University of Alberta

Characterization of the agglutinin domain gene family in flax (*Linum usitatissimum*)

by

Kashfia Faruque

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Plant Biology

Department of Biological Sciences

Fall 2013
Edmonton, Alberta

Abstract

Lectins are proteins that bind specifically to carbohydrates and are active in plant defense and other processes. The sequencing of the flax (*Linum usitatissimum*) genome showed that it was unusually rich in genes predicted to encode one or more agglutinin domains (PFAM PF07468). This domain is characteristic of the amaranthin-type lectin family, thus the 19 predicted flax genes were named LuALLs (*Linum usitatissimum* amaranthin-like lectins). To investigate their functions, transcript expression of 19 LuALLs was measured using qRT-PCR. Some LuALLs were enriched in specific tissues, such as developing seeds, while most LuALLs were expressed at various levels throughout the vegetative and floral tissues assayed. Distinct clades of LuALLs were found to be inducible by either salicylic acid or methyl jasmonate, exclusively, consistent with a role for these genes in plant defense. Preliminary hemagglutination assays with a recombinant LuALL provided evidence that members of this family function as genuine lectins.

Acknowledgement

I would like to express my heartfelt thanks and gratitude to my supervisor, Dr. Michael Deyholos, at first for taking me in his lab and providing me with fresh opportunity to complete my masters. Throughout my research work he has provided me patient guidance, encouragement and outstanding advices, helping me out whenever I got stuck, at the same time providing me time to think and do the work. He responded to my queries and concerns very promptly and without his help, it was hard to complete the thesis work and write up in a short time.

I would also like to thank my supervisory committee, Dr. Stephen Strelkov and Dr. Christine Szymanski, for their advices and encouragement.

I would also like to thank all my lab mates. Especial thanks to Mary De Pauw for guiding me first on how to do qRT-PCR. Also to Ryan McKenzie for providing me the qRT-PCR internal control primers and showing me how to calculate the data. To the undergrad student, Kayla Lindenback for designing eight of the qRT-PCR primers in this study. A big thank you to Shanjida Khan and Haiyan for their help and suggestions during my protein extraction and purification work. To Joshua Mcdill for showing me how to generate phylogenetic tree.

Also I would like to thank Troy Locke from MBSU, for his suggestions on qRT-PCR.

I am also thankful to the department of Biological Sciences for all the support, especially during less favourable conditions.

I would like to thank the funding agencies, Genome Canada, Prairie and Alberta; and my supervisor for the financial support of the project.

Last but not the least; I would like to thank my family and friends for always supporting and encouraging me.

Contents

List of Tables

List of Figures

List of Abbreviations

RIP	ribosome inactivating protein
Gal	galactose
GalNAc	N-acetylgalactosamine
ER	endoplasmic reticulum
CRD	carbohydrate-recognition domains
GlcNAc	N-acetyglucosamine
ABA	*Agaricus biscorpus* agglutinin
EEA	*Euonymus europaeus* agglutinin
GNA	*Galanthus nivalis* agglutinin
NICTABA	*Nicotiana tabacum* agglutinin
EGFP	Enhanced Green Fluorescent Protein
CDS	coding sequence
cTP	chloroplast transit peptide
mTP	mitochondrial targeting peptide
SP	signal peptide
SVM	support vector machine
qRT-PCR	Quantitative real time RT-PCR
ANOVA	analysis of variance
SDS	sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis

TEMED	tetramethylethylenediamine
DTT	dithiothreitol
CD	Circular dichroism
HU	hemagglutination unit
PR	Pathogenesis Related
Ct	Threshold cycle
MeJA	methyl jasmonate
SA	salicylic acid

Chapter 1 Introduction

1.1 Lectins

The first lectin, ricin, was discovered at the end of nineteenth century from the seeds of castor beans (*Ricinus communis*) and was observed to aggregate red blood cells (Sillmark, 1888). Hence, the term 'hemagglutinin' was introduced to describe similar proteins from plants that have this ability (Elfstrand, 1898). Later, it was observed that some of these hemagglutinins could selectively distinguish between human erythrocytes of different blood types (Boyd & Regura, 1949; Renkonen 1948). Thus the term 'lectin' was coined, from the Latin verb 'legere' meaning 'to select' (Boyd, 1954). Furthermore, it was observed that these proteins are capable of agglutinating cells, other than the erythrocytes (Peumans & Van Damme, 1995). Afterwards, the basis of this agglutination/ hemagglutination property was found to be its selective binding to specific carbohydrate structures present on the cell membrane (Watkins & Morgan, 1952).

1.2 Biochemical definition of lectins

Lectins are a heterogeneous group of proteins of non-immune origin that are known to have at least one non-catalytic domain, allowing them to selectively recognize and reversibly bind to specific free sugars or glycans present on glycoproteins and glycolipids without altering the properties of the carbohydrates (Peumans & Van Damme, 1995). Lectins with two or more carbohydrate binding sites have the ability to agglutinate cells/glycoconjugates (Kumar *et al.,* 2012) and thus can be called agglutinins. Even though the term lectin is most commonly used, the terms agglutinins and hemagglutinins are still used synonymously (Van Damme, Lannoo & Peumans, 2008).

Lectins have a higher affinity for complex carbohydrates (in the 10^{-6} to 10^{-8} M range) as compared to simple sugars (in the 10^{-3} to 10^{-4} M range), due to the increased number of molecular interactions between the proteins and larger carbohydrates (Vandenborre *et al.*, 2008; Zhao, Patwa, Lubman & Simeone, 2008). Lectin-carbohydrate interactions are usually weaker than antigen-antibody complexes (Zhao *et al.*, 2008). Although lectins were first discovered in plants, they are now well-known to be widely present in other domains of life, from viruses, fungi, and bacteria to animals. (Lannoo & Van Damme, 2010).

1.3 Ricin: The first lectin discovered

The toxicity of ricin has been known since ancient times. Stillmark in 1888 extracted this toxic protein from seeds of castor beans (*R. communis*) and named it ricin. The protein preparation could agglutinate red blood cells and was thus thought to be toxic (Lord, Robert & Robertus, 1994). Later it was found that the protein preparation of Stillmark had a mixture of a potent cytotoxin (ricin) and a hemagglutinin (*Ricinus communis* agglutinin, RCA) (Lord *et al.*, 1994). It is now known that the toxicity of ricin is due to its catalytic action on eukaryotic ribosomes (Olsnes, Refsnes & Pihl, 1974).

Other than ricin, several other cytotoxic proteins from plants have been isolated. These irreversibly inhibit protein synthesis by inactivating ribosomes and are called ribosome inactivating proteins (RIP) (Lord, Hartley & Roberts, 1991). These proteins are frequently N-glycosylated and are usually monomers of 30 kDa (type I RIPs). Type I RIPs are not cytotoxic (as they cannot enter eukaryotic cells) and are widely consumed by humans and animals. Examples of type I RIPs include proteins of wheat-germ and barley grains. However, type II RIPs are heterodimeric toxins in which a monomeric RIP

is joined by a disulphide bond to a galactose (Gal) binding lectin polypeptide (around 30 kDa). These type II RIPs can therefore bind to cell surface galactosides, enter the cytosol, inhibit protein synthesis and subsequently cause cell death (Lord *et al.,* 1991).

Ricin is a type II RIP composed of a ribosome inactivating enzyme (called A chain, 32 kDa) and a galactose/N-acetylgalactosamine (GalNAc) binding lectin (called B chain, 34 kDa) joined by a disulphide bond. Different isoforms of ricin and RCA together constitute about 5 % of the total protein present in mature *R. communis* seeds (Lord *et al.*, 1994).

Ricin and its homologue, RCA, are synthesized in the endosperm cells of maturing seeds and are stored in protein storage vacuoles (previously called protein bodies). Within a few days of germination, the toxins are destroyed by hydrolysis (Lord *et al.,* 1994).

Ricin is synthesized as a preproprotein, containing both the A and B chains (Lord & Roberts, 1993). The signal sequence at the amino terminal directs the protein to the endoplasmic reticulum (ER) where it is cleaved off by the ER luminal signal peptidase. Proricin is N-glycosylated as it enters the ER lumen. Protein disulfide isomerases catalyze formation of five disulphide bonds as the proricin molecule is folded. The B chain has four intrachain disulphide bonds and is joined to the A chain by another disulphide bond. Proricin undergoes further modifications of its oligosaccharides within the Golgi complex and then is transported in vesicles to the protein storage vacuoles (Lord, 1985a, b; Lord *et al.*, 1994). Ricin becomes activated in the protein storage vacuoles by an endopeptidase, which cleaves and liberates two chains that remain linked together by a single disulphide bond because cleavage occurs within a disulfide loop con-

necting the A and B chain sequences. This proteolytic processing step involves the removal of a 12-residue peptide that links the mature A and B chain sequences (Lord *et al.,* 1994).

Ricin, with the help of its lectin domain (B chain), binds to galactose-containing glycoproteins and glycolipids present on the surface of cells, facilitating its entry into the cytosol. The A chain is an N-glycosidase that removes a single adenine residue from the 28S ribosomal RNA loop contained within the 60S subunit. This irreversibly inactivates eukaryotic ribosomes, inhibiting protein synthesis (Audi, Belson, Patel, Schier & Osterloh, 2005).

The structure of ricin is shown as a ribbon drawing in Figure 1.1 (Lord *et al.,* 1994). The A chain (braided ribbon on the upper right) is a 267-residue globular protein. It has eight alpha helices and eight beta sheets. The substrate binding site is the cleft marked by the substrate adenine ring. The B chain (solid ribbon on the lower left) is a 262-residue protein. It is dumbbell shaped with galactose binding sites at both ends (depicted by lactose rings). These two sites allow hydrogen bonding to the membrane sugars, galactose and N-acetyl galactosamine. A disulfide bridge (-S-S-) joins chain A with B (Lord *et al.,* 1994).

Figure 1.1: The backbone of ricin. The A chain is in the upper right as a braided ribbon. The adenine ring marks the active site cleft. The B chain is the solid ribbon. Lactose is shown binding at each end of the two domain peptide (Lord *et al.,* 1994).

1.4 Classification of lectins

1.4.1 Animal lectins

Animal lectins can be classified into a number of structurally distinct families based on their carbohydrate-recognition domains (CRDs; the protein module within the lectin possessing the carbohydrate binding activity, Gupta, Gupta, & Gupta, 2009). Of the eight well-established groups, four contain predominantly intracellular lectins and four contain extracellular lectins. The intracellular lectins (i.e., calnexin family, M-type, L-

type and P-type) are located in luminal compartments of the secretory pathway and function in the trafficking, sorting and targeting of maturing glycoproteins. The extracellular lectins (i.e. C-type, R-type, siglecs and galectins) are either secreted into the extracellular matrix or body fluids, or are localized to plasma membranes, and mediate a range of functions including cell adhesion, cell signaling, glycoprotein clearance and pathogen recognition (Gupta *et al.*, 2009).

1.4.2 Bacterial lectins

Bacterial lectins can be grouped into two classes: (1) lectins (adhesins) that reside on the bacterial surface and facilitate bacterial adhesion and colonization and (2) secreted bacterial toxins (Varki, Etzler, Cummings & Esko, 2009).

As a prerequisite to infection, microorganisms typically need to adhere to the host tissue. Bacteria express lectins that interact with glycoproteins, proteoglycans and glycolipids present on the surface of the host tissues (Gupta, 2012). For example, many bacterial surface lectins are present in the form of long, hairy appendages known as fimbriae or pili that extend away from the cell. The presence of multiple glycan-binding subunits in the fimbriae allows multivalent interactions with the host cell, thus facilitating binding (Varki *et al.*, 2009).

Many secreted bacterial toxins also bind to glycans. For example the soil bacterium *Bacillus thuringiensis* produces crystal toxins (Bt toxins). Although the toxin can kill larval stages of some plant herviborous insects, the toxin is harmless to most other organisms, including humans. Bt toxins act by binding to glycolipids that line the gut in

nematodes and presumably other invertebrates. The glycolipid receptors include in their structure the characteristic ceramide-linked, mannose-containing core tetrasaccharide GalNAcβ1–4GlcNAcβ1–3Manβ1–4GlcβCer, which is found only in invertebrates and is conserved between nematodes and insects but is absent in vertebrates. Thus, the specificity of Bt toxins determines their tissue and species sensitivity (Esko & Sharon, 2009).

1.4.3 Plant lectins

Plant lectins can be divided into four major groups (Van Damme, Peumans, Barre & Rougé, 1998; Van Damme *et al.,* 2008) on the basis of their overall domain structure: merolectins, hololectins, chimerolectins and superlectins. Merolectins are monovalent lectins, consisting of a single carbohydrate-binding domain. They are therefore incapable of agglutinating cells or precipitating glycoconjugates, unless they form a higher oligomer. Examples of merolectins include monomeric mannose binding lectins from orchid (Van Damme, Balzarini, Smeets, Van Leuven & Peumans, 1994) and hevein, which is the chitin binding domain containing protein from the latex of rubber plants (Van Parijs *et al.,* 1991). Hololectins comprise two or more identical or highly similar carbohydrate binding domains that can bind the same or structurally similar sugars, e.g., wheat germ agglutinin, which has four chitin binding domains (Wright, Sandrasegaram & Wright, 1991). These di- or multivalent lectins are thus capable of agglutinating cells and/or precipitating glycoconjugates. Like hololectins, superlectins have at least two carbohydrate binding domains, but they recognize structurally unrelated sugars. An example of a superlectin is tulip lectin, TxLC-I which consists of a mannose-binding domain and an unrelated GalNAc-binding domain (Van Damme *et al.,* 1996).

Chimerolectins consist of one or more carbohydrate binding domains fused with a domain that acts independently of the lectin domains, e.g., class I chitinases, which have a chitin binding domain linked to a catalytically active chitinase domain (Peumans & Van Damme, 1995). Recent genome and transcriptome analyses in plants have revealed that the chimerolectins are the most abundant type. Thus, most lectins are bi- or multi-functional proteins as opposed to exclusively carbohydrate binding proteins (Van Damme *et al.*, 2008).

One of several attempts to classify plant lectins was based on their carbohydrate binding specificity. Based on this criterion, mannose-, mannose/glucose-, mannose/maltose-, Gal/GalNAc-, (N-aetyl glucosamine, GlcNAc)/ $(GlcNAc)_n$, fucose-, and sialic acid binding lectins have been distinguished (Goldstein and Poretz, 1986; Van Damme *et al.*, 1998). However, this system allows evolutionarily unrelated lectins to be grouped together (Jiang, Zhigang & Ramachandran, 2010), since different carbohydrate-binding motifs may recognize similar sugar structures.

Van Damme *et al*., in 2008 classified lectins taking into consideration all available sequence data. In this system, plant lectins were grouped into twelve distinct families of evolutionary and structurally related lectin domains, after a careful and detailed analysis of available genome and transcriptome data (as shown in Table 1.1).

Table 1.1: The twelve plant lectin families (according to Van Damme *et al.,* 2008) and their carbohydrate specificity .

Lectin domain	Specificity	Reference
Agaricus biscorpus agglutinin (ABA)	T antigen (Galβ1,3GalNAc)	Nakamura-Tsuruta, Kominami, Kuno & Hirabayashi, 2006
Amaranthins	T antigen (Galβ1,3GalNAc)	Rinderle, Goldstain, Matta & Ratcliffe, 1989
Class V chitinase homologs	High mannose N-glycans	Van Damme *et al.,* 2007a
Cyanovirin family	High mannose N-glycans	Bolmstedt, O'Keefe,Shenoy,Mcmahon & Boyd, 2001
Euonymus europaeus agglutinin (EEA)	Blood group B oligosaccharides, high mannose N-glycans	Teneberg, Alsén, Ångstörm, Winter & Goldstein, 2003; Yamamoto & Sakai, 1981; Fouquaert *et al.,* 2008
Galanthus nivalis agglutinin (GNA)	Mannose, oligomannosides, high mannose N-glycans, complex N-glycans	Van Damme, Allen & Peumans, 1987; Van Damme *et al.,* 2009; Van Damme *et al.,* 2007b
Hevein	(High mannose and/or complex) N –glycans, Chitin	Van Damme *et al.,* 2009
Jacalins	Galactose, T-antigen, Mannose	Peumans *et al.,* 2000
Legume lectin	Mannose/glucose, galactose, GalNAc, L-Fucose, complex N-glycans, sialic acid α (2,3) galactose	Ramose *et al.,* 2000; Konami, Yamamoto, Osawa & Irimura, 1994
LysM	Chitooligosaccharides	Mulder, Lefebvre, Cullimore & Imberty, 2006
Nicotiana tabacum agglutinin (NICTABA)	GlcNAc oligomers, high man N-glycans	Lannoo *et al.,* 2006
Ricin-B	Gal/GalNAc, Siaα2–6Gal/GalNAc	Candy *et al.,* 2001; Shibuya, 1987

Each lectin domain defined in Table 1.1 has its own characteristic fold with one or more carbohydrate binding sites. Most of these domains are widespread throughout taxa in the plant kingdom, and are present in other domains of life as well, as shown by Table 1.2.

Table 1.2: Overall taxonomical distribution of the carbohydrate binding domains found in embryophyta (Van Damme *et al.*, 2008).

Overall Taxonomical Distribution of the Carbohydrate Binding Domains Found in Embryophyta[a]

				Viridiplantae						
					Embryophyta					
Carbohydrate binding domain	Bacteria	Fungi	Animals	Chlorophyta	Liverworts	Lycophytes	Mosses	Ferns	Gymnosperms	Angiosperms
ABA	–[b]	+	–	–	+	–	(+)	–	–	–
Amaranthin	–	–	–	–	–	+	–	–	+	+
CRA	–	–	–	–	–	–	–	–	–	+
Cyanovirin	+	+	–	–	+	+	–	+	–	–
EUL	–	–	–	–	+	+	+	+	+	+
GNA	+	+	+	–	+	+	+	+	+	+
Hevein	–	+	–	+	+	+	+	+	+	+
Jacalins	+	+	+	–	–	+	+	+	+	+
Legume lectin	+	+	+	+	+	+	+	+	+	+
LysM	+	+	+	+	+	+	+	+	+	+
Nictaba	–	–	–	–	+	+	+	+	+	+
Ricin-B	+	+	+	–	+	–	–	–	–	+

[a]Based on the data deposited in the following databases: Same databases as in the legend to Table I; and some additional databases:
Chlamydomonas reinhardtii: http://genome.jgi-psf.org/Chlre3/ (v3.0 assembly)
Ostreococcus lucimarinus: http://genome.jgi-psf.org/Ost9901_3/ (Assembly v.2.0)
Ostreococcus tauri: http://genome.jgi-psf.org/Ostta4/ (v.2.0 assembly)
Volvox carteri: http://genome.jgi-psf.org/Volca1/ (Assembly v1.0)
[b]Symbols: –, absent; +, present; (+), possibly present.

A drawback of Van Damme's classification system is that it is restricted to the plant lectin information that was available at the time, and does not include animal lectin-like members in plant kingdom (Jiang *et al.*, 2010). A new and different classification system was more recently proposed by Jiang *et al.*, 2010 based on the whole genome identification and characterization of the lectin families in rice, Arabidopsis and soybean. Based on the domain structure and phylogenetic analysis, they grouped the lectins from these higher plants into twelve families (Table 1.3) and named most of them according to their domain description in the Pfam database. They detected four new lectin families: Calreticulin, Gal_binding_lectin, Gal_lectin and Lectin_C, which were not included in the classification by Van Damme *et al.*, 2008. Conversely, Jiang *et al* did not include four lectin families (Cyanovirin, *Agaricus bisporus* agglutinin (ABA), Amaranthin and Chitinase-related agglutinin (CRA)) that were present in the Van Damme classification system. Jiang *et al.*, found only limited examples of these four excluded lectin families in

the genomes surveyed and in the InterPro database whereas their twelve families were ubiquitous in higher plants. Thus, they concluded that their system is effective for classifying lectins in higher plants but is not applicable to rare, lineage-specific lectins (Jiang *et al.*, 2010).

Table 1.3: Comparison of Van Damme *et al.*, 2008 and Jiang *et al.*, 2010 classification system of lectin families.

Eight of the families in the Jiang *et al.* sytem, matched with their corresponding families in the Van Damme *et al*., classification system, and are shown side by side. The four lectin families that differ in each of the classification systems are shown in red.

Van Damme *et al.*, system	Jiang *et al.*, system
GNA	B-lectin
Legume lectin	Lectin_legB
Jacalins	Jacalin
Hevein	Chitin_bind_1
Ricin-B	Ricin_B_lectin
EEA	EEA
LysM	LysM
NICTABA	Phloem
Cyanovirin	Calreticulin
ABA	Gal_binding_lectin
Amaranthin	Gal_lectin
class V chitinase homologs / Chitinase-related agglutinin	Lectin_C

Classical plant lectins are expressed abundantly in seeds or vegetative storage tissues such as tubers, bulbs, rhizomes or bark and are generally considered to be storage or defense related proteins (Van Driessche, 1988; Cammne, Peeters & Peumans, 1985; Peumans, Allen & Cammue, 1986; Van Damme, Goldstein & Peumans, 1991; Peumans, Nsimba-Lubaki, Peeters & Broekaert, 1985; Broekaert, Nsimba-Lubaki, Peeters &

Peumans, 1984). For example, *Phaseolus vulgaris* agglutinin (PHA), the vacuolar seed lectin, comprises about 10% of the total seed protein (Pusztai & Watt, 1974; Tague & Chrispeels, 1987). Classical lectins usually are synthesized independently of external environmental conditions. Most of these lectins are secretory proteins and are localized to the vacuoles or the extracellular spaces (Etzler, 1985; Etzler, 1986).

Unlike classical lectins, inducible lectins are expressed in non-storage tissues such as leaves, roots or flowers (Van Damme *et al.,* 2011). Most of these lectins are expressed at relatively low levels (Van Damme, Barre, Rougé & Peumans, 2004) in response to certain biotic or abiotic stresses such as insect herbivory, wounding, drought, cold, high salt concentrations or on treatment with different hormones like ABA, jasmonic acid, giberellins and salicylic acid. This group of lectins is mainly localized to the cytoplasmic/nuclear compartment of the cell. For example, the expression of the tobacco lectin, NICTABA, was detected only after methyl jasmonate treatment or herbivory by different insect species (Chen *et al.,* 2002; Vandenborre *et al.,* 2009a, b).

1.5 Structure of lectin domains

The availability of purified plant lectins has facilitated their study and structural analysis. Plant lectins are divided into six broad classes based on their three dimensional structures (Bettler, Loris & Imberty, 2001; online data available at www.cermav.cnrs.fr/lectines). These are legume lectins, α-D-mannose-specific plant lectin (monocot lectin), β-prism lectins, β-trefoil lectins, Cyanovirin N homolog and agglutinin with hevein domain. Structures of some other families are also summarized here.

The structure of many lectins from the legume family has been determined. For the most part, leguminous lectins assemble into a compact β-barrel configuration devoid of α helices and are dominated by two antiparallel pleated sheets (Varki *et al.*, 1999). The binding site for carbohydrate in most leguminous lectins involves a combination of H bonds, hydrophobic interactions, and van der Waals contacts. Nearby subsites on the proteins assist in binding oligosaccharides and contribute to hydrophobic interactions with aglycon moieties. In leguminous lectins, the metal-binding sites are located on a single long loop. The metals near the binding site do not make direct contact with the sugar but help stabilize amino acid side chains required for binding. Even in cocrystals between the lectin and a more complex glycan ligand, the lectin makes contact primarily with a single monosaccharide substituent. The plant lectins appear to acquire high affinity via their multivalency in dimeric and tetrameric forms (Varki *et al.*, 1999). The legume lectin Con A is a tetramer (Figure 1.2). It consists of a dome shaped monomer built up from two β-sheets. The curved front face and the flat back face have seven and six antiparallel strands of β-sheet respectively, interconnected by turns and loops (Hardman & Ainsworth, 1972; Edelman, Cunningham, Reeke, Becker, Waxdal & Wang, 1972). A third, smaller β-sheet made of five short strands of β-sheet, also referred as the S-sheet (Banerjee, Das, Ravishankar, Suguna, Surolia & Vijayan, 1996), helps to keep together the front and back sheets.

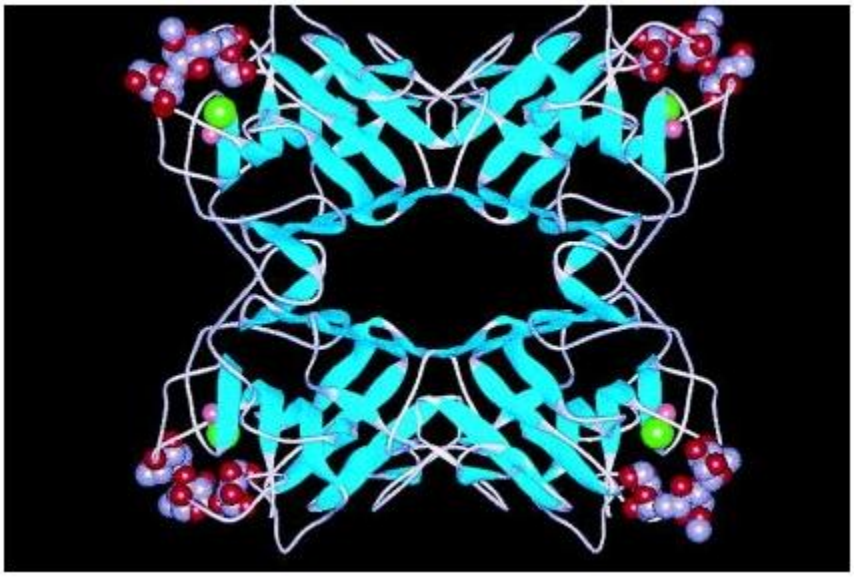

Figure 1.2: Structure of tetrameric ConA at 2.35 Å. The trimannoside ligand is indicated in space-filling mode and the coordinated Ca^{2+} and Mn^{2+} are shown as the large green balls and small pink balls, respectively. The crystal structure was originally reported as a complex of ConA and a trimannoside ligand by Naismith & Field, 1996 (Modified, with permission, from Loris *et al.*, 1996 by Varki *et al.*, 1998)

In contrast to legume lectins that interact with both mannose and glucose, the monocot mannose-binding lectins (e.g., GNA) react exclusively with mannose and mannose-containing N-glycans (Barre, Bourne, Van Damme, Peumans & Rougé, 2001). GNA is a tetramer of four covalently associated monomers (Figure 1.3). The GNA monomer consists of three four-stranded antiparallel β sheets, referred to as subdomains, connected in loops to form a triangular shaped prism structure (Hester, Kaku, Goldstein & Wright, 1995). X-ray analysis revealed that GNA binds mannose with the help of hydrogen bonds and hydrophobic interactions (Hester *et al.*, 1995).

Figure 1.3: GNA tetramer resulting from hydrophobic interaction of two dimers A-D (red-purple) and B-C (green-orange). Each dimer consists of two twelve stranded monomers linked by hydrogen bonds and C-terminal strand exchange. Stars (★) show the monosaccharide-binding sites (Barre *et al.*, 2001).

Each of the four subunits in jacalin is made of a major α-chain of 133 amino acids and a minor β-chain of 20 amino acids (Raval, Gowda, Singh & Chandra, 2004). The crystal structure of jacalin indicates that each of its subunits exhibits a type I β-prism fold, comprised of three Greek keys (four-stranded β-sheets) contributed by both the chains (Figure 1.4) (Sankaranarayanan, Sekar, Banerjee, Sharma, Surolia & Vijayan, 1996). The crystal structures of other lectins in this family (artocarpin from *Artocarpus integrifolia* (Pratap *et al*., 2002); *Helianthus tuberosus* lectin (heltuba; Bourne *et al*., 1999); *Maclura pomifera* agglutinin (MPA; Lee e*t al*., 1998); and *Calystegia sepium* lectin (Calpsa; Bourne *et al.*, 2004) confirmed this fold to be characteristic of the family, although significant differences in quaternary associations were observed. These crystal structures were also consistent with one carbohydrate-binding site per subunit. Residues forming the binding site emerge from different loops at one end of the prism.

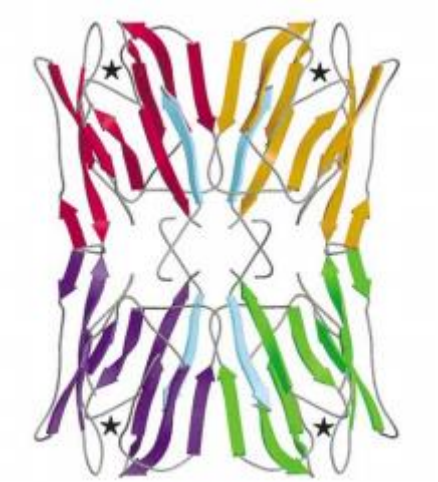

Figure 1.4: Jacalin tetramer made of four monomers, each consisting of an α-chain (orange, green, purple, red) associated to a β-chain (cyan). Stars (★) show the monosaccharide-binding sites (Barre *et al.,* 2001).

The ricin B chain of the cytotoxic protein has no regular secondary structure but displays several omega loops (Rutenber & Robertus, 1991). Ricin B has a single polypeptide chain that is comprised of two tandemly arranged globular domains with identical folding topologies (Montfort *et al.,* 1987) Each domain can be further divided into three subdomains (α, β and γ) with a fold consisting of two two-stranded hairpins and involving a representative QXW motif. The three subdomains assemble around a pseudo-threefold axis in a similar manner to the β-trefoil fold (Rutenber & Robertus, 1991; Murzin *et al.,* 1992). The crystal structure of ricin demonstrates that subdomains 1α and 2γ retain the major sugar-binding sites for galactose (Rutenber & Robertus, 1991); 1β may also have minor sugar-binding activity (Frankel *et al.,* 1996).

Other than the B-chain of Ricin, the β-trefoil family also includes Kunitz type trypsin inhibitor, amaranthin, and others (Transue *et al.,* 1997). Despite structural similarity, the

amaranthin domains have no sequence similarity to other β-trefoil proteins, and the QXW motif is absent (Wright, 1997). Amaranthin (*Amaranthus caudutus* agglutinin, ACA) is a dimeric T-antigen specific seed lectin isolated from amaranth grain (Mr=66,000) (Transue *et al.,* 1997). The two identical subunits of ACA are related by non-crystallographic two fold symmetry and are packed in a head to tail arrangement (Figure 1.5). Each of the subunits contains two homologous globular domains, designated N and C domains, linked by a short 3_{10}-helix. Each domain has the characteristic β-trefoil fold. The β-trefoil domain consists of three homologous motifs arranged about a pseudo-three-fold axis to form a six stranded anti parallel β-barrel capped at one end by three β-hairpins. The specific head-to-tail arrangement of the two amaranthin domains is necessary to establish the carbohydrate-binding site of the lectin (Figure 1.5) (Transue *et al.,* 1997).

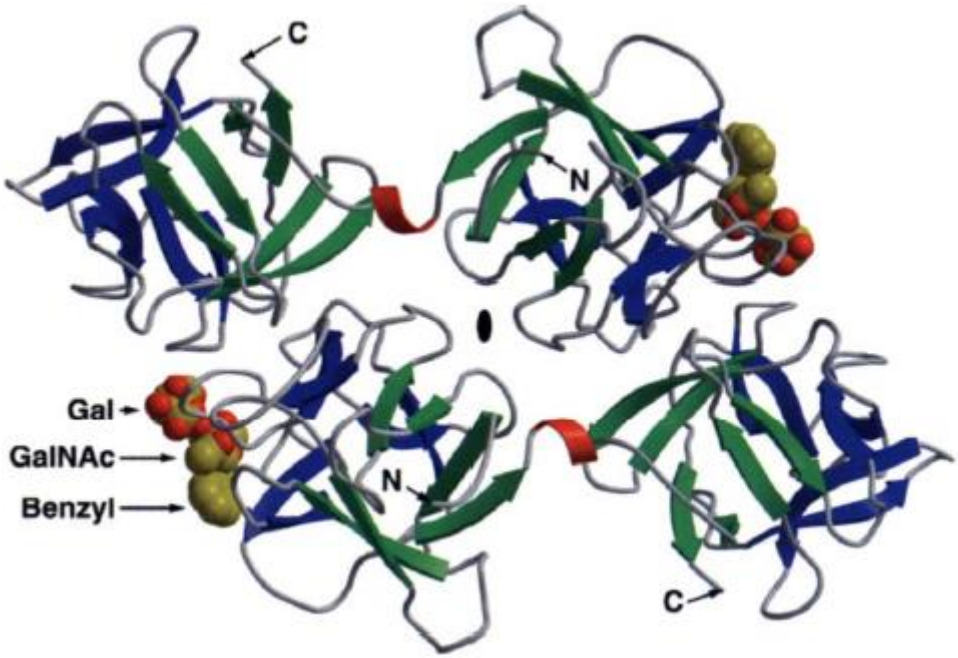

Figure 1.5: Structure of ACA dimer showing location of the two equivalent binding sites for T-antigen disaccharide (Gal-β 1,3-GalNAc-α-O-benzyl). The dimer is approximately

100 Å x 70 Å x 50Å and is comprised of two subunits related by non-crystallographic dyad (oval). Each monomer contains two β-trefoil domains. Termini of each subunit are labelled N and C. Green arrows represent β-strands forming the six-stranded β-barrel, and blue arrows are β-strands forming the three β-hair-pins in each domain. A short 3_{10}-helix in red connects the two domains of each subunit (Transue *et al.,* 1997).

UDA (*Urtica dioica* agglutinin), a member of the hevein family of lectins has two homologous 43 amino acid hevein domains. Each of the domains contains a short stretch of antiparallel β-sheet (Kom, Gochin, Peumans & Shine, 1995).

Koharudin, Viscomi, Jee, Ottonello & Gronenborn (2008) determined the NMR (Nuclear magnetic resonance) solution structures for three members of the recently discovered cyanovirin-N (CV-N) homolog family of lectins. Cyanovirin-N homologs (CVNHs) from *Tuber borchii* (fungus), *Ceratopteris richardii* (fern), and *Neurospora crassa* (fungus) were selected. Each protein adopts a similar globular fold with an ellipsoidal shape, comprised of two pseudosymmetrical halves, termed domains A and B, consistent with the definition previously described for the founding member, CV-N. The secondary structure elements in domain A are formed by a triple-stranded β sheet, a β hairpin and two 3_{10}-helical turns, again similar to the parent molecule CV-N. Likewise, domain B is composed of a triple-stranded β sheet, a β hairpin, and two 3_{10} helical turns (Koharudin *et al.,* 2008).

The structure of the LysM domain from *Escherichia coli* membrane-bound lytic murein transglycosylase D was determined. The LysM domain has a beta-alpha-alphabeta

secondary structure with the two helices packing onto the same side of an anti- parallel beta sheet (Bateman & Bycroft, 2000).

