ISOLATION, CHARACTERIZATION, AND DEVELOPMENTAL REGULATION OF A 2S SEED STORAGE PROTEIN GENE IN *Pseudotsuga menziesii* (DOUGLAS FIR)

by

David Machander
B.Sc., University of Victoria, 1991

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

We accept this thesis as conforming to the required standard

[redacted]

Dr. S. Misra, Supervisor (Department of Biochemistry and Microbiology)

[redacted]

Dr. T.W. Pearson, Departmental Member (Department of Biochemisty and Microbiology)

[redacted]

Dr. N. Sherwood, Outside Member (Department of Biology)

[redacted]

Dr. Abul Ekramoddoullah, External Examiner (Department of Biology)

Supervisor: Dr. Santosh Misra

ABSTRACT

Studies on the developmental regulation and characterization of genes from Douglas fir were undertaken. A clone (p900) was isolated from a cDNA library prepared from poly A^+ mRNA isolated from the mid-stage of Douglas fir embryogenesis. Northern hybridization studies suggested the cDNA to be a full length clone. The gene is clearly developmentally regulated at the transcriptional level in both the megagametophyte and slightly differently (temporally) in the zygotic embryo. Sequencing data confirmed the clone to be a 2S albumin, yet a unique isoform. Nucleotide sequence alignments showed an overall similarity to *Picea glauca* albumin and *Pinus strobus* albumins 1, 2, and 3 of 48.8%. An open reading frame of 140 amino acids encoding a 15.8 kDa preprotein precursor was seen from the predicted amino acid sequence data. A signal peptide of 21 aa was present as predicted by signal peptide cleavage site rules and confirmed by hydropathy plotting. The protein is predicted to have a lack of N-linked glycosylation and is a basic protein. A 30.5% overall amino acid similarity is seen between p900 and other gymnosperm albumins. The protein is high in arginine and glutamine/glutamate at 15.7% and 12.2%, respectively, indicative of the role of albumin as a nitrogen source during seed germination. Additionally, the p900 albumin has a high cysteine content (2.1%) and shows a strict conservation in the positioning of these cysteines. Unique features of this isoform include the deletion of two regions where cysteines are typically conserved.

Dr. S. Misra, Supervisor (Department of Biochemistry and Microbiology)

Dr. T.W. Pearson, Departmental Member (Department of Biochemistry and Microbiology)

Dr. N. Sherwood, Outside Member (Department of Biology)

Dr. Abul Ekramoddoullah, External Examiner (Department of Biology)

Table of Contents

page

Materials and Methods

page

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

Abbreviation	
A	absorbance
amp	ampicillin
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
bases	A: adenine
	C: cytosine
	G: guanine
	T: thymine
°C	degrees celsius (centigrade)
bp	base pairs
cm	centimeter
cpm	counts per minute
Cys, C	cysteine
DAF	days after fertilization
DAS	days after stratification
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
h, hrs.	hours
His, H	histidine
Ile, I	isoleucine
IPTG	isopropylthiogalactoside
LB	Luria-Bertani broth
Lys, K	lysine
Leu, L	leucine
kb	kilobase
kDa	kilodalton
min, '	minutes
Met, M	methionine
mRNA	messenger ribonucleic acid

mmol	millimolar
ml	milliliter
M	mole
MW	molecular weight
NZY	NZ amine (casein hydrolysate) yeast broth
OD	optical density
O/N	over night
Phe, F	phenylalanine
pfu	plaque forming units
Pro, P	proline
rpm	revolutions per minute
RT	room temperature
sec	seconds
S	svedberg sedimentation coefficient
Ser	serine
SDS	sodium dodecyl sulfate
SM	phage dilution buffer
TAE	tris-acetate EDTA buffer
TBE	tris-borate EDTA buffer
tet	tetracycline
Thr, T	threonine
Tris	tris(hydroxymethyl)aminomethane
Trp, W	tryptophan
Tyr, Y	tyrosine
UV	ultraviolet light
vol	volume
X-gal	5-dibromo 4-chloro 3-indolylgalactoside
2XYT	2X yeast tryptone broth

Acknowledgements

I am grateful to Dr. T.W. Pearson and Dr. N. Sherwood for their time. I would like to thank Michelle, Damone, Lou, and my parents for their encouragement and support. A big thanks to my room-mate Megan for her assistance with the computer. I especially would like to thank Dr. Ivan Stefanov and Tim Tranbarger for all their advice and technical assistance as well as the other members in the lab, Ben, Karia, and Malinee for all their help and friendship. Thank-you Dr. S. Misra for your support and providing the opportunity for research in your lab.

Introduction

Plants represent a diverse group of organisms that have unique reproductive, developmental, and physiological processes. Although morphologically simple, plants have molecular genetic processes that are equivalent in complexity to those found in animals. During the last 15 years there has been a major refocusing on plants as a biological system (Goldberg, 1988). The renewed interest in plants arose partly from the realization that gene transfer technology could be used to introduce novel genetic traits into plants.

Current excitement about plants as experimental organisms also derives from the perception that plants represent a new biological frontier, ready for experimentation on novel aspects of plant biology such as photosynthesis, seed development, reproduction, nitrogen fixation, fruit ripening, pollination, and light control. Although these phenomena have been studied extensively for many years, new technology, advances in molecular biology, and the development of accessible plant systems now make it possible to dissect the molecular and cellular events responsible for controlling plant-specific processes, specifically the molecular processes controlling gene expression during plant development (Goldberg, 1988).

Embryogenesis in plants appears superficially to be a relatively simple process in terms of morphological

development. This apparent simplicity allows for the exploitation of embryogenesis as a model experimental system to study the relationship between gene expression and morphogenesis. However, the large number of genes expressed (estimated to be at least 20 000 (Kamalay *et al.*, unpublished) in developing plant embryos, suggests a complexity at the molecular level that is not apparent from an anatomical perspective (Goldberg *et al.*, 1989), hence, we are only just beginning to describe changes in gene expression during embryogenesis and to identify regulatory networks. (Lindsey *et al.*, 1993)

Embryo gene expression programs occur during a developmental period that is critical for establishing the polarity and morphological pattern of the plant in which major plant tissue and cell types form (Raghaven *et al.*, 1986). In general, the overall principles governing plant gene regulation and the structure and organization of plant gene sets are similar to those studied in animals such as mouse and man (Goldberg *et al.*, 1988). Experiments on genes expressed during embryogenesis have been directed towards dissecting the regulatory circuits that control seed protein gene expression (Higgins *et al.*, 1984), and identifying genes responsible for controlling pattern formation and morphogenesis during early seed development (Meinke *et al.*, 1986).

I. Seed Development

In higher plants, the seed is the organ of dispersal and is normally produced by sexual reproduction. The end product of seed development is a mature seed, capable of surviving long periods under adverse conditions while still being able to support the growth of a seedling upon germination. In the last 10 years, with the burst of information generated on gene expression processes in higher plants, and with the ability to isolate, manipulate and re-introduce genes into other plant species, there is an enhanced interest in the seed as a model system to study plant embryogenesis. (Lindsey *et al.*,1993)

Following fertilization, the three major phases identified with the production of a mature seed are, firstly, cell division within the zygote giving rise to a globular embryo which differentiates into the embryonic axis and cotyledons. Secondly, cell expansion, results in embryo enlargement followed by storage reserve accumulation involving storage proteins. Thirdly, developmental arrest and desiccation prepares the seed for dormancy (Lindsey *et al.*, 1993). Although metabolic processes and *de novo* protein synthesis drop dramatically during this stage, synthesis of late embryogenesis abundant (LEA) proteins, coded by other gene sets, are produced to protect the seed tissues against desiccation.

A central question one might then ask is how are these processes regulated? Although most aspects of development in plants are carried out in the post-embryonic phase of the life cycle, the study of this embryonic pathway can, nevertheless, be expected to address the question of how differentiation and development are co-ordinated. In particular the expression patterns of specific genes and sets of genes during the developmental process can be investigated. Specifically, with respect to embryogenesis, what are the chemical and other environmental signals, which modulate and regulate the expression of these embryonic genes (Lindsey *et al.*, 1993) ?

II. Embryogenesis in Conifers

Because of the paucity of knowledge about the gene sets expressed in early embryogenesis, due to difficulty in obtaining sufficient amounts of plant material, most investigators studying gene expression in the developing seed have concentrated on the proteins expressed during mid-to late-maturation stages (Lindsey *et al.*, 1993). Additionally, the majority of the work comes from studies with angiosperms (flowering plants), where the resulting data are still a long way from being assembled into a complete picture of the processes that regulate embryogenesis. In conifers, an economically important species, very little is known.

1. Protein Accumulation

Embryogenesis involves a series of highly organized cell divisions to establish a cellular pattern (Lindsey *et al.*, 1993). In the seed, proteins may be considered as comprising three broad groups, according to their function (Kreis *et al.*, 1986): enzymes, such as those involved in the mobilization of food reserves; structural proteins, associated with, for example, membranes and ribosomes; and storage proteins, which may be utilized as food reserves for the germinating seedling or which may play a role in protecting the embryo during desiccation. In most dicot species studied, the endosperm (starchy tissue surrounding the embryo) degenerates during seed development, and the storage proteins accumulate in the embryo. In soybean (Goldberg *et al.*, 1989), Brassica (Blundy *et al.*, 1991), pea (Wang *et al.*, 1991) and bean (Bustos *et al.*, 1991) for example, storage proteins typically accumulate only from mid-maturation phase onwards, when cell division is complete and the basic form of the embryo has developed. The pattern of protein accumulation may also be spatially, as well as developmentally regulated (Corke *et al.*, 1990).

2. Transcript Regulation

More than 90% of the 15 000 diverse mRNAs present in mid-embryogenesis stage are also present in post-germination and in the mature plant leaf. However, a small number of genes have been identified that encode prevalent mRNAs

specific to embryogenesis (Dure *et al.*, 1985), or are present in embryos at a much higher level than at other periods of development (Dure *et al.*, 1981). Of these large numbers of genes expressed during embryogenesis, many may comprise multigene families as described for legumes (Kreis *et al.*, 1986) and cereals (Gatehouse *et al.*, 1986). Figure 1 represents the quantitative changes in gene expression programs occurring during embryogenesis in terms of the changes in abundant mRNA sets, where presumably genes in each set are responding to distinct regulatory signals. Of the collective 100 or so embryo-specific mRNA encoded proteins, generally the functions of only LEA proteins and storage proteins are known (Galau *et al.*, 1981). Diverse patterns of mRNA abundance have been described and classes of transcripts can be recognized as being restricted to specific developmental stages of embryogenesis (Lindsey *et al.*, 1993).

There is now good evidence that embryonic protein synthesis is controlled at the transcriptional level. Transcriptional and post-transcriptional control have been described for legume storage proteins (Galau *et al.*, 1981; Walling *et al.*, 1986), cereal storage proteins (Colot *et al.*, 1987), and *Brassica* storage proteins (Blundy *et al.*, 1991). Stage-specific protein synthesis has been demonstrated to reflect the abundance of encoding transcripts by analysis of cDNA (encoding specific proteins) hybridizations to RNA. Using intensive hybridization studies, Walling *et al.*, (1986) showed

that seed protein genes are transcriptionally regulated in embryogenesis. Post-transcriptional processes such as cytoplasmic entry rates and/or differential mRNA stabilities are important in regulating seed protein gene expression, and each seed protein gene family is regulated independently at both the mRNA and gene levels.

A fundamental question is what determines the temporal and cell type-specific expression of embryonic genes? One possibility is that the developmental history (lineage) of the cells may influence the gene expression pattern. It was suggested that specific cells present in the upper part of the embryo contain factors that may determine the expression pattern of the storage proteins (Fernandez *et al.*, 1991). This may represent a stable form of compartmentalization of gene expression to which specific cell types may be committed and the implication is that the biochemical or gene expression characteristics of particular cells or group of cells within an organ may be determined at a relatively early developmental stage. One possible process defined in animals that could allow cell-specific expression patterns through cell division is the 'imprinting' of relatively stable changes to the structure of chromatin domains containing embryonically-active genes. This hypothesis may account, in part at least, for the clonally inherited patterns of expression of homeobox genes in, for example, *Drosophila* and mouse (Gaunt *et al.*, 1990). One

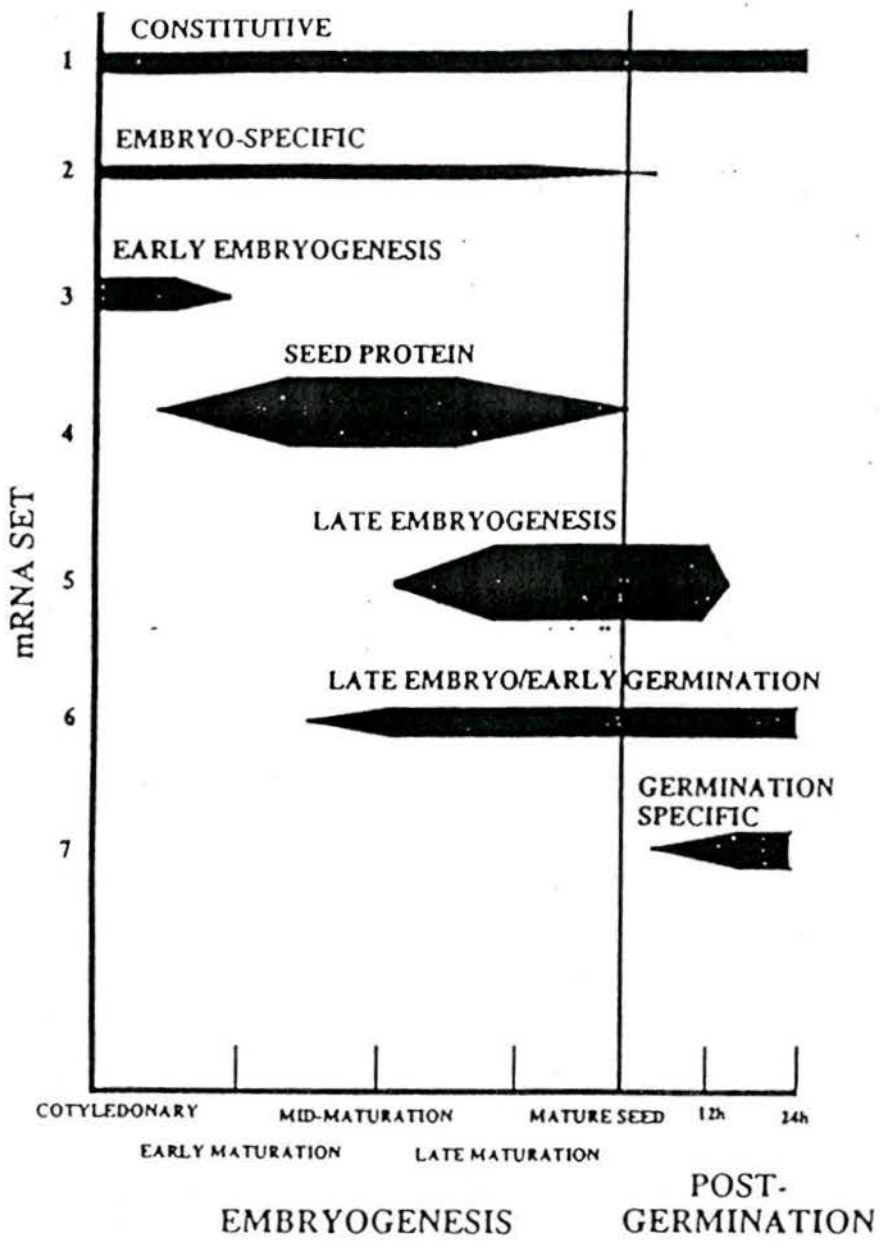


Figure 1. mRNA sets schematically represented by shaded areas indicating the timing of appearance and relative abundance of mRNA populations during seed development (after Goldberg *et al.*, 1989).

possible mechanism for the stable modulation of gene expression in development is DNA methylation. For example, an inverse correlation between maize seed storage protein gene expression and methylation in a range of tissues has been reported (Bianchi *et al.*, 1988). However, in soybean, evidence to the contrary is reported, where seed protein gene activation and repression are not correlated with detectable methylation changes (Walling *et al.*, 1986).

Alternatively, but not necessarily a mutually exclusive view would be that it is the position of the cells within the embryo, and so their spatial relationship with chemical or other signals produced by either neighboring embryonic cells or surrounding tissues such as the endosperm, or other cells of the developing seed, that determines the pattern of gene expression, rather than the lineage or history of the cell. There are arguments in favor of this hypothesis, which has also been discussed widely in animal development: gradients of specific molecules, such as regulatory proteins (St. Johnston *et al.*, 1992) or smaller molecules (Jessel *et al.*, 1992) have been correlated with pattern formation and associated gene expression. Additionally, it may be that the 'patterning' of genes revealed by mutant studies of Mayer *et al.* (1991) may play an analogous role in plants, by establishing gradients. These gradients may be of mRNAs/proteins encoding either transcription factors that may activate cell-specific gene expression patterns. The gradients may be of receptor

molecules sensitive to other regulatory molecules such as growth substances. Recent experimental work on auxin, for example, has demonstrated directly that the level of expression of a soybean auxin-inducible gene is correlated with an endogenous auxin gradient in plant tissue (Li *et al.*, 1991). These growth substances might act to activate specific sets of genes associated with, for example, the control of the correct plane of cell division. Finally, the gradients may be of components of signal transduction pathways involved in downstream processes of differentiation and development and the associated patterns of gene expression. Calcium, for example, is known to be associated, if not with the establishment of polarity, then with the maintenance of polarized growth in the developing *Fucus* embryo (Kropf *et al.*, 1987), and to be an essential component of signal transduction/protein phosphorylation pathways in plants. Furthermore, it is becoming clear that the regulation of the phosphorylation of protein kinase complexes is a critical facet of cell division control (Cyert *et al.*, 1990). The establishment of cellular pattern is a primary determinant in embryogenesis, situated at or near the top of a developmental cascade, but the regulation of spatial and temporal patterns of gene expression in development comprises both the processes of imprinting and inductive signals, superimposed one upon the other, to provide a stability of gene expression through repeated cell divisions

and an ability to respond to changing environmental conditions (Lindsey *et al.*, 1993).

III. LEA Proteins

In the vast majority of seeds, a desiccation period leading to embryo quiescence is naturally interpolated between embryo development and germination. Of the many proteins that accumulate during plant seed maturation, which occurs in the late stage of embryogenesis, some are believed to play a role in seed desiccation and are often referred to as late embryogenesis abundant (LEA) proteins or genes (Hughes *et al.*, 1991).

In many higher plants, only the embryo and other seed tissues such as the megagametophyte (surrounding maternally derived tissue), aquires the ability to withstand dehydration during seed maturation. In angiosperms, Genes have been isolated and characterized that encode transcripts accumulating in these tissues as seeds approach maturity and begin to desiccate. These LEA proteins are required for proper development and differentiation of plant embryos where they play an essential role in seed survival in controlling water exchanges during seed desiccation and imbibition (water uptake). The expression of these LEA genes has been found to peak at late stages of embryogenesis in preparation for the dormant seed stage, but also during periods of extreme water stress (Sundas *et al.*, 1992). Based on their deduced

amphiphilic secondary structure, it has been proposed that LEA proteins play a physical osmoprotectant role, thereby allowing tolerance to desiccation (Dure *et al.*, 1985). Some have been found to be responsible for the accumulation of osmolytes such as the sugar sorbitol, as in the case of barley, where the LEA protein is an aldose reductase (Bartels *et al.*, 1991).

The mechanism whereby the seed environment retains the embryo in a developmental mode and suppresses germination is not known, although two potential regulatory factors have been studied: ABA (abscisic acid) and restricted water uptake. In several higher plants it has been shown that levels of the LEA proteins can be manipulated experimentally by using ABA or osmoticum during imbibition as well as premature drying of seeds. Exogenous ABA levels have been found to correlate with LEA levels (Galau *et al.*, 1986; Hong *et al.*, 1988; Mundy *et al.*, 1988), however, the evidence is controversial, where it has been shown that endogenous ABA levels do not always correlate with LEA expression (Barratt *et al.*, 1989; Finkelstein *et al.*, 1986; Morris *et al.*, 1990). It is postulated, then, that ABA hormonally regulates the levels of desiccation tolerance (Thomann *et al.*, 1992), however, the mechanism of this modulation remains to be uncovered. The LEA genes have shown a high degree of similarity in many angiosperms but have yet to be characterized in conifers, an economically important group of higher plants.

IV. Storage Proteins

Developing embryos of plants synthesize a variety of proteins in a highly regulated fashion. The most abundant of these proteins, the seed storage proteins, are accumulated to high levels and, therefore, are of interest because they provide an excellent system for investigating different aspects of gene regulation in plants and because they code for proteins that are an important source of food for human and animal consumption. (Raghaven *et al.*, 1986)

1. Storage Protein Genes

The storage proteins are produced in large amounts at specific stages in seed development. The majority of the storage proteins are encoded by multi-gene families which differ as to their size, chromosomal location, linkage, and occurrence of non-functional alleles (Dure *et al.*, 1981). Most of the storage protein genes are specifically expressed in only one organ, the developing seed. Thus, they are examples of genes which are under strong developmental and tissue-specific control (Walling *et al.*, 1986).

2. Storage Protein Characteristics

The storage proteins show striking similarity between diverse plant species indicating evolutionary conservation. Other characteristics include the presence of a signal sequence,

are generally glycosylated, and are post-translationally modified. (Higgins *et al.*, 1984)

Storage proteins in seeds may be defined as any protein accumulated in significant quantities in the developing seed which upon germination, is rapidly hydrolyzed to provide a source of reduced nitrogen for the early stages of seedling growth. Therefore, these proteins tend to be rich in asparagine, glutamine, and arginine or proline. (Higgins *et al.*, 1984)

In mature seeds, the storage proteins are found predominantly in small (1 to 20 μm in diameter) organelles called protein bodies, which are surrounded by a single membrane of ER origin (Pernollet *et al.*, 1978; Weber *et al.*, 1980).

