# Durham E-Theses

## *Molecular studies on seed proteins from mature seeds of Phaseolus Vulgaris L.*

Martin Mühling

# MOLECULAR STUDIES ON SEED PROTEINS FROM MATURE SEEDS OF *PHASEOLUS VULGARIS* L.

## BY

## MARTIN MÜHLING

A thesis submitted to the Department of Biological Sciences

University of Durham

In accordance with the requirements for the degree of

Master of Science

August 1996

# DECLARATION

.

I declare that the work contained within this thesis submitted by me for the degree of Master of Science is my own original work, except where otherwise stated, and has not been submitted previously for a degree at this or any other University.

Parts of the work in this thesis have been presented in the following publication (see appendix F):

Mühling, M., J. Gilroy and R. R. D. Croy (1996). Legumin Proteins from Seeds of *Phaseolus vulgaris* L. Journal of Plant Physiology (in press).

Martin Mühling

# Molecular studies on seed proteins from mature seeds of *Phaseolus vulgaris* L.

Martin Mühling

MSc Thesis 1996

## Abstract

In order to elucidate the question of the presence or absence of legumin in the seeds of *Phaseolus vulgaris* L. it was decided to investigate the globulin fraction of seed proteins from mature seeds of French bean. A legumin protein was partially purified and the N-terminal sequences of both subunit polypeptides of a 75 kDa legumin subunits determined, which proved unequivocally that these polypeptides are subunits of a *P. vulgaris* legumin. The N-terminal amino acid determination of further five disulphide-linked polypeptides of lower $M_r$ revealed that all these polypeptides may well derive from the 75 kDa *P. vulgaris* legumin as the N-terminal amino acids showed complete sequence homology. The size differences of these legumin subunits might be due to peptide cleavage at the C-terminal end of the legumin α-chains. Potential mechanisms for this catalytic event are discussed.

In addition, a second protein, Glycoprotein I, an albumin type seed protein, was isolated from the albumin fraction of the seed proteins of *P. vulgaris* and the N-terminal amino acids of its subunits were determined. The obtained sequence data of the three different subunits revealed high similarity to the hyper-variable region of legumin J and K from pea. Whilst it is not inconceivable that both polypeptides represent independent protein subunits reflecting the same evolutionary mechanisms to increase the nitrogen content of storage proteins, based on the sizes of the Glycoprotein I subunits and the sequence similarity to the hyper-variable region of α-chains of legumins the hypotheses is discussed that Glycoprotein I may derive from a *P. vulgaris* legumin by cleavage of peptides from the C-terminal end of the α-chain and formation of a new polypeptide, Glycoprotein I, with new solubility characteristics (water-soluble like albumins).

Furthermore, Glycoprotein I was investigated during seed germination by immunodetection with antibodies raised against the purified protein. The results showed that Glycoprotein I serves as a reserve protein in an early stage of seed germination.

In order to obtain extensive sequence information on legumins from *P. vulgaris* it was decided to attempt to amplify a legumin gene fragment from genomic DNA of *P. vulgaris* using the polymerase chain reaction technique. Southern analysis of PCR products using a pea-legumin as a probe gave strong evidence for the successful amplification of legumin encoding genes from genomic *P. vulgaris* DNA. However, neither the cloning of these PCR products nor the repeats of this PCR experiment were successful.

A new PCR approach to amplify legumin gene fragments from genomic DNA of *P. vulgaris* using an oligonucleotide sequence for the reverse primer designed on the basis of the N-terminal amino acid sequence of the β-chain is discussed.

The nucleotide sequence of a PCR artefact of one of the unsuccessful repeats of the PCR experiment showed sequence similarity to Arcelin 2, a gene found in wild types of *P. vulgaris*. Arcelin proteins provide the plant with resistance against an important pest weevil, but have never been observed in cultivars of *P. vulgaris*. Arcelin genes are thought to have been either totally lost or inactivated during domestication of the French bean.

Further isolated and identified proteins from the albumin fraction of seed proteins from *P. vulgaris* include a superoxide dismutase (SOD). The determination of the N-terminal amino acids of the subunit poypeptide provides the first reported sequence information of a SOD in mature seeds of higher plants. The potential localisation within the seed tissue and the putative role of the presence of a SOD in mature seeds during dormancy is discussed.

Two more polypeptides were isolated from the albumin fraction and identified by N-terminal sequencing as chitinases. Whereas pvchi25 showed high sequence similarity to class II chitinases, pvchi32 was shown to be a class III chitinase. Pvchi32 provides the first reported class III chitinase in mature seeds of higher plants and the second so far described glycosylated chitinase. The potential use of these seed chitinases in crop protection is discussed. Furthermore, the enzymatic activity of both enzymes was tested as well as cross-reactivity to antibodies raised against ethylene-induced bean leaf chitinases. Pvchi25 was observed during seed germination by immunodetection confirming its role in the protection of the seed during dormancy and early germination.

# ACKNOWLEDGEMENTS

# CONTENTS

## Appendix

A. Media, buffers and antibiotics Enzymes

B. Preparation of SDS-polyacrylamide gels

C. Preparation of DNA-marker (digested with Pst I)

D. IUB codes for degenerate oligonucleotides

E  Amino acid - and nculeotide alignment of legume legumin proteins and their encoding cDNAs

F. Publication of parts of this thesis in the *Journal of Plant Physiology* (in press)

# ABBREVIATIONS

| | |
|---|---|
| Abs | Absorbance |
| Amp | Ampicillin |
| bp / kb | Base pairs / a thousand base pairs |
| CIP | Calf intestinal phosphatase |
| 2D | 2-dimensional |
| Da / kDa | Dalton / kilo Dalton |
| $dH_2O$ | Glas-distilled water |
| $ddH_2O$ | Glas-distilled and deionised water |
| DMF | N,N-Dimethylformamide |
| DMSO | Dimethylsulfoxide |
| DNA | Desoxyribonucleine Acid |
| dNTP | Deoxiribonucleoside triphosphate |
| DTT | Dithiothreitol |
| *E.coli* | *Escherichia coli* |
| EDTA | Ethylendiaminetetraacetic acid |
| ER | Endoplasmic reticulum |
| EtBr | Ethidium bromide |
| Glyco14/15/17/18 | Glycoprotein I subunit of $M_r$ 14 kDa/15 kDa/17 kDa/18 kDa |
| h | hour(s) |
| HVR | Hyper-variable region |
| IPTG | Isopropyl-$\beta$-D-thiogalactoside |
| L | liter |
| min | minute(s) |
| $M_r$ | molecular weight |
| NC | Nitrocellulose |
| OD | Optical density |
| PAGE | Polyacrylamide gel electrophorese |
| PAS | Periodic acid Schiff's reagent |
| PCR | Polymerase Chain Reaction |
| PEG | Poyethylene glycol |
| pers. communication | personnal communication |
| PHA | Phytohemagglutinin |
| pI | Isoelectric point |
| pvchi32/pvchi25 | *Phaseolus vulgaris* seed chitinases of $M_r$ 32 kDa/25 kDa |
| PVSOD | *Phaseolus vulgaris* seed superoxide dismutase |
| rel. sat. | Relative saturation |
| RER | Rough endoplasmic reticulum |
| SDS | Sodium dodecyl sulphate |
| sec | seconds |
| SOD | Superoxide dismutase |
| TEMED | N, N, N', N', - tetramethylethylenediamine |
| TE | Tris/EDTA-Buffer |
| TGB | Tris-Glycine-Buffer |
| Tris | Tris (hydroxymethyl) aminomethane |
| U | Unit (enzyme calibration) |
| UV | Ultraviolet |
| VR | Variable region |
| X-Gal | 5-bromo-4-chlor-3-indolyl-$\beta$-D-galactoside |

# FIGURES

# TABLES

# 1. Introduction

Seeds are one of the richest sources of plant proteins. They provide an important source of protein for human and livestock nutrition. Among major crop plants, seeds contain 10% to 40% protein (Murray, 1984) which is why seeds have been domesticated thousands of years. Besides the cereals, the legumes have provided a major food source for human nutrition, including *Phaseolus vulgaris*, the French bean, which - like most of the modern crop plants - served as a major food source originally only in a certain regions before it was domesticated and introduced to other countries and continents.

## 1.1 Origin of the French Bean (*Phaseolus vulgaris*)

The first "modern", meaning large-seeded *Phaseolus* beans appeared about 8,000 - 6,000 BC in South America, 5,000 BC in Mexico and between 5,000 and 3,500 BC in North America (Kaplan, 1981). These facts are in agreement with the suggestion, that the potential origin of the French bean lies in two major geographic gene pools, one distributed in Middle America (Mexico and Guatemala) and one in the Andean mountains of South America (Bolivia) (Kaplan, 1981). For these investigations, Kaplan used certain traits, such as photoperiod specificity.

More recent investigations have involved the main seed storage protein of *P. vulgaris*, phaseolin, as a phylogenetic marker. The two most prevalent phaseolin protein types among cultivars are the S- and T- types (Brown *et al.*, 1982; Gebts *et al.*, 1986). The S-type is represented in cultivars which are domesticated from wild forms of the Middle American centre of origin (Mexico, Central America and Columbia), whereas the T-type is found among cultivars evolved from wild beans in the Andean centre of origin (Southern Peru, Bolivia, Argentinia and Chile) (Brown *et al.*, 1982; Gebts *et al.*, 1986; König and Gebts, 1989). This is in agreement with the fact that wild beans of the Middle American centre show S-type phaseolin and Andean wild beans show T-phaseolin types (Gebts *et al.*, 1986; Gebts and Bliss, 1986; König *et al.*, 1990). Recently, wild *P. vulgaris* populations have been found in the mountainous area of northern Peru and Ecuador (Debouck *et al.*, 1989). The investigation of the phaseolin locus in these beans showed an I- ("Inca") phaseolin type that has not been described in other wild or cultivated French beans.

1

Allozyme analyses suggested that these populations were intermediate between the Middle American and Andean gene pools (Debouck *et al.*, 1993).

In order to prove these assumptions, analyses of phaseolin diversity at the DNA level have been performed (Kami *et al.*, 1995). These investigations are based on the presence of tandem direct repeat sequences of 15 bp in the fourth exon or 27 bp in the sixth exon of the T-phaseolin type genes and of 27 bp in the sixth exon of the S-phaseolin type genes. Up to date no S-phaseolin gene has been identified wih a 15 bp-tandem direct repeat, although this result may be due to insufficient sampling of S-phaseolin genes thus far (Anthony *et al.*, 1990; Kami and Gebts, 1994). The results of this genetic fingerprinting showed that neither the 15 bp and 27 bp tandem repeats nor a newly discovered 21 bp tandem repeat in the third intron of both T- and S-phaseolin types are present in the genome of the wild beans with the I-phaseolin type genes. This led the authors (Kami *et al.*, 1995) to the speculation that duplications of the 15 bp, 27 bp and 21 bp stretches have generated the tandem direct repeats, meaning that the I-phaseolin type represents the phaseolin ancestor. This evolutionary event seems more likely than deletions which have specifically eliminated a member of a tandem direct repeat, because the former could occur in many locations whereas the latter would occur only at the sites of the tandem direct repeats. Therefore, Kami and coworkers (1995) suppose that the origin of the modern beans lies in the mountainous areas of Ecuador and northern Peru, an area that is geographically just intermediate between that of the Middle American and the Andean gene pool. These speculations are confirmed by isozyme analyses indicating, that wild beans from the mountainous area of Ecuador and northern Peru are distinct from and intermediate between wild *P. vulgaris* forms of Middle America and the Southern Andes (König and Gebts, 1989; Debouck *et al.*, 1993).

If these speculations are right and the I-phaseolin type represents the ancestor of the phaseolin protein, the following scenario can be proposed (Kami et al., 1995): Before the divergence of the two geographic gene pools took place as suggested by Kaplan (1981), the duplication of the 21- and 27 bp tandem direct repeats took place in the third intron and the sixth exon of members of the phaseolin gene family. Later, the 15 bp duplication was introduced in the southern dissemination branch and thus came to predominate in the South American gene pool of *P. vulgaris* (Kami *et al.*, 1995).

## 1.2 Phylogeny of *Phaseolus vulgaris*

The legume family (*Leguminosae* or *Fabaceae*) is the third largest family of flowering plants, with 650 genera and nearly 20,000 species (Doyle, 1994). Most attention has been paid on agricultural and economically portant species, notably various "beans". Therefore, French bean (*Phaseolus vulgaris*), broad or field bean (*Vicia faba*) and soybean (*Glycine max*) as well as pea (*Pisum sativum*) have been investigated for a long time in order to evaluate the taxonomy of the family. For this purpose different phylogenetic markers have been used of which the molecular markers have been recognised as the most informal ones (Doyle, 1993).

The results from these phylogenetic investigations basically confirm the more classical taxonomy of the legume family in which the genus *Phaseolus*, *Vigna*, *Glycine* and others form the tribe *Phaseoleae* whereas *Vicia* and *Pisum* are members of the tribe *Vicieae* (Frohne and Jensen, 1992). Based on a 78-kb-inversion in the cpDNA, the genus *Phaseolus* and *Vigna* are regarded as monophyletic within the *Phaseoleae* and represent an own subtribe, the *Phaseolinae* (Frohne and Jensen, 1992). *Glycine* forms the subtribe *Glyciniae* (Frohne and Jensen, 1992). Besides common morphological characteristics both subtribes exhibit high similarity of secondary compounds (Frohne and Jensen, 1992).


## 1.3 Seed proteins - an overview

The main part of the protein content of seeds is made up of storage proteins which have no enzymatic activity, and simply provide a source of nitrogen, sulphur and carbon skeletons for the developing seedling until it is able to produce its own amino acids.

Because of the nutritional value of seed storage proteins much research has been devoted over the past decades to increasing the content of methionine and essential amino acids like lysine thus improving the nutritional quality of seed proteins. A fundamental milestone was produced by Osborne's investigations which provided the basis for his classification of seed proteins. According to Osborne (Osborne, 1924), seed storage proteins can be divided into four classes in relation to their solubility properties, so representing a useful extraction and fractionation scheme: (1) albumins - soluble in water and dilute buffers at neutral pH's; (2) globulins - soluble in salt solutions, but insoluble in water; (3) glutelins - soluble in dilute acids or alkali solutions; (4) prolamins - soluble in

aqueous alcohols (70-90 % ethanol). Although this classification is far from ideal, since it is only based on Osborne's physico-chemical extraction procedures, a better one has still to be devised.

### 1.3.1 Globulins

The most widely distributed protein-group is the globulin group. At the end of the last century Osborne and Campbell (1898) made the distinction between two kinds of globulins from pea, legumin and vicilin. Danielsson (1949) using ultracentrifugation was able to separate both protein fractions on the basis of size difference and determined their sedimentation-coefficients as 11S (legumins) and 7S (vicilins) (Danielsson, 1949).

### 1.3.1.1 Legumin - The 11S globulin

The legumin-like proteins have been found in the seeds of almost all angiosperms examined and represent one of the two most abundant types of seed storage proteins in the dicotyledonous plants (Derbyshire *et al.*, 1976). The second class are the vicilin-like proteins described later. The proportion of legumin of the total seed storage protein varies considerably, ranging from almost 50% in some soybean cultivars (Nielsen, 1984) to its apparent absence from barley and rye seeds (Derbyshire *et al.*, 1976). In general, these proteins are more prevalent in the legume seeds than in cereals, even though it is now well established that the oat globulins (Takwaia *et al.*, 1987) as well as the rice (Walburg *et al.*, 1986) and wheat ( Singh *et al.*, 1985, 1988, 1993) glutelins, which are the major storage proteins in the seeds of these cereals, also belong to the legumin family.

Therefore, legumin can be regarded as ubiquitous in angiosperms and gymnosperms (Häger *et al.*, 1995). DeCamp *et al.* (1994) claim to have found a sporophytic 11S globulin protein fraction in the corm of quillwort (*Isoetes echinospora* Dur.), a member of the quite ancient plant group *lycophyta*. However, the hexameric 11S proteins identified by DeCamp and which are stored in protein bodies did not show any disulfide bonds between subunit polypeptides (DeCamp *et al.*, 1994) and no (e.g. N-terminal) protein sequence was reported.

4

## A) Structure of the Legumin protein

The physical structure of the legumin proteins has been investigated using a number of techniques. Based on electron microscopy of sunflower (*Helianthus annuus*) legumin, helianthin, Reichelt and coworkers (1980) concluded that the 11S-complex is composed of six approximately spherical subunits arranged in two trimers which are stacked one on top of the other but offset by 60°, forming a trigonal antiprism (figure 1.1).

This structure was confirmed by Nielsen and coworkers (Nielsen *et al.*, 1985) investigations on soybean legumins (glycinins). Treatment of glycinin with chaotropic agents such as urea or guanidinium chloride, or with the detergent SDS leads to a dissociation which occurs gradually by formation of smaller complexes (Nielsen *et al.*, 1985). Thus the 360 kDa complex dissociates into two 180,000 Da-complexes which further dissociate gradually to six polypeptides with $M_r$'s averaging 60 kDa. The last level of the dissociation step occurred by treatment of these 60 kDa-complexes with disulphide-bond reducing agents such as 2-mercaptoethanol or dithiothreitol which finally leads to the separation of two polypeptides of approximate $M_r$'s 40 kDa and 20 kDa. Whereas the first polypeptide class exhibits acidic properties (pI in *Pisum* legumins between 4.85 and 6.15 (Matta *et al.*, 1981) and is therefore designated as α-subunit, the smaller type of polypeptides is characteristically basic (e.g. pI in *Pisum* legumin between 7.43 and 8.0 (Matta *et al.*, 1981) which led to the designation as β-subunit. Both are covalently linked by a single disulphide bond always forming a $NH_2$-acidic-basic-COOH polypetide (figure 1.2).

The position of the cysteine residues which form the disulfide-bond between the acidic and basic polypeptides are highly conserved within the primary protein structure. For instance, Staswick and coworkers (Staswick *et al.*, 1984) determined the position of the disulfide-bond in a soybean legumin. Here, Cys 86 of the α-chain is connected to the Cys 7 of the β-chain (amino acid numbering relative to the N-terminus of each polypeptide). Whereas the absolute position of the disulphide bond forming cysteine residue in the acidic subunit varies slightly in the primary structure, the seventh amino acid in the basic subunit is always the disulphide-bond building cysteine. Overall, the basic polypeptide represents the most highly conserved part of the protein. It was therefore proposed (Lawrence *et al.*, 1994 and earlier workers) that the β-subunit probably forms the core of the tertiary legumin structure. This suggestion is supported by a hydrophilicity/hydropathy plot which shows that the β-subunit is - in contrast to the α-subunit - more hydrophobic.

TRIMER
(7-8 S; Mr ≅ 180,000)

HEXAMER
(11-12 S; Mr ≅ 360,000)

MONOMER
(Mr ≅ 60 ,000)

BASIC
SUBUNIT
(Mr ≅ 20,000)

ACIDIC
SUBUNIT
(Mr ≅ 40,000)
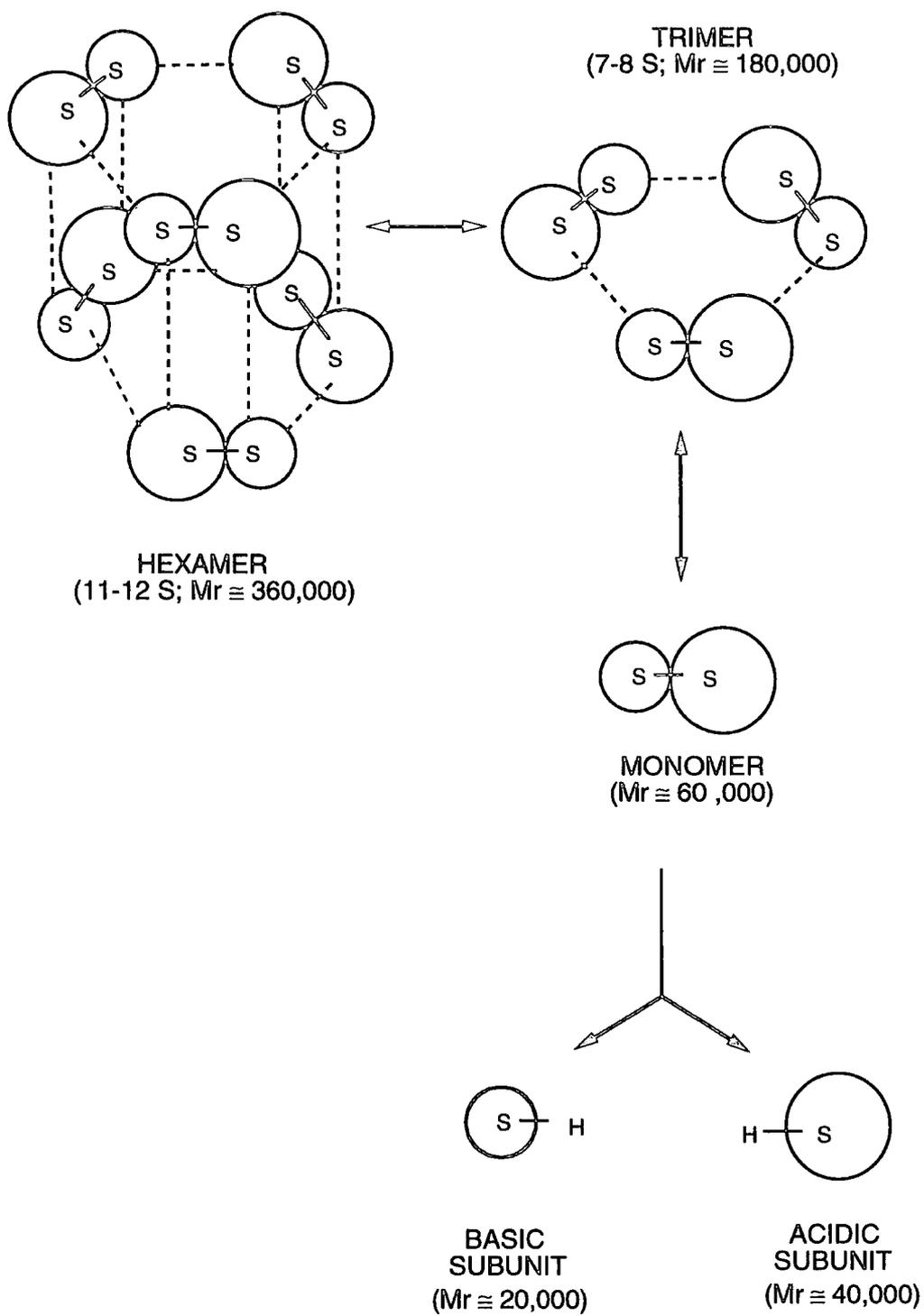
Figure 1.1: Proposed quaterny structure of legumin.
Taken from Pernollet and Mossè (1983)

6

The α-subunit is built up of six regions of which three are conserved and three are variable (figure 1.2; compare legumin alignment in Lawrence *et al.*, 1994). The variable regions (VR) are distinguished not only by variation in the amino acid sequence but also by their length. The COOH-terminal region of the α-subunit - the hyper-variable region (HVR) - represents the most variable part of the protein. These variable regions are alone responsible for size differences in legumins. Many legumins exhibit repetitive sequences which represents the results of sequence duplications. For instance, a *Brassica napus* legumin shows twelve times the amino acid sequence "GQQ" within the variable region (Rödin *et al.*, 1990) and the HVR region of the legumin-like protein from wheat seeds, triticin, contains a unique decapeptide repeat motif which is reiterated four times (Singh *et al.*, 1993).

Further investigations using highly resolving methods like 2D-electrophoresis combined with isoelectric focusing showed clear polymorphisms of the α- and β-subunit of legumin proteins which are isolated from the same species (e.g. Matta *et al.*, 1981). Such differences concern both the molecular weight and the isoelectric points revealing that the legumins are encoded by a multi-gene family (see chapter "Genetics of legumins"). In particular the α-subunits exhibits the larger range of differences.

Secondary structure predictions based on the primary amino acid sequences and the evaluation of circular dichroism measurements have permitted an estimation of the contents of α-helix, β-sheet and β-turns in the structure of legumins. However these predictions are largely inaccurate and partially contradict themselves (e.g. Zirwer et al., 1985).

## B) Genetics of legumins

The pioneering work of the Durham groups showed that the α- and β-subunits of legumins are encoded together by a single gene (Croy *et al.*, 1980a; Croy *et al.*, 1982). Therefore, legumin polypeptides of $M_r$ about 60 kDa can be synthesised by translation experiments by addition of poly(A)[+]-RNA from developing seeds to wheat germ or reticulocyte lysate systems (Evans *et al.*, 1979, Croy *et al.*, 1980a, Barton *et al.*, 1982). The resulting pre-prolegumin exhibits the structure $NH_2$-leaderpeptide-α-subunit-β-subunit-COOH (figure 1.2; e.g. Fukazawa *et al.*, 1985; Pang *et al.*, 1989; Hayashi *et al.*, 1988).

S ———————————————————————— S

NH2 ████ ▨▨ ████ ▨▨ ████ ▨▨▨ COOH    NH2 ████████ COOH

VRI     VRII        HVR

←——————— acidic ——————→          ←— basic —→
          subunit                    subunit
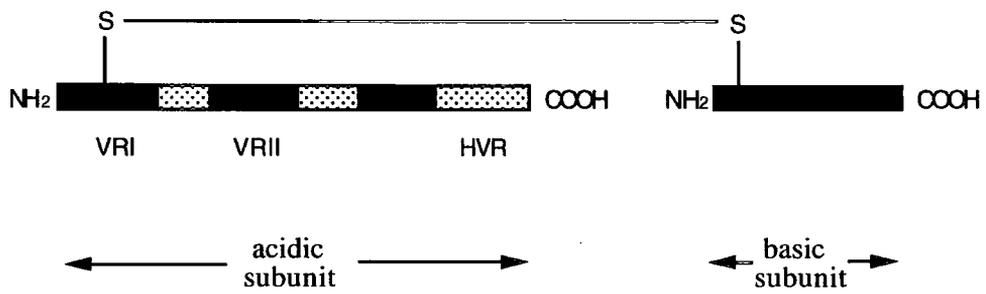
**Figure 1.2: Schematic illustration of the arrangement of polypeptide chains in legumins.**

The black areas represent conserved regions, the stippled areas variable regions (VR = variable region; HVR = hyper- variable region) (after Dank, 1994).

As shown for the legumin proteins, the legumin-encoding genes exhibit extensive heterogeneity and show highest variability in the 3'-end of the α-subunit-encoding region. The main reason for the heterogeneity of legumin encoding genes is the occurrence of a legumin multi-gene family. Based on sequence homology, legumin genes of a single species can often be grouped into different gene families. In *Ginkgo biloba*, for example, three legumin gene families could be identified (Häger *et al.*, 1995; Müller, pers. communication). The same results were obtained from investigations on other plant species such as soybean (Nielsen *et al.*, 1989) and Vicia faba (e.g. Heim *et al.*, 1994).

Further gene heterogeneity arises due to the intron/exon structure of the genes. From nucleotide sequencing of genomic DNA it has emerged that legumin genes of angiosperms generally contain three introns which interrupt the coding sequence twice in the α-subunit and once in the β-subunit (figure 1.3). In many of the legumin genes so far investigated, these introns are at homologous positions (e.g. Lycett *et al.*, 1984; Häger *et al.*, 1995). However, exceptions have been described - in a legumin gene from *Pisum sativum* the first intron was found to be missing (Gatehouse *et al.*, 1988) and in a *Helianthus annuus* legumin gene the third intron was missing (Shotwell & Larkins, 1989). In contrast Häger *et al.* (1995) showed that *Ginkgo biloba* legumin genes contained one more intron interrupting the coding region for the C-terminal part of the β-polypeptide. Whereas the first three intron positions in the legumin genes are highly conserved among legumin genes of many other species, this fourth intron could not be found in any of the legumin genes of angiosperms ivestigated. Moreover, Häger and coworkers (1996) showed that this fourth intron in the legumin genes from *Ginkgo biloba* is ancestral and was lost during evolution. The research involved an investigation of 46 cloned PCR amplifications of 22 different gymnospermous species. Each of them contained the fourth intron interrupting the region coding for the C-terminal part of the β-polypeptide, so confirming the hypothesis, that legumin evolution occurred by successive loss of the most 5' located intron in early angiosperm evolution (Häger *et al.*, 1995).

9
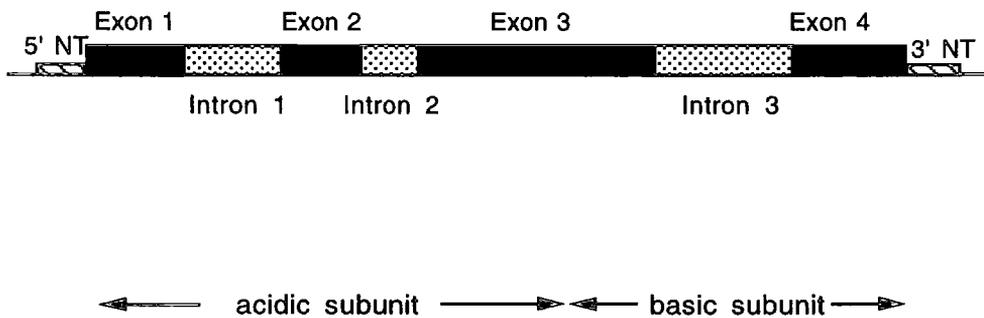
**Figure 1.3 :  Fundamental organisation of legumin genes.**

5'-NT and 3'-NT: 5'- and 3'- non-translated regions of the mRNA (after Shotwell & Larkins, 1989).

10

## C) Synthesis, assembly, transport and deposition of legumins in plant seeds

The natural synthesis and accumulation of seed storage proteins only takes place in the seed. The expression of the storage protein-encoding genes is strongly developmentally controlled and is restricted to a precise period during seed development. After a period of rapid cell division at the start of seed maturation, cell division stops. Further growth of the embryo depends then mainly on cell expansion and the storage tissues start to accumulate reserves in large quantities (Gatehouse and Shirsat, 1992). After the start of deposition of carbohydrates as starch grains, a large amount of storage protein encoding mRNA is transcribed in the cells of the endosperm which is the major storage tissue in monocots or in the cotyledonary storage tissue of most of the dicots (Gatehouse and Shirsat, 1992). The accumulation of the storage proteins occurs rapidly. Evans and coworkers (1979) estimated that a pea seed with a fresh weight of about 500 mg synthesises approximately 4.5 mg of storage protein per day midway through seed development.

The expression pase of seed storage protein-encoding genes during seed development depends very much on the plant species. Generally the amount of synthesised mRNA gives about a Gaussian-shaped curve when plotted against time after anthesis (Gatehouse *et al.*, 1986). Related to the temporal protein accumulation this observation shows, that little or no seed storage protein turnover takes place relative to the protein accumulation.

A general model has been developed for the assembly and transport of seed storage proteins. The storage proteins are synthesised by membrane-bound polysomes on the rough endoplasmic reticulum (RER) within the cells of the cotyledons or the embryo (Bollini and Chrispeels, 1979; Chrispeels, 1983). The first 20 or more amino acids represent the signal peptide and direct the transport of the newly synthesised polypeptide into the lumen of the (RER). For this purpose the signal peptide exhibits highly hydrophobic and basic characteristics (Nielsen, 1985). After the transport through the membrane of the RER the signal peptide is cotranslationally cleaved (Shotwell and Larkins, 1989). In the lumen of the endoplasmic reticulum the pro-legumin monomers assemble in a "self-assembly"-process to trimers (Dickinson *et al.*, 1990). For this assembly the β- subunits with their more hydrophobic character are of essential importance to build the core of the trimer whereas the acidic α-subunits are necessary in order to interact with solvent to keep the trimer molecules in solution. The assembly of the trimers

to the hexamers does not take place immediately as the cleavage of the pro-legumin polypeptide to the α- and β-subunit is a necessary pre-requisite for this to take place (Dickinson *et al.*, 1989).

A further process, that takes place within the lumen of the endoplasmatic reticulum is the formation of the disulphide bridge between the α- and β- subunit (Chrispeels *et al.*, 1982). For the rare case, that the mature legumin is glycosylated, the cotranslational addition of an oligosaccharide to an asparagine residue is performed within the endoplasmatic reticulum too but later modifications can occur in the Golgi complex as shown for the phytohemagglutinin and phaseolin of *P. vulgaris* (Chrispeels, 1983).

From the endoplasmic reticulum the trimers are transported to the central vacuole via the Golgi complex (Chrispeels, 1983; Harris *et al.*, 1993). In the vacuole the trimers deposit first in the periphery (Craig *et al.*, 1980). There, the final processing takes place, including the proteolytic cleavage of the prolegumin into the α- and β-subunits and their assembly to the mature 11S legumin. With further accumulation of storage proteins the central vacuole starts to subdivide forming the protein bodies. These compartments are spherical structures with a protein matrix surrounded by a single membrane and - by definition - characteristic for the storage of seed reserve proteins. In contrast to this model, more recent investigations show, that another group of protein bodies may arise directly from the endoplasmic reticulum (Robinson *et al.*, 1995). The two different groups of protein bodies, however, apparently show no differences in composition in legume cotyledons (Hinz *et al.*, 1995). Robinson *et al.* (1995) suggest that the protein storage vacuoles derived from the endoplasmatic reticulum are able to receive storage proteins directly from the endoplasmic reticulum as well as from the Golgi complex.

The general model for the assembly and transport of seed storage proteins is illustrated in figure 1.4.

### D) Legumin-like proteins in the seeds of *P. vulgaris*

It has been shown that legumin represents the major storage protein in the seeds of *Glycine*, *Vicia* and *Pisum* (Derbyshire *et al.*, 1976). In soybean, for example, depending on the cultivar, legumin protein accounts for up to about 50% of the total protein (Nielsen, 1984). A similar situation exists in almost every other legume species studied. Legumin

mRNA

RER

NH2

SIGNAL
PEPTIDE

NH2

NH2

NH
2

NH
2

NH2

Removal of signal
peptide

PRE-LEGUMIN

Disulfide bond
formation

PRE-LEGUMIN

S ——————— S

Assembly into the
8 S Trimer

8 S-TRIMER

S ——————— S

Transport to central
vacuole via Golgi
vesicels

8 S-TRIMER

S ——————— S

Proteolytic processing
and assembly into
mature legumin

11 S-HEXAMER

S ——————————— S
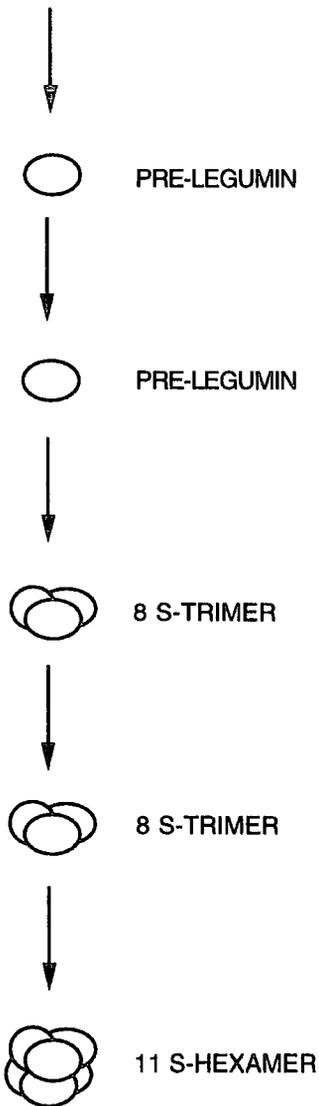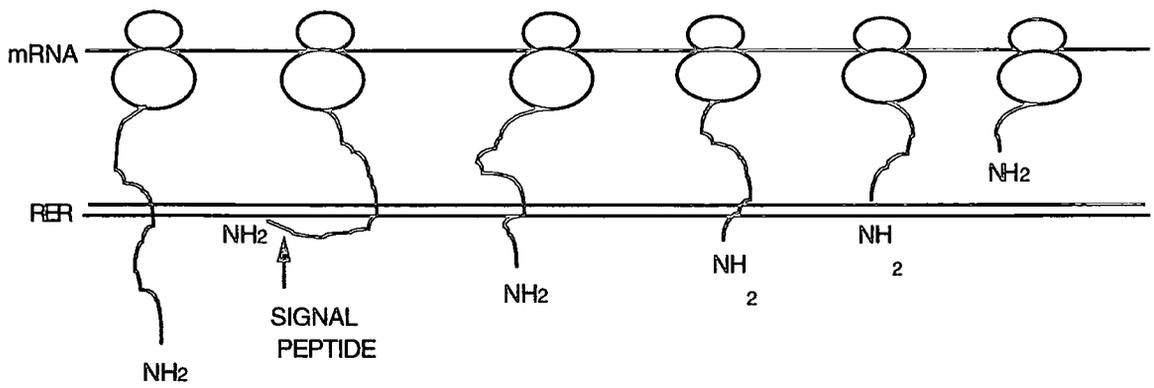
OLIGOMER-STRUCTURE
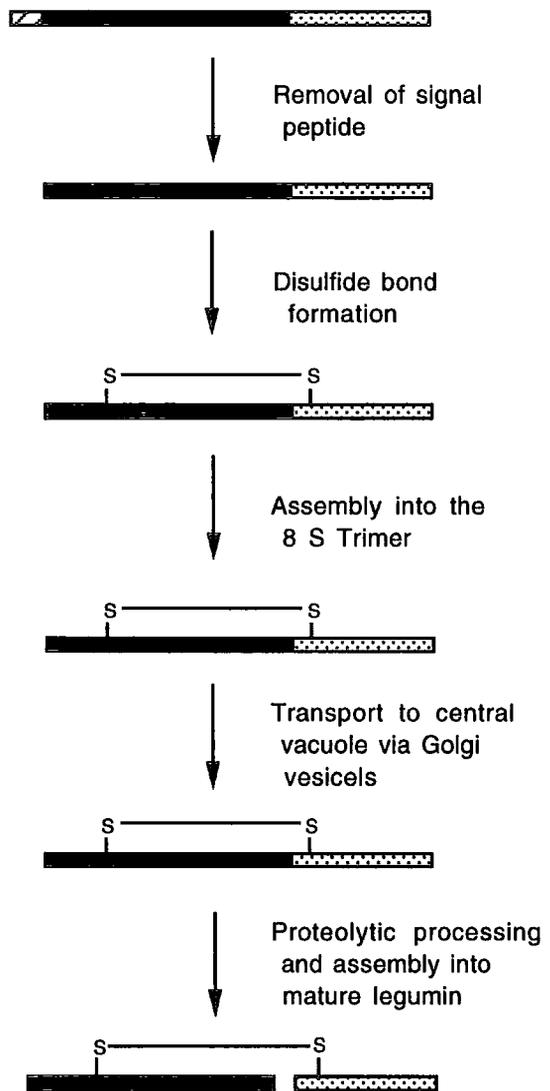
POLYPEPTIDE-STRUCTURE

has also been demonstrated in *Phaseolus aureus* (Derbyshire and Boulter, 1976) which has been later reclassified to *Vigna unguiculata*. However, the principal reserve proteins in the seeds of *P. vulgaris* are i) phaseolin, a 7S globulin which accounts for about 50% of total proteins in the cotyledons (Bollini and Chrispeels, 1978) and ii) phytohemagglutinins or lectins which make up 10% of the cotyledonary proteins (Bollini and Chrispeels, 1978). Based on ultracentrifugation experiments, Danielsson (1949) reported the presence of an 11S globulin fraction from *P. vulgaris*, the same size as the legumin-type proteins. Derbyshire and Boulter (1976) employed zonal isoelectric precipitation to separate out different *P. vulgaris* globulin fractions on the basis of their differential solubility. One globulin fraction of lower solubility than the others and containing a 340 kDa (11S) protein measured by ultracentrifugation, contained polypeptides of $M_r$ 34 kDa, 37 kDa and 21 kDa as it emerged from SDS-PAGE analysis. These size classes were similar to the legumin $\alpha$- and $\beta$-subunits and were further linked by disulphide bonds. However, the authors only succeeded in determining the N-terminal amino acid from each subunit thus failing to fully confirm the identity of the putative legumin protein. Doubt has been cast on the existence of the protein by serological investigations which have shown the absence of immunological cross-reactivity between any *P. vulgaris* seed proteins and legumins from *Vicia faba* (Dudman and Millerd, 1975) or *Pisum sativum* (Kloz and Turkova, 1963). Based on protein extraction from maturing seeds using an ascorbic acid -NaCl buffer (0.25M ascorbic acid; 0.5M NaCl) and ultrcentrifugation, McLeester *et al.* (1973) claimed to have isolated a *P. vulgaris* legumin but this was later shown to be the 7S globulin, vicilin (phaseolin) (Sun and Hall, 1975). Despite numerous other attempts to demonstrate legumin-like seed proteins in *P. vulgaris*, it has largely been an unresolved question as to whether this species and its closely related members actually contain legumin proteins or not. This and the fact, that the final unequivocable proof for the presence of legumin in French bean can only be achieved by providing sequence data, has never been accomplished, have led several authors to suggest that this species does not contain legumin (Miège, 1982, Shewry, 1995).

## 1.3.1.2 Vicilin - The 7S globulin

The second group of globulin storage proteins are the vicilins.

The general structure of vicilins is that of a trimer of three subunits of about 50 kDa, giving a native protein of molecular weight between about 150 to 190 kDa. However, in contrast to the legumins, the single subunits are not held together by covalent disulphide-bonds. This is due to the total lack of cysteines (Casey et al., 1986). However, recently a vicilin-like protein from a cycad (Zamia furfuraceae) has been described (Braun et al., 1996) which contains four cysteine residues per subunit molecule.

The vicilins, like the 11S globulins, are also very heterogenous. The size variations are determined by sequence heterogeneity as well as by differences in proteolysis and glycosylation of the protein. This provides the basis for dividing the four most intensively studied plant species into two groups.

While broad bean and pea vicilins undergo extensive post-translational proteolysis, the main reasons for the heterogeneity in phaseolin, the main storage protein of P. vulgaris, result from minor sequence variations and differences in the covalently-bound carbohydrates (Casey et al., 1986). This and small differences in the amino acid sequence of the single polypeptides explains the size differences in the phaseolin subunits observed on one- and two-dimensional polyacrylamide gels (Brown et al., 1982; Slightom et al., 1985). In contrast to vicilins of other species, there is less variation in the molecular weight of the single subunits which are designated α- (51 - 53,000), β- (47 -50,000) and γ-phaseolin (43 - 47,000) (Hall et al., 1977; Bollini and Vitale, 1981).

Similar to phaseolin, conglycinin, 7S globulin of soybean, does not undergo any apparent post-translational cleavage. Conclycinin can be divided into two forms, the so called β- and γ-conglycinin of which the β-form is quantitatively the major component. As in phaseolin all subunits of the β-conglycinin are glycosylated which leads again to a further degree of heterogeneity (Thanh and Shibasaki, 1976, 1977).

In contrast to soybean and French bean vicilins the heterogeneity in P. sativum and V. faba vicilins is further complicated by post-translational proteolysis. These proteins contain in addition to 50 kDa subunits a number of smaller (Thomson et al., 1978; Gatehouse et al., 1981: Scholz et al., 1983) which arise by proteolytic processing of a precursor of about 50 kDa. For instance in pea, the vicilins are synthesised as groups of 47 and 50 kDa

polypeptides. While some of the 50 kDa polypeptides remain intact, others and all of the 47 kDa polypeptides are processed. The proteolytic cleavage occurs at one or two sites which leads to subunits with $M_r$'s of about 33,000 (called $\alpha+\beta$), 19,000 ($\alpha$), 13,500 ($\beta$) and 12,500 ($\gamma$). Thus, native pea vicilin of about Mr 150,000 may give rise on a SDS-PAGE to a polypeptide size range from 50 kDa to 12.5 kDa (Croy *et al.*, 1980b; Higgins and Spencer, 1981; Shewry, 1995) and more discriminating techniques such as pulse-chase methods had to be used in order to confirm and further elucidate the post-translational processing of vicilin. These methods are particularly necessary in the case of *P. sativum* which contains beside the major vicilin fraction minor amounts of other precursors of about $M_r$ 68 kDa to 70 kDa which are processed *in vivo* to give 50, 18 and 13 kDa polypeptides (Domoney and Casey, 1990).

### 1.3.1.3 Legumins and vicilins derive from a common ancestor

Much research has been devoted during the last decade on the elucidation of the common ancestor of globulin storage proteins. Legumins and vicilins were first thought to derive from a common ancestor by Jackson *et al.* (1969) who's suggestion was based on biochemical properties of the two protein-types. Argos *et al.* (1985) employed DNA and protein sequence alignment and secondary structure predictions to characterise structural similarities between vicilin and legumin. The results of the secondary structure prediction indicated that the C-terminal part of vicilin subunits and the $\beta$-chain of legumins share a relatively high degree of similarity. The N-terminal halves of the vicilin subunits were found to have only a very weak similarity with $\alpha$-chains of legumins. Plietz *et al.* (1987) and Wright *et al.* (1988) demonstrated that the N-and C-terminal regions as well as the $\alpha$- and $\beta$-chains of legumins show a certain degree of similarity.

Later work included structural data obtained for vicilin proteins which provided a new means of information on the evaluation of the evolution of the globulin protein family. Gibbs *et al.* (1989) using the first available crystallographic data for a 7S globulin protein, canavalin A the vicilin from *Canavalia ensiformis*, confirmed the results of Plietz *et al.* (1987) and Wright *et al.* (1989) and showed furthermore that there exists a significant similarity between N- and C-terminal domains of the 7S protein. The sequence similarity was demonstrated to be shared also with the $\beta$-chains of legumins. For the $\alpha$-chains of

legumins, however, no sequence relationship was found. Therefore, the origin of the α-chain remained further unexplained and served as a critical point for the proposed model. More recently, a second three-dimensional structure of a 7S globulin has been described. Lawrence and coworkers (1990, 1994) determined the structure of phaseolin at 3.0 Å and later at 2.2 Å. The new data, set together with the alignment of vicilin-like sequences, led the authors to postulate a canonical structure for the vicilin proteins. Based on these crystallographic data the authors (Lawrence *et al.*, 1994) further demonstrated that the N- and C-terminal domains of vicilin subunits as well as the N-terminal α-chains and the C-terminal β-chains of legumin subunits may have originated from a common, ancestral, structural domain. This model was confirmed and extended by Shutov and coworkers (1995). The authors included in addition to Lawrence and coworker's (1994) data, sequence data from the legumin of *Ginkgo biloba* (Häger *et al.*, 1995) and results from limited proteolytic digestions of both globulin types. The analysis of these data was performed using a quantitatively evaluated cross-wise comparison based on the mutation data matrix of Dayhoff (1978). Based on this analysis the authors suggest the following scenario for the evolution of 7S and 11S globulins: Both globulin types derive from a common single-domain ancestral gene which encoded besides the structural elements, corresponding to the domains of the present globulin subunits, a short variable region (figure 1.5: "V"). Later during evolution of this ancestral gene a triplication event occurred leading after some modifications such as an insertion (figure 1.5: "I1") into the central unit to a putative legumin/vicilin ancestor. The deletion of the N-terminal or C-terminal one-third unit led to the formation of the 11S and 7S precursors, respectively. Whereas the ancestral putative 7S globulin protein resembles very much the modern vicilin-like protein, further modifications occurred at the original 11S globulin, such as the insertion of the α/β-processing site (figure 1.5: "P") and the insertion of the hyper-variable region (figure 1.5: "I2"). All other variable regions found in recent legumin and vicilin subunits derive according to Shutov and coworkers from the V-regions.

V

ancestral gene

Triplication

V        V        V

Modification

I1

V    V         V

11S/7S ancestor

Deletion of
C-terminal domain

V        V        V

7S ancestor

Deletion of N-terminal
domain

V        V

11S ancestor

V        V        V

recent 7S

V        V

recent 11S

△ △
I2 P
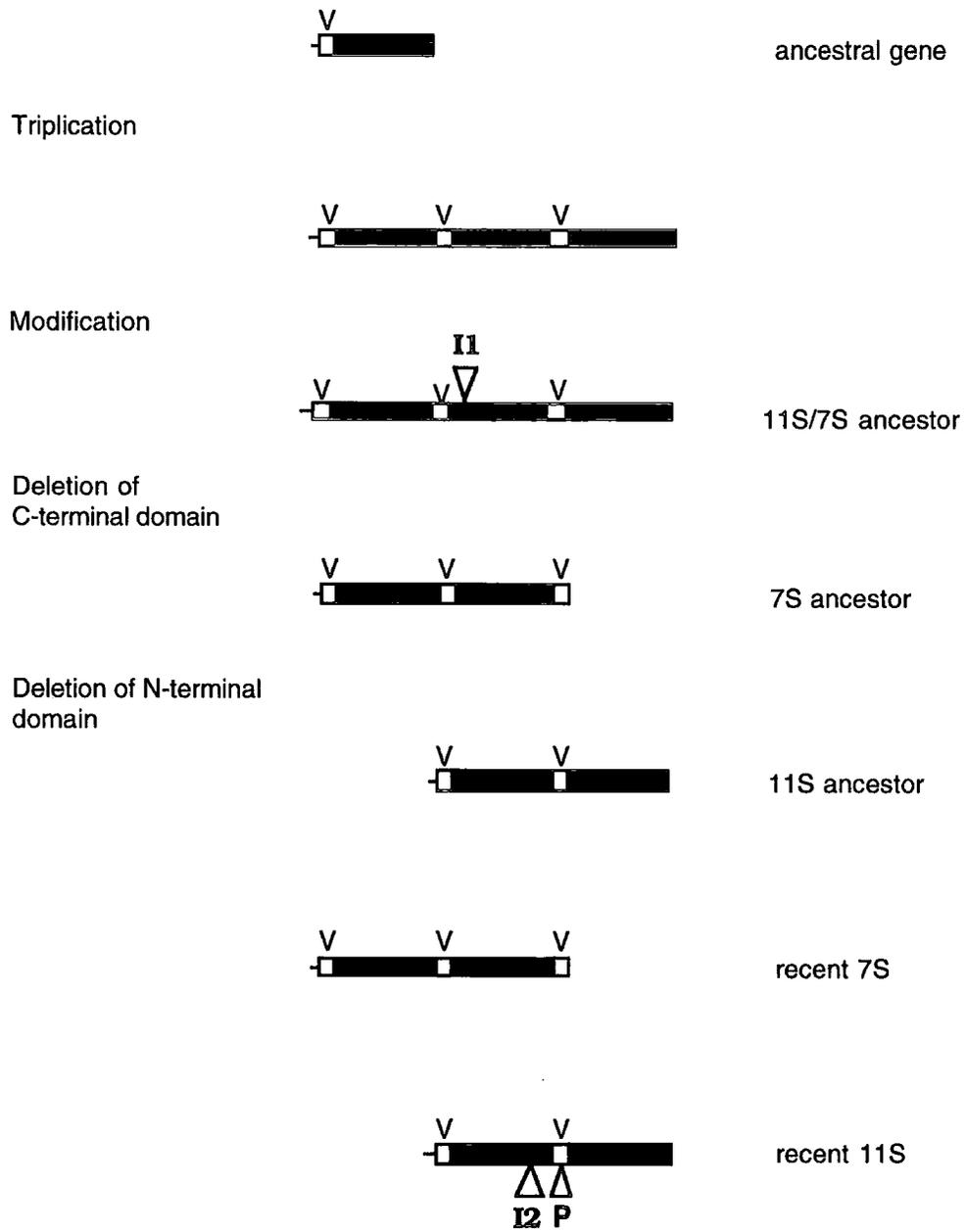
## 1.3.2 Albumins

In contrast to the globulins, the albumins, which have been defined on the basis of their solubility in water as compared with extraction in salt solutions, have not been studied as intensively as the globulins. However, during the last 15 years, the importance of these seed proteins have been recognised, first mainly because of their significant proportion of total seed proteins. For example, the albumin content in sunflower seeds has been estimated at more than 20% (Kortt and Caldwell, 1990; Youle and Huang, 1981) of the total seed proteins. This proportion refers only to the 2S albumins, which are defined as a discrete group by their sedimentation coefficient (Youle and Huang, 1981). Later, more advantageous characteristics of the 2S albumins were found and research has concentrated on this group because of two reasons:

### 1.3.2.1 Nutritional quality of seed proteins from the 2S albumin fraction

The 2S albumins have proved to be a rich source for sulphur-containing amino acids. Altenbach *et al.* (1987) reported that the sulphur content of a certain 2S albumin in the seeds of Brazil nut (*Bertholletia excelsa*) is unusually high with 8% of the codons in its gene code for cysteine and 18% specify methionine. As the main grain legume seed storage proteins, the globulins are generally poor in sulphur-containing amino acids, the 2S albumins have been identified for potential improvement of the sulphur content in the seeds of crop plants by introducing the gene encoding the 2S albumin into the genome under the control of an efficient promotor (Altenbach *et al.*, 1989, 1992; Saalbach *et al.*, 1994). For example, the accumulation of the methionine-rich 2S protein of Brazil nut in the seeds of tobacco resulted in a significant increase (30%) in the level of methionine in the seeds of the transgenic tobacco (Altenbach *et al.,* 1989) and transformation of *V. narbonensis* with the same sulphur-rich 2S gene of Brazil nut gave a 3fold increase in the methionine content of the seeds of the bean (Saalbach *et al.,* 1995a). These successes confirm the 2S albumins as promising candidates for introduction and expression in economically important crop plants. Attempts to genetically engineer the sequence of legumin or vicilin to increase the sulphur content of these proteins by addition of methionine residues into the protein, have either stopped at the level of sites proposed to insert the additional amino acids (e.g. Dyer *et al.*, 1993; Ko *et al.*, 1993; Lawrence *et al.*,

1994; Utsumi, 1992) or led mostly to a failure in the accumulation and deposition of the engineered protein (Saalbach *et al.*, 1988; 1995b). Only one successful attempt to produce a methionine-enriched vicilin has been reported (Saalbach *et al.*, 1995b). The problem with the genetic engineering of seed globulins is the fact that the tertiary structure of the polypeptides plays a very important role in processing, assembly, transport and deposition of the mature proteins in the central vacuole. To date, all of the attempts to change the structure of the storage proteins have lead to unstable proteins (Saalbach *et al.*, 1988; 1995a,b) and no promising results for food improvement in industrial standard have yet been obtained. Therefore, most efforts seem to be devoted to producing transgenic crop plants with a stable, elevated expression of genes encoding sulphur-rich protein (e.g. Altenbach *et al.*, 1989; 1992; Guerche *et al.*, 1990).

## 1.3.2.2 Biological activity

Whether the major role of the 2S albumins is clearly as storage proteins is still unknown. The fact is, that several different enzymatic active proteins are present in the 2S albumin fraction. These enzymes and other albumin proteins seem to be important for the seeds in order to protect themselves from pests and pathogens. Examples of this class of proteins are the $\alpha$-amylase and trypsin inhibitors of cereals (Kreis *et al.*, 1985 a, b). Svendsen and coworkers (1989) showed that the 2S albumins of kohlrabi (*Brassica napus*, var. *rapifera*) inhibited the enzymatic activity of bovine trypsin. $\alpha$-amylase and trypsin inhibitors decrease the digestibility of seeds by insects and vertebrates, so protecting the plant against damage through pests.

Other useful enzymes inhibit the growth of pathogenic fungi and some bacteria in seeds. Examples for these enzymes are chitinases (Schlumbaum *et al.*, 1986), beta-1, 3 - glucanases (Manners and Marshall, 1973), thionins (Bohlmann *et al.*, 1988), permatins (Roberts and Selitrennikoff, 1990). The chitinases are the most characterized enzymes and have been shown to exhibit high defence activity as their major natural role, primarily against fungal pathogens (Boller, 1988).