Despite numerous efforts, the three-dimensional structure of NICTABA and its carbohydrate-binding site could not yet be elucidated (Schouppe *et al.,* 2010). A three-dimensional model of the NICTABA structure based on the structural homology with the carbohydrate-binding module 22 of *Clostridium thermocellum* Xyn10B was proposed. According to this model, NICTABA consists of a β-sandwich composed of two β-sheets. Similar to many plant lectins the NICTABA model predicts a structure that consists mainly of β-sheet. These results are in agreement with circular dichroism analyses, which revealed that NICTABA consists of 45% β-sheet, 55% β-coil, but no α-helix (Schouppe *et al.,* 2010).

1.6 Functions of lectins

1.6.1 Lectins as storage proteins

A plant storage protein is usually defined as a protein abundantly found in the cells of reserve tissue of plants and that has no other function other than to be a nitrogen source (Pusztai & Bardocz, 1995). Although the storage proteins of plants make a large and heterogenous group, they share several important properties together. At the biochemical level, these proteins have an amino acid composition characterized by high contents of glutamate/glutamic acid, asparagine/aspartic acid, serine and glycine. Conversely, they have low contents of lysine and sulphur-containing amino acids. At the cellular level, these proteins are all processed, synthesized and transported as secretory proteins. As such they are synthesized on the rough endoplasmic reticulum and subsequently accumulate in vacuoles and vacuole-like organelles called ‘protein storage vacuoles’

(previously known as ‘protein bodies’). The expression of these genes is also regulated during development. The genes are activated in the parenchyma cells of the storage tissue in response to environmental conditions and/or the developmental stage of the plant, which triggers the induction of the massive supply of amino acids (or possibly another form of readily convertible nitrogen) in the cells of these tissues. This nitrogen supply is accumulated by the synthesis of storage proteins and sequestered in specialized organs. Usually the accumulation of storage proteins continues until the tissue begins to dessicate (e.g., ripening seeds) or enters a resting phase (e.g., bark, bulbs, tubers). After that, the storage proteins remain unchanged until the parenchyma cells are stimulated by external and internal factors (e.g., seed imbibition, shoot growth, increased day light or temperature, phytohormones) to mobilize their nitrogen reserves. This leads to proteolysis in the storage protein vacuole that continues until the storage proteins are completely degraded. Lectins present in large quantities in seeds and vegetative storage tissues are therefore similar to storage proteins with respect to their abundance, biochemical properties and developmental regulation (Pusztai & Bardocz, 1995).

The seed lectins are synthesized in a similar way as other seed storage proteins. These lectins are located in the storage parenchyma cells of cotyledons and in some cases endosperm where, like the genuine storage proteins, they accumulate in the storage protein vacuole. The seed lectins and storage proteins are also developmentally regulated in a similar manner (Figure 1.6). For example, the legume seed lectins are synthesized during seed development and are degraded during germination and seedling growth,

together with the major storage proteins. The seed lectins differ from the classical storage proteins only by their carbohydrate binding activity (Pusztai & Bardocz, 1995).

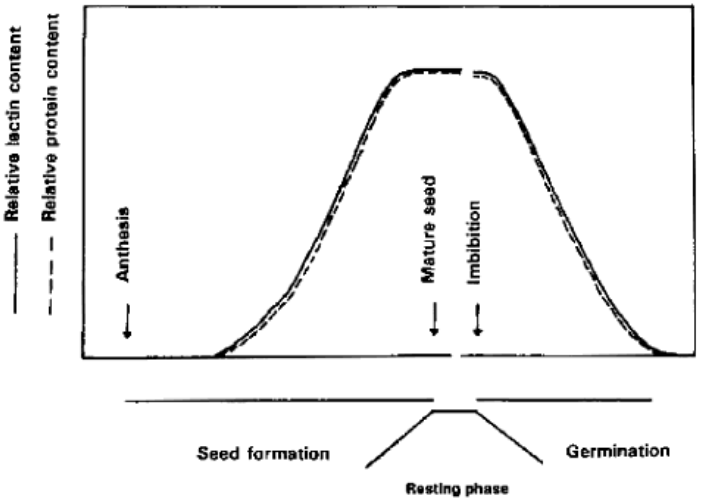

Figure 1.6: Schematic representation of the developmental regulation of lectin and storage proteins in seeds (Pusztai & Bardocz, 1995).

Available knowledge about lectins in the vegetative storage organs (e.g., bulbs, rhizomes, tubers) is scarce but suggests that these lectins also behave as storage proteins. For example, bulbs accumulate large quantities of carbohydrate and proteins at the end of the growing period when the upper part of the plant dies and the nitrogen content is stored in the underground tissues. When the new plant emerges after the resting season, the macromolecules in the old bulb are converted into simple sugars and free amino acids to be readily available to the rapidly growing and developing shoot. Studies of the changes in the lectin content during the life cycle of tulip (Van Damme & Peumans, 1989), snowdrop and daffodil (Van Damme & Peumans, 1990) have shown that these lectins accumulate and disappear during bulb formation and depletion and thus are developmentally regulated like storage proteins. The well-defined carbohydrate

specificity of these lectins distinguishes them from the classical storage proteins (Pusztai & Bardocz, 1995).

It can thus be concluded that plant lectins, whether they occur in seeds or vegetative tissues, behave as can be expected for storage proteins. When not challenged, these lectins serve as storage proteins, and when attacked with pathogens or herbivores, they take part in the defense of plants. But not all plant lectins are defense proteins. Lectins that are expressed in low quantities have been proposed to be involved in other recognition processes (Pusztai & Bardocz, 1995).

1.6.2 Lectins as defense proteins

Various reports have shown that plant lectins bind to carbohydrate structures on the surface of organisms such as viruses, bacteria, fungi, nematodes or phytophagus insects (Hopkins & Harper, 2001; Ripoll *et al.*, 2003; Wong *et al.*, 2010; Petnual *et al.*, 2010).

Many plant lectins are abundantly expressed in seeds and in storage organs that are particularly vulnerable to pathogens or insects. Therefore a role for some plant lectins in defense has been proposed. There are even examples where plant lectins interact exclusively or preferentially with foreign glycans rather than the plant specific carbohydrates (Peumans, Barre, Hao, Rouge´ & Van Damme, 2000), for example sialic acid-specific lectins from elderberry (*Sambucus* sp.) (Shibuya *et al.*, 1987) and *Muackiu amurensis* (Knibbs, Goldstein, Ratcliff & Shibuya, 1991) bind sugars that are absent in plants (Shibuya *et al.*, 1987) but are common components of animal glycoproteins (Yang, Shun, Chien & Wang, 2008).

Previous studies of the binding of plant lectins to bacterial cell wall peptidoglycans indicate that several legume seed lectins can recognize bacterial cell wall components such as muramic acid, N-acetylmuramic acid, and muramyl dipeptide (Ayouba *et al.*, 1994). Thus, lectins are potential antibacterial agents. For example, lectin from *Curcuma longa* (turmeric) rhizomes was found to inhibit a diverse selection of bacteria like *Bacillus subtilis, Staphylococcus aureus, E. coli, and Pseudomonas aeruginosa* (Petnual *et al.*, 2010).

Some lectins also have antifungal activity. The carbohydrate recognized by chitin-binding plant lectins is absent in plants but is a major component of the fungal cell wall and the exoskeleton of insects (Martínez, Falomir & Gozalbo, 2009). Lectins of the UDA and hevein families are presumed to inhibit fungal growth, probably by affecting fungal cell wall morphogenesis (Broekaert, Van Parijs, Leyns, Joos & Peumans. 1989; Van Parijs, Broekaert, Goldstein & Peumans, 1991). However, compared to other known antifungal/antimicrobial proteins like defensins, and thionins, the antifungal activity of plant lectins is generally weak (Selitrennikoff, 2001; Wong *et al.*, 2010).

Indeed, as defensive agents, plant lectins are generally more effective against phytophagus insects than microbial pathogens. It has been predicted that after ingestion of plant tissues, lectins come in contact with carbohydrate structures or glycan receptors exposed along the intestinal tract of the insects causing adverse effects that repel the herbivore. In many insects, the lumenal side of the epithelium is lined with the peritrophic membrane that is composed of chitin and glycoproteins (Hegedus, Erlandson,

Gillott, & Toprak, 2009). This is a target site for lectins. Feeding experiments with lectins like WGA (Wheat germ agglutinin) in artificial diets on European corn borer (*Ostrinia nubilalis*) have shown hypersecretion of many disorganized peritrophic membrane layers into the midgut lumen and the presence of many disintegrated microvilli (Harper, Hopkins & Czapla, 1998; Hopkins & Harper *et al.*, 2001). Thus, lectins are responsible for the formation of an abnormal peritrophic membrane and distruption of the microvilli. Lectins like GNA and ConA (Concanavalin A) were found to be present in parts of the body other than the digestive tract. They were found in hemolymph, fat tissues, malphigian tubules, ovaries, etc., indicating additional target sites for lectins inside insect body (Fitches, Woodhouse, Edwards & Gatehouse, 2001; Powell, Spence, Bharathi, Gatehouse & Gatehouse, 1998).

1.6.3 Lectins in symbiosis

It has been suggested that lectins play an important role in the symbiotic interactions between host plants and the symbiotic microorganisms such as nitrogen fixing rhizobia and mycorrhizal fungi (De Hoff, Brill & Hirsch, 2009).

The legumes are well known to establish a symbiotic relation with *Rhizobium* through root nodules (Brewin & Kardailsky, 1997). This allows the plants to exploit the bacterial capacity for nitrogen fixation, in which atmospheric nitrogen is converted into a form that can be used for plant growth. Establishing this intimate association requires a complex series of molecular interactions and signal exchanges between the host plant and the bacterium. In response to *Rhizobium*-derived signal molecules termed lipochitin oligosaccharides (Nod factors), nodule tissue is formed from a new meristem initiated in

the root cortex. The *Rhizobium* invades tissues and cells and progressively colonizes the root hair surface, the intercellular space and the infection thread (a transcellular cellulosic tube within which bacteria grow and divide). Finally, bacteria are released through the plasma membrane into the host cytoplasm, where they remain enclosed by a plant-derived peribacteroid membrane. In this 'symbiosome' compartment the rhizobia, now termed bacteroids, develop the capacity for nitrogen fixation under the low oxygen conditions that prevail in the central tissue of the nodule. This nitrogen is then excreted from symbiosomes in the form of ammonia, assimilated in the host cytoplasm and ultimately translocated to other parts of the plant. Throughout this process, from root colonization to the assimilation of the products of symbiotic nitrogen fixation, there is evidence for the involvement of legume lectins (Brewin & Kardailsky, 1997).

The abundant legume seed lectins are thought to bind carbohydrate moieties on the bacterial surface and assist in the initial attachment phase of the rhizobia to the root epidermal cells. Successful attachment of the bacteria on the surface of the root hairs facilitates infection thread formation, leading to effective root nodule formation (De hoff *et al.,* 2009). So lectins appear to be involved in determining the host-specificity during the root nodule symbiosis (Diaz, Melchers, Hooykaas, Lugtenberg & Kijne, 1989; Van Eijsden, Diaz, de Pater & Kijne, 1995; Van Rhijn, Goldberg & Hirsch, 1998). Observation with indirect immunofluorescence microscopy of the pea roots has shown the legume lectin, *Pisum sativum* lectin (PSL), to be located on the tips of emerging and growing root hairs in the root zone susceptible to infection by *Rhizobium* (Diaz *et al.*, 1986). Experiments have shown that expression of legume lectin PSL or *Glycine max lectin apyrase* (GS52) in transgenic rice (a species that establishes symbiotic mycorrhizal

associations, but is not nodulated by rhizobia) promoted the rhizobial colonization of the roots (Sreevidya *et al.,* 2005).

Symbiosis is also initiated by the attachment of the fungus to the root surface forming appressoria (Frenzel *et al.,* 2006). Once the root cortex is colonized, the fungus forms highly dichotomous branched intracellular structures referred to as arbuscules. The symbiosis is characterized by a bidirectional exchange of nutrients: carbon components are transferred from the plant to the fungus whereas the latter mainly transfers phosphorous, and also increased uptake of other minerals such as zinc and copper to its host (Frenzel *et al.,* 2006).

Through studies of genes induced in an arbuscular mycorrhiza (AM) association in the model plant *Medicago truncatula,* some AM lectin-specific gene sequences were identified (Frenzel *et al.,* 2005; Wulf *et al.,* 2003). Promoter fusion studies with a reporter gene showed an arbuscule-specific expression of two members of the AM specific lectin family, suggesting a role during arbuscule formation and functioning during late stages of the AM symbiosis (Frenzel *et al.,* 2005). One of the possible functions of AM-specific lectins could be a direct interaction with the fungal microsymbiont during arbuscule formation by binding of carbohydrates exposed to fungal cell wall surfaces (Frenzel *et al.,* 2005). The lectins could also function as storage proteins during AM symbiosis. Thus it was proposed that AM-specific lectins have specific roles during arbuscule development or function (Frenzel *et al.,* 2005).

1.6.4 Lectins can function in stress signaling

It is generally assumed that the synthesis of defense compounds is inducible, rather than constitutive, to conserve resources when the defence compounds are not needed (Zavala & Baldwin, 2004). Plants synthesize multiple signaling hormones in response to different biotic and abiotic stresses. Experimentally, phytohormone treatments can be used to induce stress responses. As described earlier, inducible lectins are expressed in low quantities in plants after being subjected to various stresses such as high salt, drought, pathogen attack, insect herbivory, and so on. An example is the mannose-specific jacalin related lectin of rice, ORYSATA, which is only expressed in roots and sheaths after exposing the plants to salt and drought stress (Zhang *et al.,* 2000; Hirano *et al.,* 2000). Later, this lectin was also shown to be expressed in response to jasmonic acid and abscisic acid treatment or after infection with the fungus *Magnaporthe grisea (*Souza Filho, Ferreira & Dias *et al.,* 2003; Kim, Cho & Yu *et al.,* 2003; Qin, Zhang, Zhao, Wang & Peng, 2003). In rice plants, proteins with EUL (*Euonymus* lectin) domain are expressed in response to abscisic acid treatment and salt stress (Moons, Gielen, Vandekerckhove, Van der Straeten, Gheysen & Van Montagu, 1997). The tobacco lectin, NICTABA, is induced by jasmonate treatment and insect herbivory (e.g., by the cotton leaf worm *Spodoptera littoralis*) (Chen *et al.,* 2002; Lannoo *et al.,* 2007a; Vandenborre, Miersch, Hause, Smagghe, Wasternack, Van Damme, 2009). Preliminary experiments have shown that the amount of NICTABA expressed by the insect herbivory is sufficient enough to exert a toxic effect on the insects (Vandenborre, Groten, Smagghe, Lannoo, Baldwin & Van Damme, 2010). Wheat plants also responded to insects by the upregulation of its lectin genes. In response to Hessian fly (*Mayetiola destructor*) larvae, mRNA transcripts

of the Hfr-1 gene was upregulated (Williams, Collier, Nemacheck, Liang & Cambron, 2002). Hfr-1 is a protein containing a C-terminal domain with sequence similarity to jacalin-related lectins (Subramanyam, Sardesai, Puthoff, Meyer, Nemacheck, Gonzalo & Williams, 2006). The insecticidal activity of Hfr-1 was clearly demonstrated when fed to larvae of the fruit fly *Drosophila melanogaster* (Subramanyam, Smith, Clemens, Webb, Sardesai & Williams, 2008). Two other Hessian fly responsive wheat genes, Hfr-2 and Hfr-3, have also been reported (Puthoff, Sardesai, Subramanyam, Nemacheck & Williams, 2005; Giovanini, Puthoff & Nemacheck *et al.*, 2006). The sequence of Hfr-2 has an N-terminal amaranthin like domain fused to a region similar to haemolytic toxin and channel forming toxins. The Hfr-3 sequence has four putative chitin-binding hevein domains. Thus ,all these examples imply that nucleocytoplasmic lectins take part in the inducible defense system and stress signaling of plants.

1.6.5 Lectin can function in glycoprotein folding and degradation

Plant lectins like calreticulin and calnexin are involved in the proper folding of glycoproteins in the endoplasmic reticulum (Banerjee *et al.*, 2007; Carameldo and Parodi, 2008). N-glycans of misfolded proteins are glucosylated and are then bound and refolded by calreticulin and/or calnexin in association with a protein disulphide isomerase. After proper folding of the glycoproteins, lectins like ERGIC-53 (ER-Golgi intermediate compartment 53 kDa protein), VIP36 (vesicular integral protein of 36 kDa), and VIPL (VIP36-like protein) bind with mannose residues in the N-glycans and then the glycoproteins are transported to the Golgi apparatus (Banerjee *et al.*, 2007; Carameldo & Parodi, 2008; Van Damme *et al.*, 2011).

When proteins are misfolded they are then directed to the ERAD (ER associated protein degradation) pathway (Benerjee *et al.,* 2007). In mammals, misfolded glycoprotein are transferred from the lumen of ER to the cytosol and are directed by the sugar binding F-box protein (Fbs) to the SCF complex (Skp, Cullin and an F-box protein) for degradation (Mizushima, Yoshida, & Kumanomidou, 2007; Petroski & Deshaies, 2005). Annotation of some plant genomes led to the identification of evolutionary related homologs of the mammalian Fbs protein, in rice, Arabidopsis, poplar and moss *Physcomitrella patens* (Lechner, Achard, Vansiri, Potuschak & Genschik, 2006; Gagne, Downes, Shiu, Durski & Vierstra, 2002; Kuroda, Takahashi, Shimada, Seki, Shinozaki & Matsui, 2002; Jain, Nijhawan & Arora; 2007; Yang, Kalluri & Jawdy, 2008). A sugar binding F-box protein from Arabidopsis was recently characterized in 2012 by Stefanowicz, Lannoo, Proost & Van Damme. This bipartite protein has an N-terminal F-box domain and a C-terminal NICTABA like domain and is located in the nucleocytoplasmic compartment of plant. Glycan array results showed that the NICTABA-like domain provided the F-box protein with carbohydrate binding activity although the specificity was substantially different from that of NICTABA. Thus, this class of proteins could be potential candidates having a similar role to the mammalian Fbs and take part in glycome regulation of the plants (Stefanowicz *et al.,* 2012).

1.6.6 Lectins may function in nucleocytoplasmic transport

It has been suggested that the transport of protein and RNA molecules in and out of the nucleus could be regulated by plant lectins (Fahrenkrog & Aebi, 2003). Several studies have shown that glycoproteins substituted with O-glycans play an important role in nuclear transport (Hanover, 2001; Miller & Hanover, 1994). It has been shown that a

GlcNAc-specific lectin, WGA capable of binding to the O-GlcNAc-residues of glycoproteins in the nuclear pore complex can prevent the import of proteins into the nucleus of animal cell (Yoneda, Imamoto-Sonobe, Yamaizumi & Uchida, 1987). Although WGA is a vacuolar lectin (Raikhel, Mishkind & Palevitz, 1984), the experimental evidence suggests that it can be involved in the nucleocytoplasmic transport.

Confocal microscopy of the subcellular localization of EGFP-NICTABA in transgenic tobacco BY-2 cells has shown the lectin to accumulate at the nuclear rim, indicating that it can interact with proteins of the nuclear pore complex (Lanoo *et al.*, 2006). Since NICTABA is partly located in the cytoplasm and the nucleus, it was hypothesized that it serves as a shuttle protein between the nuclear and the cytoplasmic compartments (Van Damme, Barre, Rougé & Peumans, 2004; Van Damme, Lannoo, Fouquaert, Peumans, 2004; Chen *et al.*, 2002).

Since NICTABA is specific for GlcNAc oligomers and N-glycans, it can be speculated that it binds to O-GlcNAc-modified or N-glycan containing nucleopore proteins (Lanoo & Van Damme, 2010). Although there was experimental evidence that NICTABA can interact with N-glycosylated proteins in nuclear extracts (Lanoo *et al.*, 2006), the nuclear receptors for NICTABA are yet to be identified.

1.7 Biotechnological applications of lectins

Lectins also have many biotechnological, diagnostic and therapeutic applications. Transfer of insecticidal lectin genes in genetically modified crops improves plant

resistance against many economically important pest insects. Engineering of GNA into crops like potatoes, rice, sugarcane, tobacco and wheat enhanced their protection against many pest insects (Couty *et al.,* 2001; Bell *et al.,* 1999; Down *et al.,* 2001; Tinjuangjun *et al.,* 2000; Foissac *et al.,* 2000; Tomov and Bernal, 2003; Wang and Guo, 1999; Stoger *et al.,* 1999). Both constitutive and phloem specific expression of GNA in rice plants significantly improved the transgenic plants' resistance against the two most important economic pests of rice, by significantly reducing nymphal survival of brown leafhopper (*Nilaparvata lugens*) and green leafhopper (*Nephotettix virescens*) (Tinjuangjun *et al.,* 2000; Foissac *et al.,* 2000). The constitutive expression of GNA and a bean chitinase in potatoes enhanced resistance of the plants towards the tomato moth (*Lacanobia oleracea L*) larvae by significant reductions in larval biomass and a retardation of their development (Down *et al.,* 2001).

None of these transgenic uses of lectin have been commercialized. This is due in large part to the 'Pusztai affair' controversy of 1998. This started when Dr. Arapad Pusztai (an expert on lectins) of the Rowett Institute, Scotland expressed his concerns publicly about the effect GM (genetically modified) potatoes have on mammalian stomach mucosa by disclosing the experimental findings of his research study to assess the safety of genetically engineered potatoes (Pusztai, 2002). The potatoes used were not a commercial variety and were not intended for human consumption (Randerson, 2008).

The potatoes transformed with a snowdrop lectin, (GNA) under the CaMV35s promoter, had been developed to increase insect and nematode resistance (Ewen & Pusztai, 1999). In the experiment, rats were fed potato diets containing GM potatoes, non-GM potatoes,

or non-GM potatoes supplemented with GNA for 10 days. Then the histological indices of their gut were examined to find out whether GNA gene insertion had affected the nutritional and physiological impact of potatoes on the mammalian gut. The findings suggested that the GM-GNA potatoes gave rise to unusual changes in the gut, such as significantly greater mucosal thickness of the stomach and crypt length in the jejunum, and thinner caecal mucosa. His team concluded that the effects could be due to the GNA transgene, other parts of the GM construct or the transformation procedure (Ewen & Pusztai, 1999).

The concerns raised by Pusztai's television interview threatened to damage the reputation of the biotech industry and its ability to create confidence in GM foods (http://www.psrast.org/pusztai.htm). Although initially praised by the institute's director, Pusztai was later suspended and gagged by the research institute. The formal reason for his suspension was that he had presented the results publicly before they had been reviewed by other scientists (peer review) as required by the Rowett Institute. Opponents of GMOs claim that Pusztai's dismissal was at the behest of the biotech industry. It was said that his results were misleading as he had mixed up the results of different studies. The Rowett Institute asked a scientific committe to review the study done by Pusztai. The committee concluded that there were important deficiencies in the study (http://www.psrast.org/pusztai.htm). Pusztai then sent his study report to independent scientists of 24 countries who refuted the conclusions of the review committee and confirmed that his research was of good quality and justified his conclusions (http://www.psrast.org/pusztai.htm). When the actual findings of the experiment were

finally published in the prestigious medical journal Lancet, the study was criticized on the grounds that the unmodified potatoes were not a fair control diet for the rats and so they would ultimately suffer from protein deficiency (Randerson, 2008). In his paper Pusztai already addressed this by mentioning that all the diets were isocaloric (Ewen & Pusztai, 1999). Pusztai's message was that GM food requires careful testing (Rhodes, 1999) before being allowed for commercialization. But the entire media storm surrounding the 'Pusztai affair' contributed to the public perception that GM crops are harmful.

Besides GNA, other lectins also have been expressed in transgenic plants. The legume lectin, PSA (*Pisum sativum* agglutinin) was expressed in the anthers and pollen of transgenic oilseed rape *(Brassica napus*) and negatively affected the larval mass gain of the pollen beetle (*Meligethes aeneus*) (Melander, Ahman, Kamnert & Strömdahl, 2003). Transgenic Indian mustard (*Brassica juncea* cv. RLM-198) plants transformed with WGA induced high mortality and significantly reduced fecundity of mustard aphid (*Lipaphis erysimi*. Kalt) (Kanrar, Venkateswari, Kirti & Chopra, 2002). Cotton plants transformed with *Amaranthus caudatus* agglutinin (ACA) under the control of a phloem-specific promoter had strongly enhanced resistance against nymphs of the cotton aphid (*Aphis gossypii*) (Wu *et al.*, 2006).

Neoplastic cells of human tumors have glycan profiles that are distinct from normal human cells (Nangia-Makker, Conklin, Hogan & Raz, 2002). Several methods have emerged where lectins were used for the detection of aberrant glycans expressed by neoplastic cells (Sobral, Rego, Cavalacanti, Carvalho & Beltrão, 2010; De Lima *et al.*,

2010; Beltrão *et al.*, 2003). Amongst these techniques, immunohistochemical assay is the most common and widespread. Leguminous lectins from from *Canavalia ensiformis* (ConA) and *Ulex europaeus* (UEA-1) were used as histochemical markers of parotid gland mucoepidermoid carcinoma. The localization of ConA in the cytoplasm and/or plasmamembrane was significantly associated with neoplastic cells from the three grades of severity: low, intermediate and high grade dysplasia, whereas UEA-1 was associated with low and intermediate grade dysplasia (Sobral *et al.*, 2010). Thus, some lectins can be used as tumor markers for differentiating between benign and malignant tumors (Kumar *et al.*, 2012). These are also used as cell markers for infectious agents such as viruses, bacteria, fungi, and parasities and for typing blood groups (Rüdiger & Gabius, 2001).

Apoptosis is a Type I programmed cell death and is used naturally to control cell proliferation or in response to irreparable DNA damage (Hengartner, 2000). Apoptosis occurs through two major pathways: the extrinsic pathway triggered by the Fas death receptors, and the mitochondria-dependent pathway that brings about the release of Cytocrome *c* (Cyto *c*) and activation of the death signals under stimulus. In both cases, the caspases, which belong to a family of cysteine proteases, have been established as major players in apoptosis-causing mechanisms (Hengartner, 2000).

In the mitochondrial-dependent pathway, ConA treatment results in a decrease of mitochondrial membrane potential, thus collapsing the transmembrane potential (Li, Xu, Liu & Bao, 2010). Cyto *c* is subsequently released, making up apoptosome with Apaf-1 and procaspase-9. After conjugating apoptosome, procaspase-3 turns into active caspase-

3 that eventually triggers apoptosis (Li *et al.,* 2010). Thus lectins can induce apoptotic cell death and can be used in the treatment of tumors. (GNA)-related lectins are also well known to function as specific inhibitors of tumor growth and retroviral infection (Chun-yang, Meng, Liu & Bao, 2009).

The anti-HIV lectin, GNA, has high affinity for α-(1-3)-D-mannose oligomers (An *et al.*, 2006). The crucial envelope glycoprotein of HIV, gp120, is extensively glycosylated with N-linked complex and high mannose carbohydrates accounting for about half of its molecular weight. GNA directly interferes with the virus-cell membrane fusion process by binding to the high-mannose glycans on gp120 and inhibiting the HIV-induced cytopathicity (An *et al.,* 2006).

Another important application of lectins is the use of lectin affinity chromatography to fractionate and purify glycoproteins (Lam & Ng, 2011). In this type of chromatography lectins are immobilized on gel beads and glycoproteins are then adsorbed by lectins and eluted with a specific carbohydrate. Immobilized ConA was applied in the isolation of glycopeptides that express biantennary and hybrid N-linked structures and high-mannose glycans, which are abundant in both embryonic stem cells and embryoid body stages (Alvarez-Manilla *et al.,* 2010). When used in conjunction with other separation techniques, lectin affinity chromatography could help to rapidly purify oligosaccharides and provide substantial information about their structural features (Lam & Ng, 2011).

1.8 Heterologous expression of lectins

Isolation of proteins from their native source (especially plant sources) is usually costly, cumbersome and lengthy (Yesilirmak & Sayers, 2009). A more convenient alternative is to express the protein in a foreign host. This would usually allow large-scale purification of a desired protein and may also provide a system to elucidate its function, structure and different biochemical and biophysical properties. Recombinant plant proteins produced by heterologous expression are also used in industrial applications (Yesilirmak & Sayers, 2009).

Plant lectins have been successfully expressed as recombinant proteins in different hosts such as bacteria (predominantly *E. coli*), yeast (often *Saccharomyces cerevisiae*), mammalian cells (Ling, Yang & Bi, 2010; Nagahora, Ishikawa, Niwa, Muraki & Jigami, 1992; Adar *et al.*, 1997). Usually the first system used in attempts to express lectins is *E. coli*. The cDNA encoding for the plant lectin is cloned into an expression vector under a strong constitutive promoter, with or without a tag (for purification purpose), for lectins with known ligands and unknown lectins respectively. The pET expression vector family is widely used for this purpose. PTA lectin of the GNA family was purified as a His-tagged protein (Ling *et al.*, 2010). AAG, BPA are some examples of unfused lectin proteins expressed in bacteria (Li *et al.*, 2000; Kusui, Yamamoto, Konami & Osawa, 1991). In the pET vectors, the genes are cloned under control of a T7 promoter. The advantage of this promoter is that it is normally not recognized by the *E. coli* RNA polymerase so that there is no leaky expression. The vector is transformed into appropriate host cells (often *E. coli* BL21 (DE3)) that have a chromosomal copy of the T7 RNA polymerase gene under the control of the lac promoter. The lac promoter, and

thus RNA polymerase production, can be induced with IPTG whereupon high levels of plant lectins may be expressed (up to several 100 mg/liter bacterial culture) (Streicher & Sharon, 2003).

The lectins expressed in bacteria are usually purified by affinity chromatography. The yield of the purified active recombinant lectin can be 0.1–20 mg per liter culture (Streicher & Sharon, 2003). Lectins from *Erythrina corallodendron* (Arango, Adar, Rozenblatt & Sharon, 1992), *Dolichos biflorus* (Chao, Casalongue, Quinn & Etzler, 1994), *Phaseolus lunatus*, lima bean (Jordan & Goldstein, 1994), *Phaseolus vulgaris* (Hoffman & Donaldson, 1987), *Canavalia ensiformis* (Yamauchi & Minimikawa, 1990), *Canavalia brasiliensis* (Nogueira *et al.,* 2002), *Ricinus communis* (Hussain, Bowler, Roberts & Lord, 1989), *Glycine Max* (Adar, Streicher, Rozenblatt & Sharon, 1997), *Galanthus nivalis* (Longstaff *et al.,* 1998), *Pinellia ternate* (Ling *et al.,* 2010) and others have been reported to be successfully expressed in *E. coli*.

Legume lectins such as GS-II (expressed without glycosylation in *E. coli*) have been reported to be biologically active even in the absence of their carbohydrate moieties (Zhu *et al.,* 1996). Recombinant unglycosylated *Erythrina corallodendron* lectin (EcorL) expressed in *E. coli* was also found to be active (Arango *et al.,* 1992). Although the structures of the unglycosylated recombinant proteins and native proteins were different (Zhu *et al.,* 1996; Arango *et al.,* 1992), it was observed that glycosylation may not play a crucial role in the activity of these lectins. So, recombinant active lectins could be expressed in bacteria. However, in some circumstances, the presence or absence of the

carbohydrate side chains can have significant effects on the functional and/or physical properties of the recombinant proteins. For example, ConA is activated by deglycosylation in plants (Sheldon & Bowles, 1992; Min, Dunn & Jones, 1992) and nonglycosylated PHA-L expressed in transgenic plants gave poor quality crystals under conditions where glycosylated PHA-L was crystallized successfully (Dao-Thi, Hamelryck, Poortmans, Voelker, Chrispeels & Wyns, 1996).

High levels of protein in the cell may lead to the formation of inclusion bodies by the bacteria to overcome this cytotoxic effect. In that case, only a small fraction of the recombinant protein will be soluble. The inclusion bodies containing the lectins can be solubilized by protein-denaturing agents and refolded. If refolded correctly, the lectin will regain its activity, but during refolding, the proteins can form aggregates (Streicher & Sharon, 2003).

Many lectins form insoluble inclusion bodies. Active lectins such as GNA (Longstaff *et al.*, 1998), ECorL, and soybean agglutinin (Adar *et al.*, 1997) have all been successfully isolated from bacterial inclusion bodies (Arango *et al.*, 1992). In each case, the recombinant protein was functionally similar to the native protein. However, the method of purifying proteins from inclusion bodies is not only time-consuming, the recovery of active protein is often poor and the lectins might not be properly folded, and will not be glycosylated.

Lectin expression in yeast (*S. cerevisiae*) can, in some circumstances, overcome some of the problems of lectin expression in bacteria, such as low yield and post-translational modification like glycosylation. But the expression of lectins in *S. cerevisiae* does not always occur as expected. PHA-L expressed in *S. cerevisiae* accumulates mainly in the vacuole (Tague & Chrispeels, 1987), due to the presence of cryptic vacuolar targeting determinants (von Schaewen & Chrispeels, 1993), and only about 1% is secreted. In addition, a significant proportion of PHA-L expressed in this host was not correctly processed. Approximately half of the PHA-L accumulating in the vacuole appeared to contain the uncleaved signal peptide, and all of the secreted PHA-L was in this unprocessed form (Tague & Chrispeels, 1987). A second related lectin, from *Dolichos biflorus*, could be directed into a secretory pathway when expressed at low levels, but accumulated in the cells (not in the vacuole) when expressed at high levels (Chao & Etzler, 1994). In both cases, the functional properties of the recombinant lectin were not reported. The only report of the correct processing and secretion of a plant lectin in *S. cerevisiae* is that of WGA, which was secreted and exhibited sugar binding activity, however, yields of protein were relatively low, of the order of 200 μg/l (Nagahora, Ishikawa, Niwa, Muraki & Jigami, 1992). Thus, it would appear that *S. cerevisiae* is not an ideal host for lectin expression.

Methylotrophic yeast, *Pichia pastoris*, is also a potential host for expressing plant lectins (Cregg, Vedvick & Raschke, 1993; Sreekrishna, 1997). Functional plant lectins, such as PHA, GNA (Raemaekers, de Muro, Gatehouse & Fordham-Skelton, 1999), *Canavalia brasiliensis* lectin (ConBr) (Bezerra, da Silveira Carvalho, Moreira & Grangeiro, 2006),

NICTABA (Lannooa, Vervecken, Proost, Rougé & Van Damme, 2007b) have been expressed in *Pichia*.

In a few cases plant lectins have been expressed in mammalian cells, predominatly in monkey COS cell lines. An example is the soybean agglutinin (SBA) (Adar *et al.,* 1997). The glycosylation and the sugar binding specificities of the recombinant protein were indistinguishable from the native protein. The yield of the protein was, however, very low: 0.1 mg/liter culture.

Plant lectins have also been expressed in cell-free systems. In the past, plant lectins like Abrin (Evensen, Mathiesen & Sundan, 1991), ConA (Carrington, Auffret & Hanke, 1985), and RCA (Roberts, Lamb, Pappin & Lord, 1985) have been expressed in commercially available rabbit reticulocyte lysates. The amounts of lectin obtained this way are small. NICTABA was also expressed in a cell free system based on *E. coli* lysate and was found to be biologically active (Vandenborre *et al*., 2008).