Storage proteins which are transported within the cell are usually synthesized as precursors. The precursor contains information specifying the transport pathway. These proteins destined for the endomembrane system are usually synthesized on membrane-bound polysomes and contain a short N-terminal amino acid sequence called the signal. The signal sequence is instrumental in directing the nascent chain to the ER and so to the first step of intracellular transport. The signal sequence is removed during translation before the nascent chain is complete (Kreil *et al.*, 1981).

Newly synthesized storage proteins have been shown to accumulate transiently within the ER prior to their transport to

the Golgi and protein bodies (Baumgartner *et al.*, 1980; Bollini *et al.*, 1982; Bollini *et al.*, 1983; Chrispeels *et al.*, 1982, 1983).

In addition to the removal of the signal peptide, it has been found that newly synthesized phaseolin and vicilin 7S proteins are post-translationally glycosylated (Bollini *et al.*, 1982; Chrispeels *et al.*, 1982). Although there is no doubt that many 7S proteins and lectins from seeds are glycosylated, there is no satisfactory explanation for the physiological function of this step. The 11S globulins (Badenoch-Jones *et al.*,1981; Gatehouse *et al.*, 1980; Hurkman *et al.*, 1980) and most prolamins are generally considered to be non-glycosylated. It might be proposed, that glycosylation could play a role in the stabilization of the protein during desiccation; however, there are numerous cases of pairs of highly homologous proteins, both within and between species, where only one member is glycosylated, such as in the lectins of pea and broad bean, the M_r 50,000 subunits of pea, and the major globulins of cotton (Davey *et al.*, 1979; Dure *et al.*, 1981), yet both members appear equally stable.

Seed storage proteins can be divided into four major solubility classes of albumins, globulins, prolamins, and glutelins, which are soluble in water, salt, aqueous alcohol, and acid or alkali solutions, respectively (Higgins *et al.*, 1984).

The major storage proteins of most legumes are the 7S and 11S globulins known as vicilin and legumin, respectively, and those of most cereals are the prolamins and glutenins.

The salt soluble globulin group is synthesized at defined stages of seed development, stored in protein bodies, is protected from desiccation upon seed maturation and is subsequently hydrolyzed upon germination, providing carbon and nitrogen skeletons for the developing seedling. Storage proteins have sedimentation coefficients of about 7S and 11S, which has led to extensive investigations designed to detect similar proteins in many different species (Danielson, 1949).

3. Homologies and Multigene Families

The storage proteins of evolutionarily diverse groups of plants show striking homology which can be detected at different levels of protein organization. The same oligomeric structure of six subunits, in which each subunit is composed of a disulfide-bonded acidic and basic polypeptide, is found in the 11S globulins of legumes (Debyshire *et al.*, 1976), rapeseed (Crouch *et al.*, 1981), and oats (Walburg *et al.*, 1983). Two prolamins of wheat and barley are homologous to legumins as shown by N-terminal sequence analysis (Shewry *et al.*, 1980), whereas prolamin cDNA sequencing from wheat, barley, and maize show strong homology between the different genera (Bartels *et al.*, 1983; Foard *et al.*, 1981; Geraghty *et al.*,1981; Pederson *et al.*, 1982; Spena *et al.*, 1982). Primary sequence homology between 11S globulins of various legumes, oats, and cucurbits has also been demonstrated by protein sequencing (Casey *et al.*, 1981; Gilroy *et al.*, 1979; Hara *et al.*, 1978).

Interestingly, N-terminal amino acid sequencing recently confirmed that the *Ginkgo biloba* and Douglas fir seeds, which are gymnosperms, contain a legumin-like storage protein with clear homology to those found in the angiosperms (Hager *et al.*, 1992; Misra *et al.*, 1993). N-terminal sequencing has also shown homologies between 7S globulins of pea, bean, and soybean by cDNA sequencing (Lycett *et al.*, 1983; Schuler *et al.*, 1982; Slightom *et al.*, 1983). Among the albumins which also serve as storage proteins, the lectins of pea, French bean, broad bean, lentil, and jack bean are all homologous (Cunningham *et al.*, 1975; Foriers *et al.*, 1981; Higgins *et al.*, 1983; Hoffman *et al.*, 1982; Hopp *et al.*, 1982). A low molecular weight albumin from castor bean shows some homology with a similar protein from lima bean (Sharief *et al.*, 1982).

The most compelling reason for the presence of homologous proteins in such a variety of plant groups may be the need for a fairly tightly constrained protein structure which must fulfill certain criteria. These could include recognition sequences that are important not only in aggregation of subunits and oligomer assembly but also in intracellular transport and deposition in protein bodies. The proteins must be able to withstand changing osmotic conditions during seed drying and imbibition, and they must be resistant to hydrolysis during seed development and be susceptible during seed germination. Once a seed protein has evolved that

meets all these criteria and is still an efficient store of nitrogen, there may be constraints against change in structure.

Small changes are tolerated, however, and it is now clear that the storage proteins, like most other proteins, are coded for by multigene families. Thus, there are probably three to ten genes for each of the prolamin families in maize (zein), and since there may be 10 families, this means 30-100 zein genes in maize (Burr *et al.*, 1982; Hagen *et al.*, 1981; Pederson *et al.*, 1982). Nucleotide sequence homology between zeins of different families is about 65% (Marks *et al.*, 1982), yet there is marked similarity between size and sequence of their protein signal sequences, "head" domain, repetitive block structure, and "tail" piece (Spena *et al.*, 1982). The repetitive block structure, made up of 6 to 9 repetitive blocks of 20 amino acids (Spena *et al.*, 1982), also occurs in the wheat prolamin (Bartels *et al.*, 1983) and may be a common feature of all prolamins.

The globulins of dicotyledonous plants are also the products of multigene families (Brown *et al.*, 1981; Casey *et al.*, 1979.; Dure *et al.*, 1981; Moreira *et al.*, 1979; Schuler *et al.*, 1982; Thompson *et al.*, 1978). In the pea, at least three families of genes coding for the M_r 50,000 subunit of vicilin (7S) have been identified and they exhibit about 80% homology. Within each family, up to six members have been identified, indicating as many as 18 genes for the M_r 50,000 family.

There are at least four homologous genes for the 11S proteins of pea (legumin) (Croy *et al.*, 1982) and soybean (glycinin) (Moreira *et al.*, 1982), and evidence from amino acid sequence data indicates that there are internal regions of homology within the acidic portion of the glycinin subunit (Moreira *et al.*, 1982) which may be equivalent to the repetitive block structure of prolamins.

The fact that storage proteins are probably all derived from multigene families, coupled with the indications showing internal regions of homology, both imply that the storage protein genes evolved by a complicated series of gene duplications. The divergence between the genes of these families is yet to be studied in detail, but a thorough analysis may reveal the role of sequence conservation, if any, in determining the essential features of a storage protein. Already it is clear that primary sequence information alone will not be sufficient; there is also a need for information on the three-dimensional structure of the proteins.

4. mRNA Regulation in Angiosperms

Much of the work attempting to elucidate the molecular basis of developmental regulation of storage proteins has been done on angiosperms, specifically *Brassica* (Crouch *et al.*, 1981) barley (Rahman *et al.*, 1982), maize (Soave *et al.*, 1981), with much work on the 11S and 7S globulin storage protein genes, where a close correlation between the level of mRNA and the

rates of polypeptide synthesis has been found. In conifers, specifically white spruce, developmental regulation has been shown in legumin proteins (Misra *et al.* 1990) which are the major storage protein (Misra *et al.*, 1991), and recently the 11S globulin (legumin-like storage protein) (Misra *et al.*, 1993).

Additional studies with angiosperms have shown that storage protein gene expression is regulated mainly at the transcriptional level (Goldberg *et al.*, 1989; Quatrano *et al.*, 1986). Recent evidence tends to point to a link between changes in ABA levels and osmoticum and deposition of storage reserves in angiosperm species (Skriver *et al.*, 1990).

Transcriptional control of a gene can either be an alteration from a state where a gene is 'off' to where it is 'on'-a switch process or an up- to down-regulation, in which the amount of a specific transcript changes through development. Most of the classical features of storage protein gene expression can be examined with reference to the pea storage protein genes.

Very early events in the embryogenesis of the pea seed (5 days after fertilization or d.a.f.) when the embryo is 0.25-2 mm in diameter and changing from the globular to the heart-shaped stage during the course of embryogenesis, have been examined by *in situ* hybridization and immunochemistry (Harris *et al.*, 1989). In this technique, radiolabelled or photobiotinylated gene fragments are used to localize the synthesis of their corresponding mRNAs in whole embryos or

in tissue sections. Messenger RNA for the 11S pea storage protein legumin was first detected using this technique in both globular and heart-shaped embryos, albeit at a low level. This legumin mRNA was located in nearly all cells throughout the embryo, but was not maintained at a later stage of development, since neither legumin mRNA nor protein could be found in endosperm at the second, cell expansion, stage of embryo development.

At about 9 d.a.f., concurrent with the cell expansion stage of seed development, a dramatic increase takes place in the levels of mRNA for vicilin and legumin (Evans *et al.*, 1984; Gatehouse *et al.*, 1986). Increased synthesis of mRNA coding for the 7S vicilin precedes the increase in legumin mRNA synthesis by about 24 h. This agrees with results obtained from transcripts isolated from nuclei at this stage of seed development (Evans *et al.*, 1984), and with the levels of protein deposited. Vicilin mRNA levels peak at 10 d.a.f., whereas, the legumin levels peak at 18-19 d.a.f., and then start to decline. This correlation between the mRNA levels for both legumin and vicilin with the levels of nuclear transcripts shows that transcriptional control is a major process regulating storage protein deposition in the pea. In addition, the pea storage protein genes appear to be upregulated rather than switched on from an inactive stage, as mRNA for the storage proteins can be detected using sensitive techniques at an early stage in embryogenesis, well before the cell expansion phase, during

which most of the increase in storage protein mRNA accumulation takes place. Similar increases in storage protein mRNAs at specific stages in seed development occur in soybean (Meinke *et al.*, 1981), maize (Bostan *et al.*, 1986), wheat (Bartels and Thompson, 1986), barley (Forde *et al.*, 1985), french bean (Sun *et al.*, 1978), sunflower (Allen *et al.*, 1985) and oats (Chesnut *et al.*, 1989), amongst others.

V. Albumins

The seed storage proteins of most dicotyledonous plants contain two major protein classes, globulins and albumins, which are distinguished on the basis of solubility as described previously (Debyshire *et al.*, 1976). Whereas the globulins are soluble in high-salt buffers and have sedimentation coefficients of 7S and 11S, many plants contain a set of small, water-soluble proteins generally referred to as 2S albumins, 1.7 S albumins, or napin-like proteins (Crouch *et al.*, 1983; Youle *et al.*, 1981). Other distinguishing features that differentiate albumins from the globulins are the lack of glycosylation in albumins (Krebbers *et al.*, 1988) and the increased number and conservation of sulfur carrying cysteine residues (Menendez-Arias *et al.*, 1988).

Initially, albumins were thought to be proteins with only enzymatic properties and metabolic functions (Altshul *et al.*, 1966; Millerd *et al.*, 1975). However, in view of the abundance and ubiquitous occurrence of albumins in various seeds (Basha

et al., 1975; Brohult *et al.*, 1954; Hill *et al.*, 1974), it was concluded that in addition to globulins, glutelins, and prolamines, albumins are important storage proteins in seeds (Youle *et al.*, 1978).

1. Composition of Protein

The 2S proteins have been shown to comprise a large portion of the total protein content in the developing seed. Comparisons of 2S protein content in the seeds of many families have shown that albumin proteins comprised between 20% (as in peanut) to 62% (as in sunflower) of the total seed protein with the globulins comprising the rest (Youle *et al.*, 1981), whereas, in gymnosperms it was found that the 2S proteins accounted for slightly less (15%) of the total protein composition (Allona *et al.*, 1993) (Table 1).

2. Amino Acid Composition

The 2S albumins have been extensively studied in angiosperms and have been found to contain high levels of nitrogen-rich glutamine/glutamate, asparagine/aspartate, and arginine, thereby, providing a major nitrogen supply to the growing seedling after germination. In soybean, for example, of the 295 amino acids, glutamine residues comprise 24% (Allen *et al.*, 1987), in brazil nut 27% (Ampe *et al.*, 1986), in *Brassica* 30% (Monsalve *et al.*, 1990), and in alfalfa 21.6%

Table 1: Amount of 2S, 7S and 11S proteins in the seeds of various species.

Family, species (common name)	% of total seed protein 2S	7S	11S
Compositae *Helianthus annus* (sunflower)	62	0	38
Cruciferae *Brassica* ssp. (mustard)	62	0	38
Linacae *Linum usitatissimum* (linseed)	42	0	58
Rosaceae *Prunus amygdalus* (almond)	25	15	60
Arachis hypogaea (peanut)	20	6	74
Cucurbitacea *Cucumis sativus* (cucumber)	56	17	27
Lecythidaceae *Bertholletia excelsa* (brazil nut)	30	9	61
Betulaceae *Corylus* spp. (hazelnut)	28	12	60
Liliaceae *Yucca* spp. (yucca)	27	16	57
Euphorbiaceae *Ricinus communis* (castor bean)	44	14	42
Pinus pinaster	15	(....85....)	

Table 2: Select amino acid composition (%) of seed 2S albumin from various species.

	Arg	Asx[a]	Glx[b]	Cys
Angiosperms				
sunflower	5.15	7.42	18.36	6.59
mustard	4.30	4.61	15.89	9.01
linseed	6.03	6.36	23.75	8.17
cucumber	11.90	5.30	20.02	8.92
Brazil nut	11.77	4.27	24.53	13.11
Hazelnut	11.82	6.59	29.69	10.27
Yucca	11.46	7.34	19.34	9.29
Castor	9.40	4.40	30.00	8.50
Cotton	10.32	8.37	27.23	7.74
Gymnosperms				
Pinus pinaster	25.46	4.28	25.03	5.99
Picea glauca	13.66	6.21	19.25	4.97
Pinus strobus (albumin 1)	14.74	4.74	16.32	4.21
Pinus strobus (albumin 2)	14.20	6.17	18.50	4.93
Pinus strobus (albumin 3)	10.63	9.38	15.63	5.63

[a] Includes aspartic acid and asparagine. [b] Includes glutamic acid and glutamine.

(Coulter *et al.*, 1990). Arginine predominates in pea at 11.5% (Higgins *et al.*, 1984), and in the conifer *Pinus pinaster* at 25.46% (Allona *et al.*, 1993). In gymnosperms, of the little information available, glutamine/glutamate predominates the amino acid composition of albumin from recently characterized *Picea glauca* (white spruce), and *Pinus strobus* albumins 1, 2, and 3 at 19.25,%, 16.32%, 18.50%, and 15.63%, respectively (Kamalay *et al.*, unpublished). High levels of these nitrogen rich amino acids are also evident in soybean (Hill *et al.*, 1974), cotton seed (Manickam *et al.*, 1980), mung bean (Youle *et al.*, 1981), mustard, sunflower, linseed, lupin, peanut, cucumber, hazelnut (Ericson *et al.*, 1986), and many others (Table 2).

In addition to high levels of nitrogen, many species have been found to contain high levels of nutritionally limiting cysteine, thereby, playing a unique role in providing a sulfur reserve in seed. This characteristic distinguishes albumins from globulins, where in the latter group of storage proteins, cysteine is generally under-represented (Higgins *et al.*, 1984; Debyshire *et al.*, 1976). For example, in alfalfa cysteine constitutes 9% (Coulter *et al.*, 1990), in castor bean 8.5% (Youle *et al.*, 1978), in pea 11.6% (Higgins *et al.*, 1984), and a very high 13.11% in Brazil nut (Youle *et al.*, 1981).

Similarly, in gymnosperms, such as *Pinus pinaster*, cysteine content was at 5.99% (Allona *et al.*, 1993). Comparably, other gymnosperm species recently characterized including *Picea glauca*, and *Pinus strobus* albumins 1, 2, and 3

show fairly high cysteine compositions of 4.97%, 4.21%, 4.93%, and 5.63%, respectively (Kamalay *et al.*, unpublished). (Table 2)

During germination, the sulfur is mobilized and utilized not only for amino acid and protein structure, but also for synthesis of important cofactors, coenzymes, and membrane sulfolipid (Youle *et al.*, 1981).

As with other storage proteins, it was found that the 2S albumins are encoded by a multigene family, where for example, 16 genes encode albumin in *Brassica napus* (Crouch *et al.*, 1987). When seed protein was fractionated, the 2S fraction was found to contain a different number of proteins for each seed, ranging from 2 to 5 in *Brassica* sp. (Monsalve *et al.*, 1990). This diversity of proteins for each seed species agrees with results reported by Scofield and Crouch (1986), and Ericson *et al.* (1986), about the existence of a multigene family for napins with different levels of expression, and thus isoforms are found in each species. (Crouch *et al.*, 1983)

Krebbers *et al.* (1988) reported the isolation and sequencing of an *Arabidopsis* 2S albumin isoform and used the sequence to isolate four genes that encode different isoforms. Hybridization studies suggested that this represented the entire gene family, however upon separation of individual 2S albumin isoforms, a previously unknown isoform differing at 10 residues was isolated and characterized giving 5 isoforms found in *Arabidopsis* (van der Klei *et al.*, 1993).

Generally it has been shown that the 2S proteins have a dimeric structure formed by 2 subunits of 7-12 kDa and 3-7 kDa, linked by disulfide bridges (Menendez-Arias *et al.*, 1988). The two components of the dimer are synthesized from a common precursor polypeptide that undergoes proteolytic processing including signal peptide cleavage required for transmembrane transport into the protein bodies for storage (Ericson *et al.*, 1986).

Several studies on the biosynthesis of 2S albumins have shown that the final storage protein product is often produced only after extensive processing of considerably larger precursors, for example, the precursor for *Brassica* napin (which is composed of 4 and 9 kDa subunits) is 20 kDa (Crouch *et al.*, 1983; Ericson *et al.*, 1986), that for the Brazil nut albumin (3 and 9 kDa subunits) is 18 kDa (Altenbach *et al.*, 1987), and that for a sunflower seed albumin (a 19 kDa monomer) is 38 kDa in size (Allen et al., 1987).

The 2S superfamily of related sulfur rich proteins from castor bean (Sharief et al., 1982), rape seed (Ericson *et al.*, 1986), mustard seed (Menendaz-Arias *et al.*, 1988), and Brazil nut (Ampe *et al.*, 1986) have the same structural organization and consist of two subunits of different sizes linked by disulfide bridges, where the number and location of cysteine residues are conserved. Sequence analysis of cDNA clones coding for the rapeseed (Scofield *et al.*, 1987) and Brazil nut (Sun *et al.*, 1987) 2S proteins has shown that these proteins are

synthesized as larger single-chain precursors which undergo post-translational modification to yield the two-chain structures of the mature proteins; whereas, in sunflower the 2S proteins are not cleaved to yield disulfide-linked large and small subunits, but instead is consisting of a single polypeptide chain (Kortt *et al.*, 1991).

Gene cloning experiments have also revealed that the two subunits are encoded on one message as part of a large precursor polypeptide (Scofield *et al.*, 1987).

3. Processing

The 2S albumins are generally made as a prepropeptide. After removal of the signal peptide, the propeptide is cleaved at four other points, giving two subunits linked by a disulfide bridge(s). Comparisons of these cleavage sites between several species suggests that while individual cleavage sites between species are conserved, the four processing sites within a species are not similar, suggesting that up to four different proteases may be involved in processing 2S albumins (Krebbers *et al.*, 1988). Hence, the majority of 2S albumins are encoded as a preprotein which is processed into a disulfide-linked heterodimeric form.

4. Homologies

The amino acid sequences among napin and other 2S proteins (Brazil nut, castor bean, yellow mustard, and

Arabidopsis) show considerable similarity in both the mature and processed regions, especially in the positioning of the cysteine residues (Menendez-Arias *et al.*, 1988). While the amino and carboxyl ends of both subunits vary, the distance between the cysteine residues is almost constant (Krebbers *et al.*, 1988). With the sequence conservation centered around the arrangement of the cysteine residues, this strongly suggests a structural homology rather than a strict sequence homology (Krebbers *et al.*, 1988). Additionally, such data suggests that disulfide-bonding patterns play an important role in the function of these proteins.

One possibility is that the disulfide bonding allows the protein to form a more compact and rigid structure as evidenced by secondary structure characterization (Allen *et al.*, 1987), which would allow the protein to withstand the harsh stress of dehydration in the dry seed stage where albumin resides in its storage form. With information accumulating on these proteins in angiosperms, it is unfortunate that so little is known about them in gymnosperms, specifically conifers.

VI. Objectives

In general, the studies presented in this thesis were designed with the objective of isolating, characterizing and analyzing developmentally regulated genes during embryogenesis in conifers. This involved the analysis of transcriptional regulation, full characterization of the coding

sequences of these genes, and assessment of similarities with previously characterized genes. Specific objectives involved the predominating studies of isolating and characterizing developmentally regulated genes from mid-embryogenesis in Douglas fir as well as initial attempts at the isolation and characterization of LEA genes from white spruce (*Picea glauca*) from the late stages of embryogenesis.

Materials and Methods

I. Material in General

Culture media for bacterial growth including tryptone, yeast extract were from Difco. Agar was supplied by Northwest. Agarose, Isopropylthiogalactoside (IPTG), ribonuclease A and lysozyme were obtained from Sigma (Mississauga, Ontario). Restriction enzymes and 10X one for all buffers for DNA digestion were obtained from Pharmacia (Uppsala, Sweden) as were modification enzymes including reverse transcriptase, DNase I, DNA ligase, and T_4 polymerase. The radionucleotides ^{32}P dCTP (specific activity, 3000 Ci/mmol (1.11×10^{14} Bq/mmol)), and ^{35}S dATP (specific activity, 100-1500 Ci/mmol ($3.70\text{-}5.55 \times 10^{14}$ Bq/mmol)) were from Dupont (Mississauga, Ont.), while the scintillation liquid Aquasol was from Du Pont. Tris was from Anachemia (Rouses Point, N.Y.), while most of the inorganic salts including NaCl were from BRL (Burlington, Ont.).

