CHARACTERIZATION OF AN 11S LEGUMIN-LIKE STORAGE PROTEIN GENE FROM THE GYMNOSPERM *PICEA GLAUCA*

by

MAGDALENA IVONNE MÁRQUEZ GARCÍA

B.Sc. (Biol.), National Autonomous University of Mexico (UNAM). 1987

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
THE FACULTY OF GRADUATE STUDIES
GENETICS

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 1994

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of GENETICS

The University of British Columbia
Vancouver, Canada

Date Sept - 9 - 94

Abstract

The amino acid sequence homologies of seed storage proteins in all seed plants, including gymnosperms, suggest that they evolved from a common ancestor. Seed storage protein genes have been extensively studied in angiosperms, however no data regarding the structural organization of the genes in gymnosperms is available. This is the first report of a gymnosperm seed storage protein gene. A λ genomic clone containing a *Picea* 11S legumin gene was isolated and characterized. The organization of the gene was found to be similar to angiosperm legumin genes. The nucleotide sequence contains five exons and four introns. The number of introns differs from those in angiosperms, however the position of the first three introns is highly conserved, as it is in angiosperms. A deletion was found in the third exon. The possibility of this deletion being a cloning artifact is discussed. The deduced amino acid sequence is 509 amino acids long and is 98.7% identical to a previously characterized legumin cDNA from *Picea glauca*. Amino acid comparisons of the legumin genes of Picea and other species showed the presence of highly conserved sequences. Putative regulatory sequences were found in the 5' flanking sequence of the *Picea* 11S legumin gene by comparisons between *Picea* 11S legumin promoter and angiosperm SSP promoters.

Table of Contents

List of Tables

List of figures

Acknowledgments

After three years of work it is difficult to list all those whom I owe thanks. So I will begin at the beginning.

I could not have undertaken these studies without assurance of financial support. I thank The Government of Canada Award Program, The CONACyT awards Program, and for the loan received from Banco de Mexico. I am perhaps most in debt by BC Research for financial support, for the use of facilities and for friends and colleagues. Special thanks to Ben who offered me his supervision and big help, to Craig who taught me as much as I could take. To John Carlson, my academic supervisor, whom I thank for accepting me as student, and specially for the help at the end of my thesis.

I would like to thank professors, for sharing their knowledge, even though I had hard time at the beginning, due to my English difficulties.

Very special thanks to my parents, to all my brothers and sisters, whose care and moral support accompanied me to the end. To Ian, who at the end of this journey brought so much happiness to my life.

To Stephanie, Melody and Sheila, for their cultural and language teachings. To my "Latin Family" in Vancouver: Victor and Lety, Nilda, Celia, Jaime, Lynn, Trini, Jorge M. and Jorge CH., Ivan, Oliva, Ricardo, Gloria, Andrea and Oscar. Specially for their care, and for their financial support when I needed most.

Chapter 1

INTRODUCTION

Seed storage proteins (SSP) are an important constituent of angiosperm and gymnosperm seeds, and of the spores of more distantly related ferns. They provide nutrients for the germination and post-germination processes necessary for the propagation of the species. Seeds contain 10 to 50% protein, most of which is storage protein (Shotwell and Larkins, 1989). These proteins have been extensively studied in angiosperms, due to their economic importance, but not many studies have concentrated on gymnosperm storage proteins. Even though angiosperms and gymnosperms diverged from one another 330 million years ago, the function and characteristics of SSP remain the same in both groups. It has been demonstrated that during the maturation process of spruce somatic embryos the pattern of accumulation of SSP is similar to that of the zygotic embryos (Flinn *et al*, 1991b). Redenbaugh *et al*, (1986) have proposed that the quality of somatic embryos depends on the extent of storage protein accumulation. It has been proposed that conifer storage proteins represent useful biochemical markers for developmental studies in zygotic and somatic embryogenesis (Flinn *et al*, 1991b,

Flinn *et al*, 1993). At the amino acid level some SSP sequences have been highly conserved among monocots and dicots. Three recent papers have shown that gymnosperms also share some of these regions of conserved sequence (Newton *et al*, 1992, Hager *et al*, 1992, Leal and Misra, 1993). However, no data regarding SSP gene regulation have been published. To date only two gymnosperm cDNA sequences have been published, both of them belonging to conifer seeds (Newton *et al*, 1992; Leal and Misra, 1993). Two recent papers suggested that at the mRNA level, SSPs in conifers are transcriptionally regulated (Leal and Misra, 1993; Flinn *et al*, 1993). These studies on gymnosperms have provided some important information regarding amino acid sequences and mRNA stability and transcription. Nevertheless, nothing is known about the structure and regulation of these genes in gymnosperms.

Plant seed storage proteins are abundant and their synthesis and accumulation is developmentally regulated. These characteristics make them an ideal model system to study gene expression and, in particular, to compare and contrast the structure and organization of the genes, the mechanisms of gene regulation and the evolutionary relatedness between angiosperms and gymnosperms. However the genes encoding these proteins in gymnosperms have not been isolated and their structural organization is unknown. Characterization of these genes would allow one to determine: Whether the SSP genes are structurally similar

(exons\introns) between angiosperms and gymnosperms, and whether the *cis*-acting regulatory elements are similar in both groups.

The development of cDNA libraries, genomic DNA libraries and a high quality embryogenesis system for interior spruce at B.C. Research provide an opportunity to study gene regulation during embryo development in spruce.

The contributions of this study include: 1) The characterization of the white spruce 11S legumin gene by sequencing of the coding region and comparison of the structure of the gene to other legumin genes. 2) Deduction of the amino acid sequence and comparison to other legumin protein sequences. 3) Cloning and sequencing of the promoter of the gene and identification of putative regulatory elements that could be important in the temporal and spatial regulation of the gene.

To date this is the first report that provides direct information on a complete SSP gene in gymnosperms. The sequence of the spruce 11S legumin gene and the comparisons with homologous genes in angiosperms, provides information about structure and gene organization. It also provides data regarding putative elements that may play a role in the regulation of the gene.

Chapter 2
LITERATURE REVIEW

2.1 SEED STORAGE PROTEIN CHARACTERISTICS AND CLASSIFICATION

Storage proteins from seeds are classified into: albumins, globulins, glutelins and prolamines (Shotwell and Larkins, 1989) due to their distinguishable physiochemical characteristics. These SSP show solubility in water (albumins), salt (globulins), acid or alkali solutions (glutelins) or aqueous alcohol (prolamines) (Bewley and Black, 1985). The predominant proteins in cereals are prolamins and glutelins. Cereal storage proteins occur predominantly in the endosperm and are limiting in lysine. In dicot seeds the most abundant storage proteins are globulins and albumins which occur in the cotyledons and are limiting in methionine and cysteine. Within the globulin group, two major forms of salt soluble proteins are resolved from one another on the basis of sedimentation characteristics, and fall into two different size classes 11S and 7S. Both are insoluble at pH 4.7 in 0.2M NaCl. Because globulin proteins have been best characterized in legumes, the 11S and the 7S fractions are referred to as legumins and vicilins, respectively. However, other trivial names derived from the genus of the plant are also given. The legumin fraction has a sedimentation constant of 11-13S and molecular weight of 360kD, composed of six identical

subunits of 60 KD. The legumin subunits, which are not glycosylated, have two components: the acidic, or α subunit of 40KD and the basic, or β subunit of 20Kd. These subunits are covalently linked by a single disulfide bond. The vicilin fraction has a 7-9 S value, and a molecular weight of 180kD. It is made up of three major subunits α, α' and β of 76KD, 72 KD and 53KD.

2.2 GLOBULIN PROTEINS IN GYMNOSPERMS

Studies based on solubility characteristics in conifer seed have shown the presence of globulin, also referred to as crystalloid protein and albumin type proteins, (Flinn *et al*, 1991a; Stabel *et al*, 1990; Hakman *et al*, 1990; Misra and Green, 1990; Green *et al*, 1991).

Stabel *et al*, (1990) described 3 major storage proteins in *Picea abies* during somatic embryogenesis. They found accumulation of storage proteins of 42, 33 and 22KD in mature embryos and degradation upon onset of germination. Hakman *et al*, (1990) described an additional storage protein of 28KD, and showed that mature embryos contain more storage proteins than immature embryos. This indicates that, as in angiosperms, synthesis and accumulation of storage proteins in gymnosperms occurs during embryo maturation. Storage proteins with similar molecular weights have been described in several *Pinus* species (Gifford, 1988) and *Picea glauca*. These findings suggest that seed storage proteins may be conserved among conifers.

Comparing interior spruce SSP from different embryo stages Flinn *et al*, (1991a) found by SDS-PAGE that the 41, 33, 24, and 23 KD storage proteins accumulated only in mature somatic and zygotic embryos. These proteins correspond to the storage proteins found in protein bodies isolated from mature seed embryos of interior spruce. The amount of these storage proteins are moderated by the influence of ABA (see ABA regulation). Misra and Green (1990) have shown that in the mature seed of white spruce, 70% of the total protein content correspond to crystalloid proteins, the major storage proteins (35 kd range). Other studies have shown similar results in different seed classes such as *Pinus*, Norway spruce, Douglas fir, etc., (Gifford, 1988; Misra and Green, 1991)

Interior spruce globulins (35, 33, 24 and 22KD) and albumins (41KD), accumulate at different developmental stages (Flinn *et al*, 1991b). Albumin-like protein accumulates at later stages of cotyledon maturation, and the rest of the storage proteins, similar to legumes, start accumulating during early embryo maturation. By two dimensional electrophoresis storage proteins appeared to be composed of various isoforms (Flinn *et al*, 1991a). By PAGE analysis under non-reducing conditions, Flinn *et al*, (1991a) found a 55-57 KDa protein. The characteristic pattern of storage proteins under reducing conditions was composed of 33, 24 and 22KD proteins, but no proteins with 55-57KD, suggesting that disulfide linkages exist between

the 33 and 24 and 23 KDa proteins, analogous to legumins in angiosperms.

Allona *et al,* (1992) have shown that the SSP content in *Pinus pinaster* differs from other conifers, in that glutelins represent 70% of total protein content while globulins and albumins constitute 26% and 4%, respectively. In this study the authors compared the structure and amino acid sequence of the glutelin protein to other plants and concluded that these glutelins are homologous to the 11S legumins. There are two basic differences between the *P. pinaster* glutelins and the 11S legumin proteins: a) the extraction requires alkali solution (similar to rice glutelins) and b) the basic character of the larger subunit which appears to be acidic in the rest of the 11S proteins. These data agree with the results from Jensen and Lixue (1991), where within 31 species of *Pinaceae* studied, all except the 12 *Abies* species were shown to contain 11S legumins. The *Abies* species lack 11S legumins but have instead, glutelin like proteins. There is no data regarding the amino acid sequence from *Abies* to compare with *Pinus pinaster* (Allona *et al,* 1992) or to define the homology between them. Jensen and Lixue (1991) suggest that the absence of legumin type proteins in *Abies* species may have something to do with the shorter period of viability of these seeds, compare to *Picea* or *Pinus.*

Legumin-like proteins in seeds of *Gingko biloba* have been reported (Jensen and Berthold, 1989). A 50 KD, dimer

separates into 28 and 21 KD subunits, that are linked by disulfide bonds. The molecular weight, the charge, subunits properties and heterogeneity correspond to legumin-like protein characteristics reported for angiosperms. It also has been demonstrated that the fern *Onclea sensibilis* (Templeman and DeMaggio, 1990), contain both globulin storage proteins, 7S and 11S, which are comparable to those reported by Templemann *et al*, (1987) for *Matteuccia struthiopteris*. The fern *Osmunda cinnamomea*, also contains globulin storage proteins of 5.5S and 11.3S (Templeman and deMaggio, 1990). The fact that globulin storage proteins share similarities between all plant groups suggest a strong conservation of seed storage proteins during the evolution of seed plants.

2.3 GLOBULIN STRUCTURE

Globulins have been extensively studied in both cereal and dicot seeds. In cereals they are not an important component whereas in most dicots they account for as much as 80% of the total seed protein. In gymnosperms, it has been shown that globulins and albumins are the major component of gymnosperm seeds [*Picea abies* (Stabel *et al*, 1990), *Picea* species (Gifford, 1988; Flinn *et al*, 1991a and 1991b; Roberts *et al*, 1990;) several *Pinus* species (Gifford, 1988; Allona *et al*, 1992), *Douglas fir*, Norway spruce (Misra and Green, 1991), *Gingko biloba* (Jensen and

Berthold, 1989), and Fern species (Templemann and deMaggio, 1990)].

The vicilin polypeptides are best characterized from various legume seeds (Nielsen, 1989). They are isolated from dilute salt extracts of seed meal as trimers with molecular weights around 180 KD and contain random combinations of non identical subunits. Each trimer has one or two N-linked glycosyl groups. The primary gene transcripts from the 7S genes are modified co-translationally and post-translationally. The proteins emanate from preproteins of 70 KD that, after loosing their signal peptide, are cleaved to produce the high molecular weight species (51 KD) and a smaller polypeptide of 20 KD.

Legumin polypeptides have also been best characterized in legume seeds, particularly from soybean and pea (Nielsen *et al*, 1989). Based on electron microscopy of sunflower 11S protein (helianthin), Richelet and co-workers (1980) concluded that each complex is composed of six subunits arranged in two trimers (Nielsen *et al*, 1989). Similar results were found for rape seed 11S globulin using x-ray scattering (Plietz *et al*, 1983). The hexamer has a molecular weight of 360 KD. Subunits in the hexamer are not glycosylated and need not all be identical. Different forms of the 11S subunits are part of the multimember families present in several species. Each subunit has two polypeptides components, one with an acidic and the other with a basic isoelectric point. The two components are

linked by a single disulfide bond. Legumin subunits in soybean (Nielsen, 1986) can be separated into two groups. Subunits in group I have uniform apparent molecular weight and contain more sulfur than members of group-II. Subunits in the same group are 88% to 90% homologous, however homology among members of different groups is 40% to 50%. The differences between the different members can also be observed at gene level (see chapter 2.8.).

2.4 DOMAIN ORGANIZATION

A relationship of predicted domain organization between 7S and 11S globulins has been proposed, based on amino acid and physical characteristics for soybean, pea and french bean (Argos *et al*, 1985). Domain I is the NH_2 terminus which differs significantly between 7S and 11S. Domain II contains common regions and domain III is the COOH-terminus half and is highly conserved. Nielsen (1986) proposed that the hydrophobic and most highly conserved domain is domain III. Argos *et al*, (1985) proposed that the single disulfide bond between domain I and III play an important role in maintenance of conformation of the subunit. Evolution of a common precursor for the vicilin and legumin families has been proposed based on amino acid sequence comparisons (Gibbs *et al*, 1989). The presence of hypervariable regions between domains II and III, accounts for the size differences between the two globulins. The insertions within these regions vary in length and consist

largely of repeated aspartate and glutamate residues, are very acidic and are predicted to exist in a helical conformation (for review see Shotwell and Larkins 1989). By comparing amino acid sequences from different species it has been shown that there are repeats of 8 to 38 amino acids corresponding to the hypervariable region at the end of domain II. These repeats contain a high proportion of polar, mainly acidic residues. Although these characteristics are a common structural feature, the inserts can vary in length, amino acid composition and location within and between species.

2.5 SYNTHESIS AND DEPOSITION OF SEED STORAGE PROTEINS DURING ANGIOSPERM AND GYMNOSPERM DEVELOPMENT

The synthesis of storage protein in seeds is regulated during development. In general gene expression starts at the end of the mitotic phase and finishes at the end of seed maturation when seed desiccation takes place. Along with the increase of storage protein formation, proliferation of the rough endoplasmic reticulum takes place (Muntz, 1989). Seed storage proteins are synthesized at membrane-bound cytoplasmic polysomes (Bollini and Chrispeels, 1979) and transferred from their site of synthesis to protein bodies (Nielsen *et al*, 1989; Bewley and Black, 1985). The mechanism of protein sorting remains unknown. Seeds contain more than one class of storage protein and each has a characteristic temporal

accumulation pattern. Despite the differences between angiosperm and gymnosperm embryo development, synthesis and deposition into storage organs is quite similar (Fig 2.1). Storage globulins are synthesized by membrane-bound polysomes as precursor polypeptides with NH_2- terminal signal sequence. The signal peptide directs the translocation of the nascent polypeptide into the lumen of the endoplasmic reticulum and is co-translationally removed. Soon after translation is complete the globulin precursors are assembled into trimers within the endoplasmic reticulum and then transported to vacuoles via Golgi apparatus. Once in the vacuole, the 11S precursors are cleaved into acidic and basic polypeptides which remain linked by disulfide bonds. After the proteolytic process, the 11S type trimers assemble into hexamers. Vacuoles subdivide to form protein bodies for the accumulation of storage proteins. Double-labeling of storage proteins of pea has shown that some protein bodies contain both 7S and 11S globulin proteins (Craig and Millerd, 1981). Microscopic analysis of protein bodies from nearly mature embryos of Interior spruce *(Picea glauca/Picea englemanii)* showed that both globulin proteins were present in the same organelles (Flinn *et al*, 1991b).

