeds from a single F_1 revealed that arcelin gene expression was inherited as a single Mendelian gene. Several backcrosses to the domestic parent, Sanilac, were employed to produce a near-isogeneic line (Romero Andreas et al, 1986).

The comparative value of the 4 arcelin variants were studied in a backcross breeding program (Cardona et al, 1990). The progenies containing the arcelin-1 and 2 genes

had the highest and intermediate arcelin contents and insect resistance, respectively. The arcelin-3 and 4 derived lines showed no resistance and low arcelin concentrations. In all the lines studied, a high arcelin/phaseolin ratio is required for insect resistance.

Common beans containing arcelin have a lower content of phaseolin (14%) than those lacking arcelin (60%) (Romero Andreas et al, 1986). When phaseolin was added to the arcelin containing seeds for insect feeding trials, the effect of arcelin was reduced (Cardona et al, 1983). This suggests that the resistant mechanism may also rely on a nutritional balance within beans as in cowpeas (Warris et al, 1983).

Certain bean cultivars containing complex lectins have less phaseolin (Osborn et al, 1985). The authors postulated that the genes controlling qualitative lectin variation may also regulate the quantitative variation of lectin and phaseolin. A similar mechanism(s) may be involved in the case of arcelin. Although the seeds have low phaseolin content, the phenotype, germination, plant growth, pollen fertility, and percentage of total seed storage protein are normal.

Insecticidal activity of arcelin. The insecticidal activity of arcelin-1 was examined in feeding experiments using seeds from bean lines with or without arcelin and using artificial seeds with different levels of purified

arcelin-1 protein (Osborn et al, 1988a). The number of days required for adult emergence and the percentage of adult emergence were taken as a measure for insecticidal activity. The level of resistance to insects increased with the introduction of arcelin-1 alleles in the cultivar 'Sanilac'. The beans from segregating lines (Arc/arc) were approximately 50% as resistant as those from homozygous lines (Arc/Arc). Similar insecticidal activity was observed for the artificial beans containing 10% (w/w) of purified arcelin-1, an amount equivalent to the arcelin level in the Arc/Arc beans.

The antimetabolic activity of arcelin was suggested by Dobie et al (1990) and Minney et al (1991) as the mechanism of insect toxicity. The antimetabolic effect may be due to arcelin's resistance to digestion by insect gut proteases. The availability of essential amino acids has also been suggested to be important (Gatehouse and Boulter, 1983). The total methionine content in the arcelin containing line is reduced, since arcelin contains less methionine than phaseolin (Osborn et a., 1988b and Bliss, 1990). Enzyme inhibition was ruled out because the antimetabolic effect was diminished when supplemented with phaseolin. As in the case of cowpea and soybean trypsin inhibitor (Warris et al, 1983, and Broadway and Duffey, 1988), protein quality significantly altered the effect of arcelin.

Arcelin genes. The arcelin genes are not linked to those controlling phaseolin but tightly linked to lectin genes (<0.30% recombination) (Osborn et al, 1986). It is likely that arcelin genes arose by duplication and diversification of existing lectin genes.

The cDNA's of arcelin-1 and 2 have been cloned and sequenced (Osborn et al, 1988a; John and Long, 1990). Each sequence codes for a polypeptide of 265 amino acids with a 21-amino-acid signal peptide. There are six single-nucleotide differences between the two clones. One of these is present in the 3' untranslated region. The other five substitutions are located in the coding region and four of these result in amino acid changes. The deduced amino acid sequences indicate that there are three possible glycosylation sites in arcelin-1 whereas arcelin-2 contains only two. The arcelin-2 transcripts, 950 nucleotides in length, began to accumulate 7 days after flowering, and continued to accumulate until 15 days after flowering (John and Long, 1990).

In spite of the 60% homology in the deduced amino acid sequences between the arcelins and lectins, several stretches of sequences in these 2 types of proteins show high levels of divergence. The longest stretch, between amino acids 172 to 181 or nucleotides 480 to 510, may serve as a probe to distinguish arcelins from lectins. These

sequences may also be important in conferring the insecticidal activity.

Phenomena of gene interaction. There are many reports indicating interactions between endogenouse and introduced genes. However, most of the studies concern the introduced genes code for the regulatory proteins of the endogenous genes. Arcelin is unlikely to be a regulatory protein of the phaseolin genes.

Some studies describe the suppression of endogenous wild-type genes by transformed homologous genes, such as the chalcone synthase gene in petunia (Napoli et al, 1990 and van der Krol, et al, 1990), the partial nopaline synthase gene in tobacco (Goring et al, 1990), and the truncated polygalacturonase gene in tomato (Smith et al, 1990). There is, however, no apparent homology between phaseolin and arcelin.

Gene interactions were observed in double transformed tobacco plants (Matzke and Matzke, 1990 and 1991). The second transformation of T-DNA (transfer DNA), containing hygromycin resistance and octopine synthase genes, inhibited the expression of the kanamycin resistance and nopaline synthase genes in the T-DNA from a prior transformation. The authors postulated that the degrees of methylation and the competition for common nuclear factors were the causes of suppression.

Another type of the interaction is that the introduced non-homologous gene alters the nutrient requirement for growth and development, resulting in suppression of regular gene expression (Lefebvre, 1990). Here, suppression is not likely due just to competition of nitrogen between arcelin and phaseolin, since arcelin and lectin are so similar and lectin does not have the same effect.

Regulatory elements of bean seed storage protein genes. Advances in molecular biology allow manipulation of genes *in vitro*. Modification of nucleotide sequences, deletions of promoter regions, fusion of regulatory sequences with other promoters, and fusion to reporter genes are common. The *in vivo* expression of these modified genes can be examined using transgenic plants. Methods such as foot printing and gel retardation assay have been used to identify regulatory elements of genes.

In eukaryotic genes, transcription regulatory sequence elements have been identified in the 5' upstream regions, as well as in the introns and the 3' downstream regions. For seed storage protein genes, such as the β-phaseolin gene, the effect of introns and intron splicing was shown not to be important (Chee et al, 1986). In the present study, both lectin and arcelin genes do not contain introns. Indeed, most studies on plant gene regulation have concentrated on the 5' region.

Table 2 lists some of the cloned storage protein genes in legumes and their regulatory elements. Among the various regulatory elements, the RY repeat may promote the formation of Z-DNA at the 5' end of the mRNA. The presence of Z-DNA at the transcription site may make the transcription initiation region more accessible to RNA polymerase (Voelker et al, 1987). A foldback at the 5' end of the mRNA is also possible and may be related to the stability of messages (Voelker et al, 1987). An alternative function is that RY repeats may confer tissue-specificity (Voelker et al, 1986). Voelker et al (1987) suggested that the RY repeats located in the upstream region of the lectin genes may serve as an enhancer. In transgenic tobacco, upstream RY repeats play a role in modulating the level of GUS expression by the soybean glycinin promoter (Lelievre et al, 1992). Gel shift experiments showed conflicting results in different plants. Ericson and co-workers (1991) found RY repeats involved in protein binding in napin genes (*Brassica napus*). However, no binding activity was detected in legA gene from pea (Meakin and Gatehouse, 1991).

The vicilin box is a conserved 42-bp sequence found in the upstream region (-120 to -130 relative to CAP site) of the vicilin gene family in pea (*Pisum sativum*) and the α- and β-phaseolin gene in common bean (Anthony et al, 1990 and Gatehouse et al, 1986). The 5' region (GCCACCTCAATTT)

Table 2. Legume seed storage protein genes and their regulatory elements.

speices	protein	regulatory element	function	ref
legumes	seed storage	RY repeat CATGCATG	enhancer	Dickinson et al, 1988
		vicilin box GCCACCTCA	enhancer	Gatehouse et al, 1986
bean Phaseolu vulgaris	phaseolin	AT rich	+	Bustos et al, 1989
		UAS1, UAS2	+	Bustos et al, 1991b
		NRS1, NRS2	-	
	lectin	AT rich (-500 to -345)	+	Riggs et al, 1989
		RY repeat	enhancer	
		AT rich (-1000 to -675)	-	
pea Pisum sativum	lugumin	leg box, 5'TCCATAGCCATGCA AGCTGCAGAATGTC-3'	tissue & developemental specific	Shirsat et al, 1989
soybean Glycine max	conglycinin	SEF 3 binding site, AACCCA	tissue & developemental specific	Allen et al, 1989
		SEF 4 binding site		
		GTGGATAG	enhancer	Chen et al, 1986
	glycinin	RY repeat	enhancer	Lelievre et al, 1992
	lectin	ATTAAAT or ATTTAAT	tissue specific	Jofuku et al, 1987
	Kunitz inhibitor	60 kd factor binding		

of these conserved sequences is identical to the adenovirus enhancer core element.

Legume storage protein genes are often flanked by A/T rich regions at both 5' and 3' ends (Gatehouse et al, 1986). The interaction of several A/T rich regions with specific nuclear factors has been demonstrated, for example, in the β-phaseolin promoter (Bustos et al, 1989), and the lectin genes in bean (Riggs et al, 1989) and soybean (Jofuku et al, 1987). These sequences may act as positive or negative regulatory elements. A/T rich regions have also been observed in the non-seed storage protein genes, including leghaemoglobin genes (Jensen et al, 1988) and several light-regulated genes (Datta and Cashmore, 1989).

Post-transcriptional regulation of seed storage protein gene expression. Current knowledge of the post-transcriptional control of plant genes is quite limited. In the studies comparing the expression of different seed storage protein genes in a single species, for example, common bean (Chappell and Chrispeels, 1986), soybean (Wallings et al, 1986, and Nielsen et al, 1989), and pea (Thompson et al, 1989), both transcriptional and post-transcriptional control were found to be important for seed storage protein gene regulation. These conclusions were drawn from correlations between the transcription rates

estimated by nuclear run-on assays and the levels of mRNAs quantitated by Northern or dot blot analyses.

The stability of eukaryotic mRNA is affected by its structure, such as the untranslated regions and the poly(A) tail (Jackson and Standart, 1990). In mammalian systems, the length of poly(A) tails decreases as the mRNA ages and some short-lived mRNAs have a shorter poly(A) tail. The sequence in the 3' untranslated region also affects the degradation of mRNA. RNA binding proteins, including poly(A) binding proteins (PABPs), protect mRNAs from RNases digestion. There is also some positive correlation between the degree of adenylation and translation. Many mRNAs, such as tubulin mRNA, that have been studied are stabilized by the presence of protein synthesis inhibitors (Atwater et al, 1990), suggesting that nucleases or other factors involved in mRNA degradation may be associated with ribosomes. Alternatively, the component of the degradation machinery may be labile, requiring continued protein synthesis. In contrast, other mRNAs, such as β-globin, are destabilized by mutations disrupting the translation process. In yeast, translation elongation is also a major process maintaining mRNA stability (Herrick et al, 1990). Stable and unstable mRNAs do not differ in their poly(A) metabolism and deadenylation does not destabilize the stable mRNAs. Also, the size of mRNA does not correlate with decay rate.

In plants, both mRNA structure and translatability influence stability of the mRNA. As mentioned earlier, the fold-back structure caused by RY repeats may related to mRNA stability (Voelker et la., 1987). When the 3' end regions from several genes were fused with the reporter gene NPTII, the resulting steady-state NPTII mRNA levels varied considerably (Ingelbrecht et al, 1989). A tRNA-like structure at the 3' end of tobacco mosaic virus increases both mRNA stability and translational efficiency (Gallia et al, 1991). The introduction of a stop codon in the patatin gene led to the premature termination of translation and a dramatic decrease in the mRNA level (Vancacanneyt et al, 1990). Similarly, the enlargement of the 3' untranslated region by several hundred bp resulted in a slight reduction in the level of mRNA. Frame-shift mutations resulting in premature appearance of stop codon have also been shown to prevent mRNA accumulation (Jofuku et al, 1989 and Voelker et al, 1990). The instability of the mRNA encoding the insecticidal protein from *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts was attributed to the difference in the codon usage between bacterial and plant cells (Murray et al, 1991). Different codon usages among the seed storage protein genes in the same plant have been reported (Shotwell et al, 1990), which might lead to the differenc in the levels of expression.

Hypothesis

The temporal and spatial regulation of seed storage protein genes have been intensively investigated at the transcriptional level. However, the regulation of the quantitative expression among different seed storage protein genes in a single plant requires more study. In this regard, the finding that the presence of the arcelin gene coincides with the reduction of phaseolin in common beans is of particular interest and the bean line may provide a unique system to study this quantitative variation of gene expression. Comparison of the sequences and the expression of these genes in bean and transgenic plants may benefit our understanding of this regulation and the evolutionary significance of arcelin.

The hypothesis to be tested in this research is: The quantitative variation of the phaseolin and arcelin gene expression is conferred by their upstream promoter regions and/or the stability of their mRNAs.

This hypothesis will be tested through the following objectives:

1. To isolate and sequence a gene(s) encoding the arcelin-1 from common beans.
2. To measure the expression of phaseolin and arcelin genes at the mRNA and transcriptional levels in common beans.
3. to transfer the wild-type and chimeric phaseolin and arcelin gene(s) into tobacco and measure the expression of

these genes at the mRNA level in the transgenic tobacco seeds.

CHAPTER II. CLONING AND EXPRESSION OF A BEAN ARCELIN-1 GENE IN COMMON BEAN PLANTS

Introduction

The major seed storage protein in common beans is phaseolin. The gene encoding β-phaseolin was the first plant gene isolated (Sun et al, 1981). In the arcelin containing bean line, Romero Andreas et al (1986) observed that the amount of phaseolin is greatly reduced and arcelin becomes the most abundant seed protein. Thus, the introduction of arcelin genes from wild accessions into domestic bean lines by breeding has a negative effect on the expression of the phaseolin genes. The goal of this study is to pinpoint at which level this interaction occurs.

The developmental expression of phaseolin and arcelin has been investigated at the protein level by Osborn et al (1988b). cDNAs of arcelin-1 and 2 have been isolated (Osborn et al, 1988a and John and Long, 1990). In order to compare the structure and regulation of phaseolin and arcelin genes, the genomic sequences of arcelin genes have to be isolated and analyzed. In this chapter, I report the isolation of an arcelin-1 gene and the comparative studies on the expression of phaseolin and arcelin genes at the RNA and transcriptional levels in the developing seeds from 2 different bean lines.

Materials and Methods

Chemicals. All chemicals (Fisher Scientific or otherwise noted) were reagent grade or better.

Plant materials. Four lines of *Phaseolus vulgaris* L. cv. were used in this study: seeds of bean line SARC1-7 (containing arcelin-1), the parent line Sanilac, and the PHA-less line were kindly provided by Dr. T. Osborn (University of Wisconsin, Madison); Tendergreen was from Olds Seeds (Madison, Wisconsin). Seeds were germinated and grown in the greenhouse. Seven stages of developing seeds, based on the seed length, were collected and stored at -80°C. Mature seeds were dried and stored at 4°C.

Protein extraction and SDS-PAGE. After seedcoat removal, mature seeds were ground in 0.2 M NaCl and 0.035 M Na-phosphate, pH 7.5, at a ratio of 20 ml buffer/g seeds. Cell debris were removed by microcentrifugation at 14,000 rpm for 15 min twice. The protein content in the supernatant was determined by the method of Lowry (1951). Protein samples (20 μg/lane) were separated by 12.5 % SDS-PAGE.

Plasmids and Synthetic Oligonucleotide. An arcelin-1 cDNA clone, pAR1-11 (Osborn et al, 1988a), containing a 845-bp XbaI insert was labeled with [α-^{32}P]dCTP by nick translation (BRL) or by random priming (Feinberg and Vogelstein, 1983) and used for library screening and Southern analysis, respectively. A 19-base

oligonucleotide, 5'-CCGAAATTGCAAGACTCCG-3', which hybridizes to pARC1-11 from nucleotide 485 to 503 (Osborn et al, 1988a) was synthesized (Biotechnology Instrumemtation and Training Facility, University of Hawaii) and labeled with [γ-^{32}P]ATP using polynucleotide kinase (Promega) for hybridization. This sequence has a maximal homology of 53% to the same regions in lectin cDNAs. Plasmid SB-9, a 4-kb SalI/BamHI phaseolin fragment in pTZ19U (Bio-Rad, Mead et al, 1986) was used as the phaseolin probe. pDLEC1 is a lectin genomic clone provided by Dr. T. Osborn.

Southern analysis. Bean genomic DNA (8 μg, provided by Dr. Osborn, University of Wisconsin) was digested with EcoRI or HindIII (Promega), and separated on a 0.7% agarose gel. Transfer and hybridization were performed by following Schleicher & Schuell NYTRAN nylon membrane protocols.

Construction and screening of the genomic library. DNA from line SARC1-7 was partially digested with MboI (1 unit/μg DNA, 37°C, 2.5 to 22.5 minutes) and the resulting fragments ranging from 9 to 23 kb were purified by NaCl gradient centrifugation (5-20%, Beckman SW28 rotor, 16K rpm, 21°C, 22hr). Vector λEMBL4 (Promega) was double digested with SalI and BamHI, dephosphorylated by calf intestinal phosphatase (CIP, Promega, 1 unit/10 μg DNA in a total volume of 50 μl, 37°C, 30 min) and the left and right

arms were isolated by NaCl gradient centrifugation. The inserts and arms were ligated (ligase from BRL), packaged (packaging extracts from Stratagene), propagated in E. coli strain K802 and amplified in strain MB406.

Fifteen potential arcelin-1 clones were isolated from 4.8 x 10^5 recombinant phages, using the 845-bp *Xba*I fragment of pAR1-11 as a probe (hybridization conditions: 6X SSC, 0.02 % Ficoll, 0.02 % polyvinylpyrrolidone-360, 0.02% BSA, 0.5 % SDS, 30 μg/ml poly(rA) (Sigma), 10 μg/ml sheared and denature salmon sperm DNA, 68°C, overnight). Phage DNA was purified from plate lysates by DEAE-cellulose column chromatography (Helms et al, 1987). The synthetic 19mer differentially hybridized to 7 out of the 15 potential clones.

Subcloning and sequencing of positive clones. A 2.3-kb EcoRI fragment from clone λ222 was subcloned into plasmid pVZ-1 (derivatives of pBluescript, Stratagene, with modified polylinker, Steve Henikoff, personal communication) and named pVZ-1.35. Two sets of deletion sub clones from the T3 and T7 primer ends, were constructed using the Erase-a-Base™ system (Promega). For deletions from the T3 primer end, the DNA was first double digested with *Pst*I and *Bam*HI, whereas *Not*I and *Bst*XI were used for the deletions from the T7 Primer end. Dideoxy sequencing was conducted by following the Sequenase Version 2.0 protocol (United States Biochemical). A 750-bp *Xba*I

fragment which overlaps the 3' end of the EcoRI fragment was also subcloned into PVZ-1 and sequenced.

RNA extraction. A small-scale RNA extraction procedure (Sunitha Midha, Enichem, personal communication) was used to isolate total RNA from developing bean seeds. Approximately 100 to 200 mg of seeds were briefly ground in liquid N_2 with a mortar and pestle and transferred into a Kontes microfuge tube kept in liquid N_2. The sample was further ground into a fine powder with a Kontes pestle, which was pre-chilled in liquid N_2 and driven by a 1/5 horsepower drill (Sears). Two-hundred-and-fifty μl extraction buffer (0.1 M LiCl, 0.1 M TrisHCl, pH 8, 0.1 M EDTA, and 1% SDS, prewarmed to 80°C) and 250 μl phenol were added to the microfuge tube. The mixture was vortexed for 30 seconds. And 250 μl of chloroform:iso-amyl alcohol (24:1) was added, vortexed for 30 seconds, and microcentrifuged for 5 min at room temperature. The aqueous phase (250 to 300 μl) was mixed with one volume of 4 M LiCl and kept at 4°C overnight. The samples were microfuged at 14,000 rpm at 4°C for 30 min. The pellet was washed with 100% ethanol, dried, and resuspended in 250 μl DEPC-treated water. The RNA was ethanol precipitated and resuspended in DEPC-treated water.