1.9 Methods used to study lectins

Characterization of a lectin involves many different techniques. Novel lectins can be classified by homology searches with the known lectin families and also with other protein sequences in the protein database. Using known lectins as models, molecular modelling can be carried out to infer the three dimensional structure of lectin, and to determine the carbohydrate recognition domains and the key residues of the binding site or the mechanism of interaction between sugars and residues in the binding site (Costa *et al.,* 2011).

The spatial expression profile of lectins in different tissues can be determined by qRT-PCR. This can allow inferences of the function of novel genes by determining their changes in expression over time in response to different factors or treatments (Puranik, Sahu, Mandal, Parida & Prasad, 2013).

Purified protein is required to determine the structural, biochemical and functional properties of proteins. Lectins can be extracted from the native tissues by techniques such as affinity chromatograpy. The lectin can also be heterologously expressed and purified in large scale for carrying out different assays.

In order to gain detailed insights into the structure of lectins, the proportion of β-strand and α-helical segments of the purified lectin can be estimated from the (Circular dichroism) CD spectra and three dimensional structure determined by X-ray crystallography and NMR.

Native molecular mass can be determined by gel filtration chromatography. A calibration curve with protein markers of known molecular weights is first obtained and by plotting the elution volume of the lectin against the molecular weights of the standards, the native molecular mass is determined (Kaur *et al.,* 2005). By comparing the band patterns in SDS-PAGE and Native PAGE, oligomerization of the protein can be observed. For example, two bands in SDS-PAGE but a single band in Native PAGE indicates that the lectin is probably a dimer (Damico *et al.,* 2003).

The purified lectin can be used in far-western blots involving crude protein extracts or proteins of specific organelles (Lannoo *et al*., 2006). Inclusion of appropriate controls will help determine the proteins that interact with the lectin. Thus potential target sites or ligands of the lectin could be identified (Lannoo *et al*., 2006).

Lectin activity is usually measured by hemagglutination of erythrocytes. Using serial dilutions of the lectin, the minimum concentration required to agglutinate erythrocytes is determined. The potential ligands binding to the lectins can be determined by hemagglutination inhibition assays with sugars. Sugars incubated with lectins prior to the addition of erythrocytes will bind the lectin and prevent it from binding the erythrocytes.

The specific carbohydrate structures recognized by lectins can be further determined by using glycan arrays. These are performed on arrays composed of various oligosaccharides and/or polysaccharides immobilized on a solid support in a spatially defined arrangement (Oyelaran & Gildersleeve, 2009), allowing the recognition of many ligands at a time.

Along with the sugar specificity of lectins, it is also important to determine the thermodynamics/kinetics of the interactions between lectins and their sugars (Duverger, Frison, Roche & Monsigny, 2003), for elucidating the mechanism of interaction and the forces involved in binding. Surface plasmon resonance has thus been used in more detailed studies of lectin-carbohydrate interactions for the determination of affinity

constants, association/dissociation contants and lectin specificity (Linman, Taylor, Yu, Chen & Cheng, 2008; Duverger *et al.,* 2003).

Antibodies against purified lectin can be raised in hosts like rabbits and used for western blot analysis of different tissue samples to confirm the presence or absence of the lectin in particular tissues (Cummings & Etzler, 2009). The level of lectin in the protein extract from different tissues can then be quantified by ELISA (Enzyme-linked immunosorbent assay).

Sub-cellular localization can be determined by transforming and expressing lectins genes fused to a reporter gene (e.g. GFP) in the desired host cell (onion epidermal cell, BY-2 cells) by transformation with *Agrobacterium tumefaciens* or biolistic particle delivery system. Also the localization of the lectins in tissues can be confirmed by immunocytochemical techniques by using an affinity-purified antiserum (the specificity of which can be confirmed by Western blot analysis).

The effects of temperature, pH and denaturants on lectin stability can be studied *in vitro*. The requirement of metal ion for its lectin activity also must be determined. By doing dialysis of the protein sample against EDTA solution metal ions can be removed and hemagglutination activity observed. In case the activity is lost, adding different metal ions ($CaCl_2$, $MnCl_2$, $FeCl_2$, $MgCl_2$) the activity can be recovered and the metal ion requirement for carbohydrate binding activity can be determined.

To determine whether a purified lectin is a glycoprotein, sugar analysis is done using the phenol sulphuric acid method (Dubois *et al.,* 1956) with glucose as standard. This method has been developed to determine submicro amounts of sugars and related substances (Dubois, Gilles, Hamilton, Rebers & Smith, 1956). Glycosylation of the lectin can also be demonstrated in gel by the method described by Zacharius, Zell, Morrison & Woodlock (1969) in SDS PAGE. This procedure stains glycoproteins in the gel.

The antifungal activity of lectins can be determined using a hyphal-extension inhibition assay. Sterile discs containing the lectin and an appropriate control are placed about 0.5 cm from the rim of the fungal mycelium (Banerjee *et al.,* 2011). Growth inhibition is observed in the form of cresents of inhibition. Antiviral assays can help determine if the lectin can inhibit viral replication such as by preventing the entry of virus in cells (van der Meer, 2007). Assays such as neutral red (NR) dye uptake assay that provides a quantitative estimation of the number of viable cells in a culture can be used in this purpose. Also, direct interaction of lectin with viral envelope glycoproteins can be assayed using ELISA (Sato, Morimoto, Kubo, Yanagihara & Seyama, 2012). Insect bioassays with artificial diets can be conducted and the abnormalities and mortality of the insects scored (Rahbé, Sauvion, Febvay, Peumans & Gatehouse, 1995).

The *in vitro* anticancer activity of lectins against human cancer cell lines like A-549 (breast), HT-29 (colon), SiHa (cervix), SNB-78 (central nervous system), and PC-3 (prostate) can be assayed by an MTT (3-(4,5-dimethylthiazol-2-yl)2,5-diphenyl tetrazolium bromide) colorimetric assay (Zuo *et al.,* 2012). This assay helps to measure

the percentage of cell growth inhibition, by reduction of the MTT dye to an insoluble form by active enzymes of the cells. The mechanism of inhibition of the cell cycle progression by lectin can be determined by flow cytometry (Zuo *et al.,* 2012), a technique that measures optical and fluorescence characteristics of single cells (or any other particle, including nuclei, microorganisms, chromosome preparations, and latex beads) (Brown & Wittwer, 2000), contributing to improved cycle phase identification (Jayat & Ratinaud, 1993). For example, the DNA in a cell can be stained with a DNA intercalating florescent dye and quantified by flow cytometry. A decrease in S DNA would indicate inhibition of DNA replication.

Certain lectins are known to be mitogens, which trigger cell division. Mitogen assays can be done where cells such as murine splenocytes are incubated with the lectins and thymidine (to determine DNA synthesis). If the lectin stimulates uptake of the thymidine by the cells then the lectin have mitogenic activity (Chan, Wong, Fang, Pan & Ng, 2012).

1.10 Flax

Flax (*Linum usitatissimum*) is a globally important agricultural crop grown mainly for its stem fiber and seed oil (Wang *et al.,* 2012). The bast fibres are used in the linen textile industry and due to their very high tensile strength, are a potential candidate for producing composite materials (Bodros *et al.,* 2007). Flax seeds and their oil also have several attributes (e.g. α-linolenic acid, lignans) that may contribute to numerous health benefits including improved cardiovascular health, treatment of certain cancers and inflammatory diseases (Singh *et al.,* 2011). The seed oil is rich in unsaturated fatty acids, especially α-linolenic acid, polymers of which are used in linoleum, paints and other

finishes (Wang *et al.,* 2012). Flax seed is also used in animal feed to increase levels of α-linolenic acid in meat or eggs (Simmons *et al.,* 2011).

Canada and China had the world's largest production of linseed and flax fibre respectively in 2011 accounting to about 368,300 tons and 126,964 tons respectively (Figure 1.7 & 1.8) (http://faostat.fao.org/). Other countries having noteworthy production are also shown in the figures.

Flax diseases are mostly caused by fungal pathogens (Rashid, 2003). Rust and Fusarium wilt had been the two diseases that limit flax production in most of the flax growing countries worldwide, caused by the fungi *Melamspora lini* and *Fusarium oxysporum* respectively (Rashid, 2003). Some of the other diseases of flax are pasmo, alternaria blight, powdery mildew, verticillium blight, sclerotinia stem rot, browning and stem break etc (Rashid, 2003). Only a small number of insect pests attack flax and these are considered minor in their economic impact on the crop (Wise & Juliana, 2003). Nevertheless insect pests can cause serious yield losses of flax. Major insect pests of flax are potato aphids, cutworms, flea beetles, flax thrips and linseed midges (Wise *et al.*, 2003).

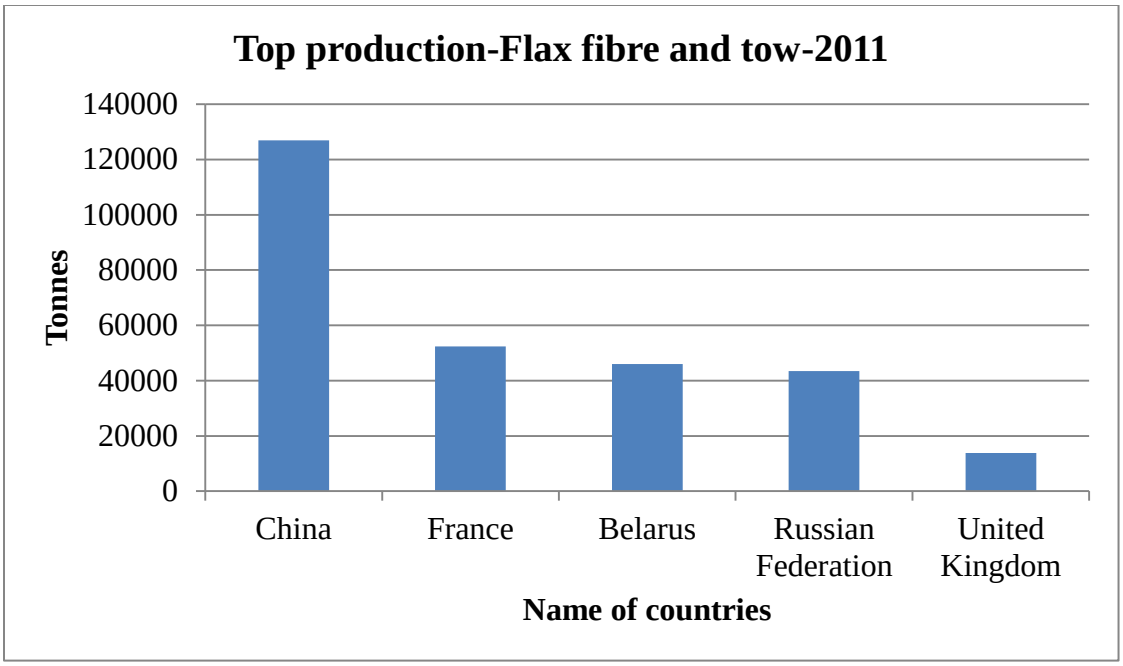

Figure 1.7: Top countries producing flax fibre and tow in the year 2011 (http://faostat.fao.org/).

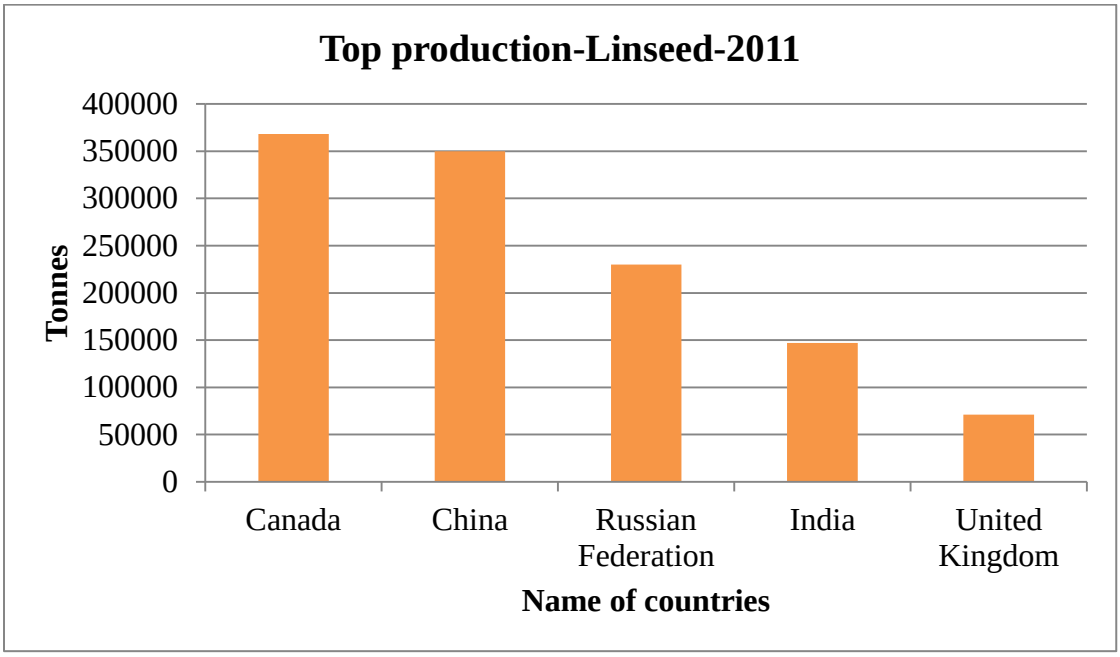

Figure1.8: Top countries producing linseed in the year 2011 (http://faostat.fao.org/).

1.11 The genome of flax

Flax belongs to the order Malpighiales, which also includes poplar (*Populus trichocarpa*), castor (*R. communis*), and cassava (*Manihot esculenta*) (Wang *et al.,* 2012). It is a self-pollinating annual diploid plant with a relatively small genome of~350 Mbp. The nuclear genome of flax has been recently sequenced using the whole genome shotgun sequencing (WGS) method. From flax DNA, seven paired end libraries were constructed that had insert sizes ranging from 300bp to 10kb, and sequenced using an Illumina genome analyzer. Low quality reads, reads containing unknown bases or the adapter sequence were filtered out before the de novo WGS assembly based on the short reads to ensure a more accurate genome sequence. The genes were predicted using two bioinformatics tools: Glimmer-HMM and Augustus. The results obtained were then compared to the flax EST database and other empirically established plant transcript sequence databases. The analysis resulted in a total of 43, 484 protein-coding genes with about 93 % of the published flax ESTs aligning with the genes. To further confirm that the predicted genes are actually legitimate genes, they were compared with the NCBI nr protein database and separately with the predicted peptide sequences from the well-characterized Arabidopsis and poplar genomes. About 89-91% of the flax WGS proteins aligned with these databases. Moreover, converse comparisons determined that about 86% of each of the Arabidopsis and poplar protein databases matched one or more predicted flax proteins. All these results indicated the excellent coverage and accuracy of the predicted genome assembly. Using Pfam-A database the functions of the predicted genes were annotated (Wang *et al.*, 2012).

During annotation of the protein-coding genes in the whole genome assembly of flax, 17 genes containing one or more agglutinin domains (PFAM PF07468) were identified out of a total of 43, 484 predicted genes (Wang *et al.*, 2012). Later analysis increased this number to 19 predicted agglutinin domain-containing proteins. Among other sequenced plant genomes available for comparison at the time, only apple (*Malus domestica*) contained a similarly large number of proteins with this PF07468 domain (22/ 57,386, PF07468) (Velasco *et al.*, 2010). Although the agglutinin domain was also detected in predicted proteins from diverse monocots (*Zea mays*, *Brachypodium distachyon*) and eudicots (e.g. *Carica papaya*, *R.communis*), it was not found in any protein from other monocots (*Sorghum bicolor*, *O.sativa*) or eudicots (*Glycine max*, *Arabidopsis thaliana*, *P. trichocarpa*) (Figure 1.9) (Wang *et al.*, 2012). The abundance of these predicted proteins in flax, and their sporadic distribution throughout the genomes of angiosperms motivated us to characterize the PF07468 domain-containing gene family of flax.

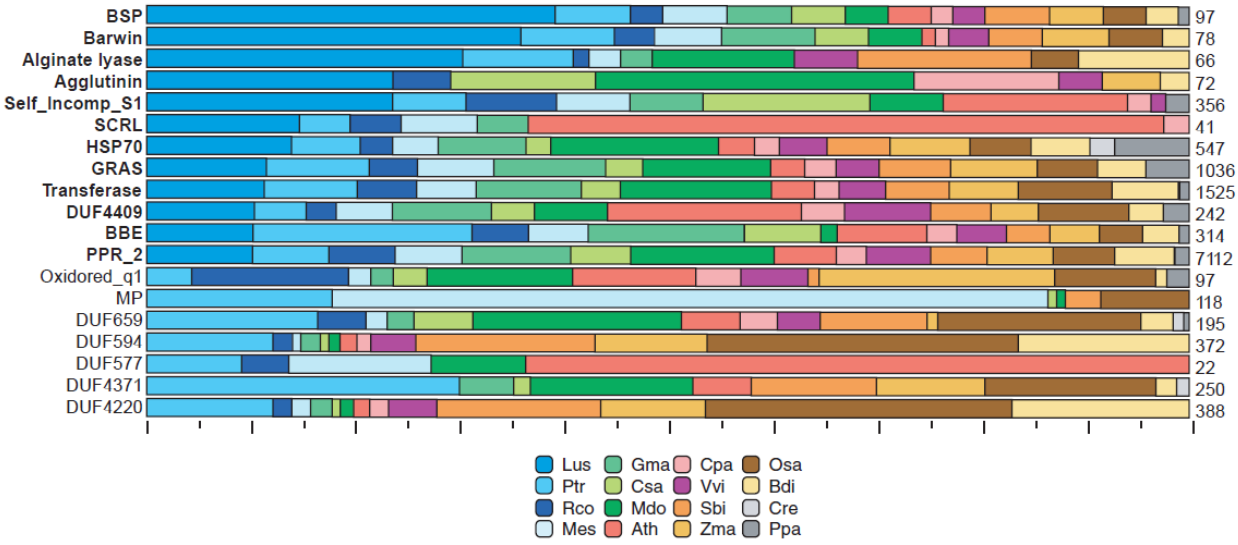

Figure 1.9: Pfam-A domain frequency in flax and other species.

All domains that were more abundant (FDR <0.05) among predicted proteins of flax compared to either *A. thaliana* or *P. trichocarpa* are shown (labels in bold). A subset of

the domains that were significantly less abundant in flax compared to either species is also shown. Additional species for which whole-genome sequence is available are shown for comparison. The width of the colored region indicates the number of genes containing a given Pfam-A domain within each species. A different scale is used for each domain, and the number to the right of each bar indicates the total number of genes represented by that bar. Redundant occurrences of the same domain within the same gene are counted only once. BSP, basic secretory protein; Barwin, PF00967; alginate lyase, PF08787; agglutinin, PF07468; Self_Incomp_S1, plant self-incompatibility protein S1 (PF05938); SCRL, plant self-incompatibility response protein (PF06876); HSP70, heat shock protein 70 (PF00012); GRAS, GRAS family transcription factor (PF03514); transferase, PF02458; DUF4409, domain of unknown function (PF14365); BBE, berberine and berberine-like (PF08031); PPR_2, PPR repeat family (PF13041); Oxidoreq_q1, oxidoreductase (PF00361); MP, viral movement protein (PF01107); DUF659 (PF04937); DUF577 (PF04510); DUF4371 (PF14291); DUF4220 (PF13968). Ath, *Arabidopsis thaliana*; Bdi, *Brachypodium distachyon*; Cpa, *Carica papaya*; Cre, *Chlamydomonas reinhardtii*; Csa, *Cucumis sativa*; Gma, *Glycine max*; Lus, *Linum usitatissimum*; Mdo, *Malus domestica*; Mes, *Manihot esculenta*; Osa, *Oryza sativa*; Ppa, *Physcomitrella patens*; Ptr, *Populus trichocarpa*; Rco, *Ricinus communis*; Sbi, *Sorghum bicolor*; Vvi, *Vitis vinifera*; Zma, *Zea mays* (Wang *et al.*, 2012)

1.12 Amaranthin

Amaranthin is a T-antigen (Galβ1-3GalNAc)-specific lectin, first isolated from the seeds of *A. caudutus* (Rinderle et al., 1989). The family consists of amaranthins derived from the Amaranthus species (genuine amaranthin-lectin) and chimeras found in other plant species. It has at least two types of hololectins (single amaranthin domain, double amaranthin domains) and three types of chimerolectins (single or double amaranthin domain with aerolysin domain, double amaranthin domain with kinase domain) (Van Damme et al., 2008; Puthoff *et al.,* 2005). However it is not known yet whether the amaranthin domains in other proteins possess a lectin activity like the genuine amaranthins. The amaranthin family is fairly widespread among plants although not ubiquitous because no corresponding genes are found in the genomes of *P. patens, A. thaliana, M. truncatula, P. trichocarpa, M. esculenta, O. sativa,* and *S. bicolor* (Van Damme *et al.,* 2008).

Genuine amaranthin protein is a homodimer of 33kDa subunits. Each of the subunits consists of about 150 amino acids and has two tandemly arrayed homologous β-trefoil domains. It appears that the single amaranthin domain does not have sugar binding activity. Furthermore, the head to tail arrangement of the two amaranthin domains is required for forming the carbohydrate binding sites (Rinderle *et al.,* 1990; Transue *et al.,* 1997). No signal peptide could be identified in proteins with amaranthin proteins. Preliminary experiments with EGFP (Enhanced Green Fluorescent Protein) tagged single amaranthin domain containing protein from *P. Persia* in tobacco BY-2 cells showed that the protein is expressed mainly in the nucleus and partially in the cytoplasm (Van

Damme *et al.,* 2011). The physiological role of amaranthin-protein is not yet known (Lannoo *et al.,* 2010).

One of the chimeras, Hfr-2 (Hessian fly responsive), has been characterized to some extent too. This is a gene identified in wheat that contains double amaranthin domains at the N-terminal end and an aerolysin domain at the C-terminal end. Expression analysis showed it to be induced upon insect infestation or following hormone treatment. Methyl jasmonate treatment resulted in a seven-fold up-regulation of Hfr-2 transcript levels 3 days after treatment. Salicylic acid treatment resulted in a small up-regulation (3.7-fold) at 4 days while abscisic acid treatment resulted in a slight increase (2.6-fold) at 2 days (Puthoff *et al.,* 2005).

1.13 Objectives of this thesis

The objective of this research is to define the agglutinin family in flax and investigate its potential functions. For this purpose:

1) The characteristics and possible evolutionary relationships of predicted flax lectins were defined by *in silico* analyses.
2) The expression profile of the 19-member gene family was determined by qRT-PCR in different tissues.
3) Etiolated flax seedlings were treated with hormones and the expression profile of the genes was determined.
4) One of the genes was heterologously expressed in *E. coli*, purified, and assayed in hemagglutination and hemagglutination inhibition assays.

Chapter 2 Materials & methods

2.1 *In silico* analysis

2.1.1 Distribution of the twelve lectin families across all species of phytozome (Deyholos 2013, unpublished data)

Amino acid sequences for predicted genes from whole genome assemblies of 38 plant species were downloaded from phytozome.org (version 9.1, accessed 2013-07-31). Each locus was represented by only one gene model in the downloaded files and in all subsequent analyses. The amino acid sequences were searched for conserved protein domains as defined by Pfam database (version 27.0, accessed 2013-07-31, Punta *et al.*, 2012) using a locally installed PfamScan.pl script (ftp.sanger.ac.uk) with default parameters. The number of genes containing at least one lectin-related domain was calculated after importing the PfamScan.pl results into a MySQL relational database. Data were represented in the same species tree as provided by phytozome.org.

2.1.2 Predicted protein sequences and domain analysis

The CDS (coding sequence) and protein sequences of all 19 agglutinin domain containing genes predicted in the whole-genome shotgun assembly (WGS) of flax (*L. usitatissimum* var. CDC Bethune) were obtained from phytozome.org (version 9.1, accessed 2013-07-31) using the search term "agglutinin". PfamScan confirmed (see section 2.1.1, above) that no other predicted flax genes contained an agglutinin (PFAM PF07468) domain. These 19 genes were designated here as LuALLs (*L. usitatissimum* amaranthin lectin like). The amino acid sequences of each of the 19 LuALLs was used to query the Pfam protein families database using default settings.

2.1.3 EST alignment to gene models

To determine which of the LuALL gene models were supported by experimental evidence, the 19 LuALL nucleotide CDS sequences were used as a query in BLASTN searches against the NCBI-EST database (accessed 2013-07-31) in CLC genomics Workbench 6 (e value $\leq 10^{-25}$; match/mismatch: match 1, mismatch -3; gap cost: existence 5, extension 2).

2.1.4 Prediction of physical and chemical properties

Length, molecular weight and theoretical pI of the LuALL proteins were computed using the Protparam tool in Expasy (Gasteiger *et al.,* 2005). The 'SOSUI' tool of Expasy was used to determine the hydrophobicity of this family of proteins (Hirokawa, Boon-Chieng & Mitaku, 1998).

2.1.5 Prediction of sub-cellular targeting

'TargetP 1.1 server'(Emanuelsson, Nielsen, Brunak & von Heijne, 2000) a neural network-based tool was used to predict the presence of chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP) in the proteins. This program returned a reliability class value ranging from 1 to 5, where 1 indicates the highest confidence (Emanuelsson *et al.,* 2000). 'Plant' was selected in the organism group and cut-off value with specificity >0.90 was chosen for the prediction. Where applicable, the cleavage site prediction (Nielsen, Engelbrecht, Brunak & von Heijne, 1997) was included.

2.1.6 Predicting glycosylation sites

Potential N and O linked glycosylation sites were predicted in the GlycoEP software suite using the Binary Profile of patterns (BPP) and Composition Profile of Patterns (CPP) programs (Chauhan, Rao & Raghava, 2013). These are support vector machine (SVM) algorithms. In both cases, the default SVM threshold value of 0 was used.

2.1.7 Predicting the secondary structure

The secondary structure of the LuALL family of proteins was predicted using YASPIN (Lin, Simossis, Taylor & Heringa, 2005). YASPIN is a HNN (Hidden Neural Network) secondary structure prediction program that uses the PSI-BLAST (Position-Specific Iterated BLAST) algorithm to produce a PSSM (position-specific scoring matrix) for the input sequence, which it then uses to perform its prediction (Lin *et al.,* 2005). NR (Non-redundant) database for PSI-BLAST and DSSP-trained option for NN (Neural network) were used for the prediction.

2.1.8 Phylogenetic analysis

A phylogenetic tree was constructed with the flax proteins and a representative protein of each of the known lectin families using MEGA 5 (Van Damme *et al.,* 2011) (Tamura *et al.,* 2011). Sequences were aligned based on only the lectin domain from each protein (except EEA, for which no significant domain was defined in Pfam). Alignments were done in MEGA 5 using ClustalW (pair-wise alignment- gap opening penalty: 10, gap extension penalty: 0.1; multiple alignment: gap opening penalty: 10, gap extension penalty: 0.2; delay divergent cut-off: 30%). A neighbour joining (NJ) phylogenetic tree was constructed using a Poisson model pair-wise deletion for the treatment of gaps/missing data.

2.2 Measurement of tissue-specific and hormone-responsive transcript expression

2.2.1 Plant growth conditions and tissue collection

Seeds were surface sterilized for 20 minutes in 15% bleach with 1 drop of Tween 20, and then rinsed three times with sterile milliQ water. Seeds were germinated in plastic containers on wet tissue paper. The boxes were placed under light with a 16 hour photoperiod at constant 22^{o}C. After one week, the cotyledons and roots were collected from the seedlings, frozen immediately in liquid nitrogen and stored at -80°C until use. For tissues harvested from older plants, seeds were sown directly on autoclaved potting mixture (peat moss and perlite) in small pots placed in trays and kept in a growth chamber for a daily cycle of 16 hours of light (22°C) and 8 hours of dark (18°C). Leaf and stem samples were collected from 3-4 week old plants. Floral buds and green seeds at the mature embryo stage were collected as they developed. Three biological replicates were obtained for each sample.

2.2.2 Hormones

Methyl jasmonate and salicylic acid (Sigma-Life Sciences) were dissolved in 1 ml absolute ethanol and subsequently diluted in water (500 ml) to get 100 µM and 1 mM concentration respectively. The control sample contained 1 ml absolute ethanol in 500ml of water. As a surfactant, Tween 20 (0.05 %) was added to all the solutions.

2.2.3 Induction experiment with etiolated seedlings

Eight days after imbibition, seedlings were sprayed with a hormone solution or a control. Samples were sprayed again every 24 h for two more days (i.e. three spray treatments in total). Samples were collected after 12 h and 72 h for RNA analysis. The seedlings were

blotted on filter paper, frozen in liquid nitrogen and stored in -80^{o}C until use. Three biological replicates were collected for each sample.

2.2.4 Isolation of total RNA and cDNA synthesis

Total RNA was isolated from cotyledons, roots, leaves, stems and floral buds using either a modified CTAB (cetyl tri-methyl ammonium bromide)-based method (Johnson *et al.*, 2012) or Tri-Reagent (Sigma) protocol combined with the RNeasy mini kit (Qiagen; 74104) or the RNeasy mini kit only. All the methods gave a good yield of RNA (260/280 $\simeq$ 2 and 260/230 $\geq$ 1.8) with most tissues, but the TRizol and the RNeasy mini kit protocols involved fewer steps than the modified CTAB-based method and gave better results with a small amount of tissues (100 mg). RNA (260/280 $\simeq$ 2 and 260/230 $\geq$ 1.6) from green seeds (mature embryo stage) was extracted using the urea-LiCl lysis method (Tai *et al.,* 2004) using 40 mg of sample. For the hormone test, the RNeasy mini kit was used with 100 mg plant tissue as input. In all cases, the residual DNA in the sample was removed using the TURBO DNA-free kit (Ambion). PCR was performed on some of the samples to ensure that no amplifiable DNA was present in the RNA samples after the use of TURBO DNA-free kit. The RNA concentration was determined using a NanoDrop ND-1000 spectrophotometer. A total of 2 µg of total RNA was used for cDNA synthesis using RevertAid H minus Reverse Transcriptase (Thermo Scientific).

2.2.5 qRT-PCR

Quantitative real time RT-PCR (qRT-PCR) was performed to measure relative transcript abundance using an Applied Biosystems 7500 Fast system with a SYBR Green master mix (SYBR® Green I Dye, AmpliTaq Gold® DNA polymerase, dNTPs with dUTP, Passive Reference, and optimized buffer components). Primers were designed based on

assembly version 1.0 of the flax genome (Wang *et al.*, 2012) using Primer Express software (version 3, Applied Biosystems) (Table 2.1). Primer sets were PCR-validated against the primer pairs of the internal control, EF1α (Lus10015070 or Lus10019899), as follows. PCR reactions were carried out in 96 well plates. Each well contained a 10µl reaction volume with 5 µl of master mix and 2.5µl of each of the template cDNA and primer mix (400 nM of each forward and reverse primer). Primer validation was done using the guidelines of Applied Biosystems with template cDNA, diluted by 4^{-1}, 4^{-2}, 4^{-3}, 4^{-4} and 4^{-5} fold. The validated primers were used for the expression analysis with a template cDNA of 62.5 ng (4^{-1} fold) in the 10 µl reaction volume. For all qRT-PCRs, the following program was used: 2 minutes at 95^{o}C followed by 40 cycles of 15 s at 95^{o}C and 1 minute at 60^{o}C. Following the PCR cycles, a dissociation curve was obtained for each reaction. In case of primer validation studies the dissociation curve program consisted of a cycle of 15s at 95^{o}C, 1 minute at 60^{o}C, 15 s at 95^{o}C and 15 s at 60^{o}C. For transcript expression experiments, the dissociation curve program consisted of a cycle of 15 s at 95^{o}C, 20 s at 60^{o}C, 15 s at 95^{o}C and 15 s at 60^{o}C. Each qRT-PCR reaction was run in triplicate. The delta Ct values obtained by the normalization of the expression of the genes to the internal control, GAPDH (Lus10006435) were used to generate a heat map with MeV version 4.9 (http://www.tm4.org/mev.html).

For analysis of transcript expression following hormone treatment, the control samples were used as the calibrator and the fold variation was determined by the delta delta Ct method (Applied Biosystems guidelines). Thus, the transcript expression of each gene for the hormone treatment was measured with respect to its expression in the control sample

at the same time point. GAPDH (Lus10006435) was used for normalizing the data. Control and hormone-treated samples was compared to control samples using ANOVA and t-test in excel.

Table 2.1: Primer pairs for the two internal controls and 19 LuALL genes.

Name of genes	Gene ID	Forward primer	Reverse primer
Control	Lus10015070/ Lus10019899	TTGGATACAACCCCGACAAAA	GGGCCCTTGTACGAGTCAAG
Control	Lus10006435	GGGCTGGAATCGCATTGA	CCCATTCGTTGTCGTACCAAGA
LuALL1	Lus10005398	TGCAGACCAACCACCAAGTG	AAGAGCTGCATCCCGTTGTC
LuALL2	Lus10005397	ACCAACGACGGAACCAGTTTC	GGATCAGCAGACTGCGCC
LuALL3	Lus10005395	CCTACACCAACGACGGAACTG	GCATCAGCACTCAGACTCTCACC
LuALL4	Lus100029186	ATGGTGGCATTCGATTTTGG	AGGTGGACCACTTCTAAACAAACTG
LuALL5	Lus100029184	TATCCAAGAGGACCGGTCTGA	CACCGCCGTTGACTTTTCAT
LuALL6	Lus10029182	GATTCGCGGGTCGAGTAATC	CTTCTGGTCGCCATTATGATCA
LuALL7	Lus10016109	GCCTACGGCTGTGCATGAG	CTGCCACAGTTCCCAAAACTC
LuALL8	Lus10031082	GAGGCACCACCCCAATCTTA	CGCATCATCAACCATTAACACA
LuALL9	Lus10041636	GTGGCATATTAAATCAGTATCCAAGG	TATCCGTGTCTAAAATTAATAGAACCACTTAC
LuALL10	Lus10010698	TGGTGATTCCAGTCAGCAAGTTACT	AGATGATGTTGAAACTCACCCGAT
LuALL11	Lus10010702	TGATCATCGCAGCAGCCA	TCCCCTTCTGATTGAGCGG
LuALL12	Lus10024084	ATTAAGCACAGCTAATATATGCGAGATC	CTTTATAATAAACATACCAAATTAACCCACAC
LuALL13	Lus10020808	GCCCACCCTAGTCTGCTTTAAA	TGTGAACATAGCCCGATGCA
LuALL14	Lus10021453	ATGCCTATAAATAGCAGGGAATCAA	AGATTTGCTCTTCAGGGTGAACTG
LuALL15	Lus10025503	CAGCAGCCGACAGTCCAAA	AGAAGCCGAACCGAGATCAA
LuALL16	Lus10022894	CGGTGGTTAGGGCTGTCAATT	TCCCTCATTATTACCCACTCACAA
LuALL17	Lus10024934	CAGAGATTGCCGAGGTTCGT	AGACCTTGCCGGATTGCATA
LuALL18	Lus10024931	ATCGTCCTTGAGCGGCG	ACGTGCCTGAAAACAACGG
LuALL19	Lus10020249	TGGAACCCATGAAGAAGGTGA	AACGGAACCTCGCACGTAATAG

2.3 Heterologous expression of LuALL in *E. coli*

Analysis of the codon usage of native LuALL12 was done using codonW (version 1.4.4, http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::codonw). The CDS sequence of LuALL12 was optimized for expression in *E. coli* using a proprietary algorithm (BioBasic, Inc.). The codon optimized sequence was synthesized by BioBasic Inc., and was cloned into pET-22b (+) vector (Novagen) using the restriction sites *Nco* I and *Xho* I, then transformed by heat-shock (Bergmans, Van Die & Hoekstra, 1981) into *E. coli* Rosetta (DE3) pLysS (Novagen). Transformants were selected on Luria-Bertani (LB) medium supplemented with chloramphenicol (34 µg/ ml) and ampicillin (100 µg/ ml). The transformants were confirmed by PCR and were stored in glycerol at -80 ˚C.