These enzymatic and inhibitory activities make the 2S albumins a promising class of protective agents and much work has been devoted during the last decade to this former under-studied group of proteins. Furthermore, a large amount of still unknown proteins can

be expected to be found by further investigations and to be identified as plant protective compounds. Investigations of certain albumin proteins, e.g. produced as a response reaction to pathogenic or pest attack, is useful in order to detect protective enzymes.

A further promising approach might be, for instance, the investigation of albumin proteins synthesised *de novo* at the start of seed germination. Boulter and Barber (1963) reported that the albumin fraction in bean seeds increased with age and the amino-acid composition of 6-day-old cotyledonous albumins differed from that of ungerminated beans. Becker *et al.* (1995) and Qi *et al.* (1992) isolated proteases from the cotyledons of germinating vetch and soybean seeds respectively. Protease activity has been shown for a long time to be responsible for the specific, endoproteolytic degradation of the reserve proteins (Chrispeels and Boulter, 1975; Chrispeels *et al.*, 1976; Baumgartner and Chrispeels, 1976). Knowledge of seed proteases and their genes will lead to a better understanding of their function and may help to improve the digestibility and the nutritional value of crop plants.

Modern research is now concentrating on the structural function of several of these proteins, including an evaluation of the 3D structure of the molecule (e.g. of chitinases: Hart *et al.*, 1993; 1995), so permitting an insight in the structure/function relationship of these proteins which may also have implications for plant engineering (Hart *et al.*, 1993).

## 1.4 Objectives and aims of the research project

Cultivated plants are the primary food source for human and livestock. Furthermore, cultivation of many crops provides an increasingly valuable source of raw material for many industrial applications. Therefore, a detailed understanding and broad knowledge of seed storage compounds and their biosynthesis, accumulation, storage and breakdown is necessary in order to find new ways of increasing their nutritional or industrial quality. On the other hand, it is equally important to protect the seeds and the plants bearing those seeds from fungal pathogens and other pests as well as from damage caused by environmental factors (abiotic stress). This is possible by either the use of herbicides, pesticides in combination with the breeding of more resistant crop plants or by developing genetically engineered plants. On a long-term basis, the latter is the more promising and successful avenue as it will enable at least the decrease of the expenditures for chemical protection and for environmental damage. Furthermore, it provides a faster way of plant

improvement as well as it offers a wider range of changes towards increased resistance against both environmental factors and pests.

The summary of the literature about seed storage proteins has shown that especially in the last two decades much work has been done on seed storage proteins and *P. vulgaris* has often been used as a model plant since French bean represents an important crop plant. However, the work has always been concentrated on certain proteins or group of proteins. So, many seed proteins, especially those, which occur only in minor amounts, are still unidentified, but could play an important role in the (nutritional) improvement, function or protection of seeds. The present work has been therefore undertaken in order to contribute to an evaluation of some of the unknown seed proteins in the seeds from *P. vulgaris*.

This work has involved studies on the two major groups of seed proteins, the albumins and globulins.

Albumins: Many different protein types are in this fraction which is why individual albumin proteins appear mainly only in small quantities. In order to obtain more information about the nature and role of these proteins, albumin proteins should be fractionated, purified and identification attempted by N-terminal sequencing ("albumin protein mapping") and enzyme assays.

Glycoprotein I, a glycosylated albumin protein originally described by Pusztai (1966), should be isolated. Pusztai described two glycoproteins, designated as Glycoprotein I (Pusztai, 1966) and Glycoprotein II (Pusztai and Watt, 1970), from mature seeds of *P. vulgaris*. Whereas Glycoprotein II (phaseolin) represents a globulin and the major storage protein (about 50% of total seed proteins; Bollini and Chrispeels, 1978) and most of the research has been put on this protein (see section 1.3.1.2), Glycoprotein I appears only as a minor amount (1-2% of total seed proteins; Pusztai, 1966) in the albumin fraction and no more research has been since then devoted on this protein, although it was found to be an interesting protein-protease with autolytic activity. Therefore, it was aimed to attempt the identification of this protein by N-terminal sequencing of the purified protein as, apart from these initial studies, no identity or detailed information is known about this protein.

Another area of investigation concerns the globulin fraction. The globulins - unlike the albumins - have been studied in some detail and occur in large amounts in the seeds of higher plants. In comparison with the albumin fraction the globulin fraction contains

relatively few types of proteins of which vicilins (phaseolins) represent the major part in French bean. However, legumin, a highly abundant seed storage protein in all other members of the legume family has not yet be unequivocally identified in *P. vulgaris*. Legumin has been found to be expressed in the seeds of nearly all other investigated dicots (e.g. Fischer *et al.*, 1995) and several monocots (e.g. Fischer *et al.*, 1996) including several cereals (Singh *et al.*, 1985, Takwaia *et al.*, 1987; Walburg *et al.*, 1986), in the seeds of many coniferous trees (Wind and Häger, 1996; Häger *et al.*, 1996) as well as in species of the chlamydosperms (Braun and Bäumlein, unpublished results/quoted in Braun *et al.*, 1996) and even in ancient plants such as the "living fossil" *Ginkgo biloba* (Häger et al., 1995). Therefore, it is not logical that legumin should be absent from the seeds of *P. vulgaris* despite the claims of other workers (Miège, 1982; Shewry, 1995) that legumin does not occur in the genus *Phaseolus*.

In order to solve this question, a major objective of this work was to attempt to identify unequivocally legumin-like proteins in the seeds of *P. vulgaris*. This is only possible by providing (N-terminal) protein sequences of the two characteristically disulfide-linked polypeptides or cDNA sequences. Further it was intended to try to isolate and sequence legumin encoding PCR-fragments amplified from genomic DNA. The success of this approach would provide evidence in the case the legumin protein could not be identified that at least a gene encoding a legumin was present in the genome of *P. vulgaris*. This would then offer an answer to the question about the presence or absence of legumin in the seeds of French bean: either legumin encoding genes are present only in form of non-translated and not functional genes (pseudogenes), or legumin is present in the seeds of *P. vulgaris* in then obviously only hardly detectable amounts.

A last aim of the work concerning legumin in the seeds of French bean was the use of molecular data for research in the plant taxonomy.

Seed proteins have been shown to be a reliable marker (Vaughan and Denford, 1968). Especially the seed storage protein and of these legumin has been proven by serological studies to exhibit systematic significance (e.g. Jensen and Greven, 1984). Therefore, using adequate mathematical algorithms the comparison of DNA sequence data of legumin encoding genes of *P. vulgaris* with those of other members of the legume family could provide considerable contribution to the further elucidation of the taxonomic position of the genus *Phaseolus* within the legume family.

# 2. Materials

## 2.1 Chemicals, solvents, kits and laboratory consumables

Most of the chemicals and solvents used in this research project were supplied by BDH Chemicals Ltd., Poole, Dorset, U.K., and SIGMA Chemical Co., Poole, Dorset, U.K., and were of the purest grade available.

Other reagents and consumable suppliers are as listed below:

3 MM chromatography paper, DEAE Cellulose: Whatman Ltd. (Maidstone, Kent, U.K.).

Hybond-N$^+$ Hybridisation membrane, ECL Detection Kit: Amersham International plc. (Little Chalfont, Buckinghamshire, U.K.).

Electrophoresis grade agarose, DIG Nucleotide Detection Kit, deoxy-nucleotides: Boehringer, Mannheim (UK) Ltd. (Lewes, U.K).

Restriction enzymes, DNA modification enzymes, IPTG, X-Gal: Northumbria Biologicals Ltd. (Cramlington, County Durham) or Boehringer Mannheim (UK) Ltd. (Lewes, U.K.).

ProBlot PVDF transfer membrane, T-vector system I, Wizard Minipreps DNA Purification System, Taq-DNA polymerase: Promega (UK) Ltd. (Southampton).

QIAquick Gel Extraction Kit: Qiagen Ltd. (Surrey, U.K.).

Bacto yeast extract, bacteriological peptone, Freunds incomplete adjuvants: DIFCO Laboratories (Detroit, Michigan U.S.A.).

Agar (bacteriological): Unipath Ltd. (Basingstoke, Hampshire, U.K.).

Antibodies raised in mice against ethylene-induced French bean leaf chitinases were a gift from Dr. Angharad Gatehouse and Dr. Lawrence Gatehouse.

Antibodies raised in rabbits against *Phaseolus vulgaris* vicilins were a gift from Dr. Ronald R. D. Croy.

Affinity-purified antibodies raised in mice against legumins from *Pisum sativum* were received as a gift from Dr. John A. Gatehouse.

Antibodies raised in mice against germins from wheat were received as gift from C. Ilett.

The legumin A-encoding genomic clone pDUB 24 (accession-no. X02982; Lycett *et al.*, 1984) was received as gift from Dr. R. R. D. Croy.

CM-chitin-RBV, the dye-labelled substrate for the detection of chitinase activity, was received as a gift from Dr. Ofelia Gozia.

Fuji RX-100 X-ray film: Fuji Photo Film Co., Ltd. (U.K.).

Dialysis tubing: Medicell International Ltd. (London, U.K.).

Paraffin wax tablets: BDH Chemicals Ltd. ( Poole, Dorset, U.K.).

Polaroid film: Polaroid (UK) Ltd. (St. Albans, Hertfordshire, U.K.).

Sephacryl S-300 Superfine: Pharmacia Fine Chemicals (Uppsala, Sweden).

Marvel (semidry milk powder): Safeways (Durham, U.K.).

Mature seeds of Phaseolus vulgaris, variety "Processor": Hurst, Gunson, Cooper & Taber
Ltd. (Witham, Essex, U.K.).


## 2.2 Bacterial strains

All bacterial strains used in this work were from laboratory stocks and were stored as
glycerol cultures at -80 °C.

*E.coli*, strain DH 5α, was used as a bacterial host for the growth of pKS$^+$ vector DNA and
for transformation experiments.

| Strain | Genotype | Reference |
|--------|----------|-----------|
| DH 5α | *supE*44 ) *lac* U169 (N80 *lacZ*) M15) *hsdR*17 *rec*A1 *end*A1 *gyr*A96thi-1 *rel*A1 | Sambrook *et al.*, 1989 |

Further, *E. coli,* strain XLBlue-1 was used as bacterial host for transformation
experiments.

| Strain | Genotype | Reference |
|--------|----------|-----------|
| XLBlue-1 | *supE*44 *hsdR*17 *rec*A1 *end*A1 *gyr*A46 *thi rel*A1 *lac⁻ F' [proAB⁺ lac^q lacZ*) M15 *TN*10 (tet$^r$)] | Sambrook *et al.*, 1989 |

# 3. Methods

## 3.1 Germination of Seeds

Seeds were treated with a 1% aqueous solution of commercial bleach for 10 minutes on a shaker followed by extensive washing with water with three changes (each 1 hour). The seeds were placed in an autoclaved metal tray with two layers of Whatman paper. The seeds were kept wet with autoclaved water over the period of germination. After each day of imbibition samples of seeds were collected and stopped from germinating by freezing in liquid nitrogen and storing at -70 °C. Any infected seeds were discarded.

## 3.2 Protein work

To avoid degradation or enzymatical breakdown of proteins it was tried to keep the plant material and protein solutions at 4 °C at every stage of investigation.

### 3.2.1 Protein extraction

Seed proteins were extracted basically according to Osborne's classification scheme (Osborne, 1924) which was based on the solubility properties of seed storage proteins. Modifications were performed according to Croy (pers. communication).

#### 3.2.1.1 Albumin proteins

Seeds were ground to a fine powder using an analytical precision mill (Janke & Kunkel, Germany). All the following steps were performed at 4°C using pre-cooled equipment. The flour was transferred to a mortar and ground to very fine powder under liquid nitrogen using a pestle. The powder was then transferred into a second mortar (4 °C) and six volumes (v/w) of cold acetate buffer (0.033M NaOAc, pH 5.0; 4 °C; Pusztai, 1966) were added. The solution was carefully stirred every 10 minutes over a period of 1.5 h using a pestle. The mixture was then centrifuged (20,000g x 30 min) in a Beckman centrifuge (J2-HC with a JA-20 rotor) to remove all solid particles and the clear supernatant was dialysed for 24 h at 4 °C against 0.033M acetate buffer, pH 5.0. Any precipitated proteins (globulins) were removed by centrifugation and the supernatant was used directly for the next purification step (ammonium sulphate precipitation; see section 3.2.2.1).

### 3.2.1.2 Globulin proteins

A total protein extract was prepared using a TGB extraction buffer (10mM Tris/HCl, pH 8.2; 80 mM glycine; 0.5M NaCl) as described by Jensen and Berthold (1989). The ratio of seed material to buffer was 1:4 (w/v). The extraction with this buffer was performed as described previously for the albumin extraction (section 3.2.1.1).

Globulin proteins of this TGB extract were precipitated from other proteins by extensive dialysis for 48 h against 0.033M acetate buffer, pH 5.0 (Pusztai and Watt, 1970). The precipitated globulins were recovered by centrifugation as before, briefly washed in acetate buffer and then dialysed against distilled water prior to freeze drying.


### 3.2.2 Methods to concentrate protein solutions

### 3.2.2.1 Ammonium sulphate precipitation

In order to increase the protein concentration after protein extraction from plant material, ammonium sulphate precipitation was selected as a mild, non-denaturing method. The addition of ions at high concentration neutralise the charges on protein molecules which leads to aggregation and precipitation of the proteins from solutions. Even at high concentrations, ammonium sulphate does not alter the pH of a protein solution and is therefore a suitable ionic compound. Further, this method allows also a crude fractionation of total protein extracts by stepwise salting in of the protein solution and separation of the precipitated proteins from the supernatant.

All ammonium precipitations were performed at 0-4 °C. The exact amount of solid ammonium sulphate required to bring a solution of known volume and initial saturation to a desired final saturation was selected from a chart (Dawson et al., 1989). The protein solution was stirred slowly with a magnetic stirrer and the ammonium sulphate was added very slowly in small amounts, allowing each aliquot to dissolve before adding more, thus avoiding high localised concentration of the salt.


### 3.2.2.2 Acetone precipitation

Acetone precipitation was found to be a convenient method for concentrating proteins from very diluted protein samples prior to analysis on SDS gels. The method, however, is

not suitable for the increase of the protein concentration for following (N-terminal) sequencing since it promotes N-terminal blockage (Gilroy, pers. communication).

Protein solutions were thoroughly mixed with 4 volumes of cold (-20 °C) acetone and then incubated at -20 °C for at least 2 h. The solution was centrifuged in a microcentrifuge tube (10,000g x 30min) and the acetone decanted. The pellet was then carefully dried under nitrogen, but was not allowed to dry completely. The still opalescent pellet was resuspended in the minimum possible volume of buffer.

### 3.2.3 Column chromatography

All chromatography steps were performed at 4 °C. Pre-swollen chromatography media (Sephacryl S-300, etc.) were degassed under vacuum prior to column packaging, an excess of buffer was decanted and the fairly thick suspension was then poured down the inside wall of the glass column 30% filled with equilibration buffer. Care was taken to avoid any air bubbles. After the gel matrix has settled down by a flow rate of about 0.2 ml/min, the column was closed with a plunger avoiding air bubbles on top of the gel bed. The plunger was put on top of the gel bed. The ready-packed column was then connected to a peristaltic pump, a UV-monitor linked to a chart recorder and a fraction collector.

### 3.2.3.1 Gel filtration chromatography

Gel filtration chromatography was performed using a Sephacryl S-300 (Pharmacia, Uppsala) column matrix (2.5cm x 110cm) equilibrated with a Tris-buffer (0.1 M Tris/HCl, pH 8.0; 0.1 M NaCl). The chromatography was performed in the same buffer using a flow rate of 0.3 ml/min, the eluat continuously monitored for protein by absorbance at 280 nm and fractions were collected every 10 min. Protein containing fractions were analysed using SDS-PAGE and/or enzyme assays (chitinase activity test).

Gel filtration under these conditions lasted about 15 hours.

### 3.2.3.2 Ion - exchange chromatography

DEAE (DE-52 or Express-Ion-Exchanger)-Cellulose (Whatman) was used as the matrix for ion-exchange chromatography. A column was filled with the pre-swollen matrix and equilibrated with a Tris-buffer ( 0.025 M Tris/HCl, pH 7.6). The final gel bed size was 2cm x 5cm. (Freeze dried) protein samples were dissolved in the equilibration buffer.

Protein solutions and ammonium sulphate precipitated protein samples were dialysed against the equilibration buffer prior to the chromatography. The protein solution was then applied onto the column. The chromatography was performed using a flow rate of 1ml/min and the eluat was monitored as described previously (section 3.2.3.1). After washing the column matrix with equilibration buffer to elute proteins which did not bind to the column matrix a linear NaCl gradient from 0M to 0.5M NaCl concentration in the starting buffer was applied in order to elute proteins according to their charge. Eluted proteins were measured and analysed as described previously (section 3.2.3.1).

### 3.2.4 Protein analysis

### 3.2.4.1 Electrophoretic analysis

SDS-PAGE was performed on a Bio-Rad "Mini-Protean II" electrophoresis apparatus (Biorad).

SDS polyacrylamide gels were prepared using a stock solution of 30% (w/v) acrylamide and 0.8% (w/v) bisacrylamide (Bio-Rad). A 4% (w/v) polyacrylamide stacking gel of 2 mm in length was poured on top of the separating gel.

The selection of the polyacrylamide concentration depended on the size range of the proteins to be analysed on the gel. The choice was based on the following table after Sambrook *et al.* (1989):

| [a]Acrylamide concentration (%) | Linear range of separation (kDa) |
| --- | --- |
| 15 | 12 - 43 |
| 10 | 16 - 68 |
| 7.5 | 36 - 94 |
| 5.0 | 57 - 212 |

[a] Molar ratio of bisacrylamide to acrylamide is 1:29.

### SDS-Polyacrylamide Gel Electrophorese (SDS-PAGE)

Polyacrylamide gel electrophoresis of proteins was performed essentially according to Laemmli (1970) on 0.5 mm thick polyacrylamide gels. Protein samples were dissolved in a sample loading buffer with (reducing conditions) or without (non-reducing conditions)

addition of 2% (v/v) 2-mercaptoethanol. Prior to the electrophoresis, protein samples (for both reducing and non-reducing conditions) were incubated at 100 °C for 5 minutes. Between non-reduced and reduced protein samples a lane on the polyacrylamide gel was left free to ensure that diffusing 2-mercaptoethanol did not reduce disulphide-linked polypeptides. Electrophoresis was performed at 50 V till the bromophenol blue band entered the separating gel. Then, the voltage was increased to 130 V. Depending on the polyacrylamide concentration, the electrophoresis was stopped when the bromophenol blue band had reached the bottom (10% polyacrylamide) or just run out of the gel (12%) or 20-30 minutes after the bromophenol blue band had run out of the gel (15%). After electrophoresis the gels were stained for all proteins by staining with 0.1% (w/v) coomassie blue R250 in 40% (v/v) methanol, 10% (v/v) acetic acid followed by destaining in the same solvent or by silver-staining (using a silver-stain kit (Biorad)). For selective staining of proteins with attached polysaccharides Schiff's staining based on the method of Fairbanks *et al.* (1971) was performed by incubation of the polyacrylamide gel in 7.5% acetic acid for 1 hour at room temperature followed by incubation in 0.2% periodic acid at 4 °C for 45 minutes. The gel was then placed immediately in Schiff's reagent and refrigerated at 4 °C for 45 minutes followed by destaining at room temperature in 10% acetic acid with several changes over a period of about 3 hours. Glycosylated proteins appeared as red protein bands.


## 2-dimensional SDS-PAGE

For 2-dimensional SDS-PAGE both dimensions were performed in vertical minigels (Biorad) using 12% (w/v) polyacrylamide gels. After the first dimension (non-reducing) the gel was cut into vertical sections, each containing a lane with the separated proteins. The gel slices were incubated in 50 mM dithiothreitol at 40 °C (water bath) for 45 min and a single gel slice was positioned horizontally on top of a 0.75 mm thick gel without a stacking gel. After the second dimension run the gel was silver stained.

### 3.2.4.2 Analysis of enzymatic chitinase activity

Chitinase activity was determined essentially by the method of Wirth and Wolf (1990).

A carboxymethyl-substituted soluble chitin and a reprecipitated colloidal chitin derivative were covalently linked with Remazol Brilliant Violet 5R (courtesy from Dr. Ofelia Gozia) to produce a suitable colorimetric substrate (CM-chitin-RBV) for the assay and detection of chitinase and lysozyme (Wirth and Wolf, 1990).

The assays were performed in 1.5 ml Eppendorf tubes. The enzyme solution or the freeze dried enzyme sample was dissolved in 100-200 µl sodium acetate buffer (0.2M, pH 5.0) giving a final volume of up to 200 µl. 100 µl of CM-chitin-RBV (2mg/ml in sodium acetate buffer (0.2M, pH 5.0)) were added and incubated at room temperature for periods ranging from 1-3 h depending on the activity of the enzyme in the extract under study. The reaction was terminated by the addition of 0.1 ml HCl (1.0M), causing the non-degraded substrate to precipitate. The reaction tubes were centrifuged in a microcentrifuge (10,000g x 5 min) and the absorbance of the supernatant was measured photometrically at 550 nm against a blank prepared similarily but without the addition of enzyme solution. The so performed enzymatic activity test was qualitative.

### 3.2.4.3 Immunological methods

__Raising Antibodies against Glycoprotein I__

Polyclonal antibodies against Glycoprotein I were raised in mice. The protein (300 µg/ml) was dissolved in PBS-buffer (1x), filtered (0.22µ filter) and emulsified with one volume of Freunds complete adjuvant. The adjuvant was added to the protein solution in 30 µl aliquots with vigorous mixing on a vortex mixer. The protein solution was then sonicated in a sonication water bath at 4 °C twice for 5-10 seconds. Prior to the injection of the protein solution into mice, blood samples for pre-immuneantiserum were taken from the animals. The protein solution was injected intraperitoneally into the mice.

After four weeks a second injection followed using a protein solution prepared as before but using Freunds incomplete adjuvant and after two more weeks blood samples were taken to prepare antisera. This was done by incubation of the blood for 2 hours at 37 °C

31

followed by incubation at 4 °C overnight. Finally the supernatant (antiserum) was decanted and NaN$_3$ was added to a final concentration of 0.02% (w/v) in order to prevent fungal or bacterial growth. The optimal antibody titre for immunodetection experiments was analysed by immunoblotting of Glycoprotein I samples on nitrocellulose membrane (see below) with different concentrations of the antiserum (e.g. 1:20,000; 1:10,000; 1:5,000; 1:2.500).

## Immunoblotting

Protein samples were separated by SDS-PAGE as described previously (section 2.2.2.4A). The separated polypeptides were then blotted onto a nitrocellulose membrane (Schuller & Schleich) by the semidry method using a SEMI-DRY Blotter II from Whatman International Ltd. (Maidstone, England). 9 pieces of Whatman 3MM paper cut to the same size as the polyacrylamide gel were soaked in the transfer buffer and placed on the anodic (lower) plate. An NC filter was cut to the same size as the gel, soaked in transfer buffer (appendix A) and placed on the 9 layers of filter paper. The polyacrylamide gel was then placed onto the NC membrane and all air bubbles between the gel and the membrane were removed. Finally, a further 9 pieces of Whatman 3MM paper soaked in transfer buffer were layed on top of the gel, the cathodic plate placed on top of the trans-unit and the proteins were blotted at 200 V for 1.5 h.

After the blotting, the filter was briefly rinsed in distilled water and proteins were stained using a Ponceau S (0.1% (w/v) Ponceau S in 1% (v/v) acetic acid) staining solution for two minutes. Destaining was performed in distilled water. The NC-membranes were either used immediately or stored at -20°C for up to 2 weeks.

## Dot-Blotting

To investigate whether anti-germin antibodies show binding to any of the *P. vulgaris* or *Pisum sativum* seed proteins, dot blotting was performed to blot the protein samples onto nitrocellulose membranes. Using this method, it was possible to check the antibodies used in a certain concentration against a large range of different protein concentrations.

A total protein extract was achieved using 8 volumes (v/w) of a TGB extraction buffer (10mM Tris/HCl, pH 8.2; 80 mM glycine; 0.5M NaCl). The mixture was centrifuged

(20,000g x 30 min) in a Beckman centrifuge (F3602 rotor) to remove all solid particles and the clear supernatant was used directly for the dot blots. The protein solutions were then diluted with water (e.g. 1:1; 1:2.5; 1:5; 1:7.5; 1:10; ...) and 150 µl of the protein samples were blotted onto a nitrocellulose membrane using a HYBAID Processing Pump. The antisera were used in appropriate concentration.

Dot blotting as described above was further used to check whether the pre-immuneantiserum contain antibodies which react with any of the seed proteins under investigation in a control experiment. None of these control experiments showed a positive immune reaction.

Immunodetection was carried out using an ECL Detection Kit from Amersham Int. plc (England) performed in a modified procedure. The NC-membrane was soaked in 2% (w/v) periodic acid for 10 min at room temperature and subsequently blocked in blocking buffer overnight at 4°C with gentle agitation. The following steps were then performed at room temperature. After blocking, the membrane was incubated with the primary antibody in antisera buffer for 2-3 h. The dilution of antiserum depended on the antibody titer of the antisera used. The filter was washed 15 minutes in antisera buffer (appendix A) with two changes and the filter was the incubated in a 1:5,000 dilution of the secondary (anti-rabbit or anti-mouse, corresponding to the species used for raising the primary antibodies) antibodies conjugated with horseradish peroxidase. After 2 h the membrane was washed in PBS/0.1% (v/v) Tween 20 for 30 minutes with 2-3 changes. In order to remove any traps of Tween 20 which inhibits the horseradish peroxidase activity, the filter was finally washed in distilled water for 10 minutes.

The following steps were performed in the darkness. Equal volumes of detection reagents 1 and 2 of the ECL detection system were mixed. The wet filter was drained and the mixed detection solution was poured over the membrane. After one minute the detection solution was drained from the filter, the filter wrapped in clingfilm and placed onto a X-ray film (Fuji). After 20 seconds exposure, the membrane was placed onto a second X-ray film which was exposed for two minutes to the membrane. The X-ray films were immediately developed after exposure by successive incubation in developing solution (Ilford, Ltd.) for 5 minutes, water for 2 minutes and fixation solution (Kodak, Ltd.) for 4 minutes.

## 3.2.4.4 Automated N-terminal sequencing

Protein samples were separated by SDS-PAGE. After the electrophoresis proteins were blotted onto a polyvinylidenedifluoride (PVDF) membrane (Promega) by the semidry blotting method as described previously (section 3.2.4.3) with a CAPS-buffer (10 mM 3-[cyclohexylamino]-1-propanesulfic acid (CAPS, Sigma), pH 11; 10% methanol) as transfer buffer. The blotted proteins were stained for 1 minute with coomassie blue (0.1% coomassie blue R250 (w/v) in 40% MeOH / 1% (v/v) acetic acid) and the membrane destained with 50% (v/v) aqueous methanol solution. The N-terminal amino acid sequence analyses were performed on a Model 477A pulsed-liquid microsequencer (Applied Biosystems) equipped with an on-line phenylthiohydantoin analyzer, Model 120A (Applied Biosystems) essentially according to Hunkapiller and Hood (1983). Cysteine was identified as Pht-Cys S-propionamide formed during the initial SDS-PAGE by reaction of residual unpolymerized acrylamide in the gel with thiol-groups of cysteine residues (Brune, 1992).

## 3.2.5 Isolation of Glycoprotein I

Glycoprotein I was isolated basically as described by Pusztai (1966).

Seeds were ground as described previously (section 3.2.1.1). The following steps were carried out at 4 °C with pre-cooled materials. A albumin extract was prepared using a 33mM sodium acetate buffer, pH 5.0 (ratio seed meal : buffer 1:8 (w/v)). The protein extract was dialysed against 10 L of extraction buffer at 4 °C for 24 h in order to precipitate globulin proteins. The protein solution was centrifuged in a Beckman centrifuge (20,000g x 20min) and the supernatant (albumin proteins) was decanted. The albumin proteins were then partially fractionated by 60% (rel. sat.) ammonium sulphate, the precipitate recovered by centrifugation (20,000g x 20min) and the pellet redissolved in a Tris-buffer (0.1M Tris/HCl, pH 8.0; 0.1 MNaCl). The solution was dialysed against 10 L of the same Tris-buffer for 24 h and any precipitate removed by centrifugation. 5 ml of the supernatant was applied onto a Sephacryl S-300 (Pharmacia) column (2.5cm x 110cm) and gel filtration carried out as described previously (section 3.2.3.1). The protein fractions were analysed by SDS-PAGE. Glycoprotein I - containing fractions were pooled and dialysed against 8 L of 0.025 M Tris/HCl buffer, pH 8.0 with one change for 36 h. 3 ml of

the protein solution was then applied onto a DEAE-cellulose column (2cm x 5cm) which was run with a salt gradient from 0 - 0.5M NaCl in the elution buffer as described previously (section 3.2.3.2). Glycoprotein I - containing fractions were pooled, dialysed against distilled water and freeze dried. For identification of the protein by N-terminal sequencing the freeze dried protein sample was dissolved in a 2-mercaptoethanol (2% v/v) containing sample buffer, proteins were separated on a 17.5% (w/v) polyacrylamide gel, transferred to a PVDF membrane and sequenced as described before (section 3.2.4.4).

### 3.2.6 Partial purification of legumin

Seeds were ground to a fine powder using an analytical precision mill (Janke & Kunkel, Germany). All following steps were performed at 4°C. A crude total protein extract was prepared using an extraction buffer (10mM Tris/HCl, pH 8.2; 80 mM glycine; 0.5M NaCl) as described by Jensen and Berthold (1989). The ratio of seed material to buffer was 1:4 (w/v). The mixture was centrifuged (20,000g x 30 min) and the supernatant dialysed against 8 L of buffer (0.1M Tris/HCl, pH 8.0; 0.1M NaCl) for 48 hours with one change. The protein extract was then applied onto a Sephacryl S-300 (Pharmacia Fine Chemicals, Uppsala) column (2.5 x 110 cm) and gel filtration was performed as described previously (section 3.2.3.1). After electrophoretic analysis (SDS-PAGE under reducing and non-reducing conditions; 2D SDS-PAGE) of the fractions obtained, the fractions containing legumin-like (disulphide-linked) proteins were pooled and globulin proteins were precipitated by dialysis against 33mM acetate buffer (pH 5.0) for 24 h at 4 °C. The precipitated proteins were recovered by centrifugation (20,000g x 20min), redissolved in $dH_2O$, dialysed against distilled water and freeze dried.

### 3.3 DNA work

### 3.3.1 Sterile working practices

Sterile working practice was maintained throughout. All media or materials were autoclaved at 15 p.s.i. for 20 minutes, unless heat sensitive, in which case they were filter sterilised by passing through a 0.22 μ nitrocellulose membrane filter.

For aliquoting LB-media or preparation of culture plates a sterile laminar flow cabinet was used. All bottle necks were flamed in a Bunsen flame before and after use, and caps were replaced on all bottles as soon as possible. Culture plates were sterilised after pouring by flaming the surface of the agar medium with a roaring Bunsen flame.

### 3.3.2 DNA precipitation with ethanol

DNA was precipitated from solutions by addition of 0.1 volume of 3M sodium acetate (pH 5.2) to the DNA solution, vortexing, followed by addition of two volumes of cold (-20 °C) 100% ethanol and vortexing. Gentle mixing was employed instead of vortexing when handling high molecular weight DNA (e.g. genomic DNA) to avoid shearing. After incubation for at least 2 h or overnight at -20°C, the precipitated DNA was recovered by centrifugation at 14,000 rpm and 4 °C for 30 minutes in a Beckman rotor (F2402 or F3602) and the pellet was washed with 70% (v/v) ethanol (-20 °C) in order to remove traces of ammonium acetate. The pellet was then dried in a vacuum desiccator for 5 minutes to remove the ethanol before being resuspended in sterile water or TE buffer.

For precipitation of DNA samples of very low concentration (<500 ng/ml), e.g. minor PCR products isolated from agarose gels, 1 µg glycogen (Boehringer) was added as a carrier.

### 3.3.3 Phenol/Chloroform extraction of DNA solutions

In order to remove proteins from DNA solutions an equal volume of phenol/chloroform/isoamylalcohol solution (25:24:1) was added and the two phases were mixed by vortexing. The mixture was centrifuged in a microcentrifuge (10,000g x 3min) and the aqueous (upper) phase containing the DNA transferred to a fresh tube. An equal volume of chloroform/isoamylalcohol mix (24:1) was added and mixed to remove traces of phenol from the aqueous DNA solution. After separating the two phases by centrifugation as before, the aqueous phase was transferred to a fresh tube and the DNA recovered by ethanol precipitation (section 3.3.2).

### 3.3.4 Agarose gel electrophresis

Agarose gel electrophoresis was performed in a Scotlab gel electrophoresis tank system (Scotlab). Agarose was added to a measured volume of 1x Tris/acetic acid/EDTA buffer (1x TAE; see appendix A) to a concentration between 0.9% to 1.2%, depending on the size of the DNA to be investigated (Sambrook *et al.*, 1989). The agarose was dissolved in a microwave oven. After cooling, ethidium bromide solution (10 mg/ml) was added to a final concentration of 0.7 µg/ml and the agarose solution poured into a mould with a well comb in place. Once the agarose had set, the gel was placed into a Scotlab gel electrophoresis tank and covered with 1x TAE buffer. DNA samples were dissolved in DNA sample buffer (appendix A) with 1 µl sample buffer per 5 µl DNA solution. The samples were then loaded into the wells of the agarose gel and electrophoresis was performed at 70 V until the DNA was separated to appropriate distances. To determine the size of the restricted DNA, a molecular weight marker was also applied onto the gel. The marker used in this work was prepared by digestion of DNA from the bacteriophage λ digested with the restriction enzyme Pst I (appendix C). The resulting nucleic acid fragments have a size range of approximately 0.5 kb to 11.5 kb (see appendix C). The gel was then placed onto a UV illuminator (310 nm) and photographed using a red filter and a polaroid camera. The sizes of the DNA fragments were determined by comparison with the molecular weight marker bands.

### 3.3.5 DNA fragment isolation from agarose gels

Two different methods were employed depending on the amount of the DNA fragment in the agarose gel and the purpose for which it was being isolated.

PCR-products were isolated by a modified electroelution method (B. Stanchev, pers. communication). DNA fragments separated in an agarose gel were visualised on UV-transilluminator, excised with a sterile razor blade and put into a short piece of dialysis tubing closed at one end with a clip. Depending on the amount of agarose gel, 0.5-1 ml of 1x TAE buffer was added, the tubing closed with a second clip and placed into an electrophoresis tank filled with 1x TAE buffer. Electrophoresis was carried out by 80 V for about 45 minutes or until the DNA was visible as a thin line on the side of the tubing when observed under UV-illumination. The direction of the current was then reversed and

37

electrophoresis carried out for about 1 min in order to detach the eluted DNA from the dialysis tubing. After electrophoresis the TAE buffer, containing the electro-eluted DNA, was transferred into a microcentrifuge tube and extracted by phenol/chloroform as described previously (section 3.3.3). The DNA was then ethanol precipitated in the presence of 1 μg glycogen (Boehringer, Mannheim) (section 3.3.2). The precipitated DNA was resuspended in an appropriate volume of sterile water or TE buffer.

When large amounts of DNA were available the QIAquick Gel Extraction Kit (QUIAGEN) was used. This was used, for example, for the preparation of plasmid DNA for sequencing. The isolation of the DNA from agarose gels was carried out using the kit according to the manufacturers instructions.

### 3.3.6 Preparation of genomic DNA from plant material

This method used was based on the method described by Murray and Thompson (1980) and was carried out with some minor modifications. In order to avoid contamination of the DNA isolated with DNA from an other source, all materials were incubated overnight in a 7% (v/v) HOCl solution (14% (w/v) available chlorine), thoroughly washed and then autoclaved.

10 g of healthy leaf tissue were put in a pre-cooled mortar, frozen with liquid nitrogen and ground to a powder. The powder was suspended in 30 ml of preheated (65 °C) buffer B (100mM Tris-HCl, pH 8.0; 1.4M NaCl; 20mM EDTA; 2% (w/v) hexadecyltrimethyl ammoniumbromide (CTAB)) in a 50 ml Falcon tube and the mixture incubated in a water bath at 65 °C for 20 minutes. After centrifugation in a bench top centrifuge at 4000 rpm for 10 minutes, the surface pellet was roughly removed with a sterile spatula and the supernatant transferred to a fresh Falcon tube. An equal volume of chloroform/isoamylalcohol (24:1) was added to the supernatant, the two phases were mixed thoroughly and then again separated by centrifugation as before for 5 minutes. The aqueous (top) phase was removed and was extracted with chloroform/isoamylalcohol two times more. The aqueous phase after the third extraction was transferred in a sterile 30 ml Corex tube. DNA was precipitated by addition of an equal volume of buffer C (50 mM Tris-HCl, pH 8.0; 10mM EDTA; 1% (w/v) CTAB) followed by incubation for 30 minutes at room temperature. After centrifugation at 9000 rpm and 4 °C for 30 minutes in a

Beckman rotor (F0850) the supernatant was removed, 2 ml of 1 M CsCl and 5 µl of RNAse (5 mg/ml) were added, the Corex tube was sealed with parafilm and placed on a shaker for 2 h to redissolve the DNA pellet and to destroy contaminating RNA. Finally, the DNA was ethanol precipitated overnight, precipitated DNA was recovered by centrifugation at 14,000 rpm and 4 °C for 30 minutes in a Beckman rotor (F8080), washed with 70% (v/v) ethanol (-20 °C) and redissolved in an appropriate volume of water for direct use or in 1x TE buffer for long time storage.

### 3.3.7 Polymerase Chain Reaction (PCR)

The polymerase chain reaction was employed for amplification of putative genes encoding legumin proteins from genomic DNA of *P. vulgaris*.

### 3.3.7.1 Selection of oligonucleotides

The selection of appropriate oligonucleotides for the polymerase chain reaction was based on sequence comparison using both amino acid sequence and a nucleotide alignment (appendix E). The sequences were retrieved from the EMBL sequence library (Heidelberg) or from the SWISSPROT data base and aligned by the clustal method using the subprogram MegAlign of the DNAStar software package (Lasergene, Ltd.). Minor corrections to this alignment were performed manually. The oligonucleotides selected were checked for potential dimerisation, self-dimerisation and hairpin-structure using the subprogram PrimerSelect of the DNAStar software package (Lasergene, Ltd.). To check potential other binding sites of the oligonucleotides within legumin genes and to analyse the expected sizes of the PCR products when using the selected primer pairs, polymerase chain reactions were imitated on a McIntosh 6100/60 computer using the program Amplify (written by W. Engels, University of Wisconsin, USA, 1992) with known legumin gene sequences derived from the EMBL database (Heidelberg) as "template DNA".

### 3.3.7.2 Oligonucleotide synthesis

Oligonucleotides were synthesised using an Applied Biosystems 381A DNA Synthesiser operated with a standard synthesis program. Oligonucleotides were synthesised using β-cyanoethylphosphoramidite chemistry and were simultaneously decyanoethylated and cleaved from the support using ammonia. After cleaving the base-protection groups with concentrated ammonium hydroxide at 55 °C overnight, the solution containing the oligonucleotide was dried using an UNIVAP vacuum centrifuge (Uniscience), redissolved in distilled water, dried again and stored at -20 °C until used. Prior to polymerase chain reactions the primers were dissolved in ddH$_2$O and its concentration was determined by spectrophoretic measuring of the OD at 260 nm. For single stranded DNA a OD value of 0.05 corresponds to an DNA concentration of 1 µg/ml (Sambrook *et al.*, 1989).

For the second primer pair used in this work for amplification of putative legumin encoding genes, oligonucleotides with attached restriction sites (EcoRI) at the 5' ends of the primers were designed. Between the 5' end of the restriction site 5 non specific nucleotides (TTATC) were added to improve the enzymatic activity of the restriction enzyme (R.R.D. Croy, B. Pöpping, pers. communication).

### 3.3.7.3 Calculation of the primer annealing temperature

The highest and lowest annealing temperatures for the oligonucleotides during the PCR were roughly calculated by the following equations (Sambrook *et al.*, 1989):

$$T_{highest / lowest} = (4x \; G/C) + 2x \; (A/T)$$

G/C and A/T = the amount of guanine and cytosine or adenine and thymidine bases per oligonucleotide. Calculating the highest annealing temperature the amount of G/C was counted as all possible guanine and cytosine bases available in the degenerate oligonucleotide sequence. Calculating the lowest annealing temperature all adenine and thymidine bases available in the degenerate oligonucleotide sequence were counted as A/T.

### 3.3.7.4 Polymerase Chain Reaction

**Standard-PCR**

In order to obtain the best possible results, a variety of different amplification conditions were investigated. In particular variations in the magnesium concentration, the amount of template DNA and the annealing temperature showed best results in optimising the reaction conditions. The PCR amplifications with primer-pair 1F/1R contained 50mM KCl; 10mM Tris-HCl, pH 9.0; 0.1% Triton X-100, 1.5-4.0mM $MgCl_2$, $2\mu M$ of each of the respective primers, 0.2mM of each dNTP, 1 unit Taq polymerase and 100-300 ng of genomic DNA from *P. vulgaris*. The template-DNA was always added at last component to the reaction mixture in order to avoid unspecific amplification at low temperatures, as Taq-polymerases show also at room temperatures certain enzymatic activity (B. Deakin, pers. communication) but the temperature of the PCR reaction mixture is below the annealing temperatur, so enabling unspecific annealing of the olgonucleotides. The amount of genomic DNA from *Pisum sativum* did not play any role in the amplification of pea legumin genes in control experiment, thus confirming the specificity of the primers.

PCR was carried out in 25 $\mu l$ volumes in 0.5 ml ultra-thin wall reaction tubes (NBL Gene Sciences, Ltd., Cramlington, England) using a Perkin Elmer DNA Thermal Cycler. After initial template denaturing at 94 °C for 4 minutes, 30 to 40 cycles of amplification were carried out with a cycling regime of 1 min denaturation at 94 °C, 45-90 sec annealing at 45-50 °C and 45-50 sec extension at 72 °C. Finally, an additional polymerisation step was performed at 72 °C for 4-10 min.

PCR products were electrophoretically separated on 0.9-1.2% (w/v) agarose gels and isolated as described previously (section 3.3.5).

**Hot-Start-PCR**

Hot-start PCRs were performed by separation of one of the primers from the second primer and the template DNA until the temperature within the reaction tube reached at least the annealing temperature of the primers, so avoiding unspecific annealing of the

41

oligonucleotides to the template DNA resulting in amplification of misprimed PCR products (Horton *et al.*, 1994).

The hot-start PCRs were carried out basically as the standard PCRs. After pipetting the mangnesium solution, the buffer (10x: 500mM KCl; 100mM Tris-HCl, pH 9.0; 1% Triton X-100), dNTPs, primer 1F or 2F, respectively, and the Taq-polymerase, each component in the same concentration as described for the standard PCR, into the reaction tube, a paraffin wax tablet (BDH Chemicals Ltd.) was added and melted by incubating the tube at 80 °C (water bath). After chilling on ice the wax became solid and ddH$_2$O, primer 1R or 2R, respectively, and the template-DNA were added on top of the solid wax layer. The tubes were then at once transferred onto the PCR cycler and the reaction started, basically under the same cycling conditions as for the standard PCRs.

## Re-PCR

Re-PCR was generally carried out using the isolated specific PCR product from the original PCR as template in the re-amplification of this particular PCR product. This, however, implied the isolation of the original PCR product after electrophoretic separation from the agarose gel. When the first PCR did not provide sufficient product, isolation of these minor amounts of PCR products from agarose gels was difficult or not possible. In these cases a small piece of the agarose gel containing the specific PCR product under investigation was taken using a sterile needle. The agarose gel containing the template DNA for the re-PCR was then at once added to the re-PCR reaction mix in order to avoid contamination during transport and storage of the agarose gel. During the initial denaturation step of the re-PCR the agarose gel was melted and the DNA released into the reaction mix.

### 3.3.8  Southern analysis of PCR products

Southern analyses of PCR products were performed basically to the method of Southern (1975) with modifications of Sambrook *et al.* (1989).

### 3.3.8.1 Preparation of a legumin gene probe

A genomic legumin clone from legumin A, a legumin encoding gene of pea (clone pDUB24; Lycett *et al.*, 1984; accession-number X02982), was used as a heterologous probe. The legumin coding region was isolated by digestion with Xho I of the legumin A gene cloned into the pUC8 vector:

| | |
|---|---|
| ddH$_2$O | 44 µl |
| Restriction enzyme buffer (10x) | 5.0 µl |
| DNA (2µg/µl) | 1.0 µl |
| XhoI (10U/µl) | 1.0 µl |

This digestion resulted in an 1.8 kb fragment containing besides the 4 exons (about 200 bp of the first exon were cleaved off from the 3' end) three legumin gene introns and about 100 bp of the 5' NT-region (compare Lycett *et al.*, 1984).

After 3 hours incubation at 37 °C the DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in an appropriate volume for the labelling reaction (see section 3.3.8.3).

### 3.3.8.2 Southern blots of restricted DNA

Following electrophoresis the agarose gel was photographed with a fluorescent ruler laid at the side of the gel. The DNA fragments were then transferred to and immobilised onto Hybond-N$^+$ nylon membranes (Amersham) essentially by the capillary method of Southern (1975) with the modifications of Sambrook *et al.* (1989).

The DNA was denatured by soaking the agarose gel in denaturing solution (0.5M NaOH; 1.5M NaCl) for 30 minutes. The agarose gel was placed onto 10 pieces of Whatman 3MM paper cut to the same size of the gel and soaked in denaturing solution. An appropriately sized piece of Hybond-N$^+$ membrane was placed on top and air bubbles thoroughly removed between the gel and the membrane. A further 10 pieces of Whatman 3MM paper and two pieces of absorbent nappies of the same size as the gel were placed on top of the membrane. Finally, a glass plate with one kilogram lead weight was put on top of the blot

and the DNA was transferred by capillary action overnight in a reservoir of denaturing solution. The membrane was then removed and the right lower corner was cut to mark the orientation. The filter was first placed on a 3MM Whatman paper soaked in neutralisation solution (1M Tris-HCl, pH 7.5; 1.5M NaCl) and then on Whatman paper soaked in 9x SSC (1.35M NaCl; 0.14M sodium citrate), each for 10 minutes. Finally, the membrane was dried on Whatman paper and the DNA fixed onto the nylon filter by cross-linking by exposure of the membrane to a UV-transilluminator (310 nm) for 30 seconds (DNA side down) or by baking of the membrane for 1-2 h at 80 °C in a vacuum oven.

### 3.3.8.3 Southern hybridisation

The DIG Luminescent Detection Kit (Boehringer) was used for the detection of nucleic acid fragments hybridising to the probe used. This non-radioactive detection system uses digoxigenin, a steroid hapten, coupled to dUTP to label the DNA used as probe for hybridisation and subsequent luminescence detection. The DIG detection system had the advantage that the whole procedure was much faster than isotopic labelling methods. Reasons for this were the shorter pre-hybridisation time, the shorter wash procedure, the possibility of long-term application of the labelled probe, since it was not radioactively labelled, and the short exposure time.

Labelling of the probe and hybridisation and detection procedures were carried out exactly according to the manufacturers instructions. A hybridisation buffer containing 0.1% (w/v) N-lauroylsarcosine, 0.02% (w/v) SDS and 1% blocking reagent (Boehringer) was used for prehybridisation and hybridisation of the PCR products with the labelled probe.

The hybridisation stringency was 60 °C. The final wash stringency of the post-hybridisation washes was 0.5x SSC at 60 °C. The labeled probe hybridised to DNA fragments on the membrane was immunodetected with anti-dioxigen, Fab fragments conjugated to alkaline phosphatase. The probe-DNA was then visualised with the chemiluminescence substrate CSPD. Enzymatic dephosphorylation of CSPD by alkaline phosphatase leads to light emission at a maximum wavelength of 477 nm which can be recorded on X-ray film. Exposure of an X-ray film (Fuji) to the membrane was usually for 30 minutes and 1 hour or overnight.

The signals obtained on the X-ray film could be related to the PCR products on the agarose gel by comparison of the distances of these bands from the start of electrophoresis (gel wells visible on the X-ray film after exposure) with the distances of the DNA bands on the gel from the gel wells, determined by using the photographed ruler (see section 3.3.8.2). Together with comparison of the bands of the molecular weight marker applied onto the gel, a good estimation of the sizes of the DNA fragments hybridised to the probe is possible.

### 3.3.9 Ligation of PCR fragments into plasmids

#### 3.3.9.1 Ligation of PCR products into pKS$^+$ digested with EcoRI

**Preparation of pKS$^+$ Vector**

DH5$\alpha$ transformed with the pKS$^+$ plasmid was grown in 10 ml LB-medium with 50 $\mu$g/ml ampicillin overnight at 37 °C with gentle agitation. The plasmid DNA was isolated by use of a Wizard Minipreps Kit (Promega) according to the manufacturers instructions and the DNA was then directly used for the restriction digestion with EcoRI. The following digestion reaction was performed:

| | |
|---|---|
| Plasmid DNA (250 ng/$\mu$l) | 16 $\mu$l |
| ddH$_2$O | 19 $\mu$l |
| Restriction enzyme buffer (10x) | 4 $\mu$l |
| EcoRI (10U/$\mu$l) | 1 $\mu$l |

The reaction volume was mixed and incubated at 37 °C for 3 h. After phenol/chloroform extraction (section 3.3.3) and ethanol precipitation (section 3.3.2), the DNA was resuspended in 40 $\mu$l ddH$_2$O. To remove the terminal 5' phosphate groups from the vector DNA an aliquot was used for the following phosphatase reaction using the enzyme calf intestinal phosphatase (CIP). This treatment prevents the linearised vector against self-ligation.

| | |
|---|---|
| DNA (1µg/12µl) | 12 µl |
| Restriction enzyme buffer (10x) | 2 µl |
| ddH$_2$O | 5 µl |
| CIP (1U/µl) | 1 µl |

The reaction mixture was incubated at 37 °C for 1 h. The CIP was removed from the reaction solution by phenol/chloroform extraction (section 3.3.3) and ethanol precipitation (section 3.3.2) since CIP is not inactivated by heating at 65 °C (Sambrook *et al.*, 1989). The plasmid DNA was resuspended in 20 µl of water so that 1 µl was equal to about 45 ng digested and dephosphorylated plasmid DNA (T. Gibbons, pers. communication). A small aliquot was used to check by agarose gel electrophoresis whether the digestion and dephosphorylation was successful (a single band, 2.96 kb in size).

## Digestion of PCR products

PCR products that were to be cloned, were electro-eluted from the agarose gel, dissolved in 5-20 µl water and used for the restriction digestion with EcoR I:

| | |
|---|---|
| DNA | 5 µl |
| ddH$_2$O | 3.8 µl |
| Restriction enzyme buffer (10x) | 1 µl |
| EcoRI (10U/µl) | 0.2 µl |

After incubation for 3 h at 37 °C the enzyme activity was destroyed by incubation at 65 °C for 10 minutes. The whole reaction volume was used directly for the following ligation reaction.

## Ligation of PCR products into pKS$^+$

The ligation was performed in a 0.5 ml microcentrifuge tube:

| PCR product | 10 µl |
|---|---|
| pKS⁺ plasmid DNA | 1 µl |
| Restriction enzyme buffer (10x) | 1.5 µl |
| ddH₂O | 2.5 µl |
| T4 DNA Ligase (4U/µl) | 1 µl |

The ligation reaction was carried out at 15 °C overnight. The ligation mix was stored at 4°C until used for transformation.


### 3.3.9.2 Ligation into a T-Vector

The pGEM®-T Vector System was also used for the ligation of PCR fragments into plasmid. The T-vector is prepared by cutting the pGEM-5Zf(+) vector with EcoRV and adding a 3' terminal thymidine to both ends which allows the direct cloning of PCR products because of the non-template dependent addition of a single deoxyadenosine to the 3' ends of PCR products by many Taq-polymerases during the PCR.

The following reaction mix was set up in a 0.5 ml microcentrifuge tube:

| | |
|---|---|
| dH₂O | 0-6 µl |
| PCR product | 1-7 µl |
| T4 DNA Ligase buffer (10x) | 1 µl |
| pGEM-T Vector (50 ng/µl) | 1 µl |
| T4 DNA Ligase (1 Weiss unit/µl) | 1 µl |
| | Σ 10 µl |

The reaction mix was incubated at 15 °C overnight and stored at 4 °C. The next day the ligase was inactivated by incubation at 65 °C for 10 min. The ligation mix was then stored at 4 °C until used for transformation.

## 3.3.10 Transformation of plasmid DNA into *E.coli*

### 3.3.10.1 Preparation of competent cells

The polyethyleneglycol/dimethylsulfoxide method described by Chung *et al.* (1989) was used in order to prepare competent *E.coli* bacteria for transformation experiments.

10 ml of an overnight culture of XLBlue-1 or DH5α was made by inoculating 10 ml of LB-medium with bacteria taken with an inoculation loop directly from the frozen stock bacterial culture in glycerol (-70 °C) without thawing the culture. 1 ml of the overnight culture was then used to inoculate 50 ml LB medium in an Erlenmeyer beaker closed with cotton wool and the bacteria were allowed to grow to the exponential stage at 37 °C with vigorous shaking. The OD of the bacterial suspension was checked every 30 minutes and incubation was continued until the OD reached 0.2 at 650 nm measured on a spectrophotometer. At this stage the cells were pelleted by centrifugation at 4,000 rpm in a benchcentrifuge for 5 minutes. The supernatant was discarded and the pellet resuspended in 0.1 of the original volume of the culture using ice-cold transformation and storage solution buffer (LB-medium containing 10% (w/v) PEG 6000, 50mM MgCl and 5% (v/v) DMSO; pH 6.5). The competent cells were distributed in 200 μl aliquots in 1.5 ml eppendorf tubes and stored at -70 °C. For every transformation reaction the whole aliquot was used.

### 3.3.10.2 Transformation reaction of plasmid DNA into *E.coli*

The transformation of plasmid DNA was carried out by the heat shock method basically according to Sambrook *et al.* (1989).

Two tubes with frozen (-70 °C) competent cells (200 μl) were thawed in an ice bath. 2 μl and 8 μl of the ligation reaction (10 μl) were added, each aliquot to a tube of competent cells, carefully mixed and incubated on ice for 30 min. The *E.coli* cells were then exposed to a heat shock for 45 sec at exactly 42 °C and then placed immediately on ice for 2 min followed by addition of 1 ml of LB broth (pre-heated to 37 °C). The cells were then incubated on a shaker at 37 °C for 1 h.

Finally, the transformation mix was centrifuged in a microcentrifuge (10,000g x 2min) and the pellet resuspended in 50 μl of LB medium. The LB medium was then spread on a X-gal plate with 75 μg/ml ampicillin using a Trigalski-spatula and the petri plate incubated overnight at 37 °C. *E.coli* transformations containing a plasmid with an insert where identified as white colony caused by the lack of the β-galactosidase activity (Sambrook *et al.*, 1989).

### 3.3.10.3 Preparation of plasmid DNA from *E.coli*

The alkaline method described by Sambrook *et al.* (1989) was employed to isolate plasmid DNA from isolated clones.

Often, *E.coli* colonies were growing on ampicillin containing LB-plates without having an insert or blue colonies contained inserts. Therefore, *E.coli* colonies were first checked for inserts by a small scale miniprep prior to performing a large scale preparation.