Primer extension - Determination of transcription start site. The procedure of Heinrich Albert (personal communication) was followed. The 19-base oligonucleotide

(see Plasmids and synthetic oligonucleotide) was end-labelled with ^{32}P to high specific activity with T4 polynucleotide kinase (Promega) following the manufacturer's instructions. Eighteen ng (2 pmol) of the labelled 19mer was mixed with 45 μg of total RNA from developing bean seeds (SARC1-7, 11-13 mm stage) in a total volume of 25 μl. The mixture was heated at 90°C for 5 min, mixed with 2.2 μl of 1.2 M NaCl, and heated for another 20 min. The primer was annealed to the RNA by slow cooling in a turned-off heat block overnight. The following were added to the annealed mixture to extend the DNA by reverse transcription: 10 μl 5X salts (BRL), 0.5 μl RNAsin (40 unit/μl, Promega), 5 μl dNTP's (10 mM each), 5 μl 0.1 M DTT, and 2.5 μl M-MLV reverse transcriptase (200 unit/μl, BRL). After 1 hr of incubation at 37°C, another 1 μl of M-MLV reverse transcriptase was added and the reaction was incubated for an additional hour. The reaction was stopped by heating at 65°C for 10 min and ethanol precipitated. One half of the product was run on a 5% denatured sequencing gel ("Long Ranger", HydroLink) against sequencing ladders using the same primer. No visible difference from the ^{35}S-labelled sequencing ladder was observed when using ^{32}P-end-labelled primer for sequencing.

Northern analysis. Total RNA (3 μg) from developing bean seeds of various stages was separated on a 1.5% agarose/formaldehyde gel. Transfer, hybridization, and

probe removal were performed following the Schleicher & Schuell NYTRAN nylon membrane protocols. The blot was first probed with the 19-mer for arcelin-1, stripped, and then probed with the 4 kb phaseolin probe (pSB9, see Plasmids and synthetic oligonucleotide).

In vitro **translation.** A wheat germ system similar to that described by Davis and Kaesberg (1973) and used by Sun et al (1975) was used. Each translation reaction contained 3 μg of total RNA from developing Sanilac or SARC1-7 seeds (5-7, 7-9, 9-11, and 11-13 mm). "No RNA" was used as control. Seven and one half μl of the reaction, representing translation product from equal amount of RNA, was analyzed by 20% SDS/PAGE and autoradiography.

Nuclei isolation. Nuclei from freshly harvested developing seeds (same stages as in "*In vitro* translation") were isolated using the method described by Luthe and Quatrano (1980), with a few modifications. The pHs of all buffers were changed to 8.5 according to Walling et al (1986). Three to 4 g of tissue were ground in liquid nitrogen with a mortar and pestle and 2 ml/g tissue of a modified Honda buffer [0.44 M sucrose, 2.5% Ficoll (m.w. 400,000), 5% Dextran 40, 25 mM Tris-HCl, pH 8.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 2 mM spermine and 0.5% Triton X-100] was added. The mixture was allowed to thaw on ice and ground further. Another 4 volumes of buffer was added and the homogenate was sequentially filtered through

4 layers of cheesecloth, 1 layer of Miracloth and 2 nylon meshes (83 and 73 μm meshes were used instead of 80 and 61 um as specified by Luthe and Quatrano, 1980). The filtrate was centrifuged at 5,850 g (Sorvall SS34 rotor, 7,000 rpm) and the pellet was gently resuspended in 1 ml of nuclei resuspension buffer (NRB, 50 mM Tris-HCl, pH 7.8, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol and 20% glycerol). This crude nuclear extract was further fractionated on a Percoll gradient. The gradients contained 4 ml each of 40, 60, and 80% Percoll, containing 0.44 M sucrose, 25 mM Tris-HCl, pH8.5, and 10 mM $MgCl_2$, and a 2 M sucrose cushion, containing 25 mM Tris-HCl, pH 8.5 and 10 mM $MgCl_2$. A Beckman SW28.1 swinging bucket rotor was used for the centrifugation. The centrifugation was performed at 5,300 rpm (4,080Xg). The fraction at the bottom of 80% Percoll, right above the 2 M sucrose layer was collected, washed, resuspended in NRB (120 to 160 μl) and stored at -80°C.

RNA synthesis in isolated nuclei. A modification of the procedures of Chapell and Chrispeels (1986) was used. RNA synthesis was performed in a 20 to 80 μl reaction volume, containing 10 to 40 μl of nuclear extract, 100 mM $(NH_4)_2SO_4$, 4 mM $MgCl_2$, 0.5 mM each GTP, CTP, and ATP (Promega, in 10 mM Tris-HCl, pH 8), and 20 to 60 μCi of α^{32}P-UTP (3000 Ci/mmol). The volumes of reagents specified below were for a 20-μl reaction. For higher volume reactions, they were adjusted proportionally. After

incubation at 30°C for 30 min, yeast tRNA (10 mg/ml) and 0.5 μl of RQ DNase (Promega) were added and the reaction mix was incubated at 30°C for 5 min. Two and one half μl of 10X stop buffer (5 mM EDTA and 10% SDS) and 2.5 μl of proteinase K (1 μg/μl, pretreated at 37°C for 30 min) were added and the mix was incubated for another 15 min at 30°C. The reaction mix was then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), and twice with chloroform:isoamyl alcohol (24:1). The unincorporated nucleotides were removed by a Sephadex G-50 spin column. The first 2 fractions from the column were pooled and ethanol precipitated. The RNA pellets were washed with 100% ethanol, dried, and resuspend in hybridization buffer (see below). The amount of synthesized RNA was quantitated by scintillation counting after TCA (Trichloroacetic acid) precipitation. Duplicates of a 10-μl resuspension were added to 1 ml of cold 5% TCA, respectively, and incubated on ice for at least 15 min. The mixture was then spotted on a GF-C filter under vacuum. The filter was washed 5 times with 5 ml of cold 5% TCA each time and once with 1 ml cold 100% ethanol, dried and counted in a scintillation counter (Cerenkov).

Detection of arcelin-1 and phaseolin transcripts. Two pTZ18U based arcelin-1 and phaseolin clones, pAase and pPase, as well as pTZ18U were spotted on nylon membranes (Hybond N +, Amersham) using the dot blot procedure

provided by the manufacturer. The membranes were cut into 0.7 X 3.5 cm^2 strips with each strip containing one dot each of the above 3 plasmids. One μg of pAase (containing 200 ng of the arcelin coding region), 1.25 μg of pPase (containing 345 ng of the phaseolin coding region), or 0.53 μg of pTZ18U was used. These represent equal moles of each plasmid. A strip with 5-fold DNA on the membrane was used to show that the DNA was not limiting. The hybridization was performed in 5X SSPE, 5X Denhardt's solution, 0.5% SDS, 20 μg/ml denatured, sheared salmon sperm DNA (Amersham), 50% formamide and labelled RNA from developing seeds, at 42°C for 72 hrs. The strips were then washed as suggested by Amersham at high stringency (0.1X SSPE, 65°C), and further washed at room temperature twice with 2X SSPE, once with 2X SSPE containing RNase A (Promega), 1 μg/ml, and twice with 2X SSPE containing 0.5% SDS, 10 min for each wash. The hybridization signals were then visualized by autoradiography and quantitated using a densitometer (Quick Scan Auto Scanner, Helena Laboratories).

Results and Discussion

Protein profiles of different bean lines. The major seed protein in Sanilac and Tendergreen, is phaseolin (Figure 2), representing 40 to 60% of the total seed protein (Osborn, 1988). In Tendergreen, phaseolin consists of 3 subunits, α, β, and γ (γa and γb), with relative

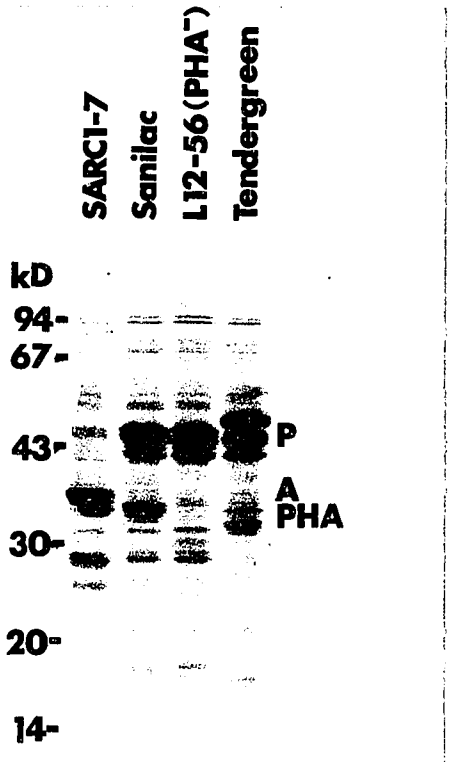


Figure 2. SDS-PAGE of total seed storage proteins from common bean lines. Phaseolin (P), lectin (PHA), and arcelin (A) are identified.

molecular weights (M_r) of 51, 48, and 45.5 kD. The phaseolin α-subunit in Sanilac has a lower M_r, as shown in Figure 2. Lectins comprise a group of polypeptides with M_r of 34 to 36 kD and are not present in the PHA deficient line (PHA⁻). In seeds of SARC1-7, the level of phaseolin is greatly reduced. In agreement with the observation by Romero Andreas et al (1986), arcelin-1 (about 37 kD) is the most predominent seed protein.

Genomic Southern analysis and gene copy number estimation. Seed storage proteins are usually encoded by

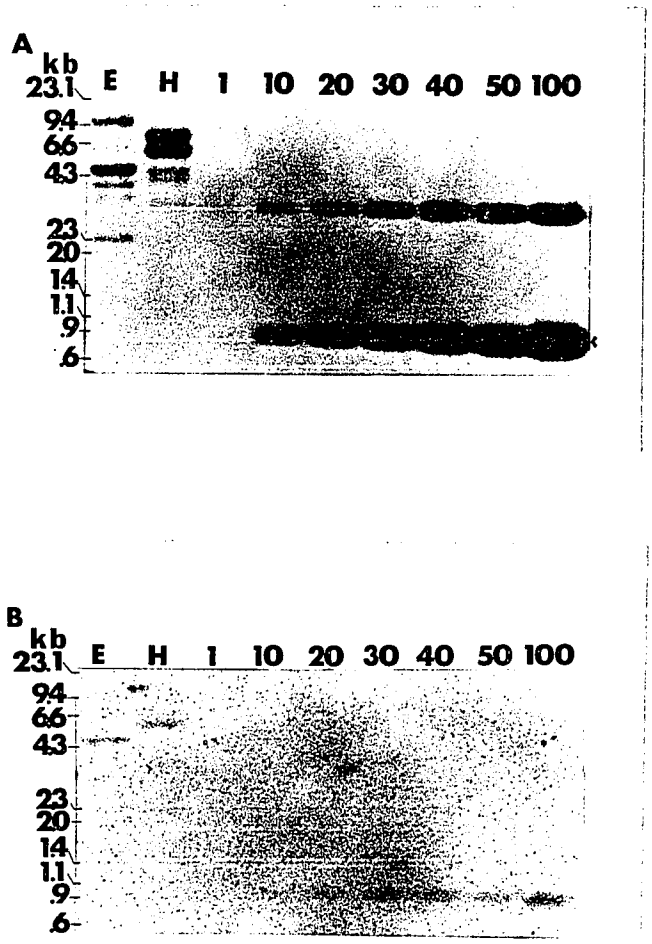


Figure 3. Southern blot analysis of bean (SARC1-7) genomic DNA. A, probed with pARC1-11, and B, with 19mer. E & H, bean genomic DNA digested with EcoRI & HindIII; 1 to 100, XbaI-digested pARC1-11, equivalent to 1 to 100 copies of arcelin cDNA.

multigene families. Southern blot analysis was used to examine the complexity of the arcelin gene family (Figure 3). The cDNA of arcelin-1 (pARC1-11) hybridized to a number of fragments in the genomic DNA. These include the 23.1, 10, 6, 4.6, 4.2, 3.8, 3.5, and 2.3 kb EcoRI fragments and the 7.5, 5.7, 4.5, 4.2, and 3.4 kb HindIII fragments. Compariosn of the intensity of these bands with that of the XbaI inserts of pARC1-11 (1 to 100 copies, inserts indicated by arrowhead "<" in Figure 3a) suggests that there are approximately 25 copies of the genomic sequences. This figure presumably includes arcelin and lectin genes, as the arcelin-1 gene shares a high sequence homology with the lectin genes (Osborn et al, 1988a). The 4.5 kb EcoRI and the 5.7 kb HindIII fragments are the most abundant species. The blot was stripped and probed with the synthetic oligonucleotide specific for the arcelin genes. Only one band in each digest hybridized (EcoRI, 4.5 kb; and HindIII, 5.7 kb, Figure 3B) and the copy number was estimated to be 10 to 20. However, this genome reconstruction experiment did not detect any species with less than 10 copies. Lectin genes are arranged as tandem repeats in the genome of beans (Hoffman and Donaldson, 1985). Since arcelin genes are likely to be diverged from lectin genes, it is possible that arcelin genes are organized in the same way.

Construction and screening of genomic library. The characteristics of the genomic library constructed for SARC1-7 bean line are summarized in Table 3. The genome size of *P. vulagaris* is 1.9 X 10^9 bp or 1.8 pg per haploid genome. If the average insert size is 20 kb, the number of clones (4.3 X 10^5) required to have a 99% chance to isolate a single copy gene can be calculated using the following equation:

$$N = \frac{\ln (1 - P)}{\ln (1 - f)}$$

N: number of clones required to have the probability P to isolate a single copy gene.

Table 3. Genomic library of common bean, *Phaseolus vulgaris*, line SARC1-7.

Insert size: 9-23 kb
Vector: Bacteriophage λEMBL4
Host strain: K802
Library size: 1.2 X 10^6 pfu
No. of clones required to isolate a single copy gene at 99% probability: 4.4 X 10^5 pfu
No. of clones screened: 4.8 X 10^5 pfu

P: possibility. When P = 99%, N = one genome equivalent.

$$f: \frac{\text{insert size}}{\text{genome size}}$$

Fifteen potential clones were identified from 4.8 X 10^5 recombinant phage using the full-length cDNA sequence as a probe. Figure 4 shows the Southern blot of these 15 clones digested with EcoRI and probed with the arcelin-specific 19-mer. Lanes A and B were the positive and lanes C and D were the negative controls to assure that the synthetic

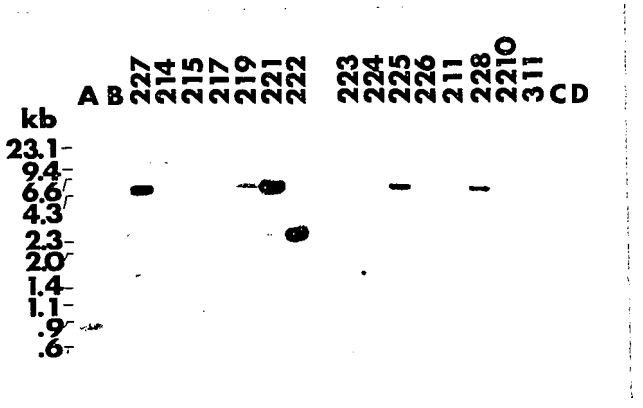


Figure 4. Southern blot analysis of potential arcelin-1 clones. Lane A, pARC1-11 digested with XbaI; B, uncut pARC1-11; 227 to 311, λPv clones digested with EcoRI; C, pDLEC1 digested with PstI; and D, uncut pDLEC1. The blot was probed with the 19mer specific to arcelin.

oligonucleotide was able to differentiate between arcelin and lectin sequences. Indeed, this probe only hybridized to pAR1-11 (lanes A and B) but not to pDLEC1 (lanes C and D). Seven out of the 15 clones contain fragments that hybridized to the arcelin specific probe. The size of the hybridized fragments in five of the clones, λPv227, λPv219, λPv221, λPv224, and λPv228, corresponds to the major band (4.5 kb) shown in the genomic Southern blot (Figure 3B). Two other clones, λPv222 and λPv2210, containing fragments of 2.3 and 6 kb, respectively, may represent low-copy classes of the arcelin genes in the genome. The 2.3-kb fragment from λPv222 and the 4-kb EcoRI/HindIII fragments from λPv219 and λPv228 were subcloned and the location and orientation of the coding region were identified by restriction mapping and hybridization (data not shown). They are similar at the 3' end of the genes and the latter 2 have longer 5' ends. The 2.3-kb subclone (pVZ-1.35) contains about 1.1 kb of upstream sequence and was selected for sequencing.

Sequence of arcelin-1 gene. A total of 2692 bp were determined from the 2.3 kb subclon (pVZ-1.35) and an overlapping 0.7 kb subclone (pVZ-1.X700). This arcelin-1 sequence, PHVARC1-1, contains 1167 bp in the 5' upstream region, 798 bp in the coding region, and 727 bp in the 3' downstream region (Figure 5). As in lectin genes (Hoffman and Donaldson, 1985), PHVARC1-1 lacks intron. Anthony et

