5. The Biochemistry and Cell Biology of Embryo Storage Proteins

NIELS C. NIELSEN, RONALD BASSÜNER and TODD BEAMAN
USDA, Agricultural Research Service, and the Departments of Agronomy and Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA

ABSTRACT. Seed legumes accumulate massive amounts of certain characteristic proteins. These include the 7S and 11S seed storage proteins, lectins and various 2S proteins. With the advent of molecular cloning techniques, a large and increasing number of primary structures of these proteins from many plants are available in the international databases. This review summarizes and compares sequences of the seed proteins presently available. To the extent that they are known, the genes which encode the proteins and mechanisms involved in their expression are summarized. Strategies involved in the assembly of the 7S and 11S protein oligomers are discussed, and data that relate existing storage protein to their evolutionary progenitors are described.

1. Introduction

Seeds synthesize and store large amounts of certain proteins for use as a source of carbon, nitrogen, and sulfur during seedling growth and development. Most of these proteins accumulate in subcellular vacuoles located in either the seed endosperm or in cotyledons of the embryo and then are degraded upon germination. The vacuoles into which the proteins are deposited are referred to either as protein storage vacuoles or protein bodies. This chapter will focus on the proteins stored in vacuoles in seed cotyledons, while chapters that follow this one concern themselves with characteristics of those proteins found in the seed endosperm. The chapter is not intended to be an exhaustive discussion about the proteins found in seed cotyledons, because this topic has been the subject of a number of earlier reviews (Derbyshire et al., 1976; Higgins, 1984; Casey et al., 1986, 1993; Rerie et al., 1992; Vitale and Bollini, 1995). Rather, it is intended as an overview about the properties of these proteins and a discussion of the biological events involved in their synthesis and deposition during embryogenesis.

Certain generalizations can be made about seed storage proteins. They usually consist of one or more families of structurally related proteins. The two that are most widely distributed in cotyledons of both angiosperms and

B.A. Larkins and I.K. Vasil (eds.), Cellular and Molecular Biology of Plant Seed Development, 151–220.
© 1997 *Kluwer Academic Publishers.*

gymnosperms are oligomers with sedimentation coefficients of 7–8S and 11–12S. Although trivial names are frequently used to identify proteins from specific plant species (Tables 1–3), the terms vicilin-like and legumin-like, respectively, are commonly used. Because the two classes of proteins account for such a large proportion of total protein in seeds, they exert a profound influence on the rheological and functional properties of foods made from seeds. In soybeans, for example, the 7S plus 11S proteins can account for in excess of 70 percent of the total protein and close to 50 percent of the seed mass. The 7S and 11S storage proteins in cotyledons of dicot plants are devoid of enzymatic activity and are soluble in dilute salt solutions. Because of their solubility properties, the proteins are considered globulins in accordance with the nomenclature developed by Osborne (1924). Indeed, the terms storage globulins and storage proteins are used interchangeably. The 7S and 11S seed storage globulins typically contain a high proportion of nitrogen due to elevated amounts of amino acids containing amide-nitrogen. They also have low amounts of sulfur-containing amino acids, and it is this property that has long attracted the attention of scientists interested in improvement of seed nutritional quality.

A number of albumins can be found in seeds of diverse species, and though present at significant levels, they have long been considered to be mainly metabolic proteins (Danielsson, 1949; Boulter and Derbyshire, 1971; Millerd, 1975; Ashton, 1976). Characterization of proteins from rape seed (Lõnnerdahl and Janson, 1972), castor bean (Youle and Huang, 1978a,b) and cottonseed (Youle and Huang, 1979) revealed a number of 2S albumins with amino acid compositions similar to the storage globulins and which were degraded during germination. As discussed in more detail later, similar proteins are widely distributed among plant species. Therefore, the 2S albumins compose a third distinct class of seed storage proteins and are found in both angiosperms and gymnosperms. Frequently, the 2S storage proteins contain elevated amounts of sulfur amino acids compared to storage globulins.

Other seed proteins with biological activities accumulate in seeds in amounts greater than most enzymes involved in cellular metabolic processes. Although not consistently found in each species of every division, they are of considerable agronomic interest because of their influence on the properties of foods derived from seeds. For example, some seeds accumulate certain protease inhibitors and lectins in amounts of 1 to 5% of the total protein. These proteins exert antinutritional effects on animals or insects that consume either seeds or food products from seeds. Seeds also contain lipoxygenases responsible for the development of undesirable flavors and aromas in foods. While the biological roles of such proteins in the seed are often not known with certainty, protection against predators is frequently a function ascribed to them. In addition, however, these proteins seem to have evolved to play a storage role in the seed; not because of their composition, but because of their abundance. Brief descriptions of seed lipoxygenases and lectins will be found

in this chapter, while the protease inhibitors are discussed elsewhere in this book.

2. Structure and molecular heterogeneity of seed storage proteins

Most early descriptions of seed globulins were derived from studies on agronomically important legumes, particularly soybean (*Glycine max*), pea (*Pisum sativum*), common bean (*Phaseolus vulgaris*), and the faba bean (*Vicia faba*). The studies revealed that the 7S and 11S proteins were oligomers composed of homologous, yet non-identical subunits. For example, at least five different subunits of the 11S protein are found in soybean (Nielsen et al., 1989), and there can be as many as 15 members of the 7S family of soybean proteins (Harada et al., 1989). The classical methods for purification of these proteins, together with a description about the physical characteristics of the isolated proteins are reported in important reviews by Derbyshire et al. (1976) and Casey et al. (1986) and will not be repeated here.

With the advent of molecular techniques, genes encoding the seed storage globulins were among the first from plants to be cloned, and an extensive library of amino acid sequence information now exists for both 7S and 11S proteins. As data accumulated, it became clear that the 11S and 7S proteins are not confined to the dicots but are found in monocots such as maize, oats, wheat, and rice. They have also been found in at least four divisions of gymnosperms. Citations with information about this diverse family of related proteins are located in Tables 1, 2, and 3. The availability of amino acid sequence data for these various proteins has permitted their comparison and revealed a remarkable level of structural conservation among examples from widely divergent plant species. The important concepts to emerge from these comparisons are that the 11S and 7S storage globulins are related (Argos et al., 1985; Gibbs et al., 1989; Lawrence, 1994; Shutov et al., 1995), and the families of genes that encode the proteins evolved from a common ancestral gene that predates the speciation of angiosperms, and perhaps gymnosperms (Borroto and Dure, 1987; Jensen and Berthold, 1989; Arahira and Fukazawa, 1994; Häger et al., 1995). Indeed, it will be interesting to determine if the ancestral gene responsible for these two groups of proteins predated the appearance of multicellular organisms.

Despite the homology of the 7S and 11S proteins among plant species, substantial variability can be observed in the relative amounts of the two globulins in seeds. Although many dicot plants contain sizable amounts of both proteins, a few contain predominantly either vicilin-like or legumin-like globulins. For example, seeds from *P. vulgaris* (Ergland et al., 1983) and winged bean (*Psophocarpus tetragonolobus*) contain predominantly 7S globulins. At the other extreme, seeds from members of the Brassicaceae, such as rapeseed (*Brassica napus*), radish (*Raphanus sativus*), and thale cress

(*Arabidopsis thaliana*), contain mainly 11S globulins (Crouch and Sussex, 1981; Laroche et al., 1984; Heath et al., 1986). The 11S globulins also account for more than 80% of the total seed protein in the *Cucurbita* species (Hara et al., 1976) and in sunflower (*Helianthus annuus*) (Dalgarrondo et al., 1984). Although many non-legume species lack 7S proteins, a 2S storage albumin frequently is accumulated instead. The lack of 7S proteins in other dicots besides legumes is not a general rule, however, because some have retained their capacity to accumulate this protein. For example, cotton (*Gossypium hirsutum*) produces a vicilin-like protein in addition to legumin-like proteins (Dure and Chlan, 1981), as does tobacco (*Nicotiana tabacum*) (Sano and Kawashima, 1983). Interestingly, alfalfa (*Medicago sativa*), a legume, contains both the 11S and 7S globulins and a 2S albumin (Coulter and Bewley, 1990; Krochko and Bewley, 1990). These differences in protein composition are not confined to interspecies variation but can also be found within species. For example, Medeiros (1982) reported that the ratio of 11S to 7S proteins varied between about 1:1 and 2.9:1 among 32 progeny lines from a cross between soybean parents with high and low protein. Mutation breeding programs have also been used to successfully eliminate the 7S storage globulins from *P. vulgaris* (Burrow et al., 1993). The amount of variation that can be observed within and among species suggests that considerable opportunity exists to manipulate the content of seed storage proteins.