2.3.1 Protein expression and extraction

2.3.1.1 Small scale protein expression and identification using western-blot

Transformed *E. coli* cells from glycerol stocks were inoculated in 3 ml 2xYT (Yeast Extract Tryptone) liquid medium supplemented with 34 µg /ml of chloramphenicol and 100 µg/ml of ampicillin and cultured overnight at 37 ˚C. 1 ml of this overnight culture was used to inoculate 100 ml of the 2xYT medium with 34 µg /ml of chloramphenicol and 100 µg/ml of ampicillin and was grown at 37 ˚C until an OD_{600} of 0.75 was attained. Then, isopropyl-β-d-thiogalactopyranoside (IPTG, 0.5 mM final concentration) was added to induce protein expression, and the culture was incubated for 2 h. Protein was extracted from 2 ml of this culture. Bacteria were pelleted at 9000X gravity for 20 minutes at 4 ˚C. The supernatant was discarded and the cells were suspended in 100 µl of

Tris-sucrose extraction buffer (200 mM Tris-HCl pH 8.0, 20 % (w/v) Sucrose). The mixture was kept on ice for 1 h with occasional shaking. Then the mixture was centrifuged at 15000X gravity for 30 minutes at 4 ˚C. The supernatant contained the soluble periplasmic extract and the pellet contained the inclusion bodies. The supernatant and the pellet were subjected to SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) and western-blot analysis.

For SDS-PAGE, a discontinuous polyacrylamide gel was prepared with 12.5% resolving gel (3.1 mL of polyacrylamide solution (40%), 2.5 mL of separating gel buffer (1.5 M Tris pH 8.8), 50 µl of ammonium persulfate, AP,(10% w/v), 100 µl of Sodium dodecyl sulphate, SDS (10%), 4.2 mL of milliQ water and 60 µl of tetramethylethylenediamine, TEMED) and 6.7% stacking gel (0.3 ml of polyacrylamide solution (40%), 0.44 ml of stacking gel buffer (1.5 M Tris pH 6.8), 10 µl of AP (10% w/v), 18 µl of sodium dodecyl sulphate, SDS (10% w/v solution), 0.98 mL of milliQ water and 10 µl of tetramethylethylenediamine, (TEMED). For gel electrophoresis, the protein solution was mixed with an equal volume of 2x sample solution (100 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.2% of bromophenol blue and 200 mM dithiothreitol, DTT) and boiled at 95˚C for 10 minutes. The gel was run in electrophoresis buffer (14.4 g glycine, 10 g SDS, 3 g of Tris-HCl in milliQ water to a final volume of 1 L) using the Mini-Protean Tetra Cell (Bio-Rad) at 200 V in room temperature untill the dye reached the end of the gel. The gel was stained in staining buffer (90 ml methanol, 90 ml water, 20 ml acetic acid and 0.5 g Coomassie blue) for at least 2 h and then destained with destaining buffer (90 ml methanol, 90 ml water, 20 ml acetic acid) at room temperature.

Proteins in both fractions were separated by electrophoresis and electrotansferred to PVDF membrane at 25 V for 1 h (transfer buffer: 14.4 g glycine, 3 g of Tris, 1 g of SDS, 200 ml of methanol to attain a final volume of 1 L). The membrane was immersed in 5% skimmed milk in a TTBS solution (0.02 M Tris-HCl (pH 7.6), 0.15 M NaCl, 0.05% Tween-20), overnight at 4 ˚C. Then the membrane was washed for 10 min in a TTBS solution. The membrane was then incubated with the anti-6X HIS epitope tag (Rabbit) antibody (GenScript,1:5000 dilution) in 1% skimmed milk in a TBS solution (0.02 M Tris-HCl, pH 7.6, 0.15 M NaCl) for 1 h at room temperature and subjected to four 5 min washes in a TTBS solution. Next, it was incubated with horseradish peroxidase-conjugated goat anti rabbit antibody (Rockland) (1: 8000) for 45 min at room temperature, and subjected to four 5 min washes in a TTBS solution. Finally the membrane was washed with TBS solution for 5 min. The blot was developed with chromogenic substrate TMB (3, 3', 5, 5'-tetramethylbenzidine) (TMB Substrate kit for Peroxidase; Vector Laboratories Inc).

2.3.1.2 Determining conditions for producing proteins in the periplasmic fraction

The bacterial cells from glycerol were inoculated in 3 ml starter culture of 2xYT liquid medium supplemented with 34 µg /ml of chloramphenicol and 100 µg/ml of ampicillin. Then it was cultured at 37 ˚C overnight. 700 µl of the overnight culture was then transferred to 70 ml of 2xYT medium and cultured until an OD_{600} of 0.47 was obtained. This was then divided into three portions, induced by 0.5 mM IPTG and cultured at 37 ˚C, 23 ˚C and 18 ˚C for different time periods. For each temperature there were control

samples with the uninduced, transformed bacteria. Protein extraction was done as described in 2.3.1.1.

2.3.1.3 Protein expression in 1L 2xYT medium and extraction of the soluble protein from periplasmic extract

Bacterial cells from glycerol stocks were inoculated in a 3 ml starter culture of 2xYT liquid medium supplemented with 34 µg /ml of chloramphenicol and 100 µg/ml of ampicillin. This starter culture was grown at 37 ˚C for 3 h. After 3 h, 30 µl of the starter culture was added to 30 ml of 2xYT medium with the same concentration of antibiotics and allowed to grow overnight at 37 ˚C. A 10 ml aliquot of the overnight culture was then transferred to 1000 ml of 2xYT media and cultured until an OD_{600} of 0.6-0.7 was obtained. IPTG was then added to the culture medium at a final concentration of 0.5 mM to induce the expression of the protein and allowed to grow for 20 h at 18 ˚C. Protein extraction was done as described in 2.3.1.1.

2.3.2 Purification of the protein and confirmation by peptide mass fingerprinting

The extracted protein was purified by nickel affinity chromatography as follows. The supernatant containing the soluble periplasmic extract was subjected to dialysis against the binding buffer (50 mM Tris-HCl, 300mM NaCl and 10 mM imidazole). After that, centrifugation was done at 15000X gravity for 30 min and the supernatant was collected. To the protein solution, 0.05 % Tween-20, 15 mM β-mercaptoethanol and 500 uL of Ni-NTA agarose (Qiagen) was added and left to rock overnight at 4 ˚C.

For purification, a Biorad poly-prep chromatographic column was first equilibrated with equilibration buffer (50 mM Tris-HCl, 300 mM NaCl and 10 mM imidazole). The entire

protein solution was then passed through the column. Next, the column was washed four times with 1 CV (column bed volumes) of wash buffer I (50 mM Tris-HCl, 300 mM NaCl and 20 mM imidazole). Then the column was washed twice with 1 CV of wash buffer II (50 mM Tris-HCl, 300 mM NaCl and 40 mM imidazole). This was followed by one elution with elution buffer I (50 mM Tris-HCl, 300mM NaCl and 100 mM imidazole) of 0.1 CV. Then a second elution was performed with four 0.1 CV of elute buffer II (50 mM Tris-HCl, 300mM NaCl and 250 mM imidazole). All the fractions were collected for SDS-PAGE analysis.

The protein in elution buffer II was cut from the gel and sent for peptide mass fingerprinting (The Institute for Biomolecular Design, University of Alberta). Briefly, the excised gel bands were destained twice in 100 mM ammonium bicarbonate/acetonitrile (50:50). The samples were then reduced (10 mm BME in 100mm bicarbonate) and alkylated (55mM iodoacetamide in 100 mm bicarbonate). After dehydration, enough trypsin (6ng/ul) was added to just cover the gel pieces and the digestion was allowed to proceed overnight (~16 hrs.) at room temperature. Tryptic peptides were first extracted from the gel using 97% water/2% acetonitrile/1% formic acid followed by a second extraction using 50% of the first extraction buffer and 50% acetonitrile.

Fractions containing tryptic peptides dissolved in aqueous 25% v/v ACN and 1% v/v formic acid were resolved and ionized by using nanoflow HPLC (Easy-nLC II, Thermo Scientific) coupled to the LTQ XL-Orbitrap hybrid mass spectrometer (Thermo Scientific). Nanoflow chromatography and electrospray ionization were accomplished by

using a PicoFrit fused silica capillary column (ProteoPepII, C18) with 100 μm inner diameter (300 Å, 5 μm, New Objective). Peptide mixtures were injected onto the column at a flow rate of 3000 nL/min and resolved at 500 nL/min using 60 min linear gradients from 0 to 45% v/v aqueous ACN in 0.2% v/v formic acid. The mass spectrometer was operated in data-dependent acquisition mode, recording high-accuracy and high-resolution survey Orbitrap spectra using external mass calibration, with a resolution of 60 000 and *m*/*z* range of 400–2000. The ten most intense multiply charged ions were sequentially fragmented by using collision induced dissociation, and spectra of their fragments were recorded in the linear ion trap; after two fragmentations all precursors selected for dissociation were dynamically excluded for 60 s. Data were processed using Proteome Discoverer 1.3 (Thermo Scientific) and the flax and Uniprot *E. coli* protein database was searched using SEQUEST (Thermo Scientific). Search parameters included a precursor mass tolerance of 10ppm and a fragment mass tolerance of 0.8Da. Peptides were searched with carbamidomethyl cysteine as a static modification and oxidized methionine and deamidated glutamine and asparagine as dynamic modifications.

2.3.3 Quantification of protein

The concentration of the purified protein was determined by Qubit® 2.0 Fluorometer (Invitrogen) according to the manufacture's instruction.

2.4 Assays performed with the purified protein

2.4.1 Circular dichroism (CD)

CD spectra were generated at 20 ˚C with an Olis Cary-17 spectrophotometer fitted with an Olis circular dichroism module, calibrated with camphorsulphonic acid. Cylindrical quartz cells of 0.2 mm path length were used for measurements in the region of 260-185 nm. A 2 nm bandwidth for the slit was used, which was kept constant throughout the scanning region by automatic adjustment of slit width. For CD, a concentration of 0.75 mg/ml of the protein was used in 50 mM Na_2HPO_4 buffer. The data was expressed as mean residue ellipticities in degrees cm^2 d mol^{-1}, with a mean residual weight of 111.46 g/mol for the protein and analysed using CAPITO (CD Anaylsis and Plotting Tool) (Wiedemann, Bellstedt & Görlach, 2013).

2.4.2 Hemagglutination assay

Hemagglutination assays were performed using rabbit erythrocytes (Rockland). These were washed and resuspended to make a 2% suspension in PBS buffer pH 6.8 (0.15M NaCl and 0.02 M potassium phosphate buffer, pH 6.8). 100 µl of the protein sample (400 µg) was serially diluted from 2^{-1} to 2^{-12} with the PBS solution in a microtitre plate. In each of the wells, 100 µl of red blood cells were added and the plate was incubated at room temperature for 1.5 h to observe hemagglutination and quantify the protein in terms of hemagglutination units (HU). One HU is defined as the lowest concentration of lectin that causes visible erythrocyte agglutination (Zuo, Fan, Wang, Zhou, & Li, 2012). GNA

(Sigma) was used as the positive control and protein buffer (50 mM Na_2HPO_4, 150 mM NaCl) without protein in PBS was used as the negative control.

2.4.3 Hemagglutination inhibition assay

Amino sugars, N-acetylgalactosamine (GalNAc) and N-acetylglucosamine (GlcNAc), were dissolved in PBS pH 6.8. 25 µl of protein (100 µg) was added to 25 µl of the amino sugar solution (400 mM) and incubated at room temperature for an hour. 50 µl of 2% rabbit red blood cells was added to the mixture and was incubated at room temperature for an hour. The wells were then compared to a non-sugar sample. A negative control was used as described in 2.4.2.

2.4.4 Native PAGE

A continuous (10%) PAGE gel (2.5 mL of polyacrylamide solution (40%), 2.5 mL of separating gel buffer (1.5 M Tris pH 8.8), 100 µl of AP (10% w/v), 5 mL of milliQ water and 20 µl of TEMED) was prepared as described by Arndt, Koristka, Bartsch & Bachmann, 2012. For native gel electrophoresis, the protein solution was mixed with an equal volume of 2x sample solution (0.187M Tris-HCl pH 6.8, 30% glycerol and 80 µg/ml of Bromophenol Blue). The gel was run in cold electrophoresis buffer (14.4 g of glycine and 3 g of Tris-HCl in milliQ water to a final volume of 2 L) at 150 V at 4˚C untill the dye reached the end of the gel. Staining and destaining were done as described for SDS-PAGE.

Chapter 3 Results

3.1 *In silico* analysis

3.1.1 Distribution of lectin domains across plant species of phytozome (Deyholos 2013, unpublished data)

The distribution of major lectin families among a broad sample of plant species (including algae) was analysed. Predicted genes from 38 whole genome assemblies were scanned for the presence of conserved domains typical of each of 12 different families of lectins (classification of Van Damme *et al.*, 2008). Results showed that although most lectin families are represented in most angiosperms (and also in some non-seed plants), a few lectin families were restricted to particular lineages (Figure 3.1). Notably, proteins containing one or more agglutinin domains (PF07468) were found in only 14/38 of the species surveyed (including a lycophyte and 13 angiosperms). The occurrence of agglutinin domain-containing proteins appeared to be a homoplastic trait; for example, these proteins were found in species as diverse as the spikemoss *Selaginella moellendorffii* and flax (*Linum usitatissimum*), but not in species more closely related to flax, such as poplar (*Populus trichocarpa*) or cassava (*Manihot esculenta*). Furthermore, the number of agglutinin domain genes varied greatly between genome species, from a minimum of two to a maximum of 22, suggesting several independent expansions of this gene family. Flax contained 19 predicted genes with one or more agglutinin domains. The agglutinin domain is typical of amaranthin-type lectins. Therefore, we will refer to these genes as LuALLs (*Linum usitatissimum* amaranthin lectin like). Note, however, that even in species such as flax where the agglutinin-domain containing gene family has

greatly expanded, these represent only a minority of all of the lectin-encoding genes predicted in the genome.

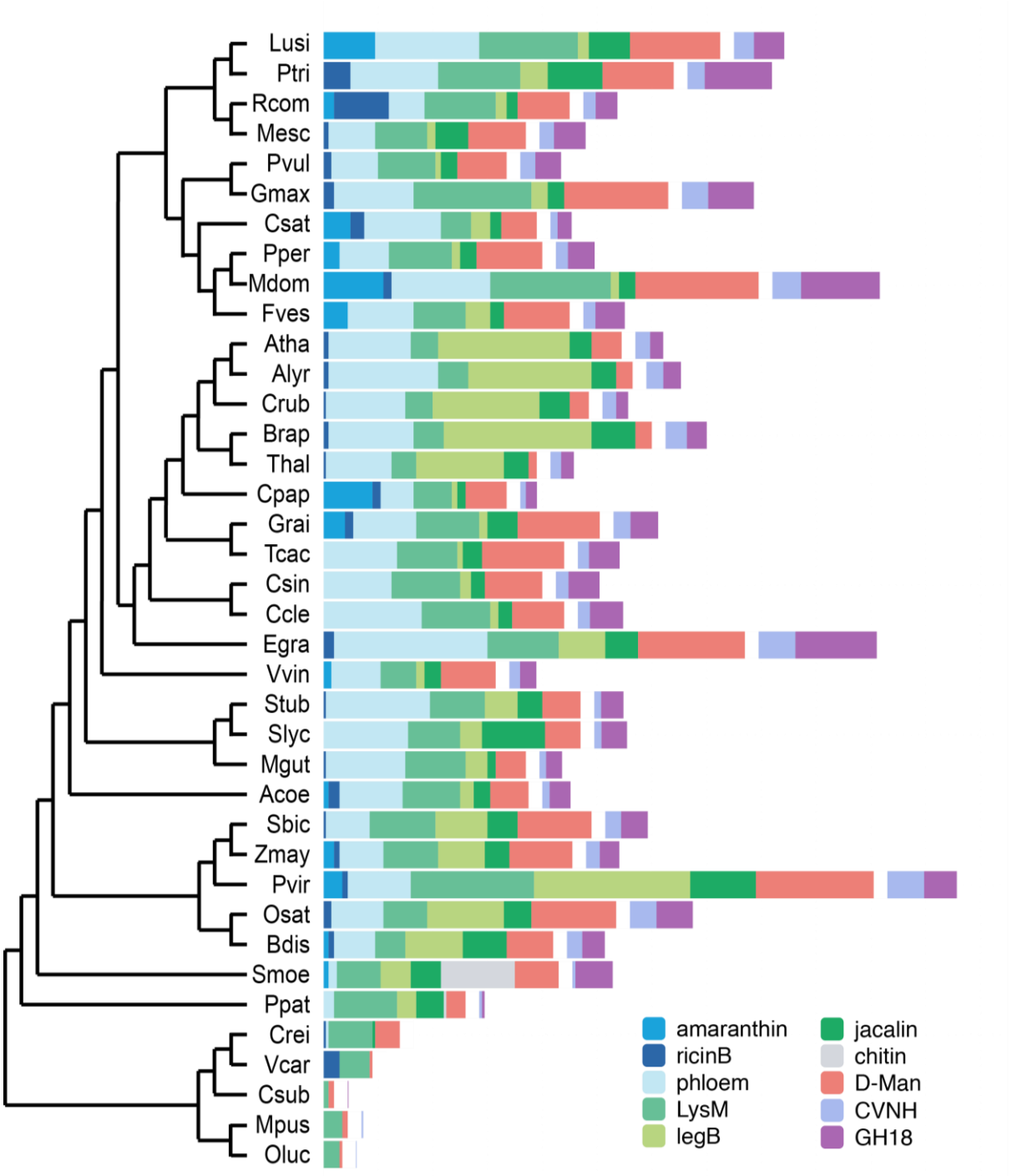

Figure 3.1: Lectin domain frequency in flax and other species. Phloem, NICTABA; legB, legume; chitin, hevein; D-Man, GNA; CVNH, cyanovirin; GH18, class V chitinase homologs; Lusi, *Linum usitatissimum*; Ptri, *Populus trichocarpa*; Rcom, *Ricinus communis*; Mesc, *Manihot esculenta*; Pvul, *Phaseolus vulgaris*; Gmax, *Glycine max*;

Csat, *Cucumis sativus*; Pper, *Prunus persica*; Mdom, *Malus domestica*; Fves, *Fragaria vesca*; Atha, *Arabidopsis thaliana*; Alyr, *Arabidopsis lyrata*; Crub, *Capsella rubella*; Brap, *Brasicca rapa*; Thal, *Thellungiella halophila*; Cpap, *Carica papaya*; Grai, *Gossypium raimondii*; Tcac, *Theobroma cacao*; Csin, Citrus sinensis; Ccle, Citrus clementina; Egra, Eucalyptus grandis; Vvin, *Vitis vinifera*; Stub, *Solanum tuberosum*; Slyc, *Solanum lycopersicum*; Mgut, *Mimulus guttatus*; Acoe, *Aquilegia coerulea*; Sbic, *Sorghum bicor*; Zmay, *Zea mays*; Pvir, *Panicum virgatum*; Osat, *Oryza sativa*; Bdis, *Brachypodium distachyon*; Smoe, *Selaginella moellendorffii*; Ppat, *Physcomitrella patens*; Crei, *Chalamydomonas reinhardtii*; Vcar, *Volvox carteri*; Csub, *Coccomyxa subellipsoidea*; Mpus, *Micromonas pusilla*; Oluc, *Ostreococcus lucimarinus* (Deyholos 2013, unpublished data).

3.1.2 Domain structure of predicted flax agglutinins

The 19 LuALL amino acid sequences were searched for conserved domains that might provide insights into their functions. As expected, each of the LuALLs contained at least one PF07468 agglutinin domain, and moreover, four LuALLs contained two of these domains (Table 3.1). Some of the LuALLs contained other conserved domains, including a Pathogenesis Related (PR) protein Bet v I domain (PF00407), which was present in five genes, and an aerolysin domain (PF01117), which was present in one gene. PR Bet v I proteins are expressed in seed plants under various conditions, especially in response to biotic and abiotic stress (van Loon, Rep & Pieterse, 2006; Srivastava, Fristensky & Kav, 2004; Liu & Ekramoddoullah, 2006**;** Krishnaswamy *et al.*, 2008). Aerolysin is a cytolytic bacterial toxin that is capable of forming pores in the cell membrane and causing cell lysis (Buckley, Halasa & MacIntyre, 1981; Parker, van der

Goot & Buckley, 1996). Thus the LuALLs can be described as containing merolectins, hololectins, and three types of chimerolectins (Figure 3.2). The merolectins and hololectins included proteins with only the single and double amaranthin agglutinin domains respectively. The chimerolectins consisted of single or double amaranthin agglutinin domain fused to Bet v I, and a single amaranthin domain fused to an aerolysin domain.

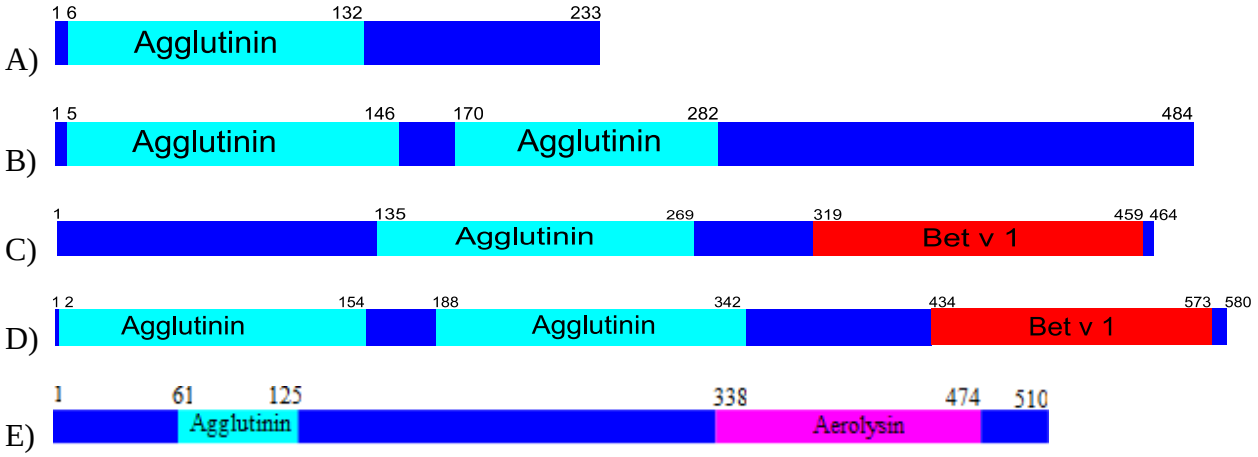

Figure 3.2: A pictorial depiction of the five different types of lectins comprising the LuALL family of proteins. A and B represents the merolectin and hololectin, respectively. C, D and E represent the three different types of chimerolectins.

3.1.3 EST evidence of gene expression

To determine which of the predicted LuALL genes were supported by evidence of transcript expression, the LuALL CDS were aligned to the NCBI-EST database, which contains 286,852 *L. usitatissimum* ESTs. 74.8% of the ESTs were from seed tissues that included developing embryos (globular, heart, torpedo, cotyledon and mature stages), seed coats (globular and torpedo stages) and endosperm (pooled globular to torpedo

stages) (Venglat *et al.*, 2011). The remainder of the ESTs were from flowers, etiolated seedlings, leaves, and stem tissues (Day *et al.*, 2005; Roach & Deyholos, 2007; Venglat *et al.*, 2011). As shown in Table 3.1, 992 ESTs aligned with one or more of a group of eight LuALLs under high stringency (BLASTN e-value $\leq 10^{-25}$). The remaining 11 LuALLs may also be expressed under conditions and in tissues not represented in the EST database.

The tissue sources of the ESTs that aligned to predicted LuALLs is shown in Table 3.2. Most of the LuALLs in this table could be detected in a wide range of the tissues sampled; five LuALLs were detected in embryos, seed coats, leaves, and stems. One LuALL (LuALL7) was detected only in leaves, and two LuALLs (LuALL16 and 18) were detected only in seeds. Both LuALL16 and 18 appeared to be abundant in seeds, with 492 and 250 ESTs found, respectively, for each.

Table 3.1: Different properties of the agglutinin-domain containing proteins in flax determined using 'Protparam', 'Pfam' and 'BLAST'.

Gene names	length (aa)	MW (kDa)	pI	Number of agglutinin domains	Other significant pfam domains	Number of hits with EST database
LuALL1	194	22.54	6.30	1	absent	7
LuALL2	219	24.75	6.53	1	absent	62
LuALL3	226	25.47	5.60	1	absent	62
LuALL4	1024	118.4	9.03	2	Pathogenesis-related protein. Bet V 1	0
LuALL5	580	66.3	5.88	2	Pathogenesis-related protein. Bet V 1	0
LuALL6	410	46.8	8.90	1	Pathogenesis-	0

					related protein. Bet V 1	
LuALL7	503	57.8	5.91	1	absent	1
LuALL8	493	55.65	5.99	2	absent	0
LuALL9	233	26	4.71	1	absent	59
LuALL10	464	51.55	8.20	1	Pathogenesis-related protein. Bet V 1	0
LuALL11	717	82.5	9.73	1	Pathogenesis-related protein. Bet V 1	0
LuALL12	231	25.7	4.81	1	absent	59
LuALL13	510	57.1	6.20	1	aerolysin	0
LuALL14	120	13.8	7.77	1	absent	0
LuALL15	484	54.85	4.75	2	absent	0
LuALL16	255	29.4	5.31	1	absent	492
LuALL17	475	53.5	6.65	1	absent	0
LuALL18	264	30.4	5.65	1	absent	250
LuALL19	483	54.65	6.73	1	absent	0

Table 3.2: Expression of agglutinin genes in different organs/tissues based on EST libraries. (✓) denotes expression. Embryo heart stage (EMH), embryo globular stage (EMG), embryo torpedo stage (EMT), embryo cotyledon stage (EMC), embryo mature stage (EMM), seed coat globular stage (SCG), seed coat torpedo stage (SCT), leaf (LF), etiolated seedling (ETS), stem (STM), stem peel (STP), Outer fibre-bearing tissue mid-flowering stage (OFM).

Gene names	Expression in organs/tissues											
	EMH	EMG	EMT	EMC	EMM	SCG	SCT	ETS	STM	STP	LF	OFM
LuALL1	✓							✓	✓	✓	✓	✓
LuALL2	✓	✓				✓	✓	✓	✓	✓	✓	✓
LuALL3	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓
LuALL7											✓	
LuALL9	✓	✓				✓	✓	✓	✓	✓	✓	✓
LuALL12	✓	✓				✓	✓	✓	✓	✓	✓	✓
LuALL16				✓	✓	✓	✓					
LuALL18				✓	✓							

3.1.4 Physical and chemical properties

The predicted isoelectric points (pI) of the LuALLs ranged from acidic to basic (4. 71 - 9.73), and the predicted sizes also varied widely in terms of both length (120 – 1024 amino acids) and molecular weight (13.8 – 118.4 kDa) (Table 3.1). For comparison, genuine amaranthin protein isolated from *A. caudutus* is a homodimer of 303 amino acids (Rinderle *et al.*, 1989; Transue *et al.*, 1997), with each subunit comprising about 150 amino acids. Predicted ALLs (PF07468 domain containing proteins) from *M. domestica* and *R. communis*, ranged in length from 109-658 and 394-476 amino acids, respectively (http://www.phytozome.net). Thus, the LuALLs had a wider range for length and molecular weight compared to other species.

Next, the predicted hydrophobicity of the LuALLs was calculated. Membrane proteins have at least one hydrophobic transmembrane helix that distinguishes them from soluble proteins (Hirokawa, Boon-Chieng & Mitaku, 1998). The average hydrophobicity of membrane proteins is ≥ 0 as calculated by the SOSUI algorithm (Hirokawa *et al.*, 1998; Mitaku & Hirokawa, 1999). The hydrophobicity of each of the LuALLs was calculated to be < 0, and thus all 19 LuALLs were predicted to be soluble proteins (Table 3.3).

Table 3.3: The solubility of the flax proteins predicted using 'SOSUI'. Average hydrophobicity was rounded to 2 decimal places.

Gene names	Average hydrophobicity	Soluble
LuALL1	-0.54	Y
LuALL2	-0.43	Y
LuALL3	-0.47	Y
LuALL4	-1.42	Y
LuALL5	-0.85	Y
LuALL6	-0.83	Y

LuALL7	-0.78	Y
LuALL8	-0.29	Y
LuALL9	-0.50	Y
LuALL10	-0.79	Y
LuALL11	-0.95	Y
LuALL12	-0.43	Y
LuALL13	-0.29	Y
LuALL14	-0.80	Y
LuALL15	-0.32	Y
LuALL16	-0.74	Y
LuALL17	-0.34	Y
LuALL18	-0.77	Y
LuALL19	-0.34	Y

3.1.5 Sub-cellular localization

Sub-cellular localization of the LuALLs proteins was predicted using TargetP (Emanuelsson *et al.,* 2000). TargetP can predict whether a given protein will be targeted to the secretory pathway via ER/Golgi, and whether it has mitochondrial targeting peptides (mTP) or chloroplast transit peptides (cTP). As shown in Table 3.4, except for two of the proteins, most LuALLs were not predicted to contain signal peptides. LuALL10 was predicted to be targeted to the chloroplast with a presequence of 34 amino acids. LuALL11 was predicted to be mitochondria-specific with a presequence of 120 amino acids.

Table 3.4: Predicted sub-cellular localization of the flax family according to 'TargetP'. mTP = mitochondrial targeting peptide cTP =chloroplast transit peptide ; 'Other' means any other location other than mitochondria, chloroplast and secretory pathway. '*' means that the category had a winning output score (the highest score amongst the three, mTP, cTP and other), below the cut-off score. Reliability class (RC) value ranges from 1 to 5, where 1 indicates the highest confidence.

Gene names	TargetP	
	Location	Reliability class (RC)
LuALL1	Other	5
LuALL2	Other	3
LuALL3	Other	3
LuALL4	* (mTP)	5
LuALL5	Other	4
LuALL6	Other	4
LuALL7	Other	4
LuALL8	*(mTP)	5
LuALL9	Other	4
LuALL10	cTP	2
LuALL11	mTP	3
LuALL12	Other	4
LuALL13	*(other)	5
LuALL14	Other	4
LuALL15	*(mTP)	4
LuALL16	Other	3
LuALL17	*(mTP)	3
LuALL18	Other	4
LuALL19	*(mTP)	5

3.1.6 Glycosylation

In plants there are two types of glycosylation: N and O-glycosylation. N-glycosylation occurs strictly on specific Asp residues of proteins that enter the secretory pathway. O-

glycosylation can occur in secreted proteins (on Ser and Thr; hydroxyproline) as well as on cytosolic/nuclear proteins (on Ser and Thr).

The potential LuALL N and O-glycosylation sites according to GlycoEP (Chauhan *et al.*, 2013) are shown in Table 3.5. As most of these proteins were not predicted to enter the secretory pathway (Table 3.4), the potential N-glycosylation sites are not likely to be glycosylated. On the other hand, the predicted O-glycosylation sites could be glycosylated in proteins that are localized to the cytosolic compartments of the cell.

Table 3.5: Predicted N and O-glycosylation sites based on 'GlycoEP'.

Gene names	Glycosylation	
	N	O
LuALL1	-	T3, T5, S57,T58, S59,T62, T89,S91,T97,T100,T102, S111, T123
LuALL2	-	T2, T3, S57, T58, T62, S84, S87, T89, S92, T116, S124, T152, S153
LuALL3	N142	T2, T3, S57, T58, T62, S84, S87, T89, S92, T116, S124, T152
LuALL4	N179,N512, N524,N596,N632,N668, N803,N825, N828, N836, N983	S73, S74, T75, S80, S82, S83, S309, S360, T445, T479, T730, T797, T883, S891, S915, S918, T1023
LuALL5	N428, N503, N541	S140, S295, T443, S451, T452, S475, S478, S483, T521, T575
LuALL6	N84, N371	S224, T229, T282, T292, S305, S308,
LuALL7	N283, N355, N378	S2, S44, S80, T99, T101, T209, T212, T224, T225, T257, T260, T325, T360, S366, S373, S374, S375, T376, T380, T381,T382,T383, S386, T387, T435, S438, T486, T502
LuALL8	N46, N137, N251, N332, N348, N349, N366, N367, N372, N412	S16, T106, S253, T257, S289, T297, T299, T307, T334, S338, T343, T350, T351, S408, T414, T415, T416, T418, S430, T442, T445, S451, T460, T465
LuALL9	-	T2, S3, T4, S16, S58, T59, S60, S88, T90, S93, T117, T125, S195, S197
LuALL10	N425	S2, S3, S4, T5, T6, T7, T27, S28, S29, T119, S122, S256, S270, S280, S295, S324, S335, T336,
LuALL11	N378, N427, N676	S2, S158, S177, T262, S300, S307, T369, T463, S487,

		T563, S564, T565, S569, S610, T716
LuALL12	-	T2, S3, T4, S16, S58, T59, S63, S88, T90, S93, T102, S103, T104, T117, T125, S193
LuALL13	N18, N101, N290, N361	S2, S5, S6, S40, T41, T184, T205, T206, T207, T208, T218, S238, T245, T277, S281, T356, S366, T367, T401, S403, S443, T444, S445, T447, T454, S508, S509
LuALL14	-	S2, S44, T114
LuALL15	N130, N223, N339, N365	T4, T5, S54, T59, S65, T153, T218, S252, T257, T358, T360, S386, S389, T419, T477, T479
LuALL16	N118	S2, S4, T5, S60, T61, S104, S105, S106, T126, T136
LuALL17	N86, N206, N243, N324, N462	S2, T219, S223, S228, T229, T230, T232, S235, S238, T327, T331, T343, S344, S372, S374, S388, S391, S423, T458, T464, S468
LuALL18	N119, N135	S2, T4, T5, S60, T61, S104, S134, T145, S176
LuALL19	N77, N94, N219, N336, N440, N471	T96, S100, T101, T240, T243, T245, T280, T284, T286, S335, S356, S381, S385, T386, S396, S397, T403, T404, T405, T425, T431, S432, T434, T473, S477, S483

3.1.7 Secondary structure

The secondary structure of LuALL family of proteins was predicted using YASPIN (Table 3.6, Lin *et al.*, 2005). Based on these results, the LuALL proteins could be divided into two classes. Class 1 structures were enriched in β-sheets, with more than twice as many predicted β-sheets as α-helices. Class 2 structures had nearly equal proportions α-helices and β-sheets. Twelve proteins, LuALL1, 4, 5, 6, 7, 8, 10, 13, 14, 15, 17 and 19 belonged to the first class. All four proteins having two agglutinin domains belonged to this class. The second class included seven proteins: LuALL2, 3, 9, 11, 12, 16 and 18.