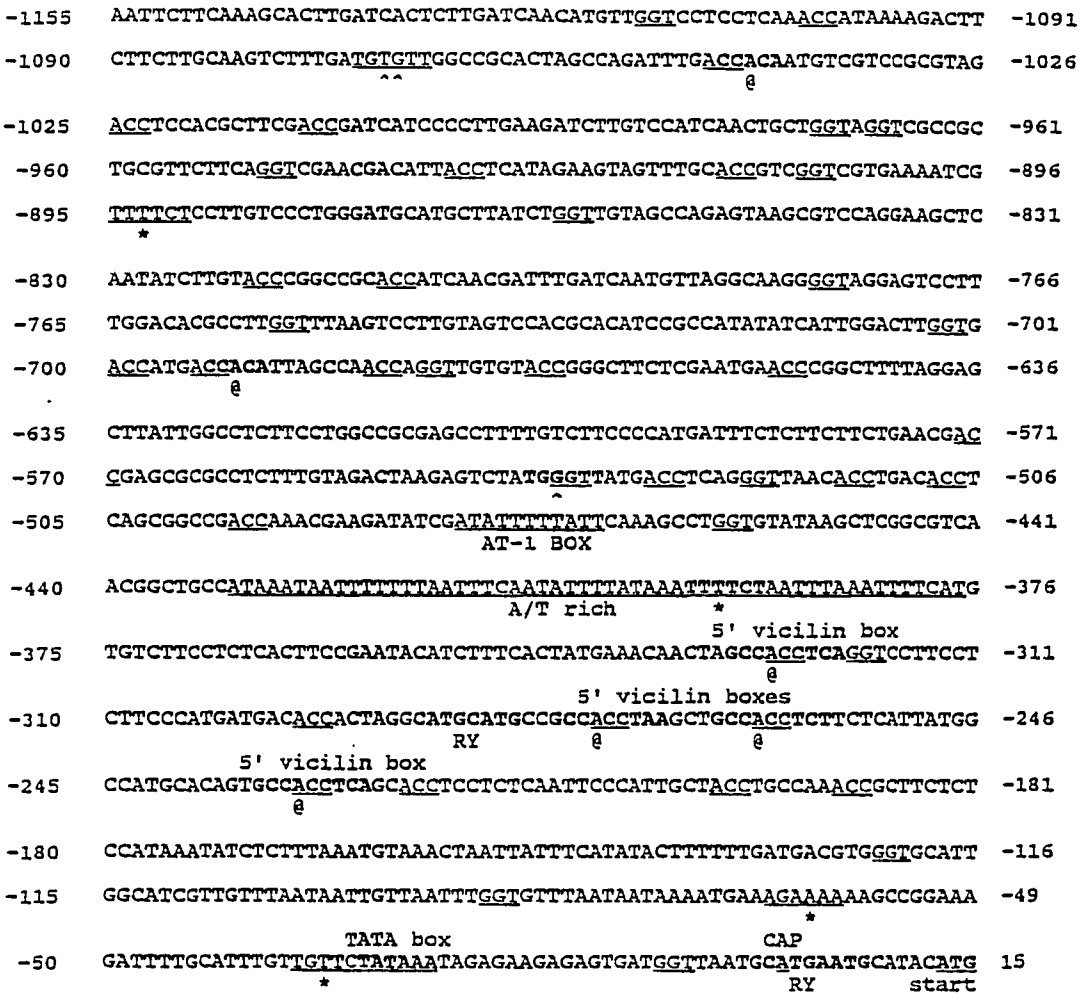


Figure 5. Complete nucleotide sequence of PHVARC1-1. Conserved sequence motifs, putative TATA box, CAP site, translation start and stop codons, poly-A signals are in bold or underlined. The single nucleotide (A) different from the published cDNA sequence (C, Osborn et al., 1988a), located right before the poly(A) signals is underlined. The 5 nucleotides different from the arcelin-2 gene (John and Long, 1990) result in 4 amino acid changes (in "()"). Trinucleotide ACC (or GGT) was noted by Bustos et al. (1991). The references for other motifs are as follow: ^^, Goldberg, 1986; @ or vicilin box, Gatehouse et al., 1986; *, Bustos et al., 1991; ^, Allen et al., 1989.

```
 16  GCTTCCTCCAACTTACTCACCCTAGCCCTCTTCCTTGTGCTTCTCACCCACGCAAACTCAAGCAA  80

 81  CGACGCCTCCTTCAACGTCGAGACGTTCAACAAAACCAACCTCATCCTCCAAGGCGATGCCACCG  145

146  TCTCATCCGAAGGCCACTTACTACTAACCAATGTTAAAGGCAACGAAGAGGACTCTATGGGCCGC  210

211  GCCTTCTACTCCGCCCCCATCCAAATCAATGACAGAACCATCGACAACCTCGCCAGCTTCTCCAC  275

                                    Ile
276  CAACTTCACATTCCGTATCAACGCTAAGAACATTGAAAATTCCGCCTATGGCCTTGCCTTTGCTC  340
                                      A (in Arc-2, Asn)
                                                                   Thr
341  TCGTCCCCGTCGGCTCTCGGCCCAAACTTAAAGGCCGTTATCTAGGTCTTTTCAACACAACCAAC  405
                                                                   G(Ala)
      C
406  TATGACCGCGACGCCCATACTGTGGCTGTGGTGTTCGACACCGTCAGCAACCGTATTGAAATCGA  470

471  CGTGAACTCCATCCGGCCTATCGCAACGGAGTCTTGCAATTTCGGCCACAACAACGGAGAAAAGG  535
                                 oligonucleotide (19mer)
                          Asp
436  CCGAGGTTCGGATCACCTATGACTCCCCCAAGAACGACTTGAGGGTTTCTCTGCTTTACCCTTCT  600
                          T (Tyr)

601  TCGGAAGAAAAGTGCCACGTCTCTGCCACAGTGCCGCTGGAGAAAGAAGTTGAGGACTGGGTGAG  665

666  CGTTGGGTTCTCTGCCACCTCAGGGTCGAAAAAAGAGACCACTGAAACGCACAACGTCCTCTCTT  730

                                   Lys
731  GGTCTTTTTCTTCCAACTTCATCAATTTTAAGGGCAAAAAATCTGAACGTTCCAACATCCTCCTC  795
                                   G (Glu)

796  AACAAGATCCTCTAGACTCCCAAAGCCAGCTTCACTGTGACAGTAAAACCTTCCTTATACGCTAA  860
               stop

861  TAATGTTCATCTGTCACACAAACTACAATAAATAAAAATGGGAGCAATAAATAAAATGGGAGCTCA  925
                                   poly(A) signals

926  TATATTTACACAATTTACACTGCCTATTATTCACCATGCCAATTATTACTGCATAATTTCAAAAT  990

991  TGTCATTTTTTAAAAGTTTATAATAATTAAGAAATATTACTATAAGTTAAAGTATAACATAGAAA  1055

1056 AAAAAACATTAAATCTTAAGAAATATTACTATAATTTACCCTTTTTTATCTGAAGAGTCTATAAT  1120

1121 TGAGAGATTGACACAAAATATTTATACCAGCTCCCTCTTTACCAAGAGCTACATTCAGTCTTCGA  1185

1186 ACGCACTAAGAATTCATTAACTAATCAACCTTGATTACTAACAATTCACCATGCCATTTATTACT  1250

1251 CGATAATCCTTTTATCTCAAAGAACAAGAAAAACTTTAATTCCTCTTACCTTATTCTTCTTTCAA  1315

1316 GTCTTGTAATCAAAATCTCAAAAAATATTCAAATCATTATTTATTTCAACTTTGTGATTTTTTAA  1380

1381 ATTAACTTTTTATTTATTGTTCTTGAACGAAGTTGGGGGCCTAGAACTATTAAGATCACTCTTCCT  1445

1446 TCGTCGGTCGGTCGAATTGTACTTCAGCTTCCCTCCTGCAAGCTCTCTCCCCCTCCTTACTCGTC  1510

1511 TCACCTCTTGGTCTTCTCGTAGTCTAG                                          1537
```

Figure 5. (continued) Complete nucleotide sequence of PHVARC1-1.

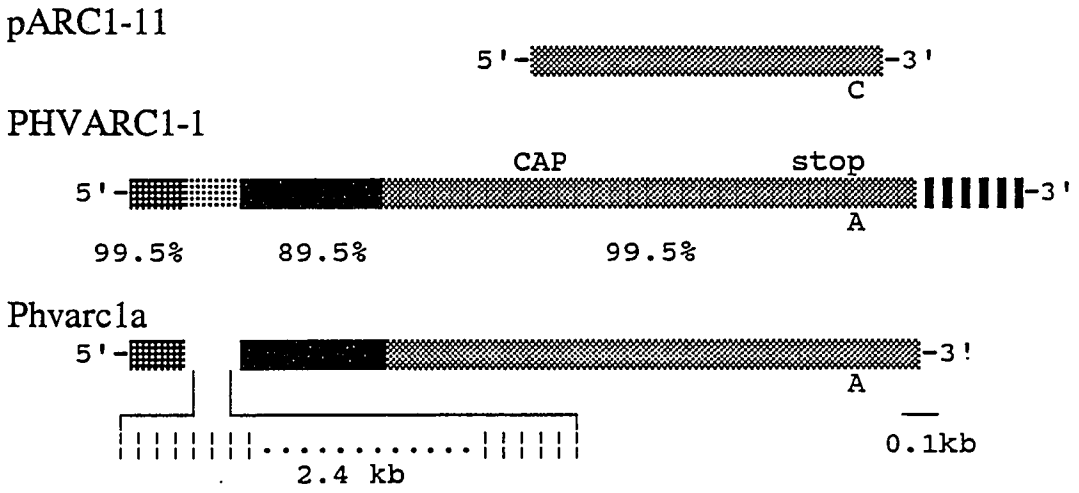


Figure 6. Schematic comparison of three arcelin-1 gene family members.

al (1991) isolated and sequenced another member (4.5 kb EcoRI class, Phvarc1a) of the arcelin gene family. Sequence comparison of these two arcelin genes reveals that there is an identity of 99.5% from -1140 to -994, 89.5% from -867 to -411, and 99.5% from -410 to 1207 (Figure 6). Phvarc1a has an extra 2.4-kb sequence between the -993 and -868 positions of PHVARC1-1.

Coding and 3' regions. With the exception of one nucleotide (+885) in the 3' untranslated region (Figure 5), the coding sequences of PHVARC1-1 and Phvarc1a are identical to the arcelin-1 cDNA sequence reported by Osborn et al (1988a). This suggests that these 3 gene are from

different members of the gene family. The open reading frame of PHVARC1-1 codes for a polypeptide of 265 amino acids. The first 21 amino acids are hydrophobic and probably function as a signal peptide. Multiple overlapping poly(A) signals appear about 80 bp downstream from the translation stop codon. In comparison to the arcelin-2 cDNA sequence (John and Long, 1990), there are 5 single nucleotide changes that result in 4 amino acid substitutions (Figure 5).

Transcription start site and TATA box. In the 5' region, a potential TATA box (TATAAA) appears 33 to 39 bp upstream from the translation start codon. The sequence around this region is almost identical to those of the lectin genes (Hoffman and Donaldson, 1985). In the case of lectin genes, the 5' untranslated regions are about 11-14 bp long. A single primer extension product of the arcelin-1 gene co-migrated with either a T (-12 from translation start) or a G (-13) residue in the sequencing ladder (Figure 7, indicated by "<"). This region (TGCATGA) has a very good match (5 out of 7 bases) with the consensus sequence for a transcription start site for higher plants (Joshi, 1987, Figure 8). The A residue (underlined in Figure 8) is the nearly invariant transcription start site. This would give rise to a 12-bp 5' untranslated region for this arcelin gene and a putative TATA box at about 23 bp upstream from the transcription start site. This is within

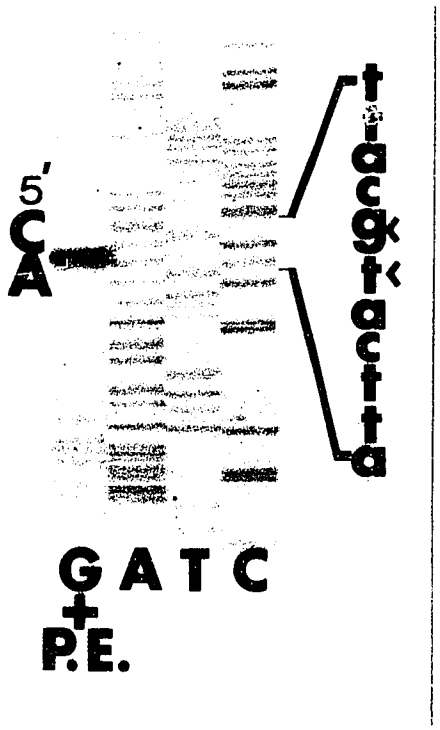


Figure 7. Determination of transcription start (CAP) site. Primer extension (P.E.) product was mixed with the sequencing reaction G. Arrows (<) indicate the potential CAP site.

```
PHVARC1-1          .G...G.
plant consensus    YTCATCA
("." = identical nucleotide)
```

Figure 8. PHVARC1-1 transcription start site vs. plant concensus.

region for plant gene TATA boxes (Joshi, 1987). Other than a TCAAT sequence at -210, no CCAAT box was found.

5' promoter region. The first 867 bp of the promoter regions of Phvarc1a and PHVARC1-1 are very similar (Figure 6). Several consensus sequences of cis-acting elements that may be involved in plant gene transcription can be found in the 5' region of PHVARC1-1. Similar to a variety of legume seed storage protein genes, including lectin genes (Dickinson et al, 1988), RY repeats (CATG(C/A)ATG) are present around the transcription start site and in the upstream region. Sequence elements similar to the 5' end of vicilin boxes (GCCACCT(C/A)(A/T), Gatehouse et al, 1986) are located around -140, -160, -170, and -230. Several seed storage protein genes contain A/T-rich activating elements (Bustos et al, 1989, Jofuku et al, 1987, and Riggs et al, 1989). In PHVARC1-1, there is a 55 bp A/T-rich region present at -384 to -439, but it is difficult to define sequence homology between this region and the A/T-rich region in the phaseolin gene. An AT-1 box consensus sequence (A(T/A)ATTTTTATT, Datta and Cashmore, 1989) is present at -470. The AT-1 box serves as an enhancer in the pea rbcS-3.6 gene (Timko et al, 1985), but as a silencer in the tobacco cab-E gene (Herrera-Estrella et al, 1984). The binding activity of the AT-1 box is regulated by phosphorylation of the protein factor AT-1 (Datta and Cashmore, 1989). Some other conserved motifs in the 5'

flanking sequences of seed storage protein genes, including the ACC trinucleotide, CCACA, AACACA, ACAAAA, and AACCCA (Figure 5) are also present in the phaseolin promoter.

Expression of arcelin-1 and phaseolin gene during seed development.

Accumulation of mRNA. In developing Sanilac seeds, phaseolin mRNA was detected when seeds reached 5 to 7 mm long (about 10 to 12 days after flowering or DAF, Figure 9A, left panel). The phaseolin mRNA increased untill the seeds reached 9 to 11 mm in length. The timing of the mRNA appearance was slightly ahead of that of phaseolin protein (Osborn et al, 1988b). A similar temporal sequence of phaseolin mRNA expression was observed during SARC1-7 seed development (Figure 9A, right panel). However, the phaseolin mRNA level decreased greatly after the seeds reached 7 to 9 mm and remained at this low level throughout the later stages of development. This reduction coincides with the onset of arcelin mRNA accumulation (Figure 9B, right panel). As expected, no arcelin mRNA was detected in developing Sanilac seeds (Figure 9B, left panel).

The phaseolin and arcelin mRNAs from developing seeds were translatable in the wheat germ *in vitro* translation system (Figure 10). In the Sanilac seeds, 2 polypeptides with molecular weights close to that of the phaseolin (about 40 kD) increase during seed development. However, in the SARC1-7 seeds, similar to the results of Northern

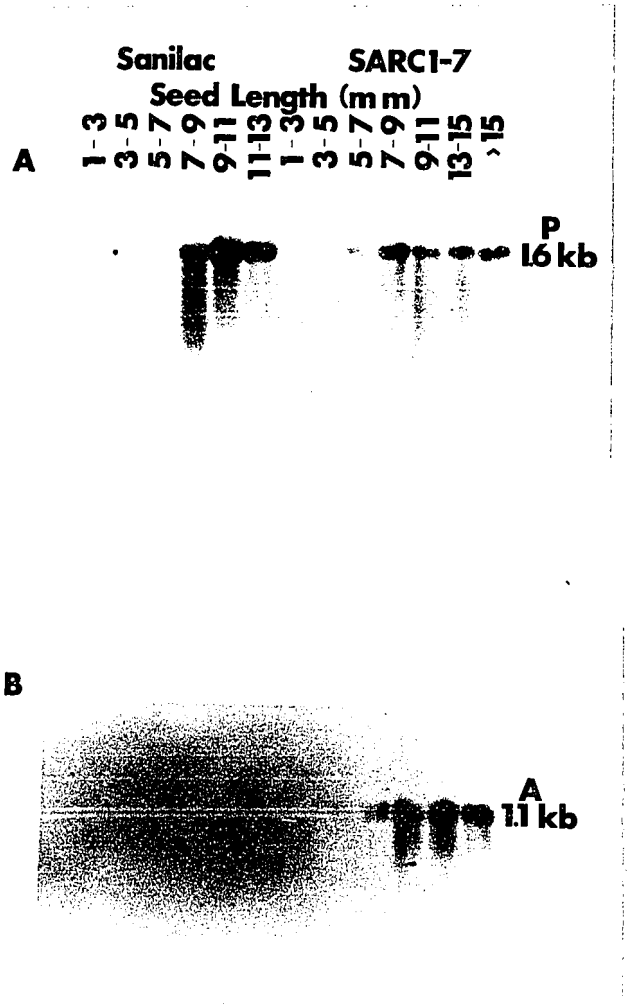


Figure 9. Northern blot analysis of developing bean seeds. A, probed with a 4-kb phaseolin gene from pSB9; B, with the arcelin-specific 19mer. Phaseolin (P) and arcelin (A) mRNA are 1.6 and 1.1 kb in length, respectively.

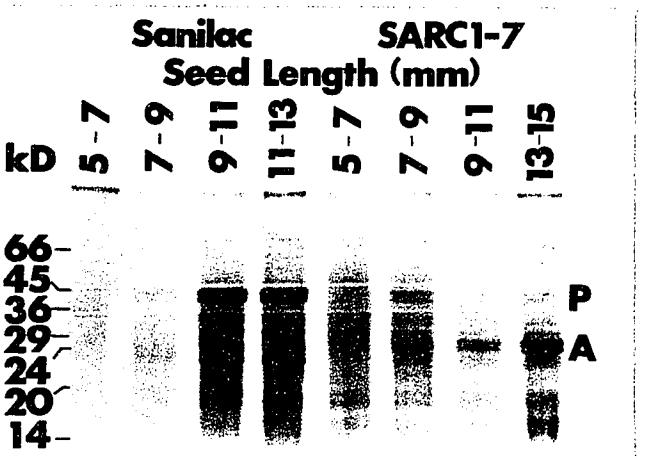


Figure 10. *In vitro* translation of RNA from developing beans. Translation products were analyzed by SDS-PAGE and autoradiography. Phaseolin (P) and arcelin (A) polypeptides are identified by their molecular weights.

the early development, but decreases at the later stages. This reduction coincides with the increase of two other polypeptides, having molecular weights similar to that of arcelin (about 30 kD).

In summary, the expression patterns of the phaseolin and arcelin genes at the RNA level during seed development of both bean lines correspond to expression at the protein level (Osborn et al, 1988a).

Relative transcription rate. The level of a mRNA is determined by the rate of transcription and its half-life. The relative transcription rates can be estimated by nuclear run-on transcription assays. Table 4, Figures 11 and 12 summarize the volumes of nucelei used and the results of 3 nuclear run-on experiments. The intensity of the hybridization signals (the dots) were corrected for amounts of input labeled RNA and size of transcripts using the following equation:

$$\text{Relative transcription rate} = \frac{\text{intensity of hybridization}}{\text{input cpm X size}}$$

The unprocessed phaseolin transcript is about 1.9 kb, containing 5 introns, whereas the arcelin transcript is about 0.8 kb with no intron.

In experiment 1, the run-on products from the 7-9 and 9-11 mm Sanilac seeds were insufficient to yield readable hybridization signals (Figure 11A and Table 4). The concentration of nuclear material in each of these 2 preparations was probably very low as compared to the other stages. Therefore, the amounts of nuclei from these 2 stages were increased for the next 2 experiments (Table 4). The hybridization buffer of 9-11 mm SARC1-7 sample leaked during incubation in experiment 3 (Figure 11C) and the experiment was repeated for the SARC1-7 seeds using the remaining run-on products (Figure 11D). The results of the 3 experiments could not be combined and averaged because

Table 4. Summary of nuclear run-on assays.

	SANILAC				SARC1-7			
	5-7 mm	7-9 mm	9-11 mm	11-13 mm	5-7 mm	7-9 mm	9-11 mm	11-13 mm
Experiment 1								
nuclei (ul)	10	20	20	20	20	10	10	30
% 32P incorporation	42.7	7.6	7.5	13.6	4.1	55.2	60.7	27.8
input cpm	43175	4988	3412	23200	21638	18788	19675	21313
hybridization signal(*):								
phaseolin	0.0296	-	-	0.0266	0.0058	0.0559	0.0726	0.0608
arcelin	0	0	0	0	0	0.0143	0.0485	0.0547
relative transcription rate:								
phaseolin	3.61	0.00	0.00	6.03	1.41	15.66	19.42	15.01
arcelin	0.00	0.00	0.00	0.00	0.00	9.51	30.81	32.08
Experiment 2								
nuclei (ul)	10	35	35	10	10	5	5	10
% 32P incorporation	29.6	11	8.2	20.1	23.9	54	35.4	24
input cpm	20775	95725	50200	21500	92425	15475	29475	141125
hybridization signal(*):								
phaseolin	0.0045	0.0584	0.0372	0.0224	0.0342	0.0103	0.0398	0.0825
arcelin	0	0	0	0	0.0139	0.0041	0.0195	0.0558
relative transcription rate:								
phaseolin	1.14	3.21	3.90	5.48	1.95	3.50	7.11	3.08
arcelin	0.00	0.00	0.00	0.00	1.88	3.31	8.27	4.94
Experiment 3								
nuclei (ul)	10	40	40	10	10	10	10	20
% 32P incorporation	26.5	4.3	18.7	ND	10.2	37.3	ND	8.6
input cpm	486925	44175	53925	82100	101975	179500	57148	58550
hybridization signal(*):								
phaseolin	0.1108	0.0248	0.0429	0.1139	0.0521	0.1045		0.0512
arcelin	0	0	0	0	0.035	0.0693		0.0583
relative transcription rate:								
phaseolin	1.20	2.95	4.19	7.30	2.69	3.06		4.60
arcelin	0.00	0.00	0.00	0.00	4.29	4.83		12.45
Experiment 4								
input cpm					80175	67100	59550	62900
hybridization signal(*):								
phaseolin					0.0503	0.0926	0.082	0.0753
arcelin					0.0196	0.0303	0.0477	0.0679
relative transcription rate:								
phaseolin					3.30	7.26	7.25	6.30
arcelin					3.06	5.64	10.01	13.49

*: The output printouts of the scanner were zeroxed. The peaks were cut and weighted as the relative intensity of hybridization signals.