3. Seed storage proteins without biological activities

A. Vicilin-like proteins in legumes. As indicated in Table 1, 7S proteins are widely distributed in the plant kingdom. However, because 7S globulins from *P. vulgaris*, *V. faba*, *G. max* and *P. sativum* are the best characterized, this discussion will concentrate on a description of their features and assume that homologous proteins in other less completely characterized 7S proteins exhibit similar features. Unfortunately, and despite the similarities among storage globulins from different species, a rather confusing nomenclature of trivial names has evolved to identify proteins from different species. Many of the names originated from the work of Osborne (1924) and are now well entrenched in the literature. To the extent possible, the trivial names associated with each protein are given in Table 1.

The 7S globulins are usually extracted from legume seeds as oligomers whose apparent molecular weights range between 140,000 and 210,000 daltons (Casey et al., 1986). The large differences in molecular mass reflect heterogeneity in the size of the subunits that comprise each trimer. As will be described shortly, many, but not all, dicots contain two general groups of 7S subunits, one around m_r 45–55,000 and the other m_r 70–80,000. Members from both groups share homology, but the members of the m_r 70–80,000 group are distinguished by the presence of a large insert near the N-terminus

TABLE 1

Seed 7S storage globulins

Species	Gene name, gene product name, group classification	Subunit molecular weight	GenBank/EMBL/DOBJ Database Accession Number
Arabidopsis thaliana	Vicilin	Fragment	Z46695
Arrachis hypogaea	Convicilin-type		
(Peanut)	Aliergen, ARA H I, AHI1	70kDa	L38853
	AH12	71kDa	L34402
Canavelis ensiformis (Jack bean)	Canavalin	50kDa	X59467, S76871
Canavalia gladiata	Canavalin	50kDa	X06733, X15076
(Sword bean)	Pseudogene		X52377
Glycine max	β-conglycinin		
(Soybean)	convicilin-type		
	α subunit	66kDa	X17698
	α' subunit	72kDa	M13759
	vicilin-type		
	β subunit	46kDa	S44893
Gossyplum hirsutum	Convicilin type		
(Cotton)	α-globulin A	71 kDa	M19878
	α-globulin B	69kDa	M16891
Lens culinaris	Vicilin A		Z48436
(Lentil)	Vicilin C		Z48434
	Vicilin D (pseudogene)		Z48435
	Vicilin E		Z48440
	Vicilin		Z48437
Lens ervoldes	Vicilin		Z48438
Pheseolus vulgaris	α-Phaseolin	49kDa	X52626
(Bean)	β-Phaseolin	48kDa	J01263, M13758
	Phaseolin	48kDa	U01131
Picea gleuce (White spruce)	Vicilin-type	50kDa	X63191

TABLE 1

(Continued)

Species	Gene name, gene product name, group classification	Subunit molecular weight	GenBank/EMBL/DOBJ Database Accession Number
Pisum sativum	Vicilin		
(Pea)	vcl	50kDa	Y00722
	vcl1	14kDa fragment	P02856*
	vclA	31.5 kDa fragment	P02855*
	vclB	46.3 kDa fragment	P02854*
	vcl	52.2 kDa	X14076
	vicJ	47 kDa	X67428
	Vicilin k	47 kDa	X67429
	Convicilin		
	cvcA	67 kdA	X06398
	cvcB	Fragment	M73805
Theobroma cacao	Vicilin-type	43 kDa	X62625, X62626
(Cocoa)	Convicilin-type	67 kDa	A20606
Triticum aestivum	Convicilin-type	71 kDa	M81719
Vicia faba	Vicilin	51 kDa	Y00462
(Broad bean)			Y00506
Zamia furfuracea	Vicilin	46 kDa	Z50791
Zea mays	Vicilin-type (Gib-2)	49 kDa	Wallace and Kriz, 1991;
	Convicilin-type (Gib-1)	65 kDa	M24845, X59083, X59084
	null alleie		U28017

Latin names are given followed by common names in brackets. Accession numbers can be found in the GenBank database, which can be obtained on the internet at http://ftp.bio.indiana.edu. The * (asterisk) denotes the SwissProt database accession number.

as compared to the smaller groups of proteins (Schuler et al., 1983; Casey et al., 1985a; Coates et al., 1985). Both groups of subunits can be subdivided into subfamilies, and the proteins are often glycosylated. Because some of the subunits may be trimmed or nicked by proteases and can be glycosylated to varying degrees, considerable heterogeneity is frequently observed in electrophoretic patterns of 7S subunits of seed extracts. For example, phaseolin (Hall et al., 1977; Bollini and Chrispeels, 1978), the 7S globulin from *P.*

TABLE 2

Seed 11S storage globuline

Species	Gene name, gene product name, group classification	Subunit molecular weight	GenBank/EMBL/DDBJ Accession Number
Arabidopsis thaliana	Cruciferin		
(Thale cross)	CRA1	52 kDa	X14312
	CRB	50 kDa	X14313
Avena sativa	12S globulin	57 kDa	J05485
(Oat)	GLAV 1 gene	61 kDa	X74740
	GLAV-3 gene	63 kDa	X74741
Brassica napus	Cruciferin		
(Rape)	Cru1	58 kDa	X62120
	Cru2/3	62 kDa	X14555
	Cru4	52 kDa	X57850, X57851
	BnC1	54 kDa	X59294
	5′ flanking region		M90109
	BnC2	55 kDa	X59259
Cucurbita pepo (Pumpkin)	Curcurbitin	55 kDa	M36407
Ephedra gerardiana (Ephedra)	Legumin	58 kDa	Z50777
Gingko biloba (Gingko)	Ginnacin	51 kDa	X75426, Z50778
Glycine max	Glycinin		
(Soybean)	Group 1		
	Gyl ($A_{1a}B_2$)	58 kDa	X02985, X15121, X02985
	Gy2 (A_2B_{1b})	58 kDa	K02646, D00216, X15122 Y00398, X02805
	Gy3 ($A_{1b}B_{1b}$)	58 kDa	X53404, X15123
	Group 2		
	Gy4 ($A_5A_4B_3$)	64 kDa	X52863, X02626
	null allele		X05651
	Gy5 (A_3B_4)	62 kDa	X79467, M10962, M35671
Glycine soya	Glycinin		
(Soybean)	Group 2		
	Gy4 ($A_5A_4B_3$)	63 kDa	X86970
Gnetum gnemon	Legumin	67 kDa	Z50779
Gossypium hirsutum	β-globulin A	58 kDa	M16905, M69188
(Cotton)	β-globulin B	59 kDa	M16936, M19389
Helianthus annuus	Helianthinin	54 kDa	M28832
(Sunflower)	5′ flanking region		X53664, X53665

TABLE 2
(Continued)

Species	Gene name, gene product name, group classification	Subunit molecular weight	GenBank/EMBL/DDBJ Accession Number
Magnolia salicifolia	LegA2	52 kDa	X82464
(Magnolia)	LegA11	52 kDa	X82463
	LegB14	52 kDa	X82465
Oryze sativa	Glutelin	56 kDa	A91367*, B91367*
(Rice)	Gt1	57 kDa	M28156
	Gt2	57 kDa	L36819
	Gt3		M28158
	Gt22	60 kDa	M28159
	GluB-1	59 kDa	X54314
	GluB-2	57 kDa	X54192
	GluB-3		X54193
	Type I	55 kDa	X05661
	Type II	54 kDa	X05664
Picea glauca (White spruce)	Coniferin	56 kDa	X63192, U19873
Pinus strobus	Globulin 1	53.8 kDa	Z11486
(Eastern white pine)	Globulin 2	45.2 kDa	Z11487
Pisum sativum	Legumin		
(Pea)	Group 1		
	LegA1	α-chain	X02802
		Fragment	J01255
	LegA2	58 kDa	X17193
	5′ flanking region		X57665
	5′ flanking region		X57666
	Group 2		
	LegC	56 kDa	X02984
	LegC 5′ region with Pis-1		X07562
	LegD (pseudogene)		X02614
	LegJ	55 kDa	X07014
	LegK		X67422, X67423, X07015
	HMW legumin	67 kDa	M16890
Pseudotsuga menziesii (Douglas fir)	11 S globulin	56 kDa	L07484
Raphanus sativus (Radish)	Cruciferin, class II	53 kDa	X59808

TABLE 2

(Continued)

Species	Gene name, gene product name, group classification	Subunit molecular weight	GenBank/EMBL/DDBJ Accession Number
Vicia faba	Legumin		
(Broad bean)	Group 1 (A-type)		
	LeA1	55 kDa	X55013
	LeA2	53 kDa	X55014
	Group 2 (B-type)		
	LeB1 (pseudogene)	α-chain	X14238
	LeB2	α-chain	X14237
	LeB3 (HMW)	65 kDa	Z25489
	LeB4	53 kDa	X03677
	5′ flanking region		X14239
	LeB5 (pseudogene)		X14239
	LeB6	Fragment	X14240
	LeB7	Fragment	X14241
	LelB161 (HMW)	Fragment	Z26487
Vicia narbonensis	Group 1		
	LegA	55 kDa	Z46803
Vicia sativa	Legumin		
(Vetch)	Group 1		
	LegA	55 kDa	Z32835
	Group 2		
	LegB	53 kDa	Z32796

Latin names are given followed by common names in brackets. Most accession numbers can be found in the GenBank database, which can be obtained on the internet at http://ftp.bio.indiana.edu. An asterisk (*) denotes the PIR database which can be searched at http://ftp.bio.indiana.edu. HMW = high molecular weight legumin. For classification into group 1 (A-type) and group 2(B-type) legumins where applicable, see text.

vulgaris, contains only the m_r 45–55,000 group of subunits. At least three subfamilies are found in this group, each with slightly different molecular characteristics (α, m_r51–53 $\times$ 10^3; β, m_r47–50$\times$$10^3$; ($\gamma$, m_r43–47 $\times$ 10^3). Because all of these subunits can potentially have either one, two, or no Asn-linked high mannose glycans, a complex pattern can be observed when they are optimally separated by two-dimensional electrophoresis.