Protein bodies are confined to the cotyledon or the triploid endosperm cells in angiosperms. In contrast the storage seed tissue in gymnosperms is haploid and homologous to the protothallium heterosporic ferns. It

develops independently and before fertilization of the egg cell (Jensen and Bethold, 1989). Protein bodies have been identified in mature and near mature seeds and have been rarely reported at very immature stages in angiosperms or gymnosperms.

Many storage proteins undergo post-translational modifications during deposition to convert them to the correct size (Muntz, 1989). The primary translation products of the legumin genes undergo co- and post-translational modifications. A signal sequence that has a hydrophobic component is removed during the synthesis of the precursors, while cleavage to form the acidic and basic polypeptides probably occurs in protein bodies. In angiosperms, cleavage has been reported always to occur between an aspargine and a glycine, with the later becoming the N-terminal of the basic polypeptide. However, recent data on legumin-like protein of the gymnosperm *Gingko biloba* has shown that there is a Asn residue at the N-terminus of the basic subunit (Hager *et al*, 1992).

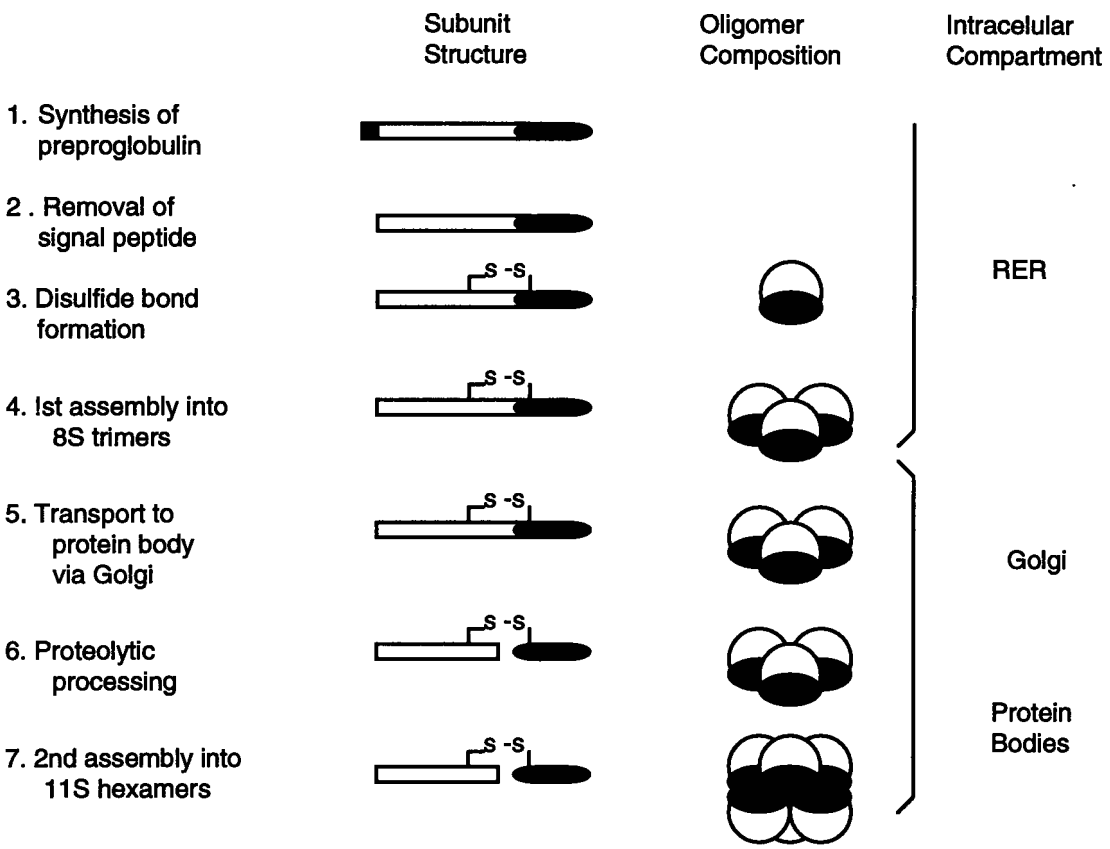


Fig 2.1 Pathway for synthesis and processing of 11S seed storage globulins. Taken from Shotwell and Larkins, 1989. (In the figures the white areas represent the α subunits, and the black areas the β subunits. RER = Rough endoplasmic reticulum).

2.6. GENE REGULATION

Genes encoding seed storage proteins have been the subject of intensive studies towards the understanding of gene expression. Seed storage proteins are encoded by a diverse gene set that is highly regulated during the plant life cycle (Goldberg *et al*, 1989). Seed storage protein genes are expressed under tissue specificity and developmental regulation and therefore are an excellent system to study the control of gene expression. The extent to which interactions between the embryo and surrounding tissues regulate development remains uncertain as are the signals that form the basis of these interactions. However many attempts to elucidate these processes have been performed in angiosperms, especially in the past 10 years. The isolation and characterization of mRNAs and their corresponding cDNAs encoding seed storage proteins have produced a vast amount of information regarding amino acid sequences, number of genes, temporal and spatial regulation, and in many cases data about the structure of the genes themselves. There is not much information published concerning the genes for SSPs in gymnosperms. The cDNA sequence for the vicilin gene of spruce (Newton *et al*, 1992), cDNA sequence for legumin and albumin genes of spruce (Newton, in preparation), cDNA sequence for legumin gene in Douglas fir (Misra and Leal, 1993) have led to interesting comparisons between angiosperms and gymnosperms at this level which permit speculation about the evolution

of these proteins. However, no data from genomic DNA clones has been published, therefore the organization of these genes in gymnosperms still remains unknown.

2.6.1 ABA REGULATION

Although the regulation of storage protein genes is influenced by the developmental stage of the seed/embryo, the details of how this occurs is not known (Baumlein *et al*, 1991). Information from phytohormone action during developmental events has provided a better understanding of embryo development. It has been shown that the phytohormone absicic acid (ABA) mediates a number of important physiological processes in plants (Finkelstein *et al*, 1985; Mundy and Chua, 1988). The mode of action of the hormone via receptors and/or transduction pathways remains obscure. Evidence at the physiological level in monocots and dicots indicates that ABA plays a major role in the control of embryo maturation and suppression of precocious germination (Mundy and Chua, 1988). It has also been demonstrated that ABA plays an important role in the proper regulation of gymnosperm embryo maturation (*Redenbaugh et al*, 1986). Roberts *et al*, (1990) and Flinn *et al*, (1991b) have shown that including ABA during the maturation of spruce somatic embryos results in accumulation of storage proteins and suppression of premature germination. Globulin proteins including

legumin and vicilin as well as albumins are accumulated in response to ABA.

ABA has been found to regulate storage protein accumulation during embryogenesis at the transcriptional level in seeds of diverse species of angiosperms (Kuhlemeir *et al*, 1987; Mundy and Chua, 1988). Exogenous ABA increases precocious accumulation of seed storage protein mRNAs in immature embryos in angiosperms (Finkelstein *et al*, 1985) and in gymnosperms (Roberts *et al*, 1990). It has been demonstrated that in the case of legumin, both protein and mRNA accumulate in response to ABA at specific developmental time in angiosperms (Finkelstein *et al*, 1985) and gymnosperms (Roberts *et al*, 1990). Sorbitol treatments produced an increase in ABA that preceded storage protein mRNA, suggesting that osmotically induced ABA stimulates storage protein expression in rapeseed (Wilen *et al*, 1990). Interior spruce zygotic and somatic embryos have shown to accumulate legumin, vicilin and albumin storage protein mRNAs, from the cotyledon stage to late embryo maturation stage, in the presence of ABA (Flinn *et al*, 1993) The amount of proteins accumulated and the transcript levels of these storage proteins in somatic embryos were ABA concentration dependent (Roberts *et al*, 1990; Flinn *et al*, 1993). Stimulation of storage protein accumulation in excised zygotic embryos by osmotic stress has been demonstrated (Finkelstein *et al*, 1985). In response to osmotic stress ABA levels in vegetative tissues are

increased (Skriver and Mundy, 1990), and it has been suggested that osmotic effects on embryo development are mediated via increased ABA levels. Vicilin and legumin proteins accumulated in broad bean cotyledons in response to high osmoticum (18% sucrose). It has been suggested that the effects of ABA can be triggered by osmotic stress in rice (Bostock and Quatrano, 1992). Recently Flinn *et al* (1993) have demonstrated osmotic stress induced storage protein and storage protein transcript accumulation in somatic embryos. The combined effect of fluoridon (an inhibitor of endogenous ABA biosynthesis) and high osmotic treatment caused the synthesis of storage proteins to be inhibited. These results suggest that similarly to angiosperms (Bostock and Quatrano, 1992), gymnosperms may have an ABA pathway that is induced by stress.

2.7 SSP ARE MEMBERS OF MULTIGENE FAMILIES

Like other eukaryotic genomes, multigene families are characteristic in plants. Genes encoding globulin storage proteins (vicilins and legumins) belong to multigene families (Ellis *et al*, 1988; Heim *et al*, 1989), varying from a few to as many as 20 members. Hybridization experiments and cDNA sequence analysis have confirmed that legumin multigene families are divided into two subfamilies A and B (Baumlein, 1986; Dure III, 1988; Turner, *et al*, 1993; Shotwell and Larkins, 1989; Breen and Crouch, 1992; Depigny-This, *et al*, 1992; Wang *et al*, 1987; Takaiwa *et al*,

1991; Shotwell *et al* 1990; Pang *et al*, 1988). The amino acid identity between the two subfamilies is 40 to 50%, however between members of the same subfamily the percentage of identity is about 80%. RFLP experiments have also confirmed for some species the presence of multigene families (Pich and Schubert, 1993; Domoney and Casey 1985; Domoney *et al*, 1986)

De Pace *et al* (1991) have shown by *in situ* hybridization that genes encoding the 2 legumin subfamilies in *Vicia faba* are arranged in two clusters: the genes encoding legumin A are located in the long arm of the two shortest subtelocentric chromosome pairs whose centromere is in a less terminal position; those coding for legumin B are located in the non-satellited arm of the longer submetacentric pair. Casey *et al* (1988) have also shown that the two legumin genes for *Pisum sativum* are located in two different chromosome pairs.

2.8 GENE STRUCTURE

Genes for 11S legumin subunits share common features. The coding region is approximately 2.7 Kb including 2 or 3 introns (Shrisat *et al*, 1989; Nielsen *et al*, 1989) in subfamily B and A, respectively. The introns are of variable sizes, 70 bp to 600 bp, however the positions are well conserved for soybean, broad bean, pea, and oilseed rape (Baumlein *et al*, 1986; Rodin *et al*, 1992; Sims and Goldberg, 1989). In angiosperms the position of introns 1

and 2 of the subfamily B, correspond to the positions of introns 2 and 3 from subfamily A genes (Baumlein *et al*, 1986). All the intron/exon junctions follow the GT/AG rule for eukaryotes.

2.9 REGULATORY SEQUENCES

Recently, attention has been given to the study of the 5' flanking sequences and to the structural and functional analysis of the upstream region that regulates seed storage protein genes. The use of transgenic plants, such as tobacco, *Petunia* and *Arabidopsis,* to investigate control sequences regulating seed storage protein gene expression has revealed an evolutionary conservation of regulatory processes. This includes tissue specificity and temporal regulation of the genes as well as correct regulation and processing for mRNAs and proteins (transient signal cleavage and glycosylation) (Bustos *et al*, 1991).

2.10 TISSUE SPECIFICITY AND TEMPORAL REGULATION

Fusion experiments of globulin genes to reporter genes and the subsequent introduction into transgenic plants have demonstrated that SSPs can be expressed in the correct size and composition only in mature seeds. Shrisat *et al*, (1989) used a T-DNA construct containing 3.4 Kb pea *LegA* fragment fused to a *nos* reporter gene which was introduced into tobacco plants via *Agrobacterium.* They demonstrated that the 3.4 Kb fragment contains all of the information

necessary for seed specificity and correct processing of the primary transcript and the legumin precursor. Ellis *et al*, (1988) showed that a 1.2 Kb upstream sequence of the pea *LegA* gene was able to direct synthesis of the legumin protein in transgenic tobacco. Shrisat *et al*, (1989) showed by deletion analysis of *LegA*, and transient expression in transgenic tobacco plants, that transgenic legumin protein was only present in seeds and absent in leaf tissues. Baumlein *et al*, (1991) have cloned a 4.7 Kb fragment of the *LegB* from *Vicia faba* containing the coding region, 2.4kb upstream and 0.3 Kb 3', and showed that it was functional after transfer into transgenic tobacco plants, and was only expressed in seed tissue. Deletion analysis of legumin genes have defined important regions for high levels of expression. Partially deleted promoter fragments of *LegB* were inserted in a vector plasmid (pGV180) in front of the *nptII* gene and transferred into tobacco via Agrobacterium. Expression was detected by *nptII* enzyme activity, and *in situ* hybridization to an antisense RNA probe. The results revealed that similar to the pea *LegA* gene (Ellis *et al* 1988), about 1.2Kb of the *LegB* flanking sequence is enough to confer high levels of expression. The possibility of minor positive elements further upstream was suggested. A construct containing only 0.2Kb of the upstream sequence resulted in a dramatic reduction of nptII activity. Shirsat *et al* (1989) showed that a 97 bp 5' fragment of pea *LegA* which contains the

CAAT and TATA boxes was not sufficient to induce expression. However the synthesis increased by increasing the 5' flanking sequence, suggesting that additional *cis*-elements must be involved. An interesting question arose from these results: What are the DNA sequences involved in the temporal and spatial regulation of these proteins? To address this question different approaches have been used: a) sequence analysis of legumin promoter regions to define *Cis*-acting sequences; b) in vitro mutagenesis of specific DNA-motifs and; c)mobility shift assays to test the binding of nuclear factors or known transcription factors.

2.11 THE ROLE OF *CIS*-ACTING ELEMENTS AND CONSERVED MOTIFS ON THE REGULATION OF LEGUMIN GENE EXPRESSION

In the search for specific motifs involved in the regulation of gene expression in seed storage proteins several putative sequences have been found. The role of these sequences in transcriptional regulation has been studied and in some cases confirmed. The legumin box is a highly conserved sequence of 28 bp, TCCATAGCCATGCAAGCTGCAGATGTC present in all legumes studied to date (Riggs *et at*, 1989; Shirsat *et al*, 1990; Ericson *et al*, 1991). These are also referred to as RY repeats for storage protein genes other than legumin (Dickinson *et al*, 1988). A 549 bp 5' flanking sequence containing CAAT, TATA and the Legumin Box could direct legumin synthesis, suggesting the involvement of the legumin box in regulation

of gene expression (Shirsat *et al*, 1989). This was also suggested by the absence of expression when using a 97 bp 5' sequence (as mentioned above) which only contained 12 bp of the legumin box. Since the legumin box is present not only in legumin genes but in all seed storage protein genes (Riggs *et al*, 1989; Chamberland *et al*, 1992), it has been suggested that presumably this sequence has a role in the regulation of tissue specificity. Many attempts to elucidate the function of this sequence have been performed. Baumlein *et al* (1991) observed a 10 fold reduction of expression when using a 200 bp 5' sequence containing the legumin box, arguing that the presence of the legumin box within this sequence plays no role in the high level of expression in developing seeds. The possibility that its function is dependent on other *Cis*-elements is not clear. However, other studies have suggested that the legumin box plays an important role as enhancer of gene expression (Lelievre *et al*, 1992). By comparing the expression of a construct containing the full *Gy2* glycinin promoter from soybean or the same promoter without the leg-box, Lelievre *et al* (1992) observed a ten-fold reduction when the element was not present, suggesting that the leg-box has a role in regulating the amount of expression of the gene.

Chamberland *et al* (1992) have shown that the legumin box plays an important role in β-conglycinin transcription. In the case of soybean β-conglycinin gene there are two

well defined legumin boxes and the mutation of both resulted in a ten-fold reduction in the transcription of the gene.

Three other regulatory elements closely related to the consensus sequences in glutelin genes in cereals TG(T/A/C)AAA(G/A)(G/T) were reported in pea *legA* between the -1203 and -549 5' flanking region (Shirsat *et al*, 1989). This sequence has been implicated in the expression of storage protein genes by nuclear DNA-binding protein experiments as well as by promoter analysis.

2.12 DNA-BINDING PROTEINS

An important step in the signal transduction pathway linking stimulus perception to alterations of eukaryotic gene expression is the binding of nuclear proteins, i.e., *trans*-acting factors to specific *Cis*-elements located primarily on sequences 5' to the gene coding region. Evidence accumulated to date indicates that sensitive regulation of transcription in a cell-type or developmentally specific manner is achieved by multiplicity of interactions between promoter enhancer sequences and *trans*-acting factors with either stimulatory or repressive functions (Meakin and Gatehouse, 1991)

Cis-acting elements controlling seed-specific expression have been identified in maize zein, wheat glutenin, barley hordein, oilseed rape napin, soybean lectin, conglycinin and french bean phaseolin genes

(Jordano *et al*, 1989, and references therein). Conserved elements have been postulated to play an important role in activation of gene transcription by the binding of *trans*-acting nuclear proteins.