Solvents were from BDH (Toronto, Ont.). Acrylamide, N,N'-Methylene bisacrylamide, sodium dodecylsulfate (SDS) were purchased from BioRad (Mississauga, Ont.), ammonium persulfate and ultrapure urea were purchased from IBI (New Haven, CT.). Herring DNA and 5-dibromo 4-chloro 3-indolylgalactoside (Xgal) were from Boehringer Mannheim (Laval, Quebec). All antibiotics used for selection purposes

were from Sigma. A mRNA purification kit using oligo-dT spun columns was purchased from Pharmacia. The Poly $(A)^+$ Tract (System II) mRNA purification kit was purchased from Promega (Madison, WI.). Random primer labelling kits were from BRL. The cDNA synthesis kit is a product from Pharmacia, the Gigapack Gold (System II) packaging extract and Lambda ZAP II was provided by Stratagene (La Jolla, CA.).

II. Bacterial Cultures

1. Bacterial strains and plasmids

The bacterial strains used for propagation of cDNA clones were *Escherichia coli* XL1-Blue (Stratagene) and *Escherichia coli* JM105. Plasmids utilized were pUC 18/19 and pBluescript (SK-) obtained from Stratagene. Selection of plasmids was possible by the addition of ampicillin (100 μg/ml), and carbenicillin (50 μg/ml).

2. Bacterial culture propagation and storage

Purified clones of XL1-blue harboring single, positive cDNA insert in pBluescript were maintained on LB+ampicillin (100 μg/ml). For long term storage, 20% glycerol was added to the culture, then rapidly frozen in liquid nitrogen, and stored at -80°C.

Table 3: Bacterial strains, bacteriophage, and plasmid.

source	genotype (description)
bacterial strain	
E.coli XL1-Blue	*sup* E 44 *hsdR* 17 *recA* 1 *endA* 1 *gyrA* 46 *thi relA*1 (lac$^-$) [F' *proAB*+ *lacI*^{q} lacZM15 Tn*10* (tet^r)]
plasmids	
pBluescript(SK-)	(f1$^-$), colE1, 2.96 kb, Ap^r
phage	
Lambda ZAP II	
R408	f1 helper phage, ssDNA, 4 kb

3. Growth of bacterial cultures

Escherichia coli XL1-Blue strains were grown in Luria-Bertani (LB) broth pH 7.5 (Sambrook et al., 1989) at 37°C with the addition of 10 mM $MgSO_4$ and 0.2% maltose to allow for susceptibility to bacteriophage infection. Maintenance of pure XL1-Blue was possible by streaking on LB (+ 1.5% agar + 12.5 µg/ml tetracycline antibiotic (Sigma) and grown O/N at 37°C. Growth in liquid media was performed in 15 ml test tubes in gyrotory shakers (approx. 250 rpm) at the indicated temperatures.

III. Bacteriophage

1. Phage type

For cDNA cloning, the Lambda ZAP II phage from Stratagene was utilized. Additionally, the R408 f1 helper phage from Stratagene was incorporated for the purpose of *in vivo* excision of pBluescript (Table 3).

IV. Media Sterilization

All media was sterilized at 121°C under 15 lb/inch^2 of pressure for 20 minutes and allowed to cool to < 60°C before addition of supplements including antibiotics.

V. Isolation of Lea Genes

1. Preparation of total RNA

In working with RNA, measures were taken to prevent extraneous RNA contamination. All glassware and metal scoopulas were baked at 180°C for a minimum of 8 hrs. Plasticware was rinsed in chloroform as an additional precaution. All solutions were made up with DEPC (0.1%) treated dH_2O. Diethyl pyrocarbonate is a strong but not absolute RNase inhibitor. Solutions were stirred O/N and then autoclaved for 15 min. at 15 lb/in^2 on liquid cycle. Electrophoresis tanks are filled with 3% H_2O_2 for 10 min. at RT, then rinsed with DEPC treated dH_2O. Gloves were always worn and changed frequently. Miscellaneous items such as pH electrodes etc. were rinsed in 0.1 N NaOH, 1 mM EDTA followed by DEPC-treated water.

For cDNA synthesis, large scale RNA preparation from white spruce dry seeds was performed initially. One gram of seed tissue was ground in liquid nitrogen with a mortar and pestle, followed by the guanidinium thiocyanate extraction protocol and then by cesium chloride purification. A 1/40 vol. of Sarkosyl was used as well as 5 ml of GuCN buffer (4 M GuCn, 0.1 M Tris/Cl (pH 7.5), β-mercaptoethanol was added to 1%. For cesium chloride purification, an 8 ml CsCl/EDTA cushion was used with centrifugation in a SW 40 rotor for 24 hrs. at 20°C at 32 000 x g. The pellet was resuspended in a total of 700 µl TE + 0.1% SDS, followed by ethanol precipitation (1/10 vol NaOAc

(3M) + 2.5 vol 95% ethanol), washed in 70% ethanol, and resuspended in 100 µl TE pH 7.6 + 0.1% SDS. Total RNA was quantitated by use of the Beckman DU-65 spectrophotometer using the Warburg Assay program #10 from the "Nucleic Acid" software package.

Small scale RNA preparation was also performed (Maniatis *et al*, 1989) which utilizes 4 M LiCl to precipitate RNA from DNA. Yields between 25 and 50 µg of total RNA result from 100 mg of seed tissue. RNA monitoring gels were 1.0% agarose gels run at 40 volts for 2 hrs.

2. Preparation of mRNA

Small scale poly $(A)^+$ RNA was prepared by pooling two small scale total RNA preparations and utilizing the Poly A Tract kit from Promega (System II), heating to 37°C to elute, resuspended in 250 µl of DEPC dH_2O, and then assayed by the Warburg program. Cuvettes were washed with 1 N NaOH/1 mM EDTA solution three times, and then DEPC dH_2O three times to prevent mRNA degradation. Additionally, for comparison poly $(A)^+$ RNA was purified by use of oligo dT-cellulose from the Pharmacia LKB Biotechnology mRNA purification kit. This involved the separation of mRNA from total RNA by the centrifugation of oligo-dT cellulose spun columns at 350 x g using high and low salt buffers.

3. Preparation of cDNA

Preparation of cDNA from mRNA entailed reverse transcription using the cDNA synthesis kit from Pharmacia LKB Biotechnology. Five µg of white spruce poly $(A)^+$ mRNA was used for each run through the protocol. The process results in cDNA with cohesive *Eco RI* ends, to allow for insertion into bacteriophage or plasmid vectors. Quantitation of cDNA throughout the protocol was performed by an ethidium bromide-agarose plate assay (Christen et al., 1989).

4. cDNA ligation into Lambda ZAP II vector arms

Ligation of cDNA product entailed using 300 ng of cDNA and 1 µg of lambda ZAP II arms. These were then co-precipitated by using 1 µl of (3M) sodium acetate and 2 vol. 95% ethanol. ATP (final concentration of 1 mM) and 0.5 U of T_4 DNA ligase were then added and ligation performed at 12°C for 16 hrs.

5. Packaging of cDNA

Packaging of the ligated product of white spruce cDNA and the lambda ZAP II vector arms was performed by adding 130 ng of the ligation product to extracts included in the Gigapack II Gold Packaging Extract. Packaged insert was diluted in SM buffer (5.8 g NaCl, 2.0 g $MgSO_4 \bullet 7H_20$, 50 ml 1 M Tris/Cl (pH 7.5), 5.0 ml 2% (w/v) gelatin per liter). XL1-Blue cells grown to OD_{600} between 0.5 and 1.0 were diluted to 0.5 with

10 mM $MgSO_4$. Ten µl of phage was incubated with 200 µl of diluted XL1-Blue for 20 min. at 37°C. Incubation mixture was added to NZY top agar at 48°C, mixed, and then plated on NZY agar plates (made according to supplier) which were allowed to harden and then incubated at 37°C O/N. Plaques were picked up using a sterile Pasteur pipette and resuspended O/N at 4°C in 1 ml SM + 1 drop chloroform.

6. Ligation of cDNA into pBluescript (SK-)

Ligations of pBluescript SK-, *Eco RI* digested, CIAP (calf intestinal alkaline phosphatase (Boehringer Mannheim) treated vector and cDNA in a ratio of 1:1 involved 100 ng of each, 1 mM ATP, and 4 U T4 DNA ligase.

Competent *E. coli* XL1-Blue and JM105 were prepared by inoculating LB media (10 g bacto-tryptone, 5 g bacto-yeast extract, and 10 g NaCl) with a single colony, incubating O/N, re-inoculating LB and incubate at 37°C at 200 rpm until OD_{600} of 0.4-0.6. Cells were prepared by the $CaCl_2$ method, aliquoted into 200 µl, and incubated at 4°C for 12 to 24 hrs. Fifty ng of the ligation product was transfected into the competent cells by heat shock at 43°C (Maniatis, 1989). One hundred microliters of the transformation mix was then plated on LB, X-Gal, IPTG, ampicillin plates (reagents from Sigma) and incubated O/N.

7. PCR Amplification

PCR amplifications were performed on Douglas fir (*Pseudotsuga menziesii*) cDNA and genomic libraries using synthetic primers made from conserved LEA gene regions using the Gene Amp PCR Reagent Kit with *Amplitaq* DNA Polymerase. The protocol followed was that according to the supplier Perkin Elmer Cetus.

VI. DNA fragment purification

DNA fragments of interest were purified by agarose gel electrophoresis (see below) and then the band of interest carefully excised under long-wavelength UV light. DNA was purified from the agarose by use of the Gene Clean II Kit (Bio/Can Scientific Inc.). The protocol followed was that according to the supplier. NaI was used to dissolve the agarose at 50°C. Five microliters of glass milk was sufficient in most cases for binding the DNA. Three rounds of washes using NEW wash were performed. Two rounds of elution at 50°C eluted the purified fragment which was then resuspended in 2 x 10 µl TE pH 8.0 buffer (10 mM Tris/Cl (pH. 8.0), 1 mM EDTA pH (8.0)).

VII. Agarose gel elctrophoresis

Agarose gel electrophoresis was used for visualization of DNA fragments (checking or preparative purposes) as well as for checking the integrity of RNA preparations. Gels ranged

from 0.8% to 1.5% agarose depending on the size of the fragments that require optimal separation.

Agarose was weighed out and 1X TAE (50X TAE: 2 M Tris-acetate, 0.5 M EDTA pH 8.0) was added to a volume dependent on the gel size required (60 ml for a small gel or 100 ml for a large gel). This solution was heated in the microwave and then allowed to cool to < 55°C after which 0.5 µg/ml of ethidium bromide was added and mixed. The gel was then poured; a comb was inserted, and the gel was allowed to solidify. Once cooled, the comb was removed and 1X TAE was also used as the running buffer. Typically, as standards, 1 µg of either the 1 kb DNA ladder (New England Biolabs) or the Lambda Hind III digest were added as reference markers. Samples to be run had 1/6 vol. of 6X DNA sample buffer (0.25% (w/v) bromophenol blue, 40% (w/v) sucrose in water) added prior to loading on the gel. DNA samples were run at 50-80 volts for approx. 1-2 hrs. or until the indicating dye in the sample buffer was approx. 1/3 from the bottom of the gel. The DNA run on the agarose gel fluoresces under UV light after ethidium bromide staining. The gel was then photographed with Polaroid 667 film at *F* 5.6 and shutter speeds of 1 to 1/4 of a second.

VIII. Albumin Gene Isolation

1. Secondary screening

1.1 Replica filters

Plates of phage (1000-2000 pfu/150 mm plate) to be screened were transferred onto Schleicher and Schuell nitrocellulose membrane solid supports. Plates were incubated at 4°C for >12 h before lifting. Duplicate filter lifts were performed with 2 and 5 min. transfer times and then allowed to dry for 1 h. Filters were denatured in 0.2 M NaOH/1.5 M NaCl for 2 min., neutralized in 0.4 M Tris/Cl pH 7.6/2 x SSC (175.3 g NaCl and 88.2 g sodium citrate) for 2 min., and washed in 2 x SSC for 30 sec., followed by baking at 80°C for 2 h.

1.2 Single stranded cDNA probe

Secondary screening of Douglas fir cDNA library clones was performed by probing with a labelled single stranded cDNA population derived from mRNA present at mid-embryogenesis. Two micrograms of mRNA were heated to 70°C for 5 min. and then 0.5 µl RNAsin, 1 µl oligo-dT, 2.5 µl random priming buffer, 0.7 of dTTP, dATP, dGTP, 1.2 µl 1:100 diluted dCTP, 6 µl-reverse transcriptase were added. Ten µCi of ^{32}P dCTP were added and the mixture incubated for 42°C for 2 h. Then 1 µl of 0.5 M EDTA (pH 8.0), 10% SDS, and 3 µl of 3 N NaOH for RNA hydrolysis were added and the mixture incubated at 68°C for 30 min. The mixture was then cooled to RT and 10 µl of 1 M Tris/Cl (pH 7.4) and 3 µl 2 N HCl were added and mixed. Phenol/chloroform

extraction (Maniatis, 1989) was then followed by ethanol precipitation (Maniatis, 1989) using a final concentration of 2.5 M NH_4^+ Acetate with the pellet resuspended in 100 µl TE. Incorporation of radionuceotide was calculated by spotting 5 µl of sample on filter paper and washing 4 times in 50 ml of ice cold 5% TCA/20 mM sodium pyrophosphate, then rinsing in 70% ethanol and dried. Washed and unwashed filters were then placed into a scintillation vial with 4 ml of Aquasol scintillation liquid and counted in the LS 5000CE Liquid Scintilation counter (Beckman Instruments). Specific activity was calculated as the ratio of cpm incorporated/total amount of product (proportion of washed/unwashed filters x total weight).

1.3 Filter hybridization

Filters were wetted and then immersed in 10 ml (1ml/filter) formamide hybridization solution (Maniatis, 1989). Using the hybridization oven (Lab-Line Instruments), the filters were incubated at 42°C for 1 h. To 10 ml of the same hybridization solution 1 x 10^6 cpm/ml solution of probe was added along with 2 µg of herring sperm, both of which were denatured by boiling for 5 min. Approximately 700 µl of the hybridization solution was spread by glass 'hockey stick' and the filters were incubated at 42°C O/N.

The filters were rinsed with 500 ml of low stringency wash buffer (2X SSC, 0.1% SDS) 3X for 15 min. at RT, followed by washing in high stringency buffer (0.2X SSC at 42°C and once at

55°C). The filters were exposed using X-Omat XK-1 (Kodak) film O/N using intensifying screens in a cassette at -80°C.

1.4 *In Vivo* Excision

Developed film was aligned with the master plates and plaques of interest showing positive signals in duplicate were picked from agar plates using sterilized pasteur pipettes and transfered to a sterile microfuge tube containing 500 µl of SM buffer and 20 µl chloroform, vortexed, and incubated O/N at 4°C to release bacteriophage. In a 15 ml sterile conical tube 200 µl of OD_{600}=1.0 XL1-Blue cells grown O/N in LB + 10 mM $MgSO_4$ + 0.2% maltose, 200 µl of Lambda ZAP II phage stock containing 4 x 10^5 phage particles, and 1 µl (7.5 x 10^4 pfu) of R408 helper phage (7.5 x 10^7 pfu/ml). The mixture was incubated at 37°C for 15', 5 ml of 2X YT media was added, incubation at 37°C for 3 h with shaking (200 rpm), heated at 70°C for 20', centrifuged at 4000g for 5' where the supernatant was decanted into a sterile 15 ml conical tube. To plate the rescued phagmid, 15 µl of phage stock (S/N) and 200 µl of OD_{600}=1.0 XL1-Blue host cell grown O/N in LB + 10 mM $MgSO_4$ + 0.2% maltose were mixed and incubated at 37°C for 15'. Fifty microliters of this mixture was plated on LB + ampicillin (100 µg/ml) plates and incubated O/N at 37°C and 42°C to prevent co-infection with helper phage (Stratagene suppliers manual, 1991).

1.5 Mini-Preparation of plasmid

Eighteen colonies were screened for each round. Isolated white colonies were picked with a sterile loop, and inoculated into 2 ml of LB + 100 µg/ml ampicillin. Cultures were grown O/N at 37°C with gentle agitation (250 rpm). Plasmid DNA was isolated on a small scale by the alkaline lysis method essentially described by Maniatis, 1989. One and a half milliliters of each culture was transferred to a sterile 1.5 ml microfuge tube, cells pelleted at 14 000 x g for 1 min., S/N removed, and the cells resuspended by vortexing in 100 µl of solution I (50 mM glucose, 25 mM Tris/HCl, 10 mM EDTA pH 8.0). Two hundred microliters of freshly made solution II (0.2 N NaOH, 1% SDS) was then added and the tubes were inverted 5-10 times. Finally, 150 ul of solution III (60 ml of 5 M potassium acetate, 11.5 ml glacial acetic acid, 28.5 ml dH_2O) was added and vortexed. The solutions were incubated on ice for 5 min. and then microfuged at 14 000 rpm for 10 min. at 4°C. To remove the supernatant was transferred to a new tube followed by extraction of proteins with the addition of 450 µl (equal volume) of phenol:chloroform (1:1), vortexed, and microfuged at 14 000 rpm for 2'. The aqueous phase (top layer) was transferred to a fresh tube and plasmid DNA was precipitated by addition of 2 vol. of 95% ethanol. The solution was vortexed and incubated on ice for 2 min. This mixture was then microfuged at 14 000 rpm for 15 min., ethanol removed, and the pellet washed in 70% ethanol which was subsequently removed. The pellet was then

vacuum dried and resuspended in 20 µl TE + 20 ng/µl RNase A. Plasmid DNA was quantitated by using the Beckman DU-65 specrophotometer utilizing the Warburg program on the "Nucleic Acids" software package.

1.6 Restriction Digestion

Five micrograms of plasmid DNA was digested with *Eco R1* to release the cDNA inserts of various sizes. One thousand units of *Eco R1* (Pharmacia) were used for every µg of plasmid to be digested. One tenth volume of 10X "one for all'" restriction buffer (Pharmacia) was then added, the mixture was quickly microfuged at 14 000 rpm for 5 sec and then incubated O/N at 37°C.

2. Southern Hybridizations

After agarose gel electophoresis of 1 µg of the restriction digested DNA clones, the gel was photographed and then immobilized by blotting onto Zeta-Probe GT nylon membrane (BioRad). The gel was first denatured by immersion in 100 ml of 0.5 N NaOH, 1 M NaCl and shaken for 30 min. at RT, rinsed with dH_2O, and then neutralized in 100 ml of 0.5 M Tris/Cl pH 7.4, 3 M NaCl for 30 min. at RT. DNA was transferred from the gel to the membrane by capillary action using a large dish filled with 500 ml 10X SSC buffer and using Whatman 3MM paper as a wick. Bubbles were removed and then two pieces of Whatman 3MM paper were placed on top of the Zeta-Probe

membrane. Approximately 15 cm of pre-cut paper towels were placed on top of the paper in order to facilitate capillary action. The transfer was then allowed to occur O/N at RT. The blotted membrane was dried in a gel dryer (BioRad) at 80°C for 30'.

2.1 Probe Labelling

All probes including the radish LEA insert P8B6, albumin 4.7 (Malinee Chatthai), legumin (Isabel Leal), and p900 (this study) were labelled using the random primer labelling method (Feinberg et al., 1983; Feinberg et al., 1984) utilizing the Random Primer Labelling kit (BRL). Twenty-five nanograms of the purified insert of interest was resuspended in 20 µl of TE, boiled for 5 min. and then placed on ice. Immediately, thereafter, 2 µl of each of dATP, dGTP, and dTTP were added in addition to 3 µl dH_2O, and 15 µl of Random Primer Buffer Mixture according to the suppliers protocol. Five microliters (approximately 50 µCi) ^{32}P dCTP 3000 Ci/mmol, 10 µCi/µl was then added along with 1 µl Klenow enzyme and incubated for at least 1 h, after which 5 µl of stop buffer was added. Counts per minute were calculated by diluting the mixture 1:250 in TE, taking 5 µl aliquots and applying them to Whatmann 3 MM paper. One filter was not washed while the other was washed 3 times in ice cold 10% (w/v) TCA containing 1% (w/v) sodium pyrophosphate, and then washed once with 95% ethanol. The papers were dried at RT and then placed in scintillation vial

with 4 ml Aquasol and counted. For the washed filter, cpm incorporated was calculated as cpm X 2750.

2.2 Hybridization

Labelled probe (10^6 cpm/ml) was added specified by the Zeta-Probe (BioRad) supplier's protocol, to the hybridization solution (0.25M Na_2HPO_4, pH 7.2, 7% SDS). The blot was immersed in a cylinder (Lab-Line Instruments) containing the hybridization solution. The amount of the solution used was approximately 150 µl /cm^2blot. The blot was pre-hybridized for 5 min. at 65°C. The probe was denatured by boiling for 5 min. and then the appropriate amount added to the pre-hybridization mixture. Hybridization was performed O/N at 65°C with gentle agitation.

The membranes were then washed twice in 20 mM Na_2HPO_4, pH 7.2, 5% SDS at 65°C using a volume of approximately 350 µl/cm^2 blot, and then twice in 20 mM Na_2HPO_4 pH 7.2, 1% SDS at 65°C in again 350 µl/cm^2 blot. The blot was wrapped in Saran Wrap and exposed, in most cases less than 2h in a cassette with an intensifying screen at -80°C with X-Omat XK-1 (Kodak) film.

In some instances the blot was re-probed. To remove the old probe, the membrane was washed twice for 20 minutes each in a large volume of 0.1X SSC/0.5% SDS at 95°C according to the Zeta-Probe supplier's instructions. The stripped membrane was then hybridized with the new probe.