Unlike the situation with phaseolin where only the m_r 50,000 subunits are found, both the large- and small-sized 7S subunit families are found in *P. sativum*, *V. faba*, *Vigna unguiculata*, *Canavalia ensiformis*, *Vigna radiata*, and *G. max* (Khan et al., 1980; Gatehouse et al., 1981; Scholz et al., 1983; Spencer

TABLE 3

Seed storage lectins

A. One-chain lectins

Species	Gene name, gene product, group classification	Molecular Weight Holoprotein or Subunit	SwissProt Database Accession Number
Arachis hypogaea (Peanut)	Peanut agglutin (PNA)	120 kDa Homotetramer	P02872
Bauhinia purpurea (Camel's foot tree)	Lectin	27 kDa Subunit	P16030
Canavalia ensiformis (Jack bean)	Concanavalin A	102 kDa Homotetramer	P02866
Canavalia gladiata (Sword bean)	Concanavalin A	126 kDa Homotetramer	P14894
Crotaliara striata (Smooth rattlebox)	Lectin	116 kDa Homotetramer	P16351
Cytisus scoparius (Scotch broom)	Lec2	27 kDa Subunit	P29257
Cytisus sessilifolius	Lec1	154 kDa Homotetramer	P22970
	Lec2	75 kDa Homodimer	P22971
Dolichos biflorus (Horse gram)	Seed lectin	30 kDa Subunit	P05045
Erythrina corallodendron (Coral tree)	Lectin	62 kDa Homodimer	P16404
Glycine max (Soybean)	Soybean lectin Le1 (Soybean agglutin)	31 kDa Subunit	P05046
Laburnum alpinum (Scotch laburnum)	Lec1	27 kDa Subunit	P23558
Lathyrus sphaericus (Spring vetchling)	Seed lectin	54 kDa Homodimer	P16349
Medicago truncatula (Barrel medic)	Lec1	31 kDa Subunit	Q01806
	Lec2	30 kDa Subunit	Q01807
Onobrychis vicifolia (Common sainfoin)	Lectin	52 kDa Homodimer	P02874
Phaseolus lunatus (Lima bean)	Lima bean lectin	96 kDa Homodimer	P16300
Phaseolus vulgaris (Bean)	Phytohaemagglutinin (PHA),		
	G2	30 kDa Subunit	P05086
	PHA-E	30 kDa Subunit	P05087, P15231
	PHA-L	29 kDa Subunit	P19329
	ARC1 (arcelin)	29 kDa Subunit	P199330
	ARC2 (arcelin)	27 kDa Subunit	P02873
	LLP (α-amylase inhibitor)		
Ulex europeus (Furze)	ANTI-H(O) LEC I (UEA-I)	27 kDa Subunit	P22972
	ANTI-H(O) LEC II (UEA-II)	27 kDa Subunit	P22973

TABLE 3

(Continued)

B. Two-chain lectins

Species	Gene name, gene product, group classification	Molecular Weight Holoprotein or Subunit	SwissProt Database Accession Number
Dioclea grandiflora (Mucana)	LecA	30 kDa Subunit	P08902
Lathyrus aphaca (Yellow vetchling)	Lectin	58 kDa ($\alpha_2\beta_2$)	P07441 (α-chain)
Lathyrus articulatus	Lectin	58 kDa ($\alpha_2\beta_2$)	P07442 (α-chain)
Lathyrus cicera (Flat-pod pea)	Lectin (α1α2)	59 kDa (α2β2)	P07440 (α-chain)
Lathyrus hirsutus (Rough pea, Hairy vetchling)	Lectin	59 kDa (α2β2)	P07443 (α-chain)
Lathyrus ochrus (Yellow flowered pea)	Lectin α1 α2 β1 and β2	59 kDa (α2β2)	P12306 P12307 P04122
Lathyrus odoratus (Sweet pea)	Lectin	58 kDa (α2β2)	P02869 (α-chain)
Lathyrus sativus (Chickling vetch)	LECA	58 kDa (α2β2)	P12308 (α-chain)
Lathyrus tingitanus (Tangier pea)	Lectin	60 kDa (α2β2)	P0744 (α-chain)
Lens culinaris (Lentil)	Lectin	46 kDa (α2β2)	P02870
Lotus tetraglobus (Winged pea)	Anti-H(O) Lectin (LTA)		P19664
Pisum sativum (Garden pea)	Pea lectin	48 kDa (α2β2)	P02867
Vicia cracca (Bird vetch)	LECA	55 kDa (α2β2)	P02868 (α-chain)
Vicia faba (Faba bean, Broad bean)	Favin (VFA)	50 kDa (α2β2)	P02871
Vicia sativa (Spring vetch, Tare)	LECA	55 kDa (α2β2)	P16350 (α-chain)

The common names for the plant source is in brackets preceded by the Latin name. The entries match the PROSITE criteria AC# PS00307 and PS00308 for the signature pattern (Sharon and Lis, 1989, 1990). The accession numbers can be found in the SwissProt database. The databases can be accessed on the internet at http://ftp.bio.indiana.edu.

et al., 1983; Casey et al., 1984; Sammour et al., 1984). In *P. sativum, V. faba*, and several other legumes, the smaller m_r 45–55,000 group of subunits are known as vicilins, while the larger m_r 70–80,000 group are called convicilins. In contrast to this situation where different names are used to distinguish the large and small groups of 7S subunits, the same name, β-conglycinin, is used to identify members of both groups of 7S subunits in *G. max*. In this species, the β-subunits correspond to the vicilins, while the larger α and α'-subunits are equivalent to convicilins. Although confusing, the term glycinin refers to 11S proteins in soybean rather than to 7S proteins, as one might expect based on the use of vicilin versus convicilin for many other legumes. Usage of the term *glycinin* to refer to 11S proteins in soybean and *conglycinin* for 7S proteins is sufficiently entrenched in the literature to preclude renaming these proteins.

Considerable effort has gone into characterizing the various members of the 7S families of genes in legumes, and some of their structural features are worth noting. At least 18 genes encode vicilin subunits in pea (Higgins, 1984; Casey et al., 1986), and similar small families of genes are responsible for the m_r 50,000 class of subunits in *V. faba* (Weschke et al., 1988) and *G. max* (Coates et al., 1985). As is important from a nutritional standpoint, many of these subunits are devoid of sulfur amino acids. The 18 vicilin genes in pea are encoded by three small multigene subfamilies. One family encodes m_r 47–48,000 subunits, while the other two encode m_r 50–52,000 globulins (Casey et al., 1985a; Rerie et al., 1992). Whereas some vicilin subunits remain intact after synthesis, others contain two potential internal proteolytic cleavage sites. None, one, or both of these sites can be cleaved to yield up to five peptide fragments (Hirano et al., 1982; Gatehouse et al., 1983; Spencer et al., 1983; Boulter, 1984).

The m_r 45–55,000 group of vicilin subunits from pea can be further subdivided into three families (Casey et al., 1988). Vicilins encoded by members of the same subfamily are less than 5% divergent, whereas 15–25% divergence in nucleotide sequence occurs among members of different subfamilies (Rerie et al., 1992). Although pseudogenes exist, those genes that are transcriptionally active appear equally transcribed as judged by the prevalences of their respective cDNAs (Domoney and Casey, 1985; Ellis et al., 1986) and their relative transcription rates (Beach et al., 1985). To the extent such information is available, a similar situation apparently exists among members in the subunit subfamilies from *P. vulgaris* (Brown et al., 1981a,b), *V. faba* (Wobus et al., 1986), and *G. max* (Harada et al., 1989).