One ml of LB-medium containing 50 μg/ml ampicilin in eppendorf tube was inoculated with a single bacterial colony and incubated overnight with agitation. The culture was centrifuged in a microcentrifuge (10,000g x 2min) and the supernatant was totally removed using a pipette. The pellet was resuspended in 50 μl of cell resuspension solution (50mM Tris/HCl, pH 7.5; 10mM EDTA, pH 8.0; 100 μg/ml RNase A). The cell were lysed by addition of an equal volume (50 μl) of cell lysis solution (0.2M NaOH, 1% (w/v) SDS) and inverting the eppendorf tube several times. The cell suspension cleared almost immediately during this treatment. The lysate solution was neutralised by adding 50 μl of neutralisation solution (potassium acetate, pH 1.32M, pH 4.8) and inverting the microcentrifuge tube several times. After centrifugation for 10 minutes the clear supernatant was transferred to a new microcentrifuge tube. The plasmid DNA was precipitated by ethanol precipitation at -20 °C for 30 minutes (section 3.3.2). After centrifugation and drying of the pellet, the plasmid DNA was resuspended in 15 μl of distilled water. The plasmid DNA was stored at -20 °C until used for sequencing.

In order to check for the presence and the size of the inserts, the plasmid DNA isolated from the transformed *E.coli* was digested with appropriate restriction enzymes and

separated on an agarose gel to investigate the presence of a successful ligated insert. For example, when using the pGEM T-vector for ligation, a double digestion was performed with Sac I (Boehringer) and Sph I (Boehringer) for 2.5 h at 37 °C using the Sph I restriction enzyme buffer (buffer M: 10mM Tris/HCl, ph 7.5; 10 mM MgCl$_2$; 50mM NaCl; 1mM DTT):

| | |
|---|---|
| DNA | 15 µl |
| Restriction enzyme buffer | 1.8 µl |
| Sac I | 0.3 µl (3U) |
| Sph I | 0.3 µl (3U) |

After the digestion the DNA sample was applied onto a 1% (w/v) agarose gel and electrophoresed.

For a large scale plasmid DNA preparation, for example for subsequent DNA sequencing, the Wizard Minipreps - Kit (Promega) was used following the producer's instructions.


### 3.3.11 Sequencing of DNA

DNA sequencing was performed by an automated dideoxy-sequencing method based on the method of Sanger *et al.* (1977), using a 373A Automatic DNA Sequencer (Applied Biosystem) and fluorescent dye linked primers appropriate for the plasmid vector used. The automatic DNA sequencer was based on fluorescent detection of electrophoretically separated DNA fragments. Four dye-labelled primers, each paired with one of the four dideoxynucleotide triphosphate (ddNTPs) chain terminators, were used in the sequencing reaction. Each of the four ddNTPs are paired with different dyes. After electrophoresis the separated reaction products can be distinguished by the colour of the fluorescence detected with a laser scanner. The sequencing was performed by Ms Julia Bartley.

### 3.3.12 Methods for analysing sequence data

#### 3.3.12.1 Sequence comparison with data bases

In order to identify sequences obtained from automated N-terminal protein sequencing or DNA sequencing, the sequences were compared with known sequences in the EMBL sequence library (Heidelberg). For this purpose the Netscape web browser software was used for the internet access.

#### 3.3.12.2 Sequence alignments

Protein and nucleotide sequence alignments were compiled using the MegAlign subprogram of the DNAStar software package (Lasergene, Ltd.) run on an IBM-compatible PC (DX4-100; 16Mb). Sequences for the alignment were retrieved either from the SWISSPROT or EMBL sequence library. To this alignment the obtained sequence data were aligned by the Martinez/Needleman-Wunsch-algorithm or the CLUSTAL method (Higgins and Sharp, 1988; Williams *et al.*, 1992 and references herein). Minor corrections to the alignment were performed manually.

## 4.1 Isolation and identification of albumin proteins

### 4.1.1 Glycoprotein I

Glycoprotein I was reported to be a 60 kDa albumin-type protein which undergoes autodigestion to smaller fragments (Pusztai and Duncan, 1971). No other enzymatic activity could be determined for Glycoprotein I (Pusztai and Duncan, 1971). Croy (1977) identified for the first time the subunit pattern of Glycoprotein I on a SDS-gel. He described the protein as a tetramer equally built up of two different subunits in the size ranging from 14 kDa to 19 kDa. Due to the lack of any known enzymatic activity the analysis of the single purification steps was based on SDS-PAGE.

**4.1.1.1 Isolation of Glycoprotein I**

The purification of Glycoprotein I was based on the method developed by Pusztai (Pusztai, 1966) and led finally to the isolation of this protein.

The albumin fraction of seed proteins from *P. vulgaris* was obtained by using a low salt extraction buffer of acidic pH (0.033M sodium acetat, pH 5.0). Contaminating globulin proteins were precipitated from the albumins by dialysis against the same acidic buffer (e.g. compare figure 4.1.4 - lane 1 and 2).

The first enrichment of Glycoprotein I within the albumin fraction was obtained by ammonium sulphate fractionation of the albumin extract. Figure 4.1.1 shows that the proteins with subunits in the size range between 14 and 20 kDa were clearly enriched in the 60% (rel. sat.) ammonium sulphate fraction. Phytohemagglutinins, the most abundant albumin-type seed proteins in the seeds of *P. vulgaris*, were mainly precipitated at 80% (rel. sat.) ammonium sulphate. After dialysis against distilled water (36 h at 4 °C) the protein solution of the 60% (rel. sat.) ammonium sulphate fraction was freeze dried, an aliquot resuspended in a Tris buffer (0.1M Tris/HCl, pH 8.0; 0.1M NaCl) and applied onto a gel filtration column (Sephacryl S-300) through which a further separation of Glycoprotein I from other seed albumin proteins and particularly from residual phytohemagglutinins was achieved (figure 4.1.2). The Glycoprotein I-containing fractions 53 to 58 from the gel filtration experiment were pooled and used for further purification. These fractions were dialysed against distilled water, freeze dried, dissolved in 0.025M

Figure 4.1.1: SDS-PAGE analysis of ammonium sulphate fractionation of albumin proteins from seeds of *P. vulgaris*.

Proteins were extracted from seed flour using a 0.033M sodium acetate buffer, pH 5.0. Contaminating globulin proteins were precipitated from the protein extract by dialysis for 48 hours against the same sodium acetate buffer and removed by centrifugation. The albumin proteins were precipitated from the albumin protein solution with

lane 1:   60% (rel. sat.) $NH_4SO_4$

lane 2:   60-80% (rel. sat.) $NH_4SO_4$

lane 3:   80-100% (rel. sat.) $NH_4SO_4$

Aliquots (approximately 10 μg protein) of each of the precipitated proteins were redissolved in sample buffer and applied onto a 15% (w/v) SDS-polyacrylamide gel. After electrophoresis the gel was stained with coomassie blue R250. Phytohemagglutinins were mainly precipitated at 60-80% (rel. sat.) $NH_4SO_4$ but also within the 80-100% (rel. sat.) $NH_4SO_4$ fraction.

M indicates the separated standard proteins.

kDa

66

45
36

29
24

20

14

← phyto-
hemagglutinin

Glycoprotein I
subunits

M        1        2        3
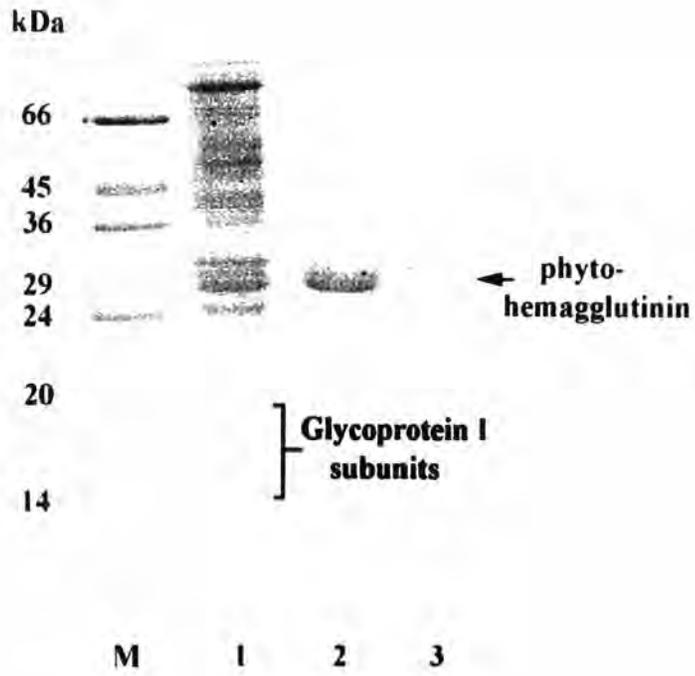
**Figure 4.1.2:** Gel filtration of the 60% (rel. sat.) ammonium sulphate fraction of albumin proteins from seeds of *P. vulgaris*.

A) Elution profile of proteins from the gel filtration on Sephacryl S-300.

The albumin proteins precipitated at 60% (rel. sat.) ammonium sulphate were redissolved in equilibration buffer (0.1M Tris/HCl, pH 8.0; 0.1M NaCl), dialysed for 24 h against the same buffer and 5 ml were applied onto the Sephacryl S-300 column (2.5cm x 110cm). Chromatography was performed at a flow rate of 0.3ml/min and fractions were collected every 10 min.

Fractions are numbered above the elution profile.

B) SDS-PAGE analysis of gel filtration on Sephacryl S-300.

An aliquot (500 µl) of each of the fractions investigated (compare fraction number with elution profile) were acetone-precipitated (section 3.2.2.2) and applied onto a 15% (w/v) SDS-polyacrylamide gel. After electrophoresis the gel was stained with coomassie blue R250. The arrowheads indicate the chitinases identified in this work (section 4.1.3). M indicates the standard proteins.

**A)**
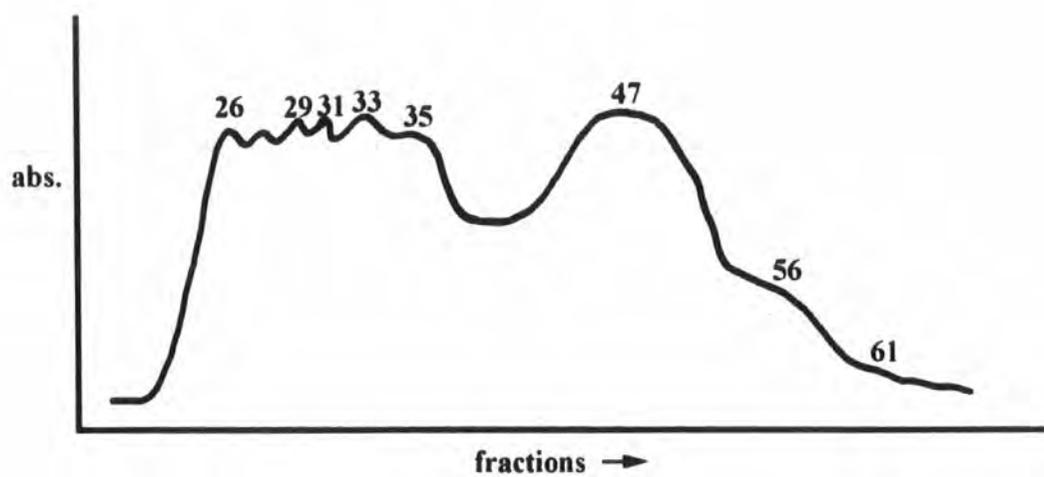


**B)**
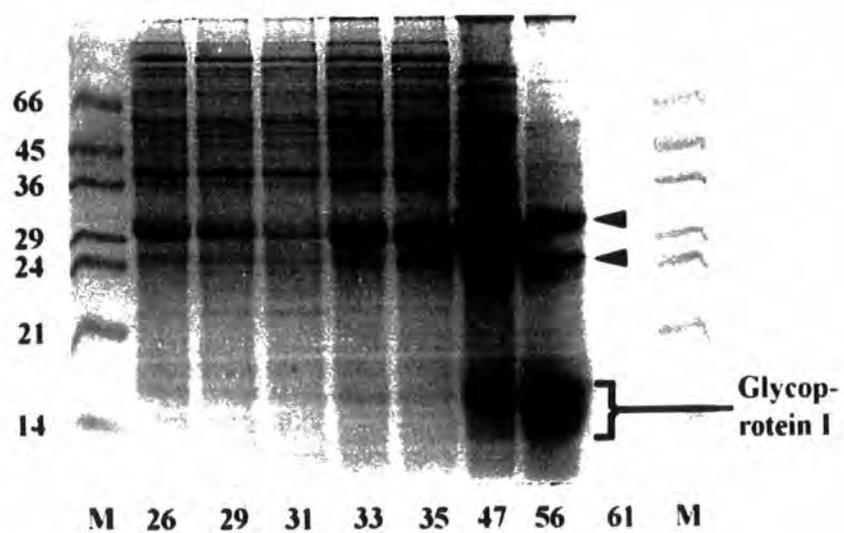
**Figure 4.1.3: Ion-exchange chromatography of the Glycoprotein I containing fraction after gel filtration**

A) Elution profile of proteins from the ion-exchange chromatography.

Fractions 53 to 58 from the gel filtration was dialysed for 48 h against a Tris buffer (0.025M Tris/HCl, pH 7.6). 2.5 ml of the protein solution were applied onto the DEAE-cellulose column (2.5cm x 5cm) equilibrated with 0.025M Tris/HCl buffer (pH 7.6). Proteins which did not bind to the cellulose were washed through with the equilibration buffer. Then a linear NaCl gradient from 0M to 0.5M NaCl in the equilibration buffer was applied at a flow rate of 1 ml/min and fractions were collected every 2 minutes. Fractions are numbered above the elution profile.

B)/C) SDS-PAGE analysis of the ion-exchange chromatography on DEAE-cellulose.

15 µl of different fractions of the eluted proteins were applied onto a 15% (w/v) SDS-polyacrylamide gel. After electrophoresis the gels were silver stained.

M indicates the separated standard proteins.

B) Separation of both chitinases

C) Separation of PVSOD (arrowhead) and Glycoprotein I subunits (arrows: Glyco14 and Glyco18). Fraction 40 contains both PVSOD and Glycoprotein I.

D) Later repeat of the ion-exchange chromatography which led to the separation of two further size classes of Glycoprotein I subunits (Glyco15 and Glyco17).

**A)**

0.5M NaCl

abs.

9

28  30  32

38  40

42

35

46      50

0M NaCl

fractions ⟶

**B)**

kDa

66

45
36
29
24
21

14

27 28 29 30 31 32 33 34 35  M

**C)**

kDa

66
45
36

29
24

20

14

M  37  38  40  42  46  50  M

**D)**

SOD
G18
G17
G15

G14
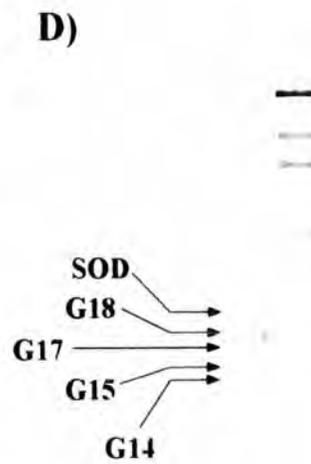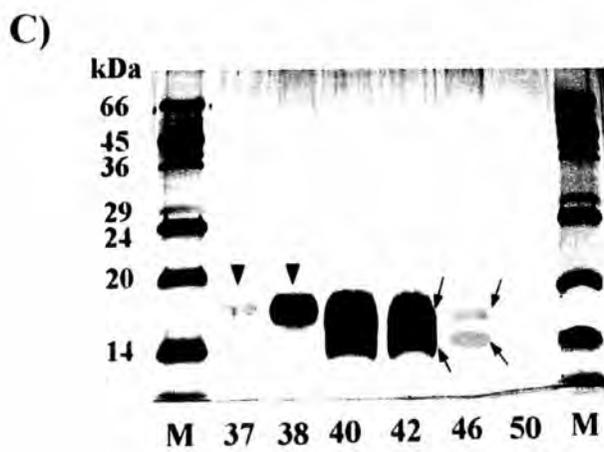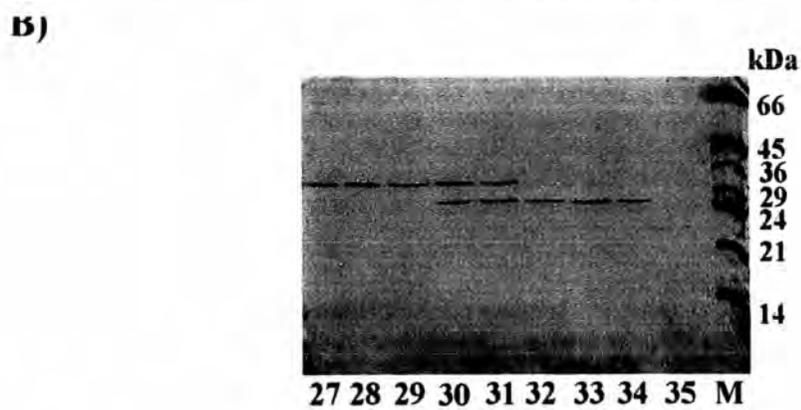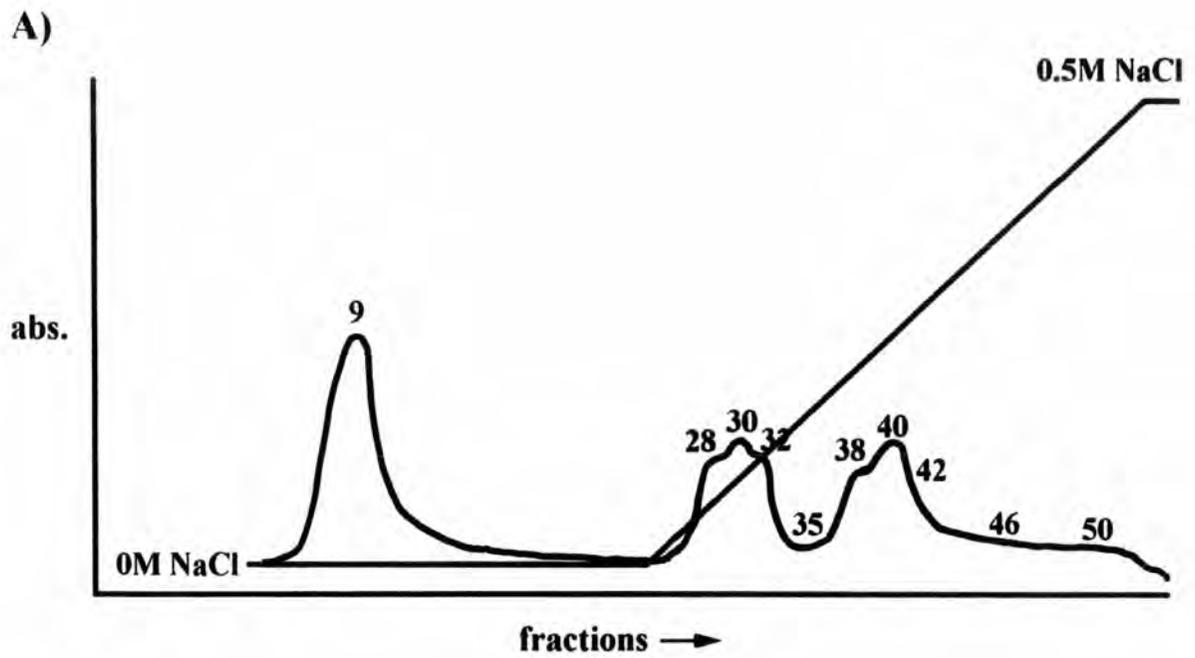
Tris/HCl, pH 7.6 and applied onto a DEAE (DE-52) ion-exchange column. The ion-exchange chromatography led directly to the separation of Glycoprotein I from the majority of contaminants (figure 4.1.3). Whereas the two major contaminating proteins after gel filtration (25 kDa and 32 kDa in size; see figure 4.1.2B: arrowheads) were eluted at a low salt concentration (0.03M and 0.09M NaCl; figure 4.1.3B), Glycoprotein I was only eluted with higher salt concentration (0.17 - 0.2M NaCl) in the elution buffer (figure 4.1.3C). The Glycoprotein I subunits proved to represent a 18 kDa and a 14 kDa polypeptide. Furthermore, a 19 kDa protein was also separated from Glycoprotein I by this procedure (figure 4.1.3C: arrowheads; see section 4.1.2). A later repeat of the isolation of Glycoprotein I resulted in two putative Glycoprotein I subunits of slightly different $M_r$ (15 kDa and 17 kDa; figure 4.1.3D). In figure 4.1.4 protein samples of all isolation steps were separated by SDS-PAGE and stained for glycoproteins (Schiff's staining) and subsequently for all proteins (commassie blue staining). In the size range between 14 and 18 kDa there are two glycosylated protein bands visible which represent glycosylated Glycoprotein I subunits. The upper glycosylated protein band is probably formed by two single polypeptide bands as judged by the intensity of the band (figure 4.1.4 - lane 5: two upper arrows) meaning that three different size classes of Glycoprotein I subunit polypeptides are glycosylated. The 19 kDa polypeptide was most likely not part of this upper, glycosylated polypeptide band as it revealed by N-terminal sequencing to be a superoxide dismutase (SOD; see section 4.1.2). SODs are known not to be glycosylated.

### 4.1.1.2 N-terminal sequencing of Glycoprotein I subunits

A major objective of this work was to obtain amino acid sequence information from purified Glycoprotein I. Pusztai and Duncan (1971) described besides Glycoprotein I a contaminating protein obtained within the first fractions of the Glycoprotein I elution peak from the ion-exchange chromatography column. To possibly elucidate also the nature of this polypeptide, the front part of the Glycoprotein I elution peak (fractions 37 and 38) representing the 19 kDa subunit polypeptide (figure 4.1.3C - arrowheads) was further investigated besides Glycoprotein I. Aliquots of the freeze dried protein samples of the combined fractions 37 to 38 and 43 to 45 from the ionic exchange column were dissolved in SDS sample buffer, applied on a 15% (w/v) SDS-polyacrylamide gel and separated by electrophoresis. After blotting onto a PVDF membrane the appropriate bands

**Figure 4.1.4:  PAS-staining of electrophoretically separated proteins at each of the purification steps of Glycoprotein I**

Aliquots (approximately 10-20 μg protein (lane 1-3) or 5-10 μg protein (lane 4 and 5)) of proteins after each of the purification steps performed to isolate Glycoprotein I were dissolved in sample buffer and applied onto a 15% (w/v) SDS-polyacrylamide gel. The proteins were electrophoretically separated and stained for attached oligosaccharides using PAS staining. Subsequently the gel was stained with coomassie blue R250. The red bands indicate glycosylated proteins. The three arrows in lane 5 indicate the three glycosylated Glycoprotein I subunits (Glyco18, Glyco17 and Glyco15).

M:  Standard protein marker.

Lane 1:  Total protein (SDS-) extract.

Lane 2:  Albumin proteins

Lane 3:  60% (rel. sat.) $NH_4SO_4$ precipitated albumin fraction.

Lane 4:  Fraction 56 from the gel filtration on Sephacryl S-300.

Lane 5:  Glycoprotein I containing fraction from ion-exchange chromatography.
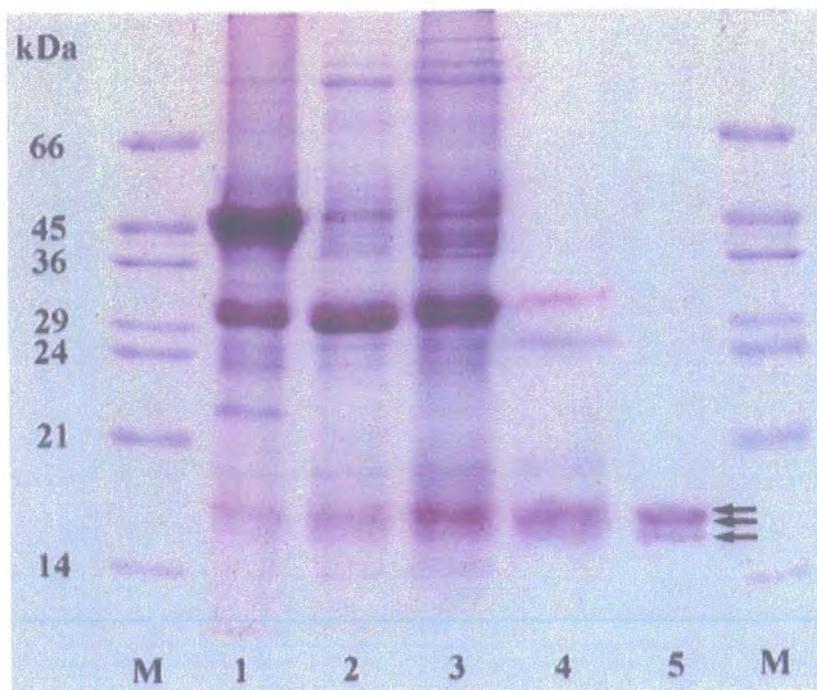
(approximately 19 kDa and 14 kDa and 18 kDa in size, respectively) were located and excised with a sterile blade. The N-terminal sequences of the Glycoprotein I subunits were determined by automated N-terminal sequencing. The results from the protein sequencing are presented in table 4.1.1. In order to find any similarities between these N-terminal amino acid sequences with already known protein sequences, the obtained sequences were used to scan the EMBL protein sequence library (Heidelberg). Whereas the 19 kDa protein revealed to be highly similar to the N-termini of superoxide dismutases (SOD) proteins from different plant species (see section 4.1.2) both Glycoprotein I bands of sizes 14 kDa (designated as "Glyco14") and 18 kDa ("Glyco18") showed strong homology to sequences within the α-subunits of legumin J and K, seed proteins from pea (figure 4.1.5). The repeat of the purification of Glycoprotein I and subsequent SDS-PAGE analysis led to the separation of two additional Glycoprotein I subunits of $M_r$ 15 kDa ("Glyco15") and 17 kDa ("Glyco17") (figure 4.1.3D). The N-terminal sequencing of these two protein bands revealed exactly the same N-terminal amino acid sequence for both polypeptides, but which was different from the first two sequences (for Glyco14 and Glyco18). The amino acid sequence obtained for Glyco15 and Glyco17 was also similar to the same sequence within legumin J and K as found for Glyco14 and Glyco18 (figure 4.1.5). Table 4.1.1 shows the obtained N-terminal sequences and the results of the database search.

| Subunit $M_r$ | N-terminal amino acid sequence | Sequence similarity to |
|---|---|---|
| 14 kDa | SHHDRREEEE | legumin J and K (pea) |
| 15 kDa | LSGHRHEEED | legumin J and K (pea) |
| 17 kDa | LSGHRHEEED | legumin J and K (pea) |
| 18 kDa | SGHHHEEEEDDPE | legumin J and K (pea) |
| 19 kDa | VSAVAVLGSSEDVTA | superoxide dismutases (different plant species) |

Table 4.1.1  N-terminal amino acid sequences obtained for Glycoprotein I subunits and a
19 kDa superoxide dismutase-subunit polypeptide.

Figure 4.1.5 shows a schematic diagram of legumin J protein showing the position of the three different N-terminal Glycoprotein I sequences. This is the first report of a amino acid sequence of this protein.
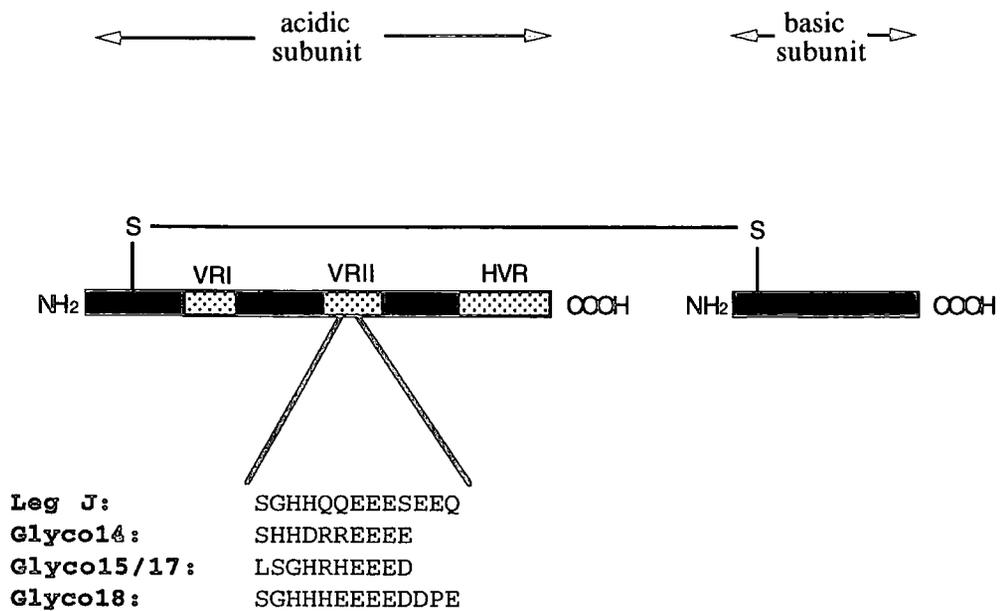
Figure 4.1.5: Alignment of the N-terminal sequences obtained for the three different Glycoprotein I subunits as it reveals by the database search.

The black areas represent conserved regions, the stippled areas variable regions (VRI/II = variable region I/II; HVR = hyper-variable region).

### 4.1.1.3 Glycoprotein I during seed germination

In contrast to the biochemical characterisation of Glycoprotein I (e.g. Pusztai, 1968; Pusztai and Duncan, 1971), no information is available for the behaviour of the protein during seed germination.

In order to investigate this aspect of Glycoprotein I, polyclonal antibodies against the purified protein were raised in mice (section 3.2.4.3). Bean seeds were germinated for periods of up to nine days and crude albumin protein extractions prepared from seeds of each day of germination using an acetate buffer (0.033M sodium acetate, pH 5.0, 4 °C; four volumes (v/w ) cold acetate buffer were added to the seed meal). After removing of all solid particles by centrifugation (20,000g x 30 min) no further purification steps such as precipitation of globulin proteins were performed. Equal amounts of this crude albumin extract from each stage of seed germination was separated by SDS-PAGE on a 15% (w/v) polyacrylamide gel and blotted onto nitrocellulose. Glycoprotein I was immuno-detected with the anti-Glycoprotein I antiserum. The antiserum was used at a 1:5000 dilution. Figure 4.1.6 shows the result of this experiment. The protein is rapidly broken down and was barely visible after only 4 days. 5 to 7 days after imbibition no antibody binding was detected to any components in the crude albumin extract. Furthermore, no breakdown products of the Glycoprotein I subunits could be detected as judged by the absence of any bands smaller than the Glycoprotein subunit bands (figure 4.1.6B). It is further noteworthy, that probably three Glycoprotein I subunits show cross reactivity to the Glycoprotein I antibodies, two forming together the upper strong signal and one the faint signal (figure 4.1.6B).

In order to compare this result with the breakdown pattern of phaseolin, the major seed storage proteins in the seeds, aliquots of a crude protein extract from each of the 9 days of germination were separated on a 12% (w/v) SDS-polyacrylamide gel and immunodetection of the vicilin-like globulins was performed with an antiserum raised against *P. vulgaris* vicilins. The first degradation products of phaseolin proteins are visible at 3 days after imbibition (figure 4.1.7). 9 days after start of germination phaseolin proteins were still present in the bean seeds in the non-degraded form indicating that the 7S vicilins serve as reserve proteins for a later stage of seed germination than Glycoprotein I.

**Figure 4.1.6: Glycoprotein I during seed germination**

Approximately equal amounts of albumin extract (about 20-25 μg proteins) from each day of germination (up to day nine) were separated on two 15% (w/v) SDS-polyacrylamide gels. Proteins of one gel were coomassie blue stained, the proteins of the second gel were blotted onto nitrocellulose and the blotted proteins were analysed by immunodetection with antiserum raised against Glycoprotein I. The primary antiserum was used as a 1:5000 dilution (section 4.1.1.3).

A)  shows the SDS-polyacrylamide gel after coomassie blue staining.

B)  resulting autoradiogram after hybridisation of the blotted proteins with anti-Glycoprotein I antiserum.

M indicates the standard proteins.

**A)**



kDa

66

45
36

29

24

20

14

Glycoprotein I
subunits

M  0  1  2  3  4  5  7  8  9

days after imbibition

→ phaseolin

→ phyto-
hemagglutinin

→ phaseolin
degradation
products

**B)**



Glycoprotein I
subunits

0  1  2  3  4  5  7  8  9

days after imbibition
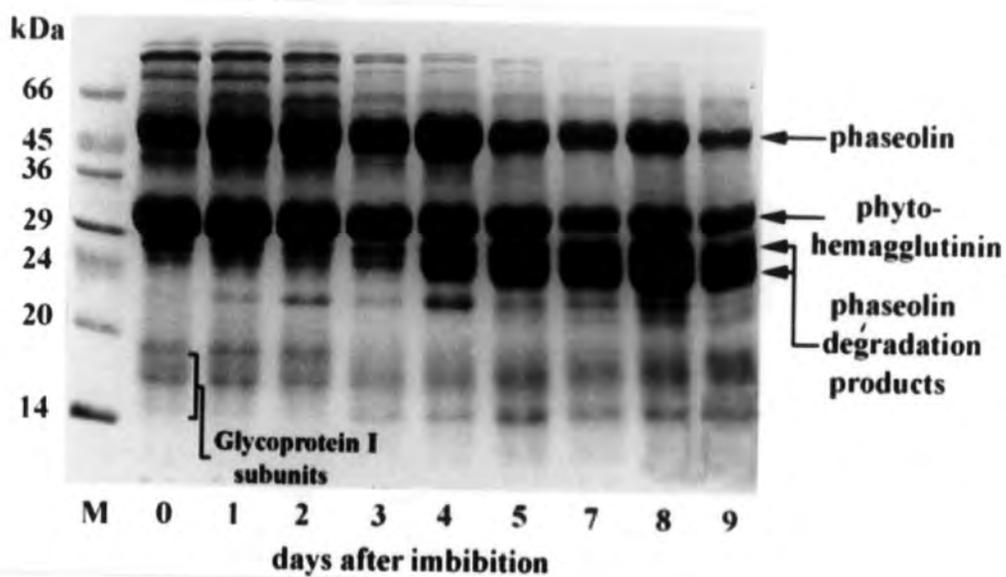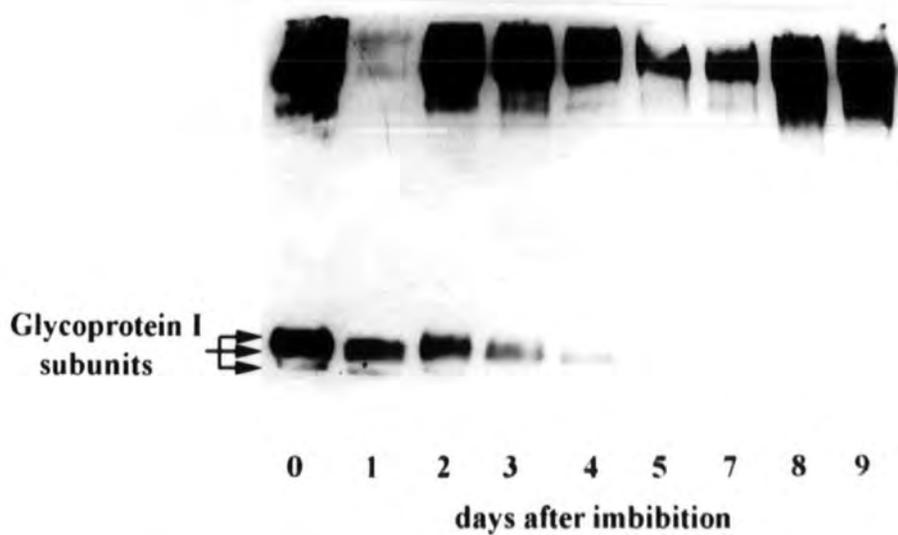
**Figure 4.1.7:** *P. vulgaris* 7S globulins during seed germination

Proteins were extracted from *P. vulgaris* seeds up to 9 days after imbibition using a 0.033M sodium acetate buffer (pH 5.0). Approximately 20-25 μg protein extract of each of the stages of germination were applied on two 12% (w/v) SDS-polyacrylamide gels. After electrophoresis proteins of one of the two SDS-polyacrylamide gels were stained with coomassie blue R250 (A). Proteins of the second gel were electro-blotted onto nitrocellulose and reacted with primary antiserum (1:10,000 dilution) raised against *P. vulgaris* vicilins.

A) shows thecoomassie blue stained gel.

B) shows the resulting auto-radiogram of the immunodetection.

M indicates the lane with the standard proteins.

A)

kDa

66
45
36

29
24

20

M   0   1   2   3   4   5   7   8   9

days after imbibition

phaseolin

phyto-
hemagglutinin

phaseolin
degradation
products

B)

kDa

66 →
45 →
36 →

29 →
24 →

20 →

0   1   2   3   4   5   7   8   9

days after imbibition

phaseolin

phaseolin
degradation
products
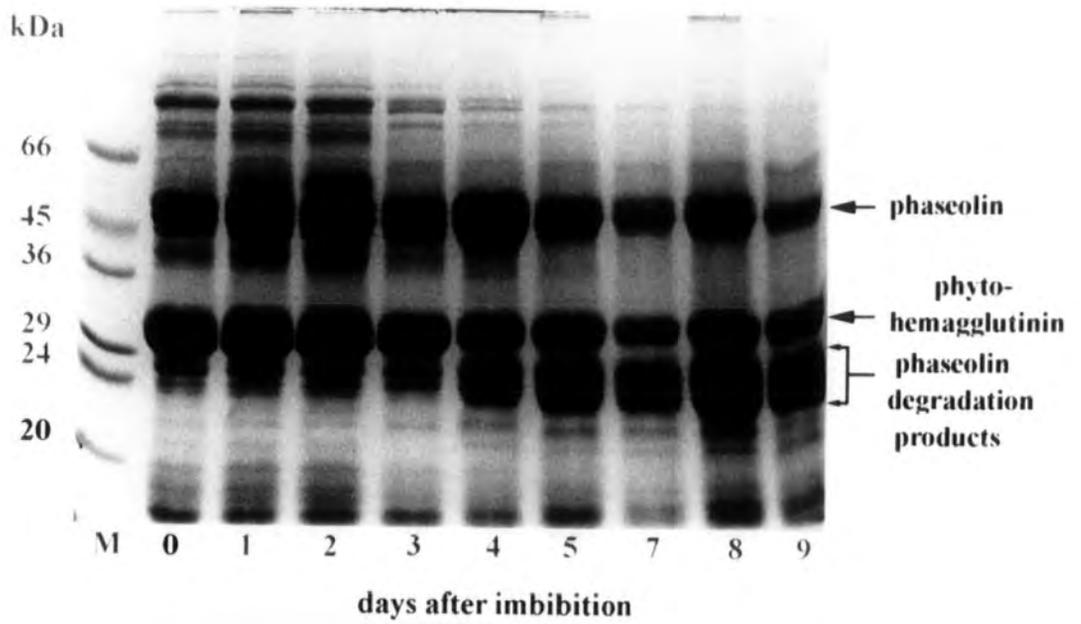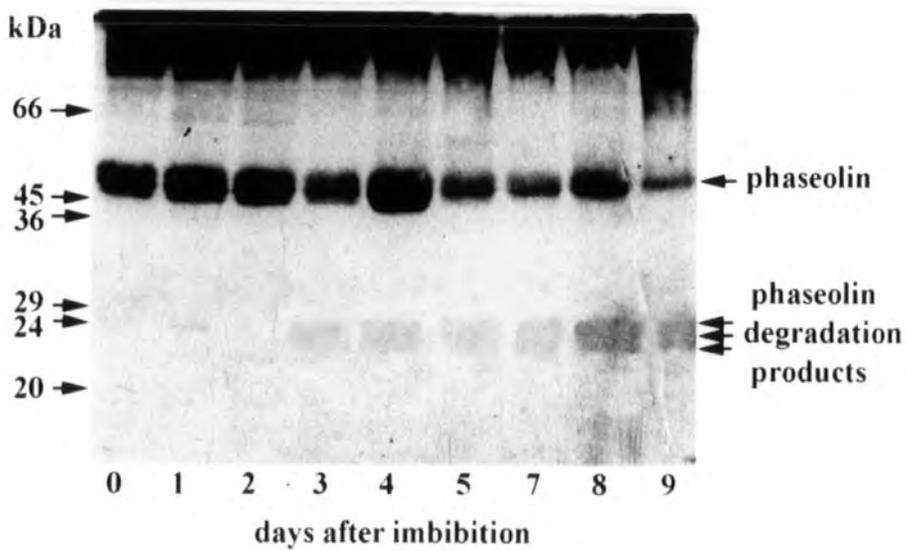
## 4.1.2 Isolation and identification of a superoxide dismutase from mature seeds of *P. vulgaris*

The methodology for fractionation and separation of seed proteins used for the purification of Glycoprotein I also proved useful for the isolation of other albumin proteins. A protein of 19 kDa in size was isolated by the same purification procedure as it was employed for Glycoprotein I (figure 4.1.3C: arrowhead).

N-terminal sequencing of the 19 kDa polypeptide identified this protein as a superoxide dismutase (SOD) with high sequence similarity to the N-terminal amino acids of SODs from monocotyledonous (maize, rice) as well as from dicotyledonous (tomato, tabacco, *Arabidopsis* and *Ipomea*) plants. The *P. vulgaris* seed SOD was designated as PVSOD. SODs localised in the cytosol of the plant cell were found to have the highest similarity to PVSOD. In figure 4.1.8 cytosolic SODs and SODs from other cell compartments were retrieved from the SWISSPROT or EMBL (Heidelberg) sequence libraries and aligned to PVSOD.

This is the first report of a protein sequence of a SOD in mature seeds.

| Plant species | Location | N-terminal protein sequence | Reference/ Accession |
|---|---|---|---|
| *Phaseolus vulgaris* | | VSAVAVLGSSEDVTA | this work |
| *Ipomea batatas* | cytosol | VKAVAVLSSSEGVSG | Q07796 |
| *Lycopersicon esculentum* | cytosol | VKAVAVLNSSEGVSG | P14830 |
| *Nicotiana plumbaginifolia* | cytosol | VKAVAVLSSSEGVSG | P27082 |
| *Oryza sativa* | cytosol | VKAVAVLASSEGVKG | P28757 |
| *Pinus sylvestris* | cytosol | GLLKAVVVLNGAADVKG | P24669 |
| *Spinacia oleracea* | chloroplast | ATKKAVAVLKGTSNVEG | P07505 |
| *Lycopersicon esculentum* | chloroplast | ATKKAVAVLKGNSNVEG | P14831 |
| *Citrullus vulgaris* | peroxisome | ATXXAVAVLXGTSNVEG | Bueno *et al.*, 1995 |
| *Nicotiana plumbaginifolia* | mitochondrion | HHQTYVTNYNKALEQ | G19693 |
| *Pisum sativum* | mitochondrion | HHQTYITNYNKALEQ | G20902 |

**Figure 4.1.8: Alignment of the N-terminal amino acid sequence of PVSOD with superoxide dismutases from other plants.**

Sequences were retrieved from either the EMBL database (Heidelberg) or the SWISSPROT sequence library and represent either N-terminal protein sequences (all with the exception of the two mitochondrial SODs) or cDNAs (both mitochondrial SODs). The N-terminal amino acid sequence of the peroxisomal SOD from *C. vulgaris* was taken from Bueno *et al.* (1995). X indicates unidentified residues.

### 4.1.3 Chitinases

#### 4.1.3.1 Isolation of two albumin proteins

During the isolation of Glycoprotein I two proteins of $M_r$ 25 kDa and 32 kDa were obtained together with Glycoprotein I in the same fractions after gel filtration (figure 4.1.2B; arrowheads). These proteins were however successfully separated from Glycoprotein I and PVSOD by ion-exchange chromatography (figure 4.1.3B). The two proteins were eluted at a low salt concentration (0.03M - 0.09M NaCl), so separating from Glycoprotein I which was eluted at a higher salt concentration (0.17 - 0.2M NaCl) in the elution buffer. Due to slightly different isoelectric points both polypeptides were eluted separately, the 32 kDa protein first indicating a higher pI than that of the 25 kDa protein.

#### 4.1.3.2 N-terminal sequencing of the two albumin proteins

To identify these proteins the N-terminal amino acid sequences of both polypeptides were determined by automated sequencing after separation by SDS-PAGE and blotting onto a PVFD membrane. The sequencing resulted in 20 determined residues for both polypeptides. The amino acid sequences are shown in table 4.1.2. The proteins were identified as chitinases by searching the EMBL protein database (Heidelberg) with the two N-terminal sequences.

| $M_r$ of Chitinase | N-terminal amino acid sequence |
|---|---|
| 25 kDa | DDVGSVISASLFEQLLKHRN |
| 32 kDa | WRDNAGIAVYWGQDAREGDL |

Table 4.1.2  N-terminal amino acid sequences of the two proteins identified as chitinases. For each of the polypeptides 20 residues from the N-terminus were identified by automated sequencing.

The alignment of both N-termini (figure 4.1.9) with chitinases showing high homology to the two isolated *Phaseolus* vulgaris chitinases, reveals that the French bean seed chitinases are mature chitinase proteins without signal peptides. Based on the level of similarity the 32 kDa protein (designated as "pvchi32") represents a class III chitinase, the 25 kDa chitinase ("pvchi25") can be regarded as a class II chitinase as judged by the high sequence

A)

Class III -
chitinases
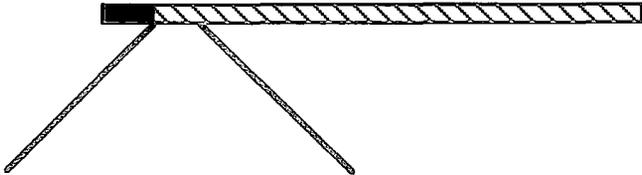
```
pvchI32:        WRDNA--GIAVYWGQDAREGDL
Cucumber:       SSDAA--GIAIYWGQNGNEGSL
Chickpea:       SSNAA--GIAVYWGQNGNEGSL
Arabidopsis:    PSDASRGGIAIYWGQNGNEGNL
```

B)

Class II -
chitinases
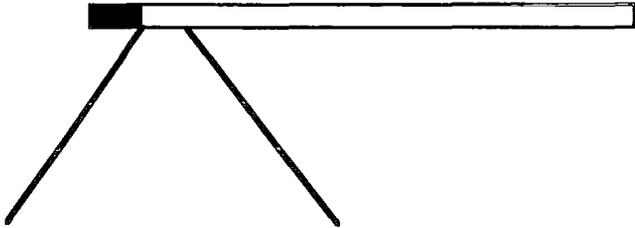
```
pvchI25:    DDVGSVISASLFEQLLKHRN
Potato:     GDIGSVISNSMFDQMLKHRN
Tobacco:    GDLGSIISSSMFDQMLKHRN
```

**Figure 4.1.9:** **Schematic illustration of the N-terminal amino acid sequences of pvchi32 and pvchi25 to class III and class II chitinases, respectively, from other plants.**

The chitinase sequences were retrieved from the SWISSPROT sequence library or from the EMBL database (Heidelberg) and represent either amino acid sequences or in amino acid sequences translated cDNAs. Accession-numbers: cucumber: P17541, chickpea: P36908, Arabidopsis: P19172, potato: U02608, tobacco: X51599.

66

similarity of the 20 determined N-terminal amino acids to class II chitinases from other plants and the lack of the cysteine-rich domain at the N-terminal end of the mature polypeptide which is characteristically for class I chitinases.

This is the first report of a class III chitinase in mature seeds.

### 4.1.3.3 Characterisation of *P. vulgaris* seed chitinases

#### A) Enzymatic activity

In order to confirm the enzymatic activity of the two chitinases in bean seeds, the enzymatic activity of the chitinases was tested. For this purpose a dye-labelled substrate for chitinases developed by Wirth and Wolf (1990) was used (section 3.2.4.2) in a qualitative assay. Both chitinases proved to be enzymatically active and hydrolysed the substrate (data not presented). The degree of activity was not determined by the method used.

#### B) Cross-reactivity with antibodies raised against bean leaf chitinases

Antibodies were available which had been raised against ethylene-induced *P. vulgaris* leaf chitinases.

To further confirm the nature of both proteins as chitinases and assess their relationship to the leaf enzymes, immunoblotting was performed with these antibodies. Figure 4.1.10 shows the result of the Western blot of protein fractions after each of the single isolation steps with bean chitinase antibodies. Both seed chitinases bound to the ethylene-induced anti-leaf chitinase antibodies, however pvchi32 showed less stronger binding than pvchi25. Pvchi32 was not detected in a total (SDS) protein extract and only as a very faint band in the total albumin extract. Both chitinases occur about in the same amount in the seeds of *P. vulgaris* as judged by the coomassie blue stained protein bands (e.g. 4.1.10A - lane 4) indicating that the reason for the different cross-reactivity lies in differences between bean seed chitinases and the bean leaf chitinases against which the antibodies were raised.

#### C) Seed chitinases during germination

It is suggested that the natural role of chitinases in seed tissue is the protection of the seed during dormancy and the beginning of germination and thus it might be expected that these chitinases are maintained during initial germination.

**Figure 4.1.10:  Western blot analysis of different stages of the purification of *P. vulgaris* seed chitinases.**

A) Aliquots (approximately 10-15 μg protein) of protein fractions from different stages of the chitinase purification were applied on a 15% (w/v) SDS-polyacrylamide gel, electrophoresed and stained with coomassie blue R250.

    M:  Standard proteins

    Lane 1:  Total (SDS) protein extract

    Lane 2:  Albumin proteins

    Lane 3:  60% (rel. sat.) fraction after ammonium sulphate precipitation of the albumin extract

    Lane 4:  Fraction 56 from gel filtration on Sephacryl S-300

B) The same protein fractions as in A) were separated on a second 15% (w/v) SDS-polyacrylamide gel and blotted onto nitrocellulose membrane. The blotted proteins were reacted with antibodies raised against ethylene-induced bean leaf chitinases (1:5000 dilution).

A)
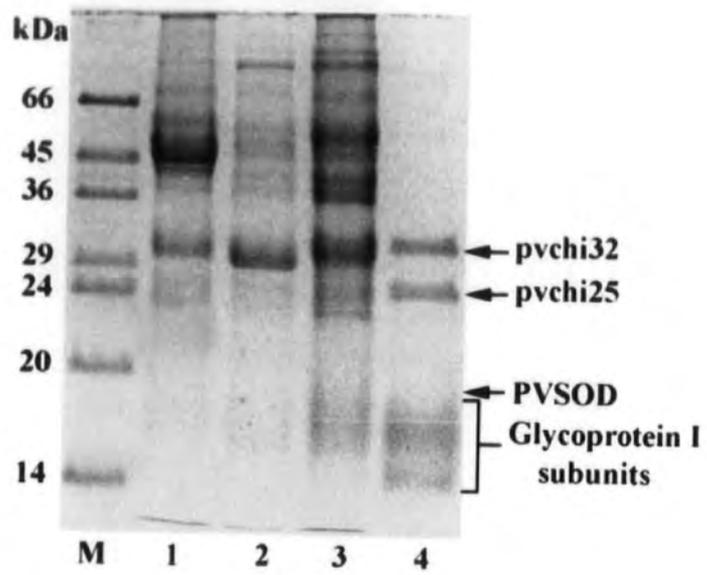
B)
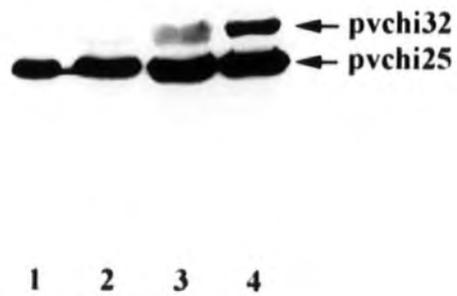
To investigate this hypothesis of the natural role of seed chitinases as antipathogenic protection during seed dormancy and early seed germination, the level of the chitinase proteins were observed during seed germination. For this purpose, seeds of *P. vulgaris* were germinated for periods of up to nine days and crude total albumin protein extractions prepared from seeds of each day of germination (see section 4.1.1.3). The presence of chitinases was tested for the first nine days of germination by immunoblotting using antibodies raised against *P. vulgaris* leaf-chitinases (section 4.1.3.3B). The results of this experiment are summarised in figure 4.1.11. Due to the low cross-reactivity of pvchi32 to the antibodies used, only pvchi25 could be observed during germination. The level of chitinase protein (pvchi25) decreased with increasing time after start of the germination of the seeds. 6-7 days after imbibition the amount of chitinase protein decreased to a level at which no cross-reactivity could be detected.

## D) Staining of the isolated chitinases for attached carbohydrates

In order to investigate whether the isolated chitinases were glycosylated aliquots of protein fractions from different stages of the protein purification were separated by SDS-PAGE and successively stained for glycoproteins with Schiff's reagent and for all proteins with coomassie blue R250. The result shows clearly that, in contrast to pvchi25, pvchi32 represents a glycoprotein (figure 4.1.4 - lane 4). Currently, only one chitinase, EP3, a 32 kDa protein from carrot embryo cell culture, has been described as a glycoprotein (De Jong *et al.*, 1992). Therefore, pvchi32 represents the second known glycosylated chitinase.

**Figure 4.1.11: Western blot analysis of *P. vulgaris* seed chitinases during seed germination**

Approximately equal amounts of albumin proteins (approximately $25\,\mu g$) from each day of seed germination (up to day nine) were separated on two 12% (w/v) SDS-polyacrylamide gels. One gel was coomassie blue stained, the second gel was blotted onto nitrocellulose and the blotted proteins were analysed by immunodetection with antiserum raised against ethylene-induced bean leaf chitinases. The antiserum was used as a 1:5000 dilution.

A)  shows the commassie blue stained polyacrylamide gel (M indicates the standard proteins).

B)  Resulting autoradiogram after hybridisation of the blotted membrane with anti-chitinase antibodies.