Examining sequence specific DNA-protein interactions, by DNA-protein binding and mobility shift assays, Shirsat *et al*, (1990) demonstrated that nuclear proteins strongly bound the -549bp flanking sequence. However a truncated -124bp *LegA* construct fragment containing the complete leg-box sequence with 6 additional 5' bases did not bind nuclear proteins.

DNA footprinting experiments demonstrated interaction of a nuclear protein from pea seed (LABF1) with the -549 to -316 fragment of *LegA* 5' flanking region (Meankin and Gatehouse, 1991). Gel retardation assays showed the specific interaction between two *LegA* promoter fragments (-540 to -316 and -833 to -584) and pea seed nuclear proteins. The promoter sequence of *LegA* between -316 to +40 did not form stable complexes with seed nuclear protein. Developmental regulation and tissue specificity between nuclear proteins and *legA* promoter was demonstrated by gel retardation assays. The nuclear protein binding the promoter region showed a molecular weight of 84 - 116 KD that was determined by elution and renaturation of protein from PAGE-SDS. Its function as DNA-binding protein was confirmed by competition assays. Meankin and Gatehouse (1991) showed the tissue specificity and developmental

regulation of this binding factor. Extracts from peas during development were tested with a probe consisting of -549 to -316 *LegA* promoter. The factor was detected in seed extracts 12 to 19 days after anthesis (DAA). The 15 DAA seed extract interacted strongly to the probe. No pea leaf extracts recognized the probe. The evidence that LABF1 was seed specific and that its binding activity was temporally correlated with synthesis of mRNA (Thompson and Larkins, 1989), suggested that it may act as a transcriptional enhancer. Since a low level of transcription occurs when LABF1 protein is not detectable, transcriptional enhancement rather than induction was suggested.

In studies on sunflower helianthinin gene expression, Jordano *et al* (1989) detected nuclear proteins that bind an A/T rich region upstream of the helianthinin promoter. Binding competition experiments showed that sunflower embryonic and somatic nuclear proteins bound to the french bean phaseolin gene and to the carrot DcG3 embryo specific genes, suggesting that binding activities are conserved between plant species. In the same report the authors showed that the sequence, containing the protein binding site, fused to a reporter gene *(GUS)* and driven by a truncated *CaMV* 35S promoter, enhanced expression in seeds in transgenic tobacco plants.

Elements that bind nuclear proteins in other species are mostly A/T rich, and do not show any particular

sequence conservation, e.g. sunflower helianthinin (Jordano *et al,* 1989), *Pha* in *phaseolus vulgaris* (Riggs *et al,* 1989), soybean lectin (Jofuku *et al,* 1987), and two soybean globulin genes (Kitamura *et al,* 1990; Itoh *et al,* 1993).

Despite its highly conserved sequence, no proteins binding to the legumin box have been detected (Meankin and Gatehouse, 1991; Shirsat *et al,* 1990; Itoh *et al*,1993). Riggs *et al,* (1989) suggested that as an alternative to regulation by soluble proteins, the CATGCATG motif may form a *Z*-DNA structure *in vivo*. One possibility is that after an activator protein binds to the upstream region (-549 to -316), the CATGCATG motif may adopt an altered conformation that enhances the recognition for or passage of transcriptional complexes. Itoh *et al* (1993), also suggested that assuming that the leg box could be a binding site for nuclear factors, these factors may be very unstable, or they may require other factors binding at a different site for interaction with these motifs as found in yeast mating type regulatory proteins.

Chapter 3
MATERIAL AND METHODS

3.1. 11S LEGUMIN GENOMIC DNA ISOLATION

The spruce 11S legumin cDNA XI5H was obtained from Dr. Craig Newton at B.C. Research. The cDNA was labeled and used as a probe to isolate the white spruce 11S legumin λ-genomic clone. The EMBL3 Eastern white spruce total genomic λ-library was constructed by L. DeVerno (PFNI). Isolated from a partial Sau3a digest, DNA was cloned into a BamH1 site.

3.2. RANDOM LABELING

20 ng of XI5H cDNA or E2.8 DNA in 11µl H_2O were heated at 100°C for 5 min. to separate the double stranded DNA. After heating the DNA sample was placed on ice and all the labeling components were added (2µl 10X labeling buffer, 2µl dNTPs (2 mM each G, T, C), 1µl pN_6, 1µl BSA (1 mg/ml), 1 µl 0.1 M DTT, 2µl α^{32}P-dATP (5000 Ci/mmol; Dupont), 1 unit of Klenow enzyme (BRL). The labeling reaction was allowed to proceed at room temperature overnight. The probe was then twice precipitated with half volume of 7.5 M NH_4OAc and 2.5 volumes of cold 95% ethanol, using 3µl of tRNA (2 mg/ml) as carrier (-80 °C; 30 min.). Sample was centrifuged at 12,000 rpm, 15 min., 4°C, The pellet was dried at room temperature and resuspended in 100 µl water. 1µl of sample

was used to verify ^{32}P-ATP incorporation in a liquid scintillation counter.

3.3. λ-GENOMIC DNA CHARACTERIZATION

Three λ clones (XI5H-1, XI5H-2, XI5H-3) provided by Dr. Craig Newton were used to characterize the genomic DNA. λ-phage dilutions (10^4-10 plaques /ml of SM (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$) were mixed with 0.1 ml of ER1647 bacteria host and incubated at 37°C in 10 ml-falcon tubes for 20 min. Following the incubation 3 ml of TB top agar (10 g/l tryptone, 5 g/l NaCl, 6 g/l agar) were added and plated on TB plates, incubated at 37°C for 7 hr and placed at 4°C overnight. One nylon hybridization filter (0.45μm) was placed on top of each plate for 2 min. to allow phage to adsorb to filter. Filters were peeled off the plates and placed DNA side up on 3M Whatman paper, soaked in denaturing solution (1.5 M NaCl; 0.5 M NaOH) for 5 min. followed by neutralizing solution (1.5 M NaCl; 0.5 M Tris-HCl pH 7.5) for another 5 min.. Filters were air dried on Whatman paper for 20 min. and exposed to UV light for 7 min. to denature DNA. Filters were placed in 10 ml pre hybridization solution (5X SSPE; 1% SLS; 0.1% Na Pyrophosphate; 200 μg/ml denatured salmon sperm DNA) for 2 hr at 65°C, the radiolabeled cDNA probe (see above) was added, and hybridization proceeded overnight at 65 °C. Filters were washed twice with 2X SSC pH 7 (0.15M NaCl,

0.015M Na-citrate) and 0.1% SDS for 30 min. at 65°C then air dried at room temperature and exposed to Kodak X-ray film with intensifying screen (-80 °C, overnight). Following overnight exposure the film was developed. The positive plaques were identified by aligning filters to X-ray film. Then by aligning filter to agar plates, three single phage plaques were identified, picked, transferred to culture tubes with 1 ml SM each, and shaken for 2 hr at room temperature. Dilutions of phage were made in SM, incubated for 20 min. with bacteria host and plated on TB agar plates. After overnight incubation at 37 °C, the number of plaque forming units (pfu/ml) was determined.

3.4. λ-DNA PREPARATION

5 x 10^9 pfu/ml of XI5H-1 and XI5H-2 were added to a 10 ml falcon tube containing 10 ml of ER1647 culture growth overnight in SM media (5 x 10^8 cell/ml) containing 4 ml of SM, mixed by inversion and placed at 37 °C for 20 min.. Lysates were added to 250 ml TB media and shaken at 2000 rpm (37 °C; 5 hr). 5 ml of chloroform were added to each 250 ml of lysed culture and shaken at 37 °C 10 min. 1µg/ml of each DNase I and RNase were added to lysates and incubated 30 min. at room temperature (RT), followed by the addition of NaCl to 1M final concentration, dissolved by swirling and placed 1 hr on ice. Lysates were centrifuged (11,000 rpm; 10 min.; 4 °C). 25 g of PEG 8000 (10% final

concentration) were added to supernatant, dissolved by stirring (RT), cooled on iced water and placed in the cold room overnight. After centrifuging (11,000 rpm; 10 min.) pellets were resuspended in 8 ml SM media. 8 ml of chloroform were added and samples centrifuged (3000 rpm; 15 min.; 4 ^{o}C), the aqueous phase recovered and bacteriophage particles collected by centrifugation (25,000 rpm; 2 hr; 4 ^{o}C) and resuspended in 0.5 ml SM (4 ^{o}C; overnight; rocking platform).

3.5. EXTRACTION OF λ-DNA

The bacteriophage solution was gently resuspended in SM media, and 20µl EDTA (0.5 M), 5µl proteinase K (5 mg/ml), 25 µl SDS 10% were added and incubated 1 hr at 56 ^{o}C). This step was followed by two chloroform:phenol (1:1) and one chloroform extractions. DNA in the aqueous phase was precipitated with half volume 7.5 M NH_4OAc and 1 volume isopropanol at -20 ^{o}C overnight. DNA was pelleted by centrifugation (14,000 rpm, 20 min.), washed with 70 % ethanol, dried and resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and CsCl purified. A 10 µl sample was digested with EcoRI and HindIII restriction enzymes following the instructions from supplier (Pharmacia), and run on a 0.8% agarose gel in TEA buffer containing EtBr (0.5 µg/ml).

3.6. CsCl DNA PURIFICATION

λ-DNA and plasmid DNA (E2.8 and S3.7) were CsCl purified. CsCl plasmid DNA preparations were used for the generation of deletion constructs, for southern blot hybridization, as well as for sequencing reactions.

After the DNA extraction step λ- or plasmid nucleic acid pellets resuspended in 2.4 ml TE buffer were purified by equilibrium sedimentation in cesium chloride-ethidium bromide (CsCl-EtBr) gradient. 4.2g of CsCl and 400 µl of 10 mg/ml EtBr were added to the plasmid solution, centrifuged (6000 rpm; RT) in a JA-21 Beckman centrifuge for 10 min.. A Ti70.1 quick seal ultracentrifuge tubes was partially filled with 8 ml of light cesium chloride solution (63 g/100 ml TE), and the DNA plasmid solution placed at the bottom of the tube. The tube was filled with cesium solution, balanced, sealed and centrifuged for 18 hr (40,000 rpm ; 20 ^{o}C) in a Ti 70 Beckman ultracentrifuge rotor. After centrifugation, the tube was protected from light, and the position of the plasmid band determined by exposing the tube to UV light. The lower DNA band was removed from the tube using a 1 ml syringe with a wide bore needle. Three volumes of TE were added and DNA extracted 4 times with an equal volume of water saturated isobutanol. The lower aqueous phase was transferred to a 30 ml corex tube and precipitated with 3

volumes of cold 95% ethanol overnight at 20 °C. The sample was centrifuged (15,000 rpm; 30 min.), the pellet rinsed with cold ethanol, dried at room temperature, and resuspended in TE buffer.

3.7. SOUTHERN BLOT

After digesting 5 µl DNA each, with EcoRI, HindIII, SalI, and combinations of them, following conditions from Promega, samples were loaded on a 1% agarose gel in TBE (0.89 mM Tris-base, 0.89 mM boric acid, 20 µM EDTA) containing 0.5 µg/ml EtBr, run for 4 hr at 75 V/H. The gel was washed in HCl solution (21.5 ml/l) 30 min., in 3M NaCl : 1M NaOH three times for 10 min., denatured in 1M NH_4OAc three times 10 min. and blotted to a nitrocellulose filter for 4 hr at room temperature. After blotting the filter was allowed to dry at room temperature and DNA fixed 5 min. under UV light. After this step the filter was hybridized to the labeled cDNA probe using the same method described above.

3.8. MAPPING THE λ-GENOMIC DNA

In order to generate a map of the legumin genomic DNA, restriction digestion of the λ-DNA and plasmid DNA were performed. In all cases CsCl purified DNA was used. 5 µg of DNA were digested with each of the following enzymes or combinations of them: EcoRI, SalI, PstI, BamHI, HindIII. All

enzymes were obtained from Promega and restriction conditions were carried out as suggested by supplier. The resulting fragments were separated according to size by electrophoresis through a 0.8% agarose gel cast in 0.5x TBE, containing EtBr (0.5 μg/ml). λ-DNA size markers were loaded next to DNA samples and were used as a reference to determine sizes. After electrophoresis was completed the gel was photographed under UV light. Sizing the different DNA fragments was done manually by measuring the distances and referring to the size markers, and also by computer scanning the negative of the photograph through a Sparc 1 (Sun) scanner. After pictures were taken gels were blotted and hybridized to cDNA probe as described above.

3.9. PLASMID DNA CLONING

The pEMBL and pGEM-3Z expression systems were convenient cloning vectors to use due to the multiple cloning sites, and usefulness for deletion experiments and sequencing reactions. Both DNA fragment and Vector DNA were digested with the same restriction enzyme, to produce compatible ends for cloning. λ-DNA was digested with EcoRI or SalI and the products were visualized in 0.8% agarose gels. An EcoRI 2.8KD fragment (E2.8) and a SalI fragment of 3.7 KD (S3.7), were gel purified (using the prep-A-gene Kit, Promega), and cloned in vector pEMBL (EcoRI 2.8 KD fragment) or pGEM-3Z vector (SalI 3.7).

3.10. VECTOR PREPARATION

pEMBL or pGEM vector (10 µg) were digested with EcoRI or SalI as needed, following the Promega instructions to get complete digestion, then treated with calf intestinal alkaline phosphatase (CIAP) (0.01 µ/pmol ends in 100 µl; 37 ^{o}C; 1 hr) to remove 5' phosphate groups and prevent recircularization of the vector during ligation. CIAP reaction was stopped by adding 2 µl of 0.5 M EDTA. Vector DNA was phenol/Chloroform extracted once, the aqueous phase extracted with chloroform:isoamyl alcohol (24:1), and DNA precipitated with 0.5 volumes of 7.5 M ammonium acetate and 2 volumes of 95% ethanol (-80 ^{o}C, 30 min.). DNA pellets were collected by centrifugation (12,000 rpm; 10 min.), washed with 95% ethanol, dried and resuspended in $H_2$0. The DNA concentration was determined by absorption spectroscopy A_{260}.

3.11. LIGATION OF VECTOR AND INSERT DNA

Vector DNA and insert DNA were mixed at 1;1 and 1:3 molar ratio in ligase mix (1 µl ligase 5x buffer, 1 unit DNA Ligase, 1 µl 10 mM ATP, $H_2$0 to 10 µl) for overnight reaction at room temperature. After ligation reactions were complete, plasmid DNA was transformed into SURE™ competent cells (see competent cells below).

Aliquots of 50 µl of competent cells were thawed on

ice and 2.5µl of the plasmid ligation reaction were added and incubated 15 min. on ice. To increase transformation efficiency a heat shock at 37 ^{o}C for 1 min. was performed, followed by 2 min. on ice. 200µl of LB medium were added to each tube and incubation at 37 ^{o}C was allow for 45 to 60 min.. Cells were plated on LB (10 g/l Bacto-tryptone, 5 g/l Bacto-yeast extract, 5 g/l NaCl, pH 7) plates containing 50 µg/ml ampicillin, 10 µl IPTG (1M), and 50 µl X-Gal (20 mg/ml) for 14 - 16 hr. Recombinant white colonies were selected and single bacteria colonies inoculated on 2 ml LB medium containing 50 µg/ml ampicillin, incubated 8 - 14 hr., followed by miniprep plasmid DNA isolation procedures.

3.12. COMPETENT CELLS PREPARATION

1ml YT/Mg (20g/l bacto-tryptone, 5 g/l yeast extract, 5 g/l NaCl 2.5 g/l $MgSO_4$) was inoculated with 1 Sure™ (Stratagene) colony and grown to mid-log phase. Bacteria were then added to a 100ml warm YT/Mg in 500 ml flask and grown to A_{600} =0.6. Bacteria cells were chilled on ice, pelleted by centrifugation (3,500 rpm, 15 min. 2^{o}C) and gently resuspended in 40 ml of cold TfBI (30 mM KOAc, 50 mM $MnCl_2$ 100 mM KCl, 10 mM $CaCl_2$, 15% glycerol). The bacterial suspension was centrifuged as above, resuspended in 5ml of cold TfBII (10 mM Na-MOPS pH 7.0, 75 mM $CaCl_2$, 10 mM KCl, 15% glycerol), aliquoted and frozen in liquid nitrogen and stored at -70^{o}C.

3.13. ISOLATION OF PLASMID DNA BY ALKALI METHOD

Plasmid DNA was isolated by the alkali lysis procedure described by Maniatis et al (1982). 1.5 ml of the plasmid cultures were centriguged at 12,000 g for 2 min. in microfuge tubes. The bacterial pellets were resuspended by vortexing in 100 µl ice cold lysis buffer (25 mM Tris-HCl, ph 8.0, 10 mM EDTA, 50 mM glucose), followed by the addition of freshly prepared solution II (0.2N NaOH, 1%SDS), mixing by inversion and incubating 2 min. at RT. 150 µl of ice-cold solution III (Potasium acetate pH 4.8) were added and mixed by inversion, incubated 5 min. on ice, and centrifuged at 12,000 g 5 min.. The supernatant was separated and one volume of phenol:chlorophorm (1:1) was added, vortexed for 1 min. and centrifuged 5 min. at 12,ooo g. The upper aqueous phase was precipitated with 2.5 volumes of ethanol (95%) for 5 min. at RT. DNA was pelleted by centrifugation at 12,ooo g 10 min., washed with 70% ethanol, vaccuum dried, and resuspended in 50 µl TE buffer.