Comparison of the deduced amino acid sequences for members of the m_r 70–80,000 group with protein sequence information determined chemically shows that this group of subunits differs from the m_r 50,000 subunits because of the N-terminal insertion mentioned earlier. The insertion, which is documented in references cited in Table 1, is located shortly after the signal sequence in the precursor. The size of the insertion is about 165 amino

acids in the case of the α-subunit of soybean β-conglycinin as compared with the β-subunit. An insertion of about 170 amino acids in the same region of the molecule has occurred in pea convicilin as compared to vicilin, and an even larger insert is found in the α'-subunit of β-conglycinin. Thus, the 7S molecules apparently tolerate considerable variation in this region.

In some species, the peptide chain that is a result of this insertion is subjected to post-translational modification. In the case of the α and α'-subunits of β-conglycinin of soybean and a 7S protein in maize (Coates et al., 1985; Doyle et al., 1986), a discrepancy exists between the N-terminal sequences of the purified protein and the deduced amino acid sequence from cDNAs. The chemically determined N-terminal amino acid sequence of the α and α' subunits begins about 40 amino acids after the putative N-terminal of the proprotein, an indication that a short 40 amino acid peptide chain is separated from the rest of the molecule post-translationally. Unlike the β-conglycinin in soybean, the m_r 70,000 convicilin subunits from pea appear not to be modified extensively post-translationally (Domoney and Casey, 1983), although Chrispeels et al. (1982b) reported an example where a pea convicilin subunit is processed from m_r 70,000 to about 50,000.

Because some 7S subunits are modified post-translationally and others are not, one must question whether these proteolytic modifications are of functional significance for maintenance of oligomeric structure. The variability in the location of the peptide bonds cut indicates that they probably are the consequence of non-lethal mutations to individual genes and result in subunits susceptible to endopeptidases that co-occupy the protein bodies with the storage globulins. The peptide fragments that result from cleavage at these sites could be stabilized in the oligomers by non-covalent protein interactions and become evident when the purified oligomers are denatured and the peptides resolved by electrophoresis. Alternately, they could be removed and digested immediately. This point has not been resolved. In addition to digestion by endopeptidases, the storage globulins can be attacked by exopeptidases. 'Ragged' amino- and carboxyterminal ends of individual purified storage proteins have been described that are consistent with proteins digested by amino- and carboxypeptidases (Higgins et al., 1983a; Slightom et al., 1983; Staswick et al., 1984a; Casey et al., 1985a). Thus, proteolytic changes such as those described can complicate identification of individual proteins, and this can become particularly acute when serological methods are used to identify peptides separated by electrophoresis.

Interestingly, far fewer genes encoding the larger molecular weight group of subunits are contained in legume genomes than those that encode the smaller m_r 45–55,000 group of globulins. For instance, only two bands of genomic DNA that hybridize with a convicilin cDNA are detected in *P. sativum* (Domoney and Casey, 1985), whereas it is estimated there are at least 18 of the smaller vicilin genes (Rerie et al., 1992). Likewise, Harada et al. (1989) found only a few genes that produced α and α' subunits of

β-conglycinin among the 15 genes from the *G. max* genome that encode 7S globulin subunits. The situation in *P. vulgaris* is even more severe, as genes encoding this group of subunits have not been detected (Hall et al., 1977). The significance, if any, for the difference in prevalence of the large and small groups of 7S globulin subunit genes is unknown.

Variation in the structure of various 7S globulin subunits as determined by gel electrophoresis has permitted establishment of genetic linkage relationships among the genes that encode them. Thompson and co-workers described variations in the apparent molecular masses of vicilin subunits in pea (Thompson and Schroeder, 1978; Thompson et al., 1980). These results and others (Mahmoud and Gatehouse, 1984; Domoney and Casey, 1985; Ellis et al., 1986) reveal that the pea vicilin genes are distributed among at least five genetic regions, each of which contain several tightly linked, closely related gene sequences. The genetically linked genes encoding structural variants located in each of these regions behave as single genes and produce simple, codominant Mendelian inheritance patterns. One of the regions is linked within about 10 map units of the *r*-locus of chromosome-2 in *Pisum* (Davies, 1980). A second region is also located on chromosome-2 but segregates independently from the *r*-locus. Structural variants of convicilin likewise exhibit simple, codominant Mendelian inheritance. At least one convicilin locus is located on linkage group 2 near the *k*-locus and segregates independently from genes encoding vicilin subunits.

A similar complex gene organization is found in *G. max*. Harada et al. (1989) identified at least 15 β-conglycinin genes that produce either 1.7 or 2.5 kb mRNAs. These were considered to encode β and α plus α' subunits, respectively. Chromosome walking experiments revealed that almost all of these genes were located in three large multigene clusters. Each cluster contained a mixture of individual and tandemly linked β-conglycinin genes that were in turn linked to functional genes whose products were not expressed in seeds. In contrast to the situation in *Pisum*, where the m_r 70,000 convicilin genes are separated from the m_r 50,000 vicilin genes, several genes in soybean encoding 1.7 and 2.5 kb transcripts were tandemly linked. The 1.7 and 2.5 kb transcripts correspond to β and the α and α' subunits, respectively. In any event, as with the 7S globulin subunits from *Pisum*, a number of size and charge variants of soybean β-conglycinins have been identified (Kitamura et al., 1984; Davies et al., 1985), and they likewise exhibit simple, codominant Mendelian inheritance patterns.

Genetic experiments reported by Hall and coworkers addressed the organization of 7S protein genes in *P. vulgaris*. Two-dimensional gel electrophoresis analysis of phaseolin revealed that differences in electrophoretic mobility exist among 7S phaseolin subunits (Romero et al., 1975; Hall et al., 1977; Brown et al., 1981a,b). No evidence for recombination among the structural variants has been reported. The interpretation of these results is that the genes encoding phaseolin are tightly linked and segregate as a single codominant

gene locus. Because at least three different classes of protein can be discerned based on cDNA sequence analysis, and there are multiple copies of each type of gene per genome, unequal crossing over and gene homogenization have likely come into play during formation of the 7S clusters in *P. vulgaris*.

The existence of multiple 7S subunits in the seed raises a question as to whether each trimer purified from legume seeds is composed of only one kind of subunit, or if it contains a heterogeneous mixture of several different subunits. This issue was first explored by Thanh and Shibasaki (1976, 1978b), who isolated multiple isomeric forms of β-conglycinin trimers. Trimers composed of each of the possible combinations of the three main subunit types of β-conglycinin (α, α', and β) were purified from seeds, either by these workers or by others (Sykes and Gayler, 1981; Yamauchi et al., 1981). The results are best explained if products from the approximately 15 β-conglycinin genes associate randomly in vivo during formation of oligomers. A similar assembly mechanism has been evoked to account for heterogeneity observed among 7S oligomers from *P. sativum* (Casey et al., 1985a). Interestingly, Thanh and Shibasaki (1978a) also demonstrated that the β-conglycinin trimers could reversibly dissociate into monomers and, under appropriate conditions, reassociate into either trimers or hexamers. Along the same line, Sun et al. (1974) reported the formation of 18S oligomers using phaseolin purified from *P. vulgaris*. The 18S oligomers apparently were formed from 12 subunits (i.e., a tetramer of trimers). Reversible disaggregation and reassembly events of this type are typical for both the 7S and 11S globulins (Wolf and Briggs, 1958). They are intriguing because they may be related to the formation of large organized structures within the protein body. Virtually nothing is known about how the 7S and 11S proteins are organized in situ.

Recently, the three dimensional crystal structures were described for 7S proteins from both *P. vulgaris* (Lawrence et al., 1990, 1994) and *C. ensiformis* (Ko et al., 1993a,b; Ng et al., 1993). Not surprising because of the homologous nature of these two proteins, nearly identical structures were proposed. Briefly, each 7S subunit features a large internal structural repeat. In both repeated structures, approximately seven β-sheets are organized into β-barrels in a 'jelly-roll' motif. The two 'jelly-roll' motifs are tightly appressed via a hydrophobic face and form the central core of the molecule. The central core is flanked on either side by α-helical domains with helix-turn-helix motifs. The helical domains appear to be involved in intra-subunit interactions involved in formation of 7S trimers, and their disruption results in subunits unable to assemble into trimers (Ceriotti et al., 1991, 1995). Availability of this detailed structural information, together with techniques that can be used to perform site-directed mutagenesis, will undoubtedly result in additional experiments to probe the function of regions in the molecules that are important for assembly of trimers and for assembly of trimers into the supra-molecular complexes found in protein bodies.