ND: not determined

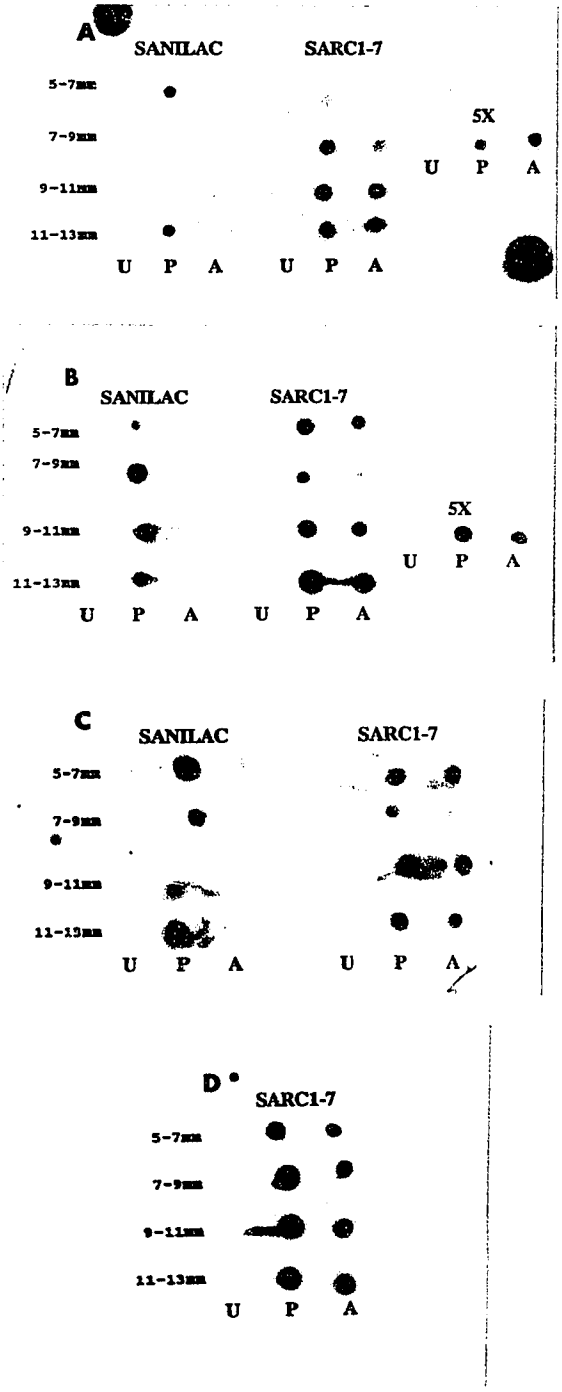


Figure 11. Nuclear run-on assays. The labeled transcripts were detected by dot blot hybridization with U, pTZ18U; P, pPase; and A, pAse (see **Detection of arcelin-1 and phaseolin transcripts**).

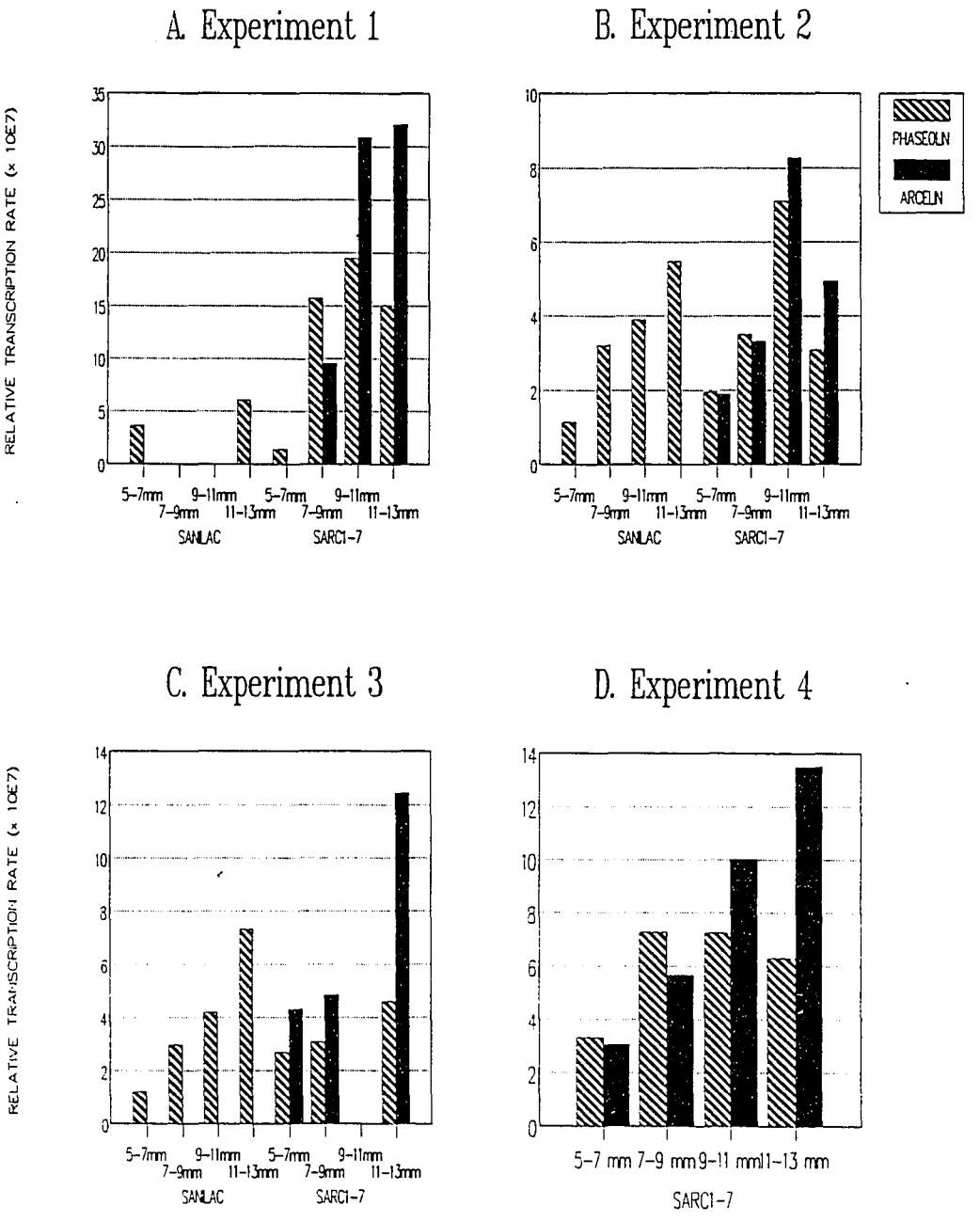


Figure 12. Relative transcription rates of phaseolin and arcelin genes during development.

the intensity of hybridization is only comparable within a single experiment but not among different experiments. This is due to the different length of exposure time for autoradiography. Also, the densitometry readings were not based on the same scale.

As arcelin and lectin genes share considerable sequence homology, there was a concern about the cross-hybridization of these 2 genes. However, the transcription rate of lectin genes was probably very low as no detectable hybridization with the arcelin DNA was observed in the Sanilac seeds (Figure 11).

In general, the phaseolin genes are transcribed with increasing rates during the development of Sanilac seeds (figure 12). In the SARC1-7 seeds, a similar increase in transcription rates was observed for the phaseolin genes during the early stages of seed development but the rates decreased during the later stages. The transcription rates of arcelin genes were relatively low at the early stages but increased at the later stages. These transcription patterns resemble what were observed at the RNA level. However, it is not certained that the differences in the transcription rates are sufficient to account for the differences at the mRNA level.

Possible mechanisms of gene interaction.

Transcriptional modulation is the major control mechanism for the spatial and temporal expression of seed storage

protein genes (Gatehouse et al, 1986). Chappell and Chrispeels (1986) showed that phaseolin genes are transcribed at a higher rate (59 ppm) than the lectin genes (36 ppm) in the cultivar Greensleeves. Arcelin-1 and phaseolin genes contain distinct sequence elements in their 5' flanking regions, that might lead to their different levels of expression. A possible example is the competition for common trans-acting factor(s) by the homologous sequence motifs. Even non-homologous motifs might compete for the same factor(s) through the binding of other factor(s). In addition, arcelin-1 genes might contain activating sequences which are not present in the phaseolin genes, or they might lack negative elements which are present in the phaseolin genes (Bustos et al, 1991).

Seed protein genes of pea (legumin and lectin, Thompson et al, 1989) and soybeans (Walling et al, 1986), however, are regulated at both transcriptional and post-transcriptional stages. Although transcribed at similar rates, β-Conglycinin, glycinin, and Kunitz trypsin inhibitor genes exibit different levels of mRNA accumulation (Walling et al, 1986). The level of an mRNA is determined both by the frequency of transcription initiation and its half-life. The phaseolin mRNA may be more labile than arcelin mRNA. This would result in a lower level of phaseolin mRNA accumulation and subsequently a lower level of protein accumulation. The 5' end of the

phaseolin mRNA lacks RY repeats that have been suggested to be related to RNA stability (Voelker et al, 1987).

Vancanneyt et al (1990) reported that the accumulation of potato patatin mRNA in transgenic tobacco was strongly affected by its translatability. In yeast, the degradation of mRNA also depends upon the concomitant translational elongation. The unstable mRNAs contain more rare codons than do stable mRNAs (Herrick et al, 1990). Codon usage also determines translation rate in *Escherichia coli* (Sorensen et al, 1989). The sequence around the translation start site and codon usage may provide some clues as to the translatability of phaseolin and arcelin mRNAs. The consensus sequence of the translation start site for plant genes has special preference for G and C residues at the +4 and +5 position, respectively, (Lütcke et al, 1987, Figure 13).

```
PHVARC1-1              .TAC.....
plant consensus (-4)AACAATGGC(+5)
phaseolin              T..T...AT
("." = identical nucleotide)
```

Figure 13. Arcelin and phaseolin gene translation start sites vs. plant consensus.

Lütcke and co-workers (1987) suggested that these two nucleotides might be involved in translation initiation. The translation start sequence in the arcelin mRNA has a slightly higher identity with the consensus, especially the

ATGGC residues. Phaseolin translation start site has a poorer match with the conserved sequence. It might be possible that the initiation of arcelin translation is more favorable.

The codon usage of phaseolin, arcelin, and two lectin genes is compared in Table 5. In general, the arcelin and lectin genes share more similar codons. All genes have similar codon usage for phenylalanine, serine, and asparagine. However, the usage for leucine, isoleucine, threonine, alanine, histidine, aspartate, and glycine, G and/or C residues are preferred in the third (wobble) position in the arcelin genes. A similar comparison has also been described for the seed storage proteins of oats, globulin and avenin (Shotwell et al, 1990). The authors speculated that developing seeds might contain a tRNA population with a certain set of anticodons. mRNAs with codons that are favored by the translation machinery would be preferentially translated and be more stable. In this case, the arcelin mRNA may be favored for translation as compared to phaseolin mRNA. Consequently, the phaseolin mRNA may be degraded faster, resulting in a lower level of expression. It is, however, rather costly to control storage protein synthesis in this manner. Furthermore, since the lectin and arcelin genes
share similar codon usage, this would not be the only controlling mechanism.

Table 5. Codon usage of arcelin (A), phaseolin (P), lectin (PHA-L & E) genes.

		P	A	PHA-L	PHA-E			P	A	PHA-L	PHA-E
Phe	TTT	8	3	4	5	Tyr	TAT	5	4	1	2
	TTC	17	12	15	9		TAC	8	2	4	4
Leu	TTA	2	2	1	1	end	TAA	0	0	0	0
	TTG	7	1	3	2		TAG	0	1	1	1
	CTT	14	6	7	8	His	CAT	5	1	2	1
	CTC	9	12	11	11		CAC	6	5	1	4
	CTA	5	4	2	7	Gln	CAA	15	2	3	8
	CTG	8	2	4	6		CAG	14	0	4	1
Ile	ATT	10	2	2	2	Asn	AAT	10	5	8	6
	ATC	9	12	10	7		AAC	20	22	17	15
	ATA	6	0	1	1	Lys	AAA	15	9	6	7
Met	ATG	5	2	1	1		AAG	10	6	4	4
Val	GTT	11	5	6	7	Asp	GAT	13	1	5	3
	GTC	4	7	6	6		GAC	8	11	11	13
	GTA	2	0	0	0	Glu	GAA	18	10	5	5
	GTG	11	7	11	8		GAG	21	7	4	4
Ser	TCT	11	11	10	8	Cys	TGT	0	0	0	0
	TCC	9	11	12	10		TGC	0	2	0	2
	TCA	4	3	2	6	end	TGA	1	0	0	0
	TCG	1	2	0	2	Trp	TGG	1	2	5	4
Pro	CCT	5	2	1	4	Arg	CGT	3	4	1	1
	CCC	4	4	8	3		CGC	1	2	2	2
	CCA	5	0	0	1		CGA	1	0	1	0
	CCG	1	1	0	1		CGG	1	3	1	0
Thr	ACT	4	2	3	4	Ser	AGT	5	0	0	3
	ACC	5	12	10	4		AGC	10	4	7	3
	ACA	2	3	5	4	Arg	AGA	6	1	0	1
	ACG	3	3	5	5		AGG	7	1	3	1
Ala	GCT	7	4	5	5	Gly	GGT	8	1	3	3
	GCC	8	12	10	9		GGC	1	9	9	5
	GCA	10	2	2	5		GGA	13	1	4	4
	GCG	0	0	0	2		GGG	3	2	4	2
Subtotal		207	149	157	154			229	118	116	109
Total		435	266	272	262						

In summary, transcriptional control is a regulatory mechanism to be considered in regarding the level of phaseolin and arcelin gene expression. However, post-transcriptional control, including mRNA degradation and translatability, may also play an important role.

CHAPTER III. EXPRESSION OF THE ARCELIN-1 AND THE β-PHASEOLIN GENES IN TRANSGENIC TOBACCO SEEDS

Introduction

The β-phaseolin gene and a number of other seed storage protein genes from various plant species have been transformed into tobacco plants and their temporal and spatial expression have been examined (see **Literature Review**). The regulatory elements and factors required for seed storage protein expression seem to be conserved between tobacco plants and other species. One of the objectives of this research is to establish a transgenic expression system so that the expression and possible interaction of the arcelin and phaseolin genes can be dissected and so that the *cis*- and *trans*-acting regulatory elements can be analyzed in the future.

This expression system also has an applied interest. Genetic engineering is an innovative way for crop improvement in agriculture. The phaseolin promoter is a very strong, seed-specific promoter that has been studied intensively and used in plant transformation. However, for crop plants such as common beans, which express high level of phaseolin, the transformed gene(s) under the control of phaseolin promoter will necessarily compete with all the members of the phaseolin gene family. It would be advantageous to have a promoter that could override the

expression of phaseolin genes. The arcelin promoter may provide such a system.

In order to compare the promoter strength of the phaseolin and arcelin genes, the two genes were fused and inserted into a binary vector, pBI121, and transformed into tobacco plants. The effect of different gene copy numbers is thus eliminated. Furthermore, position effects will be minimized, since these 2 genes are adjacent on the same construct. The genes are likely to have a similar environments in the tobacco genome. This system would allow analyses and comparisons of the expression of bean phaseolin and arcelin genes in a foreign model plant with the source plant, SARC1-7. In this study, chimeric genes were be constructed in which the promoters of the 2 genes are exchanged. This exchange allows the determination if difference in phaseolin and arcelin gene expression is caused solely by the respective promoters.

Materials and Methods

Chemicals. All chemicals (Fisher Scientific or otherwise noted) were reagent grade or better.

Plant materials. Seeds of tobacco (*Nicotiana tabacum* cv Xanthi.nc) were sterilized by shaking in 10% SDS and 10% Chlorox for 30 min. After rinsing with sterile water, the seeds were germinated on MSO medium (Murashige and Skoog salts from Sigma, containing 0.1 g/L *myo*-inositol, 0.4 mg/L

thiamine·HCl, 0.5 mg/L nicotinic acid, 0.5 mg/L pyridoxine·HCl, and 7.5 g/L bacto agar) and the seedlings were grown in a growth chamber (Percival I-37L) with continuous light at 28°C. Rooted plants were transplanted into soil and grown in a greenhouse. Developing seeds from transgenic tobacco were collected 12 to 26 days after flowering (DAF) and stored at -70°C.

Bacterial strains. Plasmid DNA's were manipulated mostly in *E. coli* DH5a (BRL). *E. coli* CJ236 (*dut*⁻ and *ung*⁻) and MV1190 (Bio-Rad) were used for site-directed mutagenesis. Helper plasmid pRK2013 in *E. coli* HB101 and *Agrobacterium tumefacien* LBA4404 (Clontech) were used for plant transformation.

Plasmid construction. A single cloning site (*Hind*III) on the binary vector pBI121 (Clontech, Jefferon et al, 1987) was chosen for the insertion of the arcelin-1 and phaseolin genes to avoid partial digestion and/or blunt end ligation. A 2.5 kb *Bgl*II/*Kpn*I (5' -> 3') arcelin-1 fragment from λPv222 was cloned into pVZ1 (designated pVZ1-A2). This fragment contains 1 kb of the 5' flanking region, 0.8 kb of the coding region, and 0.7 kb of 3' the region. The 4-kb phaseolin sequence in pSB9 contains 1 kb, 1.9 kb, and 1.1 kb of the 5', coding, and 3' regions, respectively. Both plasmids have a *Hind*III site in the polylinker region at the 5' end and a *Kpn*I site at the 3' end. To create a *Hind*III site at the 3' end, these 2

clones were digested with *Kpn*I (BRL), dephosphorylated (CIP, Boehringer Mannheim), and ligated either to each other or to a 1.4 kb *Kpn*I/*Sca*I fragment from pVZ1 (Figure 14). The resulting molecules were then digested with *Hind*III to produce a 6.5-kb phaseolin-arcelin (5'->3' + 3'->5') fragment, a 2.5-kb arcelin-1 fragment, or a 4-kb phaseolin fragment, which were gel purified and inserted into pBI121 (Figure 15A and B). The direction of the insertion was diagnosed by *Eco*RI digestion. Three clones, pPA3 [phaseolin (5'->3') + arcelin (3'->5')], pA3 (3'->5'), and pP7 (5'->3') were selected for tobacco transformation.