3. Zygotic Embryo Material

Douglas fir (*Pseudosuga menziesii*) immature cones were collected from several trees growing in the Canadian Pacific Forest Products high elevation coastal seed orchard at Saanichton, B.C., where cones were collected on a bi-weekly basis (except between collection dates June 14, 1993 and June 21, 1993) beginning May 17, 1993 and ending September 7, 1993. Cones were collected in plastic bags and then stored in our laboratory at 4°C for several weeks. Dissection of cones in order to separate the zygotic embryos from megagametophytes was performed on cones collected July 12, 1993 (31 DAF) and thereafter. These tissue were frozen quickly in liquid nitrogen and subsequently stored at -80°C. White spruce (*Picea glauca*) mature seeds were obtained from the Surrey Seed Center, B.C. Ministry of Forests.

4. Northern hybridizations

Total RNA was prepared from Douglas fir developing embryos at -24, -20 (before fertilization), 3, 10, 17, 31, 45, 59, 73 DAF (days after fertilization), and dry seed stages in development. At 31 DAF, and thereafter total RNA was prepared from the separated zygotic embryo as well as the megagametophyte.

RNA samples and RNA markers (BRL) for electrophoresis were prepared by using 20 µg RNA dissolved in 4.5 µl dH_2O 3 µl

formaldehyde, Ten µl formamide and 2 µl of 5X sample buffer (20.6 g MOPS 13.33 ml, three molar NaOAc, 800 ml dH_2O, pH to 7.0, with NaOH, and 10 ml 0.5 M EDTA). Samples were heated to 65°C, for 5 min., then cooled on ice. Next, 3 µl of sample buffer mix was added (11 µl of 1 mg/ml ethidium bromide and 22 µl sample buffer (5 ml 100% sterile glycerol 20 µl 0.5M EDTA, 25 mg bromophenol blue, 25 mg xylene cyanol)). Gels of 1% agarose containing 2.2 M formaldehyde made up with the 5X formaldehyde running buffer were pre-run for 5 min. and then run at 35 volts for 4 h in 1X formaldehyde running buffer. The gel was then photographed with Polaroid 667 film at *F*5.6 at 1/2 sec. RNA markers (BRL) were marked with India Ink on the gel as well as on the membrane. Northern transfer was performed similarily to the Southern transfer (see above), however no denaturation step was required. Gels were again transferred by capillary action to the pre-cut Zeta Probe GT membrane by blotting in 10X SSC O/N (Maniatis, 1989). The blot was rinsed in 2X SSC and then dried at 80°C for 30 min.

The p900 cDNA clone was labeled by random priming (BRL) with ^{32}P dCTP (3000 Ci/mmol) according to the manufacturer's protocols to a specific activity of approximately 1 x 10^8 cpm/µg or 10^6 cpm/ml of hybridization solution. The blot was pre-hybridized in a cylinder supplied by the hybridization oven manufacturer Lab-Line Instruments. The amount of the solution (0.25 M Na_2HPO_4, pH 7.2, 7% SDS) used was approximately 150 µl/cm^2 blot. The blot was prehybridized

for 5 min. at 65°C. The probe was denatured by boiling for 5 min. and then the appropriate amount (usually 1 x 10^7 cpm) added to the pre-hybridization mixture (10 ml in most cases). Hybridization was performed O/N at 65°C with gentle agitation.

The membranes were washed twice in 20 mM Na_2HPO_4 pH 7.2, 5% SDS at 65°C using a volume of approximately 350 µl/cm^2 blot, and then twice in 20 mM Na_2HPO_4 pH 7.2, 1% SDS at 65°C in again 350 µl/cm^2 blot (usually 100 ml/blot). The blot was exposed O/N in a cassette with an intensifying screen at -80°C with X-Omat XK-1 (Kodak) film.

5. Sequencing

DNA sequencing of the positive clone (p900) was performed by utilizing the dideoxy chain termination method of Sanger *et al.* (1977) using the Sequenase Version 2 sequencing kit (United States Biochemical). Primers used for sequencing included the T3 and T7 primers provided and custom synthetic oligonucleotide primers made in the lab with an oligonucleotide synthesizer (Applied Biosystems), thereby allowing for complete sequencing of the cDNA insert. Annealing was performed by incubating the primer:template mix at 65°C for 2 min., then allowing the mix to cool slowly to 30°C over a period of about 1 h. After annealing, 5 µCi of ^{35}S dATP (1000 Ci/mMol), 1 U of the Sequenase Version 2 enzyme, 1.5 µM of each dGTP, dCTP, and dTTP were added along with 1µl DMSO to the primer:template mix, and incubated for 2-5 min. at room temperature. Three and a half µl of this mix were then added to four tubes

containing 2.5 µl of the respective termination reaction (G, A, T, or C), where each one contains 80 µM of dCTP, dTTP, dATP, and dGTP, 50 mM NaCl and 8 µM of either ddCTP, ddTTP, ddATP or ddGTP. Additionally reactions were performed using the analogs (dITP or 7-deaza dGTP) for the purpose of eliminating secondary structures. The mixtures were then incubated at 37°C for 20 min. and then stopped with stop solution (95% formamide, 20 mM EDTA, 0.5% bromophenol blue and 0.5% xylene cyanol. Samples were heated to 80°C for 2 min, rapidly cooled on ice and 2.5 µl of each reaction mixture was loaded on a 4 mm thick, 6% polyacrylamide-8M urea sequencing gel (Davies, 1982). The running buffer was 135 mM Tris, 45 mM boric acid, 2.5 mM Na_2EDTA, pH 8.0 (TBE) and the gels were run at 40 watts (1500 volts) in a JBRL-S2 sequencing electrophoresis apparatus (Bethesda Research Labs. Life Technologies, Incubate). Samples were loaded 2 times at intervals of 2 h to increase distance of readable sequence. The gel was then fixed in 10% methanol/10% acetic acid for 15 min., transferred onto Whatman 3MM paper and dried for 1.5 hrs at 80°C. The gel was exposed O/N with X-Omat XK-1 (Kodak) film at room temperature.

6. Data analysis

Computer assisted analysis was performed using the DNA Strider 1.0 program for analyzing the p900 DNA sequence and the deduced amino acid sequence. The DNA sequence itself and

the resulting protein sequence were compared to computer data bank sequences for similarities using BlastX and BlastN programs for access to multiple data bases simultaneously including PDB, SwissProt, PIR, SPUpdate, GenPept, GenBank, and GPUpdate. Multiple nucleic acid and amino acid sequence comparisons were performed by use of the Clustal V multiple sequence alignment from SeqApp 1.9.

Results

I. Isolation of LEA genes from conifers

Initial attempts at the isolation of genes from conifers during the late stages of embryogenesis, specifically late embryogenesis abundant (LEA) genes, began by utilizing the dry seed stage of white spruce. Using a total of 2.1 g of tissue, large scale RNA preparation yielded at total of 944.8 μg of total RNA (data not shown), a 0.05% yield which agrees with the theoretical yield of apprximately 0.1%. Using oligo dT-cellulose columns, mRNA was prepared from the total RNA giving 63.9 μg, a 6.76% yield, while using Poly A Tract from Promega (System II), only 2.88 μg of mRNA resulted, a 0.3% yield. With the improved yield from the oligo dT-cellulose columns, 5.0 μg of poly A^+ RNA was then used for construction of the white spruce cDNA library.

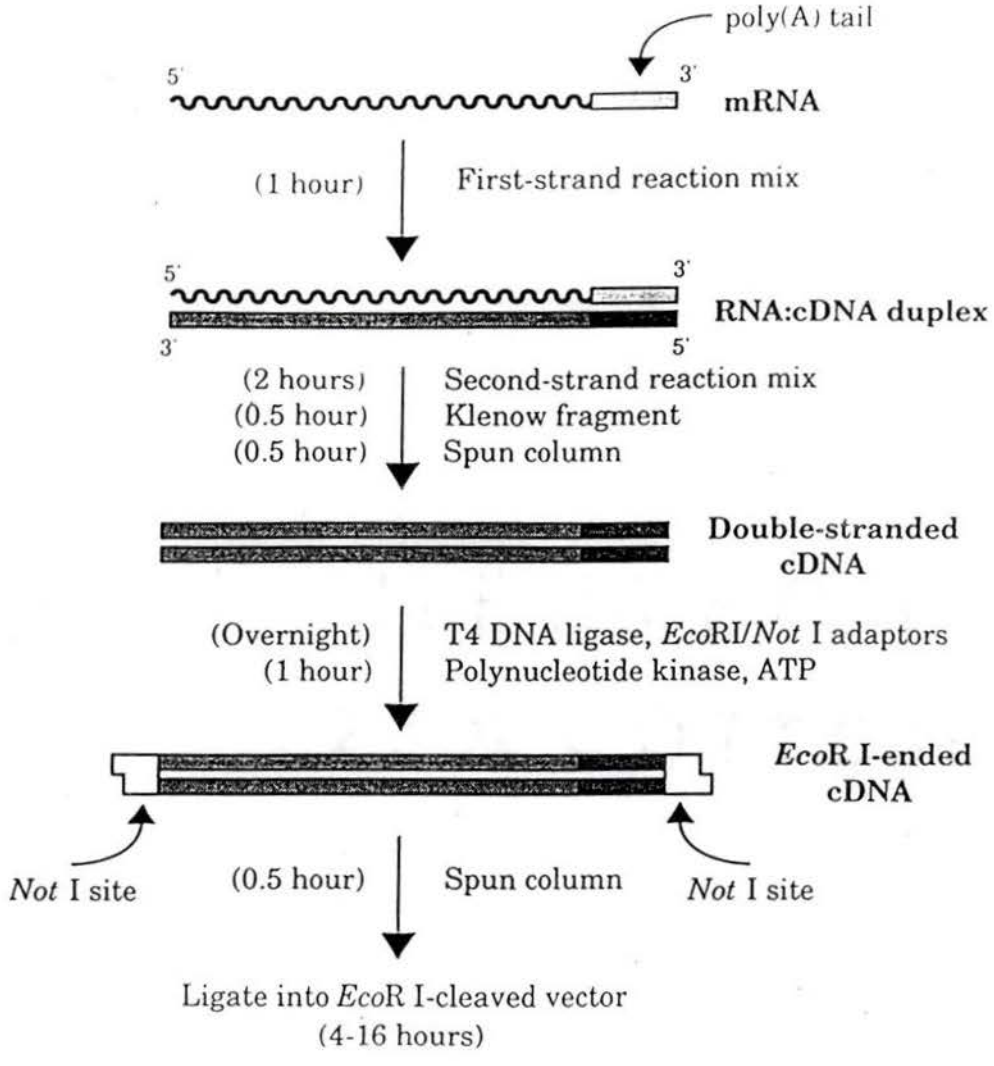


Figure 2. Flow chart showing the steps involved in cDNA synthesis from mRNA prepared from white spruce (*Picea glauca*) dry seed.

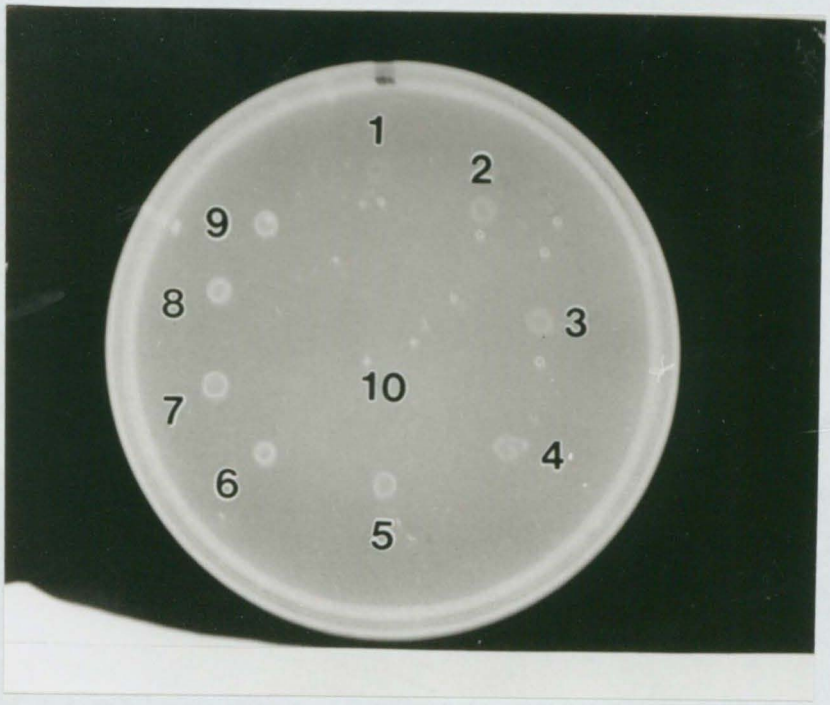


Figure 3. Ethidium bromide plate assay (Christen et al., 1989) allowing for quantitation of cDNA synthesized from white spruce mRNA (*Picea glauca*) by comparisons with standards 1-9 (10-90 ng) to the sample (10). Plate used is 1% agarose with 5 µg/ml of ethidium bromide.

Multiple sequence alignment of EM-like proteins

Figure 4. Amino acid sequence alignment showing conserved LEA regions utilized for primer synthesis for the purpose of PCR amplification.

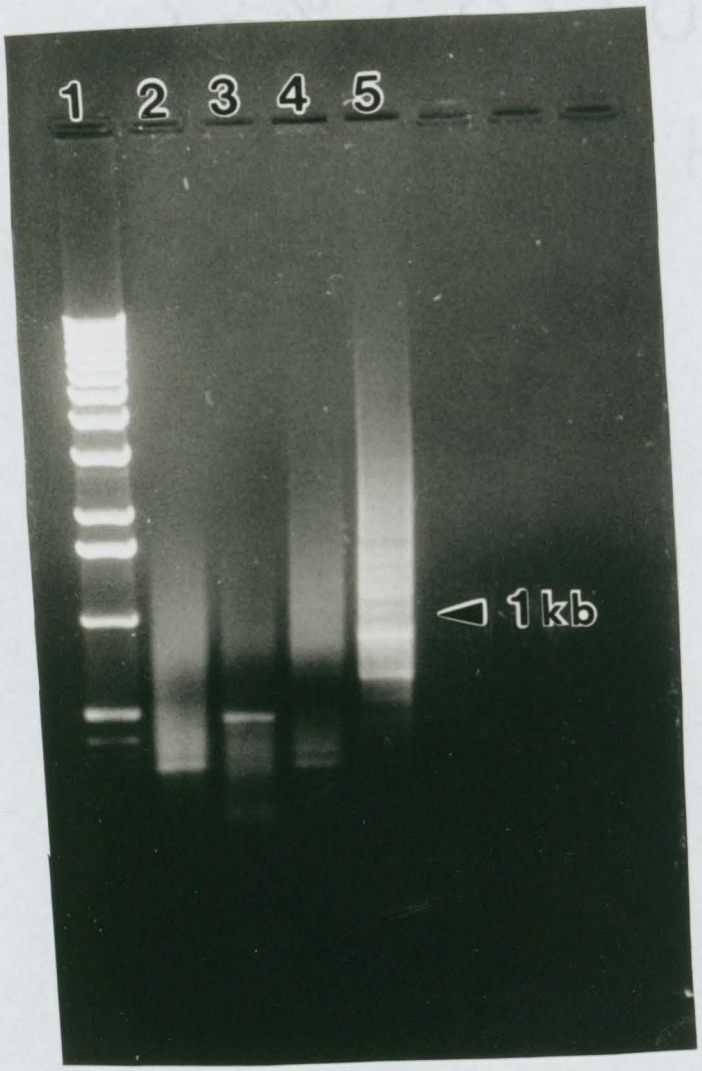


Figure 5. PCR products run on a 1% agarose gel. PCR amplification was performed using synthetic primers 1 and 2. Lane 1, 1 kb DNA ladder. Lane 2, Douglas fir genomic DNA. Lane 3, Douglas fir cDNA. Lane 4, Douglas fir cDNA. Lane 5, tobacco genomic DNA.

1. Construction of a cDNA library from white spruce

Using 5 µg of mRNA, cDNA library construction was attempted. The Pharmacia cDNA synthesis kit was employed and the strategy followed was that of synthesis of the first strand by reverse transcription, followed by second strand synthesis, and subsequent addition of adaptors (fig. 2). The cDNA product was quantitated by the nonradioactive ethidium bromide agarose plate assay using 5 µg cDNA products resulting from the initial attempts appeared to be very low in quantity (fig. 3) at <20ng. When assayed by spectrophotometry, only 0.44 ng of cDNA had been synthesized. This led to additional rounds of cDNA synthesis in order to obtain sufficient quantities for ligation. After several rounds, the cDNA products resulting from each were then pooled, however, in each case very little product was synthesized. Assaying by spectrophotometry gave an indication of the quantity of cDNA, however, differentiation between cDNA product and mRNA was not possible. Alternatively, the fluorometric assay was employed which showed <100ng of cDNA in the pooled product.

Attempts at ligating the cDNA products into Lambda Zap II vector arms and subsequent packaging into phage resulted in no plaque forming phage. Similarly, ligation into *Eco R1* digested pBluescript (SK^-) plasmid followed by transformation of XL1-Blue cells resulted in no positive clones after screening. It was concluded that the levels of cDNA were insufficient for library construction. This was likely due to the presence of inhibitors in the seed coat

which interfere with the reverse transcription of mRNA to sscDNA (S. Baker, personal communication).

2. PCR amplification of LEA sequences

An alternative to cDNA library construction, was that of PCR amplification of LEA DNA from existing Douglas fir genomic and cDNA libraries. Utilizing previously characterized LEA protein sequences from various angiosperms, conserved regions among these proteins were used to construct two oligonucleotide primers; primer 1 and primer 2 (fig. 4) assuming that the homologies with Douglas fir would be sufficient to allow primer hybridization for the purpose of amplifying the LEA DNA. After several rounds of PCR amplification of Douglas fir cDNA and genomic libraries as well as a genomic tobacco DNA control, the resulting products were run on 1% agarose gel (fig. 5). The expected sizes of amplified DNA were 475 bp and 150 bp for the cDNA library and 300 bp for the genomic library. Stronger banding can be seen at these sizes, however, it is evident that a large variety of non-specific PCR amplified product also resulted from the additional array of bands seen in each case as well as the products resulting from the negative tobacco genomic DNA control. Southern hybridization was performed on the products transferred from the gel to a Zeta-probe membrane. The probe P8B6 used was a 447 bp cDNA of radish LEA. Hybridization was attempted twice and in each case no hybridization signal was seen (except for slight non-specific binding of the probe to the DNA markers (fig. 6)).

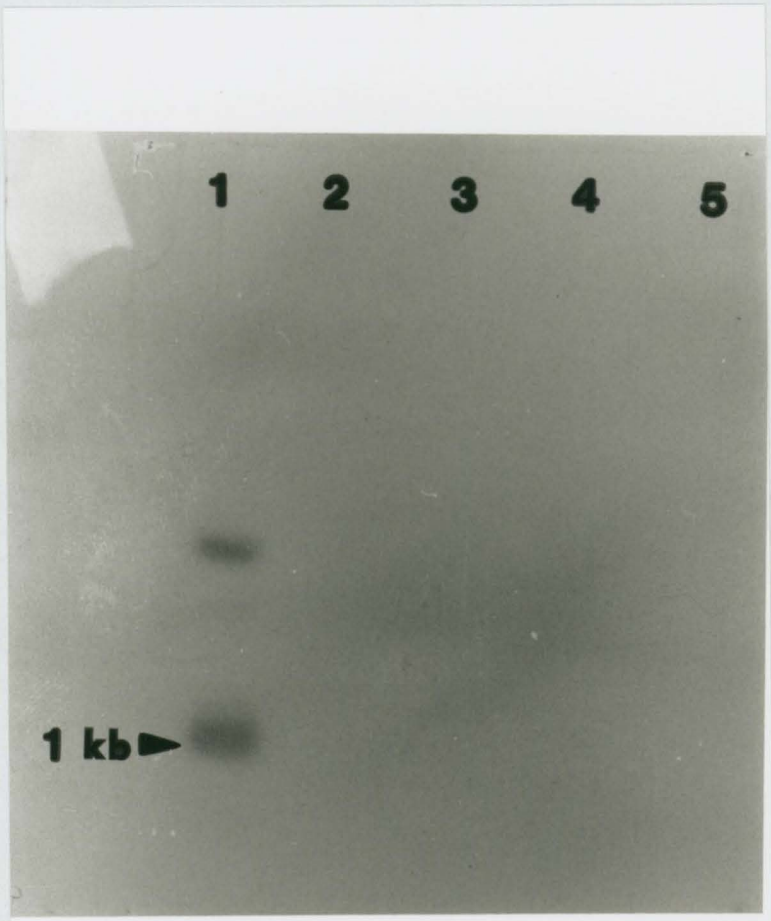


Figure 6. Southern hybridization of heterologous probe P8B6 with amplified PCR products: Lane 1, 1kb ladder. Lane 2, Douglas fir genomic DNA. Lane 3, Douglas fir cDNA. Lane 4, Douglas fir genomic DNA. Lane 5, Tobacco genomic DNA control.

A final attempt was made by excising the predicted bands from a preparatory gel, purifying the cDNA, blunt-ending the DNA, ligating into a *Sma I* (blunt-end) restricted pBluescript (SK^-) vector, and transforming into XL1-Blue cells. After screening, positive clones were characterized and unfortunately were found to be false positives. It was concluded that the PCR amplified products were the result of non-specific binding of the synthetic oligonucleotide primers. This is likely due to the fact that the primers were derivedfrom the existing regions of conserved amino acid sequences, hence the nucleotide sequence diverged enough not to allow hybridization of the primers to the Douglas fir LEA DNA.

II. Isolation of mid-embryogenesis specific clones from Douglas fir

Attempts were made to isolate clones from a cDNA library prepared from the mid-stage of Douglas fir embryogenesis. A full length cDNA was isolated and characterized.