Availability of the three dimensional structure has permitted precise alignment of the multiple vicilin-like proteins whose sequences are available (Lawrence et al., 1994), and designation of a sequence signature:

L-X-X-F-X(13)-R-X(7)-P-X(5)-P* -X(3)-D-X(9)-G*

where X refers to any amino acid, while P* and G* identify the two amino acids of the P/G motif defined by Lawrence et al. (1994) that is found twice in both the 7S and 11S globulin subunits. This signature characterizes the vicilin-like molecules sufficiently well so that it can be used to establish relationships among diverse sequences in the existing data bases. Although the phylogenetic origin of the vicilins is not yet evident, a search of the Swiss Protein data base with the program 'Blitz' (Blitz@EBI.AC.UK) reveals a high degree of similarity to a sucrose-binding protein from soybean (Grimes et al., 1992). This similarity is particularly evident in the case of vicilin from the gymnosperms *Zamia* (Braun et al., 1995b) and *Picea* (Newton et al., 1992). Interestingly, an even greater degree of homology is observed between the vicilin-like protein from *Zamia* and *Picea* and 11S legumins from angiosperms, than where similar comparisons are made among those proteins originating only from angiosperms. The observation is consistent with and supports the conclusion by Lawrence and others (Argos et al., 1985; Gibbs et al., 1989; Lawrence et al., 1994; Shutov et al., 1995) that the 7S and 11S proteins are related to a common ancestral gene. In addition to the obvious homology with sucrose-binding protein, the vicilin from *Picea* also has some sequence similarity to germins, a class of proteins synthesized de novo during seed germination which will be discussed later with the legumin-like proteins (Section 3B).

B. Legumin-like proteins. The legumin-like proteins summarized in Table 2 comprise the second major storage globulin component in many seeds. The 11S globulins of the legumes *G. max*, *P. sativum*, *V. faba* have been particularly well characterized. The 11S oligomers are isolated in dilute salt solutions as hexamers with molecular weights of 360,000–400,000 daltons. Each subunit in a hexamer is composed of two polypeptide chains that are linked covalently via a disulfide bond. The position of the interchain disulfide, together with that of a second probably involved in an intrachain disulfide bond in the acidic chain, have been conserved during the evolution of these globulins. As is described in more detail below, both peptides are derived from a single proglobulin precursor by a post-translational proteolytic step. In legumes, the larger of the two peptide chains that result from this cleavage has an acidic isoelectric point, whereas the smaller of the two has a basic isoelectric point. They are frequently referred to as acidic or basic 11S polypeptides, respectively. As is the case of the vicilin trimers, the 11S globulins are the products of small families of related genes, and the subunit products from these

genes apparently are assembled randomly into the storage protein hexamers (Horstmann et al., 1993).

The 11S storage globulins from different legumes display considerable heterogeneity with respect to both charge and size. As indicated in Tables 2 and 4, at least five subunits are present in *G. max*, and these can be divided into two groups based on amino acid sequence homology. The subunits from *G. max* are referred to as glycinins, and the two subfamilies of subunits belong to either Group-1 or Group-2 (Nielsen, 1984). The genes that encode each subunit (Fischer and Goldberg, 1982; Marco et al., 1984; Scallon et al., 1987; Nielsen et al., 1989), as well as the proteins produced from them (Moreira et al., 1981; Staswick et al., 1981; 1984 a,b), have been purified and characterized. Subunits that belong to the same subfamily share greater than 90% homology, but there is only about 50% homology among members from different subfamilies. In soybeans, the Group-1 subunits (G1, G2, and G3) are of a lower molecular weight and contain larger amounts of the sulfur amino acids than those in Group-2 (G4 and G5). The size variation between the two groups of subunits is due to changes in a region at the COOH-terminal of the acidic chain, a region referred to as the hypervariable region. Because of its high concentration of charged amino acids and its proximity to the post-translational cleavage site in the subunit precursor, the hypervariable region must be located at the surface of the proglobulin molecule. The considerable natural variation in the hypervariable region makes it an attractive one into which mutations can be introduced to increase sulfur amino acid content (Nielsen et al., 1990).

As in soybean, two major subfamilies of 11S legumin are found in seeds of both *V. faba* (Bassüner et al., 1983; Horstmann, 1983; Wobus et al., 1984; Bäumlein et al., 1986; Schlesier et al., 1990) and *P. sativum* (Casey et al., 1993), but these are referred to as A- and B-type subunits in these species. They are the structural homologues to the Group-1 and Group-2 subunits of soybean, respectively. Unlike the situation in soybean where both Group-2 subunits are larger than those found in Group-1, some B-type subunits are about the same size as the A-type subunits. Nonetheless, they can easily be assigned to their respective family on the basis of sequence homology and amino acid content (Table 2). Casey (1979), on the basis of two-dimensional gel electrophoresis analysis of the *Pisum* legumins, distinguished among major (a^M) and minor (a^m) forms of legumin acidic polypeptides based on their apparent prevalence after separation.

Many of the acidic and basic chains from 11S proteins have been identified using either various electrophoretic techniques or by direct purification of individual chains. Unfortunately, laboratories working with the same species have developed slightly different nomenclatures to refer to these proteins, and these differences in nomenclature are undoubtedly confusing to the non-specialist who encounters them while reading the literature. A simplification would be desirable, although much of the nomenclature is well entrenched

TABLE 4

Nomenclature identifying primary glycinin and β-conglycinin subunits of soybean

Gene	Subunit	CX635-1-1-1 peptide designation[1]	Clone described
Glycinin			
Gy1	G1	A1aB1b	Nielsen, 1989
Gy2	G2	A2B1a	Marco et al., 1984
			Kim and Choi, 1989
			Kitamura et al., 1990
			Momma et al., 1985a
Gy3	G3	A1bB1a	Cho et al., 1989a
Gy4	G4	A5A4B3	Scallon et al., 1985
			Momma et al., 1985b
Gy5	G5	A3B4	Fukazawa et al., 1985
β-Conglycinin			
Cgy1	α'	α'	Schuler et al., 1982
			Lelievre et al., 1992a
Cgy2	α	α	Sebastiani et al., 1990
			Lelievre et al., 1992a
Cgy3	β	β	Harada et al., 1989
			Lelievre et al., 1992a

[1]Nomenclature refers only to subunits characterized in CX635-1-1-1x. Because subunits from other soybean cultivars may have different primary sequences, it may lead to confusion if this nomenclature is used for cultivars other than CX635-1-1-1. See Moreira et al. (1979), Staswick et al. (1981) and Staswick et al. (1984a, b) for details about CX635-1-1-1.

in the literature. Because the acidic and basic chains are derived from a proglobulin precursor (Croy et al., 1983), specific acidic and basic chains are invariably paired with one another (Staswick et al., 1981; Horstmann, 1983, 1993), and it seems desirable to use nomenclature that relates the two chains when this information is known. Many of the genes that encode the various 11S proglobulins have been cloned and, in a number of instances, genomic Southern blot experiments or genetic inheritance studies have permitted individual genes to be associated with a specific gene locus. Less ambiguity would result if the gene symbols were used in the literature in those cases where the gene has been unambiguously identified. In some cases, for example the one from soybean shown in Table 4, the number of major genes that

encode 11S glycinin is small, and it has been possible to relate the cloned genes to specific loci (Nielsen, 1995). As new gene sequences are cloned and studied, genetic inheritance (segregation) studies should identify which gene is involved and whether allelic versions of the same gene are involved. Such studies should be completed before the variant sequences are identified as new genes. This approach is best suited for genomes like soybean which are devoid of large numbers of tandemly linked genes. In other species, for example *Pisum* and *Vicia*, in which a large number of genes are involved, many of these are tightly linked in small gene clusters. This type of arrangement could obviously complicate genetic segregation studies. Nonetheless, such studies will permit identification of which gene cluster is involved, and chromosome walking experiments may permit deduction of gene order. Despite these difficulties, however, development of simpler nomenclature would enhance communication among nonspecialists. Similar arguments can be advanced for the nomenclature used to describe products from vicilin genes.

As indicated in Table 4, at least five genes contribute proglycinin subunits in soybean, which are denoted Gy1 through Gy5. Each of these have been cloned and sequenced (Fischer and Goldberg, 1982; Hirano et al., 1984; Marco et al., 1984; Fukazawa et al., 1985; Momma et al., 1985a; Scallon et al., 1987; Nielsen et al., 1989). In addition, null-alleles that cause each glycinin subunit to be absent from the seed have been identified (Staswick and Nielsen, 1983; Kitamura et al., 1984; Kitamura, 1993), and RFLP markers that identify each of the five have been identified (Cho et al., 1989a,b; Diers et al., 1994). The latter have been used to explore linkage relations among these genes. Gy4 and Gy5, Group-2 glycinin genes, segregate both independently from one another and from the Group-1 genes. They are located on linkage groups O and F on the public linkage map (Diers et al., 1994). The linkage-map positions of the three Group-1 genes have not yet been established. However, Gy1 and Gy2 were isolated in one clone and shown to be arranged in a direct tandem linkage (Nielsen et al., 1989). These studies showed that the Gy1/Gy2 locus is flanked by several genes that are expressed in the leaf, and that Gy3 is found in a gene cluster that appears to be a duplication of the one containing Gy1 and Gy2. Cho et al. (1989a,b) used RFLP markers to identify the two genetic loci containing Group-1 glycinin genes and demonstrated that they segregated independently.