Site-directed mutagenesis. The *Muta-Gene Phagemid in vitro Mutagenesis* kit (Bio-Rad) was used to create *Ase*I sites (ATTAAT) at the transcription start sites of the arcelin-1 and phaseolin genes for promoter switching. This method was originally developed by Kunkel (1985) and Kunkel et al (1987). The 2.5-kb *Bgl*II/*Kpn*I arcelin-1 fragment and the 4-kb *Sal*I/*Bam*HI fragment were cloned into the pTZ18U and the resulting plasmids were named pA9 and pP4, respectively. The single strand, uracil-containing template DNA's were purified after growth in the CJ236. An 18-base oligonucleotide (Oligo.AseIA18.2) with the sequence of 5'-ATGGATGATTAATGCATG-3' was synthesized (by NBI) to create a G-to-A change for pA9 (Figure 16). For pP4, a 25mer (5'-TCACTCACTTCATTAATCATCCATC-3', Oligo.AseIP25) was synthesized (by NBI) to create a T-to-A and a C-to-A

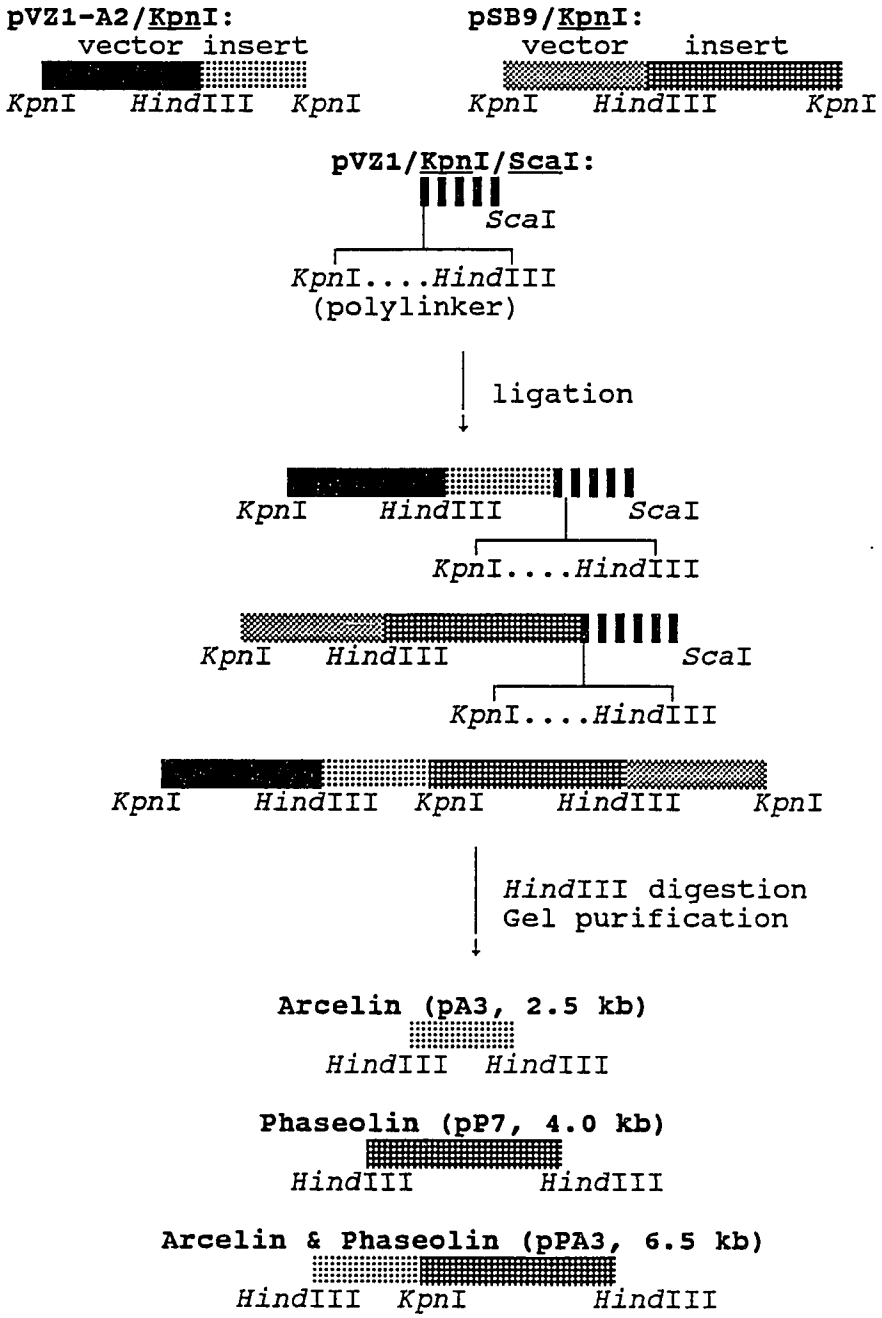
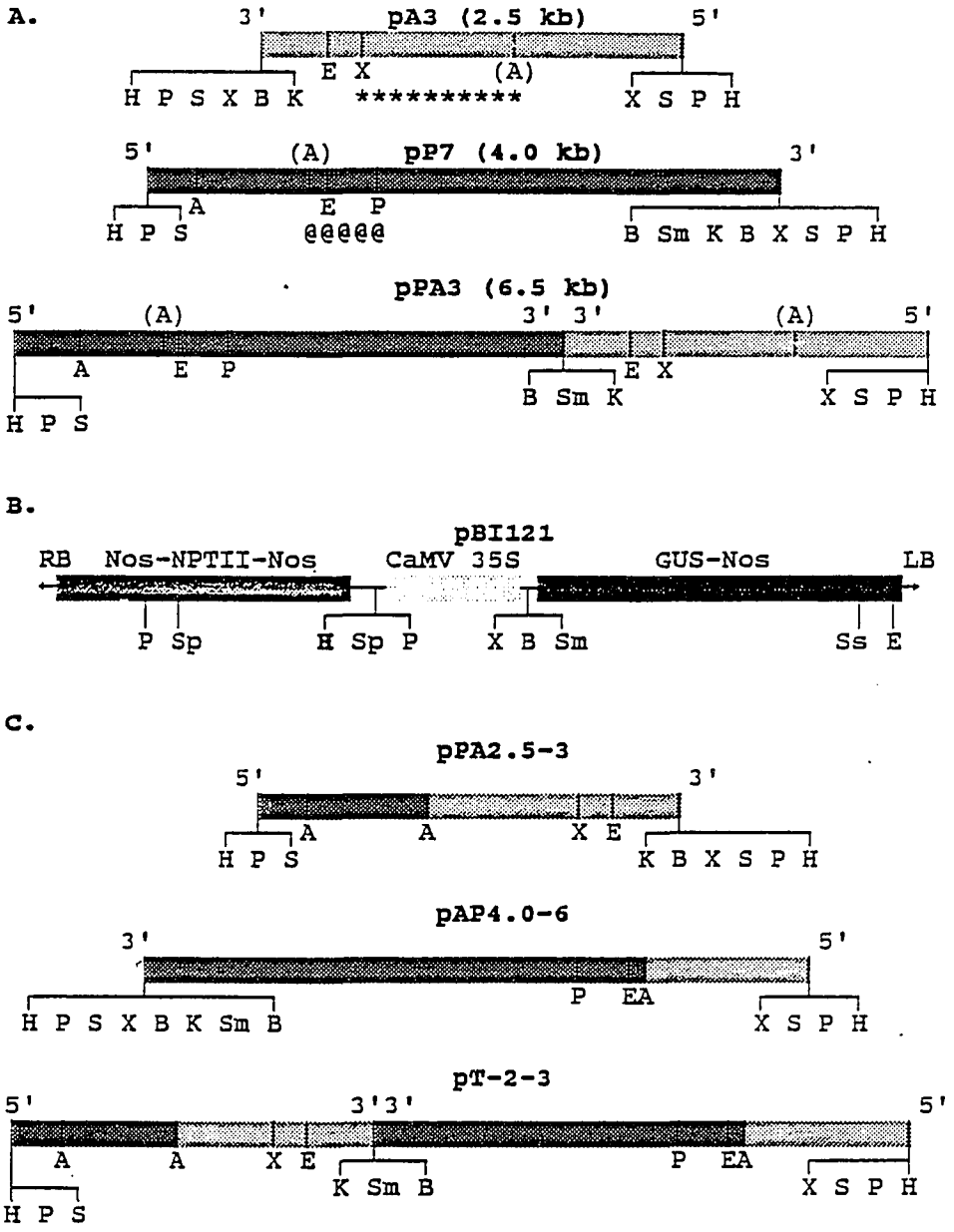


Figure 14. Strategy to create HindIII sites at the 3' ends of the inserts containing the phaseolin and/or arcelin genes (see **Plasmid construction**).



H, HindIII; P, PstI; S, SalI; A, AseI; E, EcoRI; B, BamHI; Sm, SmaI; K, KpnI; X, XbaI; (A), position of AseI site created by site-directed mutagenesis; Sp, SphI; Ss, SstI; ***, a 0.8 kb fragment used as an arcelin probe; and @@@, a 0.5 kb fragment used as a phaseolin probe. pPA3 and pT-2-3 are not drawn to scale.

Figure 15. Restriction maps of the inserts of A, pA3, pP7, & pPA3; B, pBI121; and C, pPA2.5-3, pAP4.0-6, & pT-2-3.

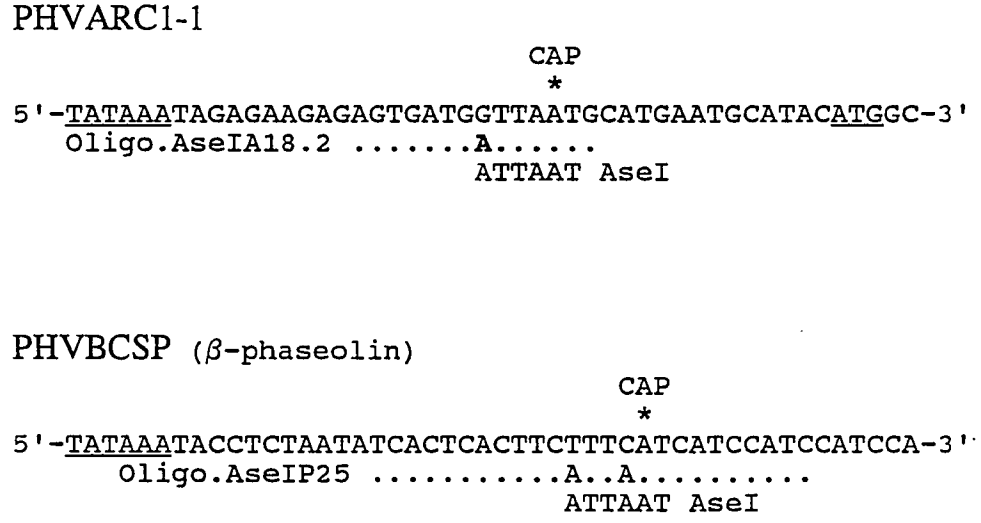


Figure 15. Oligonucleotides for site-directed mutagenesis.

change. The oligonucleotides were purified using the method described by Sambrook et al (1989). The mutant clones (named pAase and pPase) were identified by *Ase*I digestion and confirmed by sequencing. The 1-kb *Hind*III/*Ase*I arcelin 5' region from pAse was ligated with the 3-kb *Ase*I/*Kpn*I phaseolin coding and the 3' region from pPase and cloned into the pTZ18U (pA'P'29). Since pPase contains 1 existing *Ase*I site upstream from the new site, partial digestion (1 unit/9.9 μg DNA) was performed to isolate the 1-kb *Hind*III/*Ase*I phaseolin 5' region. This fragment was then fused with the 1.5-kb *Ase*I/*Kpn*I arcelin

coding region and 3' region and cloned into the pTZ18U (pP'A'9). The inserts of pA'P'29 and pP'A'9 were cloned into the pBI121 using the strategy described above (see Plasmid construction), except that a 1.2-kb *Kpn*I/*Sca*I fragment from the pBluesript II KS +/- instead of the 1.4-kb fragment from the pVZ1 was used to create the HindIII sites at the 3' ends (Figure 15C). Three clones, pPA2.5-3 (5'->3'), pAP4.0-6 (3'->5'), and pT-2-3 (5'->3' + 3'->5') were selected for tobacco transformation.

Triparental mating between *Agrobacterium* **and** *E. coli*. This method was originally described by Margossian (ARCO, PCRI) and revised by Sun (personal communication). The donor cells, DH5 containing pBI121 or its derivatives, and the helper, HB101 containing pRK2013, were grown in LB with 50 mg/L of kanamycin at 30°C. The recipient, *Agrobacterium* LBA4404, was grown at 30 °C in Min A/sucrose medium (see p.78). Overnight cultures (late log phase, O.D.660=1.2 to 1.8) from single colonies were harvested by centrifugation at 4,000 rpm (Sorvall SS34 rotor) for 5 min and resuspended to an equal O.D. at 600 nm in drug free media. Mating was performed by mixing 200 μl each of the 3 cell types. The mixture was allowed to incubate undisturbed at room temperature for at least 15 min. Two 100 μl aliquots of the mating mix were spotted on two 0.45 μm nitrocellulose (2 X 2 cm^2, MSI) filters which were placed on a dry LB plate. Each plate also included 1 filter with the donor

alone and 1 filter with the recipient alone. The plates were incubated at 30°C overnight with the cover side up. The filters were then vortexed in 2 ml of 0.9% NaCl vigorously to remove the cells. The suspensions were diluted with 0.9% NaCl and 10^{-5} to 10^{-7} dilutions were plated on ABS plates (see below) containing 50 mg/L of kanamycin. A 10^{-7} and a 10^{-8} dilution were also plated on LB plates without antibiotics to check viability. The colonies of LBA4404 containing pBI121 or the derivatives appeared after 2 to 3 days whereas *E. coli* colonies appeared after only 1 day. In general, the counts on the LB plates were 10 to 40 times higher than those on the ABS/kanamycin plates. The LBA4404 colonies were colony-purified twice before verification by boiling plasmid mini preparation and restriction enzyme digestion.

Medium compositions:

Min A medium (per liter):
- 200 ml 5X min A salts:
  - 52.5 g/L K_2HPO_4
  - 22.5 g/L KH_2PO_4
  - 5 g/L $(NH_4)_2SO_4$
  - 2.5 g/L Na-citrate
- 790 ml distilled water
- 15 g Difco agar
- autoclaved 20 minutes and add sterile:
- 1 ml 1 M $MgSO_4 \cdot 7H_2O$
- 10 ml 20% sucrose

ABS medium (per liter):
- 1 ml 0.013 g/ml $CaCl_2 \cdot 2H_2O$
- 0.31 g $MgSO_4 \cdot 7H_2O$
- 1 ml 2.5 mg/ml $FeSO_4 \cdot 7H_2O$
- 5 g sucrose
- 20 g Difco agar
- add distilled water to 900 ml, autoclave for 20

minutes, cool to 50 to 55 °C, and add 100 ml autoclaved 10X ABS salts:

39.3 g/L $K_2HPO_4 \cdot 3H_2O$
10 g/L NaH_2PO_4
10 g/L NH_4Cl
1.5 g/L KCl

Tobacco transformation and regeneration. The method originally reported by Horsch et al (1985) and modified by Chen and Kunhle (personal communication) was used. A single colony of LBA4404 harboring pBI121 or its derivatives was cultured in 5 ml of LB containing 50 mg/L of kanamycin at 28°C with shaking (220 rpm). On the second day, 2 μl of 0.5 M acetosyringone (in DMSO) was added and the culture was incubated under the same condition for another day. The culture (1 to 8 X 10^9 cfu/ml) was then diluted 10 fold with LB. Tobacco leaves from 6 to 8 week old plants were cut in the bacterial suspension into 0.5 X 1 cm^2 pieces (avoid mid ribs), blotted on sterile paper towels, and placed on 1/2MSO (same as MSO media, Sigma, except with half concentration of the MS major salts) plates. After co-cultivation at room temperature in the dark for 2 to 3 days, the explants were transferred to 1/2 MS104 plates (1/2 MSO with 1 mg/L N^6-benzyladenine and 0.1 mg/L α-naphthaleneacetic acid) containing 500 mg/L of carbenicillin (Geopen) and 300 mg/L of kanamycin. The plates were incubated in a growth chamber and the explants were transferred to fresh media every 3 weeks. After 6 to 8 weeks, the tips (with 2 to 3 expanded leaves) of the

regenerated shoots were transferred to 1/2 MSO medium with 500 mg/L carbenicillin and 100 mg/L kanamycin in Magenta boxes. The shoots were then transferred again to 1/2 MSO with 100 mg/L kanamycin but without carbenicillin. Rooted plantlets without bacterial contamination were then transplanted to 4-inch pots with soil (soil:vermiculite:perlite = 3:1:1), covered with plastic cups, and grown in the growth chamber (Percival I-37LLVL) with 16 hr light and 8 hr darkness at 28°C. After 2 to 3 days, they were transferred to the greenhouse and the plastic cups were removed. The plants were later transplanted to 1 gallon pots.

GUS assay. Leaves from tissue cultured plants were cut into approximately 0.3 X 0.3 cm^2 pieces. Two to 3 pieces from each plant were placed into a white sterile microfuge tube containing GUS staining buffer [100 mM Na-phosphate, pH 7, 0.1% Triton X-100, 0.5 mM K-ferricyanide ($K_3Fe(CN)_6$), 0.5 mM K-ferrocyanide ($K_4Fe(CN)_6$), 0.05% X-glucuronide (X-gluc, dissolved in DMF), and 1 mM EDTA] and incubated at 37°C overnight (Sun, personal communication). To eliminate the interference of chlorophyll, 0.5 ml of 100% ethanol was added and the mix was incubated at room temperature overnight.

DNA extraction. Genomic DNA was isolated from tobacco leaves, using the CTAB protocol reported by Doyle et al (1989). The RNase treatment and the final precipitation

step were omitted. RNase A was added to the restriction digests (12.5 to 50 ng/μl final concentration) and incubated with the restriction reactions.

Southern analysis. Genomic DNA (8 or 12 μg) was digested with *Hind*III and separated on 0.7% agarose gels. The DNA was transferred to "Hybond N +" membrane (Amersham) by the alkaline transfer method (Reed and Mann, 1985). The hybridization was performed according to the instructions from Amersham.

RNA extraction. RNA from developing seeds was extracted by the method of Altenbach et al (1989). For initial screening, seeds of 20 DAF from each plant were used. Two plants from each of the constructs were selected for the developmental study.

Northern and dot blot analysis. One third to 1/2 of the total RNA extracted from 20-DAF tobacco seeds (from 1 to 2 seed pods) were separated on 1.5% agarose/formaldehyde gels. One μg each of the RNAs from developing seeds (12 to 26 DAF) were used for dot blots. The manufacturer's (Amersham) instruction for "Hybond N +" were followed for Northern transfer, dot blotting, hybridization, and probe stripping.

Results and Discussion

Tobacco transformation and regeneration. A pBI121 vector control and 6 constructs containing phaseolin and/or

arcelin gene(s) under the control of its own or the other's promoter were used to transform tobacco plants (Figure 15). Three to 30 shoots were regenerated from 40 to 60 leaf explants on selecting media (Table 6). Forty five to 77 % of the shoots showed positive GUS staining. Most of the positive staining had a very strong blue color (Figure 17B) with an even darker staining in the veins. This is the typical expression pattern for CaMV 35S promoter (Jefferson et al, 1987). Most of the GUS positive and some of the GUS negative plants were transferred to the greenhouse. The flowers of some of the plants did not set seeds (Table 7). Some plants produced small seedpods with very few or no seeds. These and other abnormalities are expected since the insertion of foreign genes into the tobacco genome may interrupt the genes controlling the normal development of the plants.

Southern analysis and gene copy number estimation. Almost all of the GUS positive plants contain sequences of the phaseolin and/or arcelin genes in their genome (Table 7, Figures 18 and 19).

There are 1 to 5 copies of the phaseolin or arcelin genes in the transformants of pA3, pPA2.5-3, pP7, and pAP4.0-6. However, pPA3 and pT-2-3 transformants have greater than 10 copies of the phaseolin and arcelin genes. The expected sizes of fragments from the *Hind*III digests of genomic DNA were 2.5 kb for the transformants of pA3 and

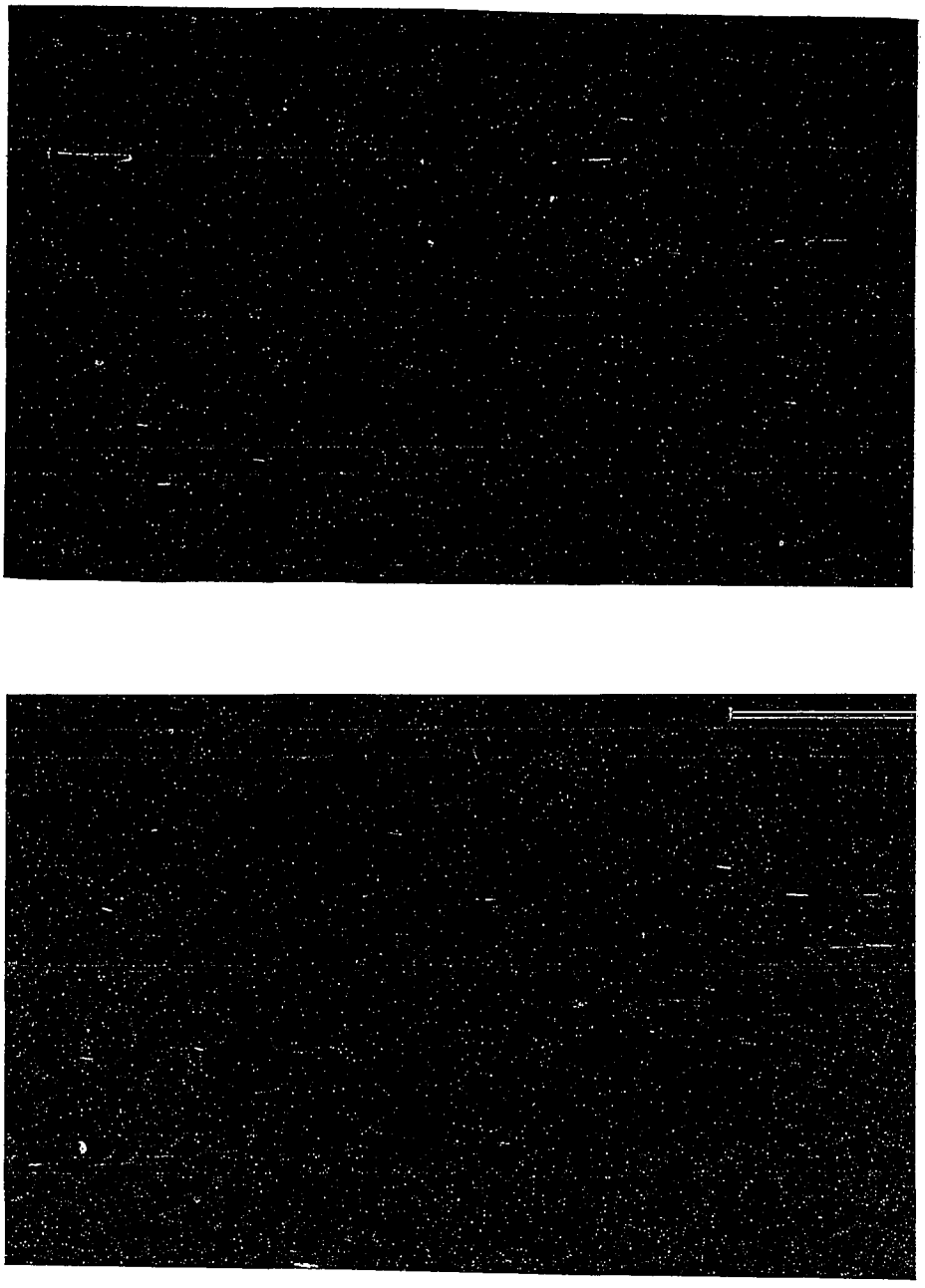


Figure 17. GUS staining of wild-type (A) and pA3 (B) transformed tobacco leaves.