3.14. GENERATION OF UNIDIRECTIONAL DELETION CONSTRUCTS FOR SEQUENCING

The erase a-base-system™ (Promega) was used for the construction of subclones containing progressive deletions of the legumin gene and promoter, to facilitate the sequence analysis. The system is based on the use of

exonuclease III to digest DNA from a 5' protruding or blunt end, while leaving a 4 base 3' protruding end or an α-phosphotioate filled end intact. The digestion produced a series of deletions of increasing size that were exposed to SI nuclease which removed the single stranded tails remainded from the Exo III digestion. SI nuclease was neutralized and heat inactivated. Klenow DNA polymerase was added to the reaction to generate blunt ends, that were ligated to circularize the deletion containing plasmids. Half of each reaction was used to transform Sure™ (Stratagene) competent cells. After transformation 4 to 10 colonies of each deletion time were selected and plasmid preparations were performed followed by enzymatic restriction to determine the samples to be sequenced.

Two CsCl-purified DNA inserts were used for sequencing purposes, the E2.8 (containing the complete coding region and 0.7kb of 3' flanking sequence) and the S3.7 (containing 1.4kb of the promoter sequence and 2.3Kb of the coding region). Bacteria containing the E2.8 or S3.7 insert were grown overnight in two hundred and fifty ml of YT broth (8 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 5 g/l NaCl) with ampicillin added at 100 μg/ml. Following the incubation at 37 ° nucleic acids were isolated using the alkaline lysis method.

3.15. SEQUENCING METHODOLOGY

The Promega fmol™ sequencing system was used to sequence the 11S legumin gene. The fmol system uses Taq DNA polymerase which is stable at 95°C and which replicates DNA at 70 °C, and allows use of a thermocycling apparatus.(Twin block ™ system, ERICOMP). Three different primers (27mer legumin specific (5'-GCCTAGGCGTTAATTGTCATAGACGTA-3'), 24mer Forward, 20mer Reverse) were end labeled (10 pmoles primer, 10 pmoles γ-ATP, 1µl 10X T4 buffer (500 mM Tris-HCl ph7.5, 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine) 5 units T4 polynuclotide kinase, 30 min. 37 °C. The kinase was inactivated at 90°C 2 min. and the labeled primers were then used for sequencing proposes. 1-2µl template DNA were mixed with 4.5 µl of sequencing buffer (250 mM Tris-HCl pH9.0, 10mM $MgCl_2$), 1.5µl labeled primer, H_2O to 18 µl, and 1µl TaqDNA polymerase (5u/µl). For each set of reactions 4 µl of the enzyme/primer/template were added to each 0.5 ml eppendorf tube containing 1 µl of each of the four d/ddNTP mixes (G [40µM 7-Deaza dGTP, 40µM dATP, 40µM dTTP, 40µM dCTP, 60µM ddGTP], A [40µM 7-Deaza dGTP, 40µM dATP, 40µM dTTP, 40µM dCTP, 700µM ddATP], T [40µM 7-Deaza dGTP, 40µM dATP, 40µM dTTP, 40µM dCTP, 1200mM ddTTP], C [40µM 7-Deaza dGTP, 40µM dATP, 40µM dTTP, 40µM dCTP, 400µM ddCTP]). One drop of mineral oil was added to each tube, spun for 2 sec. and placed in the thermal cycler preheated at 95°C for 2

min. The PCR program used for the sequencing reactions was as follow: 95 ^{o}C 30 sec. (denaturation), 60 ^{o}C 30 sec. (annealing), 70 ^{o}C 1 min. (extension) for 30 cycles total. After reactions were completed 3 µl of stop solution (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanole) were added to each tube. Samples were heated for 2 min. at 70 ^{o}C just before loading on sequencing gels.

3.16. SEQUENCING GELS AND ELECTROPHORESIS

5 ml of 50% long ranger solution (J.T.Baker), 21 g Urea, 6 ml 10x TBE, 25 ml H_2O were mixed and filtered. 25 µl TEMED and 250 µl of 10% ammonium persulfate were added and the solution was transferred to a 50-60 ml syringe and injected in between the sequencing gel glass plates. After gel polymerization (1-2 hr) sequencing reactions were loaded. Electrophoresis was performed using 0.6X TBE running buffer at 30 watts for 3-6 hrs. Once electrophoresis was completed plates were separated and the gel transferred onto Whatman 3M paper, covered with saran wrap, vacuum dried at 80^{o}C for 1 hr and exposed to a Kodak x-ray film overnight.

3.17 PRIMER EXTENSION

mRNA from white spruce proembryoand mature embryo stages was obtained from Dr. Dave Cyr (BCResearch). 3 µg of proembryo and mature embryo mRNAs were used per reaction.

Proembry mRNA was used as a control and no RNA was used as a negative control.

Three samples of mature mRNA one sample of proembryo mRNA and no RNA sample were mixed each with 10 ng of 27mer primer previously labelled (see primer labelling), in 0.3M NaCl heated at 80 °C for 60 secs. and immediately after each of the mature embryo mRNAs were placed at 42 °C, 55 °C and 65 °C for 15 min. Pro embryo and no RNA mixtures were placed at 55 °C for 15 min. All samples were removed and the Reverse Transcriptase mixture (0.1 M KCl, 0.1 M Tris pH8.5, 0.1 mg/ml BSA, 0.01 M DTT, 0.01 M $MgCl_2$,0.05 u/ml RNase inhibitor, 250 uM dNTPs, 0.5 units Reverse Transcriptase enzyme) was added, reactions were allowed for 1 hr. at 42 °C. Reactions were stopped by the addition of RNase A (37 °C 10 min.) followed by ethanol precipitation (2 vol. ethanol 95% and 1/2 vol. 7.5 M NH_4Ac) at -80 °C for 20 min. Samples were redissolve in sequencing running buffer (see sequencing methodology) and ran along with S3.7 sequencing reactions using the same 27mer primer.

Chapter 4

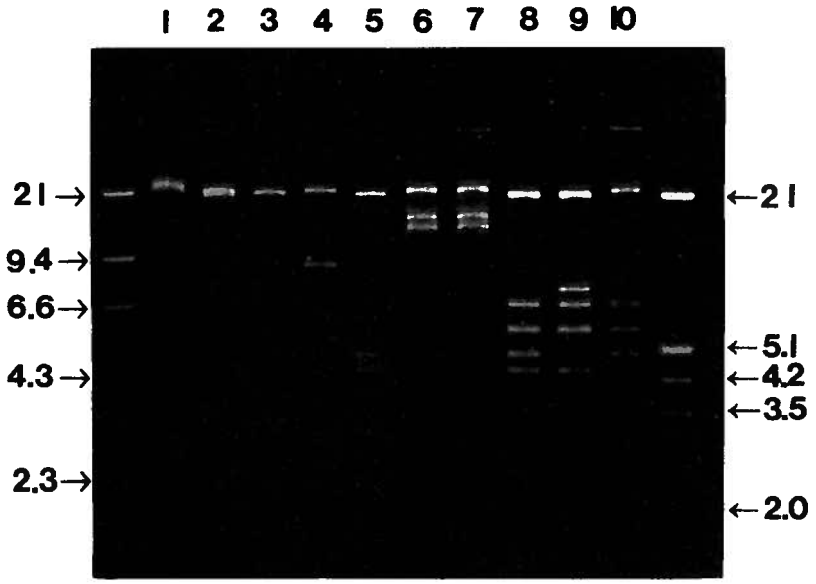
RESULTS

4.1 IDENTIFICATION OF A GENOMIC CLONE CONTAINING THE 11S LEGUMIN GENE

The EMBL3 Eastern white spruce λ-genomic library was screened with the spruce XI5H cDNA probe (11S legumin like)(C. Newton, unpublished results). Three plaques (-1,-2,-3) that strongly hybridized to the probe were selected. A partial restriction map was obtained for each DNA sample followed by a Southern blot hybridization using the 5' end of the XI5H cDNA as a probe. Phages XI5H-2 and XI5H-3 show the same restriction pattern and are therefore considered to be identical clones (data not shown). Only one was used for the southern blot. Fig 4.1 shows restriction digests and southern blot of clones λ XI5H-1 and -2. The two clones exhibited different restriction patterns. Nevertheless both hybridized the probe, although very weakly for XI5H-2, suggesting the presence of at least two members of the legumin family in *Picea glauca*. Since XI5H-1 strongly hybridized the probe, it was selected for further characterization. A restriction map for λ XI5H-1 was constructed which showed that this clone has an insert of 17.9Kb containing a 2.8 Kb EcoRI fragment (E2.8) and a 3.7

Kb Sal I fragment (S3.7) that strongly hybridize the cDNA probe. The λ XI5H-1 restriction map is shown in figure 4.3. Three fragments from XI5H-1 were subcloned in *E.coli* sequencing vectors, E2.8 was subcloned in a pEMBL vector, S3.7 in a pGEM-Z3 vector and S4.7 which contained 3' flanking sequence was subcloned in pUC9 vector. After subcloning, the 3 plasmid DNA samples were CsCl purified. In order to obtain a full length genomic DNA the E2.8 and S3.7 were selected. Both of them strongly hybridized the probe and they were large enough to contain a complete 11S gene based on legumin genes in angiosperms (Ellis *et al*, 1988; Nielsen *et al*, 1989). These two fragments were sequenced by the PCR method and only when the sequence was not clear the sequenase Kit from USB was used for confirmation. The E2.8 DNA was only partially sequenced. E2.8 has a SalI site at approximately 0.3kb from the 3' end. The E2.8 DNA was sequenced from the 5' end to the SalI site, 2350 bp in length containing 18 bp of 5' flanking region, 483 bp of 3' non-coding region and an open reading frame of 1867 bp. The S3.7 clone contains 1.4 Kb of promoter region, 1867 bp of coding region and 483bp of 3' non coding sequence.

A



B

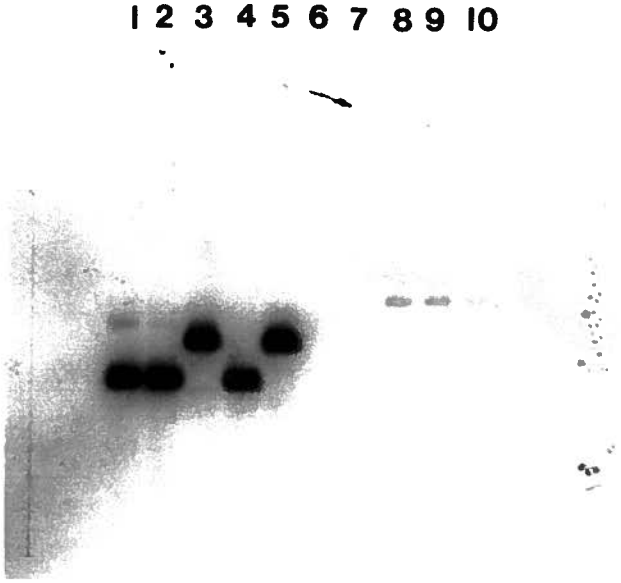
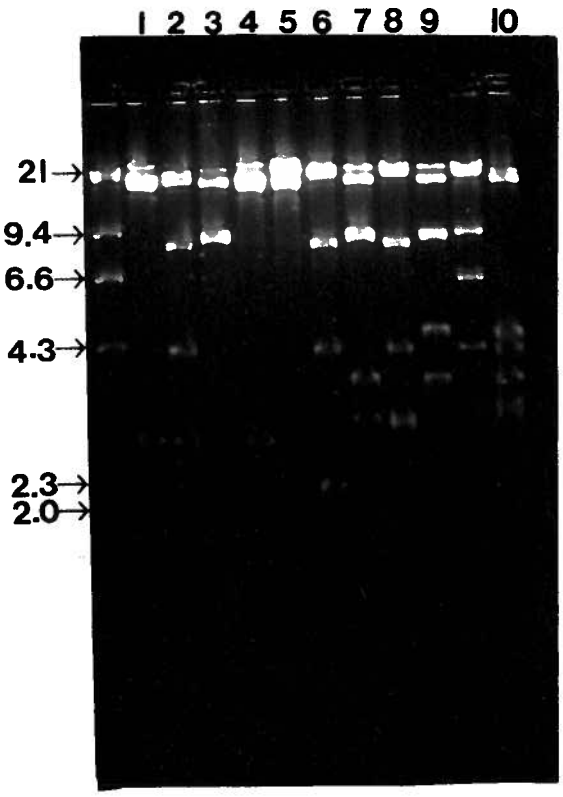


Fig 4.1. Restriction enzyme digests (A) and Southern hybridization (B) of 2 λ clones (1-5) XI5H-1 and (6-10) XI5H-2 presumably containing the 11S legumin gene from *Picea glauca*. 3μg of λ-DNA were digested EcoRI (lane 1,2,6 and 7); HindIII (lane 3 ,5,8 and 10); EcoRI/HindIII (lane 4 and 9); separated in 0.8% agarose gels (A) in presence of EtBr, and (B) blotted on Hybond N and hybridized to the ^{32}P-labeled 5' cDNA probe. Bordering lanes show DNA λ markers.

A



B

Fig 4.2. Restriction digests and Southern analysis of λ genomic clone (XI5H-1) containing *Picea glauca* 11S legumin gene. 2μg of λ-DNA were digested with either EcoRI (lane 1); EcoRI/HindIII (lane 2), EcoRI/SalI (lane 3); EcoRI/PstI (lane 4); HindIII(lane 5); HindIII/PstI (Lane 6); HindIII/SalI (lane 7); PstI (lane 8); SalI (lane 9); SalI/PstI (lane 10); and were separated in 0.8% agarose gels (A) in presence of EtBr, and (B) blotted on Hybond N and hybridized to the ^{32}P-labeled E2.8 probe. λ-DNA markers are shown.

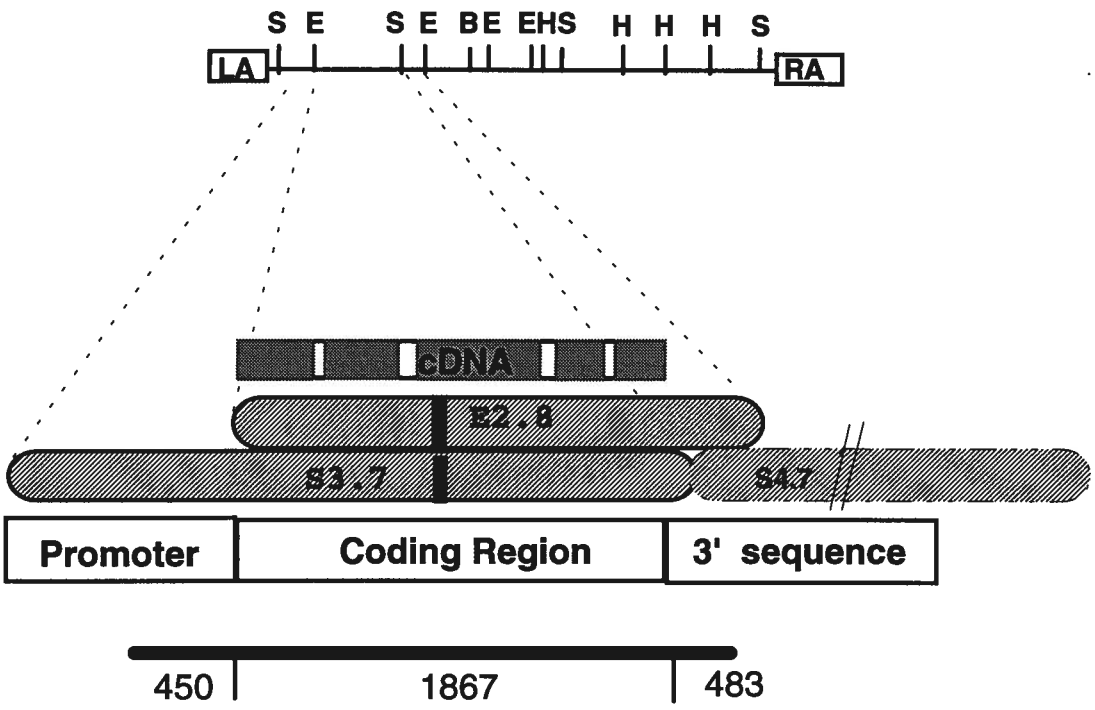


Fig 4.3 Restriction map of the λ-genomic spruce XI5H-1. The clones that strongly hybridize the spruce 11S legumin cDNA are shown (E2.8 and S3.7). S3.7 and E2.8 contain a deletion 77 bp long, marked as a solid bar across them. The cDNA is 1738 bp in length and is shown in the figure for comparison. The empty bars inserted in the cDNA sequence correspond to the introns. The line at the bottom represents the sequence obtained from the three regions; numbers correspond to length in bp. Note that the restriction map do not correspond to the cDNA. LA= λ left arm; RA= λ right arm; E= EcoRI, S= SalI, H= HindIII, B=BamHI