1. Identification of cDNA clones

Initially, a Douglas fir cDNA library constructed from the RNA transcripts from the mid-embryogenesis stage of the developing seed was utilized (Leal and Misra, 1993). Primary differential screening previously done had given many potentially positive phage harboring cDNA inserts (fig. 7). Secondary screening was performed with six of these clones using radiolabelled single stranded cDNA probes reverse transcribed from the mRNA population of the mid-embryogenesis stage of Douglas fir seed development. Positive

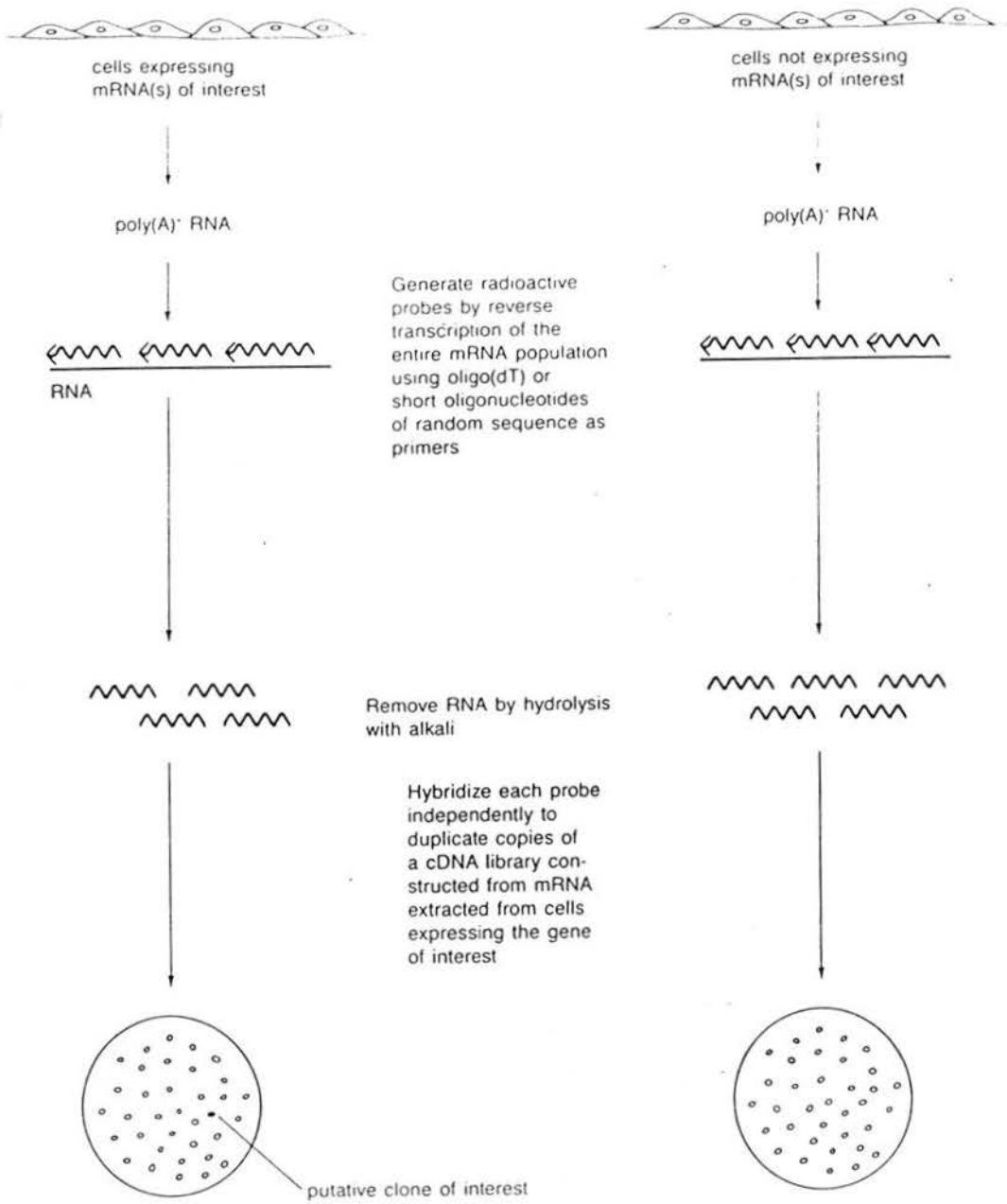


Figure 7. Flow chart of the strategy involved in differential screening of the Douglas fir cDNA library obtained from the mid-embryogenesis stage of seed development (Maniatis, 1989).

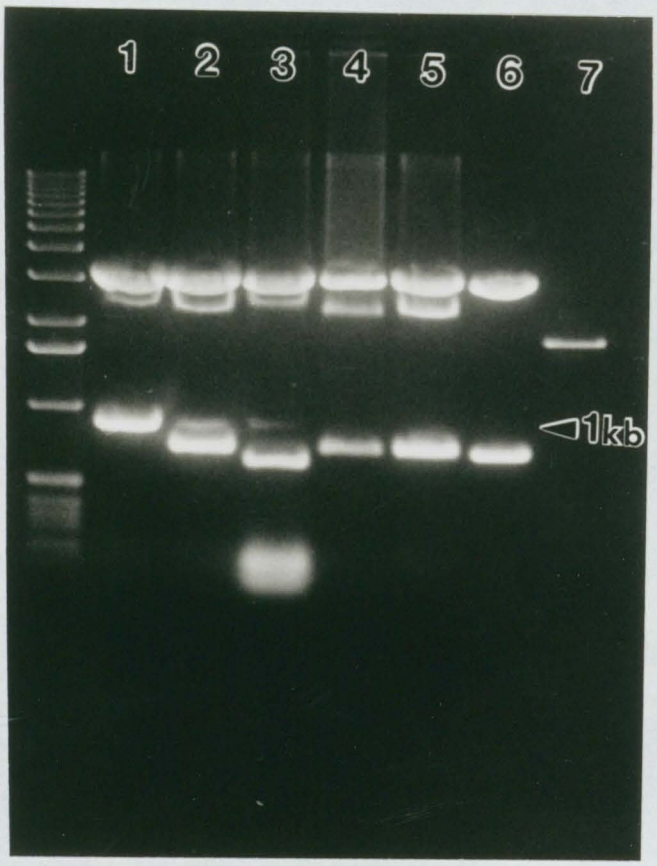


Figure 8. Agarose gel analysis of cDNA clones obtained from the mid-embryogenesis stage of seed development. One microgram of *Eco R1* digested DNA from five clones was run. Lane 1, clone 1 (p900). Lane 2, clone 2. Lane 3, clone 3. Lane 4, clone 4. Lane 5, clone 5. Lane 6, albumin 4.7. Lane 7, legumin cDNA.

Figure 9. Southern hybridization of ^{32}P labelled legumin cDNA with the 5 clones. Left margin (non-radioactive marker). Lane 1, clone 1 (p900). Lane 2, clone 2. Lane 3, clone 3. Lane 4, clone 4. Lane 5, clone 5. Lane 6, albumin 4.7 cDNA. Lane 7, legumin cDNA.

clones identified by signals on duplicated filters (data not shown) were isolated. Using successive rounds of in vivo excision of the phagemids, followed by *Eco R1* digestion in order to release the cDNA inserts, 5 clones were isolated (fig. 8). The approximate sizes of the cDNA inserts were as follows: clone 1, 900 bp, clone 2, 700 bp, clone 3, 600 bp insert and a 200 bp insert, clone 4, 700 bp insert, and clone 5, 700 bp insert (fig. 8). Cross-hybridization studies were done to eliminate previously characterized cDNAs. Probing of the blot containing all cDNA inserts from five selected clones and additional controls of previously characterized legumin and albumin (4.7) from Douglas fir was undertaken. A Southern hybridization using ^{32}P labeled Douglas fir legumin cDNA showed the control legumin insert as a signal approximately 1.8 kb. Clone 3 gave a positive signal with its 600 bp cDNA insert, but did not show any hybridization signal with the smaller 200 bp insert (fig. 9). This indicates that clone 3 is a partial legumin cDNA clone fused to an approximately 200 bp cDNA of non-legumin origin. Another Southern hybridization was performed on a similar blot with the inserts from selected cDNAs as well as controls using a previously characterized Douglas fir albumin 4.7 cDNA as the probe which is 700 bp in size. The probe 4.7 cross-hybridized to clones 2, 4, and 5 indicating sequence similarities to an albumin. However, clone 1 showed no hybridization signal with this probe (fig. 10). These two Southern blots, thereby, eliminate the possibility that the clone 1 cDNA insert is either the previously characterized legumin, or the albumin 4.7.

To confirm this observation, two Southern analyses were performed using the clone 1 cDNA insert as the radiolabelled probe. One Southern involved washing the blot under the standard high stringency conditions, while the second involved washing of the blot under only low stringency conditions in order to allow for the possibility of weak hybridizations between the probe and the cDNA inserts. The initial high stringency Southern blot showed a very strong signal of the clone 1 cDNA insert at approximately 900 bp (fig. 11). No other signals were evident, including the legumin or albumin 4.7 controls. Therefore, it appears that the 900 bp cDNA insert of clone 1 (p900) is unique and not previously characterized.

The next Southern blot was performed using the same p900 labelled insert as the probe, but the blot was washed under low stringency conditions only. Again, the results show that p900 hybridized strongly to clone 1, but there appeared to be a weak hybridization signal with the albumin 4.7 control insert and extremely weak signals with the cDNA inserts of clones 2, 4, and 5 (fig. 12). This indicates that there is slight homology to the albumin 4.7 possibly identifying p900 as an albumin not previously characterized in Douglas fir.

2. Developmental regulation of p900 albumin

The albumin p900 transcript size and its relative abundance was examined by Northern hybridization. Total RNA was prepared from both the zygotic embryo and the megagametophyte (maternal tissue surrounding the embryo) at mid-embryogenesis (45 DAF) as

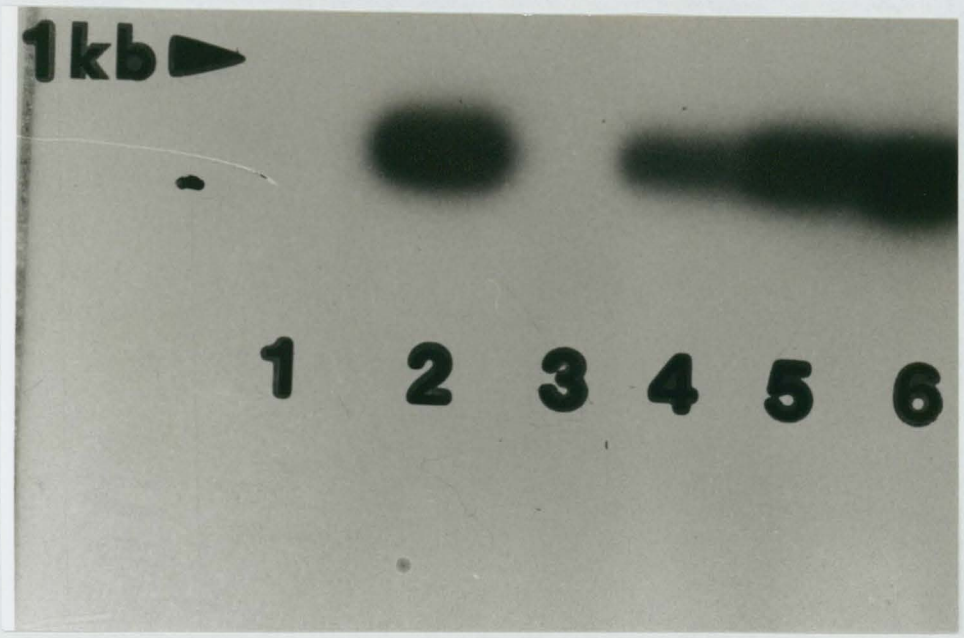


Figure 10. Southern hybridization of ^{32}P labelled albumin 4.7 cDNA insert with the 5 clones. 1 kb marker (non-radioactive in margin). Lane 1, clone 1 (p900). Lane 2, clone 2. Lane 3, clone 3. Lane 4, clone 4. Lane 5, clone 5. Lane 6, albumin 4.7.

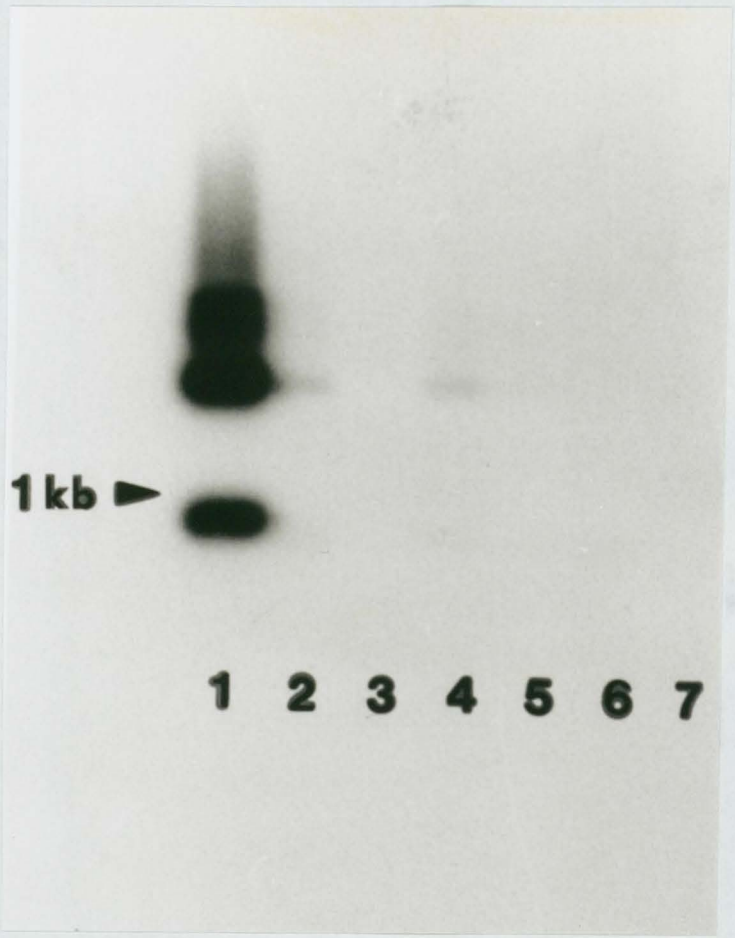


Figure 11. Southern hybridization of ^{32}P labelled clone 1 (p900) with the 5 clones digested with *Eco R1* (1 µg/lane). Blot was washed under high stringency conditions. Margin, 1 kb ladder (non-radioactive). Lane 1, clone 1 (p900). Lane 2, clone 2. Lane 3, clone 3. Lane 4, clone 4. Lane 5, clone 5. Lane 6, albumin 4.7 cDNA. Lane 7, legumin cDNA.

Figure 12. Southern hybridization of ^{32}P labelled clone 1 (p900) with the 5 clones digested with *Eco R1* (1µg/lane). Blot was washed under low stringency conditions only. Lane 1 clone 1 (p900). Lane 2, clone 2. Lane 3, clone 3. Lane 4, clone 4. Lane 5, clone 5. Lane 6, albumin 4.7 cDNA. Lane 7, legumin cDNA. Margin, 1 kb ladder (non-radioactive).

The transcripts are absent in the dry seed and post-germination stages of seed development, thereby, indicating that this albumin gene is expressed during embryogenesis only.

To further explore this developmental regulation, a time course over the entire period of embryogenesis was set up in order to accurately assess relative transcript abundance. In order to ensure equivalent amounts of mRNA at each time point during embryogenesis, the blot was probed with a ribosomal RNA probe (data not shown). The resulting Northern hybridization with radiolabelled p900 insert confirms the transcript size to be approximately 900 bases. More importantly, is clear from this blot that there is a definite developmental regulation of the p900 albumin transcript (fig. 14). Transcript levels are low during early stages of embryogenesis but start increasing during mid-embryogenesis (10 DAF), reaching peak levels at 31 and 45 DAF (fig. 14, lanes 6 and 7). Transcript levels decreased by 59 DAF (lane 8), and are completely absent at 73 DAF and dry seed (fig. 14, lanes 9 and 10). In the zygotic embryo, the transcripts are much less abundant at 31 DAF, peaks at 45 DAF (however its abundance is much less than in the megagametophye at this time point), decreases at 59 DAF, and then is non-existent thereafter (73 DAF and dry seed).

Table 4: Dates of Douglas fir cone collection during spring/summer 1993 and corresponding stages of zygotic embryo development.

Date of Collection, (DAF)[a]	Developmental stage
May 17 (-24)	Prefertilization
May 31 (-10)	Prefertilization
June 14 (3)	Embryonal mass
June 21 (10)	Embryonal mass
June 28 (17)	Club-shaped embryo
July 12 (31)	Early cotyledonary
July 26 (45)	Mid-cotyledonary
August 9 (59)	Late cotyledonary
August 23 (73)	Late cotyledonary
September 6 (87)	Mature embryo (dry seed)

[a] Days after fertilization.

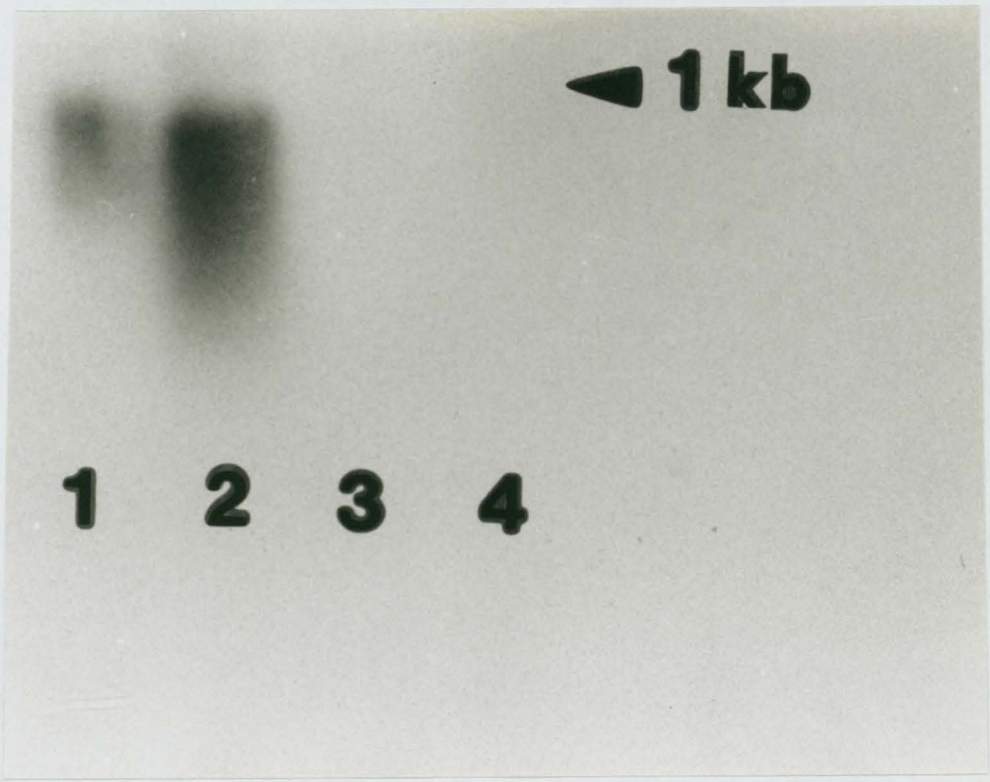


Figure 13. Northern hybridization of p900 cDNA insert with total RNA obtained from Douglas fir. Lane 1, zygotic embryos at 45 DAF; Lane 2, megagametophytes at 45 DAF; Lane 3, the dry seed stage; Lane 4, 2-day-old seedlings. Twenty micrograms of total RNA was loaded in each lane.

4. Characterization and analysis of p900 albumin

Sequence characterization of full length (836 bp) cDNA clone p900 showed the ATG initiation codon at nucleotide position 33, thereby giving a 5' non-coding region of 32 nucleotides. The termination codon of TGA, giving an opal termination signal, is found at position 451, giving a 3' non-coding region of 383 nucleotides (fig. 15). This results in a 420 bp open reading frame that when translated, gives a 140 aa, 15.8 kDa preprotein. Additionally, the polyadenylation signal AATAAA is located at position 707 (fig. 15). Beginning at the N-terminus of the predicted precursor polypeptide, there is a hydrophobic region with the characteristics of a typical signal peptide. Utilizing the theorem for determining the cleavage site between the signal peptide and the remaining protein (Von Heijne et al., 1983), it is determined that cleavage occurs between Ala-21 and Ala-22 giving a signal peptide of 2.4 kDa and a processed protein of 13.4 kDa. Scanning the sequence, It was found that the Asn-x-Ser/Thr motif is not present, thereby indicating that there is no N-linked glycosylation.