Similar genetic linkage information is available for *Pisum* and *Vicia*. Hybridization analysis with cloned legumin DNA sequences (Domoney and Casey, 1985; Domoney et al., 1986b), together with sequence information from both genomic clones (Brown et al., 1985; Lycett et al., 1985; Gatehouse et al., 1988; Rerie et al., 1990, 1991; Thompson et al., 1991) and purified legumin polypeptides (March et al., 1988), indicate that there are more than ten legumin genes in the haploid genome of *Pisum*. Domoney and Casey (1984, 1985) have used three classes of cDNAs to distinguish among these genes. Clone pCD43 is representative of genes whose products yield the prevalent

α^M legumin chains. These are derived from m_r 57–60,000 proglobulins and are similar to clones described by other workers (Chandler et al., 1983; Croy et al., 1983). Genes that encode the α^M legumin polypeptides map to linkage group 7, about 10 map units from the *r*-locus (Davies, 1980; Matta and Gatehouse, 1982). Some cultivars of *P. sativum* contain multiple α^M subunits that segregate as a single locus. Because recombinant genotypes have not been detected among the genes in this cluster, they are considered to be tightly linked (Thompson and Schroeder, 1978; Casey, 1979). Although information about the organization of genes within the cluster has become available by chromosome walking experiments (Lycett et al., 1984; Brown et al., 1985; Domoney and Casey, 1985; Casey et al., 1986), it is still incomplete. Nonetheless, those genes that contain the α^M peptide chain belong to the A-type family of legumin genes, and are the structural equivalents to the Group-1 glycinin genes of soybean.

Clones pCD40 and pCD32, used by Domoney and Casey (1984, 1985) to distinguish among the pea legumin genes, select mRNAs for larger legumin precursor subunits (M_r 63–65,000 and 80,000, respectively). Products from these genes contain the minor α^m acidic chains and correspond to B-type legumin genes. These are the structural equivalents to the Group-2 glycinin genes of soybean. There are approximately three copies of genes corresponding to pCD40 per haploid genome (Domoney and Casey, 1985), and these are probably clustered in linkage group 1 near the *a*-locus (Domoney et al., 1986b; Rerie et al., 1992; Casey et al., 1993). The genes recognized by clone pCD32, which is about 80% homologous to pCD40, also appear to map close to the *a*-locus on chromosome 1 and is represented by more than one gene (Casey et al., 1993). Thus, clusters of genes on at least two chromosomes encode prolegumin subunits. Because homologous recombination can occur among genes in clusters, the number of members in such clusters theoretically can be expected to vary among genotypes. Indeed, Turner et al. (1990, 1993) have presented data consistent with this phenomenon for genes encoding the prolegumins.

A technique of two-dimensional electrophoresis in SDS gels, first under non-reducing conditions and then under reducing conditions, has been used to characterize the pairing between large acidic and smaller basic chains of legumin from both *P. sativum* and *V. faba* (Matta et al., 1981a,b). As many as eight pea and 10 pea legumins can be resolved with this technique. As in soybean, the size of basic chains from both species are in the range of m_r 20–25,000, while the size of the acidic chains from both species varies from m_r 35 to 59,000. This technique may be useful in genetic studies directed toward understanding linkage relationships among the legumin genes of pea.

Unlike the situations in *Glycine* and *Pisum,* where the genetic inheritance of the genes has been described in some detail, this is difficult with *Vicia*. High rates of recombination occur in this species, probably due to the extremely large genome size (30pg DNA/2C). The high rates of recombina-

tion that accompany the large genome size have frustrated accurate linkage determinations. Nonetheless, because the genes are located in clusters, *in situ* techniques have permitted localization of some legumin and vicilin genes (Macas et al., 1993; Fuchs et al., 1994; Fuchs and Schubert, 1995). The LeA genes are located in the distal half of the long arm of chromosome V; LeB3 probably resides at the distal end of the small arm of chromosome II, and LeB4 is on the long arm of chromosome III.

Despite homology to the 7S proteins, at the present time the three dimensional structure and shape of the 11–12S storage proteins cannot be described with any degree of certainty. Plietz et al. (1983), on the basis of small angle x-ray scattering analysis of sunflower 12S protein in solution, concluded that these subunits have an ellipsoidal shape with axis of 12.6×8.82 nm and a molecular weight of 365 kD. They proposed that the subunits are arranged in a trigonal antiprism or two trimeric rings superimposed and rotated about 60deg with respect to one another. Utsumi and McPhearson's groups have produced crystals of 12S proteins from soybean and hemp, respectively (Utsumi et al., 1993; Patel et al., 1994). Consequently, there is hope that a more complete physical description of these oligomers may become available in the near term.

Bairoch and Bucher (1994) proposed a sequence signature for the 11S storage proteins. In accordance with the alignments provided by Lawrence et al. (1994), this signature, which is located near the N-terminus of the basic chain, can be expanded somewhat to include:

N-G-X-X-E-X-X-C″-X(6)-N-X(7)-D-X(6)-G-X(10)-P-X-L-X(6)-A-X(12)-
P*-X(4)-N-A-X(8)-G*-X(6)-V-X(3)-G-X(10)-G-X(5)-P-Q

and a signature for a conserved region near the N-terminus of the acidic chain can also be identified:

C′-X(17)-P*-X(13)-G*-X(4)-G-X(3)-P-[GS]-C″

where X corresponds to any amino acid, C′ denotes the cysteine residues involved in the disulfide bond between the acidic and basic chains, C″ denotes a cysteine residue probably involved in an intrachain disulfide bond in the acidic peptide, and P* and G* refer to a conserved P/G motif identified by Lawrence et al. (1994) that is found twice in the legumins and in homologous regions of the 7S globulin subunits. Interestingly, when these signatures are used to search protein data bases, they permit identification of several other groups of proteins. The C-terminal half of the basic legumin chains exhibit substantial sequence homology to the low molecular weight avenins from certain cereals (Fabijanski et al., 1988) (Figure 1, panel C). The avenins are considered prolamins due to their solubility in alcohol. The signatures also share some homology with germins, a family of homopentameric cereal proteins expressed during seed germination and in response to salt stress and plant dehydration (Lane, 1991; Hurkman et al., 1994). The germins may play

```
A. Alignment of germin, spherulin and helianthinin

U21743        AGYPCIRPIHVKASDFVFSLGTPG-NT-TNIISAAVTPGFVAQFPALNGLGISTAR   87
M18428        APSELDRIKLLKDNQFVFDFKNSKLGV-TQGTGGKTVATSRTNFPAVIGHNVAMTV  111
M28832        GGWSNGVEETICSMKFKVNIDNPSQADFVNPQAGSIANLNSFKFPILEHLRLSVER  356
              .    |    . . .*  .. ..     ..  .. ..     .** .    .. .

U21743        LDLAPKGV-IPMHTHPGASEVLFVLDGSITAGFISSANSVYV-QTLKPGQVMVFPQ  141
M18428        GFIEACGINLP-HTHPRATEINFIASGKFEAGFFLENQAKFIGHTLEAGMATVFPQ  166
M28832        GELRPNAIQSP-HWTINAHNLLYVTEGALRVQIVDNQGNSVFDNELREGQVVVIPQ  411
                . . ..  @ *    * .. .. .@ . .  .. .    ..*  * . *.**

U21743        GLLHFQ-INAGKTPAAALVTFSSASPGLQILD---FALFANTLS  181
M18428        GAIHFE-INMNCEPAMFVAAFNNEDPGVQTTASSFFGLPADVVG  209
M28832        ---NFAVIKRANEQGSRWVSFKTNDNAMIANLAGRVSASAASPL  452
                 .*. *. . ...   ..*.... ..        .  *.
```

```
B. Alignment of sucrose-binding protein with two 7S storage globulins

L06038        GESKPQFNIFSKRPTISNGYGRLTEVGPDDDEKSWLQRLNLMLTFTNITQRSMST  360
M24845        GESRGPYSLLDQRPSIANQHGQLYEADARSFHD--LAEHDVSVSFANITAGSMSA  357
M19378        GSEGYAFNLLSQTPRYSNQNGRFYEACPRNFQQQLREVDSSVVAF-EINKGSIFV  434
              *..  .......  *  .*  *.. *. . . ..   .  .  ..* .*.  *. .

L06038        IHYNSHATKIALVIDGRGHLQISCPHMSSRSSHSKHDKS----------------  399
M24845        PLYNTRSFKIAYVPNGKGYAEIVCPHRQSQGGESERERGKGRRSEEEEESSEEQE  412
M19378        PHYNSKATFVVLVTEGNGHVEMVCPHLSRQSSDWSS-------REEEEQEEQEVE  482
              @ **...  .. * .@.*. .. ***  ..... .

L06038        --SPSYHRISSDLKPGMVFVVPPGHPFVTIASNKENLLMICFEV-NARDNKKFTF  451
M24845        EVGQGYHTIRARLSPGTAFVVPAGHPFVAVASRDSNLQIVCFEV-HADRNEKVFL  466
M19378        RRSGQYKRVRAQLSTGNLFVVPAGHPVTFVASQNEDLGLLGFGLYNGQDNKRIFV  537
                .  *. ... *..*  ****.*** . .**....* .. *.. ..  *..

L06038        AGKDNIVSSLDN  463
M24845        AGADNVLQKLDR  478
M19378        AGKTNNVRQWDR  549
              ** .* .   *.
```

```
C. Alignment of avenin, puroindoline, 2S albumin and legumin

                                                    .* ***.*. **
X63193                             |            60  RLSSCRDYL-ER  70
M38446   MKIFFFLALLALVVSATFAQYAESDGSYEEVEGSHDRCQQHQMKLDSCREYVAERCTTMR    60
X69913   MKALFLIGLLALVASTAFAQYSEVVGSYDVAGGGGAQQCPVETKLNSCRNYLLDRCSTMK    60
         **. * . *****.*  **** *  ***    *   .    . **.*** *...**.**.

                               ** * .*...***..*
X63193                     79  CCEELQRMSPQCRCQAIQ  96
M38446   DFPITWP-WKWWKGGCEELRNECCQLLGQMPSECRCDAIWRSIQRELGGFFGTQQGLIGKR   120
X69913   DFPVTWRWWKWWKGGCQELLGECCSRLGQMPPQCRCNIIQGSIQGDLGGIFGFQRDRASKV   121
         ***.**  ********.**  ***  ***** .***..*  *** .* * ** *.  ..*

M38446   LKIAKSLPTQSTWALSAISPNSMVSHIAGKSSILRALPVDVLANAYRISRQEARNLKNNR  180
X69913   IQEAKNLP                                                     128
J05485                      PNSMVSYIAGKTSILRALPVDVLANAYRISRQESQNLKNNR  484
         .. **.**           ****** ****.*********************..******

M38446   GQESGVFTPKFTQTSFQPYPEGEDESSLINKASE  214
J05485   GEEFGAFTPKFAQTGSQSYQD-EGESSSTEKASE  518
         *.* * *****.**. *.* . * ***  .****
```