4.2. 11S LEGUMIN CODING REGION

In order to sequence the complete coding region, nested deletion experiments using the Promega erase-a-base system were performed. All deletion products were re-ligated and ligation reactions were used to transform SURE™ competent cells. White containing inserts over blue colonies were selected, grown in YT media and DNA extracted as detailed in materials and methods. 2 µg of DNA of each sample were restricted with EcoRI and separated in agarose gels, to select samples to be sequenced. All sequence data was compared to the cDNA sequence using NA-align and NA-compare programs. These comparisons showed a 98.7% homology between the cDNA and gDNA, with the exception of a 77bp deletion. Figure 4.4 gives the complete sequence of the gene including 5' and 3' flanking sequences. 10 nucleotides in the genomic DNA are different from the cDNA. 5 of these substitutions are present in the third position of codons, four of which do not produce a change in amino acid, while one changes a serine for an arginine. Another 3 of these substitutions are in the second position of codons and result in a change to a similar amino acid. One nucleotide that differs from the cDNA in the first position resulted in a change from an arginine to a cysteine. None of these changes are found in any of the highly conserved regions, and therefore do not greatly affect amino acid homologies with other legumin genes.

```
AATATTAACA TTAAAAAAAT TTATGTAGGA ATATTTAAGC CAATAAAAAA TATAAATATT      60

TAAGTAATAA AAAATAAAAA ATATAAAATT TAAGTAATAA ATTTTTTCCT CGTGGAACGT     120

ATTTTTTCCT CGTTAGATGT GAACACATAC ATTGACAGCA GCATTTCCTT AAACAAACAC     180

TCAACTTTTT ACACGTCGAA TCGTACGACA TTACACGACA CGCCGGAGAG TAGCCGCATC     240

ACACGTGATG AAGATTCCCT TTGGCCTTAA GCCCATGTGG CTCTCAGGAG TAGATATAGC     300
                                                                +1
CTTAATCATA TCGCCCTTCG CATGCTATAA AGCTAATAAT ATTCAACAAC AGCAGGGACA     360

CAGCCTGTGT ATAAAAACAC GAAGAAGCAT CTAGGAATTC AAAACGAAGC AAGAGAAATG     420

AAGGGGAAGA TGATGAGATC AGCGCGTTGT CCACTGATGC AGATACTGTT AATTGCCTCT     480
                           ⇒          M   Q   I   L   L   I   A   S       8

GCCTGCTTTC TTTTTCTCTC CCTGTCAACT GTATCACCTG TAACTGCAAT TTCCCAGCAA     540
 A   C   F   L   F   L   S   L   S   T   V   S   P   V   T   A   I   S   Q   Q      28

AGAAGAGGAA GAGGTCGTCG TTACGATGAG CAATCATCGT CATGTCGGAG GCTGCGGCGG     600
 R   R   G   R   G   R   R   Y   D   E   Q   S   S   S   C   R   R   L   R   R      48

CTAAGCGCCC ACGAACCGTC TGAATCGGAG ACGATAAGAT CGGATGGTGG CACCTTCGAA     660
 L   S   A   H   E   P   S   E   S   E   T   I   R   S   D   G   G   T   F   E      68

TTGTCCACTG GAGAAGACAA CGAGGAATTA GAGTGCGCAG GCGTTGCCTT CTTCAGAAAG     720
 L   S   T   G   E   D   N   E   E   L   E   C   A   G   V   A   F   F   R   K      88

ACGATCGAAA GCAACGCCAT CTTGTTGCCC CGATATCCCA GCGCCGATCT GTTGCTTTAC     780
 T   I   E   S   N   A   I   L   L   P   R   Y   P   S   A   D   L   L   L   Y     108

GTTGTCCGAG GTAGGTTAAT ACATGATTGT GTATGGCACA TGATTGCCTA AAATTGTCAT     840
 V   V   R---------------------------- intron 1 ----------------------     111

TATAATTGTG TATGCAG-TG AGGGCAGACT GGGAATTGTT TTCCCCGGAT GTCCGGAGAC     899

-------------------G   E   G   R   L   G   I   V   F   P   G   C   P   E   T     126

TTTCAGAGAT CATTCCTCGT TTCAAGGGCG ATCAAGGCAC AGATCAGAGG GACGACGGGA     959
  F   R   D   H   S   S   F   Q   G   R   S   R   H   R   S   E   G   R   R   E     146

GGAAGAGGAA GAGGAAGAAG AGGACTCAAG TCAGAAGGTG AGGCGAGTGA GGAGAGGAGA    1019
  E   E   E   E   E   E   E   D   S   S   Q   K   V   R   R   V   R   R   G   D     166

CGTAATAGCG ATATTTGCAG GAGCAGCCTA CTGGTCGTAC AACGATGGCA ACGAGCCTCT    1079
  V   I   A   I   F   A   G   A   A   Y   W   S   Y   N   D   G   N   E   P   L     186

CCAAATCGTA GGCATTGCCG ACACATCCAG CCGTCGAAAT CAGGGCCGCA GCAGGAGTTA    1149
  Q   I   V   G   I   A   D   T   S   S   R   R   N   Q   G   R   S   R   S   Y     206
```

```
CCGCGTAAGA ATCCCGACCA ATTAACTAAT AATCATCTTC AGTTATATTA TAGATTTTTT     1199
  R ---------------------------- intron 2 ----------------------     207

CGTTTCTTTT ATAGTTGATT GATGGGGTAG AGATATATAC ATGTACAGCC CTTCTCTTTG     1259
---------------------------------------------------- P   F  S  L     211

GCTGGGCCAG GCTCATCATC TCGTCGTGAG GAGGGAGAAG GAAAAGGAAG AGGAATTGGG     1319
 A  G  P   G  S  S  S   R  R  E   E  G  E   G  K  G  R   G  I  G     231

AGTAATATTT TTGCAGGTTT TAGCACTCGC ACTTTGGCTG AAACATTGGG GGTGGAGATT     1379
 S  N  I   F  A  G  F   S  T  R   T  L  A   E  T  L  G   V  E  I     251

GAAACTGCAA GGAAGCTTCA AGAGAATCAG CAATCGCGAC TGTTTGCGAG GGTTGAACGG  1439
 E  T  A   R  K  L  Q   E  N  Q   Q  S  R   L  F  A  R   V  E  R     271

GGCCAACGAC TGAGCTTACC CGGCCCTCGA TCTCGCTCTC GCTCTCCTTA CGAGAGGGAG     1499
 G  Q  R   L  S  L  P   G  P  R   S  R  S   R  S  P  Y   E  R  E     291

ACTGAGAGGG ATGATGTTGC TGGTGGATTG CAGGGATATT ATTCATCTGG AGATGAGAAT     1559
 T  E  R   D  D  V  A   G  G  L   Q  G  Y   Y  S  S  G   D  E  N     311

GGCGTTGAAG AGCTTGTGTG CCCACTGCGT GTAAAGCACA ATGCTGACAA TCCCGAGGAT     1619
 G  V  E   E  L  V  C   P  L  R   V  K  H   N  A  D  N   P  E  D     331

GCCGATGTCT ACGTAAGAGA TGGGGGACGA TTGAATAGAG TCAACCGCTT CAAACTTCCT     1679
 A  D  V   Y  V  R  D   G  G  R   L  N  R   V  N  R  F   K  L  P     351

GTACTCAAGT ATTTGAGATT AGGAGCCGAG AGGGTTGTTC TCCACCCGGT AAGCAATAAC     1739
 V  L  K   Y  L  R  L   G  A  E   R  V  V   L  H  P ----------     367

TTTTTATTCG CTTCACTTAA TGTCAATTTT CAAGTCCAGT GAATGAATTA ATCTGGTTGC     1799
---------------------------------- intron 3 ----------------------

AGAGAGCATC GTGTGTTCCT TCGTGGAGGA TGAACGCGCA TGGCATAATG TACGTGACGA     1859
-- R  A  S   C  V  P   S  W  R   M  N  A  H   G  I  M   Y  V  T     386

GAGGGGAGGG GAGAATTGAG GTGGTGGGAG ACGAAGGCAG GAGCGTGTTT GATGGGCGTG     1919
R  G  E  G   R  I  E   V  V  G   D  E  G  R   S  V  F   D  G  R     406

TGAGAGAGGG TCAGTTCATC GTCATTCCCC AATTCTACGC AGTCATCAAA CAGGCAGGAG     1979
V  R  E  G   Q  F  I   V  I  P   Q  F  Y  A   V  I  K   Q  A  G     426

GCGAGGGGTT TGAGTGGATA ACGTTCACAA CATCGGACAT GTAAGTATAA CATAATTAGC     2039
G  E  G  F   E  W  I   T  F  T   T  S  D  I ----------------------     440


ATTGCACATG TCATGTACTG ATTGTTATAC TCATCATAAC TGGTATGCAT CTCAGTTCTT     2099
---------------------------------- intron 4 ----------------- S     441

TCCAGTCGTT TTTGGCGGGA AGGCAATCAG TTTTGAAGGC AATGCCGGAG GAAGTGTTGA     2159
F  Q  S  F   L  A  G   R  Q  S   V  L  K  A   M  P  E   E  V  L     461
```

```
GTGCCGCTTA CAGGATGGAC CGAACTGAAG TCCGTCAGAT TATGAGTAAC AGAGAATGCG      2219
S  A  A  Y   R  M  D   R  T  E   V  R  Q  I   M  S  N   R  E  C        481

ACACCCTCAT TCTGCCTCCA TCATCCCTTG GACGTGACCA AGAACAACAG CACAACATCA      2279
D  T  L  I   L  P  P   S  S  L   G  R  D  Q   E  Q  Q   H  N  I        501

CATCTCTTCTGCACCAAGTGGAAtagggcgtttgaatgaatattatggaataaggcgttt
T  S  L  L  H  Q  V  E  -                                            509
gaatgaatattatcgagacagctctctgcttcacgcggtgtcctgtttgcgctgcat
ggttcggcttagtagctagctacccaatattacaataaaacaatgataaggctgtaa
tagatattataataaggatgttgctttctatgtgtctacaatttcgatggaactttc
tccattatattcacatgcagctacgccctcagcgttttcgttttctccatatttcca
aattccatcccaaagttataaaacatttgacgtgatttatatagcaaactcttttca
catggagcatgtcattaatgacctgggtttgtattaatattcttatcaaattaagaa
aacactaccacatcggtcaaacattgtag.....
```

Fig. 4.4 Nucleotide sequence of genomic DNA clones S3.7 and E2.8 from *Picea glauca* 11S legumin storage protein, and deduced amino acid sequence. The +1 site and the begining of the cDNA (⇒) are indicated. Putative regulatory sequences are underlined. The aminoacid sequence is printed below the nucleotide sequence. The nucleotide and amino acid sequences corresponding to the deletion are printed in italic characters. The coding region is printed in bold characters. Introns are indicated. The aminoacid sequence is explained in fig 4.7. On the 3' non-coding region the first stop codon and the polyadenylation signals are underlined.

4.3. STRUCTURAL ORGANIZATION OF THE *PICEA* 11S LEGUMIN GENE

In order to align the cDNA to the gDNA, 4 gaps were introduced on the cDNA sequence corresponding to introns and 1 gap introduced on the gDNA that corresponds to a putative deletion. The sequence of the E2.8 plasmid contains the complete coding region of the 11S legumin gene from *Picea glauca*. The coding region is interrupted by 4 short introns of 67, 1104, 74 and 75 nt respectively. The *Picea* gene contains at least one additional intron, when compared to subfamily A legumin genes, and two more introns when compared to subfamily B legumin genes from angiosperms. The first two introns are located in the region of the gene encoding the acidic or α protein subunit, and the last two are located on the region encoding the basic or β protein subunit. Introns 1, 2 and 3 correspond in position to introns 1 to 3 of subfamily A legumin genes (figure 4.5). All four introns are flanked by canonical border sequences (fig 4.6) and no direct repeats were detected in their sequence. A remarkable feature of plant introns is their elevated AT content compared to the surrounding exons. Dicot introns have an average AT content of 72%, versus 56% in monocots. The content of A/T within the *Picea* introns is in average 68.1% (67.7, 71.1, 67.5 and 64.0% for each intron respectively).

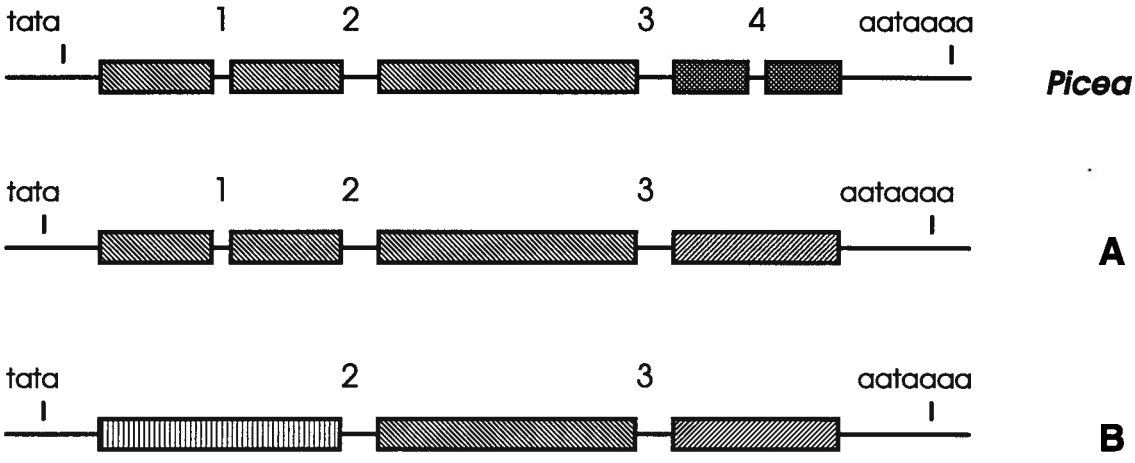


Fig 4.5 Comparison of 11S legumin genes from *Picea glauca* and angiosperm subfamilies A and B. The introns are numbered 1-4, and exons are shaded. The TATA box and polyadenylation signals (aataaaa) are indicated. The introns interrupt the coding regions at the same relative positions in each gene, but are variable in size.

Eukaryote consensus	GT$^{A}/_{G}$AG ----------N$^{C}/_{T}$AG
	INTRON 1
Picea	GTAGGT----------67 nt--------GTATGCAG
Radish	GTCCAT----------134 nt -------TTTTATAG
Soybean A	GTCCAT----------237 nt--------ATGAATAG
Soybean B	----------
Pea B	----------
	INTRON 2
Picea	GTAAGA----------104 nt-------ATGTACAG
Radish	GTAATC----------96 nt-------TGCTGTAG
Soybean A	GTGAGA- ----------282 nt-------CTTGGCAG
Soybean B	GTGAGC----------270 nt-------CTTGGCAG
Pea B	GTAAGT----------75 nt-------TTTTTCAG
	INTRON 3
Picea	GTAAGC----------74 nt-------GGTTGCAG
Radish	GTAAGT----------265 nt-------TCTTTCAG
Soybean A	GTACGT----------624 nt-------TGGTGCAG
Soybean B	GTACGT----------395 nt-------CGATGCAG
Pea B	GTACGT----------80 nt-------ATATGC
	INTRON 4
Picea	GTAAGT----------75 nt-------CATCTCAG
Radish	----------
Soybean A	----------
Soybean B	----------
Pea B	----------

Fig 4.6. Comparison of 11S legumin intron flanking sequences. The 11S legumin gene introns are flanked by canonical border sequences. The size of each intron is indicated in nucleotides. Radish and Soybean A correspond to subfamily A legumin genes and Soybean B and Pea B to subfamily B genes.

The gap introduced when aligning the cDNA to the gDNA showed a deletion that is present in the genomic DNA. The 77 nucleotide sequence is located in the region encoding the α subunit. Among the 26 deduced amino acids of this sequence, 4 are highly conserved among legumin genes but absent on the genomic DNA. The deletion shifted the sequence out of frame. No other deletions or stop signals were found in the rest of the gene. The fact that this is the only deletion on the sequence and the fact that the positions of 3 of the four introns are conserved relative to angiosperms, suggests that the deletion may be an artifact of the subcloning. A HindIII restriction site was located on the cDNA within the sequence that was absent on the genomic clone. To examine this possibility the other clone, the S3.7 plasmid DNA was tested for the presence of HindIII site (see fig 4.2). The S3.7 contains a 2.3 Kb fragment that overlaps the E2.8, however it does not have the HindIII site. The HindIII site is not present on the λ-DNA either, as corroborated by enzymatic digestion and mapping of the λ-DNA (fig 4.2). These results suggested that this could be a cloning artifact that could have arisen during the phage propagation. If this is a real deletion in the original genome and this gene is a pseudo gene this would have been the first deletion accumulated in the gene, and probably a very recent event based on the high sequence conservation with the cDNA. The possibilities of an artifact arising from cloning are difficult to explore because it would

require re-screening the library and obtaining exactly the same gene.

4.4. *PICEA GLAUCA* 11S LEGUMIN AMINO ACID SEQUENCE

Because of the deletion present on the third exon two different approaches can be taken to analyze the amino acid sequence using the genomic DNA nucleotide sequence : a) deduce the amino acid sequence from the genomic DNA that includes the deletion or b) substitute the deleted region in the genomic with the corresponding cDNA nucleotide sequence and deduce the amino acid sequence. For the purpose of this study the second approach will be used, because if this deletion is real, it would be the first accumulated deletion, which means that the rest of the gene remains intact and is therefore suitable for comparison. Also, the rest of the gene, promoter, coding region, intron positions and intron / exon junctions are consistent when compared to angiosperms. Furthermore since there are no other gymnosperm SSP gene sequences, this presents a good opportunity to study gene structure and to compare the gene and its product to genes and proteins from angiosperms. The possibility of this deletion being a cloning artifact is also to be considered.