**A)**

kDa

66

45

36

29

24

20

M  0  1  2  3  4  5  7  8  9

days after imbibition

← phascolin

phyto-
hemagglutinin

phascolin
degradation
products

**B)**

pvchi25 →

0  1  2  3  4  5  7  8  9

days after imbibition
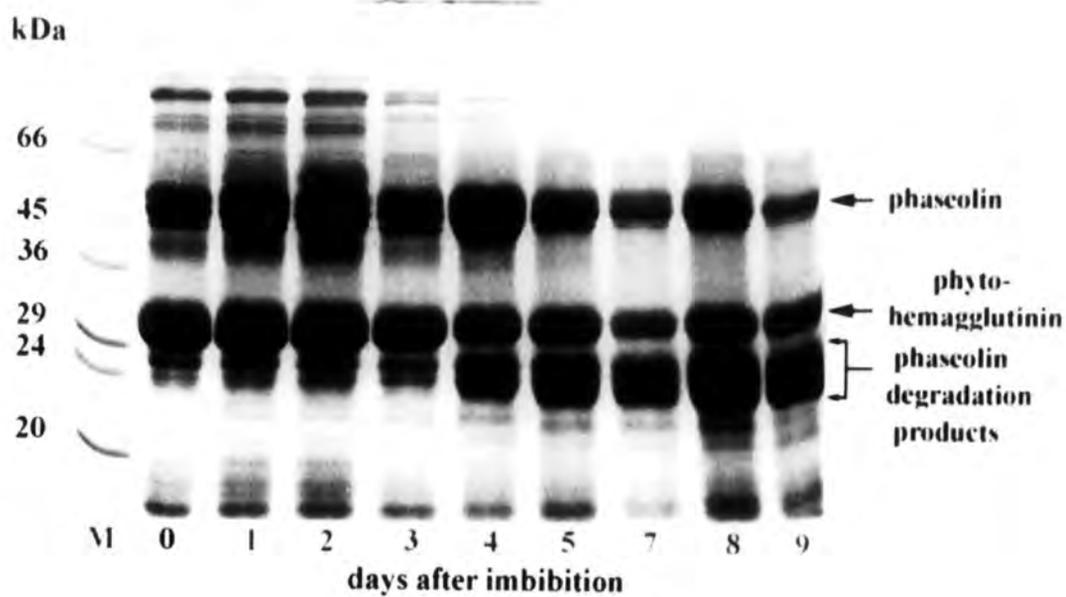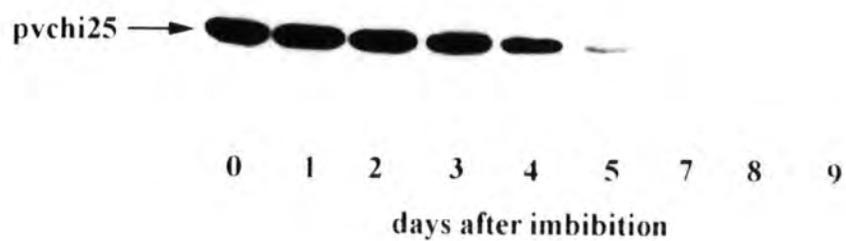
## 4.2 Legumin in the seeds of *Phaseolus vulgaris*

### 4.2.1 Fractionation of *P. vulgaris* seed proteins

Another major objective of this work was to try to confirm the presence of legumin as a storage protein in mature seeds of *P. vulgaris*. For this purpose a fractionation of the amount of seed proteins was necessary to identify proteins which may occur only in minor amounts.

Fractionation of the seed proteins from seeds of *P. vulgaris* was performed on a Sephacryl S-300 column. This experiment resulted in the separation of several high molecular weight proteins from the major storage proteins, phaseolins and phytohemagglutinins. Figure 4.2.1A shows the elution profile from the gel filtration. The protein containing fractions were analysed by SDS-PAGE on 12% (w/v) polyacrylamide gels under reducing and non-reducing conditions to identify proteins with disulphide -linked subunits. To avoid reduction of proteins in unreduced samples caused by 2-mercaptoethanol diffusing between lanes, a lane was left free between reduced and non-reduced protein samples. Figures 4.2.1B and C show the results of the SDS-PAGE analyses after silver staining of the gels. In order to apply approximately equal amounts of proteins, different volumes were applied in proportion to the amount of protein obtained for the single fractions as judged by the heights of the peaks (figure 4.2.1A). Phaseolins and phytohemagglutinins were obtained within the major protein elution peak (figure 4.2.1B - fractions 61 and 65) but were not present in the fractions of the first two peaks (figure 4.2.1B/C - fractions 38, 39, 41 and 49) indicating a successful separation of the two major seed protein fractions from proteins of higher $M_r$.

### 4.2.2 Identification of a legumin subunit

#### 4.2.2.1 Disulphide-linked subunits

Fraction 49 showed several protein bands on SDS-PAGE with $M_r$ ranging from 75 kDa to 55 kDa (figure 4.2.1C) in the absence of reducing agent. These bands are not present in the reduced protein sample and therefore represent potentially reduceable polypeptides. Several bands in the $M_r$ size range from about 54 kDa down to 35 kDa and a 21 kDa band appear in the reduced sample indicating to represent the breakdown products after the reduction of disulphide-linked polypeptides (fig 4.2.1C).

71

**Figure 4.2.1:** Gel filtration chromatography of a total seed protein extract from seeds of *P. vulgaris* on Sephacryl S-300.

A) Elution profile of the gel filtration chromatography on Sephacryl S-300.

A total seed protein extract was obtained using a Tris-Glycine extraction buffer (10mM Tris/HCl, pH 8.0; 80mM glycine; 500mM NaCl). The protein solution was dialysed for 48 h against a Tris buffer (0.1M Tris/HCl, pH 8.0; 0.1M NaCl) and 4 ml were applied onto the Sephacryl S-300 column (2.5cm x 110cm) equilibrated with the same Tris-buffer. Chromatography was performed at a flow rate of 0.3ml/min and fractions were collected every 10 minutes.

B)/C) SDS-PAGE analysis of gel filtration on Sephacryl S-300.

Protein containing fractions (compare fraction numbers of elution profile) were analysed by SDS-PAGE on 12% (w/v) SDS-polyacrylamide gels under reducing and non-reducing conditions. To avoid reduction of proteins in non-reduced samples by 2-mercaptoethanol diffusing between lanes, a lane was left free between reduced and non-reduced protein samples. In order to apply approximately equal amounts of proteins, different volumes of each fraction investigated were applied in proportion to the amount of protein per fraction as judged by the heights of the peaks (figure 4.2.1A): Fractions 38, 39, 41 and 49 12$\mu$l protein solution each; fraction 61 7$\mu$l and fraction 65 5$\mu$l protein solution each.

After electrophoresis the gels were silver stained. M indicates standard proteins.

**A)**

abs.

38  39 41  49  61  65

fractions →

**B)**

150 kDa →

−  +  +  −  −  +

kDa

66
45
36
29
24
20

41    41  61    61  65    65  M

**C)**

54 kDa    150 kDa

−  +  +  −  −  +

220 kDa →
150 kDa →

75 kDa →

kDa

66
45
36
29
24
20

49    49  39    39  38    38  M

21 kDa
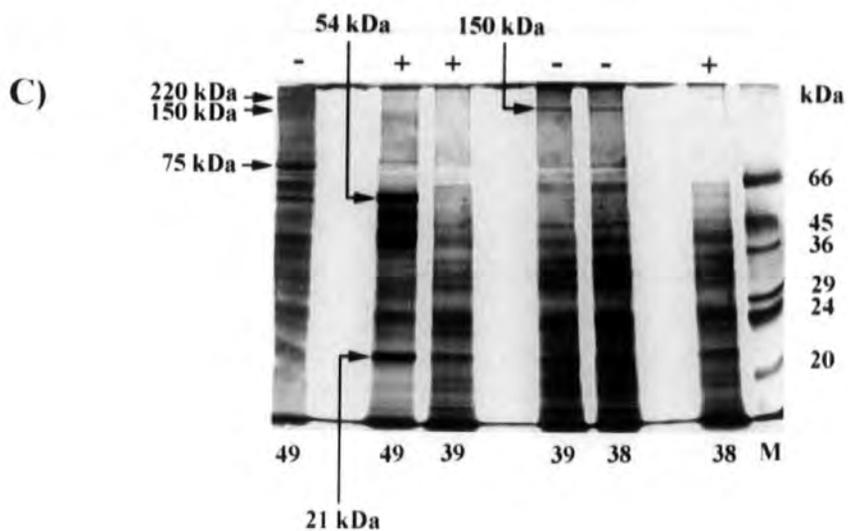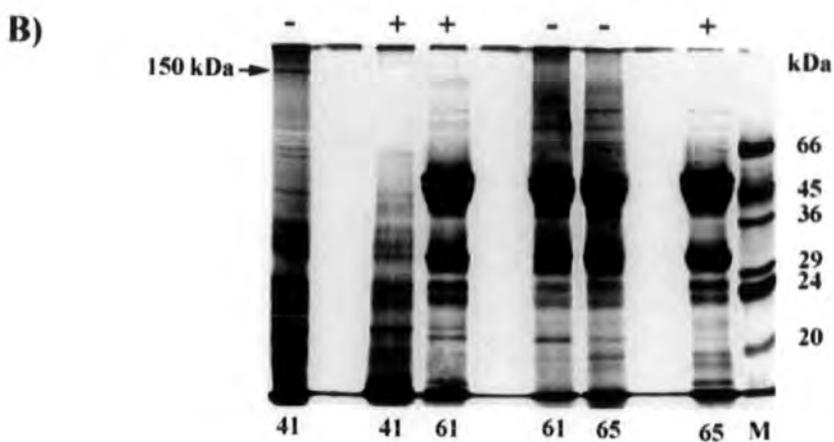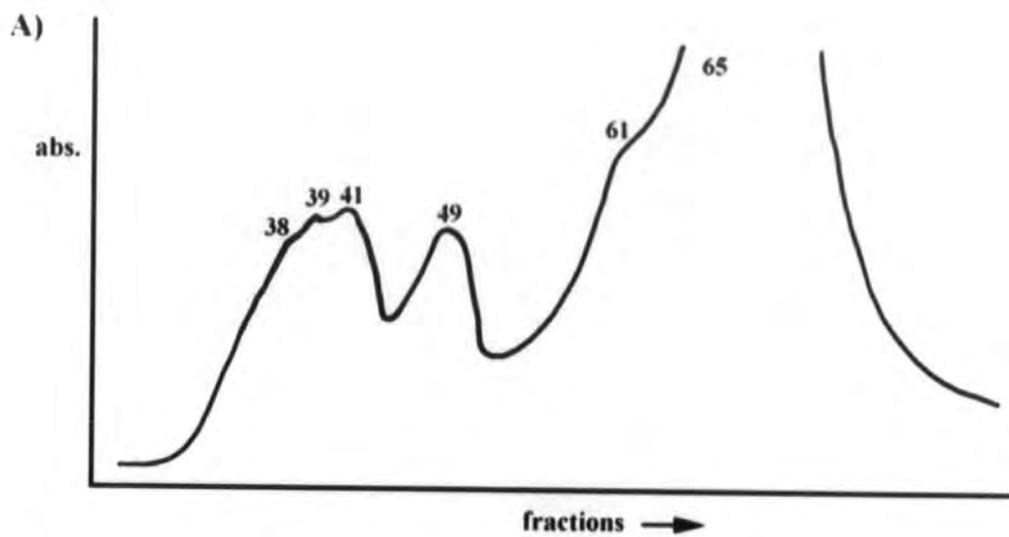
Protein bands of about 220 kDa and 150 kDa are present in fraction 49 (figure 4.2.1C) only under non-reducing conditions. The 150 kDa band appears also in the fractions 38, 39 and 41.

Fractions 61 to 65 show few differences between non-reduced and reduced protein samples and thus contain no recognisable disulphide-linked polypeptides.

In order to determine exactly which of the protein bands (55-75 kDa) in the non-reduced protein sample of fraction 49 shows legumin-like reduction to bands in the $M_r$ size range between 35 to 54 kDa and to the 21 kDa band (figure 4.2.1C), 2D SDS-PAGE analysis was employed. The first dimension was run under non-reducing conditions and the second dimension under reducing conditions as described previously (section 3.2.4.1). Fig. 4.2.2 indicates one example (fraction 49) of such an analysis. The results showed that there are about six putative α-subunits present in the size range from 54 kDa down to about 35 kDa (filled arrows) and only one size class of putative β-subunit (filled arrowhead), all of which arise from single unreduced proteins. The 150 kDa band described previously, also undergoes reduction into a potential single size class (21 kDa) of putative β-subunit (figure 4.2.2 - open arrow) but shows several different potential α-subunits distributed over the whole size range from 54 kDa to 35 kDa with single protein spots (figure 4.2.2; open arrowheads) at the same sizes as α-subunits arising from the other legumin-like proteins (figure 4.2.2: filled arrows). This behaviour was reproducible (data not presented), but the exact reason for it is unclear. The 220 kDa component did not run far enough into the polyacrylamide gel of the first dimension so that it may have been lost during the preparation for the second dimension (i.e. during the removal of the stacking gel). The exact nature of this component and its relation to the *P. vulgaris* legumin is unclear.

Based on the SDS-PAGE analyses and the elution profile from the gel filtration chromatography, fractions 36 to 43 and 44 to 50 were pooled and the globulin proteins were separated from other proteins by precipitation with 0.033M acetate buffer (pH 5.0). After dialysis against water and freeze drying, both globulin fractions were analysed on a 10% (w/v) polyacrylamide-SDS gel under reducing and non-reducing conditions (figure 4.2.3). Both peaks from the gel filtration contain several proteins of different size with apparent disulphide bonds. However, whereas the 150 kDa band appears to be a main component of fractions 36 to 43 (figure 4.2.3; lane 5 - open arrow), the major proteins of

**Figure 4.2.2: 2D SDS-PAGE analysis of fraction 49 after gel filtration of a total seed protein extract.**

Both dimensions of the 2D SDS-PAGE were performed on 12% (w/v) SDS-polyacrylamide gels. The first dimension was run under non-reducing, the second dimension under reducing conditions after incubation of a gel slice in 50mM DTT for 45 minutes at 40 °C. Filled arrows indicate putative α-subunit polypeptides and the filled arrowhead indicates a putative β-subunit polypeptide of a *P. vulgaris* legumin. Furthermore, the polypeptides arising from the 150 kDa protein band are indicated by open arrowheads (putative α-subunits) or an open arrow (putative β-subunits).
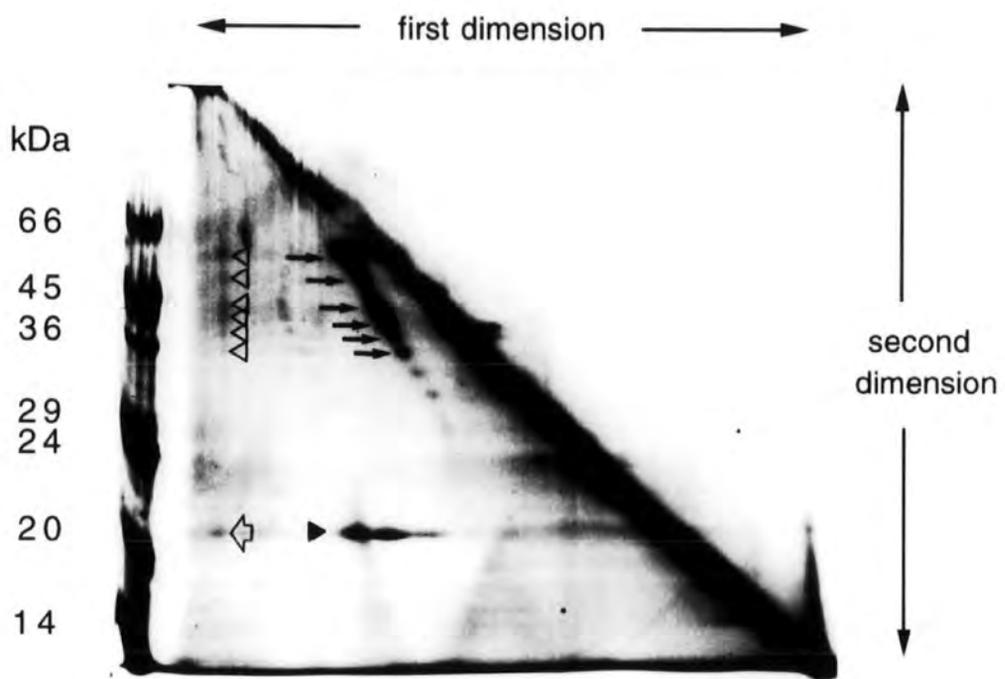
**Figure 4.2.3: SDS-PAGE analysis of disulphide-linked polypeptides in the globulin fraction of *P. vulgaris* seed proteins**

Fractions 44 to 50 (lanes 1-4) and fractions 36 to 43 (lane 5, 6) from the gel filtration chromatography of total seed proteins from *P. vulgaris* were pooled and globulin proteins were isolated by precipitation with 0.033M sodium acetate buffer (pH 5.0) from other proteins, dialysed against distilled water and freeze dried. Aliquots of both protein fractions (approximately 8-10 µg (lanes 1,3,5 and 6 or 15 µg (lanes 2 and 4)) were applied on a 10% (w/v) SDS-polyacrylamide gel and proteins were electrophoresed under reducing (+) and non-reducing (-) conditions. After electrophoresis the gel was stained with coomassie blue R250.

The polypeptides which were submitted for N-terminal amino acids sequencing are indicated by small arrows. Sequencing of the most upper of these protein bands (75 kDa) resulted in 20 determined residues for both $\alpha$- and $\beta$-subunit. The 150 kDa protein band is indicated by an arrowhead, the $\beta$-subunit polypeptide by a circle.
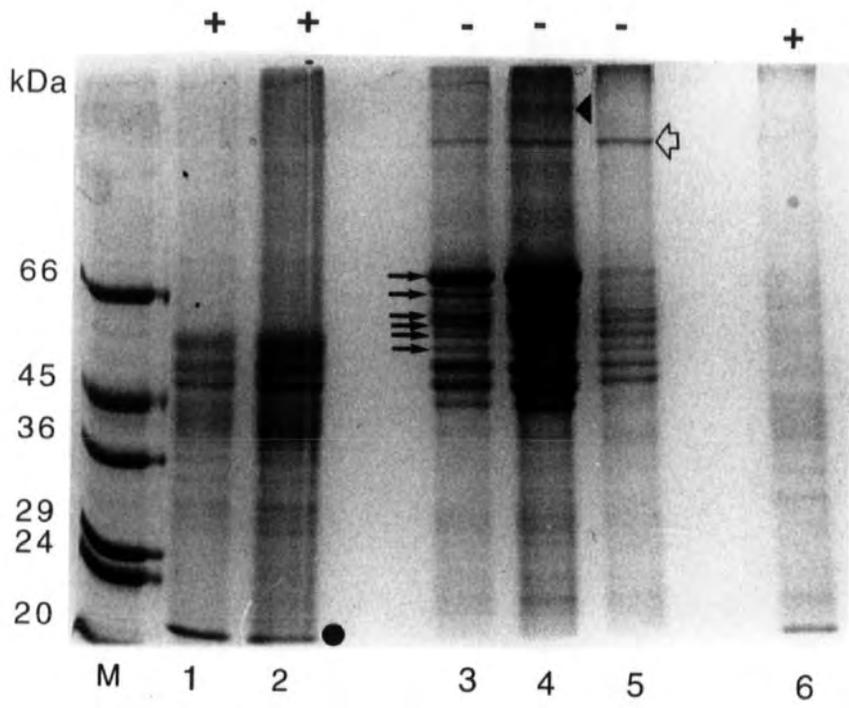
the fractions 44 to 50 are again proteins ranging in size from about 54 kDa down to about 35 kDa. The putative β-subunit is indicated by a circle. Furthermore, the 220 kDa band is in fractions 44 to 50 - compared with figure 4.2.1 - only visible as a faint protein band (figure 4.2.3: lane 4 - arrowhead) indicating that it may very well have been an artefact of the SDS-PAGE (see section 5.4.1.2).

### 4.2.2.2 Immunological cross reactivity to pea legumins

To test for similarity of the *P. vulgaris* disulphide-linked polypeptides to legumin from *Pisum sativum*, Western blot analysis using affinity-purified anti-*Pisum* legumin antibodies was performed. For this purpose reduced and non-reduced globulin protein samples from *P. vulgaris* as shown in figure 4.2.3 were separated on a 12% (w/v) SDS-polyacrylamide gel, blotted onto a nitrocellulose membrane and reacted with anti-*Pisum* legumin antibodies as described previously (section 3.2.4.3). No cross-reactivity with any of the *Phaseolus* disulphide-linked polypeptides was observed (data not presented).

### 4.2.2.3 N-terminal sequencing

To prove the identity of the disulphide-linked polypeptides as legumin-like protein, N-terminal sequencing of the 21 kDa putative β-subunit was performed. The 2-dimensional SDS-PAGE was carried out in a modified way, so that as much as possible purified protein of the separated legumin subunits could be obtained on the blot. The first dimension gel was run in vertical minigels (0.5 mm thick) under non-reducing conditions with a single well 4.5 cm in width. After the electrophoresis the gel was cut into a small horizontal section containing the 75 kDa polypeptide. This gel slice was used for performance of the second dimension. The second dimension involved reduction of disulphide bonds within the 75 kDa protein followed by electrophoretic separation of the resulting free polypeptides. For this purpose the gel slice was incubated in 50mM DTT at 40 °C for 45 min as described previously (section 3.2.4.1) and the gel slice was positioned horizontally on top of an 0.75 mm thick gel without a stacking gel and electrophoresed. The second dimension gel was then blotted onto a PVDF membrane and briefly stained (section 3.2.4.4).

The blot of the second dimension, prepared for the N-terminal protein sequencing, is shown in fig. 4.2.4. The 21 kDa band (arrow) excised from the filter yielded an N-terminal sequence of 20 amino acids (table 4.2.1).

**Figure 4.2.4:  Modified 2D SDS-PAGE for separation of disulphide-linked legumin subunit polypeptides for N-terminal amino acids determination**

The 2D SDS-PAGE was performed on 12% (w/v) SDS-polyacrylamide gels in vertical minigels with two tracks each, one (0.5 cm broad) for the standard proteins and one (7 cm broad) for the protein sample. After the first dimension (non-reducing) separation, a horizontal gel slice containing the 75 kDa protein band was excised over the whole gel width and incubated in 50mM dithiothreitol at 40 °C for 45 min to reduce disulphide-linked polypeptides. The gel slice was laid horizontally on top of a 0.75 mm thick separating gel without a stacking gel and electrophoresed as before. The second dimension gel was blotted onto PVDF membrane and proteins were located by briefly staining the membrane with 0.1% (w/v) coomassie blue R250 in 40% (v/v) MeOH, 1% (v/v) acetic acid followed by destaining in 50% (v/v) aqueous MeOH solution.

The 21 kDa band (arrow) was excised from the filter for sequencing and yielded an N-terminal sequence of 20 identified amino acids (table 4.2.1) showing high similarity to the N-terminus of the $\beta$-subunit of legume legumins. Probably due to N-terminal blocking no sequence information could be obtained from the putative legumin $\alpha$-subunit (54 kDa) of the same blot (arrowhead).
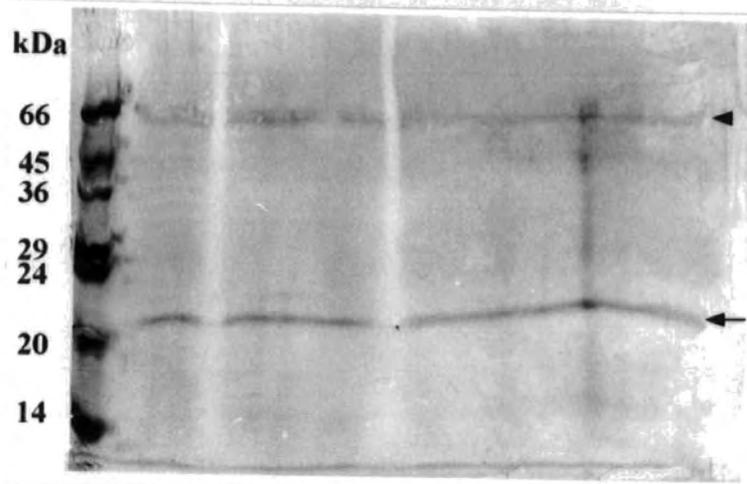
Unfortunately due most probably to N-terminal blocking no sequence information could be obtained from the putative α-subunit (54 kDa) of the same blot (4.2.4: arrowhead).

N-terminal sequencing of the putative α-subunits was achieved using the 75 kDa band from the unreduced protein sample of the pooled fractions 44 to 50 (separated as in figure 4.2.3: lane 3 - most upper arrow). Under these conditions both the α- and the β-subunits were isolated in a single protein band from the PVDF-membrane blot. Sequencing of this band yielded N-terminal sequences from both subunits simultaneously, two amino acids were obtained in each sequencing cycle. The β-subunit amino acid sequence obtained from the 21 kDa band after reduction of the 75 kDa protein (figure 4.2.4: arrow) was in agreement with one of the two amino acids sequences provided at each sequencing cycle. The α-subunit was therefore determined by subtraction of the N-terminal sequence derived for the β-subunit from the total sequence derived.

The first six amino acids of each of the unreduced proteins of lower molecular weight (figure 4.2.3: lane 3 - lower 5 arrows) were also determined by automated N-terminal sequencing. These all showed exactly the same protein sequences as determined previously for the α- and β-subunits (table 4.2.1).

| $M_r$ of polypeptide | Subunit-type | N-terminal amino acid sequence |
|---|---|---|
| 21 kDa | β | TTNLFNRCRINELNALKPDH |
| 54 kDa | α | GIEETLCTLKLQHNIARASS |
| 51 kDa, 47 kDa, 43 kDa, 37 kDa and 35 kDa | α β | TTNLFN GIEETL |

Table 4.2.1  N-terminal amino acid sequences of the α- and β-subunits of the identified *Phaseolus* legumin.

In figure 4.2.5 the partial sequences of both α- and β-subunits are aligned to legumin sequences retrieved from the SWISSPROT database. These amino acids sequences were obtained as the result of a sequence search with the N-terminal legumin sequences from French bean.

78

**Figure 4.2.5:** Alignment of the N-terminal sequences of the *P. vulgaris* legumin subunits with other legume legumins.

The obtained N-terminal amino acid sequences of the *P. vulgaris* legumin $\alpha$-subunit and the $\beta$-subunit (B) were manually aligned to the N-terminal sequences of other legume legumin subunits.

Bold letters indicate highly conserved residues, * = positions showing strictly conserved residues, + = positions showing conservative residue replacement. The protein sequences were retrieved from the SWISSPROT sequence library.

A)

| Species | α - subunit | accession-no. |
|---|---|---|
| *Phaseolus vulgaris* | TTNLFNRCRINELNALKPDH | —— |
| Glycinin 1 | FSSREQPQQNECQIQKLNALKPDN | P04776 |
| Glycinin 2 | LREQAQQNECQIQKLNALKPDN | P04405 |
| Glycinin 3 | FREQPQQNECQIQRLNALKPDN | P11828 |
| Glycinin 4 | ISSSKLNECQLNNLNALEPDH | P02858 |
| Glycinin 5 | ITSSKFNECQLNNLNALEPDH | P04347 |
| *Pisum sativum* LegA2 | LREQPQQNECQLERLNALEPDN | X17193 |
| *Pisum sativum* Leg J | TSSEFDRLNQCQLDSINALEPDH | X07014 |
| *Vicia faba* Leg 4 | TSSEFDRLNQCRLDNINALEPDH | X03677 |
|  | * * + +*** ** |  |


B)

| Species | α - subunit | accession-no. |
|---|---|---|
| *Phaseolus vulgaris* | GIEETLCTLKLQHNIARASS | —— |
| Glycinin 1 | GIDETICTMRLRHNIGQTSS | P04776 |
| Glycinin 2 | GIDETICTMRLRHNIGQTSS | P04405 |
| Glycinin 3 | GIDETICTMRLRHNIGQTSS | P11828 |
| Glycinin 4 | GVEENICTLKLHENIARPSR | P02858 |
| Glycinin 5 | GVEENICTMKLHENIARPSR | P04347 |
| *Pisum sativum* LegA2 | GLEETVCTAKLRLNIGPSSS | X17193 |
| *Pisum sativum* Leg J | GLEETICSAKIRENIADAAR | X07014 |
| *Vicia faba* Leg 4 | GLEETICSLKIRENIAQPAR | X03677 |
|  | *++* +*++++ ** |  |

This sequence comparison confirms beyond doubt that the N-terminal sequences of both α- and β-subunits were derived from legumin-type proteins and that they represent a unique legumin protein. Furthermore, this represents the first sequence description of a legumin protein from *P. vulgaris*.

### 4.2.3 *Phaseolus* legumin is a glycoprotein

Some legumin proteins have been reported to be glycosylated (e.g. from *Lupinus albus* (Duranti *et al.*, 1988, 1995); *Magnolia salicifolia* (Fischer *et al.*, 1995)). To investigate the possible glycosylation of *Phaseolus* legumin, Schiff's staining of the partially purified protein (sample from fractions 44-50; figure 4.2.3) was performed followed by coomassie blue staining. The results (figure 4.2.6) indicate that Phaseolus legumin is a glycoprotein. Two legumin subunits of $M_r$ 72 and 75 kDa (figure 4.2.6A: lane 2 - arrows) were stained by the Schiff's reagent as well as the 150 kDa subunit polypeptide (arrowhead). In the reduced protein sample (figure 4.2.6A: lane 1) a 39 kDa polypeptide was stained by the PAS method. This polypeptide was also present in the non-reduced protein sample (figure 4.2.6A: lane 2) and does therefore not represent a legumin subunit. As the β-subunit did not be stained by the Schiff's reagent (see figure 4.2.6A: lane 1) but was stained by the coomassie blue stain as a strong polypeptide band (figure 4.2.6B: lane 1 - arrowhead) it is suggested that the α-chain of the legumin subunit polypeptides contains the glycosylation site. Similar results have been observed for the *Magnolia*- and *Lupinus*- glycosylation-sites (Duranti *et al.*, 1988, 1995; Fischer *et al.*, 1995). Most probably due to too low an amount of protein no α-subunits of the reduced protein sample were visible after PAS staining (figure 4.2.6A: lane 1).

### 4.2.4 Immunological cross-reactivity to anti-germin antibodies

Recently several reports have been published on the origin and evolution of seed storage proteins (e.g. Bäumlein *et al.*, 1995, Braun *et al.*, 1996, Shewry *et al.*, 1995). Major efforts were put into classifying the different seed storage proteins in groups. These classifications of seed storage proteins were based on similarities of structurally important sequence stretches. One of the hypotheses described in these papers was a report proposing common

**Figure 4.2.6: PAS staining of *P. vulgaris* legumin subunits**

Aliquots (approximately 7-15 μg) of the partially purified legumin protein after gel filtration chromatography were dissolved in reducing (+) or non-reducing (-) sample buffer and applied onto a 10% (w/v) SDS-polyacrylamide gel. After electrophoresis the proteins were stained for attached oligosaccharides using PAS staining. After scanning the gel image (A), the SDS-polyacrylamide gel was stained with coomassie blue R250 (B). M indicates the lanes with the standard protein marker.

A)  PAS staining of two legumin subunits in the unreduced protein sample (lane 2 - arrows). These two protein bands are not present in the reduced protein sample, nor is the 150 kDa band (lane 2 - arrowhead), which is also stained by the Schiff's reagent. No other legumin subunit polypeptides were stained in the reduced protein sample (lane 1): probably due to too low an amount of protein.

B)  Coomassie blue R250 staining of the same gel as shown in A). Besides the two legumin subunits and the 150 kDa protein band in the non-reduced protein sample some minor contaminating proteins are present in the legumin fraction (lane 2). Whereas two α-subunit polypeptides appear after coomassie blue staining as faint protein bands in the reduced protein sample (lane 1: arrows), the β-subunit appears as strong protein band after reduction of disulphide-bonds (lane 1 - arrowhead).

A)
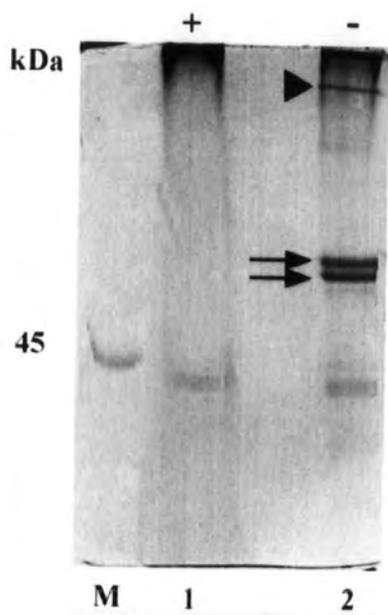
B)
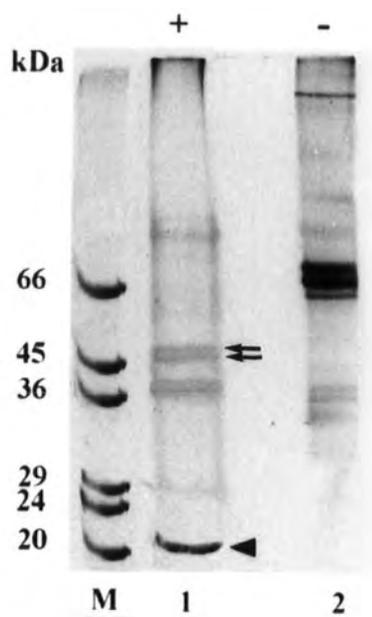
ancestry of 7S and 11S globulin seed proteins with germin proteins from wheat (Bäumlein *et al.*, 1995). To investigate, as to whether there exist some structural homologies in terms of common epitopes between germin proteins and vicilins and legumins a Westernblot analysis of the seed proteins from *Pisum sativum* and *P. vulgaris* were performed using available affinity-purified antibodies raised against germin-like proteins isolated from wheat.

To find out whether there is any cross-reactivity between germin and seed storage proteins from legume species, a dot blot analysis was performed as described previously (section 3.2.4.3) using different amounts of total seed protein extract from *P. sativum* and *P. vulgaris*, so investigating both 11S globulins (main seed proteins in pea) and 7S globulins (main seed proteins in French bean). The result (figure 4.2.7A) showed strong cross-reactivity of the French bean seed proteins with the anti-germin antibodies. Only the lower amounts of the French bean seed proteins reacted with the anti-germin antibodies, confirming a highly specificity of the antibody-protein binding. The pea seed proteins did not show any cross-reactivity with the anti-germin antibodies (figure 4.2.7A).

To investigate which of the *P. vulgaris* seed proteins bind to the anti-germin antibodies, the French bean seed proteins were separated on a 12% (w/v) SDS-polyacrylamide gel, blotted onto a nitrucellulose membrane and analysed by immunoblotting using the anti-germin antibodies as described previously (section 3.2.4.3). The result (figure 4.2.7B) revealed, that the main cross-reactivity was caused by reaction with phytohemagglutinin. Two of the phaseolin bands showed binding to the antibodies as well as a 66 kDa polypeptide and several polyeptides in the size range between 70 and 100 kDa.

**Figure 4.2.7: Analysis of immunological cross-reactivity of total protein extracts from seeds of *P. vulgaris* and *Pisum sativum* with an anti-germin antiserum**

Total proteins were extracted from seeds of *P. vulgaris* and *Pisum sativum* using a Tris-Glycine extraction buffer (10mM Tris/HCl, pH 8.0; 80mM glycine; 500mM NaCl).
Different amounts of seed proteins were directly blotted onto nitrocellulose membrane by the dot blot technique (A). For the Western blot analysis of seed proteins from *P. vulgaris* a total protein extract was separated on a 12% (w/v) SDS-polyacrylamide gel and the proteins were blotted onto NC filter by the semidry blotting technique (B). Proteins on the NC-membranes (dot blot and Western blot) were analysed by immunodetection with anti-germin antiserum in a 1:10,000 dilution.

A) Autoradiography film of the dot blot after immunodetection with ant-germin antiserum. Whereas *Pisum* seed proteins do not show any cross-reactivity to antiserum, the *Phaseolus* seed proteins show a high cross-reactivity with the antibodies.

B) Autoradiography film of different amounts of electrophoretically separated *Phaseolus* seed proteins (approximately 8 μg (lane 1), 15 μg (lane 2) 20 μg (lane 3)) after immunodetection with anti-germin antiserum. M indcates the standard proteins and Co the positive control (approximately 0.5 μg of purified germin protein).

**A)**

| | 20 | 15 | 10 | 7.5 | 5.0 | 2.5 | 1.0 | 0.75 | 0.5 | 0.2.5 | µg protein |

Control

*Pisum sativum*

*Phaseolus vulgaris*

**B)**

kDa

66→

45→

36→

29→
24→

20→

Co   1   2   3

← phaseolin

← phyto-
hemagglutinin
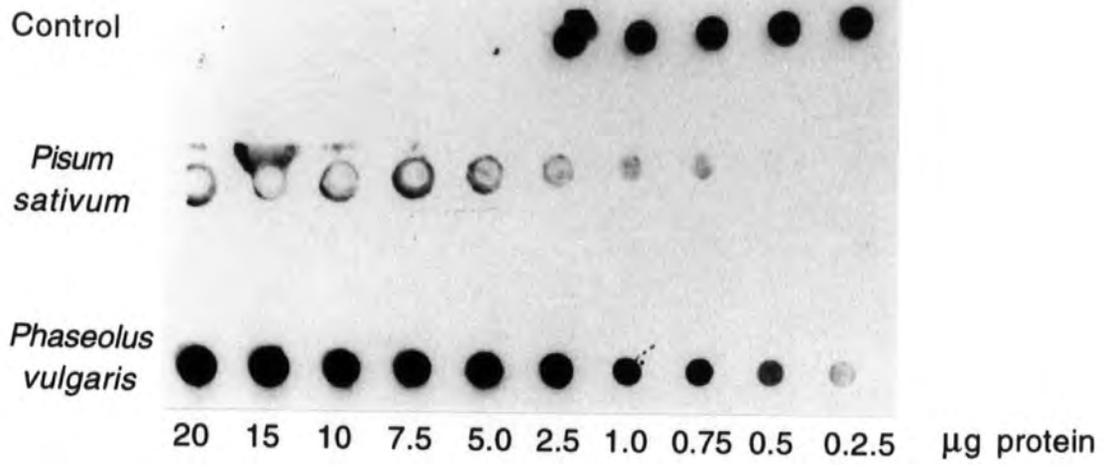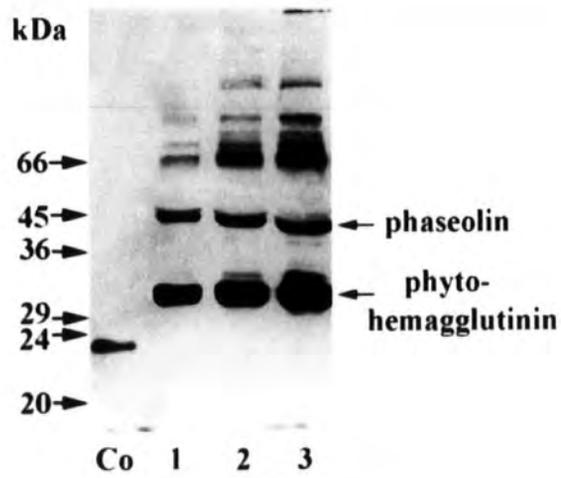
### 4.2.5 PCR-strategy for the search of legumin-encoding genes

Polymerase Chain Reaction (PCR) was employed in an extensive series of experiments designed to obtain detailed sequence information on legumin genes in *P. vulgaris*. This approach was expected to provide much more sequence information which, when translated into a protein sequence, would then help to characterise the whole legumin protein and allow an analysis of its potential protein structure. Furthermore, the amplified PCR-fragment(s) could be used as homologous probes to investigate the legumin gene family of French bean by performing genomic Southern analyses. Finally, the nucleotide sequences derived could be used as phylogenetic markers for the elucidation of the taxonomic position of *P. vulgaris*.

### 4.2.5.1 Selection of oligonucleotides

The selection of the primers for the PCR was based on extensive sequence alignments of both protein- and nucleotide sequences. The alignments (appendix E) were made using MegAlign subprogram of the DNAStar software package (Lasergene, Ltd) (see section 3.3.12.2). The nucleotide sequences were retrieved from the EMBL sequence library (Heidelberg) and protein sequences from the SWISSPROT data base.

The search for the most appropriate positions for the oligonucleotide primer pair within the legumin genes was identified based on the protein alignment followed by analysis of these regions within the translated nucleotide sequence alignment. Finally, the selected primers were tested for hairpin-structures, primer-dimers and primer-self-dimers using the subprogram PrimerSelect of the DNAStar software package (Lasergene, Ltd). Furthermore, the oligonucleotides selected for the PCRs were checked for other potential binding sites within the legumin genes by simulating polymerase chain reactions by computer using the Amplify software as described previously (section 3.3.12.2). Furthermore, the expected sizes of the PCR products were also analysed using the Amplify program and legumin gene sequences as template for the computer-based PCR simulations.

<u>Selection of primer-pair 1F/1R</u>

The search for a primer-pair resulted in the following oligonucleotides:

| Primer | Position | Oligonucleotide sequence (5' to 3' direction) |
|---|---|---|
| Primer 1F | 499-515 | ARCWYYRMWAACCAGCT |
| Primer 1R | 1270-1287 | KYTRTSRTKKGTCTTGAA |

Table 4.2.2  First primer pair (1F/1R) used in this work. The IUB codes (appendix D) are used for primer positions with degeneracy. Primer 1F represents the forward primer (sense strand) and primer 1R the reverse primer (antisense strand; translated in the opposite strand). Positions are relative to the glycinin 1 cDNA sequence (accession number P04776).

The calculations for the highest and lowest annealing temperatures (section 3.3.7.3) revealed the following values:

| | Highest annealing temperature | Lowest annealing temperature | Length | Degeneracy |
|---|---|---|---|---|
| Primer 1F | 54 °C | 44 °C | 17 bp | 128-fold |
| Primer 1R | 56 °C | 44 °C | 18 bp | 128-fold |

Table 4.2.3  Characteristics of the first primer pair (primer 1F/1R) used in this work.

Selection of primer-pair 2F/2R

As the PCRs of genomic DNA from *P. vulgaris* using the first pair of oligonucleotide primers did not result in amplifications of any legumin gene fragments (section 4.2.5.2A) a second oligonucleotide set was synthesised. In order to overcome problems concerning the ligation of potential PCR products for cloning, oligonucleotide linkers with EcoRI restriction sites were synthesised at the 5' ends of the specific legumin sequences. At the 5' end of the restriction site 5 non-specific nucleotides were added to displace the restriction site to allow more efficient restriction. Considering the different legume sequences, the following degenerate oligonucleotides were designed as primers.

| Primer | Position | Oligonucleotide sequence (5' to 3' direction) |
|--------|----------|-----------------------------------------------|
| Primer 2F | 133-156 | TTATC<u>GAATTC</u>GAAGSTGGDYTCAYTGAGACITGG |
| Primer 2R | 1270-1292 | TTATC<u>GAATTC</u>SYAGSTCTRTCATTKGTCTTGAA |

Table 4.2.4   Second primer-pair used in this work. The IUB codes (appendix D) are used for primer positions with degeneracy. EcoRI restriction sites are underlined. Primer 2F represents the forward primer (sense strand) and primer 2R the reverse primer (antisense strand; translated in the opposite strand). Positions are relative to the glycinin 1 cDNA sequence (accession number X15121).

The calculations for the highest and lowest annealing temperatures revealed the following values:

| | Highest annealing temperature | Lowest annealing temperature | Length (with EcoRI restriction site) | Degeneracy |
|--|--|--|--|--|
| Primer 2F | 74 °C | 68 °C | 35 bp | 24-fold |
| Primer 2R | 68 °C | 62 °C | 34 bp | 32-fold |

Table 4.2.5   Characteristics of the second primer pair (2F/2R) used in this work. The restriction site and the five additional 5' nucleotides as well as the inosine nucleotide replacement of primer 2F were not included in this calculation.

Computer simulations of the polymerase chain reaction using the Amplify software package (section 3.3.12.2) with different legume legumin encoding genes as template and with the primer-pairs 1F/1R and 2F/2R revealed that a successful amplified pea legumin J encoding gene was expected to be between 0.9 - 1.2 kb (primer pair 1F/1R) or 1.4 - 1.8 kb (primer pair 2F/2R) in size (data not shown). The positions and the distances between the primers within the genomic DNA sequence of legumin A from pea (accession-number: X02982; Lycett *et al.*, 1984) are illustrated in figure 4.2.8. The exact primer positions are further marked in the legumin alignments in appendix E (1F/1R; red boxed; 2F/2R: blue boxed).

5' NT                                                                                          3' NT

2F                                                                              2R

← · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·→
1.6 kb

1F                                                                              1R

← · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·→
1.1 kb

intron                                        Eco RI oligo-
                                              nucleotide linker

exon
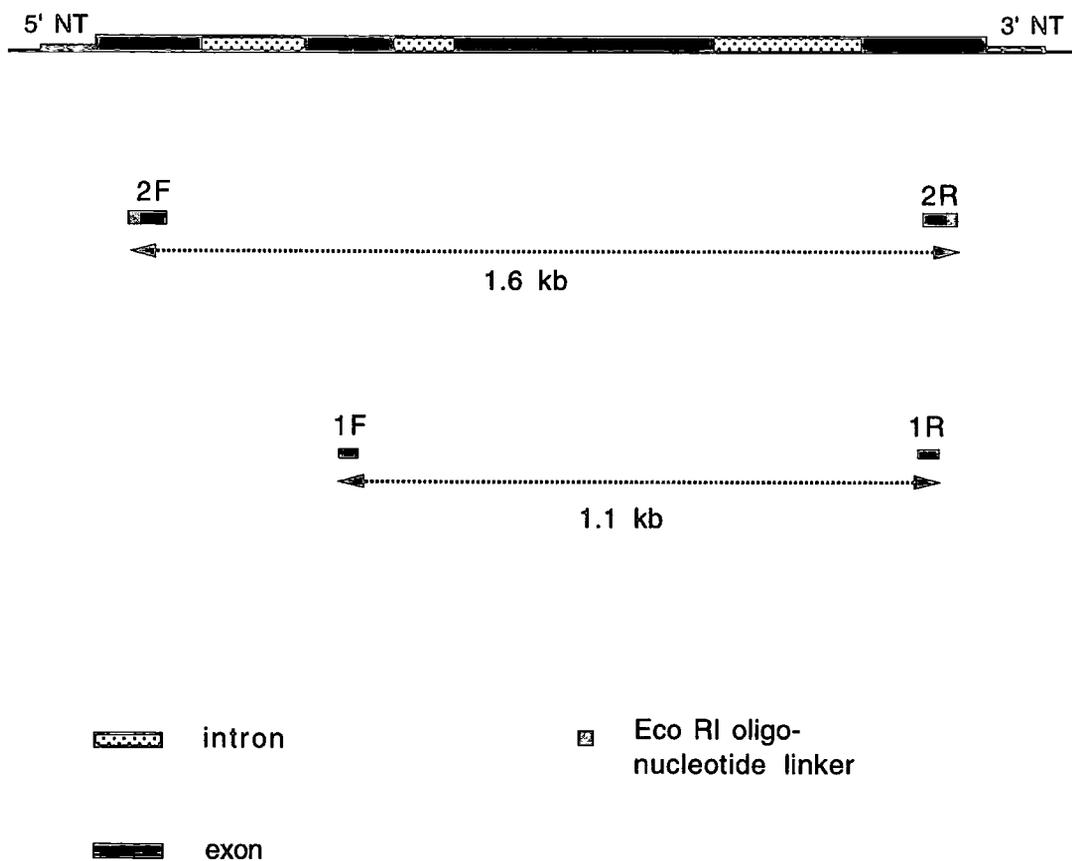
**Figure 4.2.8:** **Schematic illustration of the positions of the primer pairs within the legumin gene.**

The schematic legumin gene corresponds to legumin A from pea (accession-no. X02982; Lycett *et al.*, 1984). 5'-NT and 3'-NT: 5'- and 3'- non-translated regions.

### 4.2.5.2 Polymerase Chain Reaction

#### A) PCR using primer pair 1F//1R

The primers were first checked for their function by performing a positive control using pea genomic DNA as the template. This PCR performed under standard conditions (see below) led to the amplification of two major fragments of about 1kb in size similar to those obtained for the hot-start PCR (compare figure 4.2.10B: lanes 3 and 4). This result agreed with the sizes of the PCR fragments as it was predicted by PCR simulations using the Amplify program (section 4.2.5.1).

Standard-PCR

PCR with genomic DNA from *P. vulgaris* was carried out under a series of different conditions in order to evaluate the best parameters for the PCR. However, only very small amounts of PCR products were obtained using a standard PCR protocol (data not shown). Attempts to increase the amounts of PCR products of a specific DNA band using re-PCR (section 3.3.7.4) did not lead to a sufficient increase of amount of PCR product for cloning. Furthermore, this approach includes the danger of selecting a wrong PCR fragment as template for the re-PCR when several PCR products are obtained within the expected size range of the PCR products. Therefore, another strategy was chosen to attempt to amplify the legumin genes.

Low-/high-annealing temperature combination

A second method for amplification of genomic plant-DNA using degenerate oligonucleotides was employed (C. Illet, E. Broomfield, pers. communication). After the initial denaturation step at 94 °C for 4 minutes, 3 PCR cycles were carried out at low annealing temperature (35 °C) followed by 35 cycles of a relatively high annealing temperature (60 °C) taking into account the calculated annealing temperature of the primers. This approach resulted in amplification of a clear 700 bp product and seemingly minor unspecific products (figure 4.2.9A). Although this approach did not lead to any major PCR product using pea genomic DNA as template in a positive-control experiment

**Figure 4.2.9: Results of the PCR and re-PCR using a low-/high-annealing temperature PCR technique and primer pair 1F/1R**

Low-/high-annealing temperature PCR technique was used in order to amplify legumin encoding gene fragments from genomic DNA of *P. vulgaris*. 3 PCR cycles were performed at low annealing temperature (35 °C) followed by 35 cycles at a relatively high annealing temperature (60 °C).

A) Results of the PCR.

The PCRs were performed under the same conditions except the following differences:

Lane 1:  2mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

Lane 2:  4mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

Lane 3:  2mM $Mg^{2+}$ and 250 ng genomic DNA from *P. vulgaris*

Lane 4:  4mM $Mg^{2+}$ and 250 ng genomic DNA from *P. vulgaris*

Lane 5:  2mM $Mg^{2+}$ and 100 ng genomic DNA from *Pisum sativum*

Lane 6:  4mM $Mg^{2+}$ and 100 ng genomic DNA from *Pisum sativum*

Lane 7:  Negative control - no addition of any genomic DNA to the PCR mix

10 µl of each of the PCRs were applied on a 1% (w/v) agarose gel and separated electrophoretically. M indicates the standard DNA marker (λ-DNA digested with PstI).

B) The DNA of the major PCR products of the low-/high-annealing temperature PCR (A: arrowhead) was isolated from the agarose gel and an aliquot (approximately 10 ng) was used as template for re-amplification. Lane 1 shows the result of the re-PCR. 7 µl (out of 50 µl PCR volume) were applied on a 1% (w/v) agarose gel and separated electrophoretically (lane 1). Lane 2 shows the negative control (no addition of template DNA for the re-PCR).

Sequencing of the cloned 700 bp re-PCR product revealed that it does not represent an amplified legumin gene fragment.

A)

11.5 —
5.1 —
4.8 —
4.5 —
2.8 —
2.5 —
2.1 —
2.0 —
1.7 —
1.2 —
1.1 —
0.8 —
0.5 —
0.4 —

M    1    2    3    4    5    6    7    M

B)

11.5 —
5.1 —
4.8 —
4.6 —
2.8 —
2.5 —
2.1 —
2.0 —
1.7 —
1.2 —
1.1 —
0.8 —
0.5 —
0.4 —

M    1    2
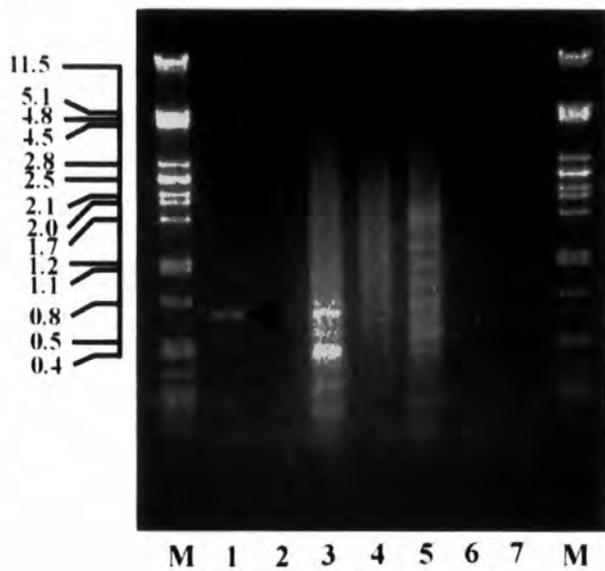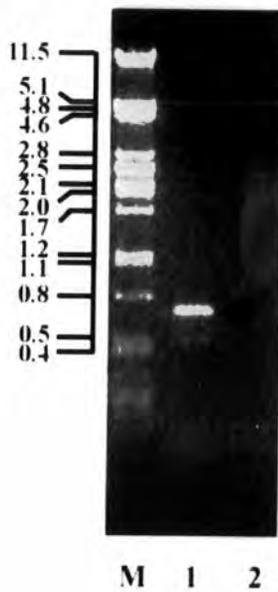
(figure 4.2.9A: lanes 5,6), the 700 bp PCR fragment amplified from *P. vulgaris* genomic DNA was further investigated. Re-PCR was carried out at an annealing temperature of 60 °C and using the 700 bp PCR fragment (figure 4.2.9A-lane 1: arrowhead) isolated from the agarose gel as template resulted in the amplification of sufficient PCR-product of the 700 bp fragment for cloning (figure 4.2.9B: lane 1 - arrowhead). Sequencing of the 700 bp fragment cloned into pGEM T-vector was performed from one site of the cloned fragment using the T7 Primer as sequencing primer and resulted in a nucleotide sequence of 400 bp. Comparison of the longest open-reading frame of this sequence with different sequence databases revealed homology to a phenol hydroxylase from *Pisum sativum* (accession-number P19732). The analysis of this seemingly unspecific amplified 700 bp PCR product revealed, that this approach to amplify legumin genes from genomic DNA of *P. vulgaris* did not lead to a specific amplification of legumin genes.

## Hot-Start PCR

High unspecificity and low specific product yield can occur in a polymerase chain reaction when reaction components are mixed at low temperature allowing non-specific primer annealing and extension (Chou *et al.*, 1992; Horton *et al.*, 1994). To overcome this problem which caused the misprimed products when used the low-/high-annealing temperature combination, hot-start-PCR technique was employed. In hot-start PCRs at least one essential reagent is misheld from the PCR mixture until the system has reached a temperature that favours specific primer annealing. Hot-start PCR can greatly improve sensitivity, specificity and yield of a PCR (Horton *et al.*, 1994). Hot-start PCR can be achieved by either mechanically separating a reagent (Horton *et al.*, 1994) or by initial blocking of the polymerase with an antibody (Kellog *et al.*, 1994). In this work the first technique was employed using paraffin wax tablets (BDH) to separate one primer from the template in the reaction mixture (section 3.3.7.4).

This approach led finally to the amplification of two major PCR products of 1 kb and 700 bp in size and several seemingly unspecific PCR products of smaller and bigger size (figure 4.2.10A). To avoid a repeat of the cloning of misprimed PCR products (see above)

**Figure 4.2.10:** Southern analysis of the PCR products of the hot-start PCR using primer pair 1F/1R

Hot-start PCR strategy was employed in order to avoid unspecific annealing of the oligo-nucleotides to the genomic DNA of *P. vulgaris*.

A) Results of the PCR.

   The PCRs were performed under the same conditions except the following differences:

   Lane 1: 2mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

   Lane 2: 4mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

   Lane 3: 2mM $Mg^{2+}$ and 100 ng genomic DNA from *Pisum sativum*

   Lane 4: 4mM $Mg^{2+}$ and 100 ng genomic DNA from *Pisum sativum*

   Lane 5: Negative control - no addition of any genomic DNA to the PCR mix

   10 µl of each of the PCR volume were applied on a 1% (w/v) agarose gel and separated electrophoretically. M indicates the standard DNA marker (λ-DNA digested with PstI).

B) Southern analysis of the PCR products of the hot-start PCR using legumin A from pea as a probe. The exposure time of the X-ray film to the membrane was 1 hour.