Fig. 1. Similarities in amino acid sequences among the 11S and 7S storage globulins, 2S storage albumins, germin-class proteins, sucrose-binding protein and prolamins. The computer program CLUSTAL (Higgins and Sharp, 1988, 1989) was used to generate the multiple alignments shown. Default settings were: K-tuple value = 8, window size = 10, filtering level = 2.5, open gap penalty = 12, unit gap penalty = 10. *Panel A.*: Multiple alignment of the N-terminal portion of the basic 11S chain from helianthinin (sunflower, M28832), germin (rapeseed, U21743) and spherulin (slimemold, M18428). A vertical line indicates the position of the N-terminal glycine residue of the basic chain in the mature helianthinin subunit after post-translational cleavage at the conserved N-G peptide bond by asparaginyl endopeptidase. Amino acids that correspond to the conserved P/G motif described by Lawrence et al. (1994) as characteristic of 11S and 7S storage globulins are indicated by @. Asterisks (*) indicate the position of identical residues among the sequences and dots (·) indicate the position of amino acids with conserved physical properties. *Panel B*: Multiple alignment of sucrose binding protein from soybean (L06038) with 7S globulins from corn (M24845) and cotton (M19378). Sequence homology is greatest in the portions of the molecules illustrated, but extends further into the molecules. The 7S globulins from the legumes indicated in Table 1 also align, but with slightly less homology. Interestingly, stretches of highly charged amino acids that are typical of the storage globulins and albumins are absent in the sucrose binding protein. The notation of identical, conserved and P/G motif amino acid residues are as in Panel A. *Panel C*: Alignment of avenin (oat, M38446), puroindoline (wheat, X63193), a 2S albumin from *Picea glauca* (X69913), and an 11S storage globulin (oat, J05485). The signal peptide cleavage position in oat avenin is marked by a vertical line. Other notations are as described in Panel A. The second signature sequence for 2S proteins described in the text is well preserved in both avenin and puroindoline, whereas the first signature of the avenin sequence (M38446) is only partially conserved as compared to that in the 2S albumin sequence from *Picea glauca* (X63193). Accession numbers used to identify each sequence refer to records in the EMBL/DDBJ databases, and may be accessed on the Internet at http://ftp.bio.indiana.edu.

```
                SP region            N-terminal propeptide        SMALL

P80352                                                            EPLCRR     6
P01090    MANKLFLVSATLAFFFLLTNAS   IYRTVVEFDEDDATDS   AGPFRIPKCRK    49
P80207                                                       AGPFRFPRCRK    11
P38057                                                      PAGPFRLPRCRK    12
P15322                                                      PAGPFRIPKCRK    12
P15457    MANKLFLVCAALALCFLLTNA    SIYRTVVEFEEDDATN  PIGP-KMRKCRK    48
P04403    MAKISVAAAALLVLMALGHATA     FRATVTTTVVEEEN            QEECRE    42
P23110    MARFSIVFAAAGVLLLVAMAPVSEA  STTTIITTIIEEN  PYGRGRTESGCYQ    51
P26986    MGVFSPSTTRLTLKWFSLSVALFLLFHWGIPSVDGHEDN MYGEEIQQQRRSCDP    54
P29835    MASKVVFFAAALMAAMVAISGA                   HVSESEMRFRDRQCQR    38
                                                                *

          CHAIN                Internal Propeptide     LARGE

P80352    QFQQHQHLRACQRYLRRRAQRGGLA                        DEQRGPALRL    41
P01090    EFQQAQHLRACQQWLHKQAMQSGGGPSWTLDGEFDFEDDMENPQGPQQRPPLLQQ   104
P80207    EFQQAQHLRACQQWLHKQAMQSGSG                   PQPQGPQQRPPLLQQ    51
P38057    EFQQAQHLRACQQWLHKQARQSGSG                   PSPQGPQQRPPLLQQ    52
P15322    EFQQAQHLRACQQWLHKQAMQSGSG                   PSPQGPQQRPPLLQQ    52
P15457    EFQKEQHLRACQQLMLQQARQGRSD   EFDFEDDMEN      PQGQQQEQQLFQQ    96
P04403    QMQRQQMLSHCRMYMRQQMEESPYQ                     TMPRRGMEPHMSE    80
P23110    QMEEAEMLNHCGMYLMKNLGERSQ                      VSPRMREEDHKQL    88
P26986    QRDP-QRLSSCRDYLERRREQPSER                                    78
P29835    EVQD-SPLDACRQVLDRQLTGRERF                   QPMFRRPGALGLRMQ    77
          . .    *  *   . ..

          CHAIN

P80352    CCNQLRQVNKPCVCPVLRQAAHQQLYQGQI-------------------------    71
P01090    CCNELHQEEPLCVCPTLKGASKAVKQQIQQQGQQQGK------------------   141
P80207    CCNELHQEEPLCVCPTLKGASKAVKQQIRQQGQQQGQ------------------    88
P38057    CCNELHQEEPLCVCPTLKGAAKAVKQQIQQQGQQHGQ------------------    89
P15322    CCNELHQEEPLCVCPTLKGASKAVKQQVRQQLEQQGQ------------------    89
P15457    CCNELRQEEPDCVCPTLKQAAKAVRLQGQHQPMQ---------------------   130
P04403    CCEQLEGMDESCRCEGLRMMMMRMQQEEMQP------------------------   111
P23110    CCMQLKNLDEKCMCPAIMMMLNEPMWI----------------------------   115
P26986    CCEELQRMSPQCRCQAIQQMLDQSLSYDSFMDSDSQEDAPL--------------   119
P29835    CCQQLQDVSRECRCAAIRRMVRSYEESMPMPLEQGWSSSSSEYYGGEGSSSEQGY   132
          ** .*    .  * *   .

P80352    -------------------EGPRQVRRLFRAARNLPNICKIPAVGRCQFTRW      104
P01090    ---------------------QQMVSRIYQTATHLPKVCNIPQVSVCPFQKTMPGPSY   178
P80207    ----------------QGQQLQHEISRIYQTATHLPRVCNIPRVSICPFQKTMPGPS   129
P38057    ----------------QGQQLQHEIRRIYQTATHLPKVCNIPQVQVCPFNKTMPGPS   130
P15322    -------------------QGPHVISRIYQTATHLPKVCNIPQVSVCPFKKTMPGPS   127
P15457    ------------------------VRKIYQTAKHLPNVCDIPQVDVCPFNIPSFPSFY   164
P04403    --------------------RGEQMRRMMRLAENIPSRCNLSPMRCPMGGSIAGF      146
P23110    ----------------------RMRDQVMSMAHNLPIECNLMSQPCQM             141
P26986    NQRRRRREG--RGREEEEAMERAAYLPNTCNVREPPRRCDIQRHSRYSMTGSSFK      172
P29835    YGEGSSEEGYYGEQQQQPGMTRVRLTRARQYAAQLPSMCRVEPQQCSIFAAGQY       186
                             . .          . . *  * .
```