If approach "a", is studied, the deduced amino acid sequence is truncated and the protein is 244 aminoacids in length. The 77 nucleotide deletion is in the third exon, in an area of highly conserved aminoacids. However by

substituting the deleted area with the corresponding cDNA sequence, the product of the gene is a typical 11S legumin protein comparable to other legumin proteins. The deduced amino acid sequence codes for a protein of 509 amino acids (Fig 4.7). The ATG initiation codon was determined by comparison to other legumin aminoacid sequences. The first 24 aminoacids at the N- terminus, correspond to a very hydrophobic region , the same size as those reported for GluA and GluB from rice (Takaiwa *et al*, 1991; Misra and Leal, 1993), and may represent the signal peptide. The cleavage site was estimated to lie between the alanine-24 and isoleucine-25 by comparison to the XI5H cDNA and to the Douglas fir sequences. The deduced molecular weight for the precursor is 57.18 KD and the co-translationally processed protein is 54.4 KD. The post-translational cleavage of the legumin subunit precursors into α and β polypeptides is regulated by a protease that cleaves between asparangine and glycine residues (Scott *et al*, 1992). The cleavage site for processing the precursors of 11S legumin proteins in angiosperms is highly conserved and the sequence surrounding this site is also well conserved. Recently it has been shown by cDNA sequence that the predicted cleavage site in the Douglas fir legumin-like gene is also between asparagine and glycine (Leal and Misra, 1993). The legumin gene of the gymnosperm Gingko biloba, however, has an asparagine residue at the N- terminus of the β subunit instead of glycine (Hager *et al*, 1992). By comparison to other legumin

sequences, the spruce amino acid sequence from 311 to 320 corresponds to the conserved sequence with the cleavage site between the asparagine and the glycine at positions 311 and 312, respectively. The invariant cysteine residue at position 7 of the β subunit has been shown to be involved in the formation of a disulfide bridge linking the β and α subunits of 11S legumins in angiosperms and in Gingko biloba (Hager *et al*, 1992). In the spruce 11S legumin protein the putative cysteine is located at position 318 of the precursor protein. The other cysteine that may be involved in the formation of the bridge is at position 123 of the α subunit, determined by comparison to other legumin proteins. As a result of the cleavage the α subunit is a polypeptide of 311 aminoacids with a molecular weight of 34.9 KD and the β subunit is 198 amino acids in length with a molecular weight of 22.2 KD. The amino acid composition of the deduced protein is shown in table 4.1.

MQILLIASACFLFLSLSTVSPVTAISQQRRGRGRRYDEQSSSCRRLRRLS 50

AHEPSESETIRSDGGTFELSTGEDNEELECAGVAFFRKTIESNAILLPRY 100

PSADLLLYVVRGEGRLGIVFPGCPETFRDHSSFQGRSRHRSEGRREEEEE 150

EEEDSSQKVRRVRRGDVIAIFAGAAYWSYNDGNEPLQIVGIADTSSRRNQ 200

GRSRSYRPFSLAGPGSSSRREEGEGKGRGIGSNIFAGFSTRTLAETLGVE 250

IETARKLQENQQSRLFARVERGQRLSLPGPRSRSRSPYERETERDDVAGG 300

LQYYSSGGDEN↓GVEELVCPLRVKHNADNPEDADVYVRDGGRLNRVNRFKL 350

PVLKYILRLGAERVVLHPRASCVPSWRMNAHGIMYVTRGEGRIEVVGDEGR 400

SVFDGRVREGQFIVIPQFYAVIKQAGGEGFEWITFTTSDISFQSFLAGRQ 450

SVLKAMPEEVLSAAYRMDRTEVRQIMSNRECDTLILPPSSLGRDQEQQHN 500

ITSLLHQVE 509

Fig. 4.7 Spruce 11S legumin deduced amino acid sequence. The protein sequence is in capital letters. The putative signal peptide at the N-terminus is underlined. The predicted cleavage site for the α and β subunits is shown (↓). The two cysteins predicted to be involved in the formation of disulfide bridge are circled.

Arginine	11.7%	Aspartic acid	3.9%
Serine	9.7%	Phenylalanine	3.9%
Glutamic acid	9.3%	Proline	3.7%
Glycine	9.0%	Asparangine	3.2%
Leucine	8.1%	Tyrosine	2.5
Valine	6.8%	Histidine	1.8%
Alanine	5.7%	Lysine	1.8%
Isoleucine	5.2%	Cysteine	1.6%
Glutamine	4.3%	Metionone	1.6%
Threonine	3.9%	Tryptophan	0.5%

Table 4.1 Amino acid composition of the Picea 11S legumin protein.

4.4.1. AMINO ACID COMPARISONS REVEAL PRESENCE OF HIGHLY CONSERVED SEQUENCES

The deduced amino acid sequence was compared to legumin amino acid sequences from angiosperm, monocots and dicots, and to two gymnosperm sequences. The sequence was compared to EMBL protein data banks (PCGENE release CDPROT18, Intelligenetics, Inc.CA) by using the PCOMPARE and PCLUSTAL programs (fig 4.8). The results in table 4.3 (a and b) are shown in terms of percentage of identity. The results show 69.2 % identity between spruce and Douglas fir; 68.2 % between spruce and pine and 63.9 % between D.fir and pine. The percentage of identity between the spruce and dicots varies from 28.8 % to 34.5 % and among dicots from 39.8 % within different subfamilies to 65.8 % within the same subfamilies. Similarly the identity between spruce and monocots is 31.1 % to 34.5 %. The percentage of identity between monocots and dicots varies from 35.1 % to 40.5 %.

```
LEG1_PICEA   MQILLIASACF-LFLSLS-TVSPVTAISQQRRGRGRRYDEQ----SSSCRRLRRL    49
LEG1_TSUGA   MERRMRLGSAL-LVLYFSFLVISVSSVTDAQRRSGSRYGGEDERSSRPCRRLQRL    54
LEG1_PINUS   MGK-------------EEEGVVGVTMKNHHHQRAGRQ---------------RL    26
LEG1_GOSHI   MAYTS--LLSFSVCLLVLFHGCCAQIDLVTNHHQDPPWGQPQQPQPRHQSQCQLQ    53
GLU1_ORYSA   MASINRPIVFFTVCLFLLCNGSLAQQLLG---QSTSQWQSSRRGSPR---ECRFD    49
12S1_ARATH   MARVS-SLLSFCLTLLILFHGYAAQQG---------------QQGQQFPNECQLD    39
LEGA_PEA     MAKLLALSLSFC---FLLLGGCFALR-----------------EQPQQNECQLE    34
11S3_HELAN   MASKATLLLAFTL----LFATCIAR----------------HQQRQQQQNQCQLQ    35
             **         ^              ^                        ^     ^

LEG1_PICEA   SAHEPSESETIRSDGGTFELSTGEDNEELECAGVAFFRKTIESNAILLPRYPSAD   104
LEG1_TSUGA   RAHEPSESESIRSEGGTFEFSSGEDNEELECAGVAFIRKTIESNAIAIPQYPSAP   109
LEG1_PINUS   STHEPSESESIRSEGGTFELSTEEDNEELECAGVAFFRKTIERDAMSLPQYAGAD    81
LEG1_GOSHI   NLNALQPKHRFRSEAGETEFWDQ-NEDQFQCAGVAFLRHKIQRKGLLLPSFTSAP   107
GLU1_ORYSA   RLQAFEPIRSVRSQAGTTEFFDV-SNEQFQCTGVSVVRRVIEPRGLLLPHYTNGA   103
12S1_ARATH   QLNALEPSHVLKSEAGRIEVWDH-HAPQLRCSGVSFARYIIESKGLYLPSFFNTA    93
LEGA_PEA     RLDALEPDNRIESEGGLIETWNP-NNKQFRCAGVALSRATLQRNALRRPYYSNAP    88
11S3_HELAN   NIEALEPIEVIQAEAGVTEIWDA-YDQQFQCA-----WSILFDTGFNLVAFSCLP    84
             ^ ^^         ^^^^  *  ^   ^ ^^ *^     ^  ^   ^^    ^   ^

LEG1_PICEA   LLLYVVRGEGRLGIVFPGCPETFRDHSSFQGRSRHRSEGRREEEEEEEEDSSQKVR   165
LEG1_TSUGA   ELVYVARGEGRVGIVFPGCPETFREDSSFRGRSCRRSEGRREEEDKEE-DSSQKVR   164
LEG1_PINUS   LLLYVVRGEGRLGIVFPGCPETYREDEPSFQGRRSRRRSSERRGEEEDEDSSQKVR   137
LEG1_GOSHI   MLFYVEQGEGIHGAVFPGCPETYQSQSQQNIQD--RPQ---------R-DQHQKLR   151
GLU1_ORYSA   SLVYIIQGRGITGPTFPGCPESYQQQFQQSGQA--QLTESQSQSQKFK-DEHQKIH   156
12S1_ARATH   KLSFVAKGRGLMGKVIPGCAETFQDSSEF---Q--PRFEGQGQSQRFR-DMHQKVE   143
LEGA_PEA     QEIFIQQGNGYFGMVFPGCPETFEEPQESE----------QGEGRRYR-DRHQKVN   133
11S3_HELAN   TSTPLFWPSSREGVILPGCRRTYEYSQEQQFSG--EGGRRGGGEGTFR-TVIRKLE   137
                  ^    ^  * ^^^^^  ^^                              ^  ^^^

LEG1_PICEA   RVRRGDVIAIFAGAAYWSYNDGNEPLQIVGIADTSSRRNQGRSRSYRPFSLAGPG   215
LEG1_TSUGA   RVRRGDVIAIFAGAAHWWYNDGNEPLQLIAIAHTASPHNQLGRRSYRPFSLAGPA   219
LEG1_PINUS   RIRRGDVIAIFAGAAYWSYNDGNEPLQIVAIADTSNPQNQNR-RDYRSFPLAGPA   191
LEG1_GOSHI   RLKEGDVVALPAGVAHWIFNNGRSQLVLVALVDVGNDANQLDENF-RKFFLAGSP   205
GLU1_ORYSA   RFRQGDVIALPAGVAHWCYNDGEVPVVAIYVTDLNNGANQLDPRQ-RDFLLAGN-   209
12S1_ARATH   HIRSGDTIATTPGVAQWFYNDGQQPLVIVSVFDLASHQNQLDRNP-RPFYLAGNN   197
LEGA_PEA     RFREGDIIAVPTGIVFWMYNDQDTPVIAVSLTDIRSSNNQLDQMP-RRFYLAGN-   186
11S3_HELAN   NLKEGDVVAIPTGTAHWLHNDGNTELVVVFL-DTQNHENQLDENQ-RRFFLAGNP   190
             ^ ^ **^^*  ^* ^ *  *^    ^  ^ ^ ^  ^  ***^    * * ***

LEG1_PICEA   SSSR-------------------REEGEGKGRGIGSNIFAGFSTRTLAETLGVEI   251
LEG1_TSUGA   SGSSSRSP---------------REEGEGERRDIGGNILAGFSTRSLAETLGVEL   259
LEG1_PINUS   SSSGGGG-------------RREEEGEEERGRRGVNNILAGFNTRTLAQSLDVEL   233
LEG1_GOSHI   QGGVVRGGQSRDRNQRQSRTQRGEREEEESQESGGNNVLSGFRDNLLAQAFGIDT   260
GLU1_ORYSA   -----------KRNPQAYRREVEERSQ---------NIFSGFSTELLSEALGVSG   244
12S1_ARATH   PQGQVW-------------------LQGREQQPQKNIFNGFGPEVIAQALKIDL   232
LEGA_PEA     -------------HEQEFLQYQHQQGGKQEQENEGNNIFSGFKRDYLEDAFNVNR   228
11S3_HELAN   QA-QAQSQQQQQRQPRQQSPQRQRQRQRQGQGQNAGNIFNGFTPELIAQSFNVDQ   244
                                                 *^^^^^     ^^^^^ ^^
```

```
LEG1_PICEA   ETARKLQ---ENQQSRLFARVERGQRLSLPG------------------------   279
LEG1_TSUGA   ETARKLQ---QNQRSRLFARVEQGRRLSLPG------------------------   287
LEG1_PINUS   ETARRLQ---QNQHSRFFARVERGRRLSLPA------------------------   262
LEG1_GOSHI   RLARKLQ--NERDNRGAIVRMEHGFEWPEE-------------------------   288
GLU1_ORYSA   QVARQLQ--CQNDQRGEIVRVEHGLSLLQP-------------------------   272
12S1_ARATH   QTAQQLQ--NQDDNRGNIVRVQGPFGVIRP-------------------------   260
LEGA_PEA     HIVDRLQGRNEDEEKGAIVKVKGGLSIISPPEKQARHQRGSRQEEDEDEEKQPRH   283
11S3_HELAN   ETAQKLQG--QNDQRGHIVNVGQDLQIVRPPQDR--------------------   276
               ^ ^**   ^ ^ ^  ^^^^

LEG1_PICEA   -----------------PRSRSRSPYERETERDDVAGGLQYYSSGGDENGVEELV   317
LEG1_TSUGA   ---------------------PARSGQRDNEMMQQLHETHNSFANENENDVEEVV   321
LEG1_PINUS-------------------PRSRSRS-------RSPYAGRQRQWGREDSENGVEELV   293
LEG1_GOSHI   --GQRRQGREEEGEEEREPKWQRRQESQEEGSEEEEREERGRGRRRSGNGLEETF   341
GLU1_ORYSA   --YASLQEQEQGQVQSRERY-------------QEGQYQQSQYGSGCSNGLDETF   312
12S1_ARATH   -----------------PLRGQRPQEEEEEEGR----HGRH------GNGLEETI   288
LEGA_PEA     QRGSRQEEEEDEDEERQPRHQRRRGEEEEEDKKERGGSQKGKSRRQGDNGLEETV   338
11S3_HELAN   -------------------RSPRQQQEQATSPRQQQEQQQGRRGGWSNGVEETI   311

                                                    ^       ^******^

LEG1_PICEA   CPLRVKHNADNPEDADVYVRDGGRLNRVNRFKLPVLKYLRLGAERVVLHPRASCV   372
LEG1_TSUGA   CALRVKHNADNPEDADIYVRDGGRMNIVNRFKLPVLKYLGLGAERVILRQRASTA   376
LEG1_PINUS   CPMRVKHNADNPEDADVYVRDGGRMNIVNRYKLPALKYLGLGAERVILPGRASFV   348
LEG1_GOSHI   CSMRLKHRTPASS-ADVFNPRGGRITTVNSFNLPILQYLQLSAERGVLYNNAIYA   395
GLU1_ORYSA   CTLRVRQNIDNPNRADTYNPRAGRVTNLNTQNFPILSLVQMSAVKVNLYQNALLS   367
12S1_ARATH   CSARCTDNLDDPSRADVYKPQLGYISTLNSYDLPILRFIRLSALRGSIRQNAMVL   343
LEGA_PEA     CTAKLRLNIGPSSSPDIYNPEAGRIKTVTSLDLPVLRWLKLSAEHGSLHKNAMFV   393
11S3_HELAN   CSMKFKVNIDNPSQADFVNPQAGSIANLNSFKFPILEHLRLSVERGELRPNAIQS   366
             *^ ^   ^   ^^ ^*   ^  * ^  ^^  ^^**^  ^ ^^^ ^  ^  ^*

LEG1_PICEA   PSWRMNAHGIMYVTRGEGRIEVVGDEGRSVFDGRVREGQFIVIPQFYAVIKQAGG   427
LEG1_TSUGA   PSWRMNAHGIMYVTRGEGRIEVVGEQGRSLFDGRVREGQFIVIPQFHAVIKQAGD   431
LEG1_PINUS   PSWRMNAHAIMYVTRGEGRIEVVGDEGRSVFDGRVKEGQFIVIPQFYAVVKQAGE   403
LEG1_GOSHI   PHWNMNAHSIVYITRGNGRIQIVSENGEAIFDEQVERGQVITVPQNHAVVKKAGR   450
GLU1_ORYSA   PFWNINAHSVVYITQGRARVQVVNNNGKTVFNGELRRGQLLIIPQHYAVVKKAQR   422
12S1_ARATH   PQWNANANAILYETDGEAQIQIVNDNGNRVFDGQVSQGQLIAVPQGFSVVKRATS   398
LEGA_PEA     PHYNLNANSIIYALKGRARLQVVNCNGNTVFDGELEAGRALTVPQNYAVAAKSLS   448
11S3_HELAN   PHWTINAHNLLYVTEGALRVQIVDNQGNSVFDNELREGQVVVIPQNFAVIKRANE   421
             * ^  *******   *  ^^^^^^ **  ^*^^ ^  ** ^ ^**  **  **

LEG1_PICEA   EGFEWITFTTSDISFQSFLAGRQSVLKAMPEEVLSAAYRMDRTEVRQIMSNRECD   482
LEG1_TSUGA   DGLEWITFTTSDASVRSSLAGRESVLKAMPEDVVSAAYRMDRNEVREVMRNREDD   486
LEG1_PINUS   DGLEYITFTTSDNSYRSTLAARQSVLKRRCRGSVACGLQNRPKRSPSVMRNREHD   456
LEG1_GOSHI   RGFEWIAFKTNANAKISQIAGRVSIMRGLPVDVLANSFGISREEAMRLKHNR-QE   504
GLU1_ORYSA   EGCAYIAFKTNPNSMVSHIAGKSSIFRALPNDVLANAYRISREEAQRLKHNRGDE   477
12S1_ARATH   NRFQWVEFKTNANAQINTLAGRTSVLRGLPLEVITNGFQISPEEARRVKFNTLET   453
LEGA_PEA     DRFSYVAFKTNDRAGIARLAGTSSVINNLPLDVVAATFNLQRNEARQLKSN----   499
11S3_HELAN   QGSRWVSFKTNDNAMIANLAGRVSASAASPLTLWANRYQLSREEAQQLKFSQRET   476
                ^^^^^*^  ^  ^ ^**  *   ^ * ^^ ^^ ^ ^ ^^*^  ^^ ^
```

```
LEG1_PICEA   TLILPPSS----LGRDQEQQHNITSLLHQVE        509
LEG1_TSUGA   TLILPPSP----- RHQRDIE---S-RVQVE        507
LEG1_PINUS   TLILPPSSSLSGSGRYQDQQQNVTSLLVQVA        488
LEG1_GOSHI   VSVFSP--------R-QGSQQ----------        516
GLU1_ORYSA   FGAFTPIQYKSY----QDVYNAAESS-----        499
12S1_ARATH   TLTHSSGP--ASYGRP---RVAAA-------        472
LEGA_PEA     ----NPFKFLVPA-RESENRASA--------        517
11S3_HELAN   VLFAPSFS------RGQGIRASR--------        493
                   ^
```

Fig. 4.8. Amino acid alignment of 11S legumin proteins from *Picea*, *Pseudotsuga*, *Pinus strobus*, cotton (goshi), oat(orysa), *Arabidopsis*, pea, sunflower (helianthinin). The alignment was done on 8 protein sequences using the CLUSTAL PCGENE. Character to show that a position in the alignment is perfectly conserved: '*'; Character to show that a position is well conserved: '^'.