Upon multiple sequence alignment of the cDNA nucleotide sequence with several other previously characterized conifer albumins, it was found that 408 nucleotides were conserved between the p900 albumin and *Picea glauca* albumin, *Pinus strobus* albumins 1, 2, and 3 (fig. 16). This gives an overall 48.8% identity between them. The predicted amino acid composition of the p900 albumin is shown in Table 5. It is evident that nitrogen laden amino acids

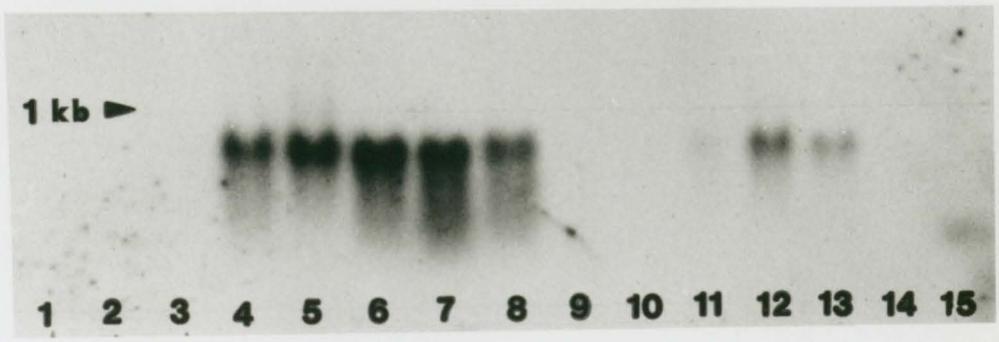


Figure 14. Time course of accumulation of p900 transcripts in Douglas fir. Lanes 1-9, megagametophyte at -24, -10, 3, 10, 17, 31, 45, 59, and 73 DAF, respectively. Lane 10, dry seed stage of the megagametophyte. Lanes 11-14, zygotic embryo at 31, 45, 59, and 73 DAF, respectively. Lane 15, the dry seed stage of the zygotic embryo. Twenty micrograms of total RNA was loaded in each lane. ^{32}P labelled p900 cDNA insert was used as a probe.

```
cctccgtttcgacatcgtgctgctgaggctaaaATGGGTGAGTTTAGGCG 50
                                  M  G  E  F  R  R

TGGCCTTCTCCTCCTCGTTCAGTGGAGCACCGCCAATGTTGATGCGGCTG 100
 G  L  L  L  L  V  Q  W  S  T  A  N  V  D  A  A  E

AAGACAATATGTTTGGAGAGGACGTCGTGCAACAACGACGACGGTTGGCC 150
  D  N  M  F  G  E  D  V  V  Q  Q  R  R  R  L  A

AACCCGCAGAGACTATCCTCTTGCCGGGACTACTTGGAGCAGCGAAGAGA 200
N  P  Q  R  L  S  S  C  R  D  Y  L  E  Q  R  R  D

CCAACCATCGCAAAGAATGTCTCCGCGATGCCGATGCCCAGCCATACAGC 250
 Q  N  P  Q  R  L  S  S  C  R  D  Y  L  E  Q  R  R

AAACGGTGGATCAATCTTATCATTCATGGATTCTGATTCTCAAGAGGATG 300
  D  Q  P  S  Q  R  M  S  P  R  C  R  C  P  A  I

CACCACTTATCAACGCCGCCGCCGCAGCCGCCGCGAAGGGCGCCGACGGG 350
Q  Q  T  V  D  Q  S  Y  H  S  W  I  L  I  L  K  R

AAGAGGACGAGGAGGAAGAGGTGCTGGACAGAGCGGCAGACCTTCCGAAT 400
 M  H  H  L  S  T  P  P  P  Q  P  P  R  R  A  P  T

ACCTGCGACGTCCGCCAGCCGGAGCCATGCGACATTCAACGTCACTCTCG 450
  G  R  G  R  G  G  R  G  A  G  Q  S  G  R  P  S

TGAgtcattcatcaatactgatcaattaattctttgccctgcgtcgtctt 500
*
agaataacgccgccatgcaaatttctttctttcagtcatgtcatttttct 550
tgttattccatggcatagtgagcgatgaatcttaagatatatgtgatgtt 600
ttactggtggtgggtggacagttattccatgacgggcagcatcatgtaag 650
tgagtgatcatgatcatgatcatgaataagagatatacataaatagtgtt 700
ttaactaAATAAAaagggaggcacttcaaccgctatgaatttgtggtaat 750
tactttatagccggatctcgccccctccatgtaacaaagaatggatgaac 800
aaagaatggatgaacgataaactgcaagagaagaag               836
```

Figure 15. Nucleotide and deduced amino acid sequence of cDNA clone p900.

predominate this protein's composition. Arginine comprises the majority of the protein at 15.7% followed by glutamine/glutamic acid at 12.2%. This protein is also fairly high in cysteine at 2.1%.

Analysis of the predicted amino acid sequence gives a hydropathy plot consistent with the predicted characteristics of an albumin. It is clear that the N-terminal portion of the protein contains a hydrophobic sequence characteristic of a signal sequence. The remainder of the protein is highly hydrophilic consistent with the hydrophilicity of albumins in general. The center of the protein is slightly hydrophobic and this characteristic is comparable to that of *Picea glauca* albumin, and *Pinus strobus* albumins 1, 2, and 3 (fig. 17). The acidity/basicity map clearly shows that the p900 albumin is comprised mainly of basic residues making it a basic protein, positively charged under physiological pH (fig. 18).

An examination of the amino acid sequence shows that when the amino acid sequences between p900 and *Picea glauca, Pinus strobus* albumins 1, 2, and 3, are aligned an overall conservation of 30.5% of the residues is seen excluding the signal peptide (fig. 19). Many of the arginines and glutamines are clearly conserved, but interestingly many of the cysteines so characteristic of albumins are also present. Many of their positions are comparable to those of other albumins.

Table 5: Amino Acid Composition of Translated p900 Albumin.

140 Amino Acids MW : 15819 Dalton

		n	n(%)	MW	MW(%)
A ala	alanine	9	6.4	639	4.0
C cys	cysteine	3	2.1	309	2.0
D asp	aspartic acid	6	4.3	690	4.4
E glu	glutamic acid	5	3.6	645	4.1
F phe	phenylalanine	2	1.4	294	1.9
G gly	glycine	11	7.9	627	4.0
H his	histidine	4	2.9	548	3.5
I ile	isoleucine	3	2.1	339	2.1
K lys	lysine	1	0.7	128	0.8
L leu	leucine	12	8.6	1357	8.6
M met	methionine	5	3.6	655	4.1
N asn	asparagine	3	2.1	342	2.2
P pro	proline	13	9.3	1261	8.0
Q gln	glutamine	12	8.6	1536	9.7
R arg	arginine	22	15.7	3434	21.7
S ser	serine	13	9.3	1131	7.2
T thr	threonine	5	3.6	505	3.2
V val	valine	5	3.6	495	3.1
W trp	tryptophan	2	1.4	372	2.4
X ---	unknown	-	-		
Y tyr	tyrosine	3	2.1	489	3.1
Z ---	STOP	1	0.7		

n: number of amino acids
%n: percentage of amino acid

```
p900        AATATGGGTGAG-TTTAGGCGTGGCCTTCTCCTCCTCGTTCAGTGGAGCA
picea       C-----------------------------------------------
P.1         TTCATGAAAGAAATCTTCAACCTCTCAGCTTCAGCTGGATCAGATTATCT
P.2         CTTATG------------------------------------------
P.3         C-----------------------------------------------

p900        CCGCCAATGTTGATGCGGCTGAAGACAATATGTTTGGAGAGGACGTCGTG
picea       --------------------------------CAATGGGTGTCTTTTCCC
P.1         CTCCATCTTTCACCGTAACGAAGAAAGAAAAGCAATGGATGTTTTCTGCC
P.2         -----------------------------------GGTGTCTTCT-CC
P.3         --------------------------------CAATG------------

p900        CAACAACGACGACGGTTGGC---CAACCCGCAGAGACTATCCTCTTGCCG
picea       CTTCGACGACGA-GGCTGACGCTCAAATGGTTCAGTTTATCCGTCGCCCT
P.1         CTCCGACTACGA-GACTGGCTCTTAAATGGGCGAGCCTAGGCGTACGCCT
P.2         C-----------------CTCTTAAATGGGTGAGCTTAGGTGTAGCCCT
P.3         --TCGA---CGA-GGTTGACGCTAAAATGGGTGACTTTAGTGGCGGCCCT
                             *    **   *   *   **        **

p900        GGACTACTTGGAGCAGCGAAG-AGACCAACCATCGCAAAGAATGTCTCCG
picea       GTTCCTCCTCTTTCACTGGGGTATTCCCAGTGTTGATGGCCATGAAGACA
P.1         TCTCGTCCTCCTTCAGTGGGGCACCCCAACTGTTGATGCGCGTGAAGGCG
P.2         TCTCCTCCTCCTTCAGTGCGCCACCCCAACTGTCGATGCGCGTGAAGGCG
P.3         TCTGTTTGTCATTCACTGTAGCACCCCCACTGTTGGTGCACATGAAGACA
                  *    **  *    *  ** *   * *        **    *

p900        CGATGCCGATGCCCAGCCATACAGCAAACGGTGGATCAATCTTATCATTC
picea       ATATGTATG-GAGAGGAGATACAACAA-CAAAGA--CGGTCGTGCGACCC
P.1         TCATGTATG-GTGAGGGCCTGCAACAA-CAACGA--CGCTCCTGCGATCC
P.2         TCATGTATG-GAGAGGGCCTGCAACAA-CAACGA--CGCTCCTGCGATCC
P.3         CCATGGATG-GAGAGGCGCTGCAACAA-CAACGA--CGGTCGTGCGA---
               ***     *    *   * ** *** *   *   *  ** *   *

p900        ATGGATTCTGATTCTCAAGAGGATGCACCACTTATCAACGCCGCCGCCGC
picea       TCAGAGA--GACCCGCA-GAGATTGT-CTTCATGCCGGGACTACT--TGG
P.1         --------------GCA-GCGACTGT-CTGAGTGCCGGGAGTACA--TGG
P.2         --------------GCA-GCGACTAT-CCCAGTGCCGGGACTACA--TGG
P.3         -----------CCCGCA-GAGATTGT-CTGATTGCCACGACTACT--TGC
                         ** * *  *   *    * *       *    *

p900        AGCCGCCGCGAAGGGCGCCGACGGGAAGAGG--ACGAGGA---GGAAGAG
picea       AGCGCCGGAGAGAGCAGCCATCGGAGAGATGCTGCGAGGAATTGCAAAGA
P.1         AGATGCGGAGAGAGCAGCCATCAGAGAGATGCTGCGAGCAATTGGAAAGA
P.2         AGATGCGGAGAGAGCAGCCATCGGAGAGATGCTGCGAGGAATTGGAAAGA
P.3         AAAGGCGGAGAGAACAGCCATCGGAAAGATGCTGCGAGGAATTGCAAAGG
            *    * * **     ***  * *  *** *   **** *   * **
```

Figure 16. Nucleic acid sequence alignment and similarities of p900 with *Picea glauca*, and *Pinus strobus* albumins 1, 2, and 3 designated by picea, P.1, P.2, and P.3, respectively. * show conserved nucleotides among all 5 albumins.

```
p900        GTG-CTGGACAGAGCGGCAGACCTTCCGAATACCTGCGACGTCCGCCAGC
picea       ATGTCTCCACAATGCCGATGCCAAGCCATACAGCAAATGC-TCGATCAAT
P.1         ATGTCTCCACAATGCCGATGCCGAGCCATACAGCAAGTAC-TTGATCAAT
P.2         ATGTCTCCACAATGCCGATGCCGAGCCATACAGCAAGTAC-TCGATCAAT
P.3         ATGTCTCCACACTGCCGATGCCGAGCCATAGAACAAACAC-TAGATCAAT
             ** **  ***  ** *  * *   **  * * *     * *    **

p900        CGGAGCCATGCGA--CATTCA---ACGTCACTCTCGTGAGTCATTCATCA
picea       CTTTATCGTATGATTCCTTCATGGATTCTGATTCTCAGGAG---GATGCA
P.1         CTCAATCATATGA--------------CCTCTTCATGGATTCTGAGGCT
P.2         CTCATTCCTATGATTCAACCACGGAGGACCTCTCCATGGATTCTGACGCT
P.3         CTTTATCGTTTGATTCATCGACAGATTCCGATTCTCAGGA---TGGTGCA
            *     * *  **                          *         *

p900        ATACT-GATCAATTAATTCTTTGCCCTGCGTCGTCTTAGAATAACGCCGC
picea       CCACTTAATCAACGCCGCCGCCGCCGCGAAGGGCGCGGAAG---AGAGGA
P.1         GCACTTAATCAGCGCCGCCGCCGC---GAATCGCGTGGAAG---AGAAGA
P.2         GCACTTAATCAGCGCCGCCGCCGCCGCGAATCGCGTGGAAG---AGAGGA
P.3         CCACTTAATCAACGCCGCCGACGCCCTGAAGGGCGCGGAAGGGAAGAAGA
              ***  ****       *   **   *    *      *     *  *

p900        CATGCAAATTTCTTTCTTTC-AGTCATGTCATTTTTCTTGTTATTCCATG
picea       GG---AGGAGGCGATGGAGAGAG-CGGCATACCTTCCGAATACCTGCAAC
P.1         GGCGGAAGAAGCGGAGGAGAGAG-CCGCATACCTTCCAGAAACGTGCAAC
P.2         GG---AGGAGGCGGAGGAGAGAGTCCGCATACCTTCCAGAAATTTGCAAC
P.3         AGAAGAGGAGGCGGTGGAGAGAGCCGG-AGAGCTTCCCGATCGATGCAAC
                 *     *         ** *     *  ** *       * **

p900        GCATAGTGAGCGATGAATC-TTAAGATATAT-GTGATGTTTT-ACTGGTG
picea       GTTCGCGAGCCCCCCCGCCGCTGCGATATTCAACGCCACTCTCGCTATTC
P.1         ATCCGCCAGCCTCCCCGCCGATGCGACGTTCAGCGCCGTTCTCGGTATTT
P.2         GTCCGCCAGCCACCC-GCCGCTGCGATGTTCAGCGCCGTTCTCGCTATTT
P.3         GTCCGCGAGTCACCCCGCCGCTGCGATATTCGACGCCACTCTCGATATTC
                      *       *  *  **  *     *    * *   *  *

p900        GTGGGTGGACAGTTATTCCATGACGGGCAGC-ATCATGTAAGTG-AGTGA
picea       CATGACGGGCAGCAG-TTTTAAGTGAGCGACGAAGAAGAAAATATAGATA
P.1         CACATCGGGCAGCGGCTTTTAAGTGCGCTACGAACA-GAAAATAGAGATA
P.2         CACATCGGGCACCGGCTTTTAAGTGAGCAACGAACA-GAAAATACAGATA
P.3         CAT------TATCGGCT-----------GAC-----------------TA
                     *    *           *                    *

p900        TCATGATCATGATCATGAATAAGAGATATACATAAATAGTGTTTTAACTA
picea       CTGCATGCATG-CTATGTATGTCC-CTAAATAAGGGAGGCACT---ACCG
P.1         CTGTGTCTATG-TAAC-TAAATAA-AGAA----GGGAGTCACTTCAGCTG
P.2         AGGTGTCTATG-TAAC-TAAATAA-ACAA-CCAGGGAGTCACTTAAGCCG
P.3         CTACGT------CAATGTATAT-----AAATAAAGATGCTGTTCGCACTA
                       *   *        *               *    *
```

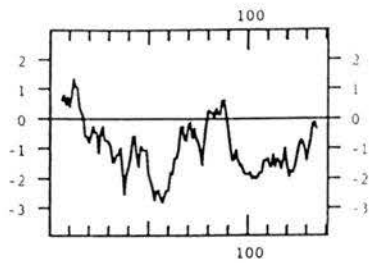
```
p900        A-ATAAAAAGGGAG-----GCACTTCAACCGCTATGAATTTGTGGTAATT
picea       CTATGTATTTTTGGTTT--CTGCTTTTATATATATAGCCTC---TCATTC
P.1         CTATGAACTTCTTGTTTTACTCTTTTTAGAAGAGTAATCTCATTTCATTT
P.2         CTATGAAATTGTAGTCTT--CCTTTTTGGAAGAGT-----TGTCTCATTT
P.3         CTAC----TTATG----TAGTGGTTTTGTAGGAG---------CTCATAT
              *                  **                    *

p900        A---CTTTATAGCCGGATCTCGCCCCCTCCATGTAACAAAGAATGGATGA
picea       AATGC-------------------------------------------
P.1         AATGCCTAACACTTTTATGGTAGCCGGCCGATGTGTTCCCTTTTATGTAA
P.2         AATGCCCGCCACTCAGATGGTAACC---CGATGTATTCCCTTTTATGTAA
P.3         AGTA---GGCTCTACAACA-TTGCCGGCTG------TCTCTAGCATGTAA
            *

p900        ACAAAGAATGGATGAACGA-TAAACTGCAA---GAGAAGAAGG
picea       TAAA-----------------------------------AAAA-
P.1         TAAAGGACTGAATGAACGACCAAATT-CATTTTGTTTCAAAA-
P.2         TAAAGTACTGAATCAACGAGCA---------------------
P.3         TAAACGAGTGAATGAATGAACAAATTGCATTATAAGTAAAAAA
              **

p900        AATATGGGTGAG-TTTAGGCGTGGCCTTCTCCTCCTCGTTCAGTGGAGCA
picea       C--------------------------------------------------
P.1         TTCATGAAAGAAATCTTCAACCTCTCAGCTTCAGCTGGATCAGATTATCT
P.2         CTTATG-----------------------------------------------
P.3         C--------------------------------------------------

p900        CCGCCAATGTTGATGCGGCTGAAGACAATATGTTTGGAGAGGACGTCGTG
picea       -------------------------------CAATGGGTGTCTTTTCCC
P.1         CTCCATCTTTCACCGTAACGAAGAAAGAAAAGCAATGGATGTTTTCTGCC
P.2         ------------------------------------GGTGTCTTCT-CC
P.3         -------------------------------CAATG------------

p900        CAACAACGACGACGGTTGGC---CAACCCGCAGAGACTATCCTCTTGCCG
picea       CTTCGACGACGA-GGCTGACGCTCAAATGGTTCAGTTTATCCGTCGCCCT
P.1         CTCCGACTACGA-GACTGGCTCTTAAATGGGCGAGCCTAGGCGTACGCCT
P.2         C------------------CTCTTAAATGGGTGAGCTTAGGTGTAGCCCT
P.3         --TCGA---CGA-GGTTGACGCTAAAATGGGTGACTTTAGTGGCGGCCCT
                             *    **   *   *   **         **

p900        GGACTACTTGGAGCAGCGAAG-AGACCAACCATCGCAAAGAATGTCTCCG
picea       GTTCCTCCTCTTTCACTGGGGTATTCCCAGTGTTGATGGCCATGAAGACA
P.1         TCTCGTCCTCCTTCAGTGGGGCACCCCAACTGTTGATGCGCGTGAAGGCG
P.2         TCTCCTCCTCCTTCAGTGCGCCACCCCAACTGTCGATGCGCGTGAAGGCG
P.3         TCTGTTTGTCATTCACTGTAGCACCCCCACTGTTGGTGCACATGAAGACA
                    *    **  *    *  ** *   * *       **    *
```

```
p900        CGATGCCGATGCCCAGCCATACAGCAAACGGTGGATCAATCTTATCATTC
picea       ATATGTATG-GAGAGGAGATACAACAA-CAAAGA--CGGTCGTGCGACCC
P.1         TCATGTATG-GTGAGGGCCTGCAACAA-CAACGA--CGCTCCTGCGATCC
P.2         TCATGTATG-GAGAGGGCCTGCAACAA-CAACGA--CGCTCCTGCGATCC
P.3         CCATGGATG-GAGAGGCGCTGCAACAA-CAACGA--CGGTCGTGCGA---
              ***     *    *   * ** *** *   *   *  ** *   *

p900        ATGGATTCTGATTCTCAAGAGGATGCACCACTTATCAACGCCGCCGCCGC
picea       TCAGAGA--GACCCGCA-GAGATTGT-CTTCATGCCGGGACTACT--TGG
P.1         --------------GCA-GCGACTGT-CTGAGTGCCGGGAGTACA--TGG
P.2         --------------GCA-GCGACTAT-CCCAGTGCCGGGACTACA--TGG
P.3         -----------CCCGCA-GAGATTGT-CTGATTGCCACGACTACT--TGC
                         ** * *  *   *    * *       *    *

p900        AGCCGCCGCGAAGGGCGCCGACGGGAAGAGG--ACGAGGA---GGAAGAG
picea       AGCGCCGGAGAGAGCAGCCATCGGAGAGATGCTGCGAGGAATTGCAAAGA
P.1         AGATGCGGAGAGAGCAGCCATCAGAGAGATGCTGCGAGCAATTGGAAAGA
P.2         AGATGCGGAGAGAGCAGCCATCGGAGAGATGCTGCGAGGAATTGGAAAGA
P.3         AAAGGCGGAGAGAACAGCCATCGGAAAGATGCTGCGAGGAATTGCAAAGG
            *    * * **     ***  * *  *** *   **** *   * **

p900        GTG-CTGGACAGAGCGGCAGACCTTCCGAATACCTGCGACGTCCGCCAGC
picea       ATGTCTCCACAATGCCGATGCCAAGCCATACAGCAAATGC-TCGATCAAT
P.1         ATGTCTCCACAATGCCGATGCCGAGCCATACAGCAAGTAC-TTGATCAAT
P.2         ATGTCTCCACAATGCCGATGCCGAGCCATACAGCAAGTAC-TCGATCAAT
P.3         ATGTCTCCACACTGCCGATGCCGAGCCATAGAACAAACAC-TAGATCAAT
             ** **  ***  ** *  * *   **  * * *     * *    **

p900        CGGAGCCATGCGA--CATTCA---ACGTCACTCTCGTGAGTCATTCATCA
picea       CTTTATCGTATGATTCCTTCATGGATTCTGATTCTCAGGAG---GATGCA
P.1         CTCAATCATATGA--------------CCTCTTCATGGATTCTGAGGCT
P.2         CTCATTCCTATGATTCAACCACGGAGGACCTCTCCATGGATTCTGACGCT
P.3         CTTTATCGTTTGATTCATCGACAGATTCCGATTCTCAGGA---TGGTGCA
            *     * *  **                          *        *

p900        ATACT-GATCAATTAATTCTTTGCCCTGCGTCGTCTTAGAATAACGCCGC
picea       CCACTTAATCAACGCCGCCGCCGCCGCGAAGGGCGCGGAAG---AGAGGA
P.1         GCACTTAATCAGCGCCGCCGCCGC---GAATCGCGTGGAAG---AGAAGA
P.2         GCACTTAATCAGCGCCGCCGCCGCCGCGAATCGCGTGGAAG---AGAGGA
P.3         CCACTTAATCAACGCCGCCGACGCCCTGAAGGGCGCGGAAGGGAAGAAGA
              ***  ****       *   **   *    *      *     *  *

p900        CATGCAAATTTCTTTCTTTC-AGTCATGTCATTTTTCTTGTTATTCCATG
picea       GG---AGGAGGCGATGGAGAGAG-CGGCATACCTTCCGAATACCTGCAAC
P.1         GGCGGAAGAAGCGGAGGAGAGAG-CCGCATACCTTCCAGAAACGTGCAAC
P.2         GG---AGGAGGCGGAGGAGAGAGTCCGCATACCTTCCAGAAATTTGCAAC
P.3         AGAAGAGGAGGCGGTGGAGAGAGCCGG-AGAGCTTCCCGATCGATGCAAC
                 *     *         ** *     *  ** *       * **
```

```
p900        GCATAGTGAGCGATGAATC-TTAAGATATAT-GTGATGTTTT-ACTGGTG
picea       GTTCGCGAGCCCCCCCGCCGCTGCGATATTCAACGCCACTCTCGCTATTC
P.1         ATCCGCCAGCCTCCCCGCCGATGCGACGTTCAGCGCCGTTCTCGGTATTT
P.2         GTCCGCCAGCCACCC-GCCGCTGCGATGTTCAGCGCCGTTCTCGCTATTT
P.3         GTCCGCGAGTCACCCCGCCGCTGCGATATTCGACGCCACTCTCGATATTC
                      *       *  *  **  *     *    * *   *  *

p900        GTGGGTGGACAGTTATTCCATGACGGGCAGC-ATCATGTAAGTG-AGTGA
picea       CATGACGGGCAGCAG-TTTTAAGTGAGCGACGAAGAAGAAAATATAGATA
P.1         CACATCGGGCAGCGGCTTTTAAGTGCGCTACGAACA-GAAAATAGAGATA
P.2         CACATCGGGCACCGGCTTTTAAGTGAGCAACGAACA-GAAAATACAGATA
P.3         CAT------TATCGGCT-----------GAC-----------------TA
                     *     *             *                     *

p900        TCATGATCATGATCATGAATAAGAGATATACATAAATAGTGTTTTAACTA
picea       CTGCATGCATG-CTATGTATGTCC-CTAAATAAGGGAGGCACT---ACCG
P.1         CTGTGTCTATG-TAAC-TAAATAA-AGAA----GGGAGTCACTTCAGCTG
P.2         AGGTGTCTATG-TAAC-TAAATAA-ACAA-CCAGGGAGTCACTTAAGCCG
P.3         CTACGT------CAATGTATAT-----AAATAAAGATGCTGTTCGCACTA
                        *   *         *                *    *

p900        A-ATAAAAAGGGAG-----GCACTTCAACCGCTATGAATTTGTGGTAATT
picea       CTATGTATTTTTGGTTT--CTGCTTTTATATATATAGCCTC---TCATTC
P.1         CTATGAACTTCTTGTTTTACTCTTTTTAGAAGAGTAATCTCATTTCATTT
P.2         CTATGAAATTGTAGTCTT--CCTTTTTGGAAGAGT-----TGTCTCATTT
P.3         CTAC----TTATG----TAGTGGTTTTGTAGGAG---------CTCATAT
              *                    **                    *

p900        A---CTTTATAGCCGGATCTCGCCCCCTCCATGTAACAAAGAATGGATGA
picea       AATGC-------------------------------------------
P.1         AATGCCTAACACTTTTATGGTAGCCGGCCGATGTGTTCCCTTTTATGTAA
P.2         AATGCCCGCCACTCAGATGGTAACC---CGATGTATTCCCTTTTATGTAA
P.3         AGTA---GGCTCTACAACA-TTGCCGGCTG------TCTCTAGCATGTAA
            *

p900        ACAAAGAATGGATGAACGA-TAAACTGCAA---GAGAAGAAGG
picea       TAAA-----------------------------------AAAA-
P.1         TAAAGGACTGAATGAACGACCAAATT-CATTTTGTTTCAAAA-
P.2         TAAAGTACTGAATCAACGAGCA---------------------
P.3         TAAACGAGTGAATGAATGAACAAATTGCATTATAAGTAAAAAA
              **
```