**A)**



11.5
5.1
4.8
4.6
2.88
2.5
2.1
2.0
1.7
1.2
1.1

M 1 2 3 4 5 M

**B)**



M 1 2 3 4 5 M
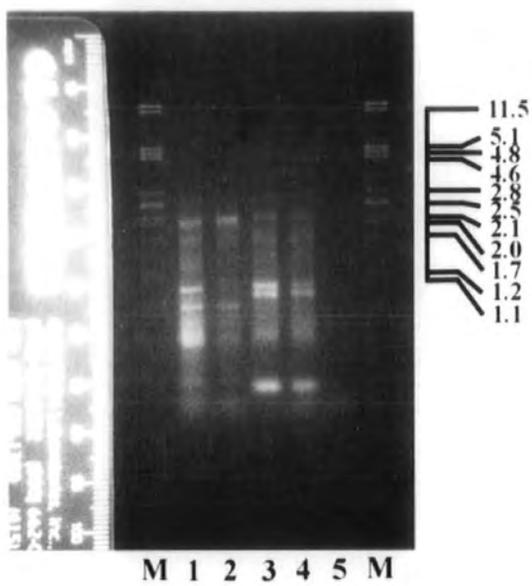
Southern analysis of the PCR-products was performed using a homologous legumin probe (legumin A from pea, section 3.3.8.1). The analysis confirmed, that the PCR products amplified in a positive-control experiment from genomic DNA of *Pisum sativum* were genuine legumin-encoding gene fragments, but none of the PCR products amplified from genomic DNA of *P. vulgaris* showed any positive hybridisation signals (figure 4.2.10B). Therefore, no attempt has been made to clone any of these PCR products.

## B) PCR using primer pair 2F/2R

Optimising the PCRs using the second primer-pair (2F/2R) it was found again that hot-start PCRs lead to the best results (data not shown). Therefore, only this PCR type was further used for the amplification of gene fragments from *P. vulgaris* genomic DNA in connection with the primer set 2F/2R.

Finally, the PCRs proved to give the largest yield of amplified product when amplifications were performed in 25 µl reaction volumes containing 50mM KCl; 10mM Tris-HCl, pH 9.0; 0.1% Triton X-100, 2mM MgCl$_2$, 1.6µM of each of the respective primers, 0.2mM of each dNTP, 1.5 units Taq polymerase and 150 ng of genomic DNA from *P. vulgaris*. Variations in the amount of genomic DNA from *Pisum sativum* (50-300 ng genomic DNA in 25 µl reation volumes) did not appear to play any role in the amplification of pea legumin genes in control experiments, so confirming the specificity of the primer pair 2F/2R (data not shown).

The cycling conditions in this PCR were the following: After initial template denaturing at 94 °C for 5 minutes, 33 PCR cycles were carried out with a cycling program of 1 min denaturation at 94 °C, 90 sec annealing at 53 °C and 90 sec extension at 72 °C. Finally, an additional polymerisation step was performed at 72 °C for 5 min.

10 µl of the PCR products were electrophoretically separated on a 1% (w/v) agarose gel, the gel photographed over UV-light (figure 4.2.11A) and then blotted onto Hybond-N$^+$ nylon membrane as described previously (section 3.3.8.2). The Southern analysis of this blot using the pea legA gene as probe revealed two strong signals in the size range of 2 kb (figure 4.2.11B), indicating that this PCR represented a successful amplification of a putative legumin encoding gene fragment from genomic DNA of *P. vulgaris*. This was

**Figure 4.2.11: PCR with pimer pair 2F/2R and Southern analysis of the PCR products**

A) Results of the PCR.

The PCRs were performed under the same conditions except the following differences:

Lane 1: 2mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

Lane 2: 4mM $Mg^{2+}$ and 100 ng genomic DNA from *P. vulgaris*

Lane 3: 2mM $Mg^{2+}$ and 200 ng genomic DNA from *P. vulgaris*

Lane 4: 4mM $Mg^{2+}$ and 200 ng genomic DNA from *P. vulgaris*

Lane 5: 2mM $Mg^{2+}$ and 300 ng genomic DNA from *P. vulgaris*

Lane 6: 4mM $Mg^{2+}$ and 300 ng genomic DNA from *P. vulgaris*

10 µl of each of the PCR volume were applied on a 1% (w/v) agarose gel and separated electrophoretically. M indicates the standard DNA marker ($\lambda$-DNA digested with PstI).

B) Southern analysis of the PCR products of the hot start PCR with primer pair 2F/2R using legumin A from pea as a probe. The exposure time of the X-ray film to the membrane was 30 minutes. The positive two signals in lane 4 derive from hybridisation of the probe to two PCR products indicated with arrows in A) - lane 4. The PCR products responsible for the signals in lane 6 on the agarose gel photograph (A) are not visible, probably due to too low an amount of DNA.

**A)**



M 1 2 3 4 5 6 M

11.5
5.1
4.6
2.1
2.0
1.7
1.2
1.1
0.8
0.5
0.4

**B)**



2.1 kb
1.7 kb

M 1 2 3 4 5 6 M
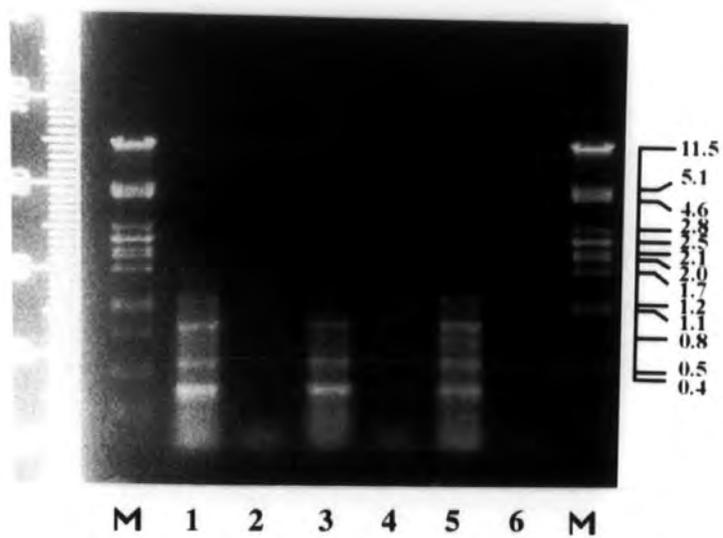
further confirmed by the hybridisation stringency of 60 °C. The final wash stringency of the post-hybridisation washes was 0.5x SSC at 60 °C (section 3.3.8.3). However, the two PCR products which hybridised to the legumin probe were present only in very minor amounts (figure 4.2.11A-lane 4: arrows; PCR products of this size are not visible in lane 6). Therefore, it was planned to firstly produce more PCR product by re-PCR and secondly to clone all PCR products in the size range of about 1.5 kb to 2.2 kb and finally to screen the transformations for putative PCR-amplified legumin gene fragments using colony screening. For this purpose, the remaining 10 μl of the original PCR were electrophoretically separated and all PCR products in the size range between 1.5 and 2.2 kb were isolated using electro-elution as described previously (section 3.3.5). An aliquot of the isolated PCR fragments was then used for a re-PCR carried out as described previously (section 3.3.7.4). The rest of the isolated DNA was digested with EcoR I to produce restriction sites for the ligation into pKS$^+$ plasmid digested with Eco R I (section 3.3.9.1). However, the re-PCR did not provide any further PCR products and neither did subsequent transformation produce any transformed E.coli colonies.

To obtain again specific PCR products the original PCR was repeated many times in the same (25 μl) and in larger (50 μl and 100 μl) reaction volumes. None of these repeats of the PCR followed by Southern analysis of the PCR products led to the amplification of these two DNA bands which hybridised to the legumin probe (data not shown). Further experiments concerning the optimising of the PCR were undertaken including changes in annealing temperature, extension time, final polymerisation step, magnesium concentration and the addition of more detergents (Triton X-100) and other additives (DMSO). Also newly isolated genomic DNA from P. vulgaris used as template did not lead to any better results. The use of a variety of different types of Taq-polymerases (data not shown) did not improve the specificity and yield of the PCR amplifications. Some results obtained from all these repeats of the PCR together with the result from the Southern analysis using pea legumin A as probe (section 3.3.8.1) is shown in figure 4.2.12 representing the PCR experiment in which the addition of additives (Triton X-100, DMSO) was tested. This result was achieved using the same amounts and concentrations of the PCR components as described for the original PCR (25 μl reaction volume containing 50mM KCl; 10mM Tris-HCl, pH 9.0; 0.1% Triton X-100, 2mM MgCl$_2$, 1.6μM of each of the respective primers, 0.2mM of each dNTP, 1.5 units Taq polymerase

A)  Results of the PCR.

The PCRs were performed under the same conditions except the following differences:

Lane 1-4:  2mM $Mg^{2+}$ and Triton X-100 as detergent

(lane 1: 5% (v/v), 2: 2.5% (v/v), 3: 1% (v/v), 4: no additional detergent)

Lane 5-8:  2mM $Mg^{2+}$ and DMSO as additive

(lane 1: 5% (v/v), 2: 2.5% (v/v), 3: 1% (v/v), 4: no additional detergent)

Lane 8:     Positive control PCR with 100 ng genomic DNA from pea

10 µl of each of the PCR volumes were applied on a 1% (w/v) agarose gel and separated electrophoretically. M indicates the standard DNA marker (λ-DNA digested with Pst I).

B)  Southern analysis of the PCR products using legumin A from pea as a probe. The X-ray film was exposed overnight to the membrane. The positive signals in lane 2 indicated with arrowheads were isolated and sequenced directly. The 1.65 kb fragment indicated with an arrow in lane 6 was ligated into pGEM T-vector and sequenced.

A)

B)

(Promega) and 150 ng of genomic DNA) and the addition of DMSO or Triton X-100 to the reaction mix of the hot-start PCR (figure 4.2.12A), which was carried out with an initial denaturation step at 94 °C for 4 minutes, followed by 33 cycles of 1 min denaturing at 94 °C, 53 °C annealing for 2 min and 2 min 2.15 min extension at 72 °C. The final polymerisation step was performed at 72 °C for 10 min. These conditions led to major PCR products of 400 bp, 600 bp, 900 bp and 1.65 kb in size (figure 4.2.12A). The Southern analysis however revealed again that the two PCR-products in the size range of about 2 kb were missing. The four major PCR products showed unspecific hybridisation to the legumin probe (figure 4.2.12B). This might be due to the large amount of these PCR products and due to the long exposure time of the X-ray film to the membrane (overnight).

Despite the failure of the amplification of PCR products which hybridised with high specificity to the legumin probe used, it was finally decided to investigate the nature of the PCR products of 600 bp, 900 bp (figure 4.2.12A/B: arrowheads) and 1.65 kb (figure 4.2.12A/B: arrow) in size by cloning and sequencing of these fragments. The PCR fragments of 600 bp and 900 bp in size were sequenced to investigate their identity. To avoid the frequent problems with the cloning of PCR fragments direct sequencing of the PCR fragments was performed using primer 2F as sequencing primer. In the case that these fragments should be of further interest, the whole sequence could be obtained using primer 2R as sequencing primer, since about 500 bp can be expected to be sequenced by such direct sequencing (Stanchev, pers. communication). The 1.65 kb fragment was cloned into the pGEM T-vector as described previously (section 3.3.9.2) as the attempt failed to clone the PCR fragment after digestion with EcoRI into pKS$^+$ vector digested with EcoRI.

The sequence database search revealed for the 600 bp and the 900 bp PCR fragments no sequence similarity to any of known sequences.

The forward and reverse sequencing (ca. 700 bp from each site) of the cloned 1.65 kb PCR fragment revealed, that this product arose due to a PCR artefact, as only primer 2F was found to have participated in the polymerase chain reaction. This PCR artefact however proved to be fortuitous, because - as is discussed later (section 5.4.3) - the sequence was found to be partially highly similar at the nucleotide level to arcelin genes which up until now were thought to exist only in wild varieties of *P. vulgaris*. However, due to no major open reading frame of the nucleotide sequence of this PCR artefact no similarity to arcelin

proteins could be determined at the amino acid level. Figure 4.2.13 shows the obtained partial sequence of the 1.65 kb PCR artefact. To this sequence arcelin 2 (John and Long, 1990; accession-number: M28470) is aligned where sequence similarity exists. The missing 200 bp between the two sequence stretches were not investigated further.

| Size of PCR product | Directly sequenced | Cloned into T-vector and sequenced | Sequence similarity |
|---|---|---|---|
| 600 bp | yes | —— | no significant homology to any of the sequences in the EMBL sequence library (Heidelberg) |
| 900 bp | yes | —— | no significant homology to any of the sequences in the EMBL sequence library (Heidelberg) |
| 1650 bp | —— | yes | homology to arcelin 2 (lectin-like protein) isolated from wild type *P. vulgaris* (accession number: M28470) |

Table 4.2.6  Sequencing of three PCR products. In order to identify the PCR products which hybridised to the pea legumin A probe, the PCR products were either direct sequenced or cloned into T-vector, sequenced and identified via database search.

A)


1.65 kb PCR-artefact: <u>TTATCGAATTCGAAGGTGGGCTCATTGAGACGTGG</u>ATATCACG     44
Arcelin A:


TTCACAAGGGAAGATCTCAGGGACATTGTGCCTCATGACAACGATCCAATAGTCATCTCGCTTGT     110


TACGACGGGAAGGAAGGTCCACAGAGTACTAGTGGACCAAGGAAGCTCGGCAGATGTGATGTTCT     176


GGCCGACTTTCACACAGTTGCAGTTGCCC-CTT-GACCAGTTGAGGCCC-TA-TGGAGGGTGTTT     234
            CAG---CCAACTACGACC-GC-GACGCCCATACTGT-GGCTGTG-     424


GTATGGTTTCGCTGGCGATCAGGTGGAGGTCAGGGGGTATATAAAATTGA-G--AACT--A-CC-     292
GT--G-TT-CGA---C-ACC-G-TC-AG--CAACCG-TAT-TGAAATCGACGTGAACTCCATCCG     473


---T-TC--A-CGGA-T---G-AAG--CGGGCTCAAGAACGGAGAAAAT-CA-AG-TACC--TCA     338
GCCTATCGCAACGGAGTCTTGCAATTTCGGCCACAACAACGGAGAAAAGGCCGAGGTTCGGATCA     538


--T-TGTGAACGCTCCCTCAGCATATAAC-ATTCTGTTG-GGAAGACCAACACTCAACAGAATAG     398
CCTAT-T-A-C--TCCC-C--CA-AGAACGA--CT-T-GAGG--GTTT--CTCTG--CTT--TAC     582


GAGCTGTGCCTTCGACCAAGCACATGAAGGTGA-A-GT-T--GCCATCCA-TGGAGG-GGGTGTG     451
C--CT-T-C-TTCGG--AAG-A-A--AAG-TGCCACGTCTCTGCCA-C-AGTGCCGCTGG-----     628


TGATCACCATCCAATCTGACCAGAAAGAAGC-GAAGAA-G--TG--C-TATGAA------GA-A-     506
-------------------AGAAAGAAGTTGAGGACTGGGTGAGCGT-TGGGTTCTCTGCCAC     670


-T-AGCCT-GAAGAACAAGAGTCCG-TGAGTTACATCACAACGACGCTGCCTCCCGTGTTGAAG-     567
CTCAGGGTCGAA-AA-AAGAGACCACTGAA--ACG-CACAACGTC-CT-C-TC---T-T-G--GT     721


C-----CAAGCCAACAG-ATCACGGTTGTTG-GTGCACGATGAGT-TGCCGTCCGGGGAAACTTC     605
CTTTTTCTT-CCAACTTCATCAA--TT-TTGAGGGCAAAAA-A-TCT                   762


CCCCATGAATGATGCTGGGGCTGAAGGAAAAAAAACCCAA                          646

B)

1.65 kb PCR-artefact: <u>TTATCGAATTCGAAGGTGGGTTCATTGAGACGTGGG</u>ATGCATCT    44
Arcelin A:


T-TA-TC----T-GG-T-T-GTACCC---CG-AGAAGGC----ATCCAGGA------AACTCAGC    85
TCTCGTCCCCGTCGGCTCTCGG-CCCAAACTTA-AAGGCCGTTATCTAGGTCTTTTCAACACAGC    389


-AACTTAC-ACCCTGCAGCACTGTCAACCAGGGCGT-CTATG-CTTG-GCAAAG-GATACGA-AT    143
CAACT-ACGACC--GC-G-AC-G-C--CCA-----TACTGTGGCT-GTG----GTGTT-CGACA-    433


CCTTTGGCAAGCCTTGTTCAGATCAGTGAAATCGACGC-A-C---ATGCG-CCATTT-GC--C--    197
CCGTCAGCAA-CC--GT----AT---TGAAATCGACGTGAACTCCATCCGGCC-TATCGCAACGG    487


ATTGCTTTTC--TT-CA-CCAGCACGACATTGGCTAGCCA---CTCAG-GGTACTGGACCTC-CC    253
AGT-CTTG-CAATTTCGGCCA-CA--ACAACGG--AGAAAAGGC-C-GAGGTTC-GGA--TCACC    540


TGAT-A-TGGCCTGCAGCGAG-GAGCTTCTGGGTTTCATCC-CTGATCACCTGTCTTCTCTCCTC    314
T-ATTACTCCCC--CAA-GAACGA-CTTGAGGGTTTC-TCTGCTT-T-ACC---CT           585


GTTGAATTTCCTTCTTCTCTGTCGCACTGGTCTGACCTGGTTGTCCATTGCTAGATGATGGCACA    380


AGAAGTCGGGGTCAATTCCCGGCATGTCTGAAGCAGACCATGCGAAAGCATCCAGATGCCGCTCT    446


ATCACCTTGGCAATCTGGTCTTGGAGCTCAGCCTCCAAGGATCTTCCCAACTTGAAGACCTTTCC    512


CCTGATCTCCCTTTCGAGCCATTCCTCGACGGGTTTTGGTCTACTTCCTTGGCGATAACCGCTCT    578


GGCGATTCCCAGCTCCCTCGCTTCCTCAGCCGGTTCCTCGCCTCTTCTTCCCGCTCGGCTTTCTC    644


TCCCTCACCCCAGCATCCATCATGGCGACGTCTCCCCCCGACGGCACTCCATCGCTGCACCAACA    700


AACCGTGACTCTGTTGGCTTGGGCTTCACACC                                    732

# 5. Discussion

## 5.1 Glycoprotein I

A great deal of biochemical work has been done in the past on seed storage proteins in leguminous plants including *P. vulgaris* L. as it represents an important crop plant. Pusztai and co-workers (e.g. Pusztai, 1966, 1968; Pusztai and Watt, 1970; Pusztai and Duncan, 1971) in particular described in a series of papers details of the structural properties of several of the major seed proteins of French bean. The authors put much effort into two glycoproteins from the seeds of *P. vulgaris* which were thoroughly investigated on a biochemical basis and which - based on the glycosylated nature of the proteins - were called Glycoprotein I and Glycoprotein II. Whereas Glycoprotein II which subsequently was identified as a vicilin-type storage protein is the well characterised major storage protein phaseolin (7S globulin), no more work has been done on Glycoprotein I during the last 19 years. The only experiments on Glycoprotein I within this time period were undertaken by Croy (1977), who employed SDS-PAGE for the analysis of seed proteins. According to Croy's work, Glycoprotein I is built up of two subunits of $M_r$ 14 kDa and 17 kDa forming a tetrameric 60 kDa protein composed of equal amounts of the two subunits, though the methods used could not discriminate between homomeric and heteromeric proteins. It was further estimated, that Glycoprotein I is only present in amounts of about 1-2% (w/w) of total proteins in the seeds and was described as an albumin type protein (best soluble at pH 5.0 at very low ionic strength) and had high contents of aspartic and glutamic acid and serine, as well as of aromatic amino acids, but was low in sulphur-containing amino acids (Pusztai, 1966). The carbohydrate moiety consisted mainly of mannose and glucosamine (Pusztai, 1966). Besides its physical and chemical characterisation the most interesting property of Glycoprotein I was its apparent autodigestion at pH 9.0-10.0 leading to the formation of new amino groups and an irreversible loss of its native conformation (Pusztai, 1968). The nature of this reaction led Pusztai to the suggestion of a limited proteolytic degradation. However, he could not determine any enzymatic activity for Glycoprotein I responsible for this breakdown (Pusztai, 1968).

In a subsequent attempt to investigate whether the observed proteolytic activity resulted from a minor contamination by a proteolytically active enzyme or whether it is indeed an

intrinsic property of Glycoprotein I, Pusztai and Duncan (1971) separated the isolated Glycoprotein I band from the DEAE-cellulose chromatography into two distinguishable fractions. Whereas the first fraction of the Glycoprotein I peak eluted from the ion-exchange column showed a high activity in hydrolysing α-N-benzoyl-L-arginine p-nitroanilide (BAPNA), a characteristic substrate for proteolytic enzymes such as trypsin, only a very weak activity (100-fold less) was observed for the main part of the Glycoprotein I peak, indicating that the first part of the Glycoprotein I peak contained a contaminating enzyme but the main part was nearly free of this contaminant. Further investigation of the separated main part of the Glycoprotein I fraction revealed, that the autodigestion of Glycoprotein I is indeed an intrinsic property of the protein (Pusztai and Duncan, 1971). The exact nature of this activity was however unknown and no amino acid sequence data was obtained.

Besides the biochemical properties of Glycoprotein I, not much information is available about this protein. Therefore, one aim of the present work was the characterisation and possible identification of Glycoprotein I through (N-terminal) protein sequencing of each of the two constituent subunits described by (Croy, 1977) and also to further study its behaviour during germination.

### 5.1.1 Further evidence that Glycoprotein I undergoes limited autocatalytic degradation

Croy (1977) described Glycoprotein I as a tetrameric protein formed of two subunits each of the 14 and 18 kDa polypeptides observed on SDS-polyacrylamide gels. These two polypeptide bands were also observed on SDS-polyacrylamide gels at each of the many repeats carried out in this work (e.g. figure 4.1.3C and D). However, in addition to these two putative subunits, two more size classes of polypeptides, 15 and 17 kDa in size, were separated from the 14 and 18 kDa subunits and their N-terminal sequences determined (figure 4.1.3D). Based on the high sequence similarities of these four subunit-polypeptides to the same protein (legumin J from pea) it is very likely that they belong to the same protein, Glycoprotein I.

Glyco17 and Glyco15 were found to have an identical N-terminal sequence. Therefore, the size difference of about 2 kDa as determined by SDS-PAGE (figure 4.1.3D) is possibly due to proteolytically cleavage of a C-terminal peptide from Glyco17 resulting in Glyco15. This suggestion is confirmed by the observation of glycosylated polypeptide bands on SDS gels at 18 and 17 kDa as well as at 15 kDa (figure 4.1.4: lane 5 - arrows) indicating that the size difference between Glyco17 and Glyco15 is not due to glycosylation of the larger polypeptide and the lack of oligosaccharides at the smaller polypeptide, so confirming the suggestion of Pusztai and Duncan (1971) that Glycoprotein I undergoes (auto-) proteolytic degradation.

## 5.1.2 Glycoprotein I is a seed storage protein

When seeds are germinating in the absence of an external nitrogen source an extensive breakdown of reserve proteins takes place which is accompanied by the interconversion and utilisation of amino acids in order to produce new nitrogen compounds (Webster, 1958). Therefore, a fast breakdown of seed storage proteins occur within the time course of germination. The disappearance of reserve proteins in the germinated seeds can be expected to depend on the initial amount of the protein in the seed.

In order to elucidate whether Glycoprotein I is a seed storage protein or not, the protein was investigated during the germination of bean seeds. For this purpose, antibodies against the protein were raised in mice and used for immunodetection of Glycoprotein I at different days after imbibition (section 4.1.1.3). This experiment revealed, that Glycoprotein I is largely broken down by about 4 days after the start of the germination and after 5-7 days after imbibition no further cross reactivity between any seed proteins and the anti-Glycoprotein I antiserum was detected (figure 4.1.6). Whereas phaseolin breakdown products were first observed after 4 days after imbibition (figure 4.1.7), Glycoprotein I appears to serve as amino acid donor only during the very beginning of germination. However, in contrast to phaseolin, no breakdown products were detected for Glycoprotein I during germination (figure 4.1.6). This may well be due to the small molecular size of the proteolytically degradated polypeptides. Degradation products of Glycoprotein I subunits would be expected to be relatively small polypeptides which may well have run out of the gel during the SDS-PAGE.

The Western blot analysis (figure 4.1.6B) showed further, that mainly three polypeptide bands were responsible for the antibody-Glycoprotein I cross-reactivity. These polypeptides might represent the "subunits" or degradation products Glyco18, Glyco17 (both together causing the strong upper signal) and Glyco15 (faint lower band). No signal was obtained for the Glyco14 polypeptide band. This might be due to the fact that this Glycoprotein I subunit is not glycosylated (compare figure 4.1.4). Glycosylation of polypeptides has been shown to serve as additional epitopes so resulting in stronger cross-reactivity (Croy, pers. communication).

The positive signals obtained at the top of the SDS gel after immunodetection are either due to high molecular weight proteins with which the anti-Glycoprotein I antiserum shows cross-reactivity or artefacts, for example due to non-dissolved protein aggregates containing Glycoprotein I as well as other proteins. The latter explanation appears to be more likely, as a similar result was observed after the immunodetection of phaseolin with anti-vicilin antiserum (figure 4.1.7B) and is commonly found in Western blot experiments.

### 5.1.3 Is Glycoprotein I formed by polypeptides cleaved from the C-terminal end of the α-chains of *P. vulgaris* legumins?

The three N-terminal sequences obtained for Glycoprotein I showed high homology to a certain domain within the VRII of the α-chain of legumin J from *Pisum sativum* (figure 4.1.5). Therefore, it is hypothesised, that Glycoprotein I may derive from a *Phaseolus* legumin. The basis for this hypothesis is discussed.

Whereas the results of the database search proposed homology of each of the three different N-terminal sequences of Glycoprotein I subunits to the same sequence part of legumin J from pea (amino acids 220 to 233 based on numbering of the preprolegumin, figure 4.1.5; see alignmet in appendix E), in figure 5.1.1 the three different Glycoprotein I sequences were manually aligned to three different stretches within the hyper-variable region of the α-chains of legumins (legumin J (pea) and glycinin 4 and 5 (soybean)). Glyco18 shows high homology to the N-terminal start of the hyper-variable region whereas Glyco15/17 and Glyco14 are located within the hyper-variable region. Considering the length variability of legumins within the hyper-variable region, the molecular weight differences between Glycoprotein I subunits observed on SDS gels reflect very well the $M_r$-distances derived from the legumin alignment (compare figure

acidic subunit ⟵————————⟶   basic subunit ⟵———⟶

S ————————————————————— S

VRI   VRII   HVR

NH2 [      ▓▓▓      ▓▓▓    ▓▓▓▓ ] COOH    NH2 [            ] COOH

```
Leg J:  -----------------------------GKEEEEKEQSH----------SHSHREEKEEEEEEEEDEEE-
Glc 5:  -WQEQ--EDEDEDEDE-EYEQTPSYPPRRPSHGKHEDDEDEDEEEDQPRPDHPPQ-----------------
Glc 4:  -WQEQQDEDEDEDEDD-EDEQIPSHPPRRPSHGKREQDEDEDEDEDKPRPSRPSQGKRNKTGQDEDEDEDEDQ-
         SGHHHEEEEDDPE            LS-GHRHEEED                    SHHDRREEEE
            Glyco18               Glyco15/17                      Glyco14
```

|———————————————|
     appr. 2 kDa
            |——————————————————————|
                  appr. 3 kDa

|———————————————————————————————————————————————|
                  appr. 8 kDa                        -NG-
                                                      ⇧
                                                 α/β-processing
                                                      site

**Figure 5.1.1:** Schematic illustration of the alignment of the three different N-terminal sequences of Glycoprotein I subunits to legumin J from pea and glycinin 4 and 5, legumins from soybean.

Sequences were obtained from the SWISSPROT sequence library (legumin J: P05692, glycinin 4: P02858, glycinin 5: P04347). The underlined residues within the glycinin 5 sequence represent the protease C1 cleavage site as determined by Qi *et al.* (1994). The bold letters indicate the glycosylation site in the glycinin 4 sequence. The black areas represent conserved regions, the stippled areas variable regions (VRI/II = variable region I/II; HVR = hyper-variable region).

5.1.1 and figures 4.1.3C and D). This calculation is based on the relatively large legumin polypeptides glycinin 4 and glycinin 5 to which the *Phaseolus* legumin described in this work (section 5.4) is most related. The largest α-subunit (54 kDa) of the *Phaseolus* legumin (figure 4.2.2) is about 12 kDa larger than the α-subunit of even the relatively large $A_4$-subunit (42 kDa; Staswick and Nielsen, 1983; Nielsen, 1984) from soybean legumin. The peptide causing the larger size of the *Phaseolus* legumin α-subunit may well be localised at the C-terminal end of the α-chain, so giving rise for an approximate 20 kDa peptide for Glyco18, an approximate 18 kDa peptide for Glyco17/15 and an approximate 15 kDa peptide for Glyco14. These sizes are in good agreement with the $M_r$ observed on SDS-gels (e.g. figure 4.1.3C/D). Furthermore, Glycoprotein I is glycosylated. Therefore, a glycosylation site within the *Phaseolus* legumin from which Glycoprotein I may derive would be expected within the C-terminal part of the α-chain. As shown previously (section 4.2.3) the legumin in the seeds of *P. vulgaris* described in the present work is a glycoprotein with a glycosylation site probably within the α-subunit. The exact localisation, however, could not be determined in this work. The α-subunit of glycinin 4 (figure 5.1.1-bold letters) and a legumin from *Picea* (Newton, unpublished results - quoted in Fischer *et al.*, 1995) have glycosylation sites at the C-terminal end of the α-subunits which are, due to the hydrophilic amino acid composition of the hyper-variable region of the α-chains, accessible to modifying enzymes. Figure 5.1.1 (bold residues) shows that the glycosylation site of glycinin 4 lies within the C-terminal part of the α-chain showing homology to the Glycoprotein I subunits Glyco17/15. Therefore, Glyco18 would also be glycosylated. Figure 4.1.4 shows that the Glycoprotein I subunit polypeptide bands of $M_r$ 15 to 18 kDa are glycosylated. All three different N-terminal sequences show a high content of charged residues. In total, 33 amino acids were identified from the three different Glycoprotein I subunit bands (Glyco18, Glyco15/17 and Glyco14). These sequences are mainly composed of glutamic acid (36%), histidine (21%), aspartic acid (12%), arginine (9%) and serine (9%). These parts of the protein would therefore be expected to lie on the surface of a protein which would also then be accessible to the glycosylation machinery.

The similarity to the α-chain of legumins, however, raises the question as to how it is then possible to obtain these polypeptides from the C-terminal end of legumins.

As discussed later (section 5.4.1.4), the legumin in the seeds of *P. vulgaris* described in this work may well have undergone proteolytic degradation or cleavage at the C-terminus of the α-chain (figure 4.2.2). Basis for this hypothesis is the fact that the six different size classes of legumin polypeptides (figure 4.2.3: arrows) showed exactly the same N-terminal amino acid sequence for both α- and β-subunit (see section 4.2.2.3). Whilst it is not unconceivable that all six subunit pairs are derived from different genes but still share the same N-terminal sequences this is somewhat unusual since legumins from the same legume species usually show some sequence similarity even within the initial six N-terminal residues (compare soybean legumin alignment: figure 4.2.5). The resulting α-polypeptides have a molecular size of 34 to 54 kDa (figure 4.2.2). The (poly-)peptides proteolytically cleaved from the N-terminal end of the α-chain would be expected to have a size of about 20 kDa and smaller so representing approximately the size classes of the identified Glycoprotein I subunits. The same results are obtained when sizes are determined theoretically (see above). Further evidence for this hypothesis is provided by the fact, that this region of the legumin lies due to the high content of charged amino acids on the very surface of the 3-dimensional protein structure (Lawrence *et al.*, 1994) and is therefore susceptible for proteolytic enzymes (figure 5.1.1). The amino acid composition of the three N-terminal sequences of Glycoprotein I subunits is also distinguished by a very high proportion of hydrophilic residues (see above) confirming this suggestion. Lawrence and coworkers (1994) suggested a protein structure for legumins similar to the crystal-structure determined for phaseolin at 2.2Å resolution. Based on structural very important amino acids which are globally conserved in vicilins as well as in legumins, the authors propose an alignment of the primary protein sequence of phaseolin and glycinin 2, a soybean legumin, which suggests that the hyper-variable region of the α-subunit lies at the surface of the protein within the loop region of the J'→ helix 4 link of the corresponding phaseolin structure (figure 5.1.2 and Lawrence *et al.*, 1994). Therefore, during seed maturation proteolytically active enzymes may well be able to cleave the protein chain at different parts within this loop resulting in small polypeptides being released. Duranti and coworkers (1992) showed that proteolytic processing of lupin legumin and prolegumin by enzymes present in developing seeds can cause a degradation event leading to several α-polypeptides of different sizes. Qi and coworkers (1994) found that the potential cleavage site within the α-chain of glycinin 5 for the soybean β-

```
                           ┌─────────────────────────────────────────┐                    ┌──
Glycinin 4:        MGKPFT--LSLSSLCLLL-LSSACFAIS---SS--KLNECQLNNLNALEPDHRV
glycinin 2:        MAK-----LVL-SLCFLL-FSGC---FA--LREQAQQNECQIQKLNALKPDNRI
β-phaseolin:MMRARVPLLLL-GILFLASLSAS---FATSLREEEESQDNPFYFNSDNSWNTLF
                                                              Z           A'


┐              ┌─┐
ESEGGLI---QTWNSQHPELKC-AGVTVSKLTLNRNGLHSPSYSPYPRMIIIAQGKGALGVAIPGCPETFE
ESEGGFI---ETWNPNNKPFQC-AGVALSRCTLNRNALRRPSYTNGPQEIYIQQGNGIFGMIFPGCPSTYQ
KNQYGHIRVLQRFDQQSKRLQNLEDYRLVEFRSKPETLLLPQQADAELLLVVRSGSAILVLVKPDDRREYF
     A                        B         C        D        E           F


     ┌───────────────────────────┐
EPQ------EQSNRRGSRSSQKQQLQDSHQKIRHFNEGDVLVIPPSVPYWTYNTGDEPVVAISLLDTSNFN
EPQ-------ESQQR----GRSQRPQDRHQKVHRFREGDLIAVPTGVAWWMYNNEDTPVVAVSIIDTNSLE
FLTSDNPIF--------------------------------SDHQKIPAGTIFYLVNPDPKEDLRIIQLAMPVNN
  F                                        G        H           I


                   ┌──────────────────────────────────────┐
NQLDQTPRVFYLAGNPDIEYPETMQQQQQQKSH---GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTN
NQLDQMPRRFYLAGNQEQEFLKYQ-QQQQGGSQSQK-----GK-QQEEENEGSNILSGFAPEFLKEAFGVN
PQI----HDFFLSSTEAQQ------------------------------------SYLQEFSKHILEASFNSK
     J                                                   1        2


        ┌─┐
        ┌──┐                        ┌──────────────────────────────────
                                      Glyco18          Glyco15/17
                                    SGHHHEEEEDDPE         LS-GHRHEE
EDIAEKLESPDDE--RKQIVTVEGGLSVISPKWQEQQDEDEDEDEDD-EDEQIPSHPPRRPSHGKREQD
MQIVRNLQGENEEEDSGAIVTVKGGLRVTAPAMRKPQQ-------EE-DDDDEEEQP--QCV-------
FEEINRVLFEEEGQQEGVIVNIDSE-----------------------------------------
   3             J'


      ┌──────────────────────────────────────────────────────────┐
                            Glyco14
ED                        SHHDRREEEE
EDEDEDEDKPRPSRPSQGKRNKTGQDEDEDEDEDQPRKSREWRSKKTQPRRPRQEEPRERGCETRN
------ETDKGCQR--QSKRSRN-------------------------------------------
---------QIKE--LSKHAKS--------------------------------------------
        └──    ──┘
            4
```

conglycinin degrading protease C1 (Qi *et al.*, 1992; see also section 5.4.1.4) is located between tryptophan residue 275 and glutamine residue 276 (numbers are counted from the N-terminal of the preprolegumin). This site exactly matches the position of Glyco18 when its N-terminal amino acid sequence is aligned to glycinin 5 and other legume legumins (figure 5.1.1: underlined residues of glycinin 5 sequence; Qi *et al.*, 1994). The basis Qi and coworkers (1994) determined the potential protease C1 cleavage sites was a sequence cluster consisting of at least six acidic amino acid residues in tandem. This has been shown for soybean β-conglycinin to be a necessary prerequisite for a protease C1 cleavage site (Qi *et al.*, 1994). Each of the N-terminal amino acid sequences of Glycoprotein I also show such a cluster of at least six acidic residues one after the other. The polypeptides arising from such cleavage could then, due to their amino acid composition, become water soluble (albumins) rather than salt soluble (globulins) as the *Phaseolus* legumin is.

The cleavage could also be achieved similar and by the same enzyme responsible for the post-translational processing of the legumin precursor. Recently Hara-Nishimura and coworkers (1995) showed that the processing sites of vacuolar processing enzymes are not as highly conserved as often believed. Furthermore, Wind and Häger (1996) demonstrated, that the translational cleavage site of legumins in most of the gymnosperms investigated in their work is not between an asparagine and glycinine residue as it was thought universal till then but rather between an asparagine and proline residue.

But how is it then possible that the peptides cleaved from the C-terminal end of the α-subunit of the *Phaseolus* legumin, form Glycoprotein I, a 60 kDa polypeptide?

Staswick and coworkers (1984) as well as Momma *et al.* (1985) suggest that non-covalent interactions within legumin proteins play an important role for the interpolypeptide binding of the protein structure. This type of binding, which is based on different potentials of the involved polypeptides, was described to occur between the C-terminus of the $A_4$ acidic subunit and the C-terminus of the $B_3$ basic subunit of glycinin $A_5A_4B_3$ (Staswick *et al.*, 1984; Momma *et al.*, 1985). The authors suggested that these interpolypeptide bonds are responsible for the retension of the $A_4$ subunit to the mature legumin, as $A_4$ lacks a disulphide bond. Glycoprotein I may have a similar origin. The single polypeptides may be hold together by interpolypeptide binding. This type of protein

formation could then also be the reason for the gradual degradation of the protein as described by Pusztai and Duncan (1971). At pH 9-10 the non-covalent interpolypeptide bonds may well be disturbed leading to a degradation of the 60 kDa protein.

Furthermore, this type of protein formation could have left to protein characteristics which are different to those of legumins since Glycoprotein I is water-soluble (albumin-type) whereas legumin is only soluble in high salt buffers (globulin-type). Differences would also concern the protein structure which provide an explanation for the lack of common epitopes of both proteins as it revealed from Western blot anlysis of Glycoprotein I with affinity-purified anti-*Pisum* legumin antibodies. No cross-reactivity between both proteins was observed (data not shown).

However, despite of all arguments in favour of the hypothesis discussed above, the sequence similarity between the hyper-variable region of the $\alpha$-chain of legumin proteins and Glycoprotein I might also reflect the same evolutionary mechanisms of both reserve proteins, for example, in order to increase the nitrogen content of storage proteins.

An other explanation for the sequence similarity of legumin and Glycoprotein I may be offered when considering the common evolution of 7S and 11S globulins. Shutov and coworkers (1995) described a hypothetical scheme for a common ancestor of both types of globulin proteins (see also section 1.3.1.3). After different evolutionary events both protein types separated and developed to the modern forms. These modifications within legumins include besides the formation of the $\alpha/\beta$-processing site an insertion at the C-terminal end of the $\alpha$-chain resulting in the hyper-variable region. Mechanisms such as insertions are discussed for the evolution of the prolamin storage proteins (e.g. Shewry *et al.*, 1995). First results of the analysis of the sequence of a cDNA clone from seeds of *Sagitarria sagittifolia* indicate, that this clone is composed of a legumin-like sequence at the 3' and 5' ends of the clone, but shows a prolamin-like sequence fragment between these legumin homologous parts, possibly due to an insertion event (Fischer, unpublished results). Therefore, it appears to be possible that - similar to the events occurred during prolamin evolution - a Glycoprotein I "subunit" could have been inserted into the C-terminal part of the $\alpha$-chain resulting in the hyper-variable region during legumin evolution. The tertiary structure of the resulting (legumin) protein would then show a different surface structure

which cause the lack of cross-reactivity between anti-Glycoprotein I antibodies and *Phaseolus* legumin proteins (data not shown). Furthermore, this explanation would also allow to regard Glycoprotein I as a genuine protein.

However, despite of all attempts to explain the high sequence similarity of Glycoprotein I and *Phaseolus* legumin(s), the answer to the question about the relationship between Glycoprotein I and *Phaseolus* legumin(s) can only be given by identification of the complete sequence(s) from the legumin genes or proteins of *P. vulgaris* and comparison with the whole sequences of Glycoprotein I. The results of these experiments can further be investigated by *in vitro* digestion experiments of the *Phaseolus* legumins with different proteases as described by Qi *et al.* (1994) followed by analysis (separation by HPLC-techniques; N-terminal sequencing; size-determination by SDS-PAGE; staining for potential glycosylation sites) of the resulting digestion products. Attempts to obtain more protein sequence data might also be necessary in order to identify the true nature of Glycoprotein I.

## 5.2 Isolation and identification of a SOD in mature seeds of *P. vulgaris*

The isolation of Glycoprotein I was based on the techniques described by Pusztai (1966). At the beginning of the project a very similar experiment to that of Pusztais and Duncans (1971) was carried out, using similar equipment and the same chromatography medium (DE-52). This experiment resulted in a "rather broad and assymmetrical band" (Pusztai and Pusztai, 1971) when Glycoprotein I was eluted from DEAE-cellulose during the ion-exchange chromatography (data not shown). Subsequent experiments using a highly resolving FPLC-system (Pharmacia) and an improved chromatography medium (DEAE-Express Ion Exchanger, Whatman) resulted in the separation of a 19 kDa polypeptide that formed the front part of the elution peak of which Glycoprotein I represented the main part of the peak (figure 4.1.3). The determination of the 15 N-terminal amino acids revealed this protein subunit to be a superoxide dismutase (SOD), subsequently designated as PVSOD.

### 5.2.1 PVSOD provides the first sequence information on a SOD from mature seeds of higher plants

Much research has been devoted to plant SOD's, mainly with respect to their enzymatic function in photosynthetic plant tissues. Using a method developed by Beauchamp and Fridovich (1971) which is based on the inhibition by the enzyme of nitroblue tetrazolium reduction by light, Reuzeau and Cavalie (1995) recently showed, that SOD's are also enzymatically active in mature seeds. However, until now no sequence information on SODs from mature seed tissues have been published in the literature or databases. Sakamoto and co-workers (1993) sequenced a cDNA of a SOD-encoding gene isolated from immature seeds of rice. But no further description of the state of seed maturation was made. Furthermore, the presence of a SOD-encoding mRNA in immature seeds does not necessarily mean, that the encoded specific SOD enzyme will still be present in the mature seed. Therefore, the present work represents the first protein sequence information on a SOD from mature seeds in the form of an N-terminal amino acid sequence of PVSOD from *P. vulgaris* seeds.

SODs occur as monomers or dimers built up of subunits in the size range between 17 to 19 kDa, that is a polypeptide of about 150 amino acids (Kanematsu and Asada, 1989). The size of the native protein as determined by gel filtration is usually less than double the size of the molecular weight of a subunit as estimated by SDS-PAGE (Kanematsu and Asada, 1989). The reason for this discrepancy is not known but may well be caused by the structure of the polypeptide affecting both the results from gel filtration and from the SDS-PAGE. PVSOD was obtained from the Sephacryl S-300 column within the same size fraction as the two chitinases (monomers of molecular weight 25 and 32 kDa) and Glycoprotein I (60 kDa). Therefore, it is suggested, that PVSOD represents a homodimer built up of two of the 19 kDa subunits. The size of the enzyme in the native form is estimated - considering the size discrepancy discussed above - to be about 33 kDa.

### 5.2.2 PVSOD might be localised in the cytosol or represent a new type of SODs

The known SODs are found in chloroplastic, cytosolic and mitochondrial compartments of the seed tissue (Bowler *et al.*, 1992) as well as in peroxisomes (Bueno and del Rio, 1992; Bueno *et al.*, 1995). The database sequence comparison revealed that PVSOD showed the

highest similarity to the N-termini of cytosolic SODs (figure 4.1.8). The localisation of these SODs within the cytosol is based on sequence similarity of the enzymes to SODs for which the localisation has been determined. Therefore, it is suggested that PVSOD represents a superoxide dismutase localised in the cytosolic compartment of the cell. This is also in agreement with the consideration, that seeds do not contain any chloroplasts (Sitte *et al.*, 1991).

However, the comparison of the N-terminal amino acid sequence of PVSOD with those of other SODs leads to the suggestion that the overall primary protein sequence of PVSOD might be more different from the amino acid sequences of the available cytosolic SODs. This suggestion has to be considered on the basis of the fact that SODs have been evolving very slowly. For example, the amino acid sequences of SODs of monocots and dicots, which diverged about 125 million years ago, show on average 28 differences (Fitch and Ayala, 1994). Residues at the positions 2, 12 and 15 based on the numbering in the SOD alignment in figure 4.1.8 are highly conserved in cytosolic SODs but different in PVSOD (figure 4.1.8). The glycine residue at position 12 of the known SODs from angiospermous plant is replaced by the charged amino acid aspartic acid. The aspartic acid residue at position 12 of PVSOD is only observed also in pine (Karpinski *et al.*, 1992), confirming the relatively high substitution rate in PVSOD. Whereas the legumes arose about 60 million years ago, the *Pinaceae*, to which the gymnospermous pine tree belongs, are thought to have developed during the Jurassic (Stewart and Rothwell, 1993) or even Triassic quaterny period (Sitte *et al.*, 1991), that is about 150 to 250 million years ago. Therefore, a substitution appears to be much more likely in the older pine tree than in *P. vulgaris*. Further significant replacement is the substitution of the lysine residue at position 2 conserved in all known cytosolic SODs against a serine residue in PVSOD and the substitution of glycine (position 15) with an alanine residue.

An other explanation of these differences in the primary sequence may be offered by the possibility that PVSOD represents a member of a new, unknown group of SOD possibly even localised in another compartment of the cell. Considering the endosymbiont hypothesis, this alternative may well be possible. The SODs occurred already in procaryotic bacteria which - according to the endosymbiont theory - are the origin of the mitochondrial and chloroplastic SODs of the eucaryotic organisms. It is further thought that all different SOD types arose from the mitochondrial and chloroplastic organelles

from where their encoding genes have been transferred to nuclear genes during eucaryote evolution (Alberts et al., 1994). For example, the amino acid sequence of a mitochondrial SOD from chicken shows much more similarity to the corresponding bacterial enzyme than to the cytosolic SOD in the same cell (Alberts et al., 1994). As the cytosolic SODs are more different from the chloroplastic SODs than they are among themselves, it seems likely, that they all arose from the chloroplastic SODs and developed later to the different isoforms. Based on the same scenario, it may very well be that PVSOD represents a SOD developed like the cytosolic SODs from chloroplastic SODs. After transfer of the PVSOD gene to the nuclear genome it may have developed to its specific function as SOD in an other cell compartment than the cytosol, therefore developing other, specifc characteristics. As no investigations concerning the localisation of PVSOD within the seed cells has been attempted in this study, localisation experiments by, for example, immuno-histochemical detection would be further necessary to test this hypothesis.

In any case, concerning the localisation of the protein, the main question resulting from the presence of a SOD in mature seeds concerns the biological role of such an enzyme in mature seeds during dormancy when no major metabolic reactions take place. To help to elucidate this question, it is necessary to discuss the role of the enzymatic activity of SODs in the plant cell.

### 5.2.3 SODs play an important role in plant protection against oxygen stress

Much work has been devoted to the elucidation of SOD enzymatic activity in plants. The enzyme was found to be important because it represnts an integral part of the protection of the plant against oxygen stress (Bannister et al., 1987). Different types of SODs have been described, mainly distinguished by their metal cofactor: Superoxide dismutases are known which contain copper/zinc (Cu/ZnSOD), manganese (MnSOD) or iron (FeSOD) as metal cofactor (for a review see Bannister et al., 1987). The main function of SODs in plants is thought to be the protection of the plant against oxygen radicals which may be built up especially in chloroplasts. For example, the electrons arising from the photosystem I can not only be transferred to $NADP^+$ but under certain conditions as well to $O_2$ (pseudo-cyclic electron-transport, "Mehler-reaction"; Sitte et al., 1991), so producing a superoxide radical ($O_2^-$) as intermediate state (figure 5.2.1A - (1); for reviews concerning the physiology of

SODs in plants see Bowler *et al.*, 1992; Scandalios, 1993; Sitte *et al.*, 1991). This is the case, for example, during drought conditions which lead to a closing of the stomatal respiration system of the plant, so limiting the $CO_2$ availability for photosynthetic carbon assimilation. The superoxide radical itself is not very toxic for the plant (Sitte *et al.*, 1991). However, the simultaneous presence of hydrogen peroxide and superoxide radicals (and any metal ions as catalysts) leads to the formation of a hydroxyl radical (OH⁻) (figure 5.2.1A - (3)), which is one of the most toxic oxygen species for the cell (Sitte *et al.*, 1991). SODs prevent the formation of hydroxyl radicals by converting superoxide radicals ($O_2^-$) to hydrogen peroxide and oxygen (figure 5.2.1A - (2)). The hydrogen peroxide is then converted by catalase activity to water and oxygen (figure 5.2.1A - (4)).

Besides the origin of reactive oxygen species in chloroplasts, there are several metabolic activities in mitochondria and cytosol which can generate oxygen radicals as by-products of enzymatic activities (Bowler *et al.*, 1992), so necessitating the occurrence of SODs also in the cytosol, in peroxisomes and mitochondria.

However, the more it seems very logical, that the plant, in order to protect itself, produces much SOD in the stage of developing and germinating seeds when most enzymatic activities take place, the more it seems unusual that SODs occur also in mature, dormant seeds.

## 5.2.4 The role of a SOD in mature seeds during dormancy

There have been several suggestions for different possible functions of SODs in seed tissue.

One suggestion for the presence of a SOD in mature seeds is that it has a role in the defence response to pathogens. Low and Heinstein (1986) showed that major changes in membrane permeability occur as a defence mechanism within minutes after elicitor treatment of cell cultures. Later, they demonstrated, that the defence reaction was initiated by a pathway involving hydrogen peroxide as a transmitter rather than superoxide (Apostol *et al.*, 1986). Furthermore, it was shown, that this response to a pathogen elicitor was inhibited by catalase but not by SOD confirming the role of $H_2O_2$ produced by SOD as second messenger (figure 5.2.1B). Hydrogen peroxide as a second messenger in signal transduction pathways had been first suggested for the function of hormones such as

A)

$$O_2 + e^- \longrightarrow {}^\bullet O_2^- \tag{1}$$

$$\overset{\boxed{\text{SOD}}}{{}^\bullet O_2^- + {}^\bullet O_2^- + 2H^+ \xrightarrow{\hspace{2cm}} H_2O_2 + O_2} \tag{2}$$

$$H_2O_2 + {}^\bullet O_2^- \longrightarrow OH^- + {}^\bullet OH + O_2 \tag{3}$$

$$\overset{\boxed{\text{CAT}}}{H_2O_2 + H_2O_2 \xrightarrow{\hspace{2cm}} 2\, H_2O + O_2} \tag{4}$$

**B)**

physical defence by
cell wall fortification

⬆

lignification

⬆

function in peroxidase          production of
catalysed polymerisation        salicyclic acid

iduction of response reactions
to pathogenic attack
(e.g. pathogenesis-related
protein transcription)

NAD(P)H oxidase

polyamine oxidase

peroxidase

oxalate oxidase

e.g.          e.g.

second
messenger
function

$O_2$ → $°O_2^-$ → $\overset{SOD}{\longrightarrow}$ → $H_2O_2$

ascorbate          GSH          NADPH
                   (ox.)

ascorbate          dehydroascorbate   glutathionine
CAT   peroxidase    reductase          reductase

$H_2O$          dehydro-          GSH          NADP+
                ascorbate         (red.)

increased production of
toxic oxygen species (OH )

⬇

lipid (e.g. membrane)
oxidation
de-estrification of
phospholipids

⬇

membrane alterations

⬇

solute leakage through
membranes

⬇

triggering of signal
transduction and perception

storage mobilisation
germination

thioredoxin

protein activation
/inactivation

insulin in animal cells (Ramasarma, 1982). Although viral inoculation (TMV) of tobacco leaves causing expression of pathogen related proteins did not lead to an increase of $H_2O_2$ (Neuenschwander et al., 1995) it was shown that an injection of tobacco leaves with $H_2O_2$ in increasing concentration leads to a dose-dependent accumulation of total salicylic acid (Neuenschwander *et al.*, 1995) confirming the suggestions by Chen and coworkers (Chen and Klessig, 1991; Chen *et al.*, 1993) who proposed that $H_2O_2$ is a second messenger of salicylic acid formation (figure 5.2.1B). Salicylic acid was found to serve as endogenous signal in the resistance response (examples given in figure 5.2.1B) of the plant to viral infection (Malamy *et al.*, 1990). Further evidence for the suggestion of $H_2O_2$ as secondary messenger for salicyclic acid for the induction of pathogen related proteins was reported by Bi *et al.* (1995).

However, whereas many other enzymes (e.g. oxalate oxidase, NAD(P)H oxidase, polyamine oxidase etc.; see Zhang *et al.*, 1995 and references herein) were suggested as candidates for the production of extracellular $H_2O_2$ in plants (figure 5.2.1B), no evidence has yet been reported that SODs may have an impact in production of $H_2O_2$ in response to pathogenic attack.

Therefore, it is more likely that the possible main function of a SOD in seed tissue lies in the protection against oxygen stress to maintain the viability of the seed.

Hendry *et al.* (1992) showed that decreases in the protective mechanisms against activated forms of oxygen lead to increased lipid peroxidation and lower the viability of seeds. Furthermore, oxygen radicals have also implications for the desiccation tolerance of seeds (Leprince *et al.*, 1990). Lipid peroxidation leads to a de-esterification of phospholipids resulting in an accumulation of free fatty acids and increased solute diffusion from cells (Senaratna and Borochov, 1987). Whereas a low level of solute leakage represents a selective permeability of the cellular membranes working as barriers, a high level of solute leakage points to membrane damage (Parrish and Leopold, 1978).

Reuzeau and Cavalie (1995) suggested that these membrane alterations might occur also in mature seeds as they observed a decrease of germinability of dry sunflower seeds with ageing. The authors proposed oxidative changes of different cell components such as membrane constituents responsible for the inability of seeds to germinate. These changes could lead to alterations of perception and transduction of signals necessary for the start of seed germination (Reuzeau and Cavalie, 1995). Therefore, an effective oxygen-radical

scavenging system needs to be present in the seeds during maturation to break down radicals before desiccation as well as during seed dormancy in order to preserve seed viability. This is particularly important in recalcitrant and long-lived seeds.

It is likely to be that *P. vulgaris* has due to its relatively early time of evolution a high level of SODs. Investigations on sedimentary and other deposits provides evidence for the changes in the atmospheric content of oxygen. Berner and Canfield (1989) developed an oxygen model which indicated a rise in oxygen up to 35% and then a fall to 15% over a time period of just 120 million years during the late Paleozoic (about 410 to 245 million years before present). After this decline in the partial oxygen pressure, a second increase in the atmospheric oxygen level occurred during the Cretaceous period (145-65 million years before the present day). Whereas the first increase of oxygen affected mainly the Carboniferous and early Permian flora including ferns, primitive gymnosperms as well as sphenopsids and lycopods (Graham *et al.*, 1995 and references herein), the second rise in atmospheric oxygen seems to have had implications for the current flora (Hendry and Crawford, 1994). This increased oxygen level had important effects on the physiology of the plants concerned and most probably also on the level of SOD and other enzymes involved with $O_2$ metabolism in plants (Hendry, 1994) in order to protect themselves against the high oxygen pressure. For example, the late Cretaceous family Onagraceae have a higher SOD activity than more modern plants (Hendry, 1994). The legume family arose at about the same time as the Onagraceae (Sitte *et al.*, 1991). Therefore, it is not unreasonable to assume that PVSOD is a reflection of a similar origin, mechanism and function.