Table 3.6: Secondary structure prediction of the LuALL family of proteins using 'YASPIN'. The percentage of β-sheet and α-helix was calculated by dividing the number of amino acid residues with a particular conformation with the total number of amino acid residues of the protein and multiplying it with 100.

Gene name	Gene ID	β-sheet (%)	α-helix (%)	Class
LuALL1	Lus10005398	42.27	5.15	1
LuALL2	Lus10005397	31.05	25.11	2
LuALL3	Lus10005395	28.76	26.55	2
LuALL4	Lus10029186	16.8	6.05	1
LuALL5	Lus10029184	30.86	7.24	1
LuALL6	Lus10029182	25.37	11.22	1
LuALL7	Lus10016109	50.7	5.17	1
LuALL8	Lus10031082	52.33	1.42	1
LuALL9	Lus10041636	24.03	29.61	2
LuALL10	Lus10010698	28.45	10.99	1
LuALL11	Lus10010702	23.85	14.23	2
LuALL12	Lus10024084	30.74	22.08	2
LuALL13	Lus10020808	52.75	5.1	1
LuALL14	Lus10021453	40	2.5	1
LuALL15	Lus10025503	52.27	5.37	1
LuALL16	Lus10022894	21.57	33.33	2
LuALL17	Lus10024934	48.42	7.79	1

LuALL18	Lus10024931	22.72	28.03	2
LuALL19	Lus10020249	49.07	4.97	1

3.1.8 Phylogenetic analysis of the flax agglutinin domains with the lectin domains of other lectin families

To classify the LuALLs in an evolutionarily relevant manner, a phylogenetic tree was constructed using lectin domain amino acid sequences from LuALLs and from representatives of the twelve major plant lectin families (Van Damme *et al.*, 2008). It was noted that some LuALLs contained more than one lectin domain, in which case each lectin domain was aligned and analysed independently. Only eight lectin families were represented in the phylogenetic tree in Figure 3.3 because of the high divergence of the other four families. All of the LuALL domains grouped more closely with lectin domains from the amaranthin lectin family than any other family. At the highest level, the LuALL domains could be separated into two groups, which were labelled 1 and 2. Group 2 also included lectin domain 2 from *A. caudatas* amaranthin and from the *T. aestivum* Hfr-2 protein. Conversely, Group 1 included domain 1 of each of these two proteins. Within Group 1, all of the LuALLs that were detected in seed and embryo derived libraries (Table 3.2) formed a monophyletic subgroup (A.1). Furthermore, all of the LuALL chimerolectins that contained a Bet v I domain (Table 3.1) formed a monophyletic subgroup (B.1). Subgroup B.1 had as its sister a small group that included Lus10016109 (LuALL7), which is the LuALL found only in ESTs of leaf origin.

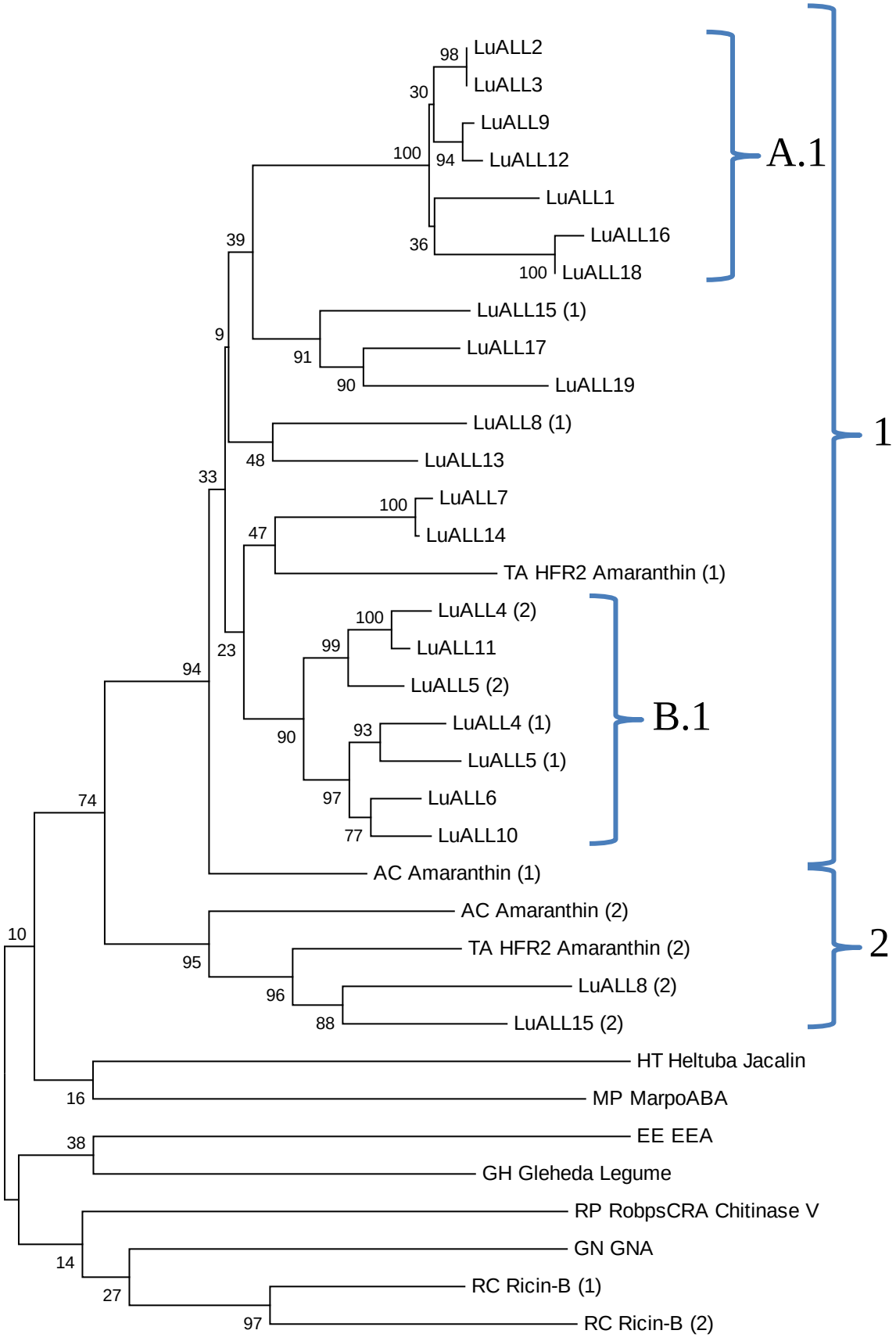
LuALL2
LuALL3
LuALL9
LuALL12
LuALL1
LuALL16
LuALL18
LuALL15 (1)
LuALL17
LuALL19
LuALL8 (1)
LuALL13
LuALL7
LuALL14
TA HFR2 Amaranthin (1)
LuALL4 (2)
LuALL11
LuALL5 (2)
LuALL4 (1)
LuALL5 (1)
LuALL6
LuALL10
AC Amaranthin (1)
AC Amaranthin (2)
TA HFR2 Amaranthin (2)
LuALL8 (2)
LuALL15 (2)
HT Heltuba Jacalin
MP MarpoABA
EE EEA
GH Gleheda Legume
RP RobpsCRA Chitinase V
GN GNA
RC Ricin-B (1)
RC Ricin-B (2)
A.1
B.1
1
2
0.2

Figure 3.3: Phylogenetic tree of flax agglutinin and lectin domains. The eight families are: Amaranthins, Jacalins, *Agaricus biscorpus* agglutinin (ABA), *Euonymus europaeus* agglutinin domain (EEA), Legume lectin domain, Class V chitinase, *Galanthus nivalis* agglutinin domain (GNA), Ricin-B domain. Lus: *L. usitatissimum*; AC: *A. caudutus*; TA: T. *aestivum* EE: E. *europaeus*; HT: *Helianthus tuberosus*; MP: *Marchantia polymorpha*; GH: *Glechoma hederacea*; RP: *Robinia pseudoacacia*; GN: *G. nivalis*; Rc: *R. communis*.

3.2 Spatial transcript expression patterns assayed by qRT-PCR

To further investigate the transcript expression patterns of the predicted LuALL genes, their expression was analyzed by qRT-PCR in six different tissues: cotyledons, roots, leaves, stems, floral buds and green seeds (i.e., mature embryo stage) (Figure 3.4). Out of 19 LuALLs, 16 genes (84 %) were expressed in at least one of the six tissues assayed (Figure 3.4). Of the three genes for which no expression was detected by qRT-PCR, two (LuALL14, LuALL17) were also not represented in ESTs databases (Table 3.2), while LuALL1 was represented by a small number of ESTs.

Several patterns were evident in the transcript heat map (Figure 3.4). In floral tissues and in all vegative tissues except leaves, LuALL3 transcripts were more abundant than the transcripts of any other LuALL gene (Figure 3.4). The maximum expression of LuALL3 occurred in roots. In leaves only, the expression of LuALL9 was higher than LuALL3. Like LuALL3, LuALL2 was also strongly expressed in all tissues, with its maximum abundance detected in roots. LuALL7, LuALL9, LuALL10 and LuALL12 transcripts were relatively abundant in all vegetative and floral tissues assayed, with maximum expression in leaves and cotyledons.

A few genes showed high tissue-specificity in their expression patterns (Figure 3.4): LuALL5 and LuALL15 were detected exclusively in cotyledons and stems, respectively, although their expression in each case was relatively low compared to the maximum of other LuALLs in these and other tissues. LuALL8 and 13 were strongly enriched in floral tissues compared to other tissues, although they were detected in vegetative tissues.

LuALL16 and LuALL18 were expressed almost exclusively in developing seeds, and their expression in these tissues was orders of magnitude stronger than any other LuALLs detected in seeds. LuALL19 had a high expression in roots compared to its expression in any other vegetative and floral tissues assayed.

Amongst the tested tissues, cotyledons had the highest number of LuALLs expressed, with 14/19 genes detected, followed by stems and floral buds that had 13/19 genes and roots and leaves with 12/19 genes detected (Figure 3.4). Conversely, green seeds (mature embryo stage) expressed the fewest number of LuALLs, with only 7/19 genes detectable.

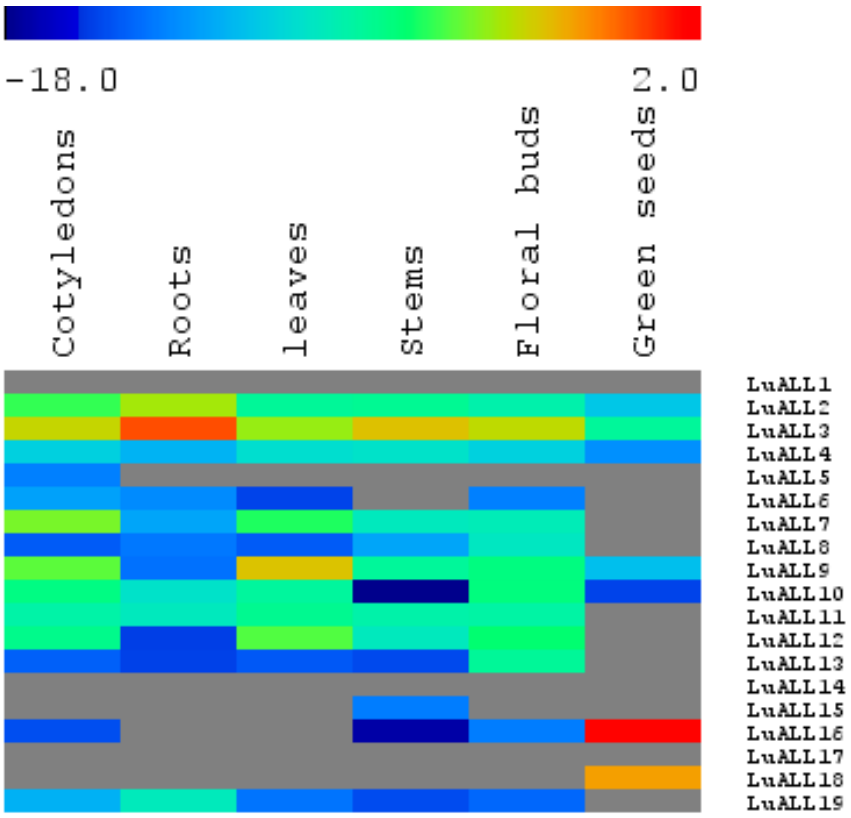

Figure 3.4: Transcript abundance of flax LuALL genes in six different tissues, as measured by qRT-PCR. Expression levels are relative to the most stable internal control

gene, GADPH (Huis, Hawkins & Neutelings, 2010). The heat map was generated using the delta Ct values calculated by subtracting the Ct value of the internal control from the Ct value of the genes and multiplying by -1. Because of the multiplication, high delta Cts represent high expression while low delta Cts represent low expression (Goni, García, & Foissac, 2009). In the map, blue indicates low expression and red indicates high expression. Gray indicates no detectable expression.

3.3 Hormone inducible gene expression

Because some lectins are known to be involved in defense, the inducibility of LuALLs at 12 h and 72 h following treatment with either of the defense-related hormones methyl jasmonate (MeJA) or salicylic acid (SA) was determined. Transcript abundance was measured by qRT-PCR and compared to mock-treated controls. Genes found to be statistically significant in ANOVA were then evaluated using a t-test to determine the time points (12 hours and 72 hours) that were statistically significant.

3.3.1 Transcript expression in response to MeJA

Transcripts of six LuALL genes (LuALL2, 3, 6, 9, 12 and 15) significantly (ANOVA and t-test) increased in abundance following MeJA treatment (Figure 3.5). None of the genes significantly decreased in transcript abundance following treatment. For all six genes, transcript expression increased until at least 72 h following treatment. LuALL2, 3, 9 and 12 were relatively abundant even in mock-treated samples (data not shown). In contrast, in the absence of hormone, LuALL6 and LuALL15 were not abundant in mock-treated seedlings or in any other tissues assayed (Figure 3.4). LuALL6 increased only 3-fold following MeJA, treatment, whereas LuALL15 increased 3,103-fold after 72h and showed the highest fold-induction of any LuALL upon MeJA treatment (Figure 3.5).

3.3.2 Transcript expression in response to SA

Six LuALL genes significantly (ANOVA and t-test) increased in transcript abundance following SA treatment. These six genes were: LuALL4, 6, 10, 11, 17 and 19. As in MeJA, most of the salicylic acid responsive genes (LuALL4, 10, 11 and 19) were also highly expressed in the mock-treated samples (data not shown).

The gene LuALL17 was not expressed in the absence of exogenous hormone in either seedlings or in the six mature tissues tested (Figure 3.4). However, LuALL17 was highly responsive to SA treatment, with a 15.9-fold increase in abundance at 72 h compared to the mock (Figure 3.6). As was observed with MeJA treatment (Figure 3.5), the SA-responsive transcripts increased in abundance until at least 72 h following treatment, although the increase in transcript abundance of LuALL6 between 12 h (6.25-fold) and 72 h (7.5 –fold) was found to be statistically insignificant. None of the genes significantly decreased in transcript abundance following SA treatment (ANOVA).

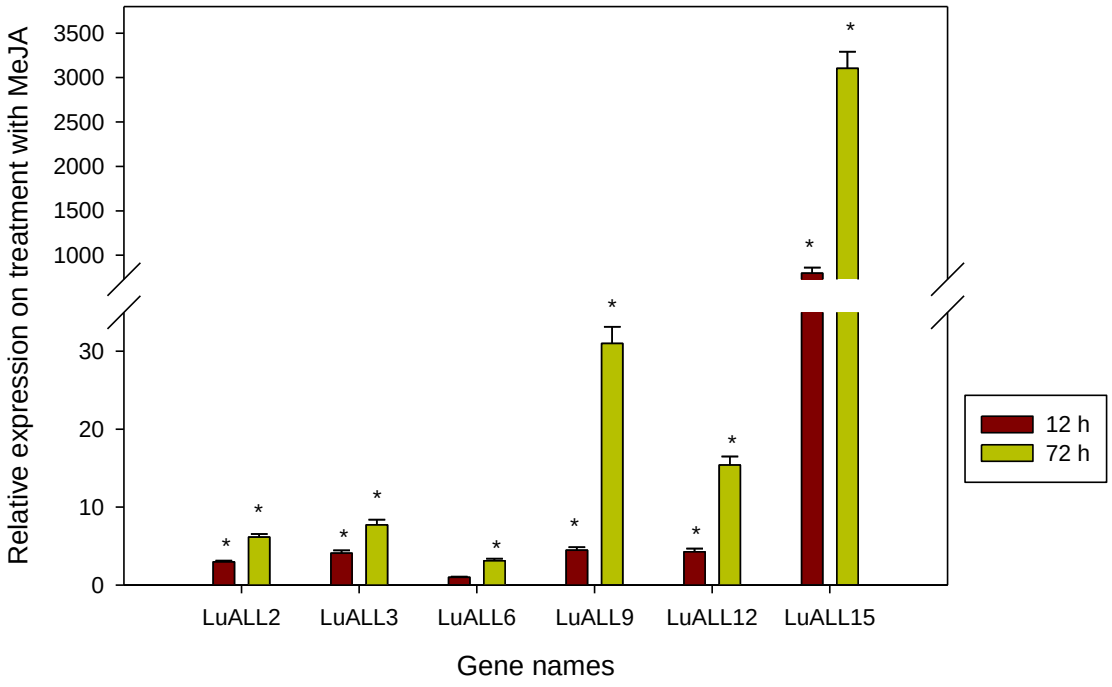

Figure 3.5: Relative expression of the five genes in etiolated seedlings up regulated significantly by methyl jasmonate (MeJA) treatment. The expressions are relative to the

mock-treated sample. Stars indicate statistically significant by both ANOVA and t-test. The standard deviations are denoted by error bars. n=3 replicates.

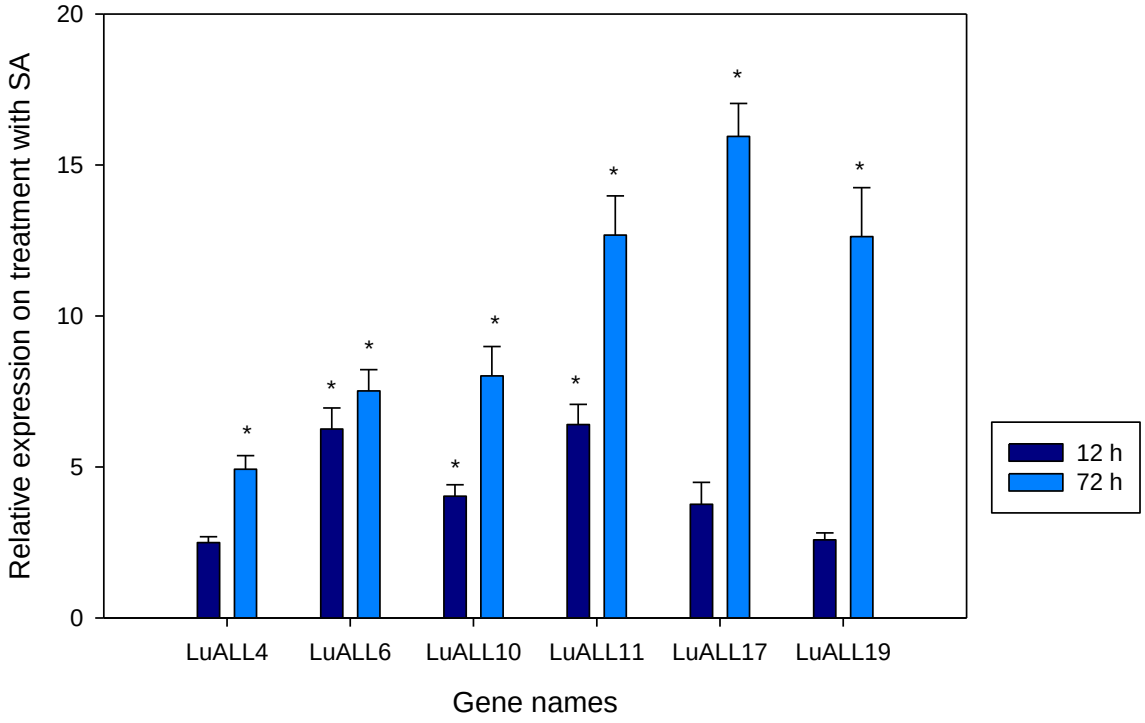

Figure 3.6: Relative expression of the six genes in etiolated seedlings up regulated significantly by salicylic acid (SA) treatment. The expressions are relative to the mock-treated sample. Stars indicate statistically significant by both ANOVA and t-test. The standard deviations are denoted by error bars. n=3 replicates except LuALL17 (72h) for which n=2.

3.4 Expression and functional characterization of a representative LuALL

To test whether any member of the LuALL family had carbohydrate-binding ability, one LuALL was selected for heterologous expression in *E. coli*. LuALL12 (Lus10024084) was chosen for this purpose because it is representative of the majority of LuALLs, in that it has a single agglutinin domain, is broadly expressed in vegetative tissues and flowers but is not abundant in seeds (Figure 3.4). There was also independent evidence of LuALL12 expression in EST databases (Table 3.2). Furthermore, LuALL12 was interesting because of its inducibility by MeJa in flax seedlings (Figure 3.5), suggesting that it might have a role in defense.

3.4.1 Codon optimization of LuALL12 for expression in *E.coli*

E.coli is a frequently used host for heterologous gene expression. However, differences in codon preferences can quantitatively limit expression of plant genes in *E. coli* (Chen & Texada, 2006). Rarely used codons of *E. coli* have been classified into two groups. Group I includes the seven codons (AGG, AGA, CGA, CUA, AUA, CCC and CGG) used at a frequency of < 0.5% and group II includes 13 codons (ACA, CCU, UCA, GGA, AGU, UCG, CCA, UCC, GGG, CUC, CUU, UCU and UUA) used at a frequency of > 0.5% but < 1.1%. The use of any codon from group I and the first six codons listed in group II can cause translational problems in *E. coli* (Chen *et al.,* 2006).

Analysis of the native LuALL12 coding sequence using codonW version 1.4.4 showed usage of rare codons from both group I and group II (Table 3.7). Thus, these rare codons were replaced as the coding sequence of LuALL12 was optimized for expression in *E. coli* (Figure 3.7). In the synthetic, optimized gene, a further change was made: threonine

(ACG) at position 2 was substituted by glycine (GGT) to allow for incorporation of an NcoI site (CCATGG) for sub-cloning.

Table 3.7: The number of each codon in LuALL12 according to codonW. The rare codons according to Chen *et al.*, 2006 are shown in red. The blue are the rest seven rare codons belonging to group II.

Phe UUU 0 UUC 9 Leu UUA 2 UUG 1	Ser UCU 1 UCC 9 UCA 1 UCG 1	Tyr UAU 1 UAC 9 TER UAA 0 UAG 0	Cys UGU 0 UGC 2 TER UGA 1 Trp UGG 3
CUU 3 CUC 7 CUA 1 CUG 3	Pro CCU 1 CCC 6 CCA 2 CCG 0	His CAU 0 CAC 4 Gln CAA 0 CAG 6	Arg CGU 0 CGC 4 CGA 0 CGG 0
Ile AUU 2 AUC 9 AUA 1 Met AUG 2	Thr ACU 0 ACC 7 ACA 2 ACG 2	Asn AAU 1 AAC 11 Lys AAA 4 AAG 16	Ser AGU 2 AGC 4 Arg AGA 0 AGG 2
Val GUU 2 GUC 8 GUA 0 GUG 8	Ala GCU 8 GCC 9 GCA 4 GCG 2	Asp GAU 5 GAC 16 Glu GAA 3 GAG 14	Gly GGU 0 GGC 5 GGA 3

			GGG 3

A

```
  1 atggttccaccgttgcgggcctgccgaaatatgttgtgctgaaa
    M  G  S  T  V  A  G  L  P  K  Y  V  V  L  K
 46 tctaaatccactggcaagtatctgcattacctgtggaacgacgaa
    S  K  S  T  G  K  Y  L  H  Y  L  W  N  D  E
 91 ttcagcgactactacaaacacatgggttctaaacgcgacgtagat
    F  S  D  Y  Y  K  H  M  G  S  K  R  D  V  D
136 ccggtaaaccctttcgtgaaactggaagttgtaccaagcaccgcg
    P  V  N  P  F  V  K  L  E  V  V  P  S  T  A
181 gatccgacgctggtacacctgatctgtagctataacagcaaattc
    D  P  T  L  V  H  L  I  C  S  Y  N  S  K  F
226 atccaactggtttctaaaggcggtgtgagctggctgtccgcaacc
    I  Q  L  V  S  K  G  G  V  S  W  L  S  A  T
271 gcagactctccggatgaagatctgactaaggaaacctccaccctg
    A  D  S  P  D  E  D  L  T  K  E  T  S  T  L
316 tttcagccgatctttccggctggcgaaccgaatactgtggaatttt
    F  Q  P  I  F  P  A  G  E  P  N  T  V  E  F
361 ctgcacgttcagaccaaccgtaacgtgcgcatcttcatcaacaaa
    L  H  V  Q  T  N  R  N  V  R  I  F  I  N  K
406 gactatggtgactctattaataacgtcgcttgcgcatactccaaa
    D  Y  G  D  S  I  N  N  V  A  C  A  Y  S  K
451 gacggtggtggcgacatcaaccgtttcgagttcgcggcgtgggtt
    D  G  G  G  D  I  N  R  F  E  F  A  A  W  V
496 tcttacgaggatgtcatcaaggcgaaagacgatgagattgaaaaa
    S  Y  E  D  V  I  K  A  K  D  D  E  I  E  K
541 ctgaaagctcaggctgcagctggcgacgatgatgaatccctgtct
    L  K  A  Q  A  A  A  G  D  D  D  E  S  L  S
586 gccgatgcaattgttgaagaactgcgtaaggacatccgtgaacag
    A  D  A  I  V  E  E  L  R  K  D  I  R  E  Q
631 aacgaacagctggaagcagcctacaacgaaattgatgccctgaaa
    N  E  Q  L  E  A  A  Y  N  E  I  D  A  L  K
676 gcggcggcgaaagctaaaagcgctctcgagcaccaccaccaccac
    A  A  A  K  A  K  S  A  L  E  H  H  H  H  H
721 cactga 726
    H  *
```

B

```
  1 atgacgtctacagtggccggattacccaagtacgtggtgctcaag
    M  T  S  T  V  A  G  L  P  K  Y  V  V  L  K
 46 tccaagtccaccgggaagtaccttcactacctatggaacgacgag
    S  K  S  T  G  K  Y  L  H  Y  L  W  N  D  E
 91 ttctccgactactacaagcacatgggatccaagcgcgacgtcgac
    F  S  D  Y  Y  K  H  M  G  S  K  R  D  V  D
136 ccagtcaaccctttcgtgaagctggaggttgtcccctccacagcc
    P  V  N  P  F  V  K  L  E  V  V  P  S  T  A
181 gacccaaccctcgtccacctcatctgctcctacaacagcaagttc
    D  P  T  L  V  H  L  I  C  S  Y  N  S  K  F
226 atccagctcgtcagcaagggcggcgtctcatggctctccgccacc
    I  Q  L  V  S  K  G  G  V  S  W  L  S  A  T
271 gccgattcccccgacgaggaccttaccaaggaaacctccaccctc
    A  D  S  P  D  E  D  L  T  K  E  T  S  T  L
316 ttccagcccatcttccccgccggcgagcccaacaccgtcgaattc
    F  Q  P  I  F  P  A  G  E  P  N  T  V  E  F
361 ctgcacgtccagacgaaccgcaacgtgcgcatcttcatcaacaag
    L  H  V  Q  T  N  R  N  V  R  I  F  I  N  K
406 gattacggcgacagcatcaacaacgtggcttgcgcctacagcaaa
    D  Y  G  D  S  I  N  N  V  A  C  A  Y  S  K
451 gacggaggggcgacatcaaccgcttcgagttcgcagcgtgggtt
    D  G  G  G  D  I  N  R  F  E  F  A  A  W  V
496 tcgtacgaggacgtgatcaaggctaaggacgacgagatcgagaag
    S  Y  E  D  V  I  K  A  K  D  D  E  I  E  K
541 ctcaaagcacaggcagctgctggggacgacgatgagagtctgagt
    L  K  A  Q  A  A  A  G  D  D  D  E  S  L  S
586 gctgatgccattgtggaagagcttaggaaggacataagggagcag
    A  D  A  I  V  E  E  L  R  K  D  I  R  E  Q
631 aacgagcagttggaggcggcttataatgagattgatgctttaaaa
    N  E  Q  L  E  A  A  Y  N  E  I  D  A  L  K
676 gctgccgcaaaagccaagtga 696
    A  A  A  K  A  K  *
```

Figure 3.7: The optimized (A) and the original codons (B) of LuALL12 with the coded amino acid generated using ORF Finder (Open reading Frame finder) (http://www.ncbi.nlm.nih.gov/projects/gorf/)

3.4.2 Expression of the recombinant protein in *E.coli*

The expression of codon-optimized, 6x-His tagged recombinant LuALL12 was induced in *E. coli* by IPTG at 37 ˚C and the soluble periplasmic (SP) and inclusion body (IB) fractions from both the recombinant and control (i.e., not induced) strains were analyzed by SDS-PAGE (Figure 3.8). A thick band was observed at the expected 27 kDa position

in the IB fraction but not in the SP fraction. Subsequent western-blotting with an anti-His antibody confirmed that the 27 kDa band contained the recombinant protein (Figure 3.9). Thus, under these conditions, LuALL12 was insoluble and accumulated in inclusion bodies.

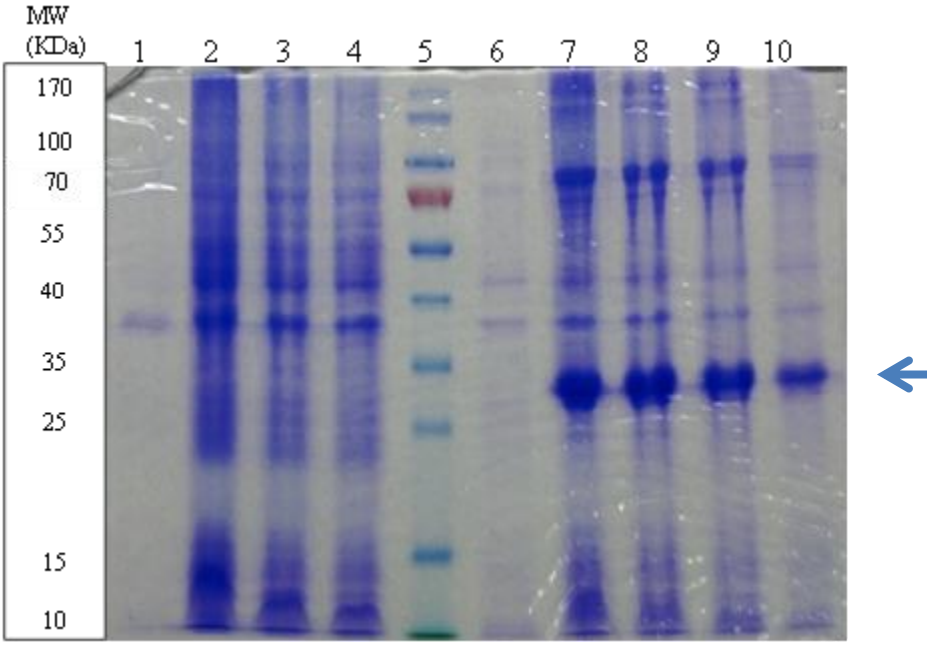

Figure 3.8: Extracted proteins from control and induced *E. coli* cultured at 37˚C. The molecular weight (MW) of the bands in the protein ladder is shown on the left hand side. Lane 1: Soluble protein (SP) of control; Lane 2-4: Inclusion bodies (IB) of control; Lane 5: Page ruler prestained protein ladder (Thermo-scientific); Lane 6: SP of induced bacteria; Lane 7-10 IB of induced bacteria. The arrow indicates the expressed protein in IB.

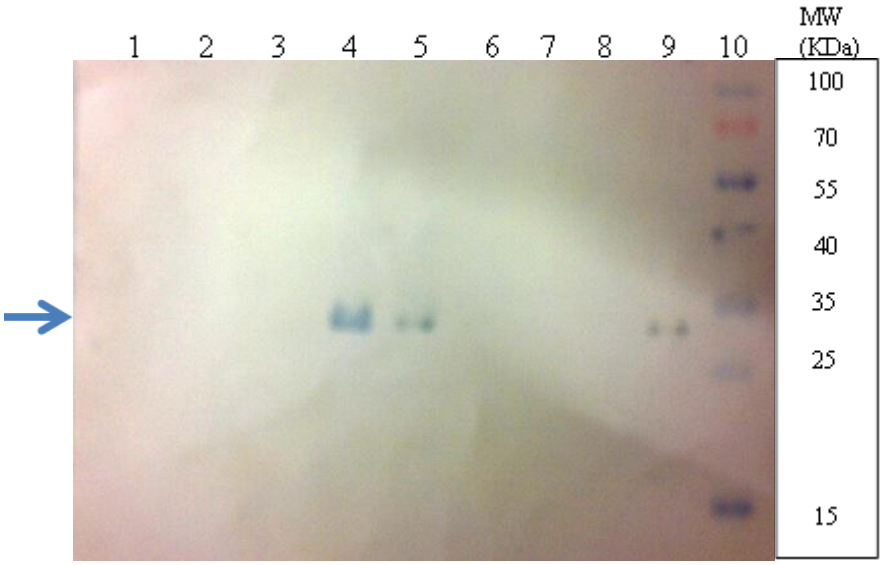

Figure 3.9: Western blot of extracted proteins from control and induced *E.coli* cultured at 37˚C. The molecular weight (MW) of the bands in the protein ladder is shown on the left hand side. Lane 1: SP of control; Lane 2-3: IB of control; Lane 4, 5, 9: IB of induced bacteria; Lane 7, 8: SP of induced bacteria; Lane 10: Page ruler prestained protein ladder (Thermo-scientific). The arrow indicates the position of the positive protein bands obtained by western blot.