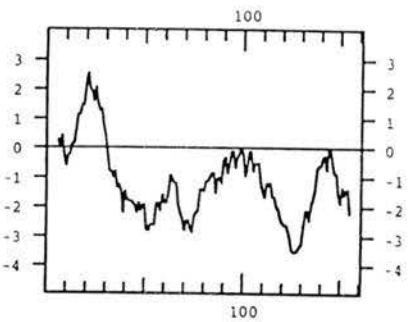
the p900 orf sequence -> KD Hydrophobicity <1/11>

Protein sequence 140 a.a. MGEPPRGLLLLD ... AGAMRHSTRLST



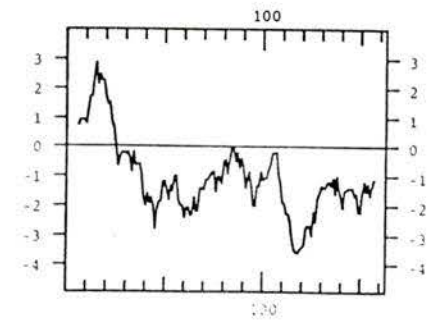
picea glauca sequ -> KD Hydrophobicity <1/11>

Protein sequence 160 a.a. MGVFSPSTTRLT ... VREPPRRCD.



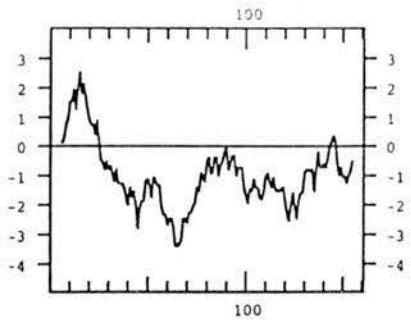
pinus strobus alb 2 -> KD Hydrophobicity <1/11>

Protein sequence 162 a.a. MGVFSPLKWVSL ... RRSRYFT



pinus strobus 3 -> KD Hydrophobicity <1/11>

Protein sequence 160 a.a. MSTFLTLKWVTL ... DIRPF



pinus strobus alb 1 -> KD Hydrophobicity <1/11>

Protein sequence 190 a.a. MKEIFNLSASAG ... RRSRYF

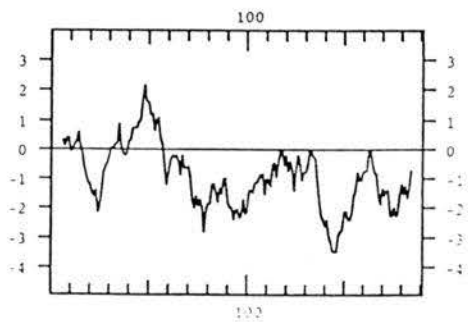


Figure 17. Kyte Doolittle hydropathy plot comparisons of a) p900 albumin, b) *Picea glauca* albumin, c) *Pinus strobus* albumin 1, d) *Pinus strobus* albumin 2, and e) *Pinus strobus* albumin 3.

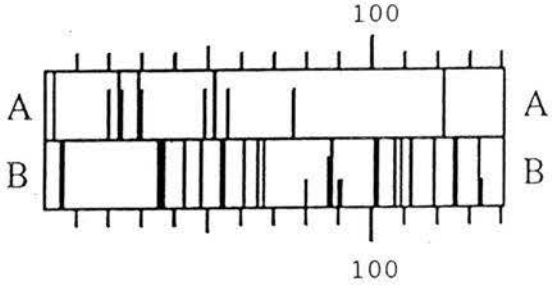


Figure 18. Graphic localization map of p900 acidic and basic residues. A: acidic B: basic

```
p900            MGE--------------FRRG-------------------------LL 50
picea           M-------------------------GVFSPSTTRLTLKWFSLSVALF
P.1             MKEIFNLSASAGSDYLSIFHRNEERKAMDVFCPPTTRLALKWASLGVRLL
P.2             M-------------------------GVFSP------LKWVSLGVALL
P.3             M-------------------------------STRLTLKWVTLVAALL
                *                                             *.

p900            LLVQWSTANVDAAEDNMFGEDVVQQRRR---LANPQRLSSCRDYLEQRRD 100
picea           LLFHWGIPSVDGHEDNMYGEEIQQQRRSCDPQRDPQRLSSCRDYLERRRE
P.1             VLLQWGTPTVDAREGVMYGEGLQQQRRSCDPQR----LSECREYMEMRRE
P.2             LLLQCATPTVDAREGVMYGEGLQQQRRSCDPQR----LSQCRDYMEMRRE
P.3             EVIHCSTPTVGAHEDTMDGEALQQQRRSCDPQR----LSQCHDYLQRRRE
                 . . . . *.. *. * ** . ****          ** *..*.. **.

p900            QPS-------QRMSPRCRCPAIQQTVDQSYHSWILILKRMHHLSTPPPQP 150
picea           QPSERCCEELQRMSPQCRCQAIQQMLDQS-LSYDSFMDSDSQE-DAPLNQ
P.1             QPSERCCEQLERMSPQCRCRAIQQVLDQS-QSYD-----LFMDSEAASNQ
P.2             QPSERCCEELERMSPQCRCRAIQQVLDQS-HSYDSTTEDLSMDSDAALNQ
P.3             QPSERCCEELQRMSPHCRCRAIEQTLDQS-LSFDSSTDSDSQDG-APLNQ
                ***       .****.*** **.* .***  *             .. .

p900            PRRAPTGRGRGGRGAGQSGR----PSEYLRRPPA---GAMRHS----TSL 200
picea           RRRRREGRGREE-EE-AMERAAYLPNTCNVREPPRRCDIQR---------
P.1             RRRR-ESRGREEAEE-AEERAAYLPETCNIRQPPRRCDVQRRSRYFTSGS
P.2             RRRRRESRGREE-EE-AEERVRIPSRNLQRPPATRRCDVQRRSRYFTSGT
P.3             RRAAPEGAGREEEEEEAVERAGELPDRCNVAESPAACDIRRHSRYSIIG-
                 *    . **         *    .       .    .  *

p900            SX
picea           --
P.1             GF
P.2             GP
P.3             --
```

Figure 19. Amino acid sequence comparisons between p900, *Picea glauca* albumin, and *Pinus strobus* albumins 1, 2, and 3

Discussion

The results indicate that a unique isoform of 2S albumin in Douglas fir has been isolated, shown to be developmentally regulated and fully characterized. Initial Southern hybridization work showed that under normal conditions, no hybridization was evident between the radiolabelled p900 cDNA insert and the other albumins, including the control albumin 4.7 (fig, 11). However, low stringency wash conditions showed that there indeed was some hybridization confirming the sequence data which concluded that p900 was an albumin. The elucidation of a unique isoform that initially eluded detection had previously been seen in *Arabidopsis thaliana* (Krebbers et al., 1988), where initial hybridization studies identified four isoforms. However, protein separation through chromatography later resulted in the discovery of a fifth isoform which eluded Southern hybridization detection because of the divergence of several amino acids (Van der Klei et al., 1993).

In terms of developmental regulation, Northern hybridizations clearly illustrate control of p900 albumin at the transcriptional level. In the megagametophyte, no transcripts are evident up to 10 DAF. Transcript levels peak between 35 and 45 DAF, decline by 59 DAF and then become undetectable thereafter including the dry seed stage (fig. 14). Interestingly, the transcriptional regulation differs temporally between the zygotic embryo and the megagametophyte itself. Transcripts were undetectable up to 35 DAF, peak at 45 DAF and then decline accordingly whereas, by 73 DAF and thereafter, the transcripts were completely undetectable. This pattern of regulation

indicates that the 2S albumin mRNA is translated during this period of time for storage in the protein bodies in the dry seed (Chrispeels et al., 1982). The maternal tissue begins the initial production of the albumin because it is mostly a storage organ responsible as the nutritional source for the developing embryo (Lindsey et al., 1993). In dry seed stage in both tissues, the translated product is stored for use as a nitrogen reserve for the germinating seed (Higgins et. al, 1984).

It is known that many albumins are comprised of large multigene families with several isoforms present (Hilde van der Klei et al., 1993). It appears that p900 is indeed an unique isoform of albumin in Douglas fir and, therefore a member of a multigene family. With several other albumins previously characterized a comparison of the developmental regulation patterns (data not shown) shows small variations in the transcript accumulation. The p900 transcripts are at a peak at 45 DAF in the megagametophyte while other albumins show a decline from 31 DAF. The p900 transcripts are non-existent at 73 DAF, whereas other albumins show transcripts not only present at 73 DAF, but some are even expressed at the dry seed stage. In the embryo, p900 transcripts are evident at 31 DAF (fig. 14) while most others do not have transcriptional activation until 45 DAF. This comparison shows that p900 is a unique member of the albumin multigene family in Douglas fir where each member is independently regulated due to different but unknown cis-acting regions. Transcript size is also different comprising 836 nt, while most other albumins in Douglas fir are

approximately 700 nt. However, the members do share similar features including high expression levels during embryogenesis and accumulation and decay of their transcripts in a precise developmental timetable.

Upon sequence characterization, the p900 albumin cDNA was found to be 836 bp in length. The ATG initiation codon was found at position 33, The termination codon of TGA, giving an opal termination signal, was found at position 450 (fig. 15). This results in a 420 bp open reading frame that when translated, gives a 140 aa, 15.8 kDa preprotein. Additionally, the polyadenylation signal AATAAA is found to be located at position 707 (fig. 15).

Beginning at the N-terminus of the predicted precursor polypeptide, there is a hydrophobic region with the characteristics of a typical signal peptide indicating that the protein is destined for secretion. Utilizing the theorem for determining the cleavage site between the signal peptide and the remaining protein (Von Heijne et al., 1983), it is likely that cleavage occurs between Ala-21 and Ala-22. The evidence of this predicted site is further strengthened by comparison with data from many angiosperm albumin signal peptides. In general, this preprotein is cotranslationally transported into the endoplasmic reticulum with the simultaneous cleavage of the 21 aa signal peptide (Chrispeels et al., 1984). The 119 aa mature albumin then passes through the ER to the Golgi complex and from there to the protein bodies for storage (Chrispeels et al., 1982).

Scanning the sequence, it is evident that no Asn-x-Ser/Thr motif is present, thereby indicating that there is no N-linked

glycosylation. This agrees with the generalization that the 2S albumins and 11S legumins are unglycosylated, whereas other storage proteins including phaseolin and vicilin 7S are post-translationally glycosylated (Chrispeels et al., 1982). To date, there is no satisfactory explanation for the physiological function of this step.

When the amino acid composition of p900 albumin is analyzed and compared to other albumins, some striking similarities as well as differences are seen. The major amino acid represented in p900 albumin is arginine at 15.7% as is the case with *Pinus pinaster*, another gymnosperm (Table 2). This is also comparable to that of *Pinus strobus* albumins 1 and 2 at 14.74% and 14.20%, respectively. However, most abundant in these two species is glutamine and glutamate (Glx) at 16.32% and 18.5%, respectively, whereas in p900 albumin Glx is the second most abundant at 12.2% (Table 2). Interestingly, in angiosperms Glx is by far the most abundant; for example, castor has 30% Glx and only 9.4% Arg. In the gymnosperms including p900, however, the distribution is much more even among the nitrogen-laden amino acids Glx and Arg (Table 2). The high levels of the nitrogen-rich Glx, Asx, and Arg in p900 albumin is consistent with the fact that albumin is indeed a storage protein providing a major nitrogen supply upon its hydrolysis to the growing seedling after germination.

Another interesting finding is the high level of cysteine in the p900 albumin. At 2.1% cysteine, this level is comparable to the high percentage of cysteine found in all albumins (Youle et al., 1981) and is found to be similar to cysteine levels in other gymnosperms. For

example, *Pinus strobus* albumins 1, 2, and 3 have cysteine levels at 4.21%, 4.93%, and 5.63%, respectively. This characteristic distinguishes albumins from globulins, where in the latter group of storage proteins, cysteine is generally under-represented; for example, the legumin-like protein in Douglas fir has a cysteine level of 0.8% (Misra et al., 1993). It is thought that with these consistently high levels of cysteine found in all albumins, these storage proteins may fulfill the role of a sulfur reserve in the seed as well as a nitrogen reserve (Youle et al., 1981). In general, amino acid compositions of angiosperm and gymnosperm albumin storage proteins appear to be very similar.

Upon multiple sequence alignment of the cDNA nucleotide sequence with several other previously characterized conifer albumins, it was found that 408 nucleotides were conserved between the p900 albumin and *Picea glauca* albumin, *Pinus strobus* albumins 1, 2, and 3 (fig. 16). This gives an overall 48.8% identity between them indicating that albumin in gymnosperms has a fair degree of conservation. Conversely, when comparisons are made to angiosperms the degree of conservation drops dramatically indicating that there is indeed an evolutionary link but that there has been a divergence between angiosperm albumin and gymnosperm albumin. If one focuses on the amino acid sequence, however, certain amino acid motifs are found to be universally conserved. Conservation is evident in the positioning of many cysteine residues. For example, between the gymnosperms, the cysteines at positions 91, 117, and 119 are absolutely conserved (fig.

18). These cysteines are also found in all angiosperm albumins and are thus the signature residues of albumin (Kortt et al., 1991). The p900 albumin does indeed have these signature residues, however, interestingly does diverge from the other gymnosperms by two clear motif deletions. The cysteine at position 79 is found in nearly all gymnosperm and angiosperm albumins but is absent from p900. The two cysteines at positions 106 and 107 are found in nearly all albumins as well but are clearly absent from p900. In fact, the amino acid motif encompassing the two cysteines of ERCCEEL found in *Picea glauca* and *Pinus strobus* 1, 2, and 3 (with *Pinus strobus* 1 slight divergence of ERCCEQL) is completely absent from p900 (fig. 18). With the position of the cysteine residues relative to each other appearing to be generally conserved amongst all albumins, it has been suggested that tertiary structure constraints imposed by disulfide bridges dominate sequence conservation. Since most of the cysteine residues of albumins are involved in disulfide bridge formation (Lonnerdahl et al., 1972), their conserved relative positions suggest that the disulfide bridges constitute a major structural element of the 2S storage proteins. As for most secreted protein, (e.g. most serum proteins), cysteine residues are known to enhance stability of proteins by interconnecting the protein regions by disulfide bonds. Similarily, the function of such bridges in storage proteins might be primarily to form a compact rigid structure, able to remain intact for a long time under severe condition such as the desiccation of the Douglas fir embryo (dry

seed), and probably only secondarily to serve as a sulfur source (Ampe et al., 1986).

When the hydropathy plot of p900 is compared to those of *Picea glauca*, and *Pinus strobus* albumins 1, 2, and 3, some common features are apparent. The initial peak illustrating hydrophobicity in the first 20 amino acids of p900 clearly illustrates the signal peptide portion of the pre-protein. This peak is clearly seen among the other gymnosperm albumins (fig. 17). The generalization that albumins are water soluble is clearly seen by the majority of each protein, including p900, showing a hydrophilic amino acid composition. A characteristic region of decreased hydrophilicity of approximately 30 amino acids localized around amino acid 100 of the pre-protein, evident in each species including p900, illustrates the folding of the protein with this region towards the interior and implies a tertiary structure conservation (fig 17).

Generally, it has been shown that the angiosperm 2S proteins have a dimeric structure formed by 2 subunits of 7-12 kDa and 3-7 kDa, linked by disulfide bridges (Menendez-Arias et al., 1988). With the exception of sunflower (Kortt et al., 1991), extensive post-translational proteolytic processing of the protein is undertaken. p900 does indeed have the key cysteines for the potential disulfide bridge formation but without N-terminal amino acid sequencing of the purified peptides separated by HPLC, it is difficult to assess whether or not p900 is a single polypeptide chain or composed of disulfide-bonded peptides.

Conclusions

The unique p900 albumin transcript mRNA is transcriptionally regulated in the embryo and megagametyophyte during Douglas fir embryogenesis. The p900 protein isoform is predicted to lack N-linked glycosylation, have a signal sequence for secretion, be hydrophilic, and to be a basic protein. A 30.5% overall amino acid conservation is seen between p900 and other gymnosperm albumins. The protein is high in N-rich amino acids including arginine and glutamine/glutamate at 15.7% and 12.2%, respectively, suggesting the role of albumin as a nitrogen source by its hydroysis upon germination. The p900 albumin has additionally a high (2.1%) cysteine content as well as showing a strict conservation in the positioning of these cysteines for the purpose of a sulfur reserve and for stability of the protein during dehydration fom disulfide bridges. However, unique features of isoform include the clear deletion of two regions where cysteines are typically conserved.

Future work could include isolation of the p900 albumin protein product followed by HPLC and N-terminal peptide sequencing. This data would confirm the predicted amino acid sequence and give further insight into the possibility of large and small disulfide-linked peptides and thus, the proteolytic processing of the Douglas fir albumin. Other work could include use of the p900 cDNA to probe a genomic library in order to obtain a genomic clone which could further elucidate

the 5' upstream regulatory regions. To date, genomic sequence data corresponding to the seed storage proteins of conifers have not been reported. Further work could reveal more about the structure, evolution, and regulation of these genes in conifers.

Literature Cited

Adachi, T., Hidehiko, Izumi, Yamada, T., Tanaka, K., Takeuchi, S., Nakamura, R., and Matsuda, T. (1993) Gene structure and expression of rice seed allergenic proteins belonging to the alpha-amylase/trypsin inhibitor family. Plant Mol. Biol. 21, 239-248.

Aguirre, P.J., and Smith, A.G. (1993) Molecular characterization of a gene encoding a cysteine-rich protein preferentially expressed in anthers of *Lycopersicon esculentum*. Plant Mol. Biol. 23, 477-487.

Allen, R.D., Cohen, E.A., Vonder haar, R.A., Adams, C.A., Ma, D.P., Nessler, C.L., and Thomas, T.L. (1987) Sequence and expression of a gene encoding an albumin storage protein in sunflower. Mol. Gen. Genet. 210, 211-218.

Altenbach, S.B., Pearson, K.W., Leung, F.W., Sun, S.S.M. (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant Mol. Biol. 8, 239-250.

Altshul, A.M., Yatsu, L.Y., Ory, R.L., and Engleman, E.M. (1966) Seed proteins. Annu. Rev. Plant Physiol. 17, 113-136.

Ampe, C., Van Damme, J., deCastro, L.A.B., Sammpio, M.J.A.M., Van Montagu, M., ande Vandekerckhove, J. (1986) The amino-acid sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). J. Biochem. 159, 597-604.

Badenoch-Jones, J., Spencer, D., Higgins, T.J.V., and Millerd, A. (1981) The role of glycosylation in storage-protein synthesis in developing pea seeds. Planta 152, 201-209.