The fact that SOD represents a key enzyme role in protection of plant tissue against oxygen stress led researcher to investigate the possibility of increased protection by SOD-overexpression in plants (see Gupta *et al.*, 1993 and references herein). Gupta and coworkers (1993) observed besides a 3-fold increase of the overexpressed pea Cu/Zn SOD in the transgenic tobacco plants a 3- to 4-fold increase in ascorbate peroxidase, an other important key enzyme in the Halliwell-Asada pathway (compare figure 5.2.1B) showing that the increased $H_2O_2$ production caused by the higher SOD level is automatically compensated by the transgenic plant by increased expression of the $H_2O_2$-degradation enzyme ascorbate peroxidase. Therefore, increasing the SOD levels in plants by genetic

engineering seems to provide a promising avenue in seed protection against oxygen stress. The first sequence information on a seed SOD provided in this work could represent a first milestone for this approach.

## 5.3 Seed - chitinases

During the last few years there has been growing interest in enzymes which protect the plant from damage by pathogenic fungi or insects and much work has been devoted to a search for these proteins. One of the main groups of such enzymes are the chitinases which have been shown to play an important role in defence reactions of plants against potential pathogens (Schlumbaum et al., 1986).

Chitinases have been found widely in different tissues of plants including leaves of dicots (e.g. Boller et al., 1983) and monocots (Kragh et al., 1990), nodules of legume plants, roots, rhizomes (Broekart et al., 1989) as well as in seeds (e.g. Leah et al., 1991; Yamagami and Funatsu, 1994). They are also present in secretions such as the latex of *Hevea brasiliensis* (Martin, 1991). Chitinases are defined as enzymes cleaving a bond between the C1 and C4 of two consecutive N-acetylglucosamines of chitin (Flach et al., 1992) and are therefore able to degrade the cell walls of pathogenic fungi (Schlumbaum et al., 1986). Furthermore, it has been demonstrated, that increased levels of chitinases in transgenic tobacco and rape plants containing a bean chitinase gene under a constitutive promoter leads to an increased resistance to certain fungal pathogens when compared with the non-transgenic control plants (Broglie et al., 1991). These promising results led finally to an increasing interest in chitinase research as an application in plant protection (e.g. Broglie et al., 1991).

Much effort has been devoted to the classification of the increasing number of chitinase sequences and to derive a generally accepted and meaningful classification. These classifications are mainly based directly on sequence comparisons. In a comprehensive classification of glycosyl hydrolases, Henrissat (1991) and Henrissat and Bairoch (1993) include plant chitinases, along with those from fungal and bacterial sources, in groups 18 and 19. Other authors (e.g. Collinge et al., 1993) proposed an own classification for plant chitinases. Currently six classes of chitinase enzymes are known - class Ia, Ib, II, III, IV and V chitinases (figure 5.3.1). The main difference between class I and class II chitinases is the presence of an N-terminal cysteine-rich region of approximately 40 amino acids in

| | |
|---|---|
| ▬ Leader sequence | ▨ Catalytic region |
| ▨ Cysteine-rich domain | ▥ C-terminal extension |
| ▨ Hinge region | Λ Deletion |
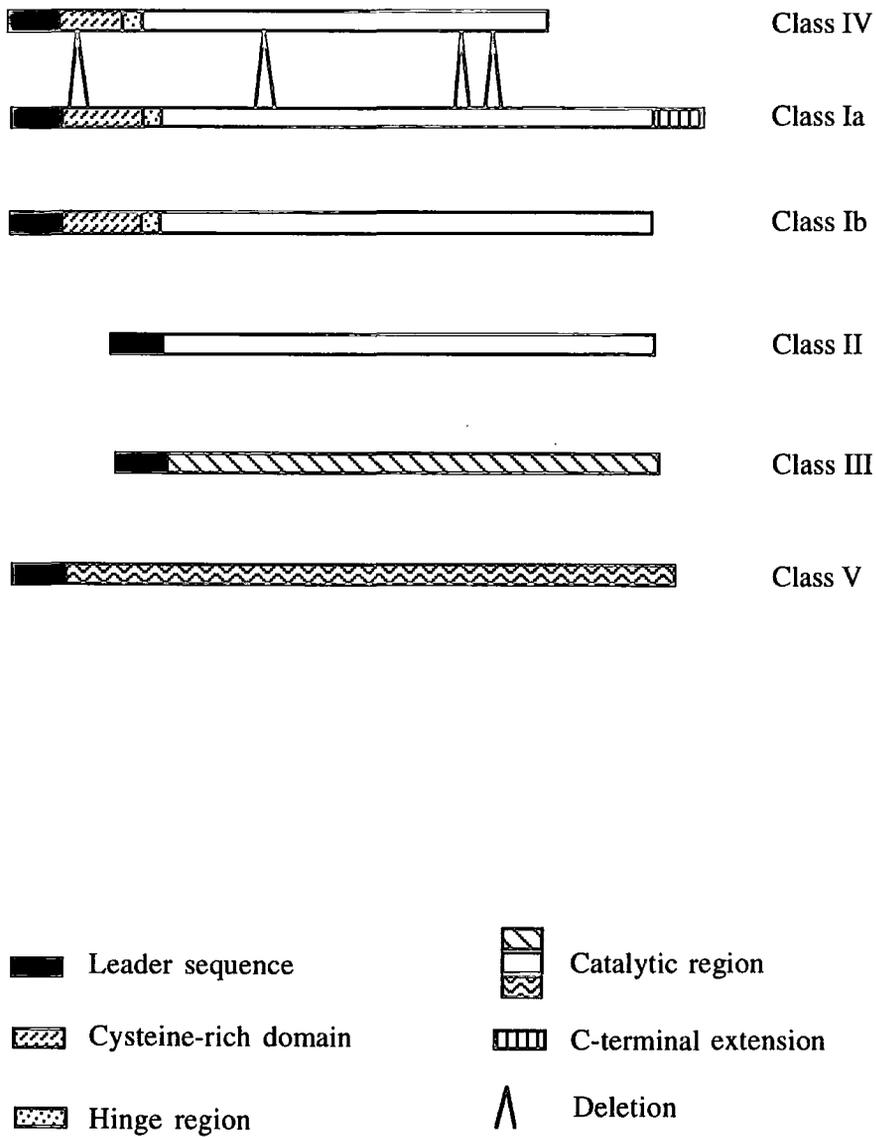
**Figure 5.3.1: Classification and structure of chitinases.**

(Based on Collinge *et al.*, 1993).

118

class I chitinases which is separated from the conserved main enzyme structural region by a highly variable "hinge" region (Collinge *et al.*, 1993; Flach *et al.*, 1992). Class Ia chitinases are distinguished from Ib chitinases by the presence of a C-terminal extension (Collinge *et al.*, 1993). Class II chitinases show a high sequence similarity to that of class I chitinases at the catalytic active site. Class III chitinases have no sequence similarity with class I or class II chitinases and lack the N-terminal cysteine-rich region. Class IV chitinases contain the cysteine-rich N-terminal domain and the conserved main structural region thus showing high homology to class I chitinases but are smaller due to four deletions (figure 5.3.1). The total sequence homology between class I and class IV chitinases is 41-47% Collinge *et al.*, 1993).

More recently, a new class of chitinases, designated as class V chitinases, has been found in tobacco after exposure to different forms of stress such as viral infection and wounding (Melchers *et al.*, 1994). The class V chitinase shows high sequence homology to bacterial chitinases but no sequence similarity to any of the class I to IV plant chitinases.

### 5.3.1 A class II and a class III chitinase are present in the mature seeds of *P. vulgaris*

During the present work active chitinase enzymes have been purified from seeds of *P. vulgaris* and identified by N-terminal sequencing, enzymatic activity test and Western blot analysis (section 4.1.3). The alignment of these N-terminal chitinase sequences to those obtained from the EMBL sequence library (Heidelberg) revealed that neither of them shows the cysteine-rich region typical for class I and class IV chitinases and that the 25 kDa (pvchi25) protein can be regarded as a class II chitinase (figure 4.1.9B). The 32 kDa (pvchi32) protein represents a class III chitinase (figure 4.1.9A) as revealed by its high sequence similarity to other class III chitinases. The enzymes are present in the albumin class of proteins and are purified by a combination of gel filtration chromatography and ion-exchange chromatography. The ion-exchange chromatography enabled also the separation of both chitinases from each other due to their slightly different pI's (figure 4.1.3B).

Both chitinases were also distinguishable by immunoblot analysis. The seed chitinases share common epitopes with ethylene-induced bean leaf chitinases (e.g. Boller *et al.*, 1983) as judged by cross-reactivity with antibodies raised against this type of chitinases. Whereas pvchi25 gave strong signals with the antibodies against bean leaf chitinases,

pvchi32 showed only a very low level of cross-reaction (figure 4.1.10). The presence of the pvchi32 enzyme was not detectable in total seed protein extract probably due to the low cross-reactivity to the antibodies used in this experiment (figure 4.1.10 - lane 1). The reason for the different cross-reactivity of pvchi25 and pvchi32 could be on the one hand an absence of class III chitinases in bean leaves combined with a low cross-reactivity of class I and/or class II chitinases with pvchi32. On the other hand, the weak signals from the Western blot analysis might be caused by just very low amounts of class III chitinases in bean leaves and therefore minor amounts of anti-class III chitinase-antibodies in the antiserum raised against total ethylene-induced bean leaf chitinases. The latter explanation seems more likely as up to date no cross-reactivity of antibodies raised against class I and class II chitinases with class III chitinases has been shown (e.g. Vogelsang and Barz, 1993).

The sequences of both chitinases lack the leucine- and valine-rich signal peptide. Neuhaus and coworkers (1991) showed that this short and very basic signal peptide is necessary for targeting a class I chitinase from tobacco to the vacuole where it is located within the plant cell. Therefore, both seed chitinases are mature proteins, transported to and located at their destined sites. The localisation of the *P. vulgaris* seed chitinases was not determined in the present study.


## 5.3.2 Pvchi32 is the first reported class III chitinase in seeds

Sequence information for chitinases from mature seeds is up to now only available from rye (Yamagami and Funatsu, 1993, 1994) and chestnut (Collada *et al.*, 1992). Both rye (chitinase-c: Yamagami and Funatsu, 1993; chitinase-a: Yamagami and Funatsu, 1994) and the three chestnut chitinases (Collada *et al.*, 1992) represent either class I or class II chitinases. Class III chitinases have, until now, only been found either in plant cell culture material (*Cicer arietinum*: Vogelsang and Barz, 1993; *Parthenocissus quinquifolia*: Bernasconi *et al.*, 1987), in leaves (*Cucumis sativus*: Metraux *et al.*, 1989; *Arabidopsis thaliana*: Samac *et al.*, 1990) or in the rubber latex of *Hevea brasiliensis* (Jekel *et al.*, 1991). Thus, pvchi32 is the first class III chitinase to be isolated and described from mature seeds.

The second distinguishing feature of pvchi32 is the presence of N-linked carbohydrates as judged by PAS staining of the electrophoretically separated protein (figure 4.1.4 - lane 4). The only so far described glycosylated chitinase is the extracellular protein 3 (EP3), a 32 kDa chitinase isolated from carrot cell culture medium (De Jong *et al.*, 1992). EP 3 has been found to have an important function in early plant somatic embryo development (De Jong *et al.*, 1992).

In the past much research has been devoted to crop plant protection. In particular during the last years strategies for control of fungal diseases with transgenic plants expressing chitinases from other plants have been developed (Broglie *et al.*, 1991; Cornelissen and Melchers, 1993; Jach *et al.*, 1995). Investigations have shown that it is very important to provide the plant tissue to be protected with different types of chitinases as higher levels of expression of a single chitinase does not necessarily lead to better resistance against pathogenic fungi (Neuhaus *et al.*, 1991; Jach *et al.*, 1995).

Pvchi32 and pvchi25 appear to be a class III and class II chitinases, respectively. The localisation of pvchi32 might be in the extracellular volume as has been described for cucumber (Metraux *et al.*, 1989) and *Arabidopsis* (Samac *et al.*, 1990) class III chitinases. Class II chitinases are also reported to be mostly localised apoplastically (Collinge *et al.*, 1993). If pvchi32 and/or pvchi25 are also localised extracellular, both seed chitinases could provide an opportunity for transferring a broad resistance against pathogenic fungi by transforming the gene encoding pvchi32 and/or pvchi25 into other crop plants containing mainly class I chitinases. A successful transformation of these chitinases would then improve protection against intercellular propagating fungi. *Nicotiana sylvestris* plants transformed with an intracellular located class I chitinase from tobacco and under the regulation of a constitutive promoter accumulated up to 120-fold more chitinases than in control plants (Neuhaus *et al.*, 1991). However, the transgenic plants did not show higher resistance to the intercellular developing fungus *Cercospora nicotiana* than the non-transformed plants. Thus a combination of different chitinase enzymes exhibiting different specificities is more important than just a high level of expression of a particular enzyme. However, further experiments are necessary in order to elucidate the localisation of both seed chitinases in the seeds of *P. vulgaris*.

### 5.3.3 The natural role of seed chitinases

Transcription of chitinase-encoding genes was shown to be inducible by various factors including infection of plant tissue with phytopathogenic fungi (Metraux and Boller, 1986), pest attack, as well as treatment with the plant stress hormone ethylene (Broglie et al., 1986; Mauch and Staehelin, 1989) or elicitor treatment (Herget et al., 1990). Therefore, it was proposed that the major natural role for chitinases is defence, especially against fungal pathogens whose cell walls are composed of chitinase components (Boller 1987, 1988). This suggestion has been confirmed by the observation that chitinases accumulate around fungal, hyphal material in plants (Benhamou et al., 1990; Wubben et al., 1992).

Both *P. vulgaris* seed chitinases showed enzymatic activity towards a chitinase-substrate confirming their possible antipathogenic function in seeds.

The protection of seeds during dormancy plays an important role in seed viability, seed dispersal and plant propagation, especially as the seeds lie on or under the surface of the soil often covered by moist soil or grass and therefore most favoured to be attacked by fungal pathogens. The plant has therefore to take precautions in order to protect its seeds during dormancy as no metabolic activity such as protein synthesis is possible in the dry state. Therefore, it is to be expected that mature seeds contain lytic active enzymes which help to defend it against fungal attack and that after germination and the growth of a seedling these enzymes are no longer necessary or are superseded by a new battery of lytic enzymes.

If this hypothesis is true, the chitinases in seeds would be expected to be of no further use once germination is under way and therefore are degraded after the start of seed-germination.

In order to try to prove this hypothesis, the seed chitinases of kidney bean were assayed during the first nine days after imbibition of the seeds (section 4.1.3.3C). As figure 4.1.11 shows, pvchi25 is fully degraded after about six days from the start of germination. Pvchi32 could not be studied due to the lack of cross-reactivity with the available antibodies (section 5.3.1). This observation suggests that the natural role of at least pvchi25 is a contribution to the defence system of the seeds against fungal pathogens during dormancy and early germination.

## 5.4 Legumin in the seeds of *P. vulgaris*

Since the first suggestion and partially superficial description of legumin proteins in *Phaseolus* species (for a review see Derbyshire *et al.*, 1976) several attempts have been undertaken to demonstrate legumin-like proteins in French bean (e.g. Derbyshire and Boulter, 1976; McLeester *et al.*, 1973; Sun and Hall, 1975). Attempts to identify legumin proteins in other species of the genus Phaseolus were also approached (e.g.. *P. coccineus*: Bernardi *et al.*, 1990). However, a final prove for the presence of legumin in the genus *Phaseolus* has never been achieved.

In order to detect such legumin-like proteins in the seeds of *P. vulgaris* two experimental approaches were attempted - a molecular biological (section 5.4.2) and a protein purification approach (section 5.4.1).

### 5.4.1 Legumin proteins from the seeds of *P. vulgaris*

#### 5.4.1.1 Fractionation of *P. vulgaris* seed proteins

As legumin was expected to appear only (if at all) in a small amount in the seeds of *P. vulgaris*, a separation strategy for the seed proteins was employed which allowed a separation of the major seed proteins, phaseolins (50% of total seed proteins; Bollini and Chrispeels, 1978) and phytohemmagglutinins (10% of total seed proteins; Bollini and Chrispeels, 1978) from the other seed proteins which occur only in minor amounts. In order to separate these major seed proteins, known to be about 130-150 kDa in size, from any legumin proteins present, which are universally of much higher molecular weight (about 360-400 kDa), size fractionation by gel filtration was employed. The SDS-PAGE analysis of the different fractions obtained from the Sephacryl S-300 column shows (figure 4.2.1), that this approach was successful and led directly to the separation of vicilins and phytohemagglutinins from the proteins of higher $M_r$ including legumin-like (disulphide-linked) proteins (figure 4.2.1A).

#### 5.4.1.2 Legumin-like proteins are present also in the seeds of *P. vulgaris*

The protein fractionation by gel filtration enabled the identification of several protein bands which are not visible on SDS-polyacrylamide gels when total seed or total globulin extracts are applied. Comparison of the SDS-PAGE patterns of reduced and unreduced protein samples of the protein-containing fractions eluted from the Sephacryl S-300

column helped to identify polypeptides linked by intermolecular disulphide-bonds, a characteristic of legumin-like proteins. Figure 4.2.1 showed that the first two peaks contained several candidates for legumin-like proteins. The 2-dimensional gel analysis of one of the protein sample containing disulphide-linked polypeptides confirmed, that about six disulphide-linked proteins of different sizes are present in the TGB-extract (figure 4.2.2). Reduction of the proteins after the first dimension (non-reducing conditions) in 50 mM DTT buffer separated several protein subunits into polypeptides of $M_r$ 54 kDa, 51 kDa, 47 kDa, 43 kDa, 37 kDa and 35 kDa (figure 4.2.2: filled arrows), each of them linked to a polypeptide of 21 kDa (figure 4.2.2: filled arrowhead), corresponding to typical legumin-like α- and β-subunits.

The protein band at 150 kDa might also represent legumin proteins since DTT treatment causes reduction into a potential single size class of putative β-subunit (21 kDa; figure 4.2.2: open arrow) and into different putative α-subunits distributed over the whole size range from 54 kDa to 35 kDa with single spots (figure 4.2.2: open arrowheads) of the same size values of the α-subunits arising from the other legumin-like proteins (figure 4.2.2: filled arrows). Based on this behaviour and the fact that the 150 kDa protein band represents double the size of the 75 kDa protein identified as legumin by N-terminal sequencing (section 4.2.2.3), it may very well be an artefact of the SDS-PAGE, due, for example, to a failure to be fully broken down by boiling the protein sample in SDS-containing sample buffer prior to applying onto the gel, so remaining as an high molecular weight aggregate, e.g. as a 150 kDa dimer. Such a case is known as an artefact by the performance of Western blotting as described previously (section 5.1.4). The same explanation is suggested for the 220 kDa protein band (figure 4.2.1A) which might represent a higher $M_r$ aggregate formed by three 75 kDa polypeptides. A further explanation for the occurrence of the 150 kDa protein may be offered by the observation that this protein is broken down by DTT-treatment to polypeptides in the size typical for legumin-like protein subunits indicating that it may be hold together by disulphide bridges formed between different subunits of the hexameric legumin protein.

Based on the gel filtration protein profile and quantitation of the SDS gel analysis (figure 4.2.1), the ratio of the legumin-like proteins to vicilin proteins (phaseolins) in the seeds of *P. vulgaris* is estimated to be about 1:15 meaning that legumin accounts for only about 3% of total seed proteins in the cotyledons of the investigated French bean variety. The

legumin to vicilin ratio in other legumes investigated is known to be much higher and may approach as much as 1:1 to 2:1 and a legumin content of 38-47% of the total seed proteins (e.g. soybean; Saio *et al.*, 1969).

### 5.4.1.3 The N-terminal sequences of the α- and β-subunit confirm legumin as a typical legume legumin

The N-terminal amino acid sequence analysis resulted in the determination of 20 amino acids for each of the subunits. The alignment of the two partial sequences to the corresponding parts of legumin protein sequences of other legumes (figure 4.2.5) confirms that *Phaseolus* legumin represents a typical legume legumin.

Dry, mature seeds were used for the protein extraction. At this stage the seed storage proteins are localised in the protein bodies. Therefore, the signal peptide, that directs the translocation of the nascent polypeptide from membrane-bound polysomes into the lumen of the RER from where it is tranferred to the protein bodies, is cotranlationally removed (Shotwell and Larkins, 1989).

It is furthermore worth mentioning in almost all of the currently known legumin sequences of gymnosperms and all angiosperms (compare legumin alignment of Lawrence et al., 1994) the highly conserved motif $C - X_{4-5} - L/I - X_4 - P$ is also represented in the α-subunit of the *Phaseolus* legumin. The last part (L-N-A-L-K-P) of this motif in the *Phaseolus* legumin is present in either exactly the same sequence (*Brassica napus*, cruciferin) or slightly different (e.g. cotton: L-N-A-L-Q-P and monocots: L-Q-A-F-E-P in rice and oat and L-Q-A-S-T-P in wheat). The fact that this conserved motif is present in even the much older gymnospermous trees such as the "living fossil" *Ginkgo biloba* (L-N/S-A-Q-E-P) indicates that this region might be of importance for the tertiary structure of the protein and has therefore not undergone any bigger changes during the evolution, although in *Pseudotsuga menzies* and *Picea glauca* legumin an additional arginine residue is added between the conserved cysteine and the leucine residue which is partially replaced by the also hydrophobic isoleucine residue (Lawrence *et al.*, 1994). Lawrence and coworkers (1994) aligned this region (using glycinin 2 sequence; figure 5.1.2) to the very N-terminal of the β-phaseolin sequence. This stretch covers the area from the Z- to the A'- strand in the tertiary protein structure of phaseolin and lies in the centre of the protein and is therefore likely to be of importance for the protein structure.

Besides some amino acid substitutions compared to other legume legumin sequences highly conserved residues within the β-chain of legumin proteins are also present in the *Phaseolus* legumin sequence, confirming that Phaseolus legumin represents a typical legume legumin but also a unique legumin protein. Western blotting and reaction with affinity-purified anti-*Pisum* legumin antibodies failed to show cross-reactivity with any of the *Phaseolus* protein components providing further evidence for the uniqueness of *Phaseolus* legumin and indicating that this protein is sufficiently structurally different as to have no common epitopes with other legume legumins. Similar observations have been made by Kloz and Turkova (1963) between soybean legumin and pea legumin and between legumins from other legume species.

Derbyshire and Boulter (1976) determined the N-terminal amino acids for the *Phaseolus* legumin subunits as threonine (corresponding to the α-subunit) and glycine (corresponding to the β-subunit) which agree with the results reported in this work. In addition, Derbyshire and Boulter found a third amino acid, leucine, which leads to the suggestion that there may be at least one other legumin-type protein present in the seeds of *P. vulgaris*.

### 5.4.1.4 *Phaseolus* legumin seems to undergo proteolytic degradation

The β-subunit (21 kDa) of the *Phaseolus* legumin represents a typical β-subunit size of legume legumins. The $M_r$ of the *Phaseolus* legumin subunit is 75 kDa meaning larger than the largest glycinin subunits (glycinin 4 and 5). Therefore, the additioal peptides causing the higher $M_r$ of the *Phaseolus* legumin are localised within the variable regions of the α-subunit, possibly within the HVR. This suggestion is confirmed by the 2D SDS-PAGE analysis (figure 4.2.2) and subsequent N-terminal sequencing of different size classes of legumins (section 4.2.2.3; figure 4.2.3: arrows) which revealed, that only one legumin type has been identified in the seeds of *P. vulgaris*. All six legumin subunits of different $M_r$ showed exactly the same N-terminal amino acid sequence for the initial six residues of both the α- and the β-subunit. Whilst it is not unconceivable that all six subunit pairs are derived from different genes but still share the same N-terminal sequences this is somewhat unusual since members of the legumin family from the same species usually show some sequence variation, even within the initial six N-terminal residues (e.g. glycinin: figure 4.2.5). The phenomenon of the different sizes for the same α-subunit could

arise from proteolytic processing of the *Phaseolus* legumin and prolegumin similar to that observed for lupin seed legumin (Duranti *et al.*, 1992). The authors explained the occurrence of different size classes of the lupin legumin α-subunit as due to proteolytic processing of lupin legumin and prolegumin by protease activity present during seed development, which then could lead to an orderly cascade degradation resulting in different size variants of the α-polypeptide. Such proteolytic degradation has also been observed for pea legumin during germination which involved the removal of C-terminal peptides from the α-subunit resulting in different sizes of α-polypeptides with the same N-terminal sequence (Croy, unpublished results). Qi and coworkers (1994) investigated the breakdown of legumin proteins from pea and soybean when digested with the soybean β-conglycinin degrading protease C1 (Qi *et al.*, 1992). The authors found that the α-polypeptides of glycinin were 2 to 5 kDa smaller after digestion than the native legumin polypeptide chains. The α-subunits of pea legumin also showed a significant size change. Based on these observation Qi and coworkers (1994) suggested possible protease C1 cleavage sites in the α-chains of soybean and pea legumins. Therefore, it is likely that one or more protease(s) present in the developing seeds of *P. vulgaris* could cleave at different positions within the hyper-variable region of the *Phaseolus* legumin resulting in the different size classes of α-subunits which all derive from the same polypeptide.

### 5.4.1.5 *Phaseolus* legumin is a glycoprotein

Legumin proteins have for a long time been thought not to be glycosylated (e.g. Casey *et al.*, 1986), but some examples of oligosaccharide substitution have been reported. Lupin legumin is glycosylated in the N-terminal region of the α-chain (Duranti *et al.*, 1995) within a conserved region which lies exposed on the surface of the protein as judged by the legumin-vicilin alignment and the 3-dimensional structure of phaseolin described by Lawrence *et al.* (1994). In contrast to the lupin legumin the glycosylation site in the *Magnolia* legumin B14 (Fischer *et al.*, 1995) lies at the C-terminal end of the β-chain. Based on hydropathy plot analysis the authors suggested that this part of the legumin might lie on the surface of the protein molecule due to its hydrophilicity and is therefore accessible to modifying enzymes.

*Phaseolus* legumin was investigated for potential glycosylation sites using Schiff's staining of oligosaccharides (Fairbanks *et al.*, 1971). The results (figure 4.2.6) indicate that the *Phaseolus* legumin is indeed glycosylated. The non-reduced protein sample shows two bands after Schiff's staining at 72 kDa and 75 kDa corresponding to the unreduced legumin subunit molecule (figure 4.2.6A: lane 2-arrows). The same protein sample under reduced conditions, however, did not show any additional glycosylated polypeptides of 54 kDa and 51 kDa or 21 kDa in size as it was expected for the α- and β-subunits (figure 4.2.6A: lane 1). The comparison with the reduced protein sample after coomassie blue staining shows an additional strong protein band for the 21 kDa β-subunit (figure 4.2.6B: arrowhead) but only very faint bands for the legumin α-subunits (figure 4.2.6B: arrows). Therefore, it is very likely, that the α-chain of the *Phaseolus* legumin is glycosylated but was not stained by the PAS staining, possibly due to too low an amount of protein. Repeats of this experiment did not lead to different results. The reason for this behaviour is unclear. Thus, glycosylation of legumins does not seem to be as exceptional as is assumed. Badenoch-Jones and coworkers (1981) demonstrated that the inhibition of glycosylation of vicilin by tunicamycin-treatment does not affect the synthesis, assembly or transport of the protein. Although Doyle *et al.* (1986) suggest that glycosylation near the C-terminus of vicilins may be involved in proteolytic processing or storage of these proteins, no solid evidence could yet been achieved. Thus, the physiological function of this kind of polypeptide modification is still unclear (Bewley and Greenwood, 1992).

### 5.4.1.6 Anti-germin antibodies show cross-reactivity with phytohemagglutinins and phaseolin

Recently several reports were published on the origin and evolution of seed storage proteins (e.g. Bäumlein *et al.*, 1995, Braun *et al.*, 1995, Shewry *et al.*, 1995). Major efforts were put in classifying the different seed storage proteins into groups. These classifications of seed storage proteins were based on similarities of structural important sequence stretches. One of the hypotheses described in these papers was the report about a common ancestor of 7S and 11S globulin seed proteins with germin-like proteins from wheat (Bäumlein *et al.*, 1995). Germins have been described in the seeds of barley and wheat and have been found to play an important role in seed germination (Lane, 1991, 1994) and in induction of response reactions to pathogenic attack (Zhang *et al.*, 1995).

To investigate similarities of germins and 11S and 7S globulins as described by Bäumlein *et al.* (1995) an immuno-blot analysis was carried out of seed proteins of *P. vulgaris* and *Pisum sativum* with available anti-wheat germin antibodies. The results of the dot-blot analysis with different amounts of protein indicated a strong cross-reactivity between the antibodies used and the seed proteins of *P. vulgaris* but not of *Pisum sativum* (figure 4.2.7A). The subsequent Western blot analysis of the electrophoretically separated *P. vulgaris* proteins revealed that the positive signals for *P. vulgaris* seed proteins obtained on the dot blots after hybridisation with the anti-germin antibodies was mainly caused by cross-reactivity of phytohemagglutinins with the antibodies (figure 4.2.7B). Thus, on the one hand the suggestion lies close that there is a high degree of homology between PHA proteins and anti-germin antibodies. However, on the other hand, this cross-reactivity could be caused by the high content of oligosaccharides (4-6%; Bollini *et al.*, 1983) attached to the PHA primary protein sequence, as oligosaccharide structures can also serve as epitopes for antibodies (Croy, pers. communication). Germins are also glycoproteins (Lane, 1991, 1994), so that antibodies raised against germins are also directed against the oligosaccharide part of the germin proteins. Thus, the latter explanation seems more likely, as phaseolin, the second protein type which shows cross-reactivity with the anti-germin antibodies, is also known to be glycosylated (4-6%; Bollini *et al.*, 1983). No binding was observed between pea seed proteins and anti-germin antibodies on the dot blot. The glycosylated pea seed proteins might not show such an oligosaccharide structure appropriate for the anti-germin antibodies.

Therefore, it is likely to be that the real basis for the cross-reactivity observed between anti-germin antibodies from wheat and seed proteins from *P. vulgaris* but not from *Pisum sativum* is the higher oligosaccharide content of the main seed proteins from *P. vulgaris.*

### 5.4.2 The search for the genes encoding a *P. vulgaris* legumin

To obtain more sequence information on legumins from *P. vulgaris* extensive attempts were made to search for the genes encoding the legumin proteins. Basically, three different methods are available for this aim. First, there, is the possibility to screen a cDNA library with a heterologous probe such as a legumin-encoding gene from a related plant species. This approach, however, may be accompanied by two major problem. First, if the gene under investigation is a regulatory gene and the level of mRNA is low, the gene might not

be represented in the cDNA library. Second, if the gene has a developmentally regulated pattern of expression the cDNA may not represent the time specificity of peak expression of the gene, thus minimising the change to find the clone in the available cDNA library. As the amount of legumin in the seeds of *P. vulgaris* proved to be very low (3% of total seed proteins; section 5.4.1.2) in comparison to the major seed proteins (phytohemagglutinins: 10% and phaseolins: 50%; Bollini *et al.*, 1978) and the expression pattern of seed storage proteins depends very much on the plant species (Gatehouse *et al.*, 1986), more than a single mRNA isolation followed by cDNA synthesis and screening of the cDNA library may have to be carried. A similar laborious approach represents the sreening of a genomic library from *P. vulgaris*. Therefore, the third approach, the polymerase chain reaction, has been employed as the strategy for the search for legumin encoding genes in the genome of *P. vulgaris*.

### 5.4.2.1 Polymerase chain reaction with the primer pairs 1F/1R and 2F/2R

PCR has been proven to be a powerful *in vitro* method for specific replication of a particular DNA segment from genomic DNA (e.g. Saiki *et al.*, 1985). The use of degenerate primers, that is oligonucleotides varying in base sequence but with the same number of bases, are often necessary when only limited information about a gene is available, for example, when primers are to be designed on the basis of an amino acid sequence or when the search is for new or uncharacterised sequences related to a known family of genes.

The first attempts to amplify legumin gene fragments by PCR were reported to have been successful (Heim *et al.*, 1994; Arahira and Fukazawa, 1994). However, the described attempts were undertaken on plant species on which legumin gene sequence information was already available from cDNA data (e.g. *Vicia faba*: Heim *et al.*, 1994) or from extensive protein data (*Ginkgo biloba*:Arahira and Fukazawa, 1994). Recently, the first legumin gene fragments PCR-amplified from genomic DNA from plant species of which no sequence information was known were reported (Wind *et al.*, 1996; Häger *et al.*, 1996). The species investigated in these studies were coniferous trees and the PCR amplified legumin gene fragments showed a high sequence similarity. In these cases the PCR technique was established (Wind, Häger; pers. communication) so that legumin gene fragments from different species can be investigated by the PCR technique without

130

legumin sequence information even from the plant genus the species belongs to (Häger *et al.*, 1996; Wind and Häger, pers. communication). In other cases, the use of degenerate primers for amplification of legumin gene fragments from genomic DNA has yet not led to any success (e.g. pepper: Fischer, Franke, pers. communication).

The present attempts to amplify a legumin-encoding gene fragment using the polymerase chain reaction resulted mostly in amplification of non-specific fragments (section 4.2.5.2). A general reason for this failure might be the fact, that the genome of *P. vulgaris* is relatively big (1.7 x $10^9$ bp; Croy *et al.*, 1993). Combined with the use of degenerate oligonucleotides and a possible low copy number of legumin genes which are usually multi-copy genes (Gatehouse and Shirsat, 1992; Häger *et al.*, 1995) might have led to a mispriming of the oligonucleotides at the beginning of the PCR followed by the amplification of these non-specific gene fragments.

This seems very likely to have been the main reason for the failure when using the primer pair 1F/1R as no other major reason, i.e. concerning the primer design, can be determined. The most important part of the primer, the very 3' end of which at least the last three 3' bp have to be absolutely homologous to the template DNA (J. Lang, B. Stanchev, pers. communication) as this is the part of the oligonucleotide, which serves as the start-point for the polymerase (Alberts *et al.*, 1994). Both primers 1F and 1R have been designed so that even the last eight nucleotides of the primer sequences lay within a very homologous region of the legume legumin genes so that no sequence variation of the oligonucleotides was necessary within these eight 3'end nucleotides (section 4.2.5.1 - table 4.2.2; compare also nucleotide sequence alignment in the appendix E). Furthermore, different PCR techniques were employed.

To overcome the amplification of misprimed gene fragments obtained using the low-/high-annealing temperature PCR method hot-start PCR was employed. However, this technique did also not lead to specific amplification of legumin gene fragments from genomic DNA of *P. vulgaris*, although the technique proved to be the appropriate way as the control experiment with genomic DNA from pea revealed (figure 4.2.10). The up to 2 kb large DNA fragments observed for both the pea and French bean PCR amplified gene fragments (figure 4.2.10A) probably resemble PCR artefacts. Such PCR products of which the 2 kb band is the major component, represent "shuffle clones" which can be produced by

amplification of multi-allelic gene families (e.g. Scharf *et al.*, 1988) when the extension of one allele is incomplete and hybridises to other alleles during subsequent cycles (Saiki *et al.*, 1988).

Therefore, a new primer pair was designed. To increase the specificity of the primers, longer oligonucleotides were chosen (section 4.2.5.1 - table 4.2.4). Furthermore, in order to overcome problems concerning the ligation of PCR-amplified gene fragments for cloning, endonuclease linkers with Eco R I restriction sites were synthesised at the 5' ends of the specific legumin sequences (section 4.2.5.1-table 4.2.4). Using again the hot start PCR technique this primer pair led to a specific amplification of legumin genes from genomic DNA from pea in the control experiments (e.g. figure 4.2.12: Co). Furthermore, after optimising the cycle conditions and the reaction mixture a seemingly successful amplification of a legumin gene fragment from genomic DNA of *P. vulgaris* was achieved (figure 4.2.11). The PCR products which hybridised to the legumin probe (figure 4.2.11A: lane 4 - arrows) were about 2 kb and therefore of the expected size. Computer imitations of PCR amplifications using the Amplify programme and pea and soybean genomic sequences as template revealed that the use of primer pair 2F/2R results in amplified fragments of 1.4 to 1.8 kb (section 4.2.5.1). Size determination of the *Phaseolus* legumin subunits by SDS-PAGE revealed that the *Phaseolus* legumin is about 12 kDa bigger (75 kDa) than for example the pea and soybean legumins. This size difference corresponds to about 300 additional bp. However, due to the low amount of putatively specific amplified PCR product neither the cloning nor the re-amplification by re-PCR of these putative legumin encoding gene fragments was successful (section 4.2.5.2B). All later repeats carried out under exactly the same conditions did not lead to the amplification of these putative legumin gene fragments as judged by Southern analysis. Also an extend series of further variations of all reaction parameters concerning both the PCR cycle conditions and the reaction mixture did not provide any specific PCR products which hybridised to the legumin probe used in this study. The reason for the failure of the repeats of an originally successful experiment is unclear. An example of PCRs carried out in order to optimise the reaction mixture and the subsequent Southern analysis are shown in figure 4.2.12. Besides seemingly unspecific PCR products (figure 4.2.12A) which were partially also amplified in the successful PCR (figure 4.2.11A) the two PCR products of about 2 kb hybridising highly specific to the legumin are missing as judged by the Southern analysis (compare

figures 4.2.11B and 4.2.12B). The hybridisation of the pea legumin probe to DNA bands of 400 bp, 600 bp, 900 bp and 1.7 kb were seemingly unspecific due to the high amount of DNA and the long exposure of the autoradiography film to the blot (figure 4.2.12B: overnight; figure 4.2.11B: 30 minutes).

Despite of the seemingly unspecific amplification three of the PCR products which hybridised to the pea legumin probe were sequenced to elucidate possible relationship between the amplified fragments and legumin genes. Sequencing of the 600 bp and 900 bp PCR products revealed that these fragments did not show any sequence similarity to any known sequences of different databases and were indeed results of unspecific PCR amplification.

The amplified 1.65 kb PCR product was found to represent a PCR artefact. The sequencing of this fragment from both sites did not only show that this PCR product does not contain any clear homology to legumin encoding genes but also that it is amplified by participation only of primer 2F. Whereas primer 2R did obviously not bind to the template DNA, primer 2F served as forward as well as reverse primer (figure 4.2.13 A/B: underlined nucleotides). A possible explanation of the bias towards primer 2F which finally led to the exclusively annealing of primer 2F and the amplification of a gene fragment of a gene which does not encode a legumin protein, might be the relatively high annealing temperature of 53 °C at which the PCR was performed. Whereas the theoretically determined annealing temperature (section 3.3.7.3) of primer 2F lies between 68 °C (lowest) and 74 °C (highest), the theoretical annealing temperature for primer 2R is between 62 °C (lowest) and 68 °C (highest) (section 4.2.5.1; table 4.2.5). Furthermore, the last three 3' nucleotides of primer 2F are "TGG" (codon for tryptophan), the 3' end of primer 2R represents with "GAA" a much less tidy 3' end binding site for an oligonucleotide and behaves more like an untidy bound "tail" at the end of primer 2R. These two factors combined might have finally led to the preferable annealing of primer 2F to the template DNA and the more successful enzymatic amplification by the Taq-polymerase at the 3' end of primer 2F which resulted in the amplification of a gene fragment with partial homology to arcelin.

133

### 5.4.2.2 Proposal for a new primer pair for the PCR-amplification of legumin gene fragments from genomic DNA of *P. vulgaris*

The amplification of the 1.65 kb PCR artefact with sequence similarity to arcelin genes showed that the putative reason for the failure of a repeat (figure 4.2.12) of the seemingly successful PCR amplification of a *P. vulgaris* legumin gene fragment (figure 4.2.11) with primer pair 2F/2R might lie in the characteristics of primer 2R (as discussed previously; section 5.4.2.1). Therefore, a new reverse primer has to be chosen for a new attempt to amplify gene fragments of *P. vulgaris* legumin genes. Whereas for the forward primer the same nucleotide sequence is suggested, it is proposed that the reverse primer nucleotide sequence is based on the obtained protein sequence data from the N-terminus of the β-subunit of the identified legumin protein (see section 4.2.2.3), so ensuring the specificity of the oligonucleotide. It is further suggested, that both primers are lacking in any additional oligonucleotide linkers as endonuclease restriction sites at the 5' ends of the specific legumin sequences as these additional oligonucleotide linkers increase the unspecificity of the primer and are often altered in their nucleotide sequence after several PCR cycles so losing their function as restriction sites (see section 4.2.5.2B). In figure 5.4.1 the suggested primer pair is illustrated. The expected specific amplified PCR product(s) would provide sequence information for the α-subunit of *P. vulgaris* legumin(s) which would help to elucidate the relationship of Glycoprotein I to legumin proteins from seeds of *P. vulgaris* (see section 5.1.5). Furthermore, this sequence information would also provide the basis for the taxonomic evaluation of *P. vulgaris* within the legume family, as information about the variable regions would be available which is needed for the evaluation of phylogenetic relationship at a low taxonomic level (e.g. within plant families).

### 5.4.2.3 Genes with homolgy to arcelin genes are also present in the genome of French bean cultivars

Plants have evolved with various mechanisms to protect seeds from predators. One of the main plant defence compounds are the lectin-like proteins (Chrispeels and Raikhel, 1991). Lectins are defined as carbohydrate-binding proteins which bind glycans of glycoproteins, glycolipids or polysaccharides with high affinity (Goldstein and Hayes, 1978). *P. vulgaris* contains especially large amounts of these lectins, which is why common bean has long

acidic
subunit

basic
subunit

S ———————————————————— S

VRI   VRII   HVR

NH2 ▮▮▮ VRI ▮▮▮ VRII ▮▮▮ HVR ▮ COOH   NH2 ▮▮▮ COOH

...E  A  G  L  T  E  T  W...

... G  I  E  E  T  L  C...

5' GGIATTGARGARACCCTITG 3'

3F:  GAAGSTGGDYTCAYTGAGACITGG

3R:  3' CCITAACTYCTYTGGGAIAC 5'

**Figure 5.4.1:** Schematic illustration of the suggested primer pair 3F/3R for a new attempt to amplify legumin gene fragments from genomic DNA of *P. vulgaris*.

The black areas represent conserved regions, the stippled areas variable regions (VRI/II = variable region I/II; HVR = hyper-variable region). The IUB codes are used for the degenerate oligonucleotides. 3F resembles the suggested forward primer (sense-strand) and 3R the suggested reverse primer (antisense-strand).

been known to be toxic toward mammals when eaten raw (Pusztai *et al.*, 1979; Jaffe and Vega-Lette, 1986) and birds (Jayne-Williams and Burgess, 1974). A considerable proportion of these lectins in French bean is formed by phytohemagglutinins (PHA) which represents about 10% of the total seed proteins (Bollini and Chrispeels, 1978). PHA has been shown to bind to the intestinal mucosa of rats, resulting in abnormal development or degradation of the microvilli which finally leads to an inhibition of the absorption of nutrients through the intestinal wall and increases the bacterial colonisation of the small intestine (for a review see Liener, 1986). However, in some cases PHA is not toxic to insects as was demonstrated for the cowpea weevil (Murdock *et al,*. 1990). Although PHA may protect bean seeds from certain insect predators, it is ineffective against the two most important bruchid pests of French bean, the bean weevils, *Acanthoscelides obtectus* (Say) and the Mexican bean weevil, *Zabrotes subfasciatus* (Boheman) (Osborn *et al.*, 1988). However, in contrast to the French bean cultivars, the seeds of many of the wild types of *P. vulgaris* are resistant against *Z. subfasciatus* which is due to the presence of a lectin-like protein, termed arcelin, which has been reported to occur in 10% of the known wild accessions (Chrispeels and Raikhel, 1991) but not in domesticated cultivars of French bean (Osborn *et al.*, 1988). Therefore, Chrispeels and Raikhel (1991) suggested, that all domesticated lines may have developed from wild types of *P. vulgaris* which did not have arcelin or that the arcelin gene was lost or inactivated during domestication.

The fortuitous amplification of a 1.65 kb large PCR product with homology to arcelin genes (figure 4.2.13) may confirm the latter assumption as the sequence regions homologous to arcelin were interrupted by sequence parts without any similarity to arcelin (figure 4.2.13). These additional sequence stretches may be due to insertions which occurred during the evolution and domestication of the modern French bean cultivars. These insertions led then to the inactivation of the arcelin genes. Such an inactivation having occurred to a wild type of *P. vulgaris* could have led to a bean bruchid pest attack and dying out of the mutants. However, due to the protection of such mutants during the domestication process these mutants were propagated due to other qualities (i.e. larger seeds) and resemble the origin of the modern crop plant. But further investigations are necessary to elucidate the possible mechanisms for the inactivation of these genes.

### 5.4.3 Legumin as a phylogenetic marker - A contribution to the elucidation of the taxonomic position of *Phaseolus vulgaris*

During the last years molecular data have been shown to be essential for a modern phylogenetic analysis of plant taxa (Doyle, 1993). In consequence of such additional data the current stage of those phylogenies, which are mainly based on morphological, chemical or immunological markers, will be confirmed or modified (Jensen *et al.*, 1995).

For these systematic investigations molecular phylogenetic markers are frequently used, for example, rbcL within higher plants (Chase *et al.*, 1993) and rRNA within lower plants (e.g. algae (e.g. Friedl *et al.*, 1994) and bryophytes (e.g. Capesius, 1995)). But for a variety of reasons efforts are now increasing to make use of alternative genes. Among several other markers the legumin gene proved to be a quite universal marker for inferring seed plant evolution. As it is already mentioned previously (section 1.4), legumin genes are widely distributed, in gymnosperms as well as in angiosperms. Further legumin genes provide a huge number of single characters like base pairs. Another reason for legumin as the molecular marker of choice is the fact, that legumin genes contain both conserved as well as variable regions. This permits the evaluation of phylogenetic relationships on different taxonomic levels. Finally Jensen and Greven (1984) have already shown by serological studies, that legumins are of high significance for the use as phylogenetic marker in plant systematics.

However, due to the failure to isolate a PCR fragment of a legumin encoding gene it is not possible to contribute to the originally attempted extent to the elucidation of the systematic position of *P. vulgaris*. Based on the amino acid data of the *Phaseolus* legumin described in this work, a certain statement concerning the taxonomic position of *P. vulgaris* is possible, although no statistical evaluation of evolutionary relatedness is achievable. The sequence search of sequence databases revealed that the new legumin sequence shows a higher similarity to soybean legumins than to either of the *Pisum sativum* or *Vicia faba* legumins (figure 4.2.5) confirming the position of the genus *Phaseolus* together with *Glycine* within the tribe *Phaseoleae* whereas the genus *Vicia* and *Pisum* are grouped together in the tribe *Viciae* (Frohne and Jensen, 1992; Sitte *et al.*, 1991; see also section 1.2).

# 6. References

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson (1994). Molecular Biology of he Cell. 3$^{rd}$ edition, Garland Publishing, Inc., New York.

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

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

Altenbach, S. B., C. C. Kuo, L. C. Staraci, K. W. Pearson, C. Wainwright, A. Georgescu and J. Townsend (1992). Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol Biol 18: 235-245.

Anthony, J. L., R. A. Vonder Haar and T. C. Hall (1990). Nucleotide sequence of an α-phaseolin gene from *Phaseolus vulgaris*. Nucl Acids Res 18: 3396.

Apostol, I., P. F. Heinstein and P. S. Low (1989). Rapid stimulation of an oxidative burstduring elicitation of cultured plant cells. Role in defence and signal transduction. Plant Physiol 90: 109-116.

Arahira, M. and C. Fukazawa (1994). Ginkgo 11S seed storage protein family mRNA: unusual Asn-Asn linkage as post-translational cleavage site. Plant Mol Biol 25: 597-605.

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

Asada, K. and M. Takahashi (1987). Production and scavenging of active oxygen in photosynthesis. *In* "Photoinhibition". (D. J. Kyle, C. B. Osmond and C. J. Arntzen, eds.), pp. 227-287. Elsevier Science, Amsterdam.

Badenoch-Jones, J., D. Spencer, T. J. V. Higgins and A. Millerd (1981). The role of glycosylation in storage protein synthesis in developing pea seeds. Planta 153: 201-209.

Bannister, J. V., W. H. Bannister and G. Rotilio (1987). Aspects of the structure, function and application of superoxide dismutase. CRC Crit Rev Biochem 22: 111-180.

Barton, K. A., J. F.Thompson, J. T. Madison, R. Rosenthal, N. P. Jarvis and R. N. Beachy (1982). The biosynthesis and processing of high molecular weight precursors of soybean glycinin subunits. J. Biol. Chem. 257: 6089-6095.

Baumgartner, B. and M. J. Chrispeels (1976). Partial characterisation of a protease inhibitor which inhibits the major endopeptidase present in the cotyledons of mung beans. Plant Physiol 58: 1-6.

Bäumlein, H., H. Braun, I. A. Kakhovskaya and A. D. Shutov (1995). Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. J Mol Evol 41: 1070-1075.

Beauchamp, C. and I. Fridovich (1971). Superoxide dismutase: Improved assay and an assay applicable to acrylamide gels. Analyt Biochem 44: 276-287.

Becker, C., A. D. Shutov, V. H. Nong, V. I. Senyuk, R. Jung, C. Horstmann, J. Fischer, N. C. Nielsen and K. Müntz (1995). Purification, cDNA cloning and characterisation of proteinase B, an asparagine-specific endopeptidase from germinating vetch (*Vicia sativa* L.) seeds. Eur J Biochem 228: 456-462.

Benhamou, N., M. H. A. J. Joosten and P. J. G. M. de Wit (1990). Subcellular localization of chitinase and of its potential substrate in tomato root tissues infected by *Fusarium oxysporium f.sp. radicis-lycopersici*. Plant Physiol. 92: 1108-1120.

Bernardi, R., M. C. Lupi and M. Durante (1990). *Phaseolus coccineus* L. storage proteins. Extraction and characterization. Biologica Plantarum 32: 198-204.

Bernasconi, P., R. Locher, P. E. Pilet, J. Jolles and P. Jolles (1987). Purification and N-terminal amino acid sequence of a basic lysozyme from *Parthenocissus quinquifolia* cultured *in vitro*. Biochim Biophys Acta 915: 254-260.

Berner, R. A. and D. E. Canfield (1989). A new model for atmospheric oxygen over Phanerozoic time. Am J of Science 289: 333-361.

Bewley, J. D. and J. S. Greenwood (1992). Protein storage and utilisation in seeds. *In* "Plant physiology, biochemistry and molecular biology" (D. T. Dennis and D. H. Turpin, eds.), Longman, Harlow.

Bi, Y.-M., P. Kenton, L. Mur, R. Darby and J. Draper (1995). Hydrogen peroxide does not function downstream of salicylic acid in the induction of PR protein expression. Plant J 8: 235-245.

Boller, T. A. Gehri, F. Mauch and U. Vogeli (1983). Chitinase in bean leaves: induction by ethylene, purification, properties, and possible function. Planta 157: 22-31.

Boller, T. (1987). Hydrolytic enzymes in plant disease resistance. *In* "Plant-Microbe Interactions, Molecular and Genetic Perspectives". Volume 3 (Koshuge, T. and E. W. Nester, eds), pp. 385-411. Macmillan, New York.

Boller, T. (1988). Ethylen and the regulation of antifungal hydrolases in plants. *In* "Surveys of Plant Molecular and Cell Biology", Volume 5. (Miflin, B. J., ed.), pp. 145-174. Oxford University Press, Oxford.

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

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

Bollini, R. and A. Vitale (1981). Genetic variability in charge microheterogeneity and polypeptide composition of phaseolin, the major storage protein of Phaseolus vulgaris; and peptide maps of its three major subunits. Physiologia Plant 52: 96-100.

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

Boulter, D. and J. T. Barber (1963). Amino-acid metabolism in germinating seeds of *Vicia faba* L. in relation to their biology. New Phytologist 62: 301-316.

Bowler, C., M. Van Montagu and D. Inze (1992). Superoxide dismutase and stress tolerance. Annu Rev Plant Physiol Plant Mol Biol 43: 83-116.

Braun, H., A. Czihal, A. D. Shutov and H. Bäumlein (1996). A vicilin-like seed protein of cycads: similarity to sucrose binding proteins. Plant Mol Biol 31: 35-44.

Broekart, W. F., L. Van Parijs, F. Leyns, H. Joos and W. J. Peumans (1989). A chitin-binding lectin from stinging nettle rhizomes with antifungal properties. Science 245: 1100-1102.

Broglie, K. E., J. J. Gaynor and R. M. Broglie (1986). Ethylene-regulated gene expression: molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris*. Proc Natl Acad Sci USA 83: 6820-6824.

Broglie, K., I. Chet, M. Holliday, R. Cressman, P. Biddle, S. Knowlton, C. J. Mauvais and R. Broglie (1991). Transgenic plants with enhanced resistance to the fungal pathogen Rhizoctonia solani. Science 254: 1194-1197.

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

Brown, J. W. S., J. R. McFerson, F. A. Bliss and T. C. Hall (1982). Genetic-divergence among commercial classes of *Phaseolus vulgaris* in relation to phaseolin pattern. Hort. Science 17: 752-754.

Brune, D. C. (1992). Alkylation of cysteine with acrylamide for protein sequence analysis. Anal Biochem 207: 285-290.

Bueno, P. and L. A. del Rio (1992). Purification and properties of glyoxisomal cuprozinc superoxide dismutase from watermelon cotyledons (*Citrullus vulgaris* Schrad.). Plant Physiol 98: 331-336.

Bueno, P., J. Varela, G. Gimenez-Gallego and L. A del Rio (1995). Peroxisomal copper, zinc superoxide dismutase - Characterisation of the isoenzyme from watermelon cotyledons. Plant Physiol 108: 1151-1160.

Capesius, I. (1995). A molecular phylogeny of bryophytes based on the nuclear encoded 18S rRNA genes. J Plant Physiol 146: 59-63.

Casey, R., C. Domoney and N. Ellis (1986). Legume storage proteins and their genes. Oxford Surv Plant Mol Cell Biol 3: 1-95.

Chase, M. W. et al. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene. Ann Missouri Bot Gard 80: 528-580.

Chen, Z. and D. Klessig (1991). Identification of a soluble salicylic acid-binding protein that may function in signal transduction in the plant disease-resistance response. Proc Natl Acad Sci USA 90: 9533-9537.

Chen, Z., H. Silva and D. Klessig (1993). Involvment of reactive oxygen species in the induction os systemic acquired resistance by salicylic acid in plants. Science 242: 883-886.

Chou, Q., M. Russel, D. E. Birch, J. Raymond and W. Bloch (1992). Prevention of pre-PCR mispriming and primer dimerization improves low-copy-number amplifications. Nucleic Acids Res 20: 1717-1723.

Chrispeels, M. J. and D. Boulter (1975). Control of storage protein metabolism in the cotyledons of germinating mung beans: role of endopeptidase. Plant Physiol 55: 1031-1037.

Chrispeels, M. J., B. Baumgartner and N. Harris (1976). Regulation of reserve protein metabolism in the cotyledons of mung bean seedlings. Proc Natl Acad Sci USA 73: 3168-3172.

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

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

Chrispeels, M. J. and N. V. Raikhel (1991). Lectins, lectin genes, and their role in plant defence. Plant Cell 3: 1-9.

Chung, C. T., S. L. Niemala and R. H. Miller (1989). One-step preparation of competent cells *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci USA 86: 2172-2175.