3.4.3 Expression of the recombinant LuALL12 protein in the periplasmic space of *E.coli* and its purification

Overexpression of a protein in *E. coli* can result in the accumulation of the recombinant protein in inclusion bodies (Pédelacq *et al*., 2002). Proteins that accumulate in inclusion bodies are typically insoluble and inactive (Rudolph & Lilie, 1996). Finding the proper conditions to refold insoluble proteins can be very challenging (Vinogradov, Kudryashova, Levashov & van Dongen, 2003). As an alternative to refolding, attempts were made to improve protein solubility by optimizing culture conditions. It was found that lower growth temperature (18˚C) could increase the proportion of soluble protein

obtained (Figure 3.10), although it was necessary to increase the incubation time to 20 h to compensate for the reduced growth rate. No protein bands near 27 kDa were detectable in the soluble fraction after 2 h, 4 h, 6 h or 20 h at either 37˚C or 23˚C (Figure 3.10 and data not shown).

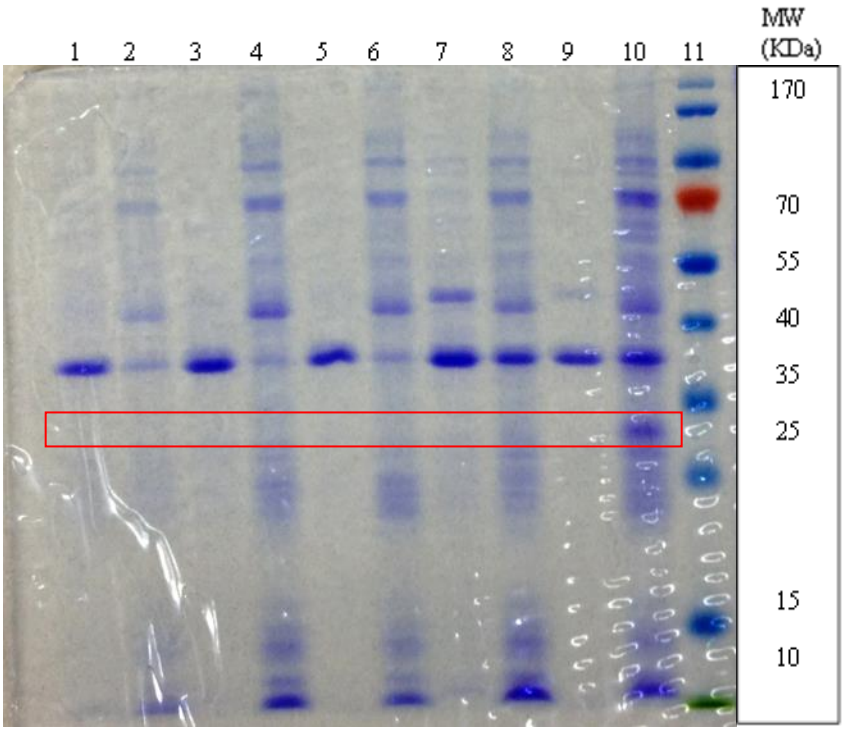

Figure 3.10: SDS-PAGE of extracted soluble proteins from control and induced samples at different temperatures for different time periods. Lanes 1, 3, 5, 7, 9: uninduced controls; lanes 2, 4, 6, 8, 10: induced control. Cultures were grown at: 37˚C for 2 h (lanes 1, 2), 4 h (lanes 3, 4), 6 h (lanes 5, 6); or 23˚C for 6 h (lanes 7, 8); or 18˚C for 20 h (lanes 9, 10). The Page ruler prestained protein ladder (Thermo-scientific) is shown in lane 11.

Figure 3.11 shows fractions obtained during purification of soluble protein from a 1 L culture of recombinant *E. coli* grown at 18˚C. Two strong bands were observed very

close together at approximately the position expected for the 27 kDa recombinant LuALL12. Peptide mass fingerprinting of both bands (excised from the gel separately) confirmed the presence of LuALL12 (Table 3.8). Thus, the protein could be successfully expressed and purified from the soluble periplasmic fraction at this temperature. The two bands obtained at 10 kDa and 75 kDa were also found by mass fingerprinting to contain LuALL12, and the 75 kDa band also contained other *E.coli* proteins (Table 3.8).

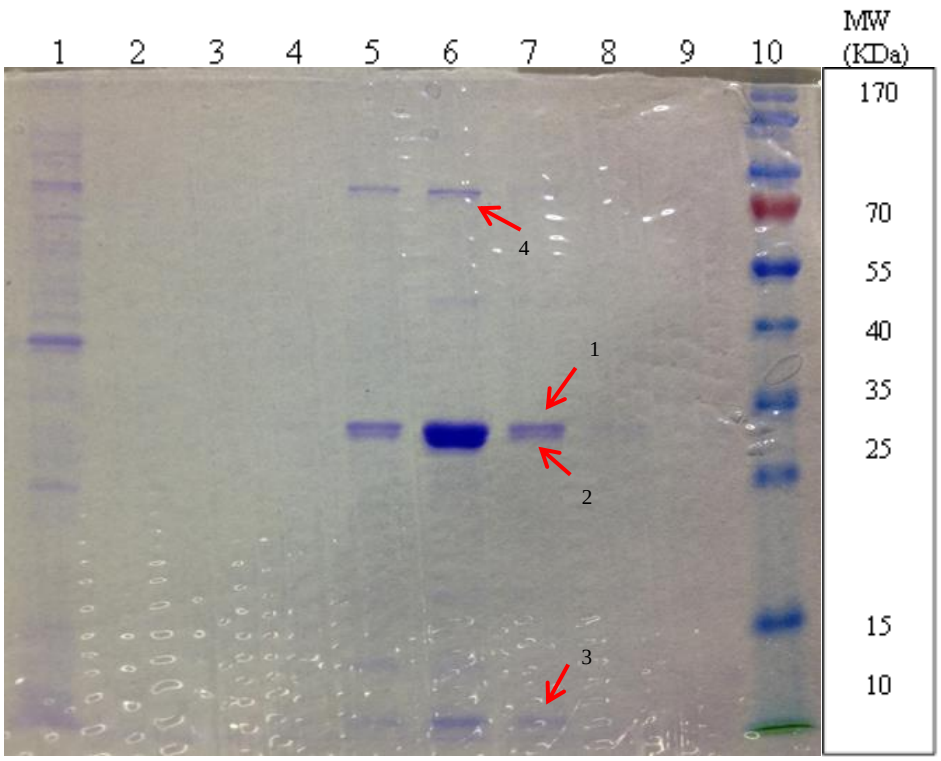

Figure 3.11: SDS-PAGE of the purified fractions of recombinant protein, LuALL12 from *E.coli*. Lane 1: protein flow-through; Lane 2: wash buffer I; Lane 3, 4: Wash buffer II; Lane 5: Elute I; Lane 6-9: Elute II (1^{st}-4^{th} flow); Lane 10: Page ruler prestained protein ladder. The arrows 1 and 2 indicate the purified protein.

Table 3.8: Bands from figure 3.11 identified by peptide mass fingerprinting.

Band	Accession	Description	Score	Coverage	Protein	Unique peptides	Peptides	MW
1	Lus10024084	flax protein	2216.42	84.85	1	23	23	25.7
2	Lus10024084	flax protein	2082.55	84.85	1	23	23	25.7
3	Lus10024084	flax protein	52.35	18.18	1	3	3	25.7
4	Lus10024084	flax protein	17.48	25.11	1	3	3	25.7
4	A7ZHA4	chaperone protein DnaK (*E.coli*)	3171.38	58.46	1	34	34	69.1
	B7N5M0	bifunctional polymyxin resistance protein ArnA (*E.coli*)	145.46	28.33	3	15	15	74.1
	B7LQ71	threonine--tRNA ligase (*E. fergusonii*)	39.16	10.12	3	7	7	74.0

3.4.4 Determination of the secondary structure of the recombinant LuALL12

The secondary structure of recombinant LuALL12 isolated from the soluble periplasmic fraction was determined by CD-spectroscopy in the far-UV spectral region (260-185 nm) and the data were analysed using CAPITO (CD Anaylsis and Plotting Tool) (Wiedemann, Bellstedt & Görlach, 2013). The analysis of the secondary structure content revealed that the recombinant LuALL12 contained both α-helix and β-sheet structures as demonstrated by the negative ellipticity at 222 and 208 nm and a positive ellipticity at 193 nm (Holzwarth & Doty, 1965); and negative ellipticity at 218 nm and positive ellipticity at 195 nm (Greenfield & Fasman, 1969) respectively (Figure 3.12). Based on the CD results, the recombinant protein was estimated to consist of 24% α-helix and 29% β-sheet. These observed values were similar to the values predicted using YASPIN (Table 3.6).

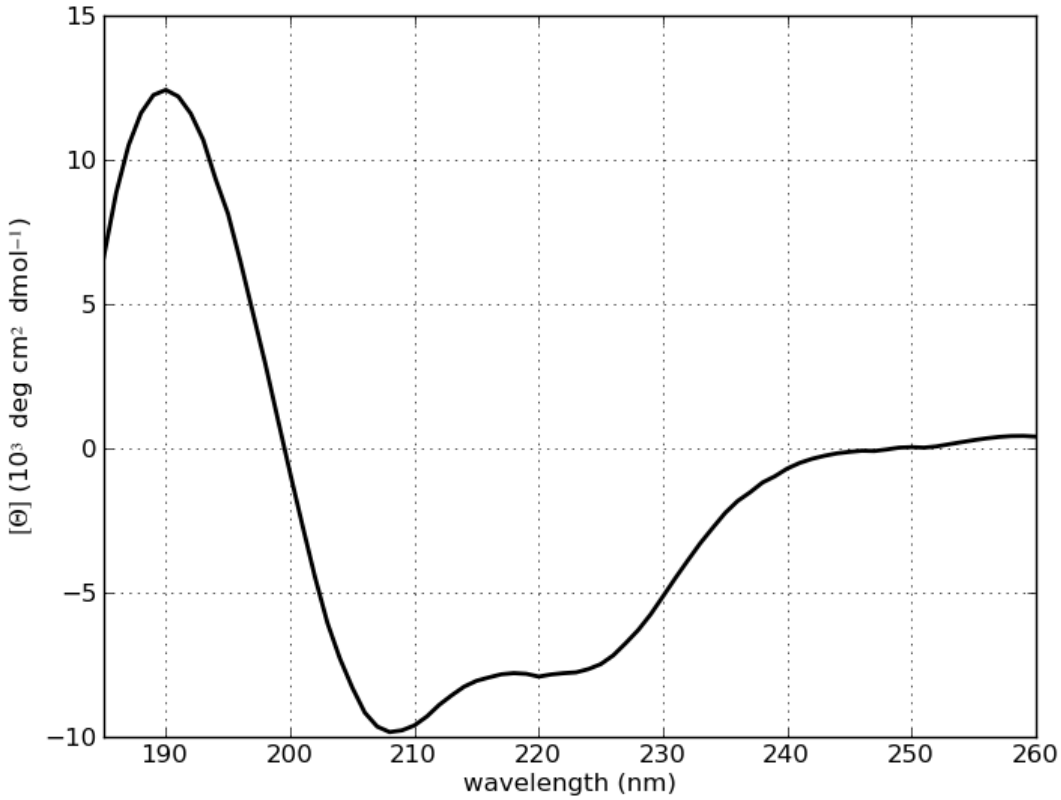

Figure 3.12: Circular dichroism of LuALL12 in the far-UV spectral region (260-185 nm). The wavelength and the mean residue ellipticity are shown on the X and Y-axis respectively.

3.4.5 Assay of hemagglutinating activity

To test for functional activity of the purified, recombinant LuALL12, a hemagglutination assay was performed using a 2 % (v/v) suspension of rabbit erythrocytes. The assay was performed in conical-bottom microtiter plates so that in the absence of agglutination, erythtocytes settle in a distinct spot at the very bottom of the conical well. On the other hand, agglutinated erythrocytes form a diffuse mat that covers the entire well. As shown in Figure 3.13, recombinant LuALL12 was able to agglutinate the erythrocytes, similar to the positive control (GNA). This is the first empirical evidence that any LuALL can agglutinate rabbit erythrocytes and is therefore a functional lectin. Based on the serial

dilution of the recombinant protein, a minimum of 0.5 mg/mL of LuALL12 was required for agglutination, and this is therefore defined as 1 hemagglutination unit of recombinant LuALL12.

The first well (Figure 3.13 A) had a concentration of 200 µg added to 100 µl of 2 % suspension of RBCs in a total 200 µl of solution. In the next two-fold dilution, 100 µg of protein could still produce visible agglutination of RBCs.

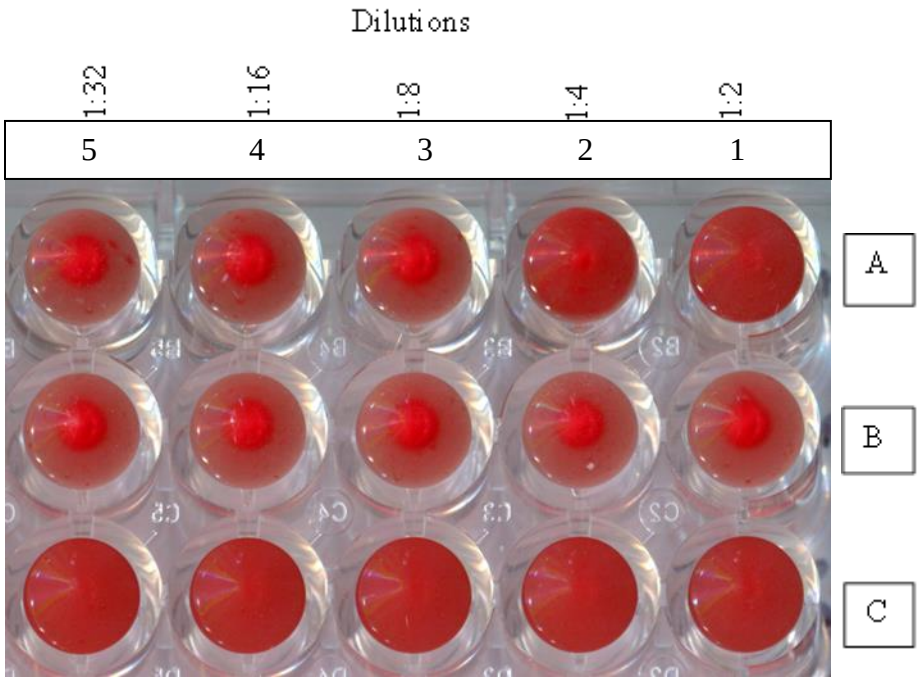

Figure 3.13: Hemagglutination assay with recombinant LuALL12 on rabbit erythrocytes. Serial two fold dilutions were used for the A) Protein sample (LuALL12); B) Negative (buffer) control; C) Positive control (GNA).

3.4.6 Hemagglutination inhibition assay

In previous studies, hemagglutination activity of amaranthin from *A. caudutus* and *A. leucocarpus* could be inhibited by N-acetylgalactosamine (GalNAc) (Rinderle *et al.*, 1989; Hernández, Debray, Jaekel, Garfias, del Carmen Jiménez, Martínez-Cairo & Zenteno, 2001). Two amino sugars, N-acetylgalactosamine (GalNAc) and N-acetylglucosamine (GlcNAc) were thus tested for their ability to inhibit LuALL12-induced hemagglutination of rabbit erythrocytes. At 200 mM, GalNAc showed only very partial inhibition of agglutination and GlcNAc apparently appeared to cause more inhibition than N-acetylgalactosamine (Figure 3.14). Thus the two amino sugars might be part of the potential ligands of LuALL12.

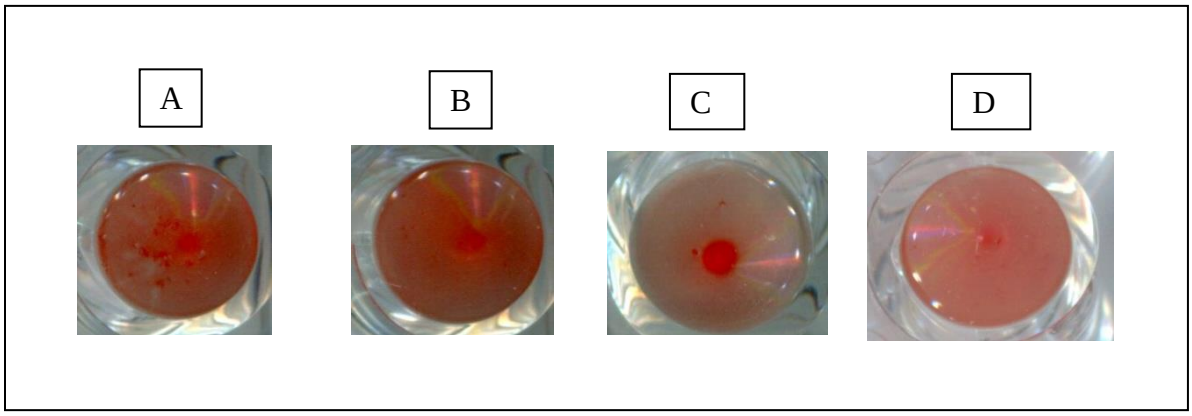

Figure 3.14: Hemagglutination inhibition assay of LuALL12 on rabbit erythrocytes for two amino sugars, GlcNAc and GalNAc. A) GlcNAc; B) GalNAc; C) Negative (buffer) control; D) Positive control (LuALL12 with erythrocytes only).

3.4.7 Native-PAGE of purified recombinant LuALL12

Native-PAGE was run with purified recombinant protein that had not been subjected to denaturation (Figure 3.15). The native-PAGE showed two protein bands very close to

each other, similar to the pattern observed on SDS-PAGE. This shows that the extracted protein was pure, but contained two highly similar isoforms, which had sufficiently similar charge, structure, and mass so that they migrated together on the native gel.

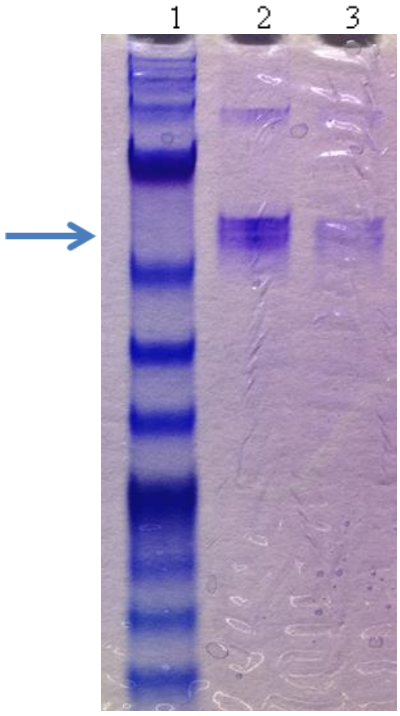

Figure 3.15: Native-PAGE gel of purified recombinant protein. Lane 1: Page ruler prestained protein ladder (Thermo-scientific) (marker weights as figure 3.11); Lane 2: 3 µg of protein in native state; Lane 3: 1µg of protein in native state.

Chapter 4 Discussion

The apparently lineage-specific amplification of agglutinin-domain containing genes in flax has resulted in this species containing more of these genes than any other species sampled thus far except apple (*Malus domesticus*) (Figure 3.1). These 19 genes were named LuALLs (*Linum usitatissimum* amaranthin lectin like), because the agglutinin domain is characteristic of amaranthin-type lectins. The unusual size of this gene family in flax compared to other species led us to characterize the structure and function of LuALLs. This will lead to a better understanding of the biology of both flax and lectins in general.

4.1 General features of the LuALL family

From the *in silico* analyses it was predicted that the LuALL family of proteins are soluble cytosolic proteins (Table 3.3 & 3.4). No member of the amaranthin family has been localized at the sub-cellular level yet (Van Damme *et al.*, 2008). But sequence analysis and preliminary experiments have suggested that the amaranthin family is devoid of signal peptides and located in the nucleo-cytoplasmic compartment (Lannoo *et al.*, 2010; Van Damme *et al.*, 2011). Amarathin from *A. caudutus* was non-glycosylated(Rinderle *et al.*, 1989). Based on analysis in GlycoEP, the LuALLs showed potential sites for N and O glycosylation (Table 3.5). N-glycosylation is rare in non-secretory proteins. Since the LuALL proteins were predicted to be non-secretory, it is likely that any glycosylation that occurs on them is O-glycosylation.

Among the 19 LuALLs proteins were eleven merolectins (i.e., one lectin domain), two hololectins (i.e., two lectin domains) and six genes with one of two types of

chimerolectins (i.e., one or more lectin domains with another protein domain) (Table 3.1). The description of the agglutinin domain hololectins in flax was novel, as these were previously known only in the family Amaranthaceae (Van Damme *et al.,* 2008). The occurrence of single domain amaranthin and Hfr-2 type chimeric protein in the rosaceae family has been reported previously (Van Damme *et al.,* 2008), but the simultaneous occurrence of both of these in one species has been reported for the first time in flax. The chimerolectins of flax with Bet v I were also novel, as described in the following section.

4.2 Relationship of ALLs to plant defense and biotic stress

4.2.1 Chimerolectins contain defense-related protein domains

Besides the agglutinin domain, the LuALL chimerolectins contained one of two conserved domains: Bet v I or aerolysin (Table 3.1). Bet v I belongs to the pathogenesis related, PR-10 family known to be expressed in the cytoplasm (Radauer, Lackner & Breiteneder, 2008) and detected in defense responses to pathogens and in other physiological processes (Marković-Housley *et al.,* 2003). The other fusion partner in the chimerolectin, aerolysin toxin domain, also has a potential role in defense. Aerolysin is produced by *Aeromonas* species, but related proteins are present in bacteria, plants and eukaryotes. The protein can bind to eukaryotic cells and polymerize to form pores leading to the cell lysis (Perker *et al.,* 1996). In plants, chimeras with this domain fused to amaranthin domain (Hfr-2), have been reported to be induced in response to insects, wounding and phytohormone treatment (Puthoff *et al.,* 2005). In flax, the aerolysin containing LuALL13 however was not up regulated by hormone treatment (Figure 3.5 & 3.6).

4.2.2 Defense hormones

Methyl jasmonate (MeJA) and its free acid jasmonic acid (JA) are well known to activate plant defense mechanisms in response to insect driven wounding, various pathogens and environmental stresses, such as drought, low temperature and salinity (Wasternack & Parthier, 1997). On the other hand, salicylic acid (SA) is induced after infection by many microbial pathogens or abiotic stresses such as salt, drought, heat, cold, ozone, UV exposure and metal stress (Fragnière, Serrano, Abou-Mansour, Métraux & L'Haridon, 2011; Yuan & Lin, 2008). The bare majority (11/19) of LuALLs were induced by either one of these defense hormones (Figure 3.5 & 3.6), indicative of a clear role for many LuALLs in defense. Interestingly, there was little overlap between the genes that were induced by either MeJA and SA, with only one gene (LuALL6) induced by both hormones (Figure 4.1). LuALL 6 is a Bet v 1 chimerolectin belonging to the PR-10 family. PR-10 genes like RPR10a, JIOsPR10 in rice have been previously shown to be up regulated both by MeJA and SA (McGee, Hamer & Hodges, 2001; Jwa, Agrawal, Rakwal, Park & Agrawal, 2001).

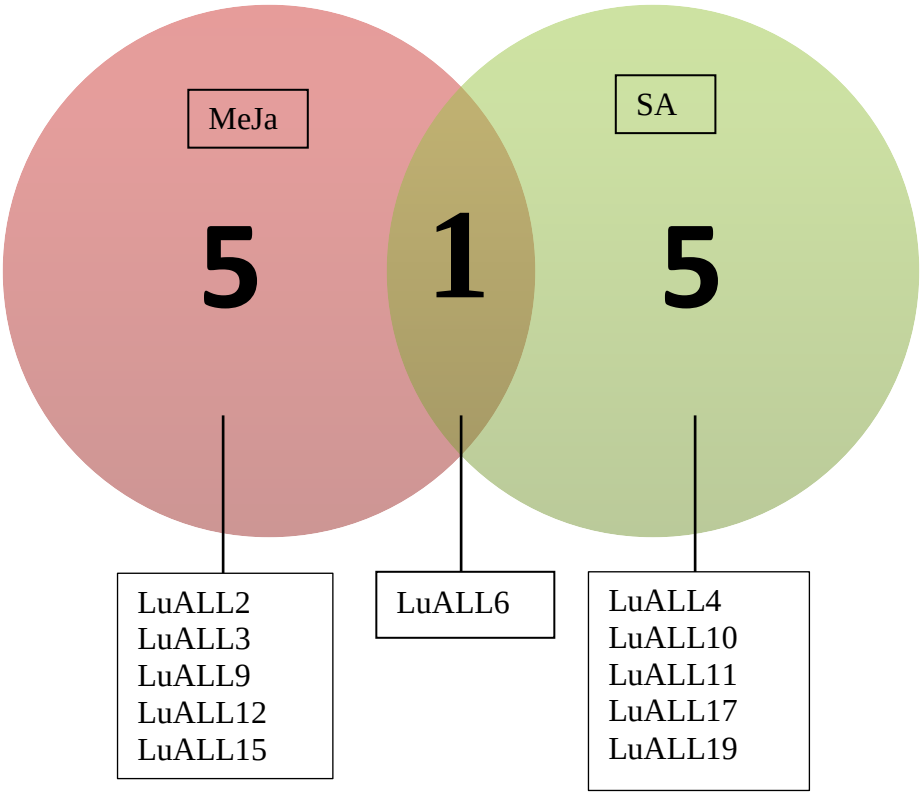

Figure 4.1: Venn diagram showing the classification of genes inducible by methyl jasmonate (MeJA) and salicylic acid (SA) based on qRT-PCR analysis. The numbers in each group is the number of inducible gene in each group. A list of the genes of the three groups is identified in the boxes below.

4.3 Sub-groups with distinct properties can be defined within the LuALL family

The phylogenetic analysis sorted the LuALLs into distinct groups based on similarities in the amino acid sequences of their respective agglutinin domains. Based on our analyses, several of these phylogenetically defined groups could be associated with distinct properties and patterns of expression as outlined here:

4.3.1 Group A.1 and the seed lectins

The four genes LuALL2, 3, 9, and 12 formed a clade (Group A.1) that was a sister group to LuALL1, 16, and 18 (Figure 3.3). These are all predicted to be merolectins (i.e. having a single lectin domain) with cytosolic localization (Table 3.4) and molecular weight under 30kDa (Table 3.1). All of these except LuALL1 have roughly similar proportions of alpha-helices and beta-sheets in their predicted structures. Together, these seven genes aligned exclusively with all but one of the 992 public ESTs sourced from a range of tissues (Table 3.2), suggesting that these are generally the most abundant LuALL transcripts in the plant body. More precisely, LuALL1 and the four genes in clade (A.1) were found in EST libraries derived from developing embryos and seed coats, leaves and stems. In contrast, LuALL16 and 18 were found only in embryo and seed coat derived libraries, where they were very abundant (Table 3.1 and 3.2), representing more than half of all ESTs (492/992 and 250/992 ESTs, respectively). qRT-PCR results (Figure 3.4) were generally consistent with the inferences drawn from EST analysis: transcripts of LuALL2, 3, 9 and 12 were expressed in all vegetative and floral tissues sampled, and all of these except LuALL12 were also detected in developing seeds. Furthermore, qRT-PCR confirmed that LuALL16 and 18 were the most abundant LuALL transcripts in developing seeds, with LuALL18 expressed exclusively in this tissue. LuALL1 was not detected above the threshold of significance in the qRT-PCR analysis, and it was represented by only 7 ESTs, suggesting that LuALL1 is expressed, but at a very low level. Finally, the four genes in clade A.1 were all inducible by MeJA, but not by SA (Figure 3.5). Thus, LuALL2, 3, 9, and 12 can be described as having transcripts that are relatively abundant in most tissues, with increased expression inducible by

MeJA. On the other hand, LuALL16 and 18 are strongly or exclusively associated with late stages of embryo development as well as seed coats. It is noteworthy that LuALL16 and 18 transcripts were highly abundant in mature embryos, but not in early embryos or in seedlings, suggesting a specific function of these genes associated with late stages of seed development. Seed lectins have been known previously that act like storage protein and accumulate at high levels in the seed, although in the vacuole (Pusztai & Bardocz, 1995). Thus the seed LuALLs in flax are probably not conventional storage proteins since they are predicted to be non-secreted proteins.

4.3.2 Group B.1

LuALL4, 5, 6, 10 and 11 formed a second clade (Group B.1, Figure 3.3) that was phylogenetically and functionally distinct within the LuALL family. All of these genes were chimerolectins, and were the only LuALLs that contained a Bet v I domain. Because the phylogenetic alignments were made using only the agglutinin domain, it is not the presence of the Bet v I domain that causes them to be grouped together but rather some shared features related to their carbohydrate binding ability. All of these except LuALL5 were also highly inducible by SA (Figure 3.6). LuALL5 was otherwise distinct for having an acid pI (5.9) and being expressed exclusively in cotyledons, whereas the others (LuALL4, 6, 10, 11) were predicted to have basic pI's and were expressed in a wider range of tissues. Because the Bet v I domain is often found in proteins associated with biotic stress (Liu *et al.,* 2006), and SA is a hormone known to be induced by various types of biotic stress (Fragnière *et al.,* 2011), it is tempting to speculate that the LuALL lectins in clade B.1 function in defense responses, and that these are functionally different from the MeJA responsive lectins in Group A.1. However, it should be noted that both

SA and Bet v I domain-containing proteins also have functions outside of biotic stress responses.

4.3.3 Other LuALLs

The remaining six LuALL genes not discussed thus far (i.e., LuALL 7, 8, 13, 14, 15, 19) did not together comprise a monophyletic group and showed a variety of characteristics, some of which were unique. LuALL8 and LuALL13 were expressed in vegetative tissues and more so in floral tissues, but were not hormone inducible and shared little else with each other or with any other LuALL. LuALL13 was a chimerolectin that contained an aerolysin domain, and LuALL8 was a hololectin (i.e., contained two agglutinin domains). LuALL7 was expressed in all vegetative and floral tissues samples, but transcripts of it presumptive paralog LuALL14 were not detected in any tissue under any conditions, raising the possibility that it is a pseudogene or that it is expressed only under conditions not tested here. Notably, LuALL14 was also unusual in having a predicted length of 120 amino acids. LuALL17 and LuALL19 are likewise paralogs, and only LuALL19 was detected in untreated tissues. However, both LuALL17 and LuALL19 were inducible by SA. Finally, in the qRT-PCR of tissues in untreated plants, LuALL15 was detected only in stems (Figure 3.4). It was also detected at relatively low abundance in untreated seedlings. Upon MeJA treatment, LuALL15 transcripts increased 800 fold after 24 h and 3100 fold after 72 h (Figure 3.5). LuALL15 was not induced by SA, again suggesting a specific role for this gene in plant defense.

4.4 Biochemical analysis of a representative LuALL

4.4.1 Heterolgous expression of a functional LuALL

To test whether any of the LuALLs could function as lectins, one of the LuALLs containing a single agglutinin domain, LuALL12, was expressed in *E. coli* (Figures 3.11). The protein was purified and found to agglutinate rabbit red blood cells at a minimum concentration of 0.5 mg/ ml (Figure 3.13). It was found that glycosylation and metal ions were not required for the agglutination activity. Two amino sugars, N-acetyl galactosamine and N-acetyl glucosamine, were found to be potential ligands of the recombinant protein (Figure 3.14).

4.4.2 Quantitative agglutination activity

The minimum concentration of recombinant LuALL12 required to produce visible agglutination of rabbit red blood cells was 0.5 mg/ ml, which is relatively high compared to the 2 µg/ ml required by homodimeric amaranthin to cause agglutination of human red blood cells. It could be that LuALL12 is more specific towards red blood cells from other species and can cause agglutination at a much lower concentration. Such as, the plant lectin, TEL isolated from *Talisia esculenta* seeds was found to agglutinate red blood cells of different species at different concentrations. It was most effective against human (type A) erythrocytes requiring 0.09 mg/ ml for agglutination, whereas it required 1 mg/ ml to cause detectable agglutination of cow erythrocytes (Freire *et al.*, 2002). Another example is the purified *Mycena* lectin (MPFA) from *Mycena pura* fresh fruit bodies. It was found to require a minimum of 2.5 mg/ ml and 9.7 µg/ ml for agglutinating cow and dog erythrocytes, respectively (Antonyuk, Yashchenko, Antonyuk & Ambarova, 2009).

4.4.3 Secondary structure

The secondary structure of the recombinant LuALL12 was determined by circular dichroism (CD) spectroscopy and was found to consist of 24% α-helix and 29% β-sheet (Figure 3.12). Thus, the structure was different from the secondary structure of amaranthin, which is known to be enriched in β-sheets (Hernández *et al.*, 2001). Nevertheless, the observed CD spectra matched the expectations based on secondary structure prediction (Table 3.6), suggesting that the secondary structure obtained in *E. coli* was correct.

4.4.4 Structural analysis

LuALL12 is a predicted merolectin, and according to previous structural analysis of proteins from the amaranthin-lectin family, two lectin domains are required to form the ligand binding site of the homodimeric protein (Rinderle *et al.*, 1990; Transue *et al.*, 1997). Thus, presumably LuALL12 needs to form an oligomer, or more precisely a dimer, for carbohydrate binding. Each protein in the doublet could be formed as a dimer that is responsible for the agglutination of the red blood cells. The native-PAGE also has the doublet (Figure 3.15), so probably the dimerization does not occur between proteins of the doublet. If the proteins in the doublet had formed a dimer, the native-PAGE would have given a single band.

Innative-PAGE of LuALL12 a 75 kDa band of protein was observed. It is possible that this was an oligomer of the LuALL12. However, it was three times the 25 KDa size. It was therefore unlikely that this high molecular weight band represented an oligomer of LuALL12.

4.5 Conclusions

This research has described for the first time the 19 members of the LuALL family, one of the largest agglutinin-domain containing protein families identified to date. It was demonstrated that at least 17 of these predicted genes were expressed at the transcript level, and three major functionally and evolutionarily related sub-groups of LuALLs have been defined: (i) the small (25kDa), moderately acidic merolectins of clade A.1, which are expressed abundantly and constitutively in many tissues and are further inducible by MeJA; (ii) the similarly small, moderately acidic, paralogous merolectins LuALL16 and LuALL18, which are highly enriched in mature-stage embryos, but are not induced by hormones and are rare or are not found in any other tissue, and (iii) the larger (46 – 110kDa), basic (pI 8.2-9.7) mero- and holo- chimerolectins of clade B.1 (excluding LuALL5), which all contain a Bet v I domain and are inducible by SA. The enrichment of LuALL16 and 18 specifically during late stage seed maturation suggests that at least one of their functions could be in seed storage proteins. On the other hand, it is likely that the hormone-inducible clade A.1 and clade B.1 LuALLs have distinct functions in plant defense, based on the specificity of their hormone inducibility and differences in their amino acid sequences. Outside of these clades, transcripts of three other LuALLs (15, 17, 19) were also inducible by either MeJA or SA, by as much as 3100-fold, and an additional gene (LuALL 13) contained a defense-related aerolysin domain. Together, these results suggest that the majority of LuALLs are involved in plant responses to biotic stress. Heterologous expression of a representative LuALL showed that it could function as a lectin by inducing blood cell agglutination, although the specific activity of the purified protein in this assay was comparatively low, due perhaps to incomplete

oligomerization or low affinity for the blood cells of the arbitrarily selected species (i.e., rabbit) tested here.