Table 6. Summary of tobacco transformation - 1
(The numbers indicate the numbers of explants, shoots, or plants.)

Construct	Explant	Shoot	GUS+	Greenhouse (GUS+)+(GUS-)
Control: pBI121	40	3	1	1+2
Wild type: pP7	40	30	14	13+5
pA3	40	30	19	18+2
pPA3	60	11	6	4+3
Mutant: pPA2.5-3	60	23	17	16+3
pAP4.0-6	60	20	9	9+2
pT-2-3	60	13	10	9+0

Table 7. Summary of tobacco transformation - 2
(The numbers of plants over the total numbers of plants in the greenhouse or the numbers of positve plants over the plants tested.)

Construct	Seed producer (/greenhouse)	Southern (+/tested)	Northern (+/tested)
Control: pBI121	3/3		
Wild type: pP7	17/18	11/14	13/17
pA3	19/20	13/17	11/15
pPA3	3/7	6/7	3/3
Mutant: pPA2.5-3	11/19	11/13	7/8
pAP4.0-6	8/11	10/11	7/10
pT-2-3	6/9	9/9	6/6

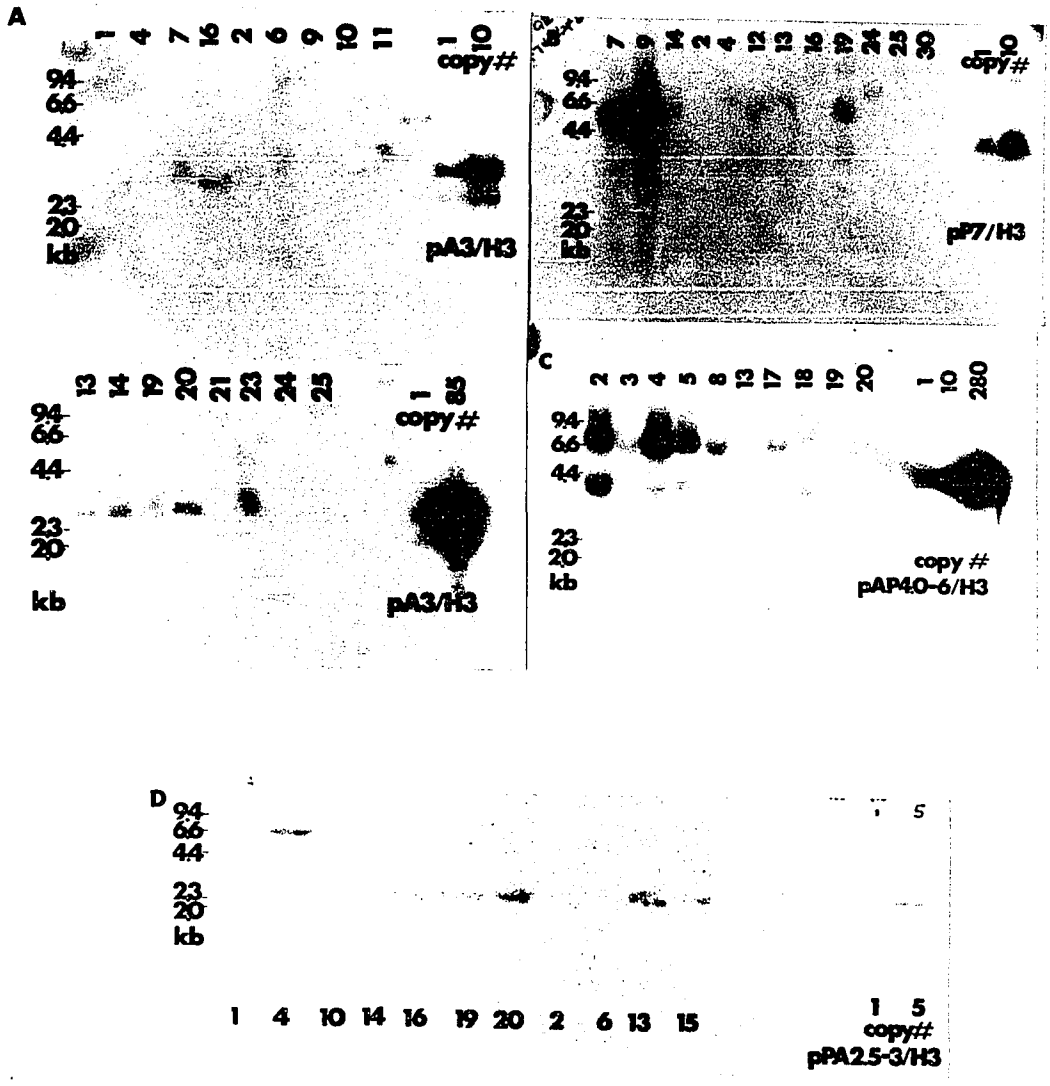


Figure 18. Southern blot analysis of tobacco genomic DNA digested with HindIII. Leaf DNAs were from plants transformed with: A, pA3; B, pP7; C, pAP4.0-6; and D, pPA2.5-3. Gene copy number reconstruction experiments are also included.

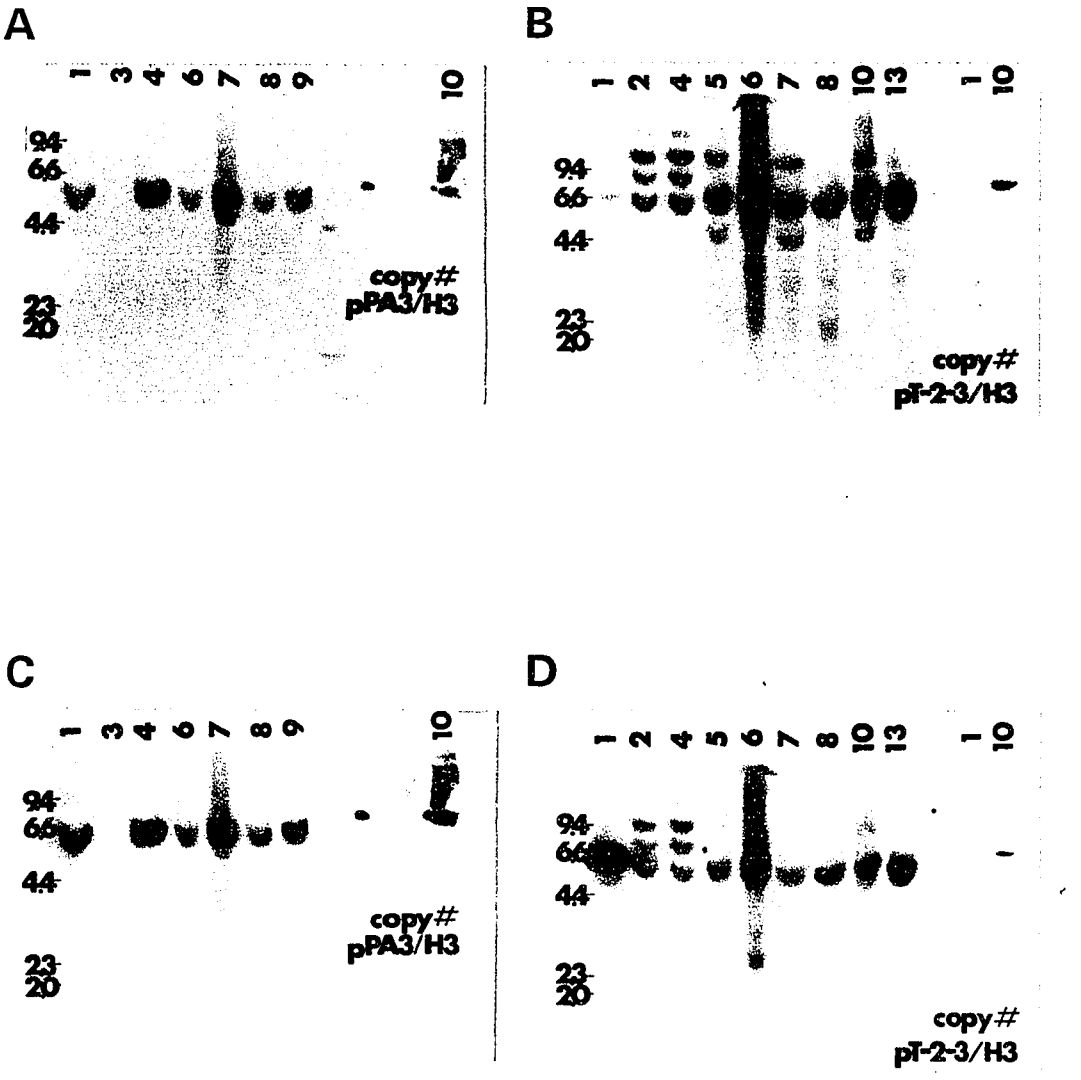


Figure 19. Southern blot analysis of pPA3 (A & C) and pT-2-3 (B & D) transformants. A & B, probed with the arcelin probe; and C & D, with the phaseolin probe.

pPA2.5-3, 4 kb for pP7 and pAP4.0-6, and 6.5 kb for pPA3 and pT-2-3, respectively. The transformants of pA3 and pPA2.5-3 have the correct size bands. However, the transformants from other constructs, in addition to the correct-sized bands, have DNA fragments larger than the expected sizes that hybridized to the phaseolin and arcelin probes. These may have resulted from DNA rearrangement. The multiple bands of pT-2-3 transformants might also be explained by incomplete digestions. However, the pP7 and

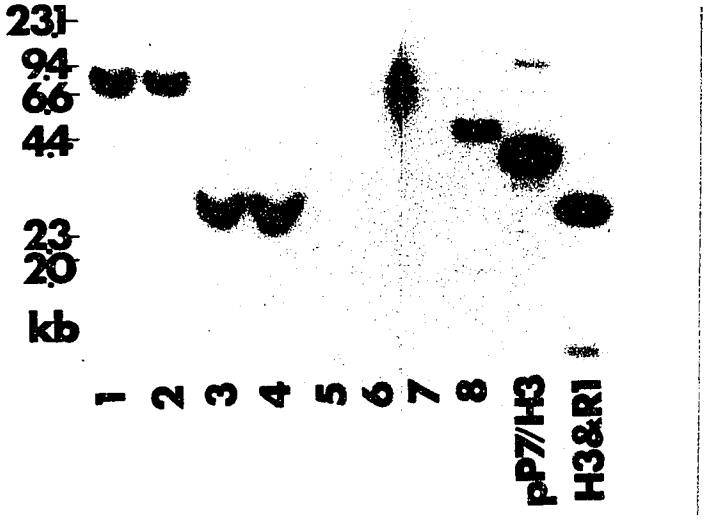


Figure 20. Southern analysis of transgenic tobacco genomic DNA. Lanes 1, 2, 5, & 7, HindIII digests of P7-12, AP4.0-6-13, pBI121-2, & wild-type genomic DNA, respectively; lanes 3, 4, & 6, HindIII & EcoRI double digests of P7-12, Ap4.0-6-13, & pBI121-2; lane 8, EcoRI, BamHI, & PstI triple digest of pBI121 (1000X in excess); pP7/H3, HindIII digest of pP7; and H3&RI, HindIII & EcoRI double digest of pP7.

pAP4.0-3 transformants repeatedly showed larger-sized bands. When double digested with *Hind*III and *Eco*RI (Figure 20), correct-sized bands appeared. This suggests that the sequences around the *Hind*III site at the 3' end of the phaseolin gene, when incorporated into the tobacco genome, may have created unfavorable conditions for the *Hind*III enzyme. Nath and Azzolina (1981) reported that there was a 14-fold difference in the cleavage rate among the *Hind*III sites on λ DNA. Methylation did not appear to be the cause of the preference of the recognition sites.

Expression of the arcelin-1 and β-phaseolin genes during seed development. The number of tobacco plants expressed the arcelin and/or phaseolin genes and the number of plants screened are summarized in Table 7. Phaseolin introns were correctly processed in the tobacco seeds as the mRNA detected was 1.6 kb in length (Figures 21 and 22). Some plants (eg. P7-24 and P7-25) did not express the phaseolin gene although the genes were detected in their genomes.

The pattern of mRNA accumulation during tobacco seed development, as obtained by the dot blot analysis, correlates with the results from the Northern analysis (Figure 23). The dot blot technique was chosen for this investigation (Figures 24 and 25). The timing of arcelin and phaseolin mRNA accumulation in transgenic tobacco seeds is similar to that obseerved in beans. For plants

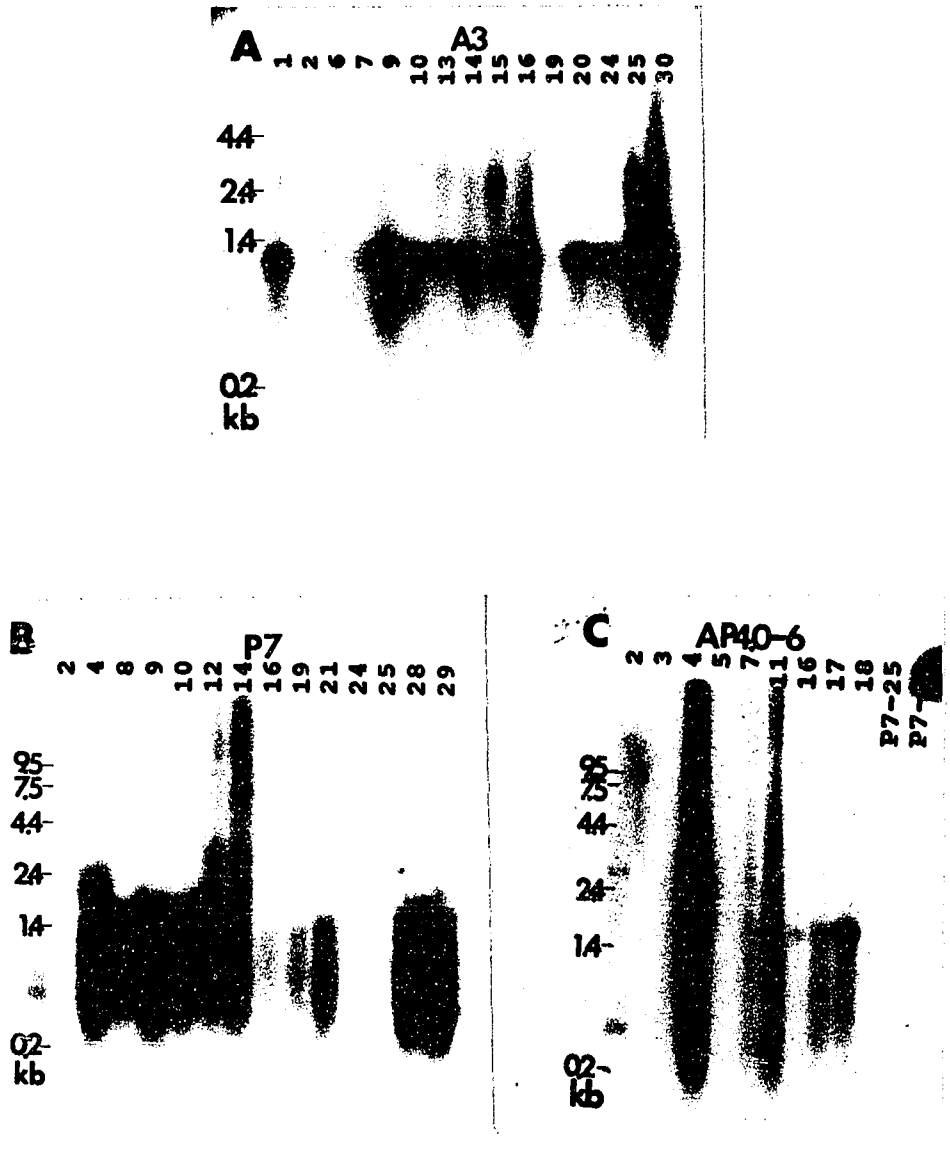


Figure 21. Northern blot analysis of RNA from 20-DAF tobacco seeds - 1. RNA samples were from plants transformed with: A, pA3; B, pP7; and C, pAP4.0-6.

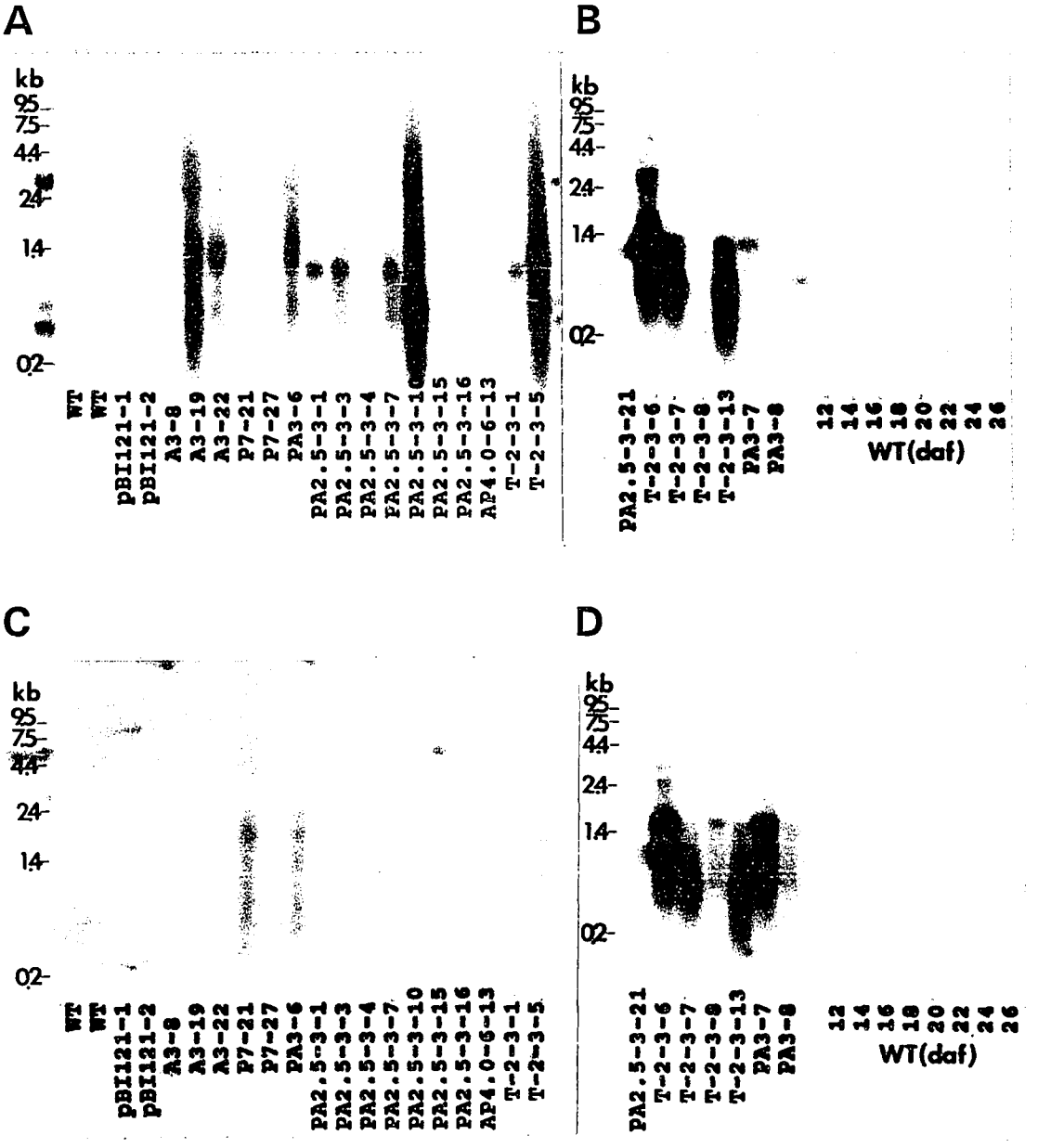


Figure 22. Northern blot analysis of RNA form 20-DAF tobacco seeds - 2. RNA samples were from plants transformed with both arcelin and phaseolin genes. A and B, probed with the arcelin probe; and C and D, with the phaseolin probe.