Collada, C., R. Casado, A. Fraile and C. Aragoncillo (1992). Basic endochitinases are major proteins in *Castanea sativa* cotyledons. Plant Physiol 100: 778-783.

Collinge, D. B., K. M. Kragh, J. D. Mikkelsen, K. K. Nielsen, U. Rasmussen and K. Vad (1993). Plant chitinases. Plant J 3: 31-40.

Cornelissen, B. J. C. and L. S. Melchers (1993). Strategies for control of fungal diseases with transgenic plants. Plant Physiol 101: 709-712.

Croy, R.R.D. (1977) .Major proteins and proteolytic enzymes in *Phaseolus vulgaris* L. at germination. Ph.D. Thesis, University of Aberdeen.

Croy, R.R.D., G. W. Lycett, J. A. Gatehouse, J. N. Yarwood and D. Boulter (1980a). Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea (*Pisum sativum* L.). I Legumin. Planta 148 49-56.

Croy, R.R.D., G. W. Lycett, J. A. Gatehouse, J. N. Yarwood and D. Boulter (1980b). Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea (*Pisum sativum* L.). II Vicilin. Planta 148: 57-63.

Croy, R. R. D., G. W. Lycett, J. A. Gatehouse, J. N. Yarwood and D. Boulter (1982). Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295: 76-79.

Croy, E. J., T. Ikemura, A. Shirsat and R. R. D. Croy (1993). Plant nucleic acids. *In* "Plant Molecular Biology Labfax". (R. R. D. Croy, ed.), Chapter 2, pp. 21-48. Bios Scientific Publishers and Blackwell Scientific Publishers, Oxford.

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


Dank, N. (1994). Konstruktion einer cDNA-Bank und Sequenzierung drei Legumin-codierender Transkripte aus *Calocedrus decurrens*. Diplomarbeit, University of Bayreuth.


Dawson, R. M. C., D. C. Elliott, W. H. Elliott and K. M. Jones (1989). Data for biochemical research. 3rd edition, p. 537, Oxford University Press, New York.


Dayhoff, M. O. (1978). A model in evolutionary change in proteins: Matrices for detecting distant relationships. *In* "Atlas of protein sequences and structure". (M. O. Dayhoff, ed.), vol. 5, suppl. 3, pp 345-358. National Biochemical Research Foundation, Washington, DC.


Debouck, D. G., O. Toro, O. M. Paredes, W. C. Johnson and P. Gebts (1993). Genetic diversity and ecological distribution of *Phaseolus vulgaris* (*Fabaceae*) in north-western South America. Econ Bot 47: 408-423.


DeCamp, J. D., D. A. Stetler and A. E. DeMaggio (1994). Expression of sporophytic storage proteins in the corm of the quillwort (*Isoetes echinospora* Dur.). Plant physiol 106: 1395-1402.


De Jong, A. J., J. Cordewener, F. L. Schiavo, M. Terzi, J. Vandekerckhove, A. V. Kammen and S. C. De Vries (1992). A carrot somatic embryo mutant is rescued by chitinase. Plant Cell 4: 425-433.


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


Derbyshire, E. D. J. and D. Boulter (1976). Isolation of legumin-like protein from *Phaseolus aureus* and *Phaseolus vulgaris*. Phytochemistry 15: 411-414.

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

Dickinson, C. G., M. P. Scott, E. H. A. Hussein, P. Argos and N. C. Nielsen (1990). Effect of structural modifications on the assembly of a glycinin subunit. Plant Cell 2: 403-413.

Domoney, C. and R. Casey (1990). Another class of vicilin gene in *Pisum*. Planta 182: 39-42.

Doyle, J. J., M. A. Schuler, W. D. Godette, V. Zenger, R. N. Beachy and J. L. Slightom (1986). The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. J Biol Chem 261: 9228-9238.

Doyle, J. J. (1993). DNA, phylogeny, and the flowering of plant systematics. BioScience 43: 380-389.

Doyle, J. J. (1994). Phylogeny of the legume family: an approach to understanding the origins of nodulation. Annu Rev Ecol Syst 25: 325-349.

Dudman, W. F. and A. Millerd (1975). Immunochemical behaviour of legumin and vicilin from *Vicia faba*: a survey of related proteins in the leguminosae subfamily Faboideae. Biochem Systemat Ecol 3: 25-29.

Duranti, M., N. Guerrieri, T. Takahashi and P. Cerletti (1988). The legumin-like storage protein of *Lupinus albus* seeds. Phytochemistry 27: 15-23.

Duranti, M., N. Guerrieri, T. Takahashi, P. Cerletti and G. Vecchio (1992). The legumin precursor from white lupin seed - identity of the subunits, assembly and proteolysis. Eur J Biochem 206: 941-947.

Duranti, M., C. Horstmann, J. Gilroy and R. R. D. Croy (1995). The molecular basis for N-glycosylation in the 11S globulin (legumin) of lupin seed. J Protein Chem 14: 107-110.

Dyer, J. M., J. W. Nelson and N. Murai (1993). Strategies for selecting mutation sites for methionine enhancement in the bean seed storage protein phaseolin. J Prot Chem 12: 545-560.

Evans, I. M., R. R. D. Croy, P. Hutchinson, D. Boulter, P. I. Payne and M. E. Gordon (1979). Cell free synthesis of some storage protein subunits by polyribosomes and RNA isolated from developing seeds of pea (*Pisum sativum* L.). Planta 144: 455-464.

Fairbanks, T., T. L. Steck and D. F. H. Wallach (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry 10: 2606-2617.

Fischer, H., V. Haake, C. Horstmann and U. Jensen (1995). Characterization and evolutionary relationships of *Magnolia* legumin-encoding cDNAs representing two divergent gene subfamilies. Eur J Biochem 229: 645-650.

Fischer, H., L. Chen and S. Wallisch (1996). The evolution of angiosperm seed proteins: A methionine-rich legumin subfamily present in lower angiosperm clades. J Mol Evol (in press).

Fitch, W. H. and F. J. Ayala (1994). The superoxide dismutase molecular clock revisited. Proc Natl Acad Sci USA 91: 6802-6807.

Flach, J., P.-E. Pilet and P. Jolles (1992). What's new in chitinase research? Experentia 48:701-716.

Fukazawa, C., T. Momma, H. Hirano, K. Harada and K. Udaka (1985). Glycinin $A_3B_4mRNA$. Cloning and sequencing of double-stranded cDNA complementary to a soybean storage protein. J Biol Chem 260: 6234-6239.

Friedl, T. (1994). Assessing the relationships of some coccoid green lichen algae and the microthamniales (Chlorophyta) with 18S ribosomal-RNA gene sequence comparison. J Phycology 30: 500-506.

Frohne, D. und U. Jensen (1992). Systematik des Pflanzenreichs. 4$^{th}$ edition, p. 159, Gustav Fischer Verlag, Stuttgart.

Gatehouse, J. A., R. R. D. Croy, H. Morton, M. Tyler and D. Boulter (1981). Characterisation and subunit structure of the vicilin storage proteins of pea (*Pisum sativum* L.). Eur J Biochem 118: 627-633.

Gatehouse, J. A., I. M. Evans, R. R. D. Cro and D. Boulter (1986). Differential expression of genes during legume seed development. Philos Trans R Soc London Ser B 314: 342-353.

Gatehouse J. A., Bown D., Gilroy J., Levasseur M., Castleton J. and T. H. N. Ellis (1988). "Two genes encoding 'minor' legumin polypeptides in pea (*Pisum sativum* L.). Characterization and complete sequence of the LegJ gene. Biochem. J. 250:15-24(1988).

Gatehouse, J. A. and A. H. Shirsat (1992). Control of expression of seed storage protein genes. *In* "Control of plant gene expression". (Verma, D. P. S., ed.). CRC Press, Boca Raton.

Gebts, P., T. C. Osborn, K. Rashka and F. A. Bliss (1986). Phaseolin-protein variability in wild forms and land races of the common bean (*Phaseolus vulgaris*) - evidence for multiple centres and domestication. Econ Bot 40: 451-461.

Gebts, P. and F. A. Bliss (1986). Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ Bot 40: 469-478.

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

Goldstein, I. J. and C. E. Hayes (1978). The lectins: Carbohydrate-binding proteins of plants and animals. Adv Carbohydr Chem Biochem 35: 127-340.

Graham, J. B., R. Dudley, N. M. Aguilar and C. Gans (1995). Implications of the late Paleozoic oxygen pulse for physiology and evolution. Nature 375: 117-120.

Guerche, P., E. R. P. DeAlmeida, M. A. Schwarztein, E. S. Gander, E. Krebbers and G. Pelletier (1990). Expression of the 2S albumin from *Bertholletia excelsa* in *Brassica napus*. Mol Gen Genet 221: 306-314.

Gupta, A. S., R. P. Webb, A. S. Holaday and R. D. Allen (1993). Overexpression of superoxide dismutase protects plants from oxidative stress. Plant Physiol 103: 1067-1073.

Häger, K.-P., H. Braun, A. Czihal, B. Müller and H. Bäumlein (1995). Evolution of seed storage proteins: Legumin genes of *Ginkgo biloba*. J Mol Evol 41: 457-466.

Häger, K.-P., B. Müller, C. Wind, S. Erbach and H. Fischer (1996). Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. FEBS Letters 387: 94-98.

Hall, T. C., R. C. McLeester and F. A. Bliss (1977). Equal expression of the maternal and paternal alleles for polypeptide subunits of the major storage protein of the bean *Phaseolus vulgaris* L. Plant Physiol 59: 1122-1124.

Halliwell, B. (1987). Oxidative damage, lipid peroxidation and antioxidant protection in chloroplasts. Chem Phys Lipids 44: 327-340.

Hara-Nishimura, I., T. Shimada, N. Hiraiwa and N. Nishimura (1995). Vacuolar processing enzyme responsible for maturation of seed proteins. J Plant Physiol 145: 632-640.

Harris, N., J. Henderson, S. J. Abott, J. Mulcrone and J. T. Davies (1993). Seed development and structure. *In* "Seed storage compounds". (Shewry, P. R. and K Stobart, eds.), pp. 1-22. Clarendon Press, Oxford.

Hart, P. J., A. F. Monzingo, M. P. Ready, S. R. Ernst and J. D. Roberts (1993). Crystal structure of an endochitinase from *Hordeum vulgare* L. seeds. J Mol Biol 229: 189-193.

Hart, P. J., H. D. Pfluger, A. F. Monzingo, T. Hollis and J. D. Robertus (1995). The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. J Mo Biol 248: 402-413.

Hayashi, M., H. Mori, M. Nishimura, T. Akazawa and I. Hara-Nishimura (1988). Nucleotide sequence of cloned cDNA for pumpkin 11S-globulin. Eur J Biochem 172: 627-632.

Heim, U., H. Bäumlein and U. Wobus (1994). The legumin gene family: a reconstructed *Vicia faba* legumin gene encoding a high-molecular-weight subunit is related to type B genes. Plant Mol Biol 25: 131-135.

Hendry, G. A. F., W. E. Finch-Savage, P. C. Thorpe, N. M. Atherton, S. M. Buckland, K. A. Nilsson and W. A. Seel (1992). Free radical processes and loss of seed viability during drying in the recalcitrant species *Quercus robur* L. New Phytolgist 122: 273-279.

Hendry, G. A. F. and R. M. M. Crawford (1994). Oxygen and environmental stress in plants - an overview. Proceedings of the Royal Society of Edinburgh 102B: 1-10.

Hendry, G. A. F. (1994). Oxygen and environmental stress in plants: an evolutionary context. Proceedings of the Royal Society of Edinburgh 102B: 155-165.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280: 309-316.

Henrissat, B. and A. Bairoch (1993). New families in the classification of clycosyl hydrolases based on amino acid sequence similarities. Biochem J 293: 781-788.

Herget, T., J. Schell and P. H. Schreier (1990). Elicitor-specific induction of one member of the chitinase genefamily in *Arachis hypogea*. Mol Gen Genet 224: 469-476.

Higgins, D. G. and P. M. Sharp (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244.

Higgins, T. J. V. and D. Spencer (1981). Precursor forms of pea vicilin subunits. Plant Physiol 67: 205-211.

Hinz, G., B. Hoh, I. Hohl and D. G. Robinson (1995). Stratification of storage proteins in the protein storage vacuole of developing cotyledons of *Pisum sativum* L. J Plant Physiol 145: 437-442.

Horton, R. M., H. P. Hoppe and B. M. Conti-Tronconi (1994). Ampligrease - hot start PCR using petroleum jelly. BioTechniques 16: 42-43.

Hunkapiller, M. W. and L. E. Hood (1983). Meth Enzymol 91: 486-493.

Jach, G., B. Gornhardt, J. Mundy, J. Logemann, E. Pinsdorf, R. Leah, J. Schell and C. Maas (1995). Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. Plant J 8: 97-109.

Jackson, P., D. Boulter and D. A. Thurman (1969). A comparison of some properties of vicilin and legumin isolated from seeds of *Pisum sativum*, *Vicia faba* and *Cicer arietinum*. New Phytol 68: 25-33.

Jaffe, W. G. And C. L. Vega Lette (1986). Heat-labile growth-inhibiting factors in bean (*Phaseolus vulgaris*). J Nutr 94: 203-210.

Jayne-Williams, D.-J. and C. D. Burgess (1974). Further observations on the toxicity of navy beans (*Phaseolus vulgaris*) for Japanese Quail (*Coturnix coturnix japonica*). J Appl Bacteriol 37: 149-169.

Jekel, P. A., B. H. Hartmann and J. J. Beintema (1991). The primary structure of hevamine, an enzyme with lysozyme/chitinase activity from *Hevea brasiliensis* latex. Eur J Biochem 200: 123-130.

Jensen, U. and B. Greven (1984). Serological aspects and phylogenetic relations of the *Magnoliidae*. Taxon 33: 563-577.

Jensen, U. and H. Berthold (1989). Legumin-like proteins in Gymnosperms. Phytochemistry 28: 1389 1394.

Jensen, U., H. Fischer, K.-P. Häger and H. Hurka (1995). Perspectives in plant systematics. Giornale Botanico Italiano 129: 29-39.

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

Kami, J. A. and P. Gebts (1994). Phaseolin nucleotide-diversity in *Phaseolus*. 1. Intra-specific diversity in *Phaseolus vulgaris*. Genome 37: 751-757.

Kami, J., V. B. Velasquez, D. G. Debouck and P. Gebts (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc. Natl. Acad. Sci. USA 92: 1101-1104.

Kanematsu, S. and K. Asada (1989). CuZn-superoxide dismutases in rice: occurrence of an active, monomeric enzyme and two types of isozyme in leaf and non-photosynthetic tissues. Plant Cell Physiol 30: 381-391.

Kanematsu, S. and K. Asada (1990). Characteristic amino acid sequences of chloroplast and cytosol isozymes of CuZn-superoxide dismutase in spinach, rice and horsetail. Plant Cell Physiol. 31: 99-112.

Kaplan, L. (1981). What is the origin of the common bean? Economic Botany 35: 240-254.

Karpinski, S., G. Wingsle, O. Olsson and J. E. Hallgren (1992). Characterization of cDNAs encoding CuZn-superoxide dismutases in Scots pine. Plant Mol Biol 18: 545-555.

Kellogg, D. E., I. Rybalkin, S. Chen, N. Mukhamedova, T. Vlasik, P. D. Siebert and A. Chenchick (1994). Taqstart antibody (TM) - Hot start PCR facilitated by a neutralizing monoclonal antibody directed against taq DNA-polymerase. BioTechniques 16. 1134-1137.

Kloz, J. and V. Turkova (1963). Biol Plant 5: 29-36.

Ko, T.-P., J. D. Ng, J. Day, A. Greenwood and A. McPherson (1993). Determination of three crystal structure of canavalin by molecular replacement. Acta Cryst D49: 478-489.

König, R. and P. Gebts (1989). Allozyme diversity in wild *Phaseolus vulgaris* - Further evidence for 2 major centers of genetic diversity. Theor Appl Genet 78: 809-817.

König, R., S. P. Singh and P. Gebts (1990). Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris, Fabaceae*). Econ Bot 44: 50-60.

Kortt, A. A. And J. B. Caldwell (1990). Low molecular weight albumins from sunflower seed: identification of a methionine-rich albumin. Phytochemistry 29: 2805-2810.

Kragh, K. M., S. Jacobsen and J. D. Mikkelsen (1990). Induction, purification and characterisation of barley leaf chitinase. Plant Sci. 71: 55-68.

153

Kreis, M., P. R. Shewry, B. G. Forde, J. Forde and B. J. Miflin (1985a). Structure and evolution of seed storage proteins and their genes, with particular reference to those of wheat, barley and rice. *In* "Oxford Surveys of Plant Cell and Molecular Biology." (B. J. Miflin, ed.), 2, pp. 253-317. Oxford University Press, Oxford.

Kreis, M., B. G. Forde, S. Rahman, B. J. Miflin and P. R. Shewry (1985b). Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183: 499-502.

Lane, B. G. (1991). Cellular desiccation and hydration: developmentally regulated proteins, and the maturation and germination of seed embryos. FASEB J 5: 2893-2901.

Lane, B. G. (1994). Oxalate, germin, and the extracellular matrix of higher plants. FASEB J 8: 294-301.

Larson, R. A. (1988). The antioxidants of higher plants. Phytochemistry 27: 969-978.

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

Lawrence, M.C., T. Izard, M. Beuchat, R.J. Blagrove, and P. M. Colman (1994). Structure of phaseolin at 2.2 Å resolution - Implications for a common Vicilin/Legumin structure and the genetic engineering of seed storage proteins. J Mol Biol 238: 748-776.

Leah, R., H. Tommerup, I. B. Svendsen and J. Mundy (1991). Biochemical and molecular characterisation of three barley seed proteins with antifungal properties. J Biol Chem 266: 1564-1573.

Leprince, O., R. Deltour, P. C. Thorpe, M. N. Atherton and G. A. F. Hendry (1990). The role of free radicals and radical processing systems in loss of desiccation tolerance in germinating maize (*Zea mays*). New Phytologist 116: 573-580.

Liener, I. E. (1986). Nutritional significance of lectins in the diet. *In* "The Lectins". (I. E. Liener, N. Sharon and I. J. Goldstein, eds), pp. 527-552. Academic Press, Inc., San Diego.

Low, P. S. and P. F. Heinstein (1986). Elicitor stimulation of the defense response in cultured plant cells monitored by fluorescent dyes. Arch Biochem Biophys 249: 472-479.

Lycett, G. W., R. R. D. Croy, A. H. Shirsat and D. Boulter (1984). The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L.). Nucleic Acids Res 12: 4493-4506.

Malamy, J., J. P. Carr, D. F. Klessig and I. Raskin (1990). Salicylic acid: A likely endogenous signal in the resistance response of tobacco to viral infection. Science 250: 1002-1004.

Martin, M. (1991). The latex of Hevea brasiliensis contains high levels of both chitinases and chitinases/lysozymes. Pl Physiol 95: 469-476.

Matta, N. K., J. A. Gatehouse and D. Boulter (1981). Molecular and subunit heterogeneity of legumin of *Pisum sativum* L. - a multi-dimensional gel electrophoretic study. J Exp Bot 32: 1295-1307.

Mauch, F. And A. Staehelin (1986). Functional implications of the subcellular localization of ethylene-induced chitinase and beta-1,3-glucanase in bean leaves. Plant Cell 1: 447-457.

McLeester, R. C., T. C. Hall, S. M. Sun and F. A. Bliss (1973). Comparison of globulin proteins from *Phaseolus vulgaris* with those from *Vicia faba*. Phytochemistry 12: 85-93.

Melchers, L. S., M. Apotheker-de Groot, J. A. van der Knaap, A. S. Ponstein, M. B. Sela-Buurlage, J. F. Bol, B. J. C. Cornelissen, P. J. M. van den Elzen and Huub J. M. Linthorst (1994). A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. Plant J 5: 469-480.

Metraux, J. P. and T. Boller (1986). Local and systemic induction of chitinase in cucumber plants in response to viral, bacterial and fungal infections. Physiol Mol Plant Pathol 28: 161-169.

Metraux, J. P., W. Burkhart, M. Moyer, S. Dincher, W. Middlestaedt, S. Williams, G. Payne, M. Carnes and J. Ryals (1989). Isolation of a complementary DNA encoding a chitinase with structural homology to a bifunctional lysozyme/chitinase. Proc Natl Acad Sci USA 86: 896-900.

Miège, M.-N. (1982). Protein types and distribution. In "Nucleic acids and proteins in plants I." (D. Boulter and B. Parthier, eds), pp. 29-345. Springer-Verlag, Berlin.

Momma, T., T. Negoro, H. Hirano, A. Matsumoto, K. Udaka and C. Fukazawa (1985). Glycinin $A_5A_4B_3$ mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean. Eur J Biochem 149: 491-496.

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

Murray, D. R. (1984). Accumulation of seed reserves of nitrogen. In "Seed Physiology" (D. R. Murray, ed), 1, pp. 83-127, Academic Press, Inc., London.

Murray, M. G. And W. F. Thompson (1980). Nucleic Acids Res 8: 4321-4325.

Neuenschwander, U., B. Vernooij, L. Friedrich, S. Uknes, H. Kessmann and J. Ryals (1995). Is hydrogen peroxide a second messenger of salicylic acid in systemic acquired resistance? Plant J 8: 235-233.

Neuhaus,J.-M., P. Ahl-Goy, U. Hinz, S. Flores and F. Meins (1991). High level expression of a tobacco chitinase gene in Nicotiana sylvestris. Susceptibility of transgenic plants to Cercospora nicotianae infection. Plant Mol Biol 16: 141-151.

Nielsen, N. C. (1984). The chemistry of legume storage proteins. Phil Trans R Soc Lond B 304:287-296.

Nielsen, N. C. (1985). Structure and complexity of the 11S polypeptides in soybeans. J Am Oil Chem Soc 62: 1680-1686.

Nielsen, N. C., C. D. Dickinson, T.-J. Cho, V. H. Tanh, B. J. Scallon, R. L. Fischer, T. L. Sims, G. N. Drews and R. B. Goldberg (1989). Characterization of the glycinin gene family in soybean. Plant Cell 1: 313-327.

Osborn, T. B. and G. G. Campbell (1898). J Am Chem Soc 20: 348-362.

Osborne, T. B. (1924). The vegetable proteins. 2nd ed., Longsman-Green, London

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

Pang, P. P., R. E. Pruitt and E. M. Meyerowitz (1989). Molecular cloning, genomic organisation, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana. Plant Mol Biol 11: 805-820.

Parrish, D. J. and A. C. Leopold (1978). On the metabolism of ageing in soybean seeds. Plant Physiol 61: 385-386.

Pernollet, J.C. and J. Mosse (1983). Secondary structure prediction of seed storage proteins. Int J Peptide protein res 22: 456-463.

Pliëtz, P., B. Drescher and G. Damaschun (1987). Relationship between the amino acid sequence and the domain structure of the subunits of the 11S seed globulins. Int J Biol Macromol 9: 161-165.

Pusztai, A. (1966). The isolation of two proteins, Glycoprotein I and a trypsin inhibitor, from the seeds of Kidney Bean (*Phaseolus vulgaris*). Biochem J 101: 379-298.

Pusztai, A. (1968). Hydrogen-ion equilibra of Glycoprotein I from Kidney Bean. Arch Biochem and Biophys 126: 289-298.

Pusztai, A. and W. B. Watt (1970). The isolation and characterization of a major antigenic and non-hemagglutinating glycoprotein from *Phaseolus vulgaris*. Biochem Biophys Acta 207: 413-431.

Pusztai, A. and I. Duncan (1971). Glycoprotein I of *Phaseolus vulgaris* - Homogeneity and enzymatic properties. Biochim Biophys Acta 229: 785-794.

Pusztai, A., E. M. W. Clarke and T. P. King (1979). The nutritional toxicity of *Phaseolus vulgaris* lectins. Proc Nutr Soc 38: 115-120.

Qi, X., K. A. Wilson and A. L. Tan-Wilson (1992). Characterization of the major protease involved in the soybean -conglycinin storage protein mobilization. Plant Physiol 99: 725-733.

Qi, X., R. Chen, K. A. Wilson and A. L. Tan-Wilson (1994). Characterization of a soybean -conglycinin-degrading protease cleavage site. Plant Physiol 104: 127-133.

Ramasarma, T. (1982). Generation of $H_2O_2$ in biomembranes. Biochem Biophys Acta 694: 69-93.

Reichelt, R., K. D. Schwenke, T. König, W. Pähtz and G. Wangermann (1980). Biochem Physiol Pflanz 175: 653-663.

Reuzeau, C. and G. Cavalie (1995). Activities of free radical processing enzymes in dry sunflower seeds. New Phytologist 130: 59-66.

Robinson, D. G, B. Hoh, G. Hinz and B.-K. Jeong (1995). One vacuole or two vacuoles: Do protein storage vacuoles arise *de novo* during pea cotyledon development? J Plant Physiol 145: 654-664.

Rödin, J. and L. Rask (1990). The relationship between mature chains and their precursors of cruciferin, the 12S storage protein of *Brassica napus*. Plant Sci 70: 57-63.

Ryan, C. A. (1989). Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores. BioEssays 10: 20-24.

Saalbach, G., R. Jung, I. Saalbach and K. Müntz (1988). Construction of storage protein genes with increased number of methionine codons and their use in transformation experiments. Biochem Physiol Pflanzen 183: 211-218.

Saalbach, I., T. Pickardt, F. Machemehl, G. Saalbach, O. Schieder and K. Müntz (1994). A chimeric gene encoding the methionine-rich 2S albumin of the Brazil nut (*Bertholletia excelsa* H.B.K.) is stably expressed and inherited in transgenic grain legumes. Mol. Gen. Genetic 242: 226-236.

Saalbach, I., D. Waddell, T. Pickardt, O. Schieder and K. Müntz (1995a). Stable expression of the sulphur-rich 2Salbumin gene in transgenic *Vicia narbonensis* increases the methionine content of seeds. J. Plant Physiol. 145, 647-681.

Saalbach, G., C. Veselin, R. Jung, I. Saalbach, R. Manteuffel, G. Kunze, K. Brambarov and K. Müntz (1995b). Stable expression of vicilin from *Vicia faba* with eight additional single methionine residues but failure of accumulation of legumin with an attached peptide segment in tobacco seeds. Molecular Breeding 1: 245-258.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim (1985). Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350-1354.

Saiki, R. K., D H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and A. H. Erlich (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Saio, K., M. Kamiya and T. Watanabe (1969). Food processing characteristics of soybean 11S and 7S protein 1. Effect of difference of protein components among soybean varieties on formation of Tofu-gel. Agric Biol Chem 33: 1301-1308.

Scharf, S. J., C. M. Long and A. H. Erlich (1988) Sequence analysis of the HLA-DRβ and HLA-DQβ loci from three *Pemphigus vulgaris* patients. Hum Immunology 22: 61-69.

Sakamoto, A., Y. Nosaka and K. Tanaka (1993). Cloning and sequencing analysis of a complementary DNA for manganese-superoxide dismutase from rice (*Oryza sativa* L.). Plant Physiol 103: 1477-1478.

Samac, D. A., C. M. Hironaka, P. E. Vallaly and D. M. Shah (1990). Isolation and characterization of the genes encoding basic and acidic chitinase in *Arabidopsis thaliana*. Plant Physiol 93: 907-914.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Scandalios, J. G. (1993). Oxygen stress and superoxide dismutases. Plant Physiol. 101: 7-12.

Schlumbaum, A., F. Mauch, U. Vogeli and T. Boller (1986). Plant chitinases are potent inhibitors of fungal growth. Nature 324, 365-367.

Scholz, G., R. Manteuffel, R. Müntz and A. Rudolph (1983). Low-molecular-weight polypeptides of vicilin from *Vicia faba* L. are products of proteolytic breakdown. Eur J Biochem 132: 103-107.

Senaratna, T. and A. Borochov (1987). Dessication and free radical mediated changes in plant membranes. J Exp Bot 38: 2005-2014.

Shewry, P. R. (1995). Plant storage proteins. Biol Rev 70: 375-426.

Shewry, P. R., J. A. Napier and A. S. Tatham (1995) Seed storage proteins: Structures and biosynthesis. Plant Cell 7: 945-956.

Shotwell, M. A. and B. A. Larkins (1989). The biochemistry and molecular biology of seed storage proteins. *In* "The biochemistry of plants, a comprehensive treatise". (E. Marcus, ed.), vol. 15, Academic Press, New York, 297-345.

Shutov, A. D.,I. A. Kakhovskaya, H. Braun, H. Bäumlein and K. Müntz (1995). Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. J Mol Evol 41: 1057-1069.

Singh, N. K. and K. W. Shepherd (1985). The structure and genetic control of a new class of disulphide-linked proteins in wheat endosperm. Theor Appl Genet 71: 79-92.

Singh, N. K., K. W. Shepherd, P. Langridge, L. C. Gruen, J. H. Skerritt and C. W. Wrigley (1988). Identification of legumin-like proteins in wheat. Plant Mol Biol 11: 633-639.

Singh, N. K., G. R. Donovan, H. C. Carpenter, J. H. Skerritt and P. Langridge (1993). Isolation and characterization of wheat triticin cDNA revealing a unique lysine-rich repetitive domain. Plant Mol Biol 22: 227-237.

Sitte, P., H. Ziegler, F. Ehrendorfer and A. Bresinsky (1991). Strassburger: Lehrbuch der Botanik. 33$^{rd}$ edition, Gustav Fischer Verlag, Stuttgart.

Slightom, J. L., R. F. Drong, R. C. Klassy and L. M. Hoffman (1985). Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucl Acids Res 13: 6483-6498.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503-517.

Sperisen, C., J. Ryals and F. Jr. Meins (1991). Comparison of cloned genes provides evidence for intergenomic exchange of DNA in the evolution of a glucan endo-1,3-β-glucosidase gene family. Proc Natl Acad Sci USA 88: 1820-1824.

Staswick, P. E., M. A. Hermodson and Niels C. Nielsen (1984). Identification of the cysteines which link the acidic and basic components of the glycinin subunits. J Biol Chem 259: 13431-1343.

Stewart, W. N. and G. W. Rothwell (1993). Paleobotany and the evolution of plants. 2$^{nd}$ edition, Cambridge University Press, New York.

Sun, S. M. and T. C. Hall (1975). Solubility characteristics of globulins from *Phaseolus* seed in regard to their isolation and characterization. J Agric Food Chem 23: 184-189.

Svendsen, I., D. Nicolova, I. Goshev and N. Genov (1989). Isolation and characterization of a trypsin inhbitor from kohlrabi *(Brassica napus*, var. *Rapifera)* belonging to the napin family of storage proteins. Carlsberg Research Communications 54: 231-239.

Takaiwa, F., S. Kikuchi and K. Ono (1987). A rice glutelin gene family - a major type of glutelin mRNAs can be divided into two classes. Mol Gen Genet 208: 15-22.

Thanh, V. H. and K. Shibasaki (1976). Heterogeneity of β-conglycinin. Biochim Biophys Acta 439: 326-338.

Thanh, V. H. and K. Shibasaki (1977). β-conglycinin from soybean proteins. Isolation and immunological and physiochemical properties of the monomeric forms. Biochim Biophys Acta 490: 370-384.

Thomson, J. A, H. E. Schroeder and W. F. Dudman (1978). Cotyledonary storage proteins in *Pisum sativum*. I. Molecular heterogeneity. Aust J Plant Physiol 5: 263-279.

Utsumi, S. (1992). Plant Food Protein Engineering. Adv Food and Nut Res 36: 89-208.

Vaughan, J. G. and K. E. Denford (1968). An acrylamide gel electrophoretic study of the seed proteins of *Brassica* and *Sinapis* species, with special reference to their taxonomic value. J Exp Bot 19: 724-732.

Vogelsang, R. and W. Barz (1992). Purification, characterization and differential hormonal regulation of a beta-1,3-glucanase and two chitinases from chickpea (*Cicer arietinum* L.). Planta 189: 60-69.

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

Webster, J. C. (1958). Nitrogen metabolism in plants. Row-Peterson, Evanston.

Williams, D. M. (1992). DNA analysis: theory. *In* "CLADISTICS - A practical course in systematics". (P. L. Forey, C. J. Humphries, I. L. Kitching, R. W. Scotland, D. J. Siebert and D. M. Williams, eds.), pp. 89 -101. Oxford University Press Inc., New York.

Wind, C. and K.-P. Häger (1996). Legumin encoding sequences from the redwood family (*Taxodiaceae* reveal precursors lacking the conserved ASN-GLY processing site. FEBS Letters 383: 46-50.

Wirth, S. J. and G. A. Wolf (1990). Dye-labelled substrates for the assay and detection of chitinase and lysozyme activity. J. Microbiol. Methods 12: 197-205.

Wright, R. J. (1988). The seed globulins - part II. *In* "Development in Food Proteins". (B. J. F. Hudson, ed.) 6, pp. 119-178. Elsevier, London.

Wubben, J. P, M. H. A. J. Joosten, J. A. L. Van Kan and P. J. G. M. de Wit (1992). Subcellular localization of plant chitinases and 1,3-beta-glucanases in *Cladosporium fulvum* (syn. *Fulvia fulva*)-infected tomato leaves. Physiol Mol Plant Pathol 41: 23-32.

Yamagami, T. and G. Funatsu (1993). The complete amino acid sequence of chitinase-c from the seeds of rye (*Secale cereale*). Biosci Biotechnol Biochem 57: 1854-1861.

Yamagami, T. and G. Funatsu (1994). The complete amino acid sequence of chitinase-a from the seeds of rye (*Secale cereale*). Biosci Biotechnol Biochem 58: 322-329.

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

Zhang, Z., D. B. Collinge and H. Thordal-Christensen (1995). Germin-like oxalate oxidase, a $H_2O_2$-producing enzyme, accumulates in barley attacked by the powdery mildew fungus. Plant J 8: 139-145.

Zirwer, D., D. Gast, H. Welfle, B. Schlesier and K. D. Schwenke (1985). Secondary structure of globulins from plant seeds: a re-evaluation from circular dichroism measurements. Int. J. Macromol. 7: 105-108.

# Appendix

## A. Media, buffers and antibiotics

Glass-distilled water (dH$_2$O) was used for all buffers used for chromatography and other protein work. For the preparation of all buffers used for DNA work the distilled water was further deionised (ddH$_2$O) prior to use.

## SDS-PAGE

SDS-Sample Loading Buffer

| | | |
|---|---|---|
| Distilled water | 3.0 ml | |
| Glycerol | 1.6 ml | |
| 10% (w/v) SDS | 1.6 ml | (2% final concentration) |
| 0.5M Tris-HCl, pH 6.5 | 1.0 ml | (0.063M final conc.) |
| 0.5% (w/v) Bromophenol blue (in water) | 0.4 ml | (0.025% (w/v) final conc.) |
| 2-mercaptomethanol (40%) | 0.4 ml | (2% (v/v) final conc.) |

When SDS-PAGE under non-reducing conditions was performed, water was added instead of 2-mercaptoethanol.

Standard proteins (protein marker)

| | |
|---|---|
| Albumin, bovine | 66 kDa |
| Albumin, egg | 45 kDa |
| Glyceraldehyde-3-phosphate dehydrogenase | 36 kDa |
| Carbonic anhydrase, bovine eruthrocytes | 29 kDa |
| Trypsinogen, PMSF treated | 24 kDa |
| Trypsin inhibitor, soybean | 20 kDa |
| α-Lactalbumin, bovine milk | 14 kDa |

<u>10x SDS-PAGE Buffer</u>

| | |
|---|---|
| Glycine | 144 g |
| Tris | 30 g |
| SDS | 10 g |

in 1L dH$_2$O


**Western blotting and Immunodetection**

<u>10x PBS (Phosphate Buffered Saline)</u>

| | |
|---|---|
| NaCl | 80 g |
| KCl | 2 g |
| Na$_2$HPO$_4$ | 14.4 g |
| KH$_2$PO$_4$ | 2.4 g |

in 1L dH$_2$O (The pH was not adjusted to any certain value)


<u>Blocking buffer</u>

| | |
|---|---|
| PBS buffer(10x) | 100ml/l) |
| Marvel (Semidry milk powder) | 50g/l |
| Tween 20 | 10ml/l |


<u>Antisera buffer:</u>

| | |
|---|---|
| PBS buffer(10x) | 100ml/l) |
| Marvel (Semidry milk powder) | 50g/l |
| Tween 20 | 1ml/l |


<u>Transfer buffer</u>

| | |
|---|---|
| Tris/HCl | 5.81 g |
| Glycine | 2.93 g |
| SDS | 0.37 g |
| Methanol | 200 ml |
| Water | 800 ml |

<u>Ponceau S staining solution</u>

(0.1% (w/v) PonceauS in 1% (v/v) acetic acid)

| | |
|---|---|
| Ponceau S | 100 mg |
| Acetic acid | 1 ml |
| Water | 99 ml |

**Stock Solutions for DNA Work**

1M Tris/HCl (pH 7.5 and 8.0)                88.1 g Tris

in 400 ml ddH$_2$O;

adjust pH to 7.5 or 8.0 and fill volume up to 500 ml

0.5M EDTA, pH 8.0                          93.05 g

solve in 400 ml ddH$_2$O while adjusting pH to 8.0 by adding 10% NaOH;

fill up to 500 ml.

3M NaCl                                    350.64 g

in 500 ml ddH$_2$O.

<u>TE Buffer pH 8.0 (10mM Tris-HCl, 1mM EDTA)</u>

| | |
|---|---|
| 1M Tris-Hcl, pH 8.0 | 1.0 ml |
| 0.5M EDTA, pH 8.0 | 0.2 ml |

in 10 ml ddH$_2$O

X-Gal                                      200 mg

in 10 ml DMF

**Antibiotic Stock Solutions**

Ampicillin was prepared as a 50 mg/ml stock solution in 50% ethanol. The solution has been filter sterilised through a 0.2μ-filter, and normally used at a final concentration of 50 μg/ml.

To avoid satellite-colonies ampicillin has been used in a final concentration of 75 μg/ml in X-Gal-plates. These satellites are caused by β-lactamase in the medium which has been produced by the ampicillin resistant colonies.

**Agarose gel electrophoresis**

<u>10x TAE Buffer</u>

| | |
|---|---|
| Tris | 48.4 g |
| 0.5M EDTA (pH 8.0) | 20 ml |
| Glacial Acetic Acid | 11.42 ml |

in 1L dH$_2$O

<u>6x Agarose Gel Loading Buffer</u>

250mg Bromophenol blue

250mg Xylene cyanol FF

3ml Glycerol

7 ml Water

EtBr-Stocksolution

| | |
|---|---|
| EtBr (aqueous) | 10 mg/ml |

**DNA Extraction from Plant Material**

<u>Buffer B</u>

| | | |
|---|---|---|
| Tris/HCl | 6.05 g | (100mM) |
| NaCl | 40.9 g | (1.4M) |
| EDTA | 3.72 g | (20mM) |
| Hexadecyltrimethyl | | |
|      ammoniumbromide (CTAB) | 10 g | (2% (w/v)) |

in 400 ml ddH$_2$O; adjust pH to pH 8.0;
fill up to 500 ml.

<u>Buffer C</u>

| | | |
|---|---|---|
| Tris/HCl | 3 g | (50 mM) |
| EDTA | 1.86 g | (10 mM) |
| CTAB | 5 g | (1% (w/v)) |

| | |
|---|---|
| RNase A | 5 mg/ml in 0.5x TE, 50 % (v/v) Glycerol |

**Minipreps by Alkaline Lysis**

<u>Cell resuspension solution</u>

| | |
|---|---|
| Tris/HCl (50 mM) | 0.5 ml 1M, pH 7.5 |
| EDTA (10 mM) | 0.2 ml 0.5M, pH8.0 |
| RNase A (100 μg/ml) | 200 μl (5 mg/ml) |

fill up to 10 ml with ddH$_2$O

<u>Cell Lysis Solution</u>

| | |
|---|---|
| NaOH (0.2M) | 1 ml 2M NaOH |
| SDS (1% (w/v)) | 0.1 g |

fill up to 10 ml with ddH$_2$O

<u>Neutralisation solution</u>

| | |
|---|---|
| Potassium acetate (5M) | 6.0 ml |
| Glacial acetic acid | 1.15 ml |
| ddH$_2$O | 2.85 ml |

fill up to 10 ml with ddH$_2$O

**Southern Blotting**

<u>Denaturing Solution</u>

0.5M NaOH

1.5M NaCl

<u>Neutralisation Solution</u>

1M Tris/HCl, pH 7.5

1.5 M NaCl

<u>9x SSC</u>

| | |
|---|---|
| NaCl | 39.45 g |
| Sodium citrate | 19.85 g |

in 500 ml ddH$_2$O.

**Growth Media**

<u>Luria-Bertani Broth</u>

| | |
|---|---|
| Bacto-Tryptone | 10g |
| Yeast Extract | 5g |
| NaCl | 5g |
| in 1L dH$_2$O | |

<u>SOC-Media</u>

| | |
|---|---|
| Bacto-Tryptone | 20g |
| Yeast Extract | 5g |
| NaCl | 0.6g |
| KCl | 0.2g |
| in 1L dH$_2$O | |

<u>Luria-Bertani-Agar Plates</u>

| | |
|---|---|
| Bacto-Tryptone | 10g |
| Yeast-Extract | 5g |
| NaCl | 5g |
| Bacto-Agar | 15g |
| NaOH (5M NaOH) | 580 µl |
| in 1L H$_2$O | |

**X-Gal-Plates**

As LB-Plates; after autoclaving and cooling down to about 40 °C add sterile to 300 ml medium:

| | |
|---|---|
| Ampicillin (50 mg/ml) | 600 µl (100 µg/ml) |
| IPTG (100 mg/ml) | 80 µl (25 µg/ml) |
| X-Gal (40 mg/ml in DMF) | 400 µl (50 µg/ml) |

**B. Preparation of SDS-polyacrylamide gels**

For the preparation of the SDS-polyacrylamide gels, the following gel solutions were prepared depending on the appropriate polyacrylamide concentration:

| | Separating Gel (0.375 M Tris/HCl, pH8.8) | | | Stacking Gel (0.125M Tris/HCl, pH 6.8) |
|---|---|---|---|---|
| Acrylamide (monomer) concentration (% T, 2.67% C)[b] | 15 % | 12 % | 10 % | 4 % |
| Acrylamide | 5.0 ml | 4.0 ml | 3.3 ml | 1.3 ml |
| Distilled water | 2.35 ml | 3.35 ml | 4.0 ml | 6.1 ml |
| 1.5 M TrisHCl, pH 8.8 | 2.5 ml | 2.5 ml | 2.5 ml | |
| 0.5 M Tris/HCl, pH 6.8 | | | | 2.5 ml |
| 10% (w/v) SDS | 100 µl | 100 µl | 100 µl | 100 µl |
| 10% Ammonium persulfate | 50 µl | 50 µl | 50 µl | 60 µl |
| TEMED | 7 µl | 7 µl | 7 µl | 12 µl |
| Total Monomer | 10 ml | 10 ml | 10 ml | 10 ml |

[b]     %T = [(g acrylamide + g bis-acrylamide)/total volume] x 100

%C = [(g bis-acrylamide/(g acrylamide + g bis-acrylamide)] x 100

## C. Preparation of DNA-marker (λ-DNA digested with Pst I)

13 µg λ-DNA were digested with 30 U Pst I in 100 µl reaction volume. After 3 h incubation at 37 °C an aliquot of 5 µl was checked on an agarose gel whether the digestion was completed. If not, another 10 U of Pst I were added followed by incubation for another hour. When the digestion was completed, the reaction volume was incubated at 65 °C for 15 min to denature the enzyme.

The digestion of the λ-DNA resulted in the following fragments which served as standard marker:

| | | |
|---|---|---|
| 11.50 kb | 2.50 kb | 1.10 kb |
| 5.10 kb | 2.14 kb | 0.80 kb |
| 4.75 kb | 1.98 kb | 0.52 kb |
| 4.60 kb | 1.7 kb | 0.45 kb |
| 2.84 kb | 1.15 kb | |

## D. IUB codes for degenerate oligonucleotides

| | | | |
|---|---|---|---|
| M = A/C | R = A/G | V = A/C/G | B = C/G/T |
| W = A/T | S = G/C | D = A/G/T | N = A/C/G/T |
| Y = C/T | K = G/T | H = A/C/T | |

## E. Amino acid - and nculeotide alignment of legume legumin proteins and their encoding cDNAs

i) Alignment of amino acid sequences of legumins from members of the legume family. Sequences were retrieved from the SWISSPROT sequence library. Codes used for individual sequences are as follows (databank accession numbers follow in parentheses): Soybean-legumins: GLC1.PRO, glycinin 1 (P04776); GLC2.PRO, glycinin 2 (P04405); GLC3.PRO, glycinin 3 (P11828); GLC4.PRO, glycinin 4 (P02858); GLC5.PRO, glycinin 5 (P04347); *Pisum*-legumins: PSLEGA2.PRO, legumin A2 (P02857); PSLEGJ.PRO, legumin J (P05692); *Vicia faba*-legumin: VICIALE4.PRO, legumin 4 (P05190).

Sequences were aligned using CLUSTAL (Higgins and Sharp, 1988). Minor corrections were performed manually.

ii) Nucleotide alignment of cDNA sequences of legumins.

Sequences were retrieved from the EMBL sequence library (Heidelberg) and represent cDNA. Codes used for individual sequences are as follows (databank accession numbers follow in parentheses): Soybean-legumins: GLC1.SEQ, glycinin 1 (X15121); GLC2.SEQ, glycinin 2 (X15122); GLC3.SEQ, glycinin 3 (15123); GLC4.SEQ, glycinin 4 (X52863); GLC5.SEQ, glycinin 5 (X79467); *Pisum*-legumins: PSLEGA2.SEQ, legumin A2 (X17193); PSLEGJ.SEQ, legumin J (X07014); *Vicia faba*-legumin: VICIALE4.SEQ, legumin 4 (X03677).

The alignment was achieved as described above (i).

i) Alignment of amino acid sequences of legumins from members of the legume family.

```
              MAKPF--LSL-SLCFLLFSGCCFAFS-REQPQQNECQLQNLNALEPDNRIESEGGLIETWNPNNKEFQCAGVALSRRTLNRNALRRPSYSNAPQEIYIQQ
                        10        20        30        40        50        60        70        80        90        100

GLC1.PRO      MAK----LVF-SLCFLLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIES EGGLIETW NPNNKPFQCAGVALSRCTLNRNALRRPSYTNGPQEIYIQQ  95
GLC2.PRO      MAK----LVL-SLCFLLFSGC---FALREQAQQNECQIQKLNALKPDNRIES EGGFIETW NPNNKPFQCAGVALSRCTLNRNALRRPSYTNGPQEIYIQQ  92
GLC3.PRO      MAK----LVL-SLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIES EGGFIETW NPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIYIQQ  95
GLC4.PRO      MGKPF-TLSLSSLCLLLLSSACFAIS-SS--KLNECQLNNLNALEPDHRVES EGGLIQTW NSQHPELKCAGVTVSKLTLNRNGLHSPSYSPYPRMIIIAQ  96
GLC5.PRO      MGKPFFTLSLSSLCLLLLSSACFAIT-SS--KFNECQLNNLNALEPDHRVES EGGLIETW NSQHPELQCAGVTVSKRTLNRNGLHLPSYSPYPQMIIIAQ  97
PSLEGAG.PRO   MA-KLLALSL-SFCFLLLGGC---FALREQPQQNECQLERLDALEPDNRIES EGGLIETW NPNNKQFRCAGVALSRATLQRNALRRPYYSNAPQEIFIQQ  95
PSLEGA2.PRO   MATKLLALSL-SFCFLLLGGC---FALREQPEQNECQLERLNALEPDNRIES EGGLIETW NPNNKQFRCAGVALSRATLQHNALRRPYYSNAPQEIFIQQ  96
PSLEGJ.PRO    MSKPF--LSLLSLSLLLFASACLATS-SEFDRLNQCQLDSINALEPDHRVES EAGLTETW NPNHPELKCAGVSLIRRTIDPNGLHLPSFSPSPQLIFIIQ  97
VICIALE4.PRO  MSKPF--LSLLSLSLLLFTSTCLATS-SEFDRLNQCRLDNINALEPDHRVES EAGLTETW NPNHPELRCAGVSLIRRTIDPNGLHLPSYSPSPQLIYIIQ  97


              GKGIFGMIFPGCPETFEEPQESQQR----GQSQRPQDRHQKIRRFREGDIIAVPTGVPYWMYNNGDTPVVAVSLLDTSSFNNQLDQMPRRFYLAGNPEQE
                        110       120       130       140       150       160       170       180       190       200

GLC1.PRO      GKGIFGMIYPGCPSTFEEPQQPQQR----GQSSRPQDRHQKIYNFREGDLIAVPTGVAWWMYNNEDTPVVAVSIIDTN SLENQLDQMPRRFYLAGNQEQE 191
GLC2.PRO      GNGIFGMIFPGCPSTYQEPQESQQR----GRSQRPQDRHQKVHRFREGDLIAVPTGVAWWMYNNEDTPVVAVSIIDTN SLENQLDQMPRRFYLAGNQEQE 188
GLC3.PRO      GSGIFGMIFPGCPSTFEEPQQ---K----GQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMYNNEDTPVVAVSLIDTN SFQNQLDQMPRRFYLAGNQEQE 188
GLC4.PRO      GKGALGVAIPGCPETFEEPQEQSNRRGSRSQKQQLQDSHQKIRHFNEGDVLVIPPSVPYWTYNTGDEPVVAISLLDTS NFNNQLDQTPRVFYLAGNPDIE 196
GLC5.PRO      GKGAIGFAFPGCPETFEKPQQQSSRRGSRSQ-QQLQDSHQKIRHFNEGDVLVIPPGVPYWTYNTGDEPVVAISLLDTS NFNNQLDQNPRVFYLAGNPDIE 196
PSLEGAG.PRO   GNGYFGMVFPGCPETFEEPQESEQ-----GEGRRYRDRHQKVNRFREGDIIAVPTGIVFWMYNDQDTPVIAVSLTDIR SSNNQLDQMPRRFYLAGNHEQE 190
PSLEGA2.PRO   GNGYFGMVFPGCPETFEEPQESEQ-----GEGRRYRDRHQKVNRFREGDIIAVPTGIVFWMYNDQDTPVIAVSLTDIR SSNNQLDQMPRRFYLAGNHEQE 191
PSLEGJ.PRO    GKGVLGLSFPGCPETYEEPRSSQSRQESR---QQQGDSHQKVRRFRKGDIIAIPSGIPYWTYNHGDEPLVAISLLDTS NIANQLDSTPRVFYLGGNPETE 194
VICIALE4.PRO  GKGVIGLTLPGCPQTYQEPRSSQSRQGSR---QQQPDSHQKIRRFRKGDIIAIPSGIPYWTYNNGDEPLVAISLLDTS NIANQLDSTPRVFYLVGNPEVE 194


              FLEYQ-QQQQGGSQ----------GQHQQEEE----NEGGSILSGFSSEFLAQAFNVNEDIAKRLQGPNEEE-KGAIVTVKGGLSVISPPTQEQQQ----
                        210       220       230       240       250       260       270       280       290       300

GLC1.PRO      FLKYQ--QEQGGHQSQK-------GKHQQEEE----NEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQ---- 275
GLC2.PRO      FLKYQ-QQQQGGSQSQK-------GK-QQEEE----NEGSNILSGFAPEFLKEAFGVNMQIVRNLQGENEEEDSGAIVTVKGGLRVTAPAMRKPQQ---- 272
GLC3.PRO      FLQYQPQKQQGGTQSQK-------GKRQQEEE----NEGGSILSGFAPEFLEHAFVVDRQIVRKLQGENEEEEKGAIVTVKGGLSVISPPTEEQQQ---- 274
GLC4.PRO      YPETMQQQQQQKSH-----GGRKQGQHQQEEE----EEGGSVLSGFSKHFLAQSFNTNEDIAEKLESPDDE--RKQIVTVEGGLSVISPKWQEQQDEDED 285
GLC5.PRO      HPETMQQQQQKSH-----GGRKQGQHQQEEE----E-GSVLSGFSKHFLAQSFNTNEDTAEKLRSPDDE--RKQIVTVEGGLSVISPKWQEQ--EDED 282
PSLEGAG.PRO   FLQY--QHQQGG---------------KQEQE----NEGNNIFSGFKRDYLEDAFNVNRHIVDRLQGRNEDEEKGAIVKVKGGLSIISPPEKQARHQRGS 269
PSLEGA2.PRO   FLRY--QHQQGG---------------KQEQE----NEGNNIFSGFKRDFLEDAFNVNRHIVDRLQGRNEDEEKGAIVKVKGGLSIISPPEKQARHQRGS 270
PSLEGJ.PRO    FPETQEEQQGRHRQKHSYPVGRRSGHHQQEEESEEQNEGNSVLSGFSSEFLAQTFNTEEDTAKRLRSPRDE--RSQIVRVEGGLRIIKPK--------- 283
VICIALE4.PRO  FPETQEEQQERHQQKHSLPVGRRGGQHQQEEESEEQKDGNSVLSGFSSEFLAHTFNTEEDTAKRLRSPRDK--RNQIVRVEGGLRIINPE--------- 283
```

```
             RDEEEEDEQEESQP--RCS-G--EEEEEEEDEEDDPRCQRP------------------------------K--R----------------------------QGRNGLEETICT
                    310         320         330         340         350         360         370         380         390         400

GLC1.PRO     RPQEEEEEEEDEKP--QCK------------GKDKHCQRP------RGSQ------------------SKSRR----------------------------NGIDETICT 318
GLC2.PRO     ---EEDDDDEEEQP--QCV------------ETDKGCQR----------Q------------------SKRSR------------------------------NGIDETICT 308
GLC3.PRO     RP------EEEEKP--DCD------------EKDKHCSQ------S--------------------------R--------------------------------NGIDETICT 304
GLC4.PRO     EDEDDEDEQIPSHPPRRPSHGKREQDEDEDEDEDKPRPSRPSQGKRNKTGQDEDEDEDEDQPRKSREWRSKKTQPRRPRQEEPRERGCETRNGVEENICT 385
GLC5.PRO     EDEDEEYEQTPSYPPRRPSHGKHEDDEDEDEEEDQPRPDHPPQ--------------------------RPSRPEQQEPRGRGCQTRNGVEENICT 352
PSLEGAG.PRO  RQEEDEDE--EKQP--RHQRGSRQEEEEDEDEERQPRHQRR------RGEEEEEDKKERGGSQKGKSRR------------------QGDNGLEETVCT 340
PSLEGA2.PRO  RQEEDEDEDEERQP--RHQRGSRQEEEEDEDEERQPRHQRR------RGEEEEEDKKERGGSQKGKSRR------------------QGDNGLEETVCT 343
PSLEGJ.PRO   GKEEEEKEQSHSHSHR---EEKEEEEEEEDEEEKQRSE------------------------------------------------ERKNGLEETICS 330
VICIALE4.PRO G-------------------QQEEEEEEEEEKQRSE---------------------------------------------------QGRNGLEETICS 311


             MKLRENIGQPSSPDIYNPQAGRISTATSLDLPALRWLKLSAEYGSLYKNAMFVPHYNLNANSIIYALRGRARVQVVNCNGNAVFDGELQEGQLLVVPQNF
                    410         420         430         440         450         460         470         480         490         500

GLC1.PRO     MRLRHNIGQTSSPDIYNPQAGSVTTATSLDFPALSWLRLSAEFGSLRKNAMFVPHYNLNANSIIYALNGRALIQVVNCNGERVFDGELQEGRVLIVPQNF 418
GLC2.PRO     MRLRQNIGQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQYGSLRKNAMFVPHYTLNANSIIYALNGRALVQVVNCNGERVFDGELQEGGVLIVPQNF 408
GLC3.PRO     MRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALSWLKLSAQFGSLRKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIVPQNF 404
GLC4.PRO     LKLHENIARPSRADFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSVIYTRGQGKVRVVNCQGNAVFDGELRRGQLLVVPQNF 485
GLC5.PRO     MKLHENIARPSRADFYNPKAGRISTLNSLTLPALRQFGLSAQYVVLYRNGIYSPHWNLNANSVIYTRGKGRVRVVNCQGNAVFDGELRRGQLLVVPQNF 452
PSLEGAG.PRO  AKLRLNIGPSSSPDIYNPEAGRIKTVTSLDLPVLRWLKLSAEHGSLHKNAMFVPHYNLNANSIIYALKGRARLQVVNCNGNTVFDGELEAGRALTVPQNY 440
PSLEGA2.PRO  AKLRLNIGPSSSPDIYNPEAGRIKTVTSLDLPVLRWLKLSAEHGSLHKNAMFVPHYNLNANSIIYALKGRARLQVVNCNGNTVFDGELEAGRALTVPQNY 443
PSLEGJ.PRO   AKIRENIADAARADLYNPRAGRISTANSLTLPVLRYLRLSAEYVRLYRNGIYAPHWNINANSLLYVIRGEGRVRIVNCQGNTVFDNKVRKGQLVVVPQNF 430
VICIALE4.PRO LKIRENIAQPARADLYNPRAGSISTANSLTLPILRYLRLSAEYVRLYRNGIYAPHWNINANSLLYVIRGEGRVRIVNSQGNAVFDNKVTKGQLVVVPQNF 411


             VVAAQSGSDN-FEYVVFKTNDRAGISNLAG-NSVFNALPSEVLANTFNLRQQQARQLKNNNPFGFLVPPQ---ESQRRAVAXX
                    510         520         530         540         550         560         570         580

GLC1.PRO     VVAARSQSDN-FEYVSFKTNDTPMIGTLAGANSLLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQ---ESQKRAVA•• 497
GLC2.PRO     AVAAKSQSDN-FEYVSFKTNDRPSIGNLAGANSLLNALPEEVIQHTFNLKSQQARQVKNNNPFSFLVPPQ---ESQRRAVA•• 487
GLC3.PRO     AVAARSQSDN-FEYVSFKTNDRPSIGNLAGANSLLNALPEEVIQQTFNLRRQQARQVKNNNPFSFLVPPK---ESQRRVVA•• 483
GLC4.PRO     VVAEQAGEQ-GFEYIVFKTHHNAVTSYL---KDVFRAIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNPESQQGSPRVKVA•• 564
GLC5.PRO     VVAEQGGEQ-GLEYVVFKTHHNAVSSYI---KDVFRAIPSEVLSNSYNLGQSQVRQLKYQGNSGPLVNP•• 519
PSLEGAG.PRO  AVAAKSLSDR-FSYVAFKTNDRAGIARLAGTSSVINNLPLDVVAATFNLQRNEARQLKSNNPFKFLVPAR---ESENRASA•• 519
PSLEGA2.PRO  AVAAKSLSDR-FSYVAFKTNDRAGIARLAGTSSVINNLPLDVVAATFNLQRNEARQLKSNNPFKFLVPAR---QSENRASA•• 522
PSLEGJ.PRO   VVAEQAGEEEGLEYVVFKTNDRAAVSHV---QQVFRATPSEVLANAFGLRQRQVTELKLSGNRGPLVHPRSQSQSH•• 505
VICIALE4.PRO VVAEQAGEEEGLEYLVFKTNDRAAVSHV---QQVFRATPADVLANAFGLRQRQVTELKLSGNRGPLVHPQSQSQSN•• 486
```