Chapter 5 Conclusion

5.1 Summary

Lectins in plants are an interesting field of study as they are known to participate in diverse functions, including protecting plants against various biotic and abiotic stresses (Hopkins *et al.,* 2001; Ripoll *et al.,* 2003; Wong *et al.,* 2010; Petnual *et al.,* 2010 Hegedus *et al.,* 2009; Zhang *et al.,* 2000; Moons *et al.,* 1997). The roles of most of these lectins in their native organisms are still not known with certainty (Peumans *et al.,* 1995). Amongst the many families of lectin, the lineage specific Amaranthin family (Figure 3.1 & Jiang *et al.,* 2010) has not been characterized much, although it was found in diverse species from lycophytes to seed plants. Recent genome analysis of flax identified a family of 19 genes that have the agglutinin domain (Wang *et al.,* 2012), which is the characteristic domain of the amaranthin lectin family. This gene family was completely absent or present scarcely in other members of Malpighiales (Figure 3.1). Thus, the project of characterizing the LuALLs was undertaken to determine their role in flax.

Analysis of the results suggested that the LuALLs had defense-related genes in specific pathways. Two clades in the phylogenetic tree, A.1 and B.1, separated the MeJA and SA inducible genes, respectively (Figure 3.3). The two clades also had other distinctive features. The A.1 genes with a single agglutinin domain were not only MeJA inducible but also were expressed abundantly under normal conditions (Figure 3.4). They were similar in structure and properties, being small (22-30 KDa), moderately acidic (pI 4.71-6.53) and having equal distribution (except LuALL1) of α-helix and β-sheets in their secondary structure (Tables 3.1 & 3.6). LuALL1 although not inducible in the etiolated

seedlings, had been found to be induced separately by both MeJA and SA in a separate hormone test (data not shown). This induction by both hormones could be because of its unique characteristics, having a small size and acidic pI like clade A.1, but a β-sheet enriched secondary structure like the LuALLs in clade B.1. LuALL 16 and 18, which formed a sister group in A.1 were not induced by MeJA, rather, were seed-specific lectins (Figure 3.4).

All genes in B.1 were Bet v 1-containing LuALLs that were upregulated by SA (except LuALL5) (Table 3.1 & Figure 3.3). Having either a single or double agglutinin domain, these LuALLs were relatively larger (46-120 KDa), basic (pI 8.2-9.7) and enriched in β-sheets (except LuALL11). Under normal conditions, the B.1 LuALLs were moderately to highly expressed (Figure 3.4). LuALL6, up regulated by both the hormone treatments, showed that the LuALL family also had genes responsive to multiple defense hormones (Figure 3.5 & 3.6).

Other than the genes in clades A.1 and B.1, there were three other hormone inducible genes (Figure 3.5 & 3.6), LuALL15, 17 and 19. These were moderately acidic, 54 kDa proteins, enriched in βsheets. Two other β-sheet lectins, LuALL8 and 13 were preferentially expressed in floral buds, with LuALL13 being a chimerolectin with an aerolysin domain known to be involved in defense response.

Because of the presence of single/double domain hololectin and chimerolectin in flax, it is tempting to speculate that the double domain Bet v 1 chimeras are mainly responsible

for the expansion of the ALLs in flax. During expansion, the hololectin chimeras must have lost the Bet v 1 domain or an agglutinin domain giving rise to the hololectin and merolectin chimera respectively. It seems likely that the parental Bet v 1 gene gave rise to the pure merolectins by losing both its Bet v 1 domain and one agglutinin domain. It could be that after the loss of an agglutinin domain, the merolectin chimera had a duplication event followed by an exchange event with an aerolysin domain containing gene giving rise to the aerolysin chimera.

Heterologous expression of one of the LuALLs allowed characterization of these genes as lectins that can agglutinate rabbit erythrocytes (Figure 3.13), providing for the first time proof of agglutination by a single amaranthin domain containing protein.

5.2 Weakness/limitations of the research work

The LuALLs varied widely in length ranging from 120-1024 amino acids. Hence, it could be that the gene prediction by the whole genome assembly had errors. Thus it will be necessary in the future to verify the gene predictions by full length cDNA clones and their sequencing.

There are pairs of genes (LuALL2 and 3; LuALL 9 and 12) that had very similar sequence and thus aligned with almost the same ESTs even under high stringency (Table 3.2 & 3.3). From the evidence obtained by the expression analysis in both the tissues under normal conditions and hormone treatment, LuALL2 and 3 had similar expression patterns. It is therefore possible that these are one gene instead of two.

LuALL 9 and 12, although similar in their expression patterns, also had some differences too. LuALL9 was found to be expressed in green seeds, whereas LuALL12 was completely absent in seeds. Also LuALL 9 and 12 were differentially up regulated by MeJA with about 31 and 15 fold respectively. Thus most probably they are distinct genes. Verification needs to be done by cDNA analysis.

For the expression analysis in the six tissues, it would have been better to test beforehand the stability of various internal control genes in all the tissues before choosing the most stable ones to work with.

To test the inducibility of the genes, hormones were used to mimic the stress conditions. This is indirect evidence of biotic stress responsiveness. For direct evidence, exposure of plants to pathogens or insects should have been used.

In heterologous expression, even after a lot of effort, absolutely pure protein could not be obtained. This might have had an impact on the hemagglutination assays as one of the matching *E.coli* proteins (Table 3.8), Dnak have mammalian homologues that are know lectins. The rest two proteins that matched with the *E.coli* database (Table 3.8) are not known to agglutinate red blood cells. One of these was bifunctional polymyxin protein that is involved in antibiotic resistance, lipid biosynthesis and metabolism andthreonine RNA ligase that is known to be involved in protein biosynthesis (Uniprot database). Also there could also have been very minute amounts of other *E.coli* contaminant proteins that did not appear in the SDS-PAGE gel.

For the hemagglutination assay, erythrocytes from only one species were tested. The hemagglutination inhibition assay could be tested with only one concentration of the saccharides due to limitations of protein quantity.

In the native-PAGE gel, a native molecular weight marker was not used. Often, it is difficult to find standard proteins that resemble the shape, partial specific volume and degree of hydration as the native protein under investigation (Amersham biosciences, Separation technique file no. 120).

5.3 Future directions

5.3.1 Confirmation of the predicted genes by cDNA

It is necessary to verify the gene predictions by full length cDNA cloning and sequencing. For cloning the full length cDNA, two approaches can be used 1) cDNA library construction and screening; 2) PCR based rapid amplification of cDNA ends (RACE) technology. Next by sequencing of the cDNA it can be confirmed whether the predicted gene is a genuine gene (Zhumabayeva, Chang, McKinley, Diatchenko & Siebert 2003).

5.3.2 Confirming the hemagglutination activity of recombinant LuALL12

Further replicates of the hemagglutination activity needs to be done to confirm the lectin activity. Also erythrocytes from different species should be used to determine the erythrocytes, the lectin is most effective against.

5.3.3 Hemagglutination inhibition assay

This assay needs to be done with different saccharides to determine the different ligands of the lectin. And the minimum concentration of the saccharides effective against the lectin should be found out as well.

5.3.4 Glycan array

A lot of protein is required for the hemagglutination inhibition assay. To find out the specific carbohydrate structures being recognized by the flax LuALLs, carbohydrate microarrays, a high-throughput technology could be used.

5.3.5 Molecular weight determination

The native molecular weight of the LuALLs should help determine if oligomer of the protein is forming or not. Size-exclusion chromatography (SEC), coupled with "on-line" static laser light scattering (LS), refractive index (RI), and ultraviolet (UV) detection can determine the molar mass and oligomeric state in solution of the native proteins as well as glycosylated proteins (Folta-Stogniew, 2006).

5.3.6 Sub-cellular localization

To find out the compartment the protein is located in a cell, sub-cellular localization of the LuALL can be done in tobacco BY-2 cells (as these are easy to transform and maintain in culture and tolerate transformation with fluorescent proteins such as the green fluorescent protein (Brandizzi, Irons, Kearns & Hawes, 2003)), using translational fusions to a fluorescent protein.

5.3.7 Characterizing the exclusively inducible genes

LuALL1 or LuALL17 could be heterologously expressed for further characterization, since these genes were found to be absent normally and expression observed only after hormone treatment.

5.3.8 Purifying the proteins from plants

Using the specific carbohydrate ligands in affinity chromatography the LuALLs can be purified from plants and compared and contrasted with the recombinant protein assays. As well, the glycosylation status of the proteins can be determined.

5.3.9 Structure determination of the single domain LuALLs

Since the single domain LuALL was found to agglutinate erythrocytes, its structure can be determined to provide insights in to the ligand binding mechanism of the LuALLs.

6 References

Adar, R., Streicher, H., Rozenblatt, S., & Sharon, N. (1997). Synthesis of soybean agglutinin in bacterial and mammalian cells. European journal of biochemistry, 249(3), 684-689.

Alvarez-Manilla, G., Warren, N. L., Atwood III, J., Orlando, R., Dalton, S., & Pierce, M. (2010). Glycoproteomic analysis of embryonic stem cells: Identification of potential glycobiomarkers using lectin affinity chromatography of glycopeptides. Journal of proteome research, 9(5), 2062-2075.

An, J., Liu, J. Z., Wu, C. F., Li, J., Dai, L., Van Damme, E., ... & Bao, J. K. (2006). Anti-HIV I/II activity and molecular cloning of a novel mannose/sialic acid-binding lectin from rhizome of *Polygonatum cyrtonema* Hua. Acta biochimica et biophysica sinica, 38(2), 70-78.

Antonyuk, V. O., Yashchenko, A. M., Antonyuk, R. V., & Ambarova, N. O. (2009). Carbohydrate specificity of lectin, purified from the fruiting bodies of *Mycena pura*/Fr./Kumm. and its use in histochemical investigation. Biopolymers and cell *(Kiev)*, *25*, 466-475.

Arango, R., Adar, R., Rozenblatt, S., & Sharon, N. (1992). Expression of *Erythrina corallodendron* lectin in *Escherichia coli*. European journal of biochemistry, 205(2), 575-581.

Arndt, C., Koristka, S., Bartsch, H., & Bachmann, M. (2012). Native polyacrylamide gels. In protein electrophoresis (pp. 49-53). Humana Press.

Audi, J., Belson, M., Patel, M., Schier, J., & Osterloh, J. (2005). Ricin poisoning. JAMA: the journal of the American medical association, 294(18), 2342-2351.

Ayouba, A., Causse, H., Van Damme, E. J., Peumans, W. J., Bourne, Y., Cambillau, C., & Rougé, P. (1994). Interactions of plant lectins with the components of the bacterial cell wall peptidoglycan. Biochemical systematics and ecology, 22(2), 153-159.

Banerjee, N., Sengupta, S., Roy, A., Ghosh, P., Das, K., & Das, S. (2011). Functional alteration of a dimeric insecticidal lectin to a monomeric antifungal protein correlated to its oligomeric status. PloS one, 6(4), e18593.

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A., & Vijayan, M. (1996). Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. Journal of molecular biology, 259, 281-296.

Banerjee, S., Vishwanath, P., Cui, J., Kelleher, D. J., Gilmore, R., Robbins, P. W., & Samuelson, J. (2007). The evolution of N-glycan-dependent endoplasmic reticulum quality control factors for glycoprotein folding and degradation.Proceedings of the national academy of sciences, 104(28), 11676-11681.

Barre, A., Bourne, Y., Van Damme, E. J., Peumans, W. J., & Rougé, P. (2001). Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. Biochimie, 83(7), 645-651.

Bateman, A., & Bycroft, M. (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). Journal of molecular biology, 299(4), 1113-1119.

Bell, H. A., Fitches, E. C., Down, R. E., Marris, G. C., Edwards, J. P., Gatehouse, J. A., & Gatehouse, A. M. R. (1999). The effect of snowdrop lectin (GNA) delivered via artificial diet and transgenic plants on *Eulophus pennicornis* (Hymenoptera: Eulophidae), a parasitoid of the tomato moth *Lacanobia oleracea* (Lepidoptera: Noctuidae). Journal of insect physiology,45(11), 983-991.

Beltrão, E. I. C., Medeiros, P. L., Rodrigues, O. G., Figueredo-Silva, J., Valença, M. M., Coelho, L. C. B. B., & Carvalho, L. B. (2009). *Parkia pendula* lectin as histochemistry marker for meningothelial tumour. European journal of histochemistry, 47(2), 139-142.

Bergmans, H. E., Van Die, I. M., & Hoekstra, W. P. (1981). Transformation in *Escherichia coli*: stages in the process. Journal of bacteriology, *146*(2), 564-570.

Bettler,E., Loris,R. and Imberty,A. (2001). 3D lectin data bank (http://www.cermav.cnrs.fr/databank/lectine).

Bezerra, W. M., Carvalho, C. P., Moreira, R. D. A., & Grangeiro, T. B. (2006). Establishment of a heterologous system for the expression of *Canavalia brasiliensis* lectin: a model for the study of protein splicing. Genetics and molecular research, 5(1), 216-223.

Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B., & Boyd, M. R. (2001). Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. Molecular pharmacology, 59(5), 949-954.

Bourne, Y., Roig-Zamboni, V., Barre, A., Peumans, W.J., Astoul, C.H., Van Damme, E.J., and Rouge, P. (2004) The crystal structure of the Calystegia sepium agglutinin reveals a novel quaternary arrangement of lectin subunits with a beta-prism fold. Journal of biological chemistry, 279 (1), 527–533.

Bourne, Y., Zamboni, V., Barre, A., Peumans, W. J., Van Damme, E. J., & Rougé, P. (1999). *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. Structure, 7(12), 1473-1482.

Boyd, W. C. & Reguera, R. M. (1949). Hemagglutinating substances for human cells in various plants. Journal of immunology, 62, 333–339.

Boyd, W.C. (1954). The proteins of immune reactions. In: Neurath H, Bailey K (ed) The proteins. Academic Press, New York pp 756–844.

Brandizzi, F., Irons, S., Kearns, A., & Hawes, C. (2003). BY-2 Cells: Culture and transformation for live cell imaging. Current protocols in cell biology, 1-7.

Brewin, N. J., & Kardailsky, I. V. (1997). Legume lectins and nodulation by *Rhizobium*. Trends in plant science, 2(3), 92-98.

Briesemeister, S., Rahnenführer, J., & Kohlbacher, O. (2010a). Going from where to why—interpretable prediction of protein subcellular localization. Bioinformatics, 26(9), 1232-1238.

Brill, L.M., Evans, C.J., & Hirsch, A.M. (2000). Expression of MsLEC1- and MsLEC2-antisense genes in alfalfa plants causes severe developmental and reproductive abnormalities. The plant journal: for cell and molecular biology, 25(4), 453-461.

Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B., & Peumans, W. J. (1984). A lectin from elder (*Sambucus nigra* L.) bark. Biochemical journal, 221(1), 163-169.

Broekaert, W. F., Van Parijs, J. A. N., Leyns, F., Joos, H., & Peumans, W. J. (1989). A chitin-binding lectin from stinging nettle rhizomes with antifungal properties. Science, 245(4922), 1100-1102.

Brown, M., & Wittwer, C. (2000). Flow cytometry: principles and clinical applications in hematology. Clinical Chemistry, 46(8), 1221-1229.

Buckley, J.T., Halasa, L.N. & MacIntyre, K.D.L. (1981) Purification and some properties of the hemolytic toxin aerolysin. Canadian journal of biochemistry, 59 (6),430–435.

Cammue, B. P., Peeters, B., & Peumans, W. J. (1985). Isolation and partial characterization of an N-acetylgalactosamine-specific lectin from winter-aconite (*Eranthis hyemalis*) root tubers. The biochemical journal, 227(3), 949-955.

Candy, L., Peumans, W. J., Menu-Bouaouiche, L., Astoul, C. H., Van Damme, J., Van Damme, E. J., ... & Rougé, P. (2001). The Gal/GalNAc-specific lectin from the plant pathogenic basidiomycete *Rhizoctonia solani* is a member of the Ricin-B family. Biochemical and biophysical research communications, 282(3), 655-661.

Carameldo, J.J., & Parodi, A.J., 2008. Getting in and out from calnexin/calreticulin cycles. The Journal of biological chemistry. 283, 10221–10225.

Carrington, D. M., Auffret, A. E. H. D., & Hanke, D. E. (1985). Polypeptide ligation occurs during post-translational modification of concanavalin A. Nature. 313, 64-67.

Chan, Y. S., Wong, J. H., Fang, E. F., Pan, W., & Ng, T. B. (2012). Isolation of a glucosamine binding leguminous lectin with mitogenic activity towards splenocytes and anti-proliferative activity towards tumor cells. PloS one, 7(6), e38961.

Chao, Q. M., Casalongue, C., Quinn, J. M., & Etzler, M. E. (1994). Expression and partial characterization of *Dolichos biflorus* seed lectin in *Escherichia coli*. Archives of biochemistry and biophysics, 313(2), 346-350.

Chao, Q., & Etzler, M. E. (1994). Incorrect targeting of plant vacuolar lectins in yeast. Journal of biological chemistry, 269(33), 20866-20871.

Charungchitrak, S., Petsom, A., Sangvanich, P., & Karnchanatat, A. (2011). Antifungal and antibacterial activities of lectin from the seeds of *Archidendron jiringa* Nielsen. Food chemistry, 126(3), 1025-1032.

Chauhan, J. S., Rao, A., & Raghava, G. P. (2013). *In silico* platform for prediction of N-, O-and C-Glycosites in eukaryotic protein sequences. PloS one, 8(6), e67008.

Chen, D., & Texada, D. E. (2006). Low-usage codons and rare codons of Escherichia coli. Gene therapy and molecular biology, 10, 1-12.

Chen, M. S. (2008). Inducible direct plant defense against insect herbivores: a review. Insect science, 15(2), 101-114.

Chen, Y., Peumans, W. J., Hause, B., Bras, J., Kumar, M., Proost, P., ... & Van Damme, E. J. (2002). Jasmonic acid methyl ester induces the synthesis of a cytoplasmic/nuclear chito-oligosaccharide binding lectin in tobacco leaves. The FASEB journal, 16(8), 905-907.

Costa, F. H. F., Valença, N. S. M. S., Silva, A. R. B. P., Bezerra, G. A., Cavada, B. S., & Rádis-Baptista, G. (2011). Cloning and molecular modeling of *Litopenaeus vannamei* (Penaeidae) C-type lectin homologs with mutated mannose binding domain-2. Genetics and molecular research, 10, 650-664.

Couty A., Down R. E., Gatehouse A. M. R., Kaiser L., Pham-Delegue M.H. & Poppy G. M. (2001). Effects of artificial diet containing GNA and GNA-expressing potatoes on the development of the aphid parasitoid, *Aphidius ervi* Haliday (Hymenoptera: Aphidiidae). Journal of insect physiology, 47(12), 1357-1366.

Cregg, J. M., Vedvick, T. S., & Raschke, W. C. (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. Nature biotechnology, 11(8), 905-910.

Cummings RD, Etzler ME (2009). Antibodies and Lectins in Glycan Analysis. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Damico, D. C. S., Freire, M. G. M., Gomes, V. M., Toyama, M. H., Marangoni, S., Novello, J. C., & Macedo, M. L. R. (2003). Isolation and characterization of a lectin from *Annona muricata* seeds. Journal of protein Chemistry, 22(7-8), 655-661.

Dao-Thi, M. H., Hamelryck, T. W., Poortmans, F., Voelker, T. A., Chrispeels, M. J., & Wyns, L. (1996). Crystallization of glycosylated and nonglycosylated phytohemagglutinin-L. Proteins: structure, function, and bioinformatics, 24(1), 134-137.

Davison, J. (2010). GM plants: science, politics and EC regulations. Plant Science, 178(2), 94-98.

Day, A., Addi, M., Kim, W., David, H., Bert, F., Mesnage, P., ... & Hawkins, S. (2005). ESTs from the fibre-bearing stem tissues of Flax (*Linum usitatissimum* L.): expression analyses of sequences related to cell wall development. Plant biology, 7(1), 23-32.

De Hoff, P. L., Brill, L. M., & Hirsch, A. M. (2009). Plant lectins: the ties that bind in root symbiosis and plant defense. Molecular genetics and genomics, 282(1), 1-15.

De Lima, A. L., Cavalcanti, C. C., Silva, M. C., Paiva, P. M., Coelho, L. C., Beltrão, E. I., & Correia, M. T. D. S. (2010). Histochemical evaluation of human prostatic tissues with *Cratylia mollis* seed lectin. Journal of biomedicine and biotechnology, 2010.

De Souza Filho, G. A., Ferreira, B. S., Dias, J. M., Queiroz, K. S., Branco, A. T., Bressan-Smith, R. E., ... & Garcia, A. B. (2003). Accumulation of SALT protein in rice plants as a response to environmental stresses. Plant science, 164(4), 623-628.

Díaz, C. L., Melchers, L. S., Hooykaas, P. J., Lugtenberg, B. J., & Kijne, J. W. (1989). Root lectin as a determinant of host–plant specificity in the *Rhizobium*–legume symbiosis.

Diaz, C. L., Van Spronsen, P. C., Bakhuizen, R., Logman, G. J. J., Lugtenberg, E. J. J., & Kijne, J. W. (1986). Correlation between infection by *Rhizobium leguminosarum* and lectin on the surface of *Pisum sativum* L. roots. Planta, 168(3), 350-359.

Down, R. E., Ford, L., Bedford, S. J., Gatehouse, L. N., Newell, C., Gatehouse, J. A., & Gatehouse, A. M. (2001). Influence of plant development and environment on transgene expression in potato and consequences for insect resistance. Transgenic research, 10(3), 223-236.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956). Colorimetric method for determination of sugars and related substances. Analytical chemistry, 28(3), 350-356.

Duverger, E., Frison, N., Roche, A. C., & Monsigny, M. (2003). Carbohydrate-lectin interactions assessed by surface plasmon resonance. Biochimie, 85(1), 167-179.

Edelman, G. M., Cunningham, B. A., Reeke, G. N., Becker, J. W., Waxdal, M. J., & Wang, J. L. (1972). The covalent and three-dimensional structure of concanavalin A. Proceedings of the national academy of sciences, 69(9), 2580-2584.

Elfstrand, M. (1898). U ¨ ber blutko¨rperchenagglutinierende Eiweisse. In ‘‘Go¨rberdorfer Vero¨Ventlichungen a. Band I’’ (R. Kobert, ed.), pp. 1–159. Enke, Stuttgart, Germany.

Emanuelsson, O., Nielsen, H., Brunak, S., & von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology, 300(4), 1005-1016.

Esko, J. D., & Sharon, N. (2009). Microbial lectins: hemagglutinins, adhesins, and toxins. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 34.

Etzler, M. E. (1985). Plant lectins: molecular and biological aspects. Annual review of plant physiology, 36(1), 209-234.

Evensen, G., Mathiesen, A., & Sundan, A. (1991). Direct molecular cloning and expression of two distinct abrin A-chains. Journal of biological chemistry,266(11), 6848-6852.

Ewen, S. W., & Pusztai, A. (1999). Effect of diets containing genetically modified potatoes expressing *Galanthus nivalis* lectin on rat small intestine.The Lancet, 354(9187), 1353-1354.

Fahrenkrog, B., & Aebi, U. (2003). The nuclear pore complex: nucleocytoplasmic transport and beyond. Nature Reviews Molecular Cell Biology, 4(10), 757-766.

Fitches, E., Woodhouse, S.D., Edwards, J.P., & Gatehouse, J.A. (2001). In vitro and in vivo binding of snowdrop (*Galanthus nivalis* agglutinin; GNA) and jackbean (*Canavalia ensiformis*; ConA) lectins within tomato moth (*Lacanobia oleracea*) larvae: mechanisms of insecticidal action. Journal of Insect Physiology. 47 (7), 777–787.

Fitchette, A. C., Dinh, O. T., Faye, L., & Bardor, M. (2007). Plant proteomics and glycosylation. In Plant Proteomics (pp. 317-342). Humana Press.

Foissac, X., Thi Loc, N., Christou, P., Gatehouse, A. M., & Gatehouse, J. A. (2000). Resistance to green leafhopper (*Nephotettix virescens*) and brown planthopper (*Nilaparvata lugens*) in transgenic rice expressing snowdrop lectin (*Galanthus nivalis* agglutinin; GNA). Journal of insect physiology, 46(4), 573-584.

Folta-Stogniew, E. (2006). Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. In new and emerging proteomic techniques (pp. 97-112). Humana Press.

Fouquaert, E., Peumans, W. J., Smith, D. F., Proost, P., Savvides, S. N., & Van Damme, E. J. (2008). The “old” Euonymus europaeus agglutinin represents a novel family of ubiquitous plant proteins. Plant physiology, 147(3), 1316-1324.

Fragnière, C., Serrano, M., Abou-Mansour, E., Métraux, J. P., & L'Haridon, F. (2011). Salicylic acid and its location in response to biotic and abiotic stress. FEBS letters, 585(12), 1847-1852.

Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J., & Willingham, M. C. (1996). Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1α, 1β, and 2γ. Biochemistry, 35(47), 14749-14756.

Freire, M. D. G. M., Gomes, V. M., Corsini, R. E., Machado, O. L., De Simone, S. G., Novello, J. C., ... & Macedo, M. L. R. (2002). Isolation and partial characterization of a novel lectin from *Talisia esculenta* seeds that interferes with fungal growth. Plant physiology and biochemistry, 40(1), 61-68.

Frenzel, A., Manthey, K., Perlick, A. M., Meyer, F., Pühler, A., Küster, H., & Krajinski, F. (2005). Combined transcriptome profiling reveals a novel family of arbuscular mycorrhizal-specific *Medicago truncatula* lectin genes. Molecular plant-microbe interactions, 18(8), 771-782.

Frenzel, A., Tiller, N., Hause, B., & Krajinski, F. (2006). The conserved arbuscular mycorrhiza-specific transcription of the secretory lectin MtLec5 is mediated by a short upstream sequence containing specific protein binding sites. Planta, 224(4), 792-800.

Gabius, H. J., Siebert, H. C., André, S., Jiménez-Barbero, J., & Rüdiger, H. (2004). Chemical biology of the sugar code. ChemBioChem, 5(6), 740-764.

Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M., & Vierstra, R. D. (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. Proceedings of the national academy of sciences, 99(17), 11519-11524.

Gao, Z.M., Zheng, B., Wang, W.Y., Li, Q., & Yuan, Q. P. (2011).Cloning and functional characterization of a GNA-like lectin from Chinese Narcissus (*Narcissus tazettavar*. Chinensis Roem). Physiologia plantarum, 142, 193–204.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., & Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, 571-607.

Goldstein, I. J., & Poretz, R. D. (1986). Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. The lectins, properties, functions, and applications in biology and medicine. Academic Press, Orlando, USA, 33-247.

Goni, R., García, P., & Foissac, S. (2009). The qPCR data statistical analysis. Integromics white paper: Integromics SL, 1-9.

Greenfield, N. J., & Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry, 8(10), 4108-4116.

Gupta, A., Gupta, R.K., & Gupta, G.S. (2009). Targeting cells for drug and gene delivery: Emerging applications of mannans and mannan binding lectins. Journal of scientific and industrial research, 68, 465-483.

Gupta, G.S. (2012). Lectins: an overview. In G.S. Gupta, Animal lectins: form, function and clinical applications (pp 3-25). Springer Vienna.

Hanover, J. A. (2001). Glycan-dependent signaling: O-linked N-acetylglucosamine. The FASEB journal, 15(11), 1865-1876.

Hardman K.D., & Ainsworth C.F. (1972). Structure of concanavalin A at 2.4 Å resolution, Biochemistry 11 (26), 4910–4919.

Harper, M. S., Hopkins, T. L., & Czapla, T. H. (1998). Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. Tissue and Cell, 30(2), 166-176.

Hart, G.W., & West, C.M. (2009). Nucleocytoplasmic Glycosylation. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 17.

Hartl, F. U., Hlodan, R., & Langer, T. (1994). Molecular chaperones in protein folding: the art of avoiding sticky situations. Trends in biochemical sciences,19(1), 20-25.

Hegedus, D., Erlandson, M., Gillott, C., & Toprak, U. (2009). New insights into peritrophic matrix synthesis, architecture, and function. Annual review of entomology, 54, 285-302.

Hengartner, M. O. (2000). The biochemistry of apoptosis. Nature, 407(6805), 770-776.

Hernández, P., Debray, H., Jaekel, H., Garfias, Y., del Carmen Jiménez, M., Martínez-Cairo, S., & Zenteno, E. (2001). Chemical characterization of the lectin from *Amaranthus leucocarpus* syn. *hypocondriacus* by 2-D proteome analysis. Glycoconjugate journal, 18(4), 321-329.

Hester, G., Kaku, H., Goldstein, I. J., & Wright, C. S. (1995). Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nature structural & molecular biology, 2(6), 472-479.

Hirano, K., Teraoka, T., Yamanaka, H., Harashima, A., Kunisaki, A., Takahashi, H., & Hosokawa, D. (2000). Novel mannose-binding rice lectin composed of some isolectins and its relation to a stress-inducible salT gene.Plant and cell physiology, 41(3), 258-267.

Hirokawa, T., Boon-Chieng, S., & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics, 14(4), 378-379.

Hoffman, L. M., & Donaldson, D. D. (1987). Synthesis of mitogenic phytohemagglutinin–L in *Escherichia Coli*. Nature biotechnology, 5(2), 157-160.

Holzwarth, G., & Doty, P. (1965). The ultraviolet circular dichroism of polypeptides. Journal of the American chemical society, 87(2), 218-228.

Hom, K., Gochin, M., Peumans, W. J., & Shine, N. (1995). Ligand-induced perturbations in *Urtica dioica* agglutinin. FEBS letters, 361(2), 157-161.

Hopkins, T. L., & Harper, M. S. (2001). Lepidopteran peritrophic membranes and effects of dietary wheat germ agglutinin on their formation and structure. Archives of Insect Biochemistry and physiology, 47(2), 100-109.

Hossain, M. (2012). Critical factors affecting the success of cloning, expression, and mass production of enzymes by recombinant *E. coli*. ISRN Biotechnology, 2013.

Huesing, J. E., Murdock, L. L., & Shade, R. E. (1991). Effect of wheat germ isolectins on development of cowpea weevil. Phytochemistry, 30(3), 785-788.

Huis, R., Hawkins, S., & Neutelings, G. (2010). Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant biology, *10*(1), 71.

Hussain, K., Bowler, C., Roberts, L. M., & Lord, J. M. (1989). Expression of ricin B chain in *Escherichia coli*. FEBS letters, 244(2), 383-387.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., ... & Khurana, J. P. (2007). F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant physiology,143(4), 1467-1483.

Jayat, C., & Ratinaud, M. H. (1993). Cell cycle analysis by flow cytometry: principles and applications. Biology of the cell, 78(1), 15-25.

Jiang, S.Y., Ma, Z.G., & Ramachandran, S. (2010). Evolutionary history and stress regulation of the lectin superfamily in higher plants. BMC Evolutionary biology, 10, 79.

Johnson, M. T., Carpenter, E. J., Tian, Z., Bruskiewich, R., Burris, J. N., Carrigan, C. T., ... & Wong, G. K. S. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PloS one, 7(11), e50226.

Jordan, E. T., & Goldstein, I. J. (1994). The sequence of a second member of the lima bean lectin gene family and the expression and characterization of recombinant lectin in *Escherichia coli*. Journal of biological chemistry, 269(10), 7674-7681.

Jwa, N.-S., Agrawal, G.K., Rakwal, R., Park, C.-H. and Agrawal, V.P. (2001) Molecular cloning and characterization of a novel Jasmonate inducible pathogenesis-related class 10 protein gene, *JIOsPR10*, from rice (*Oryza sativa* L.) seedling leaves. Biochemical and biophysical research communications, 286(5), 973-983.

Kanrar, S., Venkateswari, J., Kirti, P., & Chopra, V. (2002). Transgenic Indian mustard (*Brassica juncea*) with resistance to the mustard aphid (*Lipaphis erysimi* Kalt.). Plant cell reports, 20(10), 976-981.

Kaur, A., Singh, J., Kamboj, S. S., Sexana, A. K., Pandita, R. M., & Shamnugavel, M. (2005). Isolation of an N-acetyl-d-glucosamine specific lectin from the rhizomes of *Arundo donax* with antiproliferative activity. Phytochemistry, 66(16), 1933-1940.

Kim, S. T., Cho, K. S., Yu, S., Kim, S. G., Hong, J. C., Han, C. D., ... & Kang, K. Y. (2003). Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells. Proteomics, 3(12), 2368-2378.

Knibbs, R. N., Goldstein, I. J., Ratcliffe, R. M., & Shibuya, N. (1991). Characterization of the carbohydrate binding specificity of the leukoagglutinating lectin from *Maackia amurensis*. Comparison with other sialic acid-specific lectins. Journal of biological chemistry, 266(1), 83-88.

Koharudin, L. M., Viscomi, A. R., Jee, J. G., Ottonello, S., & Gronenborn, A. M. (2008). The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. Structure (London, England: 1993),16(4), 570-584.

Konami, Y., Yamamoto, K., Osawa, T., & Irimura, T. (1994). Strong affinity of *Maackia amurensis* hemagglutinin (MAH) for sialic acid-containing Ser/Thr-linked carbohydrate chains of N-terminal octapeptides from human glycophorin A. FEBS letters, 342(3), 334-338.

Krishnaswamy, S., Srivastava, S., Mohammadi, M., Rahman, M., Deyholos, M., & Kav, N. N. (2008). Transcriptional profiling of pea ABR17 mediated changes in gene expression in Arabidopsis thaliana. BMC plant biology, 8(1), 91.

Kumar, K.K., Chandra, K.L.p., Sumanthi, J., Reddy, G.S., Shekar, P.C., & BVR, R. (2012). Biological role of lectins: A review. Journal of orofacial sciences, 4(1), 20-25.

Kumar, U. (2007). SDS-PAGE-Sodium dodecyl sulfate polyacrylamide gel electrophoresis. The Journal of department of applied sciences & humanities, IV (2007), 98-100

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., & Matsui, M. (2002). Classification and expression analysis of Arabidopsis F-box-containing protein genes. Plant and cell physiology, 43(10), 1073-1085.