PERCENTAGE OF IDENTITY AT THE AMINO ACID LEVEL

a)

	Spruce	D. Fir	Pine
Spruce	—	69.2%	68.2%
D. Fir	—	—	63.9%

b)

	Soy2	PeaA	Cot2	Arab12	Rice1	Rice2
Spruce	29.7%	28.8%	33.7%	34.5%	34.5%	31.1%
Soy 2	—	65.8%	41.2%	39.8%	37.1%	37.5%
PeaA		—	41.2%	41.3%	35.1%	36.1%
Cot 2			—	44.1%	40.5%	36.9%
Arab12					39.6%	38.6%
Rice 1	—					62.9%

Table 4.2 Percentage of identity at amino acid level between a) *Picea glauca* and other gymnosperms; b) *Picea* versus dicots and monocots.

4.5. THE *PICEA GLAUCA* LEGUMIN PROMOTER REGION

The restriction map and the sequence of the S3.7 indicate that this clone contains 1.4Kb of 5' flanking region. The cap site was determined by primer extension using a 27mer primer that was constructed using 27 nucleotides at the beginning of the cDNA. mRNA from three different embryo developmental stages were assayed. The scanning for eukaryotic promoter elements using EUKPROM, PCGENE program showed the position of the start site and TATA box, being the same as determined by primer extension. The start site (+1) was determined to be 97 nucleotides upstream of the ATG (fig 4.9 and 4.10) and 35 nucleotides downstream from the putative TATA box (fig 4.10).

4.5.1. PUTATIVE REGULATORY SEQUENCES

Sequences responsible for the regulation of legumin gene expression have been described in angiosperms and it has been shown that the first 600 nt of the 5' flanking region are sufficient to give a level of regulated expression almost equal to that of the complete promoter (Shirsat *et al*, 1989; Itoh, *et al*, 1993). In order to investigate the presence of similar elements that could be regulating gene expression in spruce, 450 nucleotides of the promoter region were sequenced. Sequences homologous to known conserved motifs were found in the 5' flanking

sequence and the results are shown on figure 4.10 and table 4.3. The cap site and TATA box were determined by scaning for eukaryotic promoter elements using the PCGENE program. The putative TATA box is located at 35 bp from the cap site. One ACACA element, that has been described as an important element for seed specific expression on albumin genes, was found at position -208. Two legumin boxes with the consensus sequence CATGCAT, or RY repeat, are located at -40 and -87, respectively, and are similar to those present in the upstream region of other seed storage protein genes (Depigny-This, *et al*, 1992). Close to the RY-repeat at position -110 one ABRE element (ABA regulatory element) was found and other two ABRE elements are present at -138 and -160. These ABRE elements were first described in wheat (Marcotte *et al*, 1989) and rice (Mundy J. *et al*, 1990) and have been shown to bind transcriptional factors. The ABRE consensus sequence is included within the G-Box sequence. Four G-Box elements are present on this sequence, three of which contain ABRE elements at positions -132, -137 and -109 and another one at -159 position. An AGATGT element, recently identified in the sunflower promoter region that binds nuclear proteins (Nunberg *et al*, 1994) is present at position -216 of the Picea promoter . AGATGT elements are thought to be involved in binding nuclear factors and enhancing expression of sunflower helianthinin. An A/T rich region is present within the -223 to -340 nucleotides. A/T rich regions have been implicated in the binding of nuclear

factors (Meakin and Gatehouse, 1991; Itoh, *et al*, 1993; Nunberg *et al*, 1994).

Due to the high similarity to promoter elements described in angiosperm SSP genes, these various conserved sequences are likely to be involved in Picea SSP regulatory functions which have yet to be characterized. However, the expression of a specific gene depends on a combination of regulatory elements and a specific complement of trans-acting factors. Table 4.3 shows the possible roles of *Picea* elements compared to elements in other ssp genes.

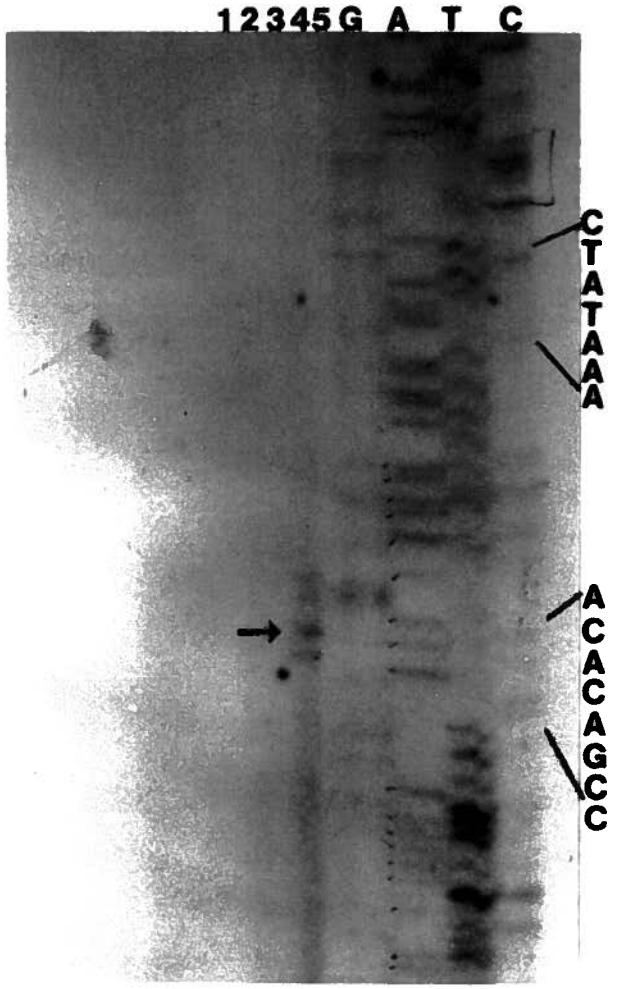

1 2 3 4 5 G A T C
CTATAAA
ACACAGCC

Fig.4.9. Determination of the transcription start site (+1) by primer extension. mRNAs from spruce mature embryo (lanes3-5) and pre-embryo (lane 2) stages were assayed. No RNA (lane 1) was used as negative control. The arrow indicates the +1 site, and the sequence is shown at the right of the S3.7 DNA sequence. The TATA box sequence is also shown.

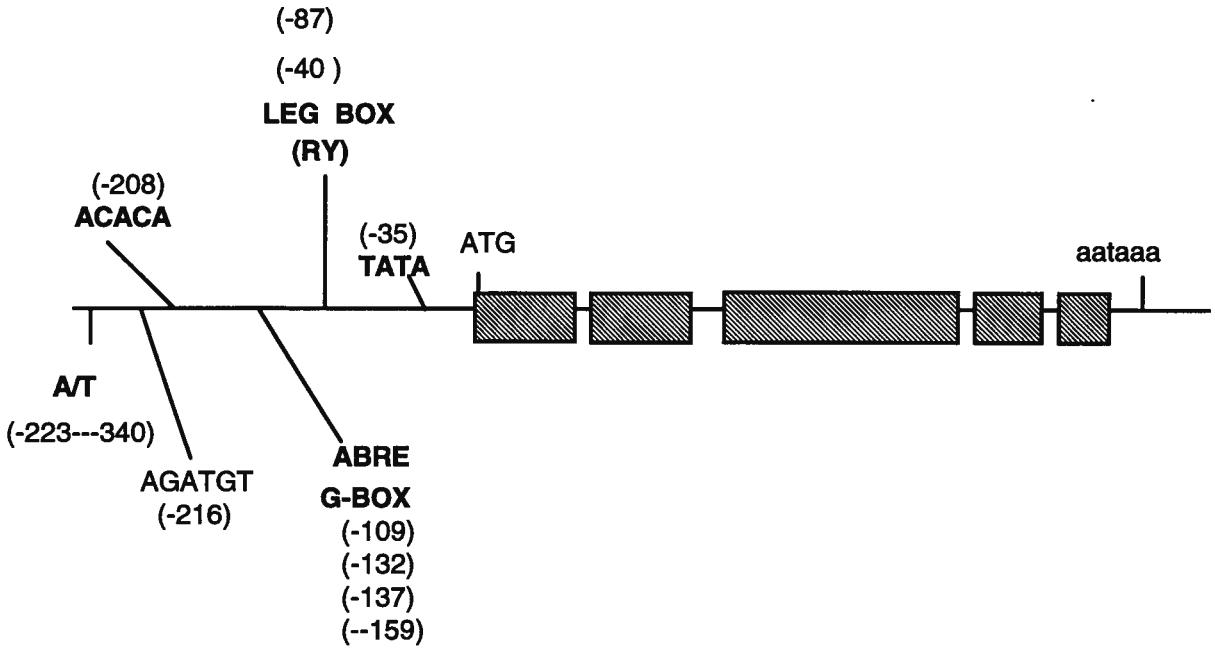


Fig. 4.10. Location of putative regulatory sequences on the promoter of *Picea* 11S legumin gene. Numbers on the figure are relative to the cap site. Abbreviations are referred to in the text.

PUTATIVE REGULATORY SEQUENCES ON THE *PICEA* LEGUMIN PROMOTER

CONSENSUS SEQUENCE	POSITION IN *PICEA*	FUNCTION	BINDING FACTOR	REFERENCE
TATA	-35	TATA-binding basal transcription	TFIID1,2	Chamberland, *et al* (1992)
LEG BOX RY repeat CATGCAT	-40 -87	Tissue specificity Enhancer of expression		Lelievre, *et al*, (1992) Chamberland, *et al* (1992)
ABRE/ G-BOX ACACGTG ACACGTCG ACACGACA	-109 -132 -137 -159	ABA responsive element and Expression enhancer	EmB-1 GCBT-1	Mundy *et al*, (1990) Guiltinan, *et al* (1990)
ACACA	-208	Albumin seed specific expression		
AGATGT	-216	Tissue specificity Legumin expression enhancer	Seed nuclear proteins from soybean	Nunberg *et al*, (1994)
A/T	-223 to -340	Seed specific expression enhancer	LABF1 Nuclear factors of early embryo genesis	Meakin and Gatehouse, (1991) Itoh, *et al* (1993)

Fig 4.3 Putative regulatory sequences in the *Picea glauca* 11S legumin promoter. Consensus sequence and positions refer to the *P. glauca* promoter sequence. The function and binding factor columns refer to other seed storage protein promoter sequences, from the literature.

Chapter 5
DISCUSSION

The coding region of the XI5H-1 *Picea glauca* 11S legumin gene characterized in the present study exhibits 98.7% homology with the cDNA clone XI5H, used to isolate this genomic DNA. The important features in the coding region have been indicated in figure 4.4. Comparison of the genomic and the cDNA sequences suggests the presence of four introns (fig 4.4 and 4.5) and a short deletion within the coding region of the *Picea* 11S legumin gene.

Based on amino acid identity the existence of two legumin subfamilies (A and B) has been shown in angiosperms. Each subfamily has a characteristic number of introns: this number is two for B and three for A (Boutler *et al*, 1987). The position of the two introns of subfamily B genes correspond to the position of the second and third introns of subfamily A genes. The position of the introns is highly conserved among legumin genes as are the surrounding sequences (Galau *et al*, 1991). This study shows that the 11S legumin gene from *Picea* has one or two extra introns compared to angiosperm genes. The position of introns one to three correspond to the position of one to three introns of subfamily A. This result demonstrates that the conservation of intron positions is extended to gymnosperms. Since this is the first report of a genomic

clone from a legumin seed storage protein gene from a gymnosperm, the possibility that this gene is a member of a different subfamily, based on the number of introns is speculated, however the possibility of *Picea* containing type A or B genes can not be excluded. Evenmore, the conservation of intron positions (one to three) suggests that these genes may have derived from a common ancestor.

The possibility that the *Picea* 11S legumin gene belongs to a different subfamily is also suggested by the percentage of homology at the amino acid level with angiosperm sequences. The percentage of identity between subfamilies A and B in angiosperms of the same or different species is about 40%. The percentage of identity among members of subfamily A, of same or different species is 65% to 85%. The percentage of identity among dicots is higher than it is between dicots and monocots. The analysis of the predicted amino acid sequences from this study shows that the percentage of identity between *Picea* and dicots or *Picea* and monocots is similar (30-34%). Since the percentage of identity between *Picea* and subfamily A is not different than that between *Picea* and subfamily B, this could suggest that they belong to three different subfamilies. Nevertheless the differences may be the result of evolution. The percentage of identity among the three gymnosperms *Picea, Pinus* and *Pseudotsuga* is around 70%. This high percentage suggests that they may belong to the same subfamily. However no data is available at the genomic

DNA level for *Pinus* and *Pseudotsuga*, so comparisons regarding intron number can not be made.

Homologous regions among legumin amino acid sequences are dispersed throughout the molecules, suggesting that the similarity is due to divergence from a common ancestor (Negoro *et al* 1985). The ancient origin of legumin genes is confirmed by the existence of homologous sequences expressed in the spores of some fern species (Templeman *et al*, 1988). Furthermore, the presence of legumin proteins in gymnosperms such as *Gingko biloba* and conifer species supports the hypothesis of a common ancestor. On the basis of structural and immunological criteria it has been speculated that the genes encoding legumin proteins in monocots and dicots have evolved from a common ancestor gene (Negoro *et al* 1985; Borroto and Dure 1987).

Recently molecular evolution data strongly suggest that the separation of monocot and dicot lineages took place in late Carboniferous (300 million years ago). The divergence time of conifers from angiosperms is estimated of 330 million years ago, and the earliest seed plants are of upper Devonian-Lower Carboniferous age (360 million years ago)(Martin *et al*, 1993). The results of this study show that the legumin gene from *Picea glauca* is structurally similar to angiosperm genes. Even though the number of introns is different, the position of three of the four introns is conserved among the gymnosperm *Picea glauca* and

angiosperms. The results also show that at the amino acid level the protein shares common regions with a 30-34% of identity to angiosperm legumin proteins and 70% of identity with other gymnosperms. This is the first report that shows that the organization of the legumin gene is highly similar between conifers and angiosperms. The notable conservation of genes and proteins support the hypothesis of a common ancestor.

The extent of conservation among 11S proteins may be based on functional constraints to evolutionary divergence, i.e., the postranscriptional processing event including the disulfide bond formation, the proteolytic cleavage, the signal for moving through the ER, the information for moving to the Golgi apparatus and the effective accumulation of the proteins in protein bodies (Borroto and Dure, 1987; Negoro *et al* 1985).