ii) Nucleotide alignment of cDNA sequences of legumins.

```
              10        20        30        40        50        60        70        80        90        100
GLC1.SEQ     ATGGCCA---AGCTTGTT-CTTTCCCTTTGTTTCCTTCTTTTCAGTGGCT----GCT---TCGCCTTGAG------AGAGCAGGCACAGCAA=---AATG  79
GLC2.SEQ     ATGGCCA---AGCTTGTT-CTTTCCCTTTGTTTCCTTCTTTTCAGTGGCT----GCT---TCGCTCTGAG------AGAGCAGGCACAGCAA----AATG  79
GLC3.SEQ     ATGGCTA---AGCTTGTT-CTTTCCCTTTGTTTTCTGCTTTTCAGTGGCT----GCTGCTTCGCTTTCAGTTTCAGAGAGCAGCCACAGCAA----AACG  88
GLC4.SEQ     ATGGGG---AAGCCCTTCACTCTCTCTCTTTCTTCCCTTTGCTTGCTACTCTTGTCGAGTGCATGCTTTGCTATTAGCTCCAG------CAAGCTCAACG  91
PSLEGA2.SEQ  ATGGCTACTAAGCTTCTTGCACTTTCTCTTTCATTCTGTTTTCTACTTTT----GGGGGGCTGTTTTGC-TTTGAGAGAACAGCCAGAGCAA----AATG  91
PSLEGJ.SEQ   ATGTCC---AAACCTT---CTCTATCTTTGTTTTCACTTTCCTTGCTACTCTTTGCAAGCGCATGTTTAGCAACTCGCTCTGAGTTTGACAGACTCAACC  94
VICIALE4.SEQ ATGTCC---AAACCTT---TTCTATCTTTGCTTTCACTTTCCTTGCTTCTCTTTACAAGCACATGTTTAGCAACTAGCTCTGAGTTTGACAGACTCAACC  94


              110       120       130       140       150       160       170       180       190       200
GLC1.SEQ     AGTGCCAGATCCAAAAGCTGAATGCCCTCAAACCAGGTAACCGTATAGAGTCG|GAAGGTGGGTTCATTGAGACCTGG|AACCCTAACAACAAGCCATTCCA  179
GLC2.SEQ     AGTGCCAGATCCAAAAGCTGAATGCCCTCAAACCGGATAACCGTATAGAGTCG|GAAGGTGGGTTCATTGAGACATGG|AACCCTAACAACAAGCCATTCCA  179
GLC3.SEQ     AGTGCCAGATCCAACGCCTCAATGCCCTAAAACCGGATAACCGTATAGAGTCA|GAAGGTGGCTTCATTGAGACATGG|AACCCTAACAACAAGCCATTCCA  188
GLC4.SEQ     AGTGCCAACTCAACAACCTCAACGCGTTGGAACCCGACCACCGCGTTGAGTTC|GAAGGTGGTTTGATTCAAACATGG|AACTCTCAACACCCTGAGCTGAA  191
PSLEGA2.SEQ  AGTGCCAGCTAGAACGCCTCAATGCCCTCGAGCCTGATAACCGTATAGAATCG|GAAGGTGGGCTCATTGAGACCTGG|AATCCCAACAACAAGCAATTCCG  191
PSLEGJ.SEQ   AATGCCAACTAGACAACATCAATGCATTGGAACCTGACCACCGTGTTGAGTCG|GAAGCCGGTCTCACTGAGACATGG|AATCCAAATAACCCTGAGCTAAA  194
VICIALE4.SEQ AATGCCGGTTAGACAACATCAATGCATTGGAACCTGACCACCGTGTGGAGTCC|GAAGCTGGTCTCACTGAGACATGG|AATCCAAATCACCCTGAGCTAAG  194


              210       220       230       240       250       260       270       280       290       300
GLC1.SEQ     GTGTGCCGGTGTTGCCCTCTCTCGCTGCACCCTTAACCGCAATGCCCTTCGTAGACCTTCCTACACCAACGGTCCCCAGGAAATCTACATACAACAAGGT  279
GLC2.SEQ     GTGTGCCGGTGTTGCCCTCTCTCGCTGCACCCTTAACCGCAATGCCCTTCGTAGACCTTCCTACACCAACGGTCCCCAGGAAATCTACATACAACAAGGT  279
GLC3.SEQ     GTGTGCCGGTGTTGCCCTCTCTCGCTGCACCCTCAACCGCAACGCCCTTCGCAGACCTTCCTACACCAACGCTCCCCAGGAGATCTACATCCAACAAGGT  288
GLC4.SEQ     ATGCGCCGGTGTCACTGTTTCCAAACTCACCCTCAACCGCAATGGCCTCCACTTGCCATCTTACTCACCTTATCCCCGGATGATCATCATCGCCCAAGGG  291
PSLEGA2.SEQ  ATGTGCTGGTGTGGCCCTCTCTCGTGCTACCCTTAACATAACGCCCTTCGCAGACCTTACTACTCCAATGCTCCCCAAGAAATTTTCATCCAACAAGGT  291
PSLEGJ.SEQ   ATGCGCTGGTGTGTCACTTATCAGACGCACCATTGACCCTAATGGACTCCACTTGCCATCTTTCTCACCCTCTCCACAGTTGATTTTCATCATCCAAGGA  294
VICIALE4.SEQ ATGTGCCGGTGTGTCTCTTATCAGACGCACCATTGACCCTAATGGACTCCACTTGCCATCTTACTCACCCTCTCCACAGTTGATTTACATCATCCAAGGA  294


              310       320       330       340       350       360       370       380       390       400
GLC1.SEQ     AATGGTATTTTTGGCATGATATTCCCGGGTTGTCCTAGCACTTATCAAGAGCCGCAAGAATC-TCAG-CAACGAGGAC--GAAG--CCAGAGGCCCCAA-  373
GLC2.SEQ     AATGGTATTTTTGGCATGATATTCCCGGGTTGTCCTAGCACTTATCAAGAGCCGCAAGAATC-TCAG-CAACGAGGAC--GAAG--CCAGAGGCCCCAA-  373
GLC3.SEQ     AGTGGTATTTTTGGCATGATATTCCCGGGTTGTCCTAGCACATTTGAAGAGCC----------TCAA-CAAAAAGGAC--AAAG--CAGCAGGCCCCAA-  373
GLC4.SEQ     AAAGGAGC-ACTGCAGTGCAAG--CCAGGATGTCCTGAGACGTTTGAGGAGCCACAAGAACAATCAAACAGAAGAGGCTCAAGGTCGCAGAAGCAGCAGC  388
PSLEGA2.SEQ  AATGGATATTTTGGGATGGTATTCCCCGGTTGTCCTGAGACCTTTGAAGAGCCACAAGAATC-TGAA-CAAGGAGAGG--GACG--C---AGGTACAGA-  382
PSLEGJ.SEQ   AAGGGTGTTCTTGGACTTTCACTTCCTGGTTGTCCCGAGACTTATGAAGAGCCACGT---TCATC--ACAATCTAGA--CAAGGATCCAG--GCAGCAAC  385
VICIALE4.SEQ AAGGGTGTTATTGGACTTACACTTCCTGGTTGTCCCCAGACTTACCAAGAGCCACGT---TCATC--ACAATCTAGA--CAAGGATCCAG--GCAGCAAC  385
```

```
                410       420       430       440       450       460       470       480       490       500
GLC1.SEQ      -----GACCGTCACCAAAAGGTACATCGCTTCAGAGAGGGTGATTTGATCGCAGTGCCTACTGGTGTTGCATGGTGGATGTACAACAATGAAGACACTCC 467
GLC2.SEQ      -----GACCGTCACCAAAAGGTACATCGCTTCAGAGAGGGTGATTTGATCGCAGTGCCTACTGGTGTTGCATGGTGGATGTACAACAATGAAGACACTCC 467
GLC3.SEQ      -----GACCGTCACCAGAAGATCTATCACTTCAGAGAGGGTGATTTGATTGCAGTGCCAACCGGTTTTGCATACTGGATGTACAACAATGAAGACACTCC 467
GLC4.SEQ      TACAGGACAGTCACCAGAAGATTCGTCACTTCAATGAAGGAGACGTACTCGTGATTCCTCCTGGTGTTCCTTACTGGACCTATAACACTGGCGATGAACC 488
PSLEGA2.SEQ   -----GACAGACATCAAAAGGTTAACCGATTCAGAGAGGGTGATATCATTGCAGTTCCTACTGGTATTGTATTTTGGATGTACAACGACCAAGACACTCC 476
PSLEGJ.SEQ    AAGGTGACAGTCACCAGAAGATTCGTCGATTCAGAAAAGGTGATATCATTGCCATTCCATCGGGAATTCCTTATTGGACATATAACCATGGCGATGAACC 485
VICIALE4.SEQ  AACCTGACAGTCACCAGAAGATTCGCCGATTCAGGAAAGGTGATATCATTGCCATTCCATCGGGAATTCCATATTGGACATATAACAATGGCGATGAACC 485

                510       520       530       540       550       560       570       580       590       600
GLC1.SEQ      TGTTGTTGCCGTTTCTATTATTGACACCAAC AGCTTGGAGAACCAGCT CGACCAGATGCCTAGGAGATTCTATCTTGCTGGGAACCAAGAGCAAGAGTTT 567
GLC2.SEQ      TGTTGTTGCCGTTTCTATTATTGACACCAAC AGCTTGGAGAACCAGCT CGACCAGATGCCTAGGAGATTCTATCTTGCTGGGAACCAAGAGCAAGAGTTT 567
GLC3.SEQ      TGTTGTTGCCGTTTCTCTTATTGACACCAAC AGCTTCCAGAACCAGCT CGACCAGATGCCTAGGAGATTCTATCTTGCTGGGAACCAAGAGCAAGAGTTT 567
GLC4.SEQ      AGTTGTTGCCATCAGTCTTCTTGACACCTCT AACTTCAATAACCAGCT TGATCAAACCCCTAGGGTATTTTACCTTGCTGGGAACCCAGATATAGAGTAC 588
PSLEGA2.SEQ   AGTTATTGCCGTCTCTCTTACTGACATTAGA AGCTCCAATAACCAGCT TGATCAGATGCCTAGGAGATTCTATCTTGCTGGGAACCACGAGCAAGAGTTT 576
PSLEGJ.SEQ    TCTTGTTGCCATTAGCCTTCTTGACACTTCC AACATTGCAAACCAGCT CGATTCAACCCCAAGAGTATTTTACCTTGGTGGAAACCCAGAAACAGAGTTC 585
VICIALE4.SEQ  GCTTGTTGCCATTAGTCTTCTTGACACTTCC AACATTGCAAACCAGCT TGATTCAACCCCAAGAGTATTTTACCTTGTTGGGAACCCGGAGGTAGAGTTC 585

                610       620       630       640       650       660       670       680       690       700
GLC1.SEQ      C---------TAAAATATCAGC---AGCA-GCAGC-----------------AAGG-A-GGTTCCCA--------------------AAGCCAGAAAG-GAAAGC------A 620
GLC2.SEQ      C---------TAAAATATCAGC---AGCA-GCAGC-----------------AAGG-A-GGTTCCCA--------------------AAGCCAGAAAG-GAAAGC------A 620
GLC3.SEQ      C---------TACAGTATCAGCCACAGAA-GCAGC-----------------AAGG-A-GGTACTCA--------------------AAGCCAGAAAG-GAAAGCGT---CA 626
GLC4.SEQ      CCAGAGACCATGCAACAACAACAACAGCAGAAAAGTCATGGTGGACGCAAGCAGGGGCAACACCAGCAGGAGGAAGAGGAAGAAGGTGGCAGCGTGCTCA 688
PSLEGA2.SEQ   C---------TACGATACCAGCATCAACAAGGAGG-----------------AAAGCA-AGAACAAG-------------------AAAATGAAGGC-AACAACATTTTCA 640
PSLEGJ.SEQ    CCCGAAAC------ACAGGAGGAACAACAAGGAAGGCATCG---GCAAAAGCATAGTTACC--CTGTTGGACGTAGGAGTGGA---------------- 658
VICIALE4.SEQ  CCTGAAAC------ACAGGAGGAGCAGCAAGAAAGACATCA---ACAAAAACATAGTTTAC--CAGTTGGACGTCGGGGTGGA---------------- 658

                710       720       730       740       750       760       770       780       790       800
GLC1.SEQ      A------CAAGAAGAA-------GAAAACG------AAGG-AAGCAA-CATAT--TGAGTGG-CTTCG--CCCCTGAATTC-----------------TTGAAA- 682
GLC2.SEQ      A------CAAGAAGAA-------GAAAACG------AAGG-AAGCAA-CATAT--TGAGTGG-CTTCG--CCCCTGAATTC-----------------TTGAAA- 682
GLC3.SEQ      G------CAAGAAGAA-------GAAAACG------AAGG-AGGCAG-CATAT--TGAGTGG-CTTCG--CCCCGGAATTC-----------------TTGGAA- 688
GLC4.SEQ      GTGGCTTCAGCAAACACTTCTTGGCACAATCCTTCAACACCAACGAGGACATAGCTGAGAAACTTCAGTCTCCAGACGACGAAAGGAAGCAGATCGTGAC 788
PSLEGA2.SEQ   GTGGCTTCAAGAGGGATTTCTTGGAAGATGCTTTCAACGTGAACAGGCATATAGTAGACAGACTTCA--AGGCAGGAATGAA--------DACGAAGA- 729
PSLEGJ.SEQ    --------------------------CATCACCAACAAGAA-------GAGGAA--------TCCGAAGAACAAAACGAAG------GTAAC 702
VICIALE4.SEQ  --------------------------CAGCACCAACAAGAA-------GAGGAA--------TCTGAAGAACAAAAGGACG------GTAAC 702
```

```
                810        820        830        840        850        860        870        880        890        900
GLC1.SEQ     -----GAAG----------CGTT--CGGCGTGAACATGCAGAT--AGTG--A--GAAACCTAC----AA--GGTGAGAACGAAGAGGAGGATA-------- 746
GLC2.SEQ     -----GAAG----------CGTT--CGGCGTGAACATGCAGAT--AGTG--A--GAAACCTAC----AA--GGTGAGAACGAAGAGGAGGATA-------- 746
GLC3.SEQ     -----CATG----------CGTT--CGTCGTGGACAGGCAGAT--AGTG--A--GAAAGCTAC----AA--GGTGAGAACGAAGAGGAAGAGA-------- 752
GLC4.SEQ     AGTGGGAAGGAGGTTCAGCGTTATCAGCCCCAAGTGGCAAGAACAACAAGATGAAGATGAAGATGAAGACGAAGATGATGAAGATGAACAAATTCCCTCT 888
PSLEGA2.SEQ  -----GAAGGGAGC----CATTGTCAAAGTGAA-AGGTGGACTCAGCATCA--TAAGCCCACCCGAGA--AGCAAGCGCGCCACCAGAGAGG-------- 807
PSLEGJ.SEQ   AGCGTG-------CTGAGTGGCGTCAGCTCAGAGTTTTTAGCACAA--ACGT-----TCAACAC-----TGAAGAGGATACAG-CGAAGAGACT-----T 777
VICIALE4.SEQ AGCGTT-------TTGAGTGGCTTCAGTTCAGAGTTTTTAGCACAC--ACAT-----TCAACAC-----CGAAGAGGATACAG-CTAAGAGACT-----T 777


                910        920        930        940        950        960        970        980        990       1000
GLC1.SEQ     ---------------GTGGAGCCATTGTGACAGTGAA---------AG-GAGGTCTAAG--AGTCACAGCTC----------CAGCCA---TGAGGAAGCC 806
GLC2.SEQ     ---------------GTGGAGCCATTGTGACAGTGAA---------AG-GAGGTCTAAG--AGTCACAGCTC----------CAGCCA---TGAGGAAGCC 806
GLC3.SEQ     ---------------AGGGTGCCATTGTGACAGTGAA---------AG-GAGGTCTCAG--CGTGATAAGCC----------CACCCA---CGGAAGAGCA 812
GLC4.SEQ     CACCCTCCTCGCCGACCAAGCCATGGAAAGCGTGAACAAGACGAGGACGAGGACGAAGATGAAGATAAACCTCGTCCTAGTCGACCAAGCCAAGGAAAGC 988
PSLEGA2.SEQ  ---------------CAGCACAAGAGGAAGATGAA---------GATGAAGATGAAG--AGAGGCAGCCG----------CGCCAC---CAGAGAGGCA 868
PSLEGJ.SEQ   CGATCTCCACGA-GAC--------GAAAGGAGT---CAAATTG----TGCG-----AGTTGAGGGAGGTCTCCGCATTA-TCAACC---CCAAGGGGA-- 851
VICIALE4.SEQ CGTTCTCCACGC-GAC--------AAAAGGAAT---CAAATCG----TGAG-----AGTTGAGGGCGGTCTCCGCATTA-TCAACC---CTGAGGGG--- 850


                1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
GLC1.SEQ     ACAG-CAAGAAGAAGATGA---TGATGATGAGGAAGAGCAGCCACAGTG---C---------------GTGGAGA----------CAGACAA---AGGTT 871
GLC2.SEQ     ACAG-CAAGAAGAAGATGA---TGATGATGAGGAAGAGCAGCCACAGTG---C---------------GTGGAGA----------CAGACAA---AGGTT 871
GLC3.SEQ     GCAA-CAA-----AGAC-------CCGAGGAAGAGGAGAAGCCAGATTG---T---------------GACGAGA----------AAGACAA---ACATT 868
GLC4.SEQ     GTGAACAAGACCAGGACCAGGACGAGGACGAAGATGAAGATGAAGATCAACCTCGCAAGAGCCGCGAATGGAGATCGAAAAAGACACAACCCAGAAGACC 1088
PSLEGA2.SEQ  GCAGACAAGAGGAAGAGGAAGATGAAGATGAAGAGAGGCAGCCGCGTCA---TCAAAGGAG----AAGAGGAGAGGAGGAAGAAGAAGACAAGAAAGAGC 961
PSLEGJ.SEQ   ---------------------AGGAAGAAGAAGAAGAAAAAGAACAGAGTCATTCTCAC-----TCTC--ACAGAGAGGAGGAGGAAGAAGA---AGAAGA-- 921
VICIALE4.SEQ ------------------------------------------------------------------------C--A------GCAGGAAGAAGAAGA---AGAAGA-- 873


                1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
GLC1.SEQ     GCC------------AACGCCAAAGCAAAAGG--AGCAGA-AATGGCATTGATGAGACCATTTGCACAATGAGACTTCGCCAAAACATTGGTCAGAATTC 956
GLC2.SEQ     GCC------------AACGCCAAAGCAAAAGG--AGCAGA-AATGGCATTGATGAGACCATTTGCACAATGAGACTTCGCCAAAACATTGGTCAGAATTC 956
GLC3.SEQ     GCC------------AAAGCCAAAGCA----------GA-AATGGCATTGACGAGACCATTTGCACAATGAGACTTCGCCACAACATTGGCCAGACTTC 944
GLC4.SEQ     -TAGACAAGAAGAACCACGTGAAAGAGGATGCGAGACAAGAAACGGGGTTGAGGAAAATATCTGCACCTTGAAGCTTCACGAGAACATTGCTCGCCCTTC 1187
PSLEGA2.SEQ  GCCGCGGCAGCCAAAAAGGCAAAAGCAGAAGGCAAGGAGACAATGGGCTTGAGGAAACAGTTTGCACTGCTAAACTTCGATTGAACATTGGCCCGTCTTC 1061
PSLEGJ.SEQ   --AGATGAGGAGAAACA-----AAGAAGTGAGGAAAGAAAGAATGGTTTGGAAGAAACTATCTGTAGTGCCAAAATTCGAGAGAACATTGCGGACGCTGC 1013
VICIALE4.SEQ --GGAGGAGGAGAAGCA-----GAGAAGTGAGCAGGGAAGGAATGGTTTGGAAGAAACCATCTGTAGTCTCAAGATTCGTGAGAACATTGCTCAGCCTGC 965
```

```
              1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
GLC1.SEQ     ATCACCTGACATCTACAACCCTCAAGCTGGTAGCATCACAACCGCCACCAGCCTTGACTTCCCAGCCCTCTGGCTTCTCAAACTCAGTGCCCAGTATGGA  1056
GLC2.SEQ     ATCACCTGACATCTACAACCCTCAAGCTGGTAGCATCACAACCGCCACCAGCCTTGACTTCCCAGCCCTCTGGCTTCTCAAACTCAGTGCCCAGTATGGA  1056
GLC3.SEQ     ATCACCTGACATCTTCAACCCTCAAGCTGGTAGCATCACAACCGCTACCAGCCTCGACTTCCCAGCCCTCTCGTGGCTCAAACTCAGTGCCCAGTTTGGA  1044
GLC4.SEQ     ACGCGCTGACTTCTACAACCCTAAAGCTGGTCGCATTAGTACCCTCAACAGCCTCACCCTCCCAGCCCTCCGCCAATTCCAACTCAGTGCCCAATATGTT  1287
PSLEGA2.SEQ  ATCACCAGACATCTACAACCCTGAAGCTGGTAGAATCAAAACTGTTACCAGCCTGGACCTCCCAGTTCTCAGGTGGCTCAAACTAAGTGCTGAGCATGGA  1161
PSLEGJ.SEQ   AGGTGCCGACCTCTATAACCCACGTGCTGGTCGTATCAGAACTGCAAACAGTTTAACTCTCCCAGTCCTCCGCTATTTACGCCTCAGCGCTGAGTATGTT  1113
VICIALE4.SEQ ACGTGCAGACCTCTATAACCCACGTGCCGGTAGTATCAGCACCGCAAACAGTTTAACCCTTCCAATCCTCCGCTATTTACGCCTCAGTGCCGAATATGTT  1065


              1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
GLC1.SEQ     TCACTCCGCAAGAATGCTATGTTCGTGCCACACTACACCCTGAACGCGAACAGCATAATATACGCATTGAATGGGCGGGCATTGGTACAAGTGGTGAATT  1156
GLC2.SEQ     TCA----------ATGCTATGTTCGTGCCACACTACACCCTGAACGCGAACAGCATAATATACGCATTGAATGGGCGGGCATTGGTACAAGTGGTGAATT  1146
GLC3.SEQ     TCACTCCGCAAGAATGCTATGTTCGTGCCACACTACAACCTGAACGCAAACAGCATAATATACGCATTGAATGGACGGGCATTGGTACAAGTGGTGAATT  1144
GLC4.SEQ     GTCCTCTACAAGAATGGAATTTACTCTCCACATTGGAATCTGAATGCAAACAGTGTGATCTATGTGACTCGAGGACAAGGAAAGGTTAGAGTTGTGAACT  1387
PSLEGA2.SEQ  TCTCTCCACAAAAATGCTATGTTTGTGCCTCACTACAACCTGAATGCAAACAGTATAATATACGCATTGAAGGGACGTGCAAGGCTACAGTAGTGAACT  1261
PSLEGJ.SEQ   CGTCTCTACAGGAATGGTATATATGCTCCACACTGGAACATAAACGCCAACAGTCTGCTGTACGTGATTAGAGGAGAAGGAAGAGTTAGGATTGTGAACT  1213
VICIALE4.SEQ CGTCTTTACAGGAATGGTATATATGCTCCACACTGGAACATAAACGCAAACAGTCTGCTGTACGTGATAAGAGGAGAAGGAAGAGTTAGGATTGTGAACT  1165


              1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
GLC1.SEQ     GCAATGGTGAGAGAGTGTTTGATGGAGAGCTGCAAGAGGGAGGGGTGCTGATCGTTCCACAAAACTTTGCGGTGGCTGCA-AAATCCCAGAGCGATAAC-  1255
GLC2.SEQ     GCAATGGTGAGAGAGTGTTTGATGGAGAGCTGCAAGAGGGAGGGGTGCTGATCGTTCCACAAAACTTTGCGGTGGCTGCA-AAATCCCAGAGCGATAAC-  1245
GLC3.SEQ     GCAATGGTGAGAGAGTGTTTGATGGAGAGCTGCAAGAGGGACAGGTGTTAATTGTGCCACAAAACTTTGCGGTGGCTGCA-AGATCACAGAGCGACAAC-  1243
GLC4.SEQ     GCCAAGGGAATGCAGTGTTCGACGGTGAGCTTAGGAGGGGACAATTGCTGGTGGTACCACAGAACTTCGTGGTGGCGGAGCAAGCCGGAGA---ACAAGG  1484
PSLEGA2.SEQ  GCAATGGCAACACCGTGTTTGATGGAGAGCTAGAAGCCGGACGTGCATTGACAGTGCCACAAAACTATGCTGTGGCTGCA-AAGTCACTAAGCGACAGG-  1360
PSLEGJ.SEQ   TCCAAGGAGACGCAGTGTTCGACAACAAGGTCAGAAAGGGACAGTTGGTGGTGGTACCACAAAACTTTGTGGTGGCGGAACAAGCTGGGGAGGAAGAAGG  1313
VICIALE4.SEQ CCCAAGGAAACGCAGTGTTCGACAACAAGGTGACAAAGGGACAGTTGGTGGTGGTACCACAAAACTTTGTGGTGGCGGAACAAGCTGGAGAGGAAGAAGG  1265


              1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
GLC1.SEQ     -TTTGAGTATGTGTCATTCAAGACCAATGATAGACCCTCGATCGGAAACCTTGCAGGGGCAAACTCATTGTTGAACGCATTGCCAGAGGAAGTGATT-CA  1352
GLC2.SEQ     -TTTGAGTATGTGTCATTCAAGACCAATGATAGACCCTCGATCGGAAACCTTGCAGGGGCAAACTCATTGTTGAACGCATTGCCAGAGGAAGTGATT-CA  1342
GLC3.SEQ     -TTCGAGTATGTTTCATTCAAGACCAATGATAGACCCTCGATCGGCAACCTTGCAGGTGCAAACTCATTGTTGAACGCATTGCCGGAGGAAGTGATT-CA  1340
GLC4.SEQ     ATTCGAATACATAGTATTCAAGACACACCACAACGCAGTCACTAGCTA-CTTGAAGGAT--------GTGTTTAGGGCAATTCCCTCAGAGGTTCTTGCC  1575
PSLEGA2.SEQ  -TTCTCATATGTAGCATTCAAGACCAATGATAGAGCTGGTATTGCAAGACTTGCAGGGACATCATCAGTTATAAATAATCTGCCGTTGGATGTTGTGGCA  1458
PSLEGJ.SEQ   ATTAGAGTATGTGGTGTTCAAGACAAATGACAGAGCTGCGGTTAGCCA-CGTACAACAG--------GTGCTTAGGGCCACTCCTGCAGAGGTTCTTGCA  1404
VICIALE4.SEQ ATTAGAGTATCTTGTGTTCAAGACAAATGACAGAGCTGCTGTTAGCCA-CGTACAGCAG--------GTGTTTAGGGCCACTCCTGCAGATGTTCTTGCA  1356
```

```
                    1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
GLC1.SEQ      GCACACTTTTAACCTAAAGAGCCAGCAGGCCAGGCAGGTGAAGAACAACAACCCTTTCAGCTTCCTTGTTCCACCTCAGGAGTCTCAGAGGAGAG-CTGT  1451
GLC2.SEQ      GCACACTTTTAACCTAAAGAGCCAGCAGGCCAGGCAGGTGAAGAACAACAACCCTTTCAGCTTCCTTGTTCCACCTCAGGAGTCTCAGAGGAGAG-CTGT  1441
GLC3.SEQ      GCAAACTTTTAACCTAAGGAGGCAGCAGGCCAGGCAGGTCAAGAACAACAACCCTTTCAGCTTCCTGGTTCCACCTAAGGAGTCTCAGAGGAGAG-TTGT  1439
GLC4.SEQ      C-ATTCTTACAACCTTCGACAGAGTCAAGTGTCTGAGCTTAAGTATGAAGGAAATTGGGGTCCTTTGGTCAACCCTGAGTCTCAACAAGGCTCACCCCGT  1674
PSLEGA2.SEQ   GC-TACATTCAACCTGCAGAGGAATGAGGCAAGGCAGCTCAAGTCCAACAATCCCTTCAAATTTCTAGTTCCAGCTCGTCAGTCTGAGAACAGAG-CTTC  1556
PSLEGJ.SEQ    A-ATGCTTTTGGTCTTCGTCAACGCCAAGTCACGGAGTTAAAGCTCAGTGGAAACCGTGGCCCTCTGGTTCACCCTCAGTCGCAAT----CTCAATCT--  1498
VICIALE4.SEQ  A-ATGCTTTTGGTCTTCGTCAGCGCCAAGTCACAGAGTTAAAGCTCAGTGGAAACCGTGGCCCTCTGGTTCACCCTCAGTCTCAAT----CTCAATCT--  1450


                    1710
GLC1.SEQ      G-------GCTTA''                                          1459
GLC2.SEQ      G-------GCTTAG''                                         1450
GLC3.SEQ      G-------GCT''                                            1445
GLC4.SEQ      GTTAAAGTCGCATAA''                                        1691
PSLEGA2.SEQ   G-------GCTTAG''                                         1565
PSLEGJ.SEQ    -----CAT''                                               1502
VICIALE4.SEQ  -----AATTG----A''                                        1457
```

F. Publication of parts of this thesis in the *Journal of Plant Physiology* (in press).

# Legumin Proteins from Seeds of *Phaseolus vulgaris* L.

**Martin Mühling, John Gilroy and Ronald R. D. Croy**

Department of Biological Sciences, University of Durham,

South Road, Durham, DH1 3LE, U.K.


Corresponding author:   Dr. Ronald R. D. Croy

Department of Biological Sciences

University of Durham

South Road, Durham,

DH1 3LE, England, U.K.

Tel. 0044-(0)191-374-2433

Fax: 0044-(0)191-374-2417

E-mail:   R.R.D.Croy@durham.ac.uk

## Summary

Legumin proteins have been partially purified from seeds of the Common or French Bean (*Phaseolus vulgaris* L.). Characteristic of legumins, these are high molecular weight globulin proteins consisting of disulphide-linked α- and β-subunits. The N-terminal sequences of both α- and β-subunits of one of the legumin proteins were determined by automated protein sequencing. The legumin sequence exhibits close similarity to the legumin sequences of other crop plants in the legume family and particularly to soybean legumin. While the β-subunits show only a single polypeptide size class of ~21kDa the α-subunits show a range of sizes from 55kDa down to 35kDa. This paper provides the first unequivocal proof of the presence of this class of globulin in *Phaseolus vulgaris*.

**Key words:** Leguminosae, *Phaseolus vulgaris*, legumin, seed storage proteins, 11S globulins.

**Abbreviations:** SDS-PAGE = Sodium dodecyl sulphate polyacrylamide gel electrophoresis

$M_r$ = relative molecular mass

**Introduction**

In addition to providing the newly germinated seedling with a source of nitrogen, carbon and sulphur, legume seed storage proteins are also rich sources of protein for human nutrition and animal feeds. Therefore much research has been devoted to investigating their properties and distribution in crop plants. The first studies were undertaken on the major crop plants in the legume family (for a review see Derbyshire *et al.*, 1976) and for this purpose seeds of *Glycine max* (soybean), *Vicia faba* (faba or broad bean), *Pisum sativum* (common pea) and *Phaseolus vulgaris* (common or french bean) have long been used as model plants and the major seed storage proteins, vicilins (7S globulins) and legumins (11S globulins) were first described in several of these species (Osborne and Campbell, 1898). These proteins are of widespread distribution occurring in seeds of angiosperms (Luthe, 1992) and conifers (Wind and Häger, 1996) as well as in 'ancient' plant species such as *Ginkgo biloba* (Häger *et al.*, 1995). As a result of the large amount of sequence data accumulated on both protein types it has been demonstrated that legumins and vicilins are most likely derived from a common ancestral gene existing before speciation and subsequently they have evolved independently to the variable modern versions found in the present day seed-bearing species (Argos *et al.*, 1985; Lawrence *et al.*, 1994; Shutov *et al.*, 1995).

Legumin proteins are hexameric molecules of average $M_r$ about 360 kDa comprising of six pairs of subunits, each pair synthesized as a contiguous, precursor molecule which is proteolytically processed into an acidic $\alpha$-subunit of about 40kDa and a smaller, basic $\beta$-subunit with an average $M_r$ of 21kDa. Both polypeptides remain linked together by a disulfide bond (Staswick *et al.*, 1985). The latter property gives legumin-like proteins the characteristic pattern when separated by 2-dimensional SDS-PAGE under non-reducing (first dimension) and reducing (second dimension) conditions and has often been used as the criterion to demonstrate the presence of legumin-like proteins in newly studied plant species (Matta *et al.*, 1981).

It has been shown that legumin represents a major storage protein in the soybean, broad bean and pea seeds (Derbyshire *et al.*, 1976). In soybean, for example, legumin and vicilin proteins together account for about 70% of the total protein (Hill and Breidenbach, 1974). The ratio of legumin to vicilin varies between 1:1 and 3:1 depending on the cultivar (Nielsen, 1984). A similar situation

2

exists in almost every other legume species studied although the ratio of the two protein types varies considerably. However, the principal protein in the seeds of *Phaseolus vulgaris* is vicilin, known by the trivial name of phaseolin, a 7S globulin which alone accounts for about 50% of total proteins in the cotyledons and phytohemagglutinins or lectins which make up 10% of the total cotyledonary proteins (Bollini and Chrispeels, 1978). Based on ultracentrifugation experiments, Danielsson (1949) reported the presence of an 11S globulin fraction from *P.vulgaris*, the same size as the legumin-type proteins. Derbyshire and Boulter, (1976), employed zonal isoelectric precipitation to separate out different *P.vulgaris* globulin fractions on the basis of differential solubility. One fraction containing a 340kDa (11S) protein, when analysed by SDS-PAGE analysis contained polypeptides of sizes (34, 37kDa and 21kDa) similar to legumin subunits, which were linked by disulphide bonds. However they only succeeded in determining the N-terminal amino acid from each subunit thus failing to confirm the identity of the putative legumin protein. Doubt has been cast on the existence of the protein by serological investigations which have showed no immunological cross-reactivity between any *P. vulgaris* seed proteins and legumins from *Vicia faba* (Dudman and Millerd, 1975) or *Pisum sativum* (Kloz and Turkova, 1963). McLeester *et al.*, (1973) claimed to have isolated a *Phaseolus vulgaris* legumin although this was later shown to be the 7S globulin, vicilin (Sun and Hall (1975). Despite numerous other attempts to demonstrate legumin-like seed proteins in *P. vulgaris*, it has largely been an unresolved question as to whether this species and its closely related members actually contain legumin proteins or not. This and the fact that the final unequivocable proof for the presence of legumin in French Bean has never be achieved have led several authors to suggest that this species does not contain legumin (Miège, 1982; Shewry, 1995). We now report for the first time unequivocal evidence for the legumin proteins in seeds of *P. vulgaris*, in the form of N-terminal sequences for the disulphide-linked α- and β- subunits. This further suggests that legumin is most likely distributed throughout all members of the genus Phaseolus. Although there are still several (especially monocotyledonous) plants in which legumin has not yet been demonstrated, this result further confirms the hypothesis of the widespread and probable universal occurrence of legumin in angiosperms (Luthe, 1992; Derbyshire *et al.*, 1976).

## Materials and Methods

*Plant materials:* Mature seeds of *Phaseolus vulgaris* L. var. Processor were obtained from Hurst, Gunson, Cooper & Taber Ltd. (Witham, Essex).

*Partial purification of legumin:* Seeds were ground to a fine powder using an analytical, precision mill (Janke & Kunkel, Germany). All extraction steps were performed at 4°C. A crude protein extract was prepared using Tris-glycine extraction buffer (10mM Tris/HCl, pH 8.2, 80 mM glycine, 0.5M NaCl) as described by Jensen and Berthold (1989). The ratio of seed material to buffer was 1:4 (w/v). The mixture was centrifuged (20,000g x 30min) and the supernatant ( 100ml ) was dialysed at 4°C against 8L of buffer (0.1M Tris/HCl, pH 8.0, 0.1M NaCl) for 48 h with one change. The dialysed protein extract was filtered (0.22μ filter) and 5ml applied onto a sephacryl S-300 (Pharmacia Fine Chemicals, Uppsala) column (2.5cm x 110cm) equilibrated with a Tris-buffer (0.1 M Tris/HCl, pH 8.0, 0.1 M NaCl). The proteins were eluted with the same buffer at a flow rate of 0.2 ml/min and 2ml fractions were collected. The fractions were measured for protein by absorbance at 280nm and analysed quantitatively by 2-dimensional SDS-PAGE. Fractions containing disulfide-linked polypeptides were pooled, dialysed against distilled water and freeze dried. Globulin proteins were precipitated from total protein extracts by dialysis for 48h, against 0.033M acetate buffer, pH5.0 (Pusztai and Watt, 1970). The precipitated proteins were recovered by centrifugation, briefly washed in acetate buffer and then dialysed against distilled water prior to freeze drying.

*Electrophoretic analysis:* SDS-PAGE was carried out according to Laemmli (1970), on 12% (w/v) polyacrylamide gels (0.5 mm thick) in a Mini-Protean II apparatus (Bio-Rad). Protein samples were dissolved by boiling for 5min with 2% (w/v) SDS in the absence (non-reducing) or the presence (reducing) of 2% (v/v) 2-mercaptoethanol. Non-reduced samples were analysed on separate gels to avoid reduction of disulfide-linked polypeptides through diffusion of 2-mercaptoethanol from adjacent samples. Following electrophoresis, gels were silver-stained using a silver-stain kit (Bio-Rad) or stained with 1% (w/v) coomassie blue R250 in 40% (v/v) methanol, 1% (v/v) acetic acid followed by destaining in the same solvent. Two-dimensional SDS-PAGE analyses were also performed in vertical minigels. After the first dimension (non-reducing) separation, the individual tracks were excised and incubated in 50 mM dithiothreitol at 40°C for 45 min to reduce disulphide-linked

4

polypeptides. Each gel track was laid horizontally on top of a 0.75 mm thick separating gel without a stacking gel and electrophoresed as before. Completed, two-dimensional gels were silver stained.

*Automated N-terminal sequencing:* Polypeptides for N-terminal sequencing were separated by PAGE under reducing or non-reducing conditions and blotted onto a polyvinylidenedifluoride (PVDF) membrane (Promega) by the semi-dry blotting method using CAPS-buffer (10mM 3-[cyclohexylamino]-1-propanesulfic acid (Sigma), pH 11; 10% MeOH) as a transfer buffer. Protein bands were located by briefly staining the membranes with 0.1% (w/v) coomassie blue R250 in 40% (v/v) MeOH, 1% (v/v) acetic acid followed by destaining in 50% (v/v) aqueous MeOH solution. N-terminal amino acid sequence analyses were performed essentially according to Hunkapiller and Hood (1983), on a Model 477A pulsed-liquid protein sequencer (Applied Biosystems) equipped with an on-line phenylthiohydantoin analyzer, Model 120A (Applied Biosystems). Cysteine was identified as Pth-Cys S-propionamide formed during PAGE by reaction with residual unpolymerized acrylamide in the gel with thiol-groups of cysteine residues (Brune, 1992).

*Data analysis:* N-terminal legumin protein sequences from representative legume species were retrieved from the SWISSPROT sequence database. The selected species were *Glycine max* L., *Vicia faba* L. and *Pisum sativum* L. ( accession numbers: glycinin 1: P04776; glycinin 2: P04121; glycinin 3: P11828; glycinin 4: P02858; glycinin 5: P04347; *Pisum sativum* - leg A2: P02857; *Pisum sativum* - leg J: P05692; *Vicia faba* - leg 4: P05190). The sequences were aligned using the Clustal method in DNAStar software package (Lasergene, Ltd.). To this alignment the new sequence data for *Phaseolus vulgaris* legumin were added and aligned manually.

## Results and Discussion

### *Legumin-like proteins exist in seeds of Phaseolus vulgaris* L.

Gel filtration of a total protein extract from seeds of *Phaseolus vulgaris* on sephacryl S-300 resulted in the separation of the two major seed proteins - vicilin and phytohaemagglutinin ( $M_r$'s 120-150kDa ), from proteins of higher molecular weight. The molecular weight range for the proteins in these higher molecular weight fractions corresponded to about 3-400kDa as calculated from standard proteins used to calibrate the gel filtration column. Subsequent examination of these high molecular weight proteins by 1D- and 2D-gel analysis under non-reducing (first dimension) and reducing (second dimension) conditions, indicated the presence of multiple disulfide-linked polypeptides (figure 1). The sizes of these were 54kDa, 47kDa, 43kDa, 37kDa and 35kDa, linked to polypeptides of 21kDa, corresponding to typical legumin-like α- (acidic) and β- (basic) subunits. In order to confirm the identity of these polypeptides as components of a putative *Phaseolus* legumin protein, N-terminal sequencing of selected polypeptides and polypeptide pairs blotted from reducing and non-reducing SDS gels, was performed. The 21kDa polypeptide identified on the 2D gels (fig 1) gave the first 20 amino acids from the N-terminus (fig 2B) which were immediately recogniseable as the highly conserved legumin β-subunit N-terminal sequence typical of most legumins (Lawrence *et al.*, 1994). The α - and β - subunit pair comprising a 75 kDa protein band on non-reducing SDS gels and producing a 54kDa (α - subunit) and 21kDa (β - subunit) band under reducing conditions, was also subjected to sequencing which provided N-terminal sequences from both polypeptides simultaneously. The same 20 N-terminal amino acids derived from the separated β - subunit were confirmed from this sequencing run and in addition 20 residues were obtained from the 54kDa polypeptide which were identified as the conserved sequence of the N-terminus of the legumin α - subunit. These data confirm the identity of a *Phaseolus vulgaris* legumin-type protein.

### *P.vulgaris legumin appears to be a typical legume legumin*

The alignment of the *P.vulgaris* legumin N-terminal sequences with those of legumin proteins from other legume species shows that there is close similarity between these protein sequences (figure 2).

6

It can be further deduced from the 2-D gel analysis of the *Phaseolus* legumin fraction that

there are at least six size classes of $\alpha$ - subunits ranging from 54kDa down to about 35kDa (figure 1),

the largest component (54kDa) appeared as the major constituent. However, there appears to be only

a single size class of $\beta$ - subunit ($M_r$ of 21kDa). This is highly characteristic of legumin proteins, in

which the $\beta$-subunits are highly conserved both in size and sequence while the $\alpha$-subunits are much

more heterogeneous exhibiting extensive size and sequence variation (Matta *et al.*, 1981; Lawrence *et*

*al.*, 1994). To investigate any sequence heterogeneity in these different size classes of $\alpha$-polypeptides,

the first six N-terminal amino acids for the $\alpha$ - and the $\beta$ - subunits of each of the five other legumin

disulphide linked subunit pairs were also determined (figure 1: arrows). All proteins showed exactly

the same N-terminal sequences for both the $\alpha$ - and $\beta$ - subunits as was elucidated for the 75 kDa

legumin (see figs 2A and 2B). Whilst it is not inconceivable that all six subunit pairs are derived

from different genes but still share the same N-terminal sequences this is somewhat unusual since

members of the legumin family from the same species usually show some sequence variation, even

within the initial six N-terminal residues. This is particularly evident within the $\alpha$-subunits, for

example as seen in the soybean legumins (fig 2A). A further explanation of this phenomenon may be

offered by the observations of Duranti *et al.*, (1992), who showed that proteolytic processing of lupin

legumin and prolegumin by enzymes present in developing seeds, could lead to an orderly cascade

degradation resulting in size variants of the $\alpha$-polypeptides. Work in our laboratory (unpublished

observations) has established that the initial stages of pea legumin degradation during germination

involve removal of a C-terminal polypeptide of about 100 amino acids, from the $\alpha$-subunits, resulting

in two different sizes of $\alpha$-polypeptides with the same N-terminal sequence. Thus mechanisms are

known whereby different size classes of $\alpha$-subunits may arise with identical N-terminal sequences and

which could therefore account for some of the observed size heterogeneity in the *Phaseolus* legumin

$\alpha$-polypeptides. Derbyshire and Boulter (1976) determined the N-terminal amino acids for the

*Phaseolus* legumin protein subunits as threonine and glycine (which agree with our results) and in

addition, leucine, from which we assume that there is at least one other legumin-type protein present.

Further experiments are necessary in order to fully elucidate the assembly and processing of

*Phaseolus* legumin and to confirm the nature of the different $\alpha$-subunit size variants.

7

Based on the gel filtration protein profiles and quantitation of the SDS gel analysis (data not shown), the ratio of these legumin-like proteins to vicilin proteins in seeds of *P. vulgaris* is estimated to be about 1:15 meaning that legumin accounts for only about 3% of total seed proteins in the cotyledons of the investigated french bean variety. Unlike the situation in other legumes where the proportion of legumin is much higher and may approach a legumin to vicilin ratio of as much as 1:1 or 2:1 and a content of 38-47 % of the total seed protein (eg. soybean and pea; Saio *et al.*, 1969); this makes *Phaseolus* legumin a relatively minor seed protein fraction and may explain the difficulty previous researchers have had in isolating these proteins.

SDS-PAGE analysis of the reduced and non-reduced proteins followed by Western blotting and reaction with affinity-purified anti-*Pisum* legumin antibodies failed to show cross-reactivity with any of the *Phaseolus* protein components. This confirms similar observations by other groups, of the absence of immunological-related proteins (Dudman and Millerd, 1975) and indicates that the *Phaseolus* legumin is sufficiently different structurally as to have no common epitopes with other legume legumins. A similar situation exists between soybean legumin and pea legumin and between legumins from other legume species (Kloz and Turkova, 1963).

*Phaseolus* legumin is a unique legumin protein. Due to the fact that only N-terminal sequences are available no statistical evaluation of evolutionary relatedness is possible. However, based on similarity searches of the EMBL sequence database (Heidelberg), the new sequence shows a higher similarity to soybean legumin than to either of the *Pisum sativum* or *Vicia faba* legumins. A full characterisation of this protein family awaits the results of gene isolation and sequencing.

**Acknowledgements**

# References

Argos, P., S. V. L. Narayana and N. C. Nielsen: EMBO J. **2**, 1111-1117 (1985).

Bollini, R. and M. J. Chrispeels: Planta **142**, 291-298 (1978).

Brune, D. C.: Anal. Biochem. **207**, 143-146 (1992).

Derbyshire, E., D. J. Wright and D. Boulter: Phytochemistry **15**, 3-24 (1976).

Derbyshire, E. and D. Boulter: Phytochemistry **15**, 411-414 (1976).

Dudman, W. F. and A. Millerd: Biochem. Systematics Ecol. **3**, 25-29 (1975).

Duranti, M., N. Guerrieri, P. Cerletti and G. Vechio: Eur. J. Biochem. **206**, 230-232 (1992).

Häger, K.P., H. Braun, A. Czihal, B. Müller and H. Bäumlein: J. Mol. Evol. **41**, 457-466 (1995).

Hill, J. E. and R. W. Breidenbach: Plant Physiol. **53**, 742-746 (1974).

Hunkapiller, M. W. and L. E. Hood: Meth. Enzymol. **91**,486-493 (1983).

Jensen, U. and H. Berthold: Phytochemistry **28**, 143-146 (1989).

Kloz, J. and V. Turkova: Biol. Plant. **5**, 29- 36 (1963).

Laemmli, U. K.: Nature **227**, 680-685 (1970).

Lawrence, M.C., T. Izard, M. Beuchat, R.J. Blagrove and P.M. Colman: J. Mol. Biol. **238**, 748-776 (1994).

Luthe, D. S.: Plant Mol. Biol. Reporter **10**, 254-262 (1992).

Matta, N.K., J.A. Gatehouse and D. Boulter: J. Exp. Bot. **32**, 1295-1307 (1981).

McLeester, R. C., T. C. Hall, S. M. Sun and F. A. Bliss: Phytochemistry **12**, 85 - 93 (1973).

Miège, M-N: Protein types and distribution. In: Nucleic Acids and Proteins in Plants I. ( D.Boulter and B.Parthier, eds.), pp.29-345. Springer-Verlag, Berlin, Heidelberg, New York (1982).

Nielsen, N. C.: Phil. Trans. R. Soc. Lond. **B 304**, 287-296 (1984).

Pusztai, A. and W.B. Watt: Biochim. Biophys. Acta. **207**, 413-431 (1970).

Osborne, T. B. and G. F. Campbell: J. Am. Chem. Soc. **20**, 348-375, 393-405 (1898).

Saio, K., M. Kamiya and T. Watanabe: Agric.Biol.Chem. **30**, 1301-1308 (1969).

Shewry, P. R.: Biol. Rev. **70**, 375-426 (1995).

Shutov, A. D., I. A. Kakhovskaya, H. Braun, H. Bäumlein and K. Müntz: J. Mol. Evol. **41**, 1057-1069 (1995).

Staswick, P. E., M. A. Hermodson and N. C. Nielsen: J. Biol. Chem. **259**, 13431-13435 (1985).

Sun, S. M. and T. C. Hall: J. Agric. Food Chem. **23**, 184 - 189 (1975).

Wind, C. and K-P. Häger: FEBS Lett. **383**, 46-50 (1996)

Wong, S. W. and T. S.-F. Wang: J. Biol Chem. **264**, 5924-5928 (1989).

Figures

Figure 1: Two dimensional-SDS-PAGE analysis of the high molecular weight protein fraction from the sephacryl-S300 gel filtration - the horizontal dimension was unreduced and the vertical dimension after reduction with dithiothreitol. The arrows indicate the positions of the legumin α- and β-subunits. The upper arrow indicates the major α-subunit (54kDa) providing the N-terminal sequence shown in fig2A. The β-subunits (21kDa) are indicated by the single arrow head. The gel was silver stained. The intense diagonal band running across the gel represents the non-disulphide linked polypeptides.

Figure 2: Alignment of the N-terminal sequences of the *Phaseolus* legumin subunits with other legume legumins - A - α-subunits and B - β-subunits. Bold letters indicate highly conserved residues, * = positions showing strictly conserved residues; + = positions showing conservative residue replacement. The protein sequences were retrieved from the SWISSPROT sequence library (see text for the sequence accession numbers).

A)

| Species | α-subunits |
|---|---|
| *Phaseolus vulgaris* | TTNLFNRCRINELNALKPDH |
| *Glycine max* Glycinin 1 | FSSREQPQQNECQIQKLNALKPDN |
| *Glycine max* Glycinin 2 | LREQAQQNECQIQKLNALKPDN |
| *Glycine max* Glycinin 3 | FREQPQQNECQIQRLNALKPDN |
| *Glycine max* Glycinin 4 | ISSSKLNECQLNNLNALEPDH |
| *Glycine max* Glycinin 5 | ITSSKFNECQLNNLNALEPDH |
| *Pisum sativum* LegA2 | LREQPQQNECQLERLNALEPDN |
| *Pisum sativum* Leg J | TSSEFDRLNQCQLDSINALEPDH |
| *Vicia faba* Leg 4 | TSSEFDRLNQCRLDNINALEPDH |
| | *  *  +    +***  ** |

B)

| Species | β-subunits |
|---|---|
| *Phaseolus vulgaris* | GIEETLCTLKLQHNIARASS |
| *Glycine max* Glycinin 1 | GIDETICTMRLRHNIGQTSS |
| *Glycine max* Glycinin 2 | GIDETICTMRLRQNIGQNSS |
| *Glycine max* Glycinin 3 | GIDETICTMRLRHNIGQTSS |
| *Glycine max* Glycinin 4 | GVEENICTLKLHENIARPSR |
| *Glycine max* Glycinin 5 | GVEENICTMKLHENIARPSR |
| *Pisum sativum* LegA2 | GLEETVCTAKLRLNIGPSSS |
| *Pisum sativum* Leg J | GLEETICSAKIRENIADAAR |
| *Vicia faba* Leg 4 | GLEETICSLKIRENIAQPAR |
| | *+_+*  +*+_+++   ** |

first dimension

second dimension

66
45
36
29
24
20.1
14.2