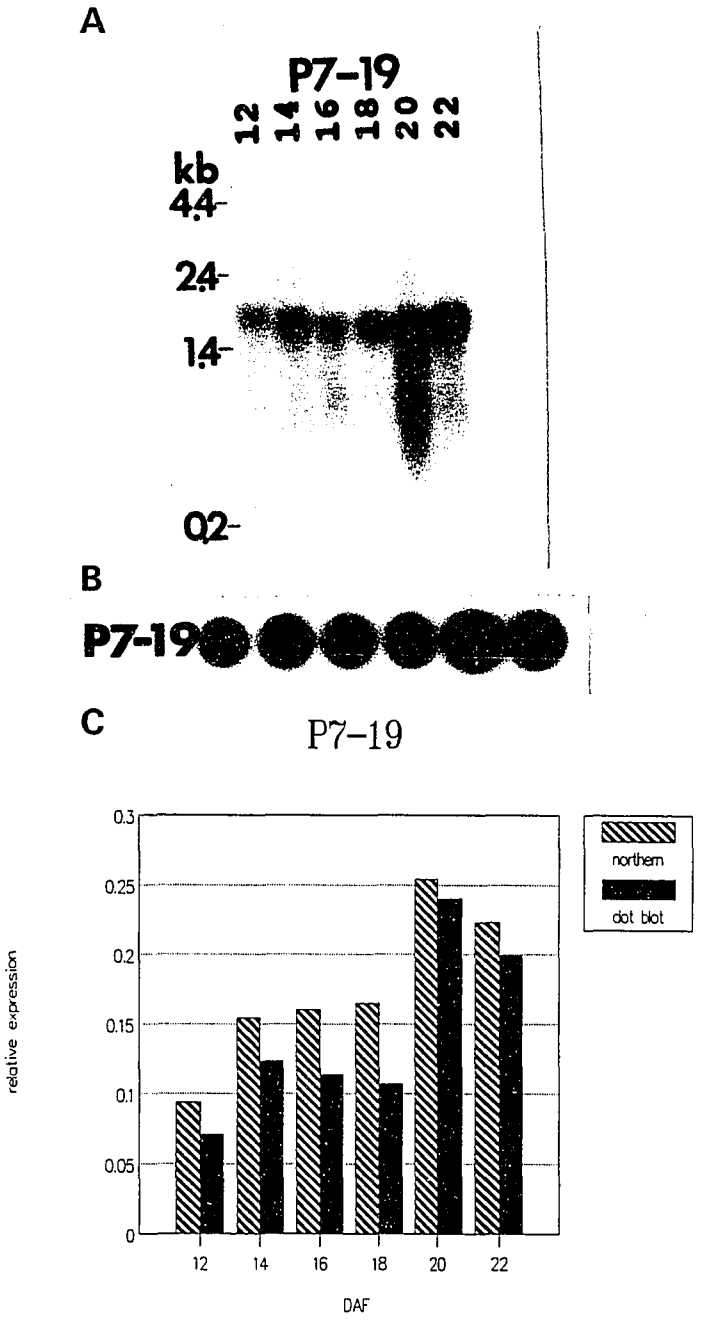


Figure 21. Comparison of Northern and dot blot analyses of developing seeds from P7-19. A, Northern; B, dot blot; and C, relative expression. Both blots were probed with a 550-bp phaseolin coding region fragment. The autorads were scanned. The peaks from the scanner prinouts were cut and weighed as the relative relative expression.

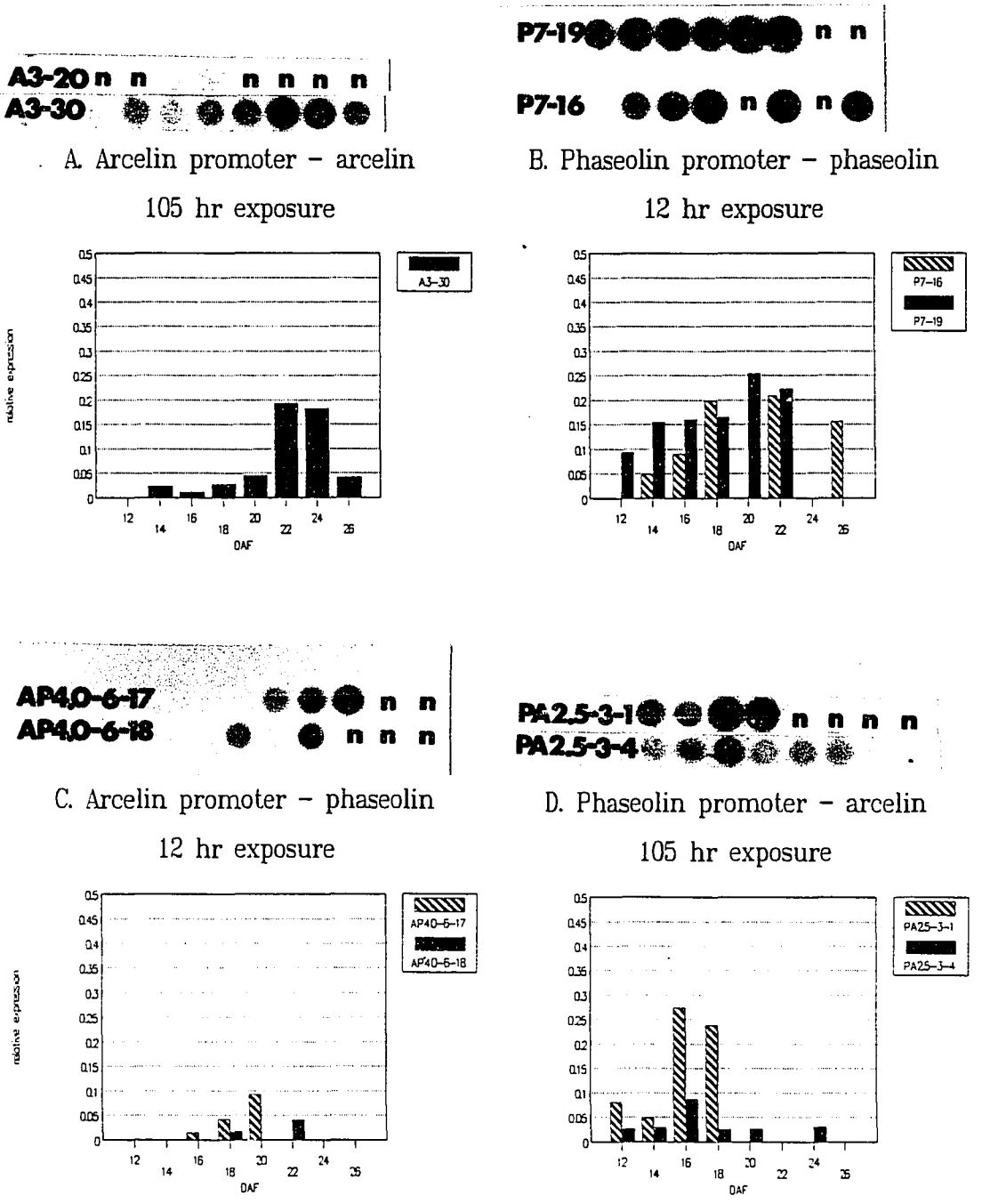
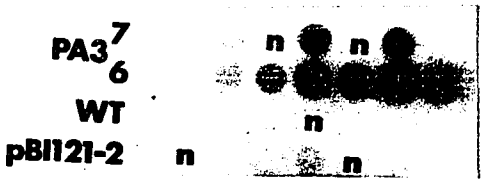
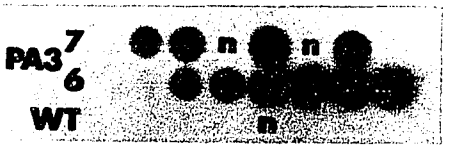


Figure 24. Developmental expression of the arcelin or phaseolin mRNA in single-gene transformed plants. Values for the bar graphs were from the dot blots and corrected for the length and specific activity of the probe. n, no RNA.



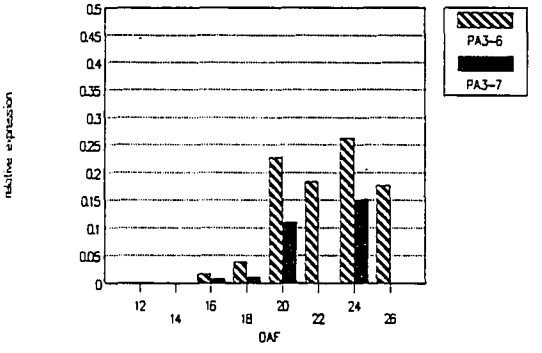
A. Arcelin promoter – arcelin

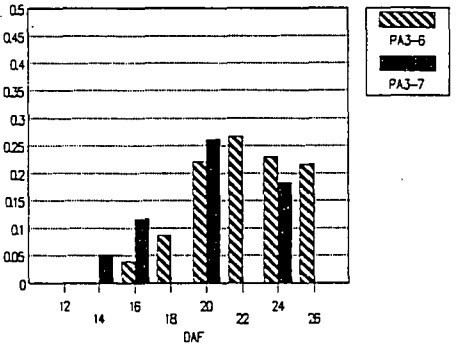
105 hr exposure

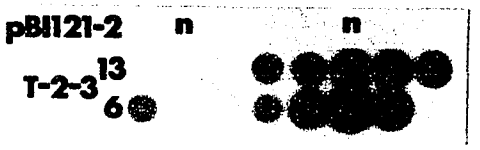


B. Phaseolin promoter – phaseolin

12 hr exposure

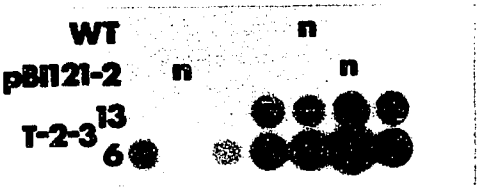






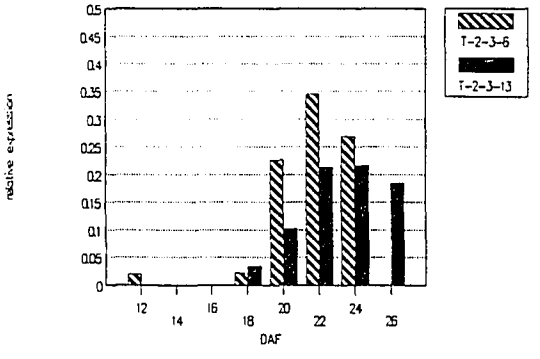
C. Arcelin promoter – phaseolin

12 hr exposure



D. Phaseolin promoter – arcelin

22 hr exposure



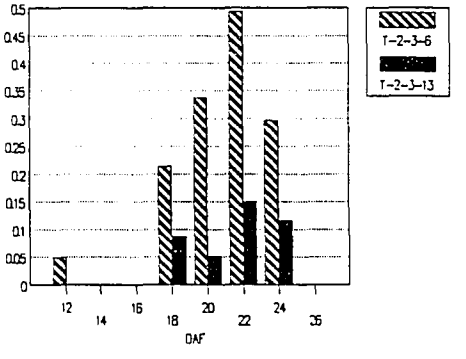


Figure 25. Developmental expression of the arcelin and phaseolin mRNA in the 2-gene transformed plants. A & B, pPA3 transformants; C & D, pT-2-3 transformants; A & D, probed with the arcelin probe; and B & C, with the phaseolin probe.

transformed with only the wild-type arcelin gene (pA3), arcelin mRNA was first detectable at 14 DAF and peaked at 22 to 24 DAF (Figure 24A). The timing for phaseolin mRNA accumulation began earlier, starting at 12 to 14 DAF and peaking at 20 to 22 DAF (Figure 24B). Unfortunately, not enough seeds were available to analyze the pAP4.0-6 transformed plants, but the limited data showed that the starting time of mRNA accumulation was more similar to that of the arcelin promoter driven expression (Figure 24C). Similarly, pPA2.5-3 transformed plants expressed arcelin mRNA with a timing more similar to the pattern driven by the phaseolin promoter (Figure 24D). For transgenic plants containing both wild-type genes (pPA3, Figure 25A and B), as well as the chimeric constructs (pT-2-3, Figure 25C and D), the timing of the mRNA accumulation also appeared to follow the expression pattern of the given promoter.

Surprisingly, the levels of the arcelin mRNA expression driven by either promoter were always lower than those of the phaseolin mRNA, as longer exposure periods were required for the arcelin mRNA to have a similar signal intensity on the autoradiograms (105 or 22 hours for the arcelin probe vs. 12 hours for the phaseolin probe). For plants with two wild-type genes, according to the two plants examined (PA3-6 and PA3-7, Figure 25A and B), the onset of arcelin mRNA accumulation did not cause an obvious reduction in phaseolin mRNA level. For plants with

the two chimeric genes (T-2-3-6 and T-2-3-13), the initiation of phaseolin mRNA expression under the control of the arcelin promoter did not affect the expression of arcelin mRNA driven by the phaseolin promoter (Figure 25C and D).

If the promoter sequences of the arcelin-1 and the β-phaseolin genes were responsible for the observed reduction in the levels of the phaseolin mRNA and protein in SARC1-7, and assuming that the tobacco seeds contain the common transcriptional components for expressing foreign seed protein genes, there would have been a lower level of phaseolin mRNA and a higher level of arcelin mRNA in the pPA3 transformed plants. Likewise, in the pT-2-3 transformants, the amounts of the phaseolin and arcelin mRNA would have been reversed. However, this was not the case. It appeared that the two promoters did not affect each other's expression in the tobacco seeds since no difference in the patterns of mRNA accumulation was observed in either single- or two-gene transformed plants.

If post-transcriptional control, namely mRNA stability, was the major mechanism involved, and assuming that the tobacco seeds provide an RNA turnover environment similar to that of beans, phaseolin mRNA level would have been lower in either type of transformants. This was not the case here either.

The amounts of arcelin mRNA were lower in the transgenic seeds, including the trasngenes driven by the phaseolin promoter. This suggests that the arcelin mRNA might not be stable in the tobacco seeds. However, the stability of arcelin and phaseolin mRNAs would have to be entirely opposite in transgenic tobacco versus in common beans, implying that their mRNA degradation mechanisms were quite different. The post-transcriptional light regulation of the mRNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase was shown to be different between soybean and petunia (Thompson and Meagher, 1990). In these 2 plants, light had different effects on the transcription rates but the rates did not correspond to the levels of mRNAs. The degradation products, examined using polyacrylamide RNA gel blots, primer extension, and S1 nuclease analyses, were similar in soybean and transgenic petunia (Thompson et al, 1992).

It is also possible that PHVARC1-1 is not the major expressed member of the gene family. Thus, it may not affect the expression of the phaseolin gene. The fact that the arcelin gene was expressed in the tobacco seeds at a very low (mRNA) level supports this possibility. The first 900 bp of the promoter sequences, however, are highly homologous between the two arcelin-1 genes sequenced so far (Phvarc1a, Anthony et al, 1991, and PHVARC1-1, this research). Thus, other possible important element(s) might

be located in the region further upstream. Since arcelin gene family members share a high homology (99.5% identity), it will be extremely difficult to distinguish the transcripts between different members and to identify the major gene(s).

Tobacco plants have a different genetic background than that of common beans. In the common beans, the arcelin genes are tightly linked to the lectin genes (PHAs) and are probably organized in a tandem repeat manner (Osborn et al, 1986). The enhancer elements in the neighboring genes, upstream or downstream of the arcelin gene, might also affect its expression. In the transgenic tobaccos, a single arcelin gene was linked to a phaseolin gene. We did not come across any reports about native tobacco lectins, lectin genes, or their genome organization.

Another possibility is that additional factor(s) are involved. Although several backcrosses to Sanilac were performed to produce SARC1-7, it is still merely a "near-isogenic" line. There may be other tightly linked genetic elements, such as trans-acting regulatory protein genes, introduced from the wild beans together with the arcelin genes into SARC1-7 that are not present in the tobaccos. Gel shift assays of nuclear extracts from developing seeds of Sanilac, SARC1-7, and tobacco might reveal the presence of these factors.

CHAPTER IV. CONCLUSION

This research represents an attempt to understand the control of the inverse correlation between the levels of phaseolin and arcelin gene expression in common beans. Seven arcelin genomic clones were isolated from the common bean line SARC1-7 and one of them (λPv222) was sequenced and named PHVARC1-1. This gene was used in the tobacco transformation study. After comparing the pattern of phaseolin and arcelin gene expression in the common bean and in the transgenic tobacco plants, I reach the following conclusions:

1. In common bean, Northern analysis shows that the developmental expression patterns of the phaseolin and arcelin genes at the mRNA level reflect the patterns at the protein level. The onset of arcelin mRNA expression coincides with the reduction of phaseolin mRNA accumulation.

2. Nuclear run-on assays indicate that transcriptional control is one of the regulatory mechanisms for these expression patterns. However, control of RNA stability may still play a role.

3. Phaseolin and arcelin genes were expressed in the transgenic tobacco seeds, either under the control of their own or the other's promoter.

4. The relative timing of phaseolin and arcelin mRNA accumulation is similar in the developing tobacco and the common bean seeds. However, the levels of arcelin mRNA controlled either by the arcelin (PHVARC1-1) promoter or by the phaseolin promoter are lower than the levels of phaseolin mRNA. Furthermore, the initiation of arcelin mRNA accumulation does not affect the expression of phaseolin mRNA in the tobacco seeds. Either the mRNA degradation systems are different in the common bean and the tobacco plants or, more likely, additional genetic elements are involved.

5. Studies which are required to elucidate further the negative effect of the arcelin gene on the accumulation of phaseolin in the common beans could include:

1). Degradation patterns of phaseolin and arcelin mRNA in the common bean and the transgenic tobacco plants, using techniques such as those reported by Thompson et al (1992).

2). Identification and characterization of the cis- and trans-acting factors involved in the arcelin gene regulation, using gel shift assays and transgenic expression studies. The promoter regions of PHVARC1-1 and other arcelin genes probably need to be analyzed.

3). Although lectin and arcelin genes have much in common, lectin genes do not affect phaseolin gene expression. Hybrid gene constructs and promoter deletion studies using the arcelin and lectin genes may reveal the

important sequence domain(s) for the expression of the arcelin genes.

4). Development of a bean (transient) expression system to analyze and compare the strength of the arcelin-1 and β-phaseolin promoter using reporter genes.

REFERENCES

Anthony, J.L., Vonder Haar, R.A., & Hall, T.C. (1991). Nucleotide sequence of a genomic clone encoding arcelin, a lectin-like seed protein from *Phaseolus vulgaris*. Plant Physiol., 97(2):839-840.