Besides introns, another difference found between the genomic clone and the cDNA is a 77 bp deletion. The deletion resides in the third exon, a highly conserved region. The deletion moves the sequence out of frame, therefore the predicted amino acid product would be a truncated protein. The deletion may be a cloning artifact. However this possibility is difficult to explore because it would be necessary to re-screen the library and obtain the same particular gene. The probability of including a specific DNA sequence in a library with a known genome size

is described by the formula N=ln(1-p)/ln(1-f), where p is the probability of containing any particular DNA sequence, f is the fraction of the genome represented by each fragment and N corresponds to the number of cloned fragments and, therefore the necessary number of recombinant clones. For *Picea glauca* the number of recombinants required to have any given DNA sequence is 2.30×10^6. Another important consideration is that legumin genes belong to multigene families. The number of gene family members for *Picea* has not been characterized. Nevertheless results from this study show the presence of at least two members of the legumin family (fig 4.1). It is important to mention that unsuccessful efforts were made to determine the number of the members of the legumin family.

However, because the deletion is the only one found within the *Picea* sequence, and it is not bordered by direct repeats or palindromic sequences, the possibility of it being a cloning artifact has to be considered. In animals, bacteria and plant genomes it has been shown that there is a close correlation between direct repeats and naturally occurring deletions upon cloning into *E.coli*, and led to the authors to favor 'slipped mispairing' as a mechanism to explain deletion formation (Heim *et al*, 1989). Another feature that suggests that this deletion is a cloning artifact is that the rest of the gene has no other modifications and the promoter region appears to be intact. The amino acid identity, if the deletion is substituted by

the corresponding sequence from the cDNA, is very high within gymnosperms (70%) and around 35% with angiosperms. The intron positions and the intron border sequences, also insinuate that the gene is functional, since they observe the rules for eukaryotic intron/exon flanking sequences. However the possibility of this being the first deletion event accumulated in the *Picea* 11S legumin gene, therefore a very recent pseudogene can not entirely discounted. Nevertheless, even if the deletion is real, since it is probably the first deletion occurring in this gene, to judge from the high sequence homology to the cDNA, the rest of the gene would still conserve the original characteristics and can be used for comparisons. The results of this present work are important in the light of the structural organization of the legumin gene since there is no data, other than that presented here, regarding genomic SSP DNA in gymnosperms. Since the cDNA sequence is available (C. Newton, unpublished results) the deleted sequence was substituted by the cDNA corresponding sequence for the purpose of comparing to other legumin genes. All the amino acid comparisons were performed after translating the gDNA with the substituted sequence.

Other important features found in the *Picea* glauca 11S legumin gene are found in the promoter region (Fig 4.4 and Table 4.3). The results of this present work show the presence of putative regulatory sequences in the 5'

flanking sequence of the gene. The presence of basic transcriptional sequences such as the TATA box, and the presence of specific sequences like the highly conserved legumin box, ABRE elements, ACACA elements, G-boxes and A/T rich regions, support the hypothesis that the *Picea* 11S legumin gene may be functional and raise the possibility that gymnosperms have similar regulatory sequences to angiosperms. The presence in *Picea* of various sequences that are highly conserved and important for transcription in angiosperm SSP genes, strongly suggest a transcriptional role. Some of these putative sequences have been determined as enhancers or binding elements for transcriptional factors in variety of seed specific genes (Weissing and Kahl, 1991). The presence of these conserved sequences suggests that the gene is functional. The objectives of this thesis do not comprise the functional characterization of the promoter region. However it would be very interesting to test whether the promoter sequence is functional or not. The functionality of the promoter could be examined by dissecting the gene promoter, fusing it to a reporter gene, and in a transgenic system determine if it directs the expression of the gene. Recent advances in genetic transformation of plants have made possible the transfer of chimeric genes into conifer genomes (Duchense and Charest, 1991). The somatic embryogenesis system developed for conifers at B.C.Research and the use of microprojectile DNA-delivery may overcome some of the

limitations for transgenic tree recovery (Ellis *et al*, 1991).

The results of the present study showed that all the structural elements of the *Picea glauca* 11S legumin gene are similar to angiosperms, with the exception of an additional intron. The extra intron and the amino acid identity suggest that the *Picea* gene belongs to a legumin subfamily different than angiosperm subfamilies A or B. Another possibility is that angiosperm genes lost one or two introns during evolution and the *Picea* gene is more similar to the ancestral gene. The conservation of legumin genes from fern species to seed plants (gymnosperms and angiosperms) strongly suggest an evolutionary relationship that may have significant consequences in regards the application of genetic technology largely developed in dicot angiosperms (i.e., tobacco, Arabidopsis) to economically important species such as conifers. The promoter region also exhibits all putative regulatory elements associated with angiosperm SSPs. While it is not possible to rule out the presence of regulatory domains unique to gymnosperms, it is likely that *cis*-acting sequences are common between gymnosperms and angiosperms. This study implies that tissue specific regulation could be achieved in transgenic conifers using angiosperm promoters.

Chapter 6

REFERENCES

Argos P., Narayana S.V.L. and Nielsen N.C. (1985) Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO J. 4:1111-1117.*

Allona I. Casado R. Aragoncillo C. (1992) Seed storage protein from *Pinus pinaster* Ait.: homology of major components with 11S proteins from angiosperms. *Plant Sci. 87:9-18.*

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

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

Bewley J.D. and Black M. (1985) Seeds: germination, structure and composition. Edited by J.D.Bewley and M. Black Plenum Press New York, pp 1-28.

Bollini R. and Chrispeels M.J. (1979) Characterization and subcellular localization of vicilin and phytohemaglutinin, the two major reserve proteins of *phaseolus vulgaris. Planta 142:291-298.*

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

Bostock R.M. and Quatrano R.S. (1992) Regulation of *Em* gene expression in rice. *Plant Physiol. 98:1356-1363.*

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

Breen J.P. and Crouch M.L. (1992). Molecular analysis of a cruciferin storage protein gene family of *Brassica napus. Plant Mol.Biol. 19:1049-1055.*

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

Casey R., Domoney C. Ellis N, and Turner S. (1988) The structure expression and arrangement of legumin genes in pea. *Bioch.Phys.Pflanz. 183: 173-180.*

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

Craig S. and Millerd A. (1981) Pea seed storage proteins: Immunocytochemical localization with prot A-gold by electron microscopy. *Protoplasma 105:333-339.*

De Pace,C.,Delre,V., Scarascia Mugnozza, G.T., Maggini, F., Cremonini, R., Frediani, M. and Cionini P.G. (1991) Legumin of Vicia faba major: accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. *Theor Appl Genet 83:17-23.*

Depigny-This D., Raynal M. Aspart L., Delsney M. and Grellet F. (1992) The cruciferin gene family in radish. *Plant Mol. Biol. 20:467-479.*

Dickinson C.D. Evans R.P. aand Nielsen N.C. (1988) RY repeats are conserved in the 5' flanking sequence of legumin protein genes. *Nuc. Acids Res. 16:371*

Domoney C. and Casey R. (1985) Measurements of gene number for seed storage protein genes in *Pisum. Nucl.Acids Res. 13:687-699.*

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

Dure III (1988) Characteristics of storage proteins of cotton. JAOCS 66:356-359.

Duchesne L.C. and Charest P.J. (1991) Transient expression of the β-glucoronidase gene in embryogenic callus of *Picea mariana* following microinjection. *Plant Cell Rep 10:191-194*

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

Ellis D.D., McCabe D.,RussellD.,Martinell B. and McCown B.H. (1991) Expression of inducible angiosoerm promoters in a gymnosperm Picea glauca. *Plant Mol Biol 17:19-27.*

Eicson M.L. Muren E., Gustavsson H.O., Josefsson L.G and Rask 1. (1991) Analysis of the prmoter region of napin genes fropm *Brassica napus* demonstrates binding of nuclear proteins *in vitro* to a conserved motif. *Eur J.Bioch. 197:741-746.*

Finkelstein R.R. Tenbarge K.M. Shumwat J.E. and Crouch M.L. (1985) Role of ABA in maturation of rapeseed embryos. *Plant Physiol. 78:630-636.*

Flinn B.S., Roberts D.R. Webb D.T. and Sutton B.C.S. (1991a) Storage protein changes during zygotic embryogenesis in interior spruce. *Tree Physiol. 8:71-81.*

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

Flinn B.S. Roberts D.R. Newton C.H. Cyr D.R. Webster F.B. Taylor I.E.P. (1993) Storage protein gene expression in zygotic and somatic embryos of interior spruce. *Phys. Plant. 89:719-730*

Galau G.A., Wang H.Y.C. and Hughes W. (1991). Sequence of the *Gossypium hirsutum* D-genome alloallele of legumin A and its mRNA. *Plant Physiol. 97:1268-1270.*

Gibbs P.E.M., Strongin K.B. and McPherson A. (1989) Evolution of legume seed storage proteins - A domain common to legumins and vicilins is duplicated in vicilins. *Mol.Biol.Evol. 6:614-623.*

Gifford D.J. (1988) An electrophoretic analysis of the seed proteins from *Pinus monticola* and eight other species of pine. *Can J.Bot. 66:1808-1812.*

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

Green M.J. McLeod J.K. and Misra S. (1991). Characterization of Douglas fir protein body composition by SDS-PAGE and electron microscopy. *Plant.Physiol.Biochem. 29:49-55.*

Gultinan M.J., Marcotte W.R. and Quatrano R.S. (1990) A plant leucine zipper protein that recognizes an absicic acid responnse element. *Science 250:267-270*

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

Hakman I., Stabel P. Engstrom P. and Erikson T. (1990) Storage protein accumulation during zygotic and somatic embryo development in *Picea abies* (Norway spruce). *Physiol Plant. 80: 441-445.*

Heim U., Schubert R., Baumlein H. Wobus U. (1989) The legumin gene family: structure and evolutionary implications of *Vicia faba* B-type genes and pseudogenes. *Plant Mol.Biol. 13:653-663.*

Itoh Y., Kitamura Y., Arahira M. and Fukazawa C. (1993) *Cis*-acting regulatory regions of the soybean seed storage 11S globulin gene and their interactions with seed embryo factors. *Plant Mol.Biol. 21:973-984.*

Jensen U. and Lixue C. (1991) *Abies* seed protein profile divergent from other *Pinaceae. Taxon 40:435-440.*

Jensen U. and Berthold H. (1989) Legumin-like proteins in gymnosperms. *Phytochem. 28:1389-1394.*

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

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

Kitamura Y., Arahiara M., Itoh Y. and Fukazawa C. (1990) The complete nucleotide sequence of soybean glicinin A_2B_1 gene. *Nuc. Acids Res. 18:4245.*

Kuhlemeir C., Green P.J. and Chua N. (1987) Regulation of gene expression in higher plants. *Ann Rev Plant Physiol. 38:221-257*

Leal I. and Misra S. (1993) Molecular cloning and characterization of a legumin-like storage protein cDNA of Douglas fir seeds. *Plant Mol.Biol. 21:708-715*

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

Maniatis T. Fritsch E.F. and Sarnbrook J. (1982) Molecular clonnig. A lab mannual. Cold spring Harbor Lab Press. N.Y.

Martin W., Lydiate D. and Brinkmann H. (1993) Molecular phylogenies in angiosperm evolution. *Mol.Biol.Evol. 10:140-162.*

Marcotte W.R Bayley C.C. and Quatrano R.S. (1988) Regulation of a wheat promoter by ABA in rice protoplasts. *Nature 335:4543-457.*

Marcotte W.R., Russell S.L. and Quatrano R.S. (1989) ABA responsive sequences from the Em gene of wheat. *Plant Cell 1:969-976.*

Meakin P.J. and Gatehouse J.A. (1991) Interaction of seed nuclear proteins with transcriptionally enhancing region of pea leg A gene promoter. *Planta 183:471-477.*

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 megagamethophyte of white spruce (*Picea glauca* (Moench) Voss.) *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

megagamethophyte of white spruce (*Picea glauca* (Moench) Voss.) *Plant Sci.78:61-71.*

Mundy J. and Chau N.(1988) ABA and water stress induce the expression of a novel rice gene. *EMBO J 7:2279-86.*

Muntz, K. (1989) Intracellular protein sorting and the formationof protein reserves in storage tissue cells of plant seeds. *Bioch Physiol Pflanzen 185:315-335.*

Negoro T., Momma T. and Fukazawa C. (1985) A cDNA clone encoding a glycinin A1a subunit precursor of soybean. *Nuc.Acids Res. 13:6719-6731.*

Newton C.H., Flinn B.S. and Sutton B.C.S. (1992) Vicilin-like seed storage proteins in the gymnospem interior spruce (*Picea glauca/englemanii*). *Plant Mol.Biol. 20:315-322.*

Nielsen N. (1986). Structural relationships between 7S and 11S legume globulins. in Molecular biology of seed storage proteins. Edited by Shannon L. and Chrispeels M.

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

Nunberg A.N., Zhuwen L., Bogue M.A. Vivekananda J., Reddy A.V. and Thomas T. (1994) Developmental and hormonal regulation of sunflower helianthinin genes: proximal promoter sequence confer regionalized seed expression. Plant Cell 6:473-486.

Pang P.P., Pruit R.E. and Meyerowitz E.M. (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage proteins of *Arabidopsis thaliana. Plant Mol.Biol. 11:805-820.*

Pich U. and Schubert I. (1993) Polymorphism of legumin genes in inbred lines of *Vicia faba. Biol.Zent.bl. 112:342-350.*

Plietz P., Damaschun G., Muller J.J. and Schlesier B. (1983) In "The biochemistry of plants. Molecular biology". (Stumfp P.K. and Conn E.E. Editors, Vol 15) pp297-345.

Riggs, D.C. Voelker, T.A. and Chispeels, M.J. (1989) Cotyledon nuclear proteins bind to DNA fragments harboring

regulatory elements of phytohhemagglutinin genes. *Plant Cell 1:609-621.*

Roberts,D.R., Flinn B.S., Webb D.T., Webster F.B. and Sutton B.C.S. (1990) ABA and IBA regulation of maturation and accumulation of storage proyteins in somatic embryos of interior spruce. *Physiol. Plant 78:355-360.*

Redenbaugh K. Paasch B.D., Nichol S.W., Kossler M., Viss P.R. and Walker K.A. (1986) Somatic seeds: Encapsulation of asexual plant embryos. *Bio-tech 4:797-781.*

Rodin J., Sjodahal S., Josefsson L. and Rask L. (1992) Characterization of *Brassica napus* gene encoding a cruciferin subunit: estimation of sizes of cruciferin gene families. *Plant Mol.Biol. 20:559-563.*

Scott M.P., Jung R., Muntz K. and Nielsen N.C. (1992) A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein in soybean. *Proc.Natl.Acad.Sci. USA 89:659-662.*

Shirsat, A.H., Meakin, P.J. and Gatehouse, J.A. (1990) Sequences 5` to the conserved 28bp Leg box element regulate the expression of pea seed storage protein gene *leg A. Plant Mol. Biol. 15:685-693.*

Shirsat A., Wilford N., Croy, R. and Boulter, D. (1989) Sequeneces responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol Gen Genet 215: 326-331.*

Shotwell M.A. and Larkins B.A. (1989) The biochemistry and molecular biology of seed storage proteins in Marcus A. The biochemistry of plants vol. 15 Acad. Press. Inc.

Shotwell M.A., Boyer S.K., Chestnut R.S. and Larkins B.A. (1990) Analysis of seed storage proteins of oat. *J.Biol.Chem. 265-9652-9658.*

Sims T.L. and Goldberg R.B. (1989). The glycinin Gy1 gene from soybean. *Nuc.Acids Res. 17:4386.*

Skriver K. and Mundy J. (1990) Gene expression in response to ABA and osmotic stress. *Plant Cell 2:503-512.*

Stabel P., Erikson T. and Engstrom, P. (1990) Cganges in protein synthesis upon cytokinin-mediated adventitious bud

induction and during seedling development in Norway spruce *Picea abies. Plant Physiol. 92: 1174-1183.*

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

Templeman T.S., DeMaggio A.E. and Stetler D.A. (1987) Biochemistry of fern spore germination: globulin storage protein in *Matteuccia struthiopteris* L. *Pl.Physiol. 85:343-349.*

Templeman T.S.,Stein D.B. and DeMaggio A.E. (1988). A fern spore storage protein is genetically similar to the 1.7S seed storage protein of *Brassica napus. Biochem.Genet. 26:595-603.*

Templeman T.S. and DeMaggio A.E.(1990) Biochemistry of fern spore proteins:globulin storage proteins in *Onclea sensibilis* and *Osmunda cinnamomea. Amer. J.Bot. 77:284-287.*

Thompson G.A. and Larkins B.A. (1989) Structurale elements regulating zein gene expression. *BioEssays 10:108-113.*

Turner L., Helllens R.P., Lee D. and Ellis T.H.N. (1993) Genetic aspects of the organization of legumin genes in pea. *Plant Mol.Biol. 22 101-112.*

Wang C., Shastri K., Wen L., Huang J. Sonthayanon B., Muthukrishnan S. and Reeck G.R. (1987) . Heterogeneity in cDNA clones encoding rice glutelin. *FEBS L. 222:135-138.*

Weissing K. and Kahl G. (1991) Towards an Understanding of plant Gene regulation. The action of nuclear factors. *Nature 46:1-11.*

Wilen R.W., Mandel R.M., Pharis R.P., Holbrook L.A. and Moloney M .M. (1990) Effects of abscicic acid and high osmoticum on storage protein gene expression in microspore embryos of *Brassica napus. Plant Physiol. 94:875-881.*