Barratt, D.H.P., Whitford, P.N., Cook, S.K., Butcher, G., and Wang TL. (1989) Analysis of seed development in *Pisum sativum* L. VIII. Does abscisic acid prevent precocious germination and control storage protein synthesis? J. Exp. Bot. 40, 1009-1014.

Bartels, D., and Thompson, R.D. (1983) The characterization of cDNA clones coding for wheat storage proteins. Nucl. Acids Res. 11, 2961-2977.

Bartels, D., Engelhardt, K., Roncarati, R., Schneider, K., Rotter, M., and Salamini, F. (1991) An ABA and GA modulated gene expressed in the barley embryo encodes an aldose reductase related protein. EMBO J. 10 (5), 1037-1043.

Basha, S.M.M., and Beevers, L. (1975) The development of proteolytic activity and protein degradation during germination. Planta 124, 77-87.

Baumgartner, B., Tokuyasu, K.T., and Chrispeels, M.J. (1980) Immunonocytochemical localization of reserve protein in the endoplasmic reticulum of developing bean (*Phaseolus vulgaris*) cotyledons. Planta 150, 419-425.

Benner, M.S., Philliips, R.L., Kirihar, J.A., and Messing, J.W. (1989) Genetic analysis of methionin-rich storage protein accumulation in maize. Theor. Appl. Genet. 78, 761-767.

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

Blundy, K.S., Blundy, A.A.C., and Crouch, M.L. (1991) Differential expression of members of the napin storage protein gene family during embryogenesis in *Brassica napus*. Plant Mol. Biol. 17, 1099-1104.

Bollini, R., Van der Wilden, W., and Chrispeels, M.J. (1982) A precursor of the reserve-protein, phaseolin, is transiently associated with the endoplasmic reticulum of developing *Phaseolus vulgaris* cotyledons. Physiol. Plant. 55, 82-92.

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

Bostan, R.S., Kodrzycki, R., and Larkins, B.A. (1986) Transcriptional regulation of maize genes. In *Molecular Biology of Seed Storage Proteins and Lectins*. Ninth symposium in Plant Physiology, eds. Sharon, L.M. and Chrispeels, M.J. American society of plant physiologists, Rockville, MP, 117

Boulter, D., and Debyshire, E. (1971) Taxonomic aspects of the structure of legume proteins. Chemotaxonomy of the *Leguminosae*. Academic Press, New York, 285-308.

Brohult, S., and Sandegren, E. (1954) Seed proteins. The Proteins. Academic Press, New York. 285-308.

Brown, J.W.S., Ma, Y., Bliss, F.A., and Hall, T.C. (1981) Genetic variation in the subunits of globulin-1 storage protein of French bean. Theor. Appl. Genet. 59, 83-88.

Burr, B., Burr, F.A., St. John, T.P., Thomas, M., and Davis, R.W. (1982) Zein storage protein gene family of maize: an assessment of heterogeneity with clones and messenger RNA sequences. J. Mol. Biol. 154, 33-49.

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

Casy, R., Domoney, C., and Ellis, N. (1986) Legume storage proteins and their genes. Oxford Surv. Plant Mol. Cell Biol. 3, 2-95.

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

Casey, R., March, J.F., and Sanger, E. (1981) N-terminal amino acid sequence of Beta-subunits of legumin from *Pisum sativum*. Phytochem. 20, 161-163.

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

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

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

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

Christen, A.A., and Montalbano, B. (1989) An ethidium bromide-agarose plate assay for the nonradioactive detection of cDNA synthesis. Anal. Biochem. 178, 269-272.

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

Corke, R.M.K., Hedley, C.L., and Wang, T.L. (1990) An analysis of seed development in *Pisum sativum*. XI. Cellular development and the deposition of storage protein in immature embryos grown *in vivo* and *in vitro*. Protoplasma 155, 127-135.

Coulter, K.M., and Bewley, D. (1990) Characterization of a small sulpher-rich storage albumin in seeds of alfalfa (*Medicago sativa* L.). J. Exp. Bot. 41 (233), 1541-1547.

Crouch, M.L. (1982) Non-zygotic embryos of *Brassica napus* contain embryo-specific storage proteins. Planta 156, 520-524.

Crouch, M.L., Tenbarge, K.M., Simon, A.E., and Ferl, R. (1983) cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cLEAved from a precursor polypeptide. J. Mol. App. Genet. 2, 273-283.

Crouch, M.L., and Sussex, I.M. (1981) Development and storage protein synthesis in *Brassica napus* L. embryos *in vivo* and *in vitro*. Planta 153, 64-74.

Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N., and Boulter, D. (1982) Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295, 76-78.

Cunningham, B.A., Wang, J.L., Waxmdal, M.J., and Edelman, G.M. (1975) The covalent and three-dimensional structure of concanavalin A.11 Amino acid sequence of cyanogen bromide fragment F_3. J. Biol. Chem. 250, 1503-1512.

Cyert, M.S., and Thorner, J. (1989) Putting it on and taking it off: phosphoprotein phosphatase involvement in cell cycle regulation. Cell 57, 891-893.

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

Davey, R.A., and Dudman, W.F. (1979) The carbohydrate of storage glycoproteins from seeds of *Pisum sativum*: characterization and distribution on component polypeptides. Aust. J. Plant Physiol. 6, 435-437.

Debyshire, E., Wright, D.J., and Boulter, D. (1976) Legumin and vicilin storage proteins of legumin seeds. Phytochem. 15, 3-24.

deCastro, L.A.B., Lacerada, Z., Aramay, R.A., Sampaio, MJ.A.M., and Gander, E.S. (1987) Evidence for a precursor molecule of Brazil nut 2S seed proteins from biosynthesis and cDNA analysis. Mol. Gen. Genet. 206, 338-343.

Dure, L. (1985) Embryogenesis and gene expression during seed formation. Oxford Surv. Plant Mol. Cell Biol. 2, 179-197.

Dure, L.S. III, and Chlan, C. (1981) Developmental biochemistry of cottonseed embryogenesis and germination. XII. Purification and properties of principal storage proteins. Plant Physiol. 68, 180-186.

Dure III, L.S. , and Galau, G.A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination, XIII. Regulation of biosynthesis of principal storage proteins. Plant Physiol. 68, 187-191.

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

Ericson, M.L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L.-G., and Rask, L. (1986) Structure of the rapeseed 1.7S storage protein, napin, and its precursor. J. Biol. Chem. 261 (31), 14576-14581.

Fernandez, D.E., Turner, F.R., and Crouch, M.L. (1991) *In situ* localization of storage protein mRNAs in developing meristems of *Brassica napus* embryos. Development 111, 299-313.

Finkelstein, R.R., Tenbarge, K.M., Schemway, J.E., Crouch, M.L. (1985) Role of ABA in maturation of rapeseed embryos. Plant Physiol. 78, 630.

Finkelstein, R.R., and Crouch, M.L. (1986) Rapeseed embryo development in culture on high osmoticum is similar to that in seeds. Plant Physiol. 81, 907-912.

Flinn, B.S., Roberts, D.R., and Taylor, I.E.P. (1991) Evaluation of somatic embryos of interior spruce. Characterization and developmental regulation of storage proteins. Physiol. Plant. 82, 624-632).

Foard, D.E., Gutay, P.A., Ladin, B., Beachy, R.N., and Larkins, B.A. (1982) *In vitro* synthesis of the Bowman-Birk and related soybean protease inhibitors. Plant Mol. Biol. 1, 227-243.

Foriers, A., Lebrun, E., Van Rapenbusch, R., de Neve, R., and Strosberg, A.D. (1981) The structure of the lentil (*Lens culinaris*) lectin. Amino acid sequence determination and prediction of the secondary structure. J. Biol. Chem. 256, 5550-5560.

Galau, G.A., Hughes, D.W., Dure III, L. (1986) Abscisic acid induction of cloned cotton late embryogenesis abundant (LEA) mRNAs. Plant Mol. Biol. 7, 155-170.

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

Gatehouse, J.A., Evans, I.M., Croy, R.R.D., and Boulter, D. (1986) Differential expression of genes during legume seed development. Philosophical Transaction of the Royal Society of London Series B 314, 367-384.

Gaunt, S.J., and Singh, P.B. (1990) Homeogene expression patterns and chromosomal imprinting. Trends Genet. 6, 208-212.

Gautier, M-F, Aleman, M-E, Guirao, A., Marion, D., and Joudrier, P. (1994) *Triticum aestivum* puroindolines, two basic cysteine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol. Biol. 25, 43-57.

Geraghty, D., Peifer, M.A., Rubenstein, I., and Messing, J. (1981) The primary structure of a plant storage protein: zein. Nucl. Acids Res. 9, 5163-5174.

Gilroy, J., Wright, D.J., Boulter, D. (1979) Homology of basic subunits of legumin from *Glycine max* and *Vicia faba*. Phytochem. 18, 315-316.

Goldberg, R.B. (1988) Plants: novel developmental processes. Science 240, 1460-1467.

Goldberg, R.B., Barker, S.J., and Perez-Grau, L.(1989) Regulation of gene expression during plant embryogenesis. Cell 56, 149-160.

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

Hagen, G., Rubenstein, I. (1981) Complex organization of zein genes in maize. Gene 13, 239-249.

Hager, K.-P., Jensen, U., Gilroy, J., and Richardson, M. (1992) The N-terminal amino acid sequence of the Beta-subunit of the legumin-like protein from seeds of *Gingko biloba*. Phytochem. 31 (2), 523-525.

Hara, I., Ohmiya, M., and Matsubara, H. (1978) Pumpkin (*Cucurbita* sp.) seed globulin. III. Comparison of subunit structures among seed

globulins of various *Cucurbita* species and characterization of peptide components. Plant Cell Physiol. 19, 237-243.

Harris, N., Mulcrone, J., and Grindley, H. (1990) Tissue preparation techniques for *in situ* hybridisation studies of storage protein gene expresion during pea seed development. In *In situ Hybridisation: Applications in Developmental Biology and Medicine,* eds. Harris, N. and Wilkinson, D.J. Cambridge University Press.

Hill, J.E., and Breidenbach, R.W. (1974) Protein of soybean seeds. I. Isolation and characterization of the major components. Plant Physiol. 53, 742-746.

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

Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C., and Spencer, D. (1983) The biosynthesis and primary structure of pea seed lectin. J. Biol. Chem. 258, 949-954.

Hoffman, I.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. Nucl. Acids Res. 10, 7819-7828.

Hong, B., Uknes, S.J., and Ho, T.-H.D. (1988) Cloning and characterization of a cDNA encoding a mRNA rapidly induced by ABA in barley aleurone layers. Plant Mol. Biol. 11, 495-506.

Hopp, T.P., Hemperly, J.J., and Cunningham, B.A. (1982) Amino acid sequence and variant forms of favin, a lectin from *Vicia faba.* J. Biol. Chem. 257, 4473-4483.

Hughes, D.W., and Galau, G.A. (1991) Developmental and environmental induction of *LEA* and *LEA* mRNAs and the post-abscission program during embryo culture. Plant Cell 3, 605-618.

Hurkman, W.J., and Beevers, L. (1980) What is pea legumin-is it glycosylated? Planta 150, 82-88.

Irwin, S.D., and Lord, J.M. (1990) Nucleotide sequence of a *Ricinus communis* 2S albumin precursor gene. Nucl. Acids Res. 18 (19), 5890.

Irwin, S.D., Keen, J.N., Findlay, J.B.C., and Lord, J.M. (1990) The *Ricinus communis* 2S albumin precursor: A single preproprotein may be processed into two different heterodimeric storage proteins. Mol. Gen. Genet. 222, 400-408.

Jessell, T.M., and Melton, D.A., (1992) Diffusible factors in vertebrate embryonic induction. Cell 68, 257-270

Josefsson, L-G., Lanman, M., Ericson, M.L., and Rask, L. (1987) Stucture of a gene encoding the 1.7 S storage protein, Napin, from *Brassica napus*. J. Biol. Chem. 262 (25), 12196-12201.

Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and Vandekerckhove, J. (1988) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol. 87, 859-866.

Kreil, G. (1981) Transfer of proteins across membranes. Ann. Rev. Biochem. 50, 317-348.

Kreis, M., Shewry, P.R., Forde, B.G., Forde, J., and Miflin, B.J. (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surveys of Plant Molecular and Cellular Biology. 2, 253-317.

Kreis, M., Williamson, M.S., Forde, J., Scmutz, D., Clark, J., Buxton, B., Pywell, J., Marris, C., Henderson, J., Harris, N., Shewry, P.R., Forde, B.G., and Miflin, B.J. (1986) Differential gene expression in the developing barley endosperm. Philosophical Transactions of the Royal Society of London Series B. 314, 355-365.

Krochko, J.E., and Bewley, J.D. (1988) Use of electrophoretic techniques in determining the composition of seed storage proteins in alfalfa. Electrophoresis 9, 751-763.

Krochko, J.E. (1990) Identification and characterization of the seed storage proteins from alfalfa (*Medicago sativa*). J. Exp. Bot. 41, 505-514.

Kropf, D.L., and Quatrano, R.S. (1987) Localization of membrane-associated calcium during development of fucoid algae using chlorotetracycline. Planta. 171, 158-170.

Kortt, A.A., Caldwell, B., Lilley, G.G., and Higgins, T.J.V. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.) Eur. J. Biochem. 195, 329-334.

Li, Y, Hagen, G., and Guilfoyle, T.J. (1991) An auxin-responsive promoter is differentially induced by auxin gradients during tropisms. Plant Cell 3, 1167-1175.

Lindsey, K., and Topping, J.F. (1993) Embryogenesis: a question of pattern. J. Exp. Bot. 44 (259), 359-374.

Lonnerdal, B., and Janson, J.C. (1972) Studies on Brassica seed proteins. 1 The low molecular weight proteins in rapeseed. Isolation and characterzation. Biochim. Biophys. Acta. 278, 175.

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

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor: New York.

Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987) Regulation of inducible and tissue-specific gene expression. Science 236, 1237-1244.

Manickam, A., and Carlier, A.R. (1980) Isolation and function of a low molecular weight protein of mung bean embryonic axes. Planta 149, 234-240.

Marks, M.D., Larkins, B.A. (1982) Analysis of sequence microheterogeneity among zein messenger RNAs. J. Biol. Chem. 257, 9976-9983.

Mayer, U., Torres Ruiz, R.A., Berleth, T., Misera, S., Jurgens, G. (1991) Mutations affecting body organization in the *Arabidopsis* embryo. Nature 353, 402-407.

Meinke, D.W., Chen, J., and Beachy, R.N. (1981) Expression of storage protein genes during soybean seed development. Planta 153, 130-139.

Meinke, D.W., (1986) Embryo-lethal mutants and the study of plant embryo development. Oxford Surv. Plant Mol. Cell Biol. 3, 122-165.

Menendez-Arias, L., Moneo, I., Dominguez, J., and Rodriguez, R. (1988) Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, Sin a I. Eur. J. Biochem. 177, 159-166.

Millerd, A. (1975) Biochemistry of legume seed proteins. Annu. Rev. Plant Physiol. 26, 53-72.

Misra, S., and Green, M.J. (1990) Developmental gene expression in conifer embryogenesis and germination I. Seed proteins and protein body composition of mature embryo and the megagametophyte of white spruce (*Picea glauca*). Plant Sci. 68, 163-173.

Misra, S., and Green, M.J. (1991) Developmental gene expression in conifer embryogenesis and germination. II. Crystalloid protein synthesis in the developing embryo and megagametophyte of white spruce (*Picea glauca* [Moench] Voss). Plant Sci. 78, 61-71.

Misra, S., and Leal, I. (1993) Molecular cloning and characterization of a legumin-like storage protein cDNA of Douglas fir seeds. Plant Mol. Biol. 21, 709-715.

Monsalve, R.I., and Rodriguez, R. (1990) Purification and characteization of proteins from the 2S fraction from seeds of the Brassicaceae family. J. Exp. Bot. 41 (222), 89-94.

Moreira, M.A., Hermodson, M.A., Larkin, B.A., and Nielsen, N.C. (1979) Partial characterization of the acidic and basic polypeptides of glycinin. J. Biol. Chem. 254, 9921-9926.

Moreira, M.A., Herodson, M.A., Larkins, B.A., and Nielsen, N.C. (1981) Comparison of the primary structure of the acidic polypeptides of glycinin. Arch. Biochm. Biophys. 210, 633-642.

Morris, P.C., Kumar, A., Bowles, D.J., and Cuming, A.C. (1990) Osmotic stress and abscisic acid induce expression of the wheat Em genes. Eur. J. Biochem. 190, 625-630.

Mundy, J. , and Chua, N.-H. (1988) Abscisic acid and wate-stress induce the expresion of a novel rice gene. EMBO J. 7, 2279-2286.

Pederson, K., Devereux, J., Wilson, D. Sheldon, R., and Larkins, B.A. (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29, 1015-1026.

Pernollet, J-C (1978) Protein bodies of seeds: Ultrastructure, biochemistry, biosynthesis and degradation. Phytochem. 17, 1473-1480.

Raghaven, V. (1986) Experimental Embryogenesis in Angiosperms (Cambridge: Cambridge Univesity Press).

Rahman, S., Shewry, P.R., and Miflin, B.J. (1982) Differential protein accumulation during barley grain development. J. Exp. Bot. 33, 717-728.

Schuler, M.A., Ladin, B.F., Placco, J.C., Greyer, G., and Beachy, R.N. (1982) Structural sequences are conserved in the genes coding for the A and B -subunits of the soybean 7S seed storage protein. Nucl. Acids Res. 10, 8245-8261.

Schuler, M.A., Schmitt, E.S., and Beachy, R.N. (1982) Closely related families of genes code for the a, a' and B-subunits of the soybean 7S seed storage protein. Nucl. Acids Res. 10, 8225-8245.

Scofield, S.R., and Crouch, M.L. (1987) Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. J. Biol. Chem. 262, 12202-12208.

Sharief, F.S., Li, S. S-L. (1982) Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. J. Biol. Chem. 257, 14753-14759.

Skriver, K., and Mundy, J. (1990) Gene Expression in response to abscisic acid and osmotic stress. Plant Cell 2, 503-512.

Shewry, P.R., Autran, J-C., Nimmo, C.C., Lew, E.J-L., and Kasarda, D.D. (1980) N-terminal amino acid sequence homology of storage protein components from barley and a diploid wheat. Nature 286, 520-522.

Shin, S., Wong, J.H., Kobrhel, K., and Buchanan, B.B. (1993) Reduction of castor-seed 2S albumin protein by thioredoxin. Planta 189, 557-560.

Slightom, J.L., Sun, S.M., Hall, T.C. (1983) Complete nucleotide sequence of a French bean storage protein gene; phaseolin. Proc. Natl. Acad. Sci. USA. 80, 1897-1901.

Soave, C., Tardini, L., Di Fonzo, N., and salamini, F. (1981) Zein level in maize endosperm depends on a protein under control of the *opaque-2* and *opaque-6* loci. Cell 27, 403-410.

Spena, A., Viotti, A., Pirrotta, V. (1982) A homologous repetitive block structure underlies the heterogeneity of heavy and light chain zein genes. EMBO J. 1, 1589-1594.

St. Johnston, D., and Nusslein-Volhard, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. Cell 68, 201-219.

Sun, S.S.M., Altenbach, S.B., and Leung, F.W. (1987) Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.). Eur. J. Biochem. 108,477-481.

Sundas, A., Tandre, K., Holmstedt, E., and Engstrom, P. (1992) Differential gene expression during germination and after the induction of adventitious bud formation in Norway spruce embryos. Plant Mol. Biol. 18, 713-724.

Terras, F.R.G., Torrkens, S., Van Leuven, F., Osborn, R.W., Vanderleyden, J., Cammune, B.P.A., and Broekaert, W.F. (1993) A new family of basic cysteine-rich plant antifungal proteins from Brassicaceae species. FEBS. 316 (3), 233-240.

Thomann, E.B., Sollinger, J., White, C., and Rivin, C.J. (1992) Accumulation of group 3 late embryogenesis abundant proteins in *Zea mays* embryos. Plant Physiol. 99, 607-614.

Thompson, J.A., and Schroeder, H.E. (1978) Cotyledonary storage proteins in *Pisum sativum* II. Hereditary variation in components of the legumin and vicilin fractions. Aust. J. Plant. Physiol. 5, 281-294

van der Klei, H., Van Damme, J., Casteels, P., and Krebbers, E. (1993) A fifth albumin isoform is present in *Arabidopsis thaliana*. 101, 1415-1416.

Von Heijne, G. (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. 133, 17-21.

Walburg, G., Larkins, B.A. (1983) Oat seed globulin: subunit characterization and demonstration of its synthesis as a precursor. Plant Physiol. 72,161-165.

Walling, L., Drews, B.N., and Goldberg, R.B. (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci. USA. 83, 2123-2127.

Wang, T.L., and Hedley, C.L. (1991) Seed development in peas: knowing your three 'r's (or four, or five). Seed Sci. Res. 1, 3-14.

Weber, E., and Neumann, D. (1980) Protein bodies, storage organelles in plant seeds. Biochem. Physiol. Pflanz. 175, 279-306.

Youle, R.J., and Huang, A.H.C. (1978) Albumin storage proteins in the protein bodies of castor bean. Plant Physiol. 61, 13-16.

Youle, R.J., and Huang, A.H.C. (1981) Occurance of low molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species. Am. J. Bot. 68 (1), 44-48.

VITA

Surname: Machander Given Names: David

Place of Birth: Brno, Czechoslovakia
Date of Birth: January 10, 1967

Educational Institutions Attended:

University of Victoria	1992 to 1994
University of Victoria	1988 to 1991
Medicine Hat College	1985 to 1988

Degrees Awarded:

B. Sc. (Major)	University of Victoria	1991

PARTIAL COPYWRITE LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis: Isolation, characterization, and developmental regulation of a 2S seed storage protein gene in Douglas fir *(Pseudotsuga menziesii)*

Author: [redacted]

David Machander

May, 1996