Kusui, K., Yamamoto, K., Konami, Y., & Osawa, T. (1991). cDNA cloning and expression of *Bauhinia purpurea* lectin. Journal of biochemistry, 109(6), 899-903.

Lam, S. K., & Ng, T. B. (2011). Lectins: production and practical applications. Applied microbiology and biotechnology, 89(1), 45-55.

Lamarre-Vincent, N., & Hsieh-Wilson, L. C. (2003). Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. Journal of the American chemical society, 125(22), 6612-6613.

Lannoo, N., & Van Damme, E.J.M. (2010). Nucleocytoplasmic plant lectins. Biochimica et biophysica acta, 1800, 190–201.

Lannoo, N., Peumans, W. J., Pamel, E. V., Alvarez, R., Xiong, T. C., Hause, G., ... & Van Damme, E. J. (2006). Localization and in vitro binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact in situ with high-mannose and complex N-glycans. FEBS letters, 580(27), 6329-6337.

Lannoo, N., Vandenborre, G., Miersch, O., Smagghe, G., Wasternack, C., Peumans, W. J., & Van Damme, E. J. (2007a). The jasmonate-induced expression of the *Nicotiana tabacum* leaf lectin. Plant and cell physiology,48(8), 1207-1218.

Lannoo, N., Vervecken, W., Proost, P., Rougé, P., & Van Damme, E. J. (2007b). Expression of the nucleocytoplasmic tobacco lectin in the yeast *Pichia pastoris*. Protein expression and purification, 53(2), 275-282.

Lechner, E., Achard, P., Vansiri, A., Potuschak, T., & Genschik, P. (2006). F-box proteins everywhere. Current opinion in plant biology, 9(6), 631-638.

Lee, X., Thompson, A., Zhang, Z., Ton-that, H., Biesterfeldt, J., Ogata, C., ... & Young, N. M. (1998). Structure of the Complex of *Maclura pomifera*Agglutinin and the T-antigen Disaccharide, Galβ1, 3GalNAc. Journal of biological chemistry, 273(11), 6312-6318.

Li, C. Y., Meng, L., Liu, B., & Bao, J. K. (2009). *Galanthus nivalis* agglutinin (GNA)-related lectins: traditional proteins, burgeoning drugs? Current chemical biology, 3(3), 323-333.

Li, C. Y., Xu, H. L., Liu, B., & Bao, J. K. (2010). Concanavalin A, from an old protein to novel candidate anti-neoplastic drug. Current molecular pharmacology, 3(3), 123-128.

Lin, K., Simossis, V. A., Taylor, W. R., & Heringa, J. (2005). A simple and fast secondary structure prediction method using hidden neural networks. Bioinformatics, *21*(2), 152-159.

Ling, L. J., Yang, Y. Z., & Bi, Y. R. (2010). Expression and characterization of two domains of *Pinellia ternata* agglutinin (PTA), a plant agglutinin from *Pinellia ternata* with antifungal activity. World journal of microbiology and biotechnology, 26(3), 545-554.

Linman, M. J., Taylor, J. D., Yu, H., Chen, X., & Cheng, Q. (2008). Surface plasmon resonance study of protein-carbohydrate interactions using biotinylated sialosides. Analytical chemistry, 80(11), 4007-4013.

Liu, C. L., Tsai, C. C., Lin, S. C., Wang, L. I., Hsu, C. I., Hwang, M. J., & Lin, J. Y. (2000). Primary structure and function analysis of the *Abrus precatorius* Agglutinin A chain by site-directed mutagenesis. Pro (199) of amphiphilic alpha-helix H impairs protein synthesis inhibitory activity. The journal of biological chemistry, 275(3), 1897-1901.

Liu, J. J., & Ekramoddoullah, A. K. (2006). The family 10 of plant pathogenesis-related proteins: their structure, regulation, and function in response to biotic and abiotic stresses. Physiological and molecular plant pathology, 68(1), 3-13.

Livak, K.J. & Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods, 25, 402–408.

Longstaff, M., Powell, K. S., Gatehouse, J. A., Raemaekers, R., Newell, C. A., & Hamilton, W. (1998). Production and purification of active snowdrop lectin in *Escherichia coli*. European journal of biochemistry, 252(1), 59-65.

Lord, J. M. (1985a). Precursors of ricin and *Ricinus communis* agglutinin. European journal of biochemistry, 146 (2), 411–416.

Lord, J. M. (1985b). Synthesis and intracellular transport of lectin and storage protein precursors in endosperm from castor bean. European journal of biochemistry, 146(2), 403-409.

Lord, J. M., & Roberts, L. M. (1993). Biosynthesis, intracellular transport and processing of ricin. In Post-Translational Modification in Plants (Battey, N. H., Dickinson, H. O., & Hetherington, A. M., eds) pp. 243-255, Cambridge University Press.

Lord, J. M., Hartley, M. R., & Roberts, L. M. (1991). Ribosome inactivating proteins of plants. In Seminars in cell biology (Vol. 2, No. 1, p. 15).

Lord, J. M., Roberts, L. M., & Robertus, J. D. (1994). Ricin: structure, mode of action, and some current applications. The FASEB journal, 8(2), 201-208.

Loris, R., Hamelryck, T., Bouckaert, J., & Wyns, L. (1998). Legume lectin structure. Biochimica et biophysica acta (BBA)-Protein structure and molecular enzymology, 1383(1), 9-36.

Loris, R., Maes, D., Poortmans, F., Wyns, L., & Bouckaert, J. (1996). A structure of the complex between concanavalin A and methyl-3, 6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside reveals two binding modes. The journal of biological chemistry, 271(48), 30614-30618.

Marković-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., ... & Breiteneder, H. (2003). Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. Journal of molecular biology, 325(1), 123-133.

Martínez, J. P., Falomir, M. P., & Gozalbo, D. (2009). Chitin: A structural biopolysaccharide. In encyclopedia of life sciences.

McGee, J. D., Hamer, J. E., & Hodges, T. K. (2001). Characterization of a PR-10 pathogenesis-related gene family induced in rice during infection with *Magnaporthe grisea*. Molecular plant-microbe interactions, 14(7), 877-886.

Melander, M., Åhman, I., Kamnert, I., & Strömdahl, A. C. (2003). Pea lectin expressed transgenically in oilseed rape reduces growth rate of pollen beetle larvae. Transgenic research, 12(5), 555-567.

Miller, M. W., & Hanover, J. A. (1994). Functional nuclear pores reconstituted with beta 1-4 galactose-modified O-linked N-acetylglucosamine glycoproteins. Journal of biological chemistry, 269(12), 9289-9297.

Min, W., Dunn, A. J., & Jones, D. H. (1992). Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. The EMBO journal,11(4), 1303-1307.

Mitaku, S., & Hirokawa, T. (1999). Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. Protein engineering, 12(11), 953-957.

Mizushima, T., Yoshida, Y., Kumanomidou, T., Hasegawa, Y., Suzuki, A., Yamane, T., & Tanaka, K. (2007). Structural basis for the selection of glycosylated substrates by SCFFbs1 ubiquitin ligase. Proceedings of the national academy of sciences, 104(14), 5777-5781.

Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S. R., Katzin, B., Rutenber, E., ... & Robertus, J. D. (1987). The three-dimensional structure of ricin at 2.8 A. Journal of biological chemistry, 262(11), 5398-5403.

Moons, A., Gielen, J., Vandekerckhove, J., Van Der Straeten, D., Gheysen, G., & Van Montagu, M. (1997). An abscisic-acid-and salt-stress-responsive rice cDNA from a novel plant gene family. Planta, 202(4), 443-454.

Mulder, L., Lefebvre, B., Cullimore, J., & Imberty, A. (2006). LysM domains of *Medicago truncatula* NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors. Glycobiology, 16(9), 801-809.

Murzin, A. G., Lesk, A. M., & Chothia, C. (1992). β-trefoil fold: patterns of structure and sequence in the kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. Journal of molecular biology, 223(2), 531-543.

Nagae, M., & Yamaguchi, Y. (2012). Function and 3D Structure of the N-Glycans on glycoproteins. International journal of molecular sciences, 13(7), 8398-8429.

Nagahora, H., Ishikawa, K., Niwa, Y., Muraki, M., & Jigami, Y. (1992). Expression and secretion of wheat germ agglutinin by *Saccharomyces cerevisiae*. European journal of biochemistry, 210(3), 989-997.

Naismith, J. H., & Field, R. A. (1996). Structural basis of trimannoside recognition by concanavalin A. Journal of biological chemistry, 271(2), 972-976.

Nakamura-Tsuruta, S., Kominami, J., Kuno, A., & Hirabayashi, J. (2006). Evidence that *Agaricus bisporus* agglutinin (ABA) has dual sugar-binding specificity. Biochemical and biophysical research communications, 347(1), 215-220.

Nangia-Makker, P., Conklin, J., Hogan, V., & Raz, A. (2002). Carbohydrate-binding proteins in cancer, and their ligands as therapeutic agents. Trends in molecular medicine, 8(4), 187-192.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering, 10(1), 1-6.

Nogueira, N. A., Grangeiro, M. B., da Cunha, R. M., Ramos, M. V., Alves, M. A., Teixeira, E. H., ... & Grangeiro, T. B. (2002). Expression and purification of the recombinant ConBr (*Canavalia brasiliensis* lectin) produced in *Escherichia coli* cells. Protein and Peptide Letters, 9(1), 59-66.

Olsnes, S., Refsnes, K., & Pihl, A. (1974). Mechanism of action of the toxic lectins abrin and ricin. Nature, 249, 627-631.

Oyelaran, O., & Gildersleeve, J. C. (2009). Glycan arrays: recent advances and future challenges. Current opinion in chemical biology, 13(4), 406-413.

Parker, M. W., Van Der Goot, F. G., & Buckley, J. T. (1996). Aerolysin—the ins and outs of a model channel-forming toxin. Molecular microbiology, 19(2), 205-212.

Pédelacq, J. D., Piltch, E., Liong, E. C., Berendzen, J., Kim, C. Y., Rho, B. S., ... & Waldo, G. S. (2002). Engineering soluble proteins for structural genomics. Nature biotechnology, 20(9), 927-932.

Petnual, P., Sangvanich, P., & Karnchanatat, A. (2010). A lectin from the rhizomes of turmeric (*Curcuma longa* L.) and its antifungal, antibacterial and alpha-glucosidase inhibitory activities. Food Science and Biotechnology, 19, 907-916.

Petroski, M. D., & Deshaies, R. J. (2005). Function and regulation of cullin–RING ubiquitin ligases. Nature Reviews Molecular Cell Biology, 6(1), 9-20.

Peumans, W. J., Allen, A. K., & Cammue, B. P. (1986). A new lectin from meadow saffron (*Colchicum automnale*). Plant physiology, 82(4), 1036-1039.

Peumans, W. J., Barre, A., Hao, Q., Rougé, P., & Van Damme, E. J. M. (2000). Higher plants developed structurally different motifs to recognize foreign glycans. Trends in Glycoscience and Glycotechnology, 12(64), 83-102.

Peumans, W. J., Nsimba-Lubaki, M., Peeters, B., & Broekaert, W. F. (1985). Isolation and partial characterization of a lectin from ground elder (*Aegopodium podagraria*) rhizomes. Planta, 164(1), 75-82.

Peumans, W.J., & Van Damme, E.J.M. (1995) Lectins as plant defense proteins, Plant Physiology. 109, 347–352.

Powell, K. S., Spence, J., Bharathi, M., Gatehouse, J. A., & Gatehouse, A. M. (1998). Immunohistochemical and developmental studies to elucidate the mechanism of action of the snowdrop lectin on the rice brown plant hopper, *Nilaparvata lugens* (Stal). Journal of Insect Physiology, 44(7), 529-539.

Pratap, J.V., Jeyaprakash, A.A., Rani, P.G., Sekar, K., Surolia, A., and Vijayan, M. (2002). Crystal structures of artocarpin, a Moraceae lectin with mannose specificity, and its complex with methyl-alpha-Dmannose: implications to the generation of carbohydrate specificity. Journal of Molecular Biology, 317(2), 237–247.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., ... & Finn, R. D. (2012). The Pfam protein families database. Nucleic acids research, 40(1), 290-301.

Puranik, S., Sahu, P. P., Mandal, S. N., Parida, S. K., & Prasad, M. (2013). Comprehensive Genome-Wide Survey, Genomic Constitution and Expression

Profiling of the NAC Transcription Factor Family in Foxtail Millet (*Setaria italica* L.). PloS one, 8(5), e64594.

Pusztai, A. (2002). GM food safety: Scientific and institutional issues. Science as culture, 11(1), 69-92.

Pusztai, A., & Bardocz, S. (Eds.). (1995). Lectins: biomedical perspectives. CRC Press.

Pusztai, A., & Watt, W. B. (1974). Isolectins of Phaseolus vulgaris. A comprehensive study of fractionation. Biochimica et biophysica acta (BBA)-Protein structure, 365(1), 57-71.

Pusztai, A., Bardocz, S., & Ewen, S.W.B. (2008). Uses of plant lectins in bioscience and biomedicine. Frontiers in bioscience, 13, 1130-1140.

Puthoff , D. P., Sardesai, N., Subramanyam, S., Nemacheck, J. A., & Williams, C. E. (2005) Hfr-2, a wheat cytolytic toxin-like gene, is up-regulated by virulent Hessian fly larval feeding. Molecular plant pathology, 6:411-423.

Qin, Q. M., Zhang, Q., Zhao, W. S., Wang, Y. Y., & Peng, Y. L. (2003). Identification of a lectin gene induced in rice in response to *Magnaporthe grisea* infection. Acta botanica sinica-Chinese Edition-, 45(1), 76-81.

Radauer, C., Lackner, P., & Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. BMC evolutionary biology, 8(1), 286.

Raemaekers, R. J., de Muro, L., Gatehouse, J. A., & Fordham-Skelton, A. P. (1999). Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. European journal of biochemistry, 265(1), 394-403.

Rahbé, Y., Sauvion, N., Febvay, G., Peumans, W. J., & Gatehouse, A. M. (1995). Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum*. Entomologia experimentalis et applicata, 76(2), 143-155.

Raikhel, N. V., Mishkind, M. L., & Palevitz, B. A. (1984). Characterization of a wheat germ agglutinin-like lectin from adult wheat plants. Planta, 162(1), 55-61.

Raina, A., & Datta, A. (2002). Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*. Proceedings of the national academy of sciences USA 89:11774.

Ramos, M. V., Grangeiro, T. B., Cavada, B. S., Shepherd, I., Lopes, R. O. D. M., & Sampaio, A. H. (2000). Carbohydrate/glycan-binding specificity of legume lectins in respect to their proposed biological functions. Brazilian archives of biology and technology, 43(4), 0-0.

Randerson, J. (2008). Arpad Pusztai: Biological divide. The guardian, 15.

Rashid, K. Y. (2003). Principal diseases of flax. Flax: The genus Linum. Taylor and Francis Ltd, London, 92-123.

Raval, S., Gowda, S. B., Singh, D. D., & Chandra, N. R. (2004). A database analysis of jacalin-like lectins: sequence–structure–function relationships. Glycobiology, 14(12), 1247-1263.

Renkonen, K. O. (1948). Studies on hemagglutinins present in seeds of some representatives of Leguminoseae. Annales medicinae experimentalis et biologiae fenniae 26, 66–72.

Rhodes, J. M. (1999). Genetically modified foods and the Pusztai affair. BMJ-British medical journal-International Edition, 318(7193), 1284.

Rinderle, S. J., Goldstein, I. J., Matta, K. L., & Ratcliffe, R. M. (1989). Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus caudatus*, that recognizes the T-(or cryptic T)-antigen. The Journal of biological chemistry, 264(27), 16123-16131.

Ripoll, C., Favery, B., Lecomte, P., Van Damme, E., Peumans, W., Abad, P., & Jouanin, L. (2003). Evaluation of the ability of lectin from snowdrop (*Galanthus nivalis*) to protect plants against root-knot nematodes. Plant science, 164(4), 517-523.

Roach, M. J., & Deyholos, M. K. (2007). Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Molecular genetics and genomics, 278(2), 149-165.

Roberts, L. M., Lamb, F. I., Pappin, D. J., & Lord, J. M. (1985). The primary sequence of *Ricinus communis* agglutinin. Comparison with ricin. The Journal of biological chemistry, 260(29), 15682-15686.

Rüdiger, H., & Gabius, H.J. (2001).Plant lectins: Occurrence, biochemistry, functions and applications. Glycoconjugate journal, 18, 589-613.

Rudolph, R., & Lilie, H. A. U. K. E. (1996). *In vitro* folding of inclusion body proteins. The FASEB journal, 10(1), 49-56.

Rutenber, E., & Robertus, J. D. (1991). Structure of ricin B-chain at 2.5 Å resolution. Proteins: Structure, Function, and Bioinformatics, 10(3), 260-269.

S.J. Rinderle, I.J. Goldstein, E.E. Remsen, Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds, Biochemistry 29 (1990) 10555–10561.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., & Vijayan, M. (1996). A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a β-prism fold. Nature structural & molecular biology, 3(7), 596-603.

Sato, Y., Morimoto, K., Kubo, T., Yanagihara, K., & Seyama, T. (2012). High mannose-binding antiviral lectin PFL from *Pseudomonas fluorescens* Pf0-1 promotes cell death of gastric cancer cell MKN28 via Interaction with α2-Integrin. PloS one, 7(9), e45922.

Schouppe, D., Rougé, P., Lasanajak, Y., Barre, A., Smith, D. F., Proost, P., & Van Damme, E. J. (2010). Mutational analysis of the carbohydrate binding activity of the tobacco lectin. Glycoconjugate journal, 27(6), 613-623.

Selitrennikoff, C. P. (2001). Antifungal proteins. Applied and environmental microbiology, 67(7), 2883-2894.

Sheldon, P. S., & Bowles, D. J. (1992). The glycoprotein precursor of concanavalin A is converted to an active lectin by deglycosylation. The EMBO journal, 11(4), 1297-1301.

Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B. & Peumans, W. J. (1987). The elderberry (*Sambucus nigra*) bark lectin recognizes the Neu5Ac(alpha 2-*6*)Gal/GalNAc sequence. The Journal of biological chemistry, 262, 1596–1601.

Sobral, A. P. V., Rego, M. J., Cavalacanti, C. L., Carvalho Jr, L. B., & Beltrão, E. I. (2010). ConA and UEA-I lectin histochemistry of parotid gland mucoepidermoid carcinoma. Journal of oral science, 52(1), 49-54.

Sreekrishna, K., Brankamp, R. G., Kropp, K. E., Blankenship, D. T., Tsay, J. T., Smith, P. L., ... & Birkenberger, L. A. (1997). Strategies for optimal synthesis and

secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. Gene, 190(1), 55-62.

Sreevidya, V. S., Hernandez-Oane, R. J., So, R. B., Sullia, S. B., Stacey, G., Ladha, J. K., & Reddy, P. M. (2005). Expression of the legume symbiotic lectin genes psl and gs52 promotes rhizobial colonization of roots in rice. Plant sScience, 169(4), 726-736.

Srivastava, S., Fristensky, B., & Kav, N. N. (2004). Constitutive expression of a PR10 protein enhances the germination of Brassica napus under saline conditions. Plant and cell physiology, 45(9), 1320-1324.

Stillmark, H. (1888). Uber Ricin, ein giftiges Ferment aus den Samen von Ricinus communis L. und einigen anderen Euphorbiaceen. Thesis University of Dorpat.

Stoger, E., Williams, S., Christou, P., Down, R. E., & Gatehouse, J. A. (1999). Expression of the insecticidal lectin from snowdrop (*Galanthus nivalis* agglutinin; GNA) in transgenic wheat plants: effects on predation by the grain aphid *Sitobion avenae*. Molecular breeding, 5(1), 65-73.

Streicher, H., & Sharon, N. (2003). Recombinant plant lectins and their mutants. Methods in enzymology, 363, 47-77.

Subramanyam, S., Sardesai, N., Puthoff, D. P., Meyer, J. M., Nemacheck, J. A., Gonzalo, M., & Williams, C. E. (2006). Expression of two wheat defense-response genes, Hfr-1 and Wci-1, under biotic and abiotic stresses. Plant Science, 170(1), 90-103.

Subramanyam, S., Smith, D. F., Clemens, J. C., Webb, M. A., Sardesai, N., & Williams, C. E. (2008). Functional characterization of HFR1, a high-mannose N-glycan-specific wheat lectin induced by Hessian fly larvae. Plant physiology,147(3), 1412-1426.

Tague, B. W., & Chrispeels, M. J. (1987). The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast. The Journal of cell biology, 105(5), 1971-1979.

Tai, H. H., Pelletier, C., & Beardmore, T. (2004). Total RNA isolation from *Picea mariana* dry seed. Plant Molecular Biology Reporter, 22(1), 93-93.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution, 28(10), 2731-2739.

Teneberg, S., Alsén, B., Ångström, J., Winter, H. C., & Goldstein, I. J. (2003). Studies on Galα3-binding proteins: comparison of the glycosphingolipid binding specificities of *Marasmius oreades* lectin and *Euonymus europaeus* lectin. Glycobiology, *13*(6), 479-486.

Tinjuangjun, P., Loc, N. T., Gatehouse, A. M. R., Gatehouse, J. A., & Christou, P. (2000). Enhanced insect resistance in Thai rice varieties generated by particle bombardment. Molecular breeding, 6(4), 391-399.

Tomov, B. W., & Bernal, J. S. (2003). Effects of GNA transgenic sugarcane on life history parameters of *Parallorhogas pyralophagus* (Marsh)(Hymenoptera: Braconidae), a parasitoid of Mexican rice borer. Journal of economic entomology, 96(3), 570-576

Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J., & Saper, M. A. (1997). Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. Nature structural biology, 4(10), 779-783.

Ukaji, N., Kuwabara, C., Takezawa, D., Arakawa, K., & Fujikawa, S. (2004). Accumulation of pathogenesis-related (PR) 10/Bet v 1 protein homologues in mulberry (*Morus bombycis* Koidz.) tree during winter. Plant, cell & environment, 27(9), 1112-1121.

Van Damme, E. J. M., & Peumans, W. J. (1989). Developmental changes and tissue distribution of lectin in Tulipa. Planta, 178(1), 10-18.

Van Damme, E. J. M., & Peumans, W. J. (1990). Developmental changes and tissue distribution of lectin in *Galanthus nivalis* L. and *Narcissus* cv. Carlton. Planta, 182(4), 605-609.

Van Damme, E. J. M., Smith, D. F., Cummings, R. and Peumans, W. J. (2009). Glycan arrays to decipher the specificity of plant lectins. In ‘‘The Molecular Immunology of Complex Carbohydrates’’ (A. M. Wu, ed.). Kluwer Academic/ Plenum Publishers, New York, in press.

Van Damme, E. J., Balzarini, J., Smeets, K., van Leuven, F., & Peumans, W. J. (1994). The monomeric and dimeric mannose-binding proteins from the Orchidaceae species *Listera ovata* and *Epipactis helleborine*: sequence homologies and differences in biological activities. Glycoconjugate journal,11(4), 321-332.

Van Damme, E. J., Barre, A., Rougé, P., & Peumans, W. J. (2004b). Cytoplasmic/nuclear plant lectins: a new story. Trends in plant science, 9(10), 484-489.

Van Damme, E. J., Briké, F., Winter, H. C., Leuven, F., Goldstein, I. J., & Peumans, W. J. (1996). Molecular cloning of two different mannose-binding lectins from tulip bulbs. European journal of biochemistry, 236(2), 419-427.

Van Damme, E. J., Culerrier, R., Barre, A., Alvarez, R., Rougé, P., & Peumans, W. J. (2007a). A novel family of lectins evolutionarily related to class V chitinases: an example of neofunctionalization in legumes. Plant physiology,144(2), 662-672.

Van Damme, E. J., Goldstein, I. J., & Peumans, W. J. (1991). A comparative study of mannose-binding lectins from the *Amaryllidaceae* and *Alliaceae*. Phytochemistry, 30(2), 509-514

Van Damme, E. J., Lannoo, N., Fouquaert, E., & Peumans, W. J. (2004a). The identification of inducible cytoplasmic/nuclear carbohydrate-binding proteins urges to develop novel concepts about the role of plant lectins. Glycoconjugate journal, 20(7-8), 449-460.

Van Damme, E. J., Nakamura-Tsuruta, S., Smith, D. F., Ongenaert, M., Winter, H. C., Rougé, P., ... & Peumans, W. J. (2007b). Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. The biochemical journal, 404(1), 51-61.

Van Damme, E.J.M., Allen, A.K., & Peumans, W.J. (1987). Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop bulbs. FEBS Letters. 215, 140-144.

Van Damme, E.J.M., Lannoo, N., & Peumans, W.J. (2008). Plant lectins. Advances in botanical research, 48, 107-209.

Van Damme, E.J.M., Peumans, W.J., Barre, A., & Rougé, P. (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related

proteins with diverse biological roles. Critical reviews in plant science, 17, 575–692.

Van Dammes, E. J., Fouquaert, E., Lannoo, N., Vandenborre, G., Schouppe, D., & Peumans, W. J. E.J.M. Van Damme, E. Fouquaert, N. Lannoo, G. Vandenborre, D. Schouppe, D., W.J. Peumans (2011). Novel concepts about the role of lectins in the plant cell. in: A.M. Wu, The molecular immunology of complex carbohydrates, Kluwer Academic/ Plenum Publishers, New York, in press.

Van Dammes, E. J., Fouquaert, E., Lannoo, N., Vandenborre, G., Schouppe, D., & Peumans, W. J. (2011). Novel concepts about the role of lectins in the plant cell. In the molecular immunology of complex carbohydrates-3 (pp. 271-294). Springer US.

Van der Meer, F. J., de Haan, C. A., Schuurman, N. M., Haijema, B. J., Peumans, W. J., Van Damme, E. J., ... & Egberink, H. F. (2007). Antiviral activity of carbohydrate-binding agents against Nidovirales in cell culture. Antiviral research, 76(1), 21.

Van Driessche, E. (1988). Structure and function of Leguminosae lectins. Advances in lectin research, 1, 73-134.

Van Eijsden, R. R., Díaz, C. L., de Pater, B. S., & Kijne, J. W. (1995). Sugar-binding activity of pea (*Pisum sativum*) lectin is essential for heterologous infection of transgenic white clover hairy roots by *Rhizobium leguminosarum* biovar viciae. Plant molecular biology, 29(3), 431-439.

Van Loon, L. C., Rep, M., & Pieterse, C. M. J. (2006). Significance of inducible defense-related proteins in infected plants. Annual review of phytopathology, 44, 135-162.

Van Parijs, J., Broekaert, W. F., Goldstein, I. J., & Peumans, W. J. (1991). Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. Planta, 183(2), 258-264.

Van Rhijn, P., Goldberg, R. B., & Hirsch, A. M. (1998). *Lotus corniculatus* nodulation specificity is changed by the presence of a soybean lectin gene. The plant cell online, 10(8), 1233-1249.

Vandenborre, G., Groten, K., Smagghe, G., Lannoo, N., Baldwin, I. T., & Van Damme, E. J. (2010). *Nicotiana tabacum* agglutinin is active against Lepidopteran pest insects. Journal of experimental botany, 61(4), 1003-1014.

Vandenborre, G., Lannoo, N., Smagghe, G., Daniel, E., Breite, A., Soin, T., ... & Van Damme, E. J. (2008). Cell-free expression and functionality analysis of the tobacco lectin. In vitro cellular & developmental biology-animal, 44(7), 228-235.

Vandenborre, G., Miersch, O., Hause, B., Smagghe, G., Wasternack, C., & Van Damme, E. J. (2009a). *Spodoptera littoralis*-induced lectin expression in tobacco. Plant and cell physiology, 50(6), 1142-1155.

Vandenborre, G., Smagghe, G., & Van Damme, E.J.M. (2011). Plant lectins as defense proteins against phytophagous insects. Phytochemistry, 72, 1538–1550.

Vandenborre, G., Van Damme, E. J., & Smagghe, G. (2009b). *Nicotiana tabacum* agglutinin expression in response to different biotic challengers. Arthropod-plant interactions, 3(4), 193-202.

Varki A, Cummings R, Esko J, et al., editors. (1999). Plant lectins. In Essentials of Glycobiology (Chapter 30). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Varki, A., Etzler, M.E., Cummings, R.D., & Esko, J.D. (2009). Discovery and classification of glycan-binding proteins. In A. Varki, R.D. Cummings, J.D Esko, H.H. Freeze, P. Stanley, C.R Bertozzi, G.W. Hart, and M.E. Etzler (Ed.), Essentials of Glycobiology (Chapter 26). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., ... & Chu, V. T. (2010). The genome of the domesticated apple (*Malus* X *domestica* Borkh.). Nature genetics, 42(10), 833-839.

Venglat , P., Xiang, D., Qiu, S., Stone, S.L., Tibiche, C., Cram, D., Alting-Mees, M., Nowak, J., Cloutier, S., Deyholos, M., Bekkaoui, F., Sharpe, A., Wang, E., Rowland, G., Selvaraj, G., & Datla, R. (2011). Gene expression analysis of flax seed development. BMC Plant biology, 11, 74.

Vinogradov, A. A., Kudryashova, E. V., Levashov, A. V., & van Dongen, W. M. (2003). Solubilization and refolding of inclusion body proteins in reverse micelles. Analytical biochemistry, 320(2), 234-238.

Von Schaewen, A., & Chrispeels, M. J. (1993). Identification of vacuolar sorting information in phytohemagglutinin, an unprocessed vacuolar protein. Journal of experimental botany, 44, 339-342.

Wang, Z., & Guo, S. (1999). Expression of two insect-resistant genescryIA (b&c)/GNA in transgenic tobacco plants results in added protection against both cotton bollworm and aphids. Chinese science bulletin, 44(22), 2051-2058.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., ... & Deyholos, M. K. (2012). The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal: for cell and molecular biology, 72(3), 461-473.

Wasternack, C., & Parthier, B. (1997). Jasmonate-signalled plant gene expression. Trends in plant science, 2(8), 302-307.

Watkins, W. M. & Morgan, W. T. J. (1952). Neutralization of the anti-H agglutinin in eel by simple sugars. Nature 169, 825–826.

Wiedemann, C., Bellstedt, P., & Görlach, M. (2013). CAPITO-A web server based analysis and plotting tool for circular dichroism data. Bioinformatics. 29 (14), 1750-1757.

Williams, C. E., Collier, C. C., Nemacheck, J. A., Liang, C., & Cambron, S. E. (2002). A lectin-like wheat gene responds systemically to attempted feeding by avirulent first-instar Hessian fly larvae. Journal of chemical ecology, 28(7), 1411-1428.

Wise, I. L., & Soroka, J. J. (2003). Principal insect pests of flax. Flax: the genus Linum, 124.

Wong, J. H., Ng, T. B., Cheung, R. C., Ye, X. J., Wang, H. X., Lam, S. K., ... & Liu, F. (2010). Proteins with antifungal properties and other medicinal applications from plants and mushrooms. Applied microbiology and biotechnology, 87(4), 1221-1235.

Wright, C. S. (1997). New folds of plant lectins. Current opinion in structural Biology, 7(5), 631-636.

Wright, H. T., Sandrasegaram, G., & Wright, C. S. (1991). Evolution of a family of N-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. Journal of molecular evolution, 33(3), 283-294.

Wu, J., Luo, X., Guo, H., Xiao, J., & Tian, Y. (2006). Transgenic cotton, expressing *Amaranthus caudatus* agglutinin, confers enhanced resistance to aphids. Plant breeding, 125(4), 390-394.

Yamamoto, S., & Sakai, I. (1981). Composition and immunochemical properties of glycoproteins with anti-B agglutinin activity isolated from *Euonymus sieboldiana* seeds. International journal of immunogenetics, 8(4), 271-279.

Yamauchi, D., & Minamikawa, T. (1990). Structure of the gene encoding concanavalin A from *Canavalia gladiate* and its expression in *Escherichia coli* cells. FEBS letters, 260(1), 127-130.

Yang, J. C., Shun, C. T., Chien, C. T., & Wang, T. H. (2008). Effective prevention and treatment of *Helicobacter pylori* infection using a combination of catechins and sialic acid in AGS cells and BALB/c mice. The Journal of nutrition, 138(11), 2084-2090.

Yang, X., Kalluri, U. C., Jawdy, S., Gunter, L. E., Yin, T., Tschaplinski, T. J., ... & Tuskan, G. A. (2008). The F-box gene family is expanded in herbaceous annual plants relative to woody perennial plants. Plant physiology, 148(3), 1189-1200.

Yesilirmak, F., & Sayers, Z. (2009). Heterelogous expression of plant genes. International journal of plant genomics, 2009.

Yoneda, Y., Imamoto-Sonobe, N., Yamaizumi, M., & Uchida, T. (1987). Reversible inhibition of protein import into the nucleus by wheat germ agglutinin injected into cultured cells. Experimental cell research, 173(2), 586-595.

Young, N. M., Watson, D. C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., ... & Thibault, P. (1995). C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in Escherichia coli characterization of "ragged ends" by mass spectrometry. Journal of biological chemistry, 270(6), 2563-2570.

Yuan, S., & Lin, H. (2008). Role of salicylic acid in plant abiotic stress. Zeitschrift fur Naturforschung c, 63(5/6), 313.

Zacharius, R. M., Zell, T. E., Morrison, J. H., & Woodlock, J. J. (1969). Glycoprotein staining following electrophoresis on acrylamide gels. Analytical biochemistry, 30(1), 148-152.

Zavala, J. A., & Baldwin, I. T. (2004). Fitness benefits of trypsin proteinase inhibitor expression in *Nicotiana attenuata* are greater than their costs when plants are attacked. BMC ecology, 4(1), 11.

Zhang, W., Peumans, W. J., Barre, A., Astoul, C. H., Rovira, P., Rougé, P., ... & Van Damme, E. J. (2000). Isolation and characterization of a jacalin-related mannose-binding lectin from salt-stressed rice (*Oryza sativa*) plants. Planta, 210(6), 970-978.

Zhao, J., Patwa, T.H., Lubman, D.M., & Simeone, D.M. (2008). Protein biomarkers in cancer: Natural glycoprotein microarray approaches. Current opinion in molecular therapeutics, 10(6), 602–610.

Zhu, K., Huesing, J. E., Shade, R. E., Bressan, R. A., Hasegawa, P. M., & Murdock, L. L. (1996). An insecticidal N-acetylglucosamine-specific lectin gene from *Griffonia simplicifolia* (Leguminosae). Plant physiology, 110(1), 195-202.

Zhumabayeva, B., Chang, C., McKinley, J., Diatchenko, L., & Siebert, P. D. (2003). Generation of full-length cDNA libraries enriched for differentially expressed genes. In generation of cDNA libraries (pp. 223-237). Humana Press.

Zuo, Z., Fan, H., Wang, X., Zhou, W., & Li, L. (2012). Purification and characterization of a novel plant lectin from *Pinellia ternata* with antineoplastic activity. SpringerPlus, 1(1), 1-9.