Allen, R.D., Bernier, F., Lessard, P.A., & Beachy, R.N. (1989). Nuclear factors interact with a soybean β-conglycinin enhancer. Plant Cell, 1:623-631.

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

Atwater, J.A., Wisdom, R., & Verma, I.M. (1990). Regulated mRNA stability. Ann. Rev. Genet., 24:519-541.

Bassüner, R., Huth, A., Manteuffel, R., & Rapoport, T.A. (1983). Secretion of plant storage globulin polypeptides by *Xenopus laevis* oocytes. Eur. J. Biochem., 133:321-326.

Bednarek, S. & Raikhel, N.V. (1991). The barley lectin carboxyl-terminal properties is a vacuolar protein sorting determinant in plants. Plant Cell, 3:1195-1206.

Bliss, F.A. (1990) Genetic alteration of legume seed proteins. Hortscience, 25(12):1517-1520.

Bollini, R., Ceriotti, A., Daminati, M.G., & Vitale, A. (1985). Glycosylation is not needed for the intracellular transport of phytohemagglutinin in developing *Phaseolus vulgaris* cotyledons and for the maintenance of its biological activities. Physiol. Plant, 65:15-22.

Bollini, R. and Chrispeel, M.J. (1979). The rough endoplasmic reticulum is the site of reserved-protein synthesis in developing *Phaseolus vulgaris* cotyledons. Planta, 146:487-501.

Bollini, R. and Chrispeel, M.J. (1978). Characterization and subcellular localization of vicilin and phytohemagglutinin, the two major reserved proteins of *Phaseolus vulgaris* L. Planta, 142:291-298.

Bollini, R., Vitale, A., & Chrispeels, M.J. (1983). *In vivo* and *in vitro* processing of seed reserve proteins in the endoplasmic reticulum: Evidence for two glycosylation steps. J. Cell Biol., 96:999-1007.

Broadway, R.M. & Duffey, S.S. (1988). The effect of plant protein quality on insect digestive physiology and the toxicity of plant proteinase inhibitors. J. Insect Physiol., 34(12):1111-1117.

Brown, J.W.S., Bliss, F.A., & Hall, T.C. (1981). Linkage relationships between genes controlling seed proteins in French beans. Theor. Appl. Genet., 60:251-259.

Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J., & Hall, T.C. (1991). Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J., 10(6):1469-1479.

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

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

Bustos, M.M., Luckow, V.A., Griffing, L.R., Summers, M.D., & Hall, T.C. (1988). Expression, glycosylation and secretion of phaseolin in a baculovirus system. Plant Mol. Biol., 10:475-488.

Cardona, C., Kornegay, J., Posso, C.E., Morales, F., & Ramirez, H. (1990). Comparative value of four arcelin variants in the development of dry bean lines resistant to the Mexican bean weevil. Entomol. Ecp. Appl., 56:197-206.

Ceriotti, A., Pedrazzini, E., Fabbrini, M.S., Zoppe, M., Bollini, R., & Vitae, A. (1991). Expression of the wild-type and mutated vacuolar storage protein phaseolin in *Xenopus* oocytes reveals relationships between assembly and intracellular transport. Eur. J. Biochem., 202:959-968.

Chappell, J. & Chrispeels, M.J. (1986). Transcriptional and Posttranscriptional control of phaseolin and phytohemagglutinin gene expression in developing

cotyledons of *Phaseolus vulgaris*. Plant Physiol., 81:50-54.

Chee, P.P., Jones, J.M., & Slightom, J.L. (1991). Expression of bean storage protein minigene in tobacco seeds: intron splicing are not required for seed specific expression. J. Plant Physiol., 137:402-408.

Chee, P.P., Klassy, R.C., & 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., & Beachy, R.N. (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA, 83:8560-8564.

Chrispeels, M.J. (1984). Biosynthesis, processing and transport of storage protein and lectin in cotyledons of developing legume seed. Phil. Trans. R. Soc. Lond. B 304:309-322.

Chrispeels, M.J. (1983). The Golgi apparatus mediates the transport of phytohemagglutinin to the protein bodies in bean cotyledons. Planta, 158:140-151.

Cramer, J.H., Lea, K., & Slightom, J.L. (1985). Expression of phaseolin cDNA genes in yeast under control of natural plant DNA sequences. Proc. Natl. Acad. Sci. USA, 82:334-338.

Cramer, J.H., Lea, K., Drong, R.F., Klassy, R.C., Slightom, J.L., Schaber, M.D., & Kramer, R. (1986). Expression of the bean seed storage protein phaseolin in *Saccharomyces cerevisiae*. In *Molecular Biology of Seed Storage Proteins and Lectins*, eds., Shannon, L.M. & Chrispeels, M.J., Proceedings of the Ninth Annual Symposium in Plant Physiology, pp.183-191.

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

Davis, J.W. & Kaesberg, P.K. (1973). Translation of virus mRNA: synthesis of bacteriophage QB proteins in a cell- free extract from wheat embryo. J. Virology, 12:1343-1441.

Dickinson, C.D., Evan, R.P., and Nielsen, N.C. (1988). RY repeats are conserved in the 5'-flanking regions of

legume seed-protein genes. Nucleic Acids Res., 16(1):371.

Dobie, P., Dendy, J., Sherman, C., Padgham, J., Wood, A., & Gatehouse, A.M.R. (1990). New sources of resistance to *Acanthoscelides obtectus* (SAY) and *Zabrotes subfasciatus* Boheman (Coleoptera: Bruchdae) in mature seeds of five species of *Phaseolus*. J. Stored Prod. Res., 26(4):177-186.

Dorel, C., Voelker, T.A., Herman, E.M., & Chrispeels, M.J. (1989). Transport of proteins to the plant vacuole is not by bulk flow through the secretory system, and required positive sorting information. J. Cell Biol., 108:327-337.

Duffey, S.S. & Felton, G.W. (1989). Plant enzymes in resistance to insects. In *Biocatalysis in Agricultural Biotechnology*, Whitaker, J.R. and Sonnet, P.E., eds., ACS Symposium Series 389, Chapt. 20, pp.289-313.

Evans, R.J., Pusztai, A., Watt, W.B., & Bauer, D.H. (1973). Isolation and properties of protein fractions from Navy beans (*Phaseolus vulgaris*) which inhibit growth of rats. Bioc. Biop. Acta, 303:175-184.

Feinberg, A.P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Biochem., 132:6-13.

Gallie, D.R., Feder, J.N., Schimke, R.T., & Walbot, V. (1991). Functional analysis of the tobacco mosaic virus tTNA-like structure in cytoplasmic gene regulation. Nucleic Acid R., 19(18): 5031-5036.

Gatehouse, A.M.R. & Boulter, D. (1983). Assessment of the antimetabolic effects of trypsin inhibitors from Cowpea (*Vigna unguiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus*. J. Sci. Food Agric., 34:345-350.

Gatehouse, A.M.R., Dewey, F.M., Dove, J., Fenton, K.A., & Pusztai, A. (1984). Effect of seed lectins from *Phaseolus vulgaris* on the development of larvae of *Callosobruchus maculatus*; mechanism of toxicity. J. Sci. Food Agric., 35:373-380.

Gatehouse, J.A., Evans, I.M., Croy, R.R., & Boulter, D. (1986). Differential expression of genes during legume seed development. Phil. Trans. R. Soc. Lond., B314:367-384.

Gentry, H.S. (1969). Origin of the common bean, *Phaseolus vulgaris*. Econ. Bot., 23:55-69.

Goldberg, R.B. (1986). Regulation of plant gene expression. Phil. Trans. R. Soc. Lond., B314:343-353.

Goring, D.R., Thomson, L., & Rothstein, S.J. (1990). Transformation of a partial nopaline synthase gene into tobacco suppresses the expression of a resident wild-type gene. Proc. Natl. Acad. Sci. U.S.A., 88(5):1770-1774.

Greenwood, J.S. & Chrispeels, M.J. (1985a). Immunocytochemical localization of phaseolin and phytohemagglutinin in the endoplasmic reticulum and Golgi complex of developing bean cotyledons. Planta, 164:295-302.

Greenwood, J.S. & Chrispeels, M.J. (1985b). Correct targeting of the bean storage protein phaseolin in the seeds of transformed tobacco. Plant Physiol., 79:65-71.

Greenwood, J.S., Keller, G.A., & Chrispeels, M.J. (1984). Localization of phytohemagglutinin in the embryonic axis of *Phaseolus vulgaris* with ultra-thin cryosection embedded in plastic after indirect immunolabelling. Planta, 162:548-555.

Guiltinan, M.J., Thomas, J.C., Nessler, C.L., & Thomas, T.L. (1989). Expression of DNA binding proteins in carrot somatic embryos that specifically interact with a *cis* regulatory element of the French bean phaseolin gene. Plant Mol. Biol., 13:605-610.

Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W. Sun, S.M., & Bliss, F.A. (1978). Messenger RNA for G1 protein of French bean seeds: Cell-free translation and product characterization. Proc. Natl. Acad. Sci. USA, 75(7):3196-3200.

Hall, T.C., Reichert, N.A., Sengupta-Gopalan, C., Cramer, J.H., Lea, K., Barker, R.F., Slightom, J.L., Klassy, R., & Kemp, J.D. (1984). Regulation of bean β-phaseolin gene expression in yeast and tobacco seed. In *Molecular Form and Function of the Plant Genome*, eds., van Vloten-Doting, L., Groot, G.S.P., & Hall, T.C., pp. 517-529.

Hall, T.C., Slightom, J.L., Ersland, D.R., Murray, M.G., Hoffman, L.M. Adang, M.J., Brown, J.W.S., Ma, Y., Mattews, J.A., Cramer, J.H., Barker, R.F., Sitton,

D.W., & Kemp, J.D. (1983). Phaseolin: nucleotide sequence explains molecular weight and charge heterogeneity of a small multigene family and also assists vector construction for gene expression in alien tissue. In *Structure and Function of Plant Genomes*, eds., Ciferri, O. & Dure III, L., Plenum Press, New York & London, NATO ASI Series A: Life Sciences, vol. 63 pp.123-142.

Hall, T.C., Sun, S.M., Buchbinder, B.U., Pyne, J.W., Bliss, F.A., & Kemp, J.D. (1980). Bean seed globulin mRNA: Translation, characterization. and its use as a probe towards genetic engineering of crop plants. In *Genome Organization and Expression in Plants*, Plenum Press, ed. Leaver, C.J., New York and London, pp.259-272.

Hartweck, L.M., Vogelzang, R.D., & Osborn, T.C. (1991). Characterization and comparison of arcelin seed protein variants from common bean. Plant Physiol., 97:204-211.

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

Helms, C., Dutchik, J.E., & Olson, M.V. (1987). A λDNA protocol based on purification of phage on DEAE-cellulose. Meth. Enz., 153:69-82.

Herrick, D., Parker, R., & Jocabson, A. (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. Mol. Cell Biol., 10(5):2269-2284.

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

Hilder, V.A., Gatehouse, A.M.R., & Boulter, D. (1989). Potential for exploiting plant genes to genetically engineered insect resistance, exemplified by the cowpea trypsin inhibitor gene. *Pestic. Sci.*, 27:165-171.

Hoffman, L.M., Donaldson, D.D. (1985). Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J., 4(4):883-889.

Hoffman, L.M., Donaldson, D.D., & Herman, E.M. (1988). A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. Plant Mol. Biol., 11:717-729.

Hoffman, L.M., Ma, Y., & Barker, R.F. (1982). Molecular cloning of *Phaseolus vulgaris* lectin mRNA and use of cDNA as a probe to estimate lectin transcript levels in various tissues. Nucleic Acids Res., 10(23):7819-7828.

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

Ingelbrecht, I.L., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C., & Depicker, A.G. (1989). Different 3' end regions strongly influence the level of gene expression in plant cells. Plant Cell, 1:671-680.

Jackson, R.J. & Standart, N. (1990). Do the poly(A) tail and 3' untranslated region control mRNA translation? Cell, 62:15-24.

Janzen, D.H. & Juster, H.B. (1976). Insecticidal action of the phytohemagglutinin in black bean on a bruchid beetle. Science, 192: 795-796.

Jefferson, R.A., Kavanaugh, T.A., & Bevan, M.W. (1987) GUS fusion: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plant. EMBO J., 6:3901.

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

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

Jofuku, K.D., Schipper, R.D., & Goldberg, R.B. (1989). A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. Plant Cell, 1(4):427-435.

John, M.E. & Long, C.M. (1990). Sequence analysis of arcelin 2, a lectin-like plant protein. Gene, 86:171-176.

Joshi, C.P. (1987). An inspection of the domain between putative TATA box and transcription start sites in 79 plant genes. Nucl. Acid Res., 15:6643-6653.

King, T.P., Pusztai, A., & Clarke, E.M.W. (1980). Kidney bean (*Phaseolus vulgaris*) lectin-induced lesions in rat small intestine: 3. Ultrastructure studies. J. Comp. Path., 92:357-373.

King, T.P., Pusztai, A., & Clarke, E.M.W. (1980). Kidney bean (*Phaseolus vulgaris*) lectin-induced lesions in rat small intestine: 1. Light microscope studies. J. Comp. Path., 90:585-595.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA, 82:488-492.

Kunkel, T.A., Roberts, J.D. & Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. Meth. Enz., 154:367-382.

Leavitt, R.D., Felsted, R.L., & Bachur, N.R. (1977). Biological and biochemical properties of *Phaseolus vulgaris isolectins*. J. Biol. Chem., 252(9):2961-2966.

Lefebvre, D.D. (1990). Expression of mammalian metalothionein suppresses glucosinolate synthesis. Plant Physiol., 93(2):522-524.

Lelievre, J.-M., Oliveria, L.O., & Nielsen, N.C. (1992). 5'-CATGCAT-3' elements modulate the expression of glycinin genes. Plant Physiol., 98:387-391.

Liener, I.E. (1976a). Phytohemagglutinin (phytolectins). Ann. Rev. Plant Physiol., 27:291-319.

Liener, I.E. (1976b). Phytohemagglutinin: Their nutritional significance. J. Agr. Food Chem., 22(1):17-22.

Lawrence, M.C., Suzuki, E., Varghese, J.N., Davis, P.C., Van Donkelaar, A., Tulloch, P.A., & Colman, P.M. (1990). The three-dimensional structure of the seed storage protein phaseolin at 3 A resolution. EMBO J., 9(1):9-15.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. (1951). Protein measurement with the folin phenol reagent. J. Biol. Chem., 193:265.

Luthe, D.S. & Quatrano, R.S. (1980). Transcription in isolated wheat nuclei. Plant Physiol., 65:305-308.

Manen, J.F. & Pusztai, A. (1982). Immunocytochemical localization of lectins in cells of *Phaseolus vulgaris* L. seeds. Planta, 155:328-334.

Matthews, J.A., Brown, J.W.S., & Hall, T.C. (1978). Phaseolin mRNA is translated to yield glycosylated polypeptides in *Xenopus* oocytes. Nature, 294:175176.

Matzke, M.A. & Matzke, A.J.M. (1990). Gene interactions and epigenetic variation in transgenic plants. Developmental Genetics, 11:214-223.

Matzke, M.A. & Matzke, A.J.M. (1991). Differential inactivation and methylation of a transgene in plants by two suppressor loci containing homologous sequences. Plant Mol. Biol., 12:821-830.

Mead, D.A., Szczesna-Skorupa, E. & Kemper, B. (1986). Protein Engineering, 1:67-74.

Meakin, P.J. & Gatehouse, J.A. (1991). Interaction of seed nuclear proteins with transcriptionally-enhancing regions of the pea (*Pisum sativam* L.) *leg*A gene promoter. Planta, 183:471-477.

Miller, J.B., Hsu, R., Heinrikson, R., & Yachnin, S. (1975). Extensive homology between the subunits of the phytohemagglutinin mitogenic proteins derived from *Phaseolus vulgaris*. Proc. Natl. Acad. Sci. USA, 72(4):1388-1391.

Minney, B.H.P., Gatehouse, A.M.R., Dobie, P., Dendy, J., Cardona, C., & Gatehouse, H.A. (1990). Biochemical bases of seed resistance to Zabrotes subfasciatus (bean weevil) in Phaseolus vulgaris (common bean); a mechanism for arcelin toxicity. J. Insect Physiol., 36(10):757-767.

Moreno, J. & Chrispeels, M.J. (1989). A lectin gene encodes the α-amylase inhibitor of the common bean. Proc. Natl. Acad. Sci., USA, 86:7885-7889.

Murai, N., Sutton, D.W., Murray, M.G., Slightom, J.L., Merlo, D.J., Reichert, N.A., Sengupta-Gopalan, C., Stock, C.A., Barker, R.F., Kemp, J.D., & Hall, T.C.

(1983). Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors. Science, 222:476-482.

Murdock, L.L., Huesing, J.E., Nielsen, S.S., Pratt, R.C., & Shade, R.E. (1990). Biological effects of plant lectins on the cowpea weevil. Phytochemistry, 29(1):85-89.

Murray, M.G. & Kennard, W.C. (1984). Altered chromatin conformation of the higher plant gene phaseolin. Biochemistry, 23:4225-4232.

Murray, E.E., Rocheleau, T., Eberle, M., Stock, C., Sekar, V, & Adang, M. (1991). Analysis of unstable RNA transcripts of insecticidal crystal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. Plant Mol. Biol., 16:1035-1050.

Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell, 2(4):279-289.

Nath, K. & Azzolina, B.A. (1981). Cleavage-properties of site-specific restriction endonuclease. In "Gene Amplification and Analysis". ed. Chirikjian, J.G., Vol. 1, pp. 113-130., Elsevier North Holland, Inc., New York, New York.

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

Osborn, T.C. (1988). Genetic control of bean seed protein. CRC Critical Reviews in Plant Sciences, 7(2):93-116.

Osborn T.C., Alexander, D.C., Sun, S.S.M., Cardona, C., & Bliss, F.A. (1988a). Insecticidal activity and lectin homology of arcelin seed protein. Science, 240:207-210.

Osborn, T.C., Ausloos, K.A., Brown, J.W.S., & Bliss, F.A. (1983). Bean lectins. III. Evidence for greater complexity in the structural model of *Phaseolus vulgaris* lectin. Plant Sci. Letters, 31:193-203.

Osborn, T.C., Blake, T., Gepts, P., & Bliss, F.A. (1986). Bean arcelin: 2. Genetic variation, inheritance and

linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. Theor. Appl. Genet., 71:847-855.

Osborn, T.C., Brown, J.W.S., & Bliss, F.A. (1985). Bean lectins: 5. Quantitative genetic variation in seed lectins of *Phaseolus vulgaris* L. and its relationship to lectin variation. Theor. Appl. Genet., 70:22-31.

Osborn, T.C., Burow, M., & Bliss, F.A. (1988b). Purification and characterization of arcelin seed protein from common bean. Plant Physiol., 86:399-405.

Pusztai, A. & Stewart, J.C. (1978). Isolectins of Phaseolus vulgaris. Physicochemical studies. Bioc. Biop. Acta, 536:38-49.

Reed, K.C. and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acid R., 13(20):7027-7221.

Riggs, C.D. & Chrispeels, M.J. (1988). Phytohemagglutinin gene expression is correlated with tissue-specific DNA methylation patterns in *Phaseolus vulgaris*. Genome, 30 (suppl. 1):95 (abstr 31.23.4).

Riggs, C.D., Hunt, D.C., Lin, J., & Chrispeels, M.J. (1989). Utilization of luciferase fusion genes to monitor differential regulation of phytohemagglutinin and phaseolin promoters in transgenic tobacco. Plant Science, 63:47-57.

Riggs, C.D., Voelker, T.A., & Chrispeels, M.J. (1989). Cotyledon nuclear proteins bind to DNA fragments harboring regulatory elements of phytohemagglutinin genes. Plant Cell, 1:609-621.

Rhodes, P.R. & Vodkin, L.O. (1985). Highly structured sequence homology between an insertion element and the gene in which it resides. Proc. Natl. Acad. Sci. USA, 82:493-497.

Romero Andreas, J., Yandell, B.S., & Bliss, F.A. (1986). Bean arcelin: 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theore. Appl. Genet., 72:123-128.

Saalbach, G., Jung, R., Kunze, G., Saalbach, I., Adler, K., & Muntz, K. (1991). Different legumin protein domains act as vacuolar targeting signals. Plant Cell, 3:695-708.