Fig. 2. Amino acid alignment of selected 2S seed storage albumins. The multiple alignment of ten 2S proteins was done using the algorithm described by Higgins and Sharp (1988) with the parameters described in Figure 1. The alignment was manually adjusted to highlight the domains of the transient signal peptide (SP), the N-terminal propeptide, and the internal propeptide that separates the small chain from the large chain. The most conserved portion of the small chain was used to derive the signature described in the text. The two most conserved parts of the large chain are separated by a hypervariable region that is extended in the cases of *Picea glauca* and *Oryza sativa* as compared to the other examples illustrated in the figure. Of the two signatures proposed for the large chain, the one from the N-terminal part of the molecule is the longest and most highly conserved among the proteins described in the figure (see text). Identical residues are marked by * (asterisk), chemically similar residues are marked · (dot). The proteins aligned are contained in the SwissProt database as follows: P80352, *Capparis masaikai* (Nirasawa et al., 1994); P01090, *Brassica napus* (Josefsson et al., 1987); P80207, *Brassica juncea* (Monsalve et al., 1993); P38057, *Sinapis arvenius* (Svendsen et al., 1994); P15322, *Sinapis alba* (Menendez-Arias et al., 1988); P15457, *Arabidopsis thaliana* (Krebbers et al., 1988); P04403, *Bertholletia excelsa* (Altenbach et al., 1987); P23110, *Helianthus annuus* (Kortt et al., 1991); P26986, *Picea glauca* (Newton, 1991); P29835, *Oryza sativa* (Shorrosh et al., 1992).

a role in altering the properties of cell walls during germinative growth (see also Lane et al., 1993, for a discussion about the relationship between wheat germins and oxalate oxidase). Similarity to germins was alluded to earlier in this discussion when vicilin from *Picea* was described (Section 3A). In this regard, the highest score of similarity was attained between germin (Saez-Vasquez, 1995) from *B. napus* and sunflower 11S helianthinin (vonder Haar et al., 1988), with 23 matches and 19 conservative amino acids within a stretch of 68 amino acids. In wheat, the germins are oxalate oxidases (Lane et al., 1993, Lane, 1994), and they exhibit homology to spherulins (Lane et al., 1991). The latter are proteins that accumulate during spherulation in slime mold. Upon spherulation, a process induced by environmental stress, the spherulin proteins are produced during an encystment process that leads to dormancy. An alignment of the conserved sequence within the N-terminal half of the basic chain of helianthinin to the homologous region of cotton vicilin, to germin from *B. napus* and sperulin is presented in Figure 1. Although structural data will be required to test the significance of these relationships, if true, they implicate proteins in much more primative organisms than recognized thus far may be related to the 7S and 11S seed proteins. The comparisons are also consistent with the common origin of these two groups of storage globulins.

C. 2S Storage Albumins. Many seeds contain a heterogeneous group of proteins frequently referred to as 2S proteins. Closer examination, however, will reveal that these can be subdivided into a number of families. One group sediments as 1.7–2.0S proteins. Unlike the storage globulins discussed previously, proteins in this group are soluble in water rather than dilute salt solutions and are more properly referred to as storage albumins. Although examples of members of this class are known for legumes (pea albumin PA1, Higgins et al., 1986), they are frequently found in elevated concentrations in seeds from non-legume plant species (Youle and Huang, 1981). The first example of this group of proteins examined in detail was napin from *B. napus* (Crouch et al., 1983; Ericson et al., 1986; Josefsson et al., 1987; Scofield and Crouch, 1987), but now at least ten examples have been reported (Figure 2). These mature 2S proteins, when isolated from seeds, consist of a small chain that is covalently linked to a larger chain by two disulfide bonds. As is characteristic of many seed storage proteins, the mature seed proteins that accumulate in the seed are derived from precursors after several post-translational modifications carried out by vacuolar endopeptidases (see Section 5). Of particular interest is the 1.8S protein from Brazil nut. This protein, like a number of the other 2S storage proteins, has a very high methionine content (Ampe et al., 1986; Altenbach et al., 1987; de Castro et al., 1987; Sun et al., 1987; Gander et al., 1991). Proteins with this trait could be used to markedly increase the seed methionine content in legumes. Unfortunately, however, the Brazil nut protein contains epitopes that are strongly antigenic in some individuals (Nordlee et al., 1994). This trait will limit the usefulness of the native pro-

tein for improving seed nutritional quality, especially in those cases where products from the seed can enter the human food chain. Other members of the 2S family besides the Brazil nut protein also have important functional properties that could impact on their usefulness for agronomic purposes. Proteins P15322 and P80207 elicit allergenic responses (Menendez-Arias et al., 1988; Monsalve et al., 1993), proteins P80352 and P30233 are heat stable and have sweetness characteristics (Liu et al., 1993; Nirasawa et al., 1994), while P38057 is a potent trypsin inhibitor (Ki $7x10^{-6}$) (Svendsen et al., 1994).

Figure 2 shows a linear amino acid alignment for ten 2S proteins and illustrates criteria to define the group and distinguish it from other proteins with similar sedimentation coefficients. The alignment reveals three fairly well conserved domains in the molecules. Each domain is separated from the others by peptides of variable length which contain charged or amidated amino acids that are characteristic of seed storage proteins. Not unexpectedly, each of the conserved domains is associated with cysteine residues in the molecules. For each of the conserved domains, a sequence signature can be assigned that may be used either individually or in combination with one another to assist in identification of this group of 1.8S proteins. These include:

Signature 1 (small chain): C′-X-X-[EQ]-X(5,6)-L-X-X-C″

Signature 2 (large chain): C″-C″-X-[EQ]-L-X(6)-C′-X-C-X-X-[LI]

Signature 3 (large chain): [AV]-X-X-[LIP]-P-X-X-C-X-[ILV]

where X is any amino acid, C′ denotes a conserved cysteine residue considered to be involved in an intrachain disulfide bond, and C″ refers to cysteine molecules involved in the interchain disulfide bonds. While these signatures are of general usefulness, not all of the 2S proteins in this family are unambiguously described by these ‘signature’ sequences. One notable exception is conglutin δ. Like other members of the 2S family, it consists of two disulfide linked peptide chains (Lilley and Inglis, 1986; Gayler et al., 1990). Although present, each of the three signatures is modified to some extent. Another interesting example of the divergence among this family of proteins is the 2S storage albumin found in jack bean (*Ricinus communis*). The jack bean 2S storage albumin is made from a large propolypeptide that probably gives rise to two binary chain 2S products (Irwin et al., 1990), one of which is derived from the N-terminal and the other from the C-terminal of the precursor encoded by this gene. Signature 2 is also widely associated with the Bowman-Birk types of protease inhibitors and the amylase/trypsin inhibitors (Rodriguez-Palenzuela et al., 1989; Garcia-Maroto et al., 1990; Gayler et al., 1990; Suzuki et al., 1993). Signature 2 also recognizes puroindolines, prolamin-like molecules from wheat (Gautier et al., 1994; Rahman et al., 1994) (Figure 1, panel D). Slightly altered signatures 1, 2, and 3 are Q00762 and Q05772, found in the tapetum from *A. thaliana* and *B. napus* (Paul et al., 1992). Cereal prolamins

have a sequence similar to signature 2 (Shewry et al., 1995). Neither the Bowman-Birk inhibitors, the tapetial proteins, nor the prolamines have the two peptide chain structure, even though they do contain sequences that bear some homology with one of the three 2S signatures. Finally, the sequences in Figure 2 bear a striking resemblance to the spore protein from ostrich fern, *Matteuccia struthiopteris*, a seedless plant species from the division *Pterophyta* (Roedin and Rask, 1990). Although much smaller than each of the other 2S proteins described in Figure 2, the spore protein not only has the two chain structure typical of the 2S protein family, but it also contains sequences homologous to 'signatures' 1 and 2. This obvious structural homology prompts one to speculate that the spore protein from *M. struthiopteris* is related to progenitors of this 2S family of proteins in angiosperms, although more definitive evidence establishing three dimensional structural similarity will be required to verify this suggestion. However, the potential relationship that exists justifies investigations to see if members of this storage protein family occur among more primitive organisms than now recognized.

4. Other prevalent seed proteins with biological activity

A. Legume Seed Storage Lectins. Numerous seeds accumulate proteins capable of agglutinating red blood cells and are referred to either as phytohaemagglutinins or lectins. A number of these accumulate in seeds in sufficient amounts so that they probably play a storage role in providing a source of carbon, nitrogen, and sulfur to be used during seedling growth and development. Because of their sugar-binding properties and ease of purification, lectins have proved useful for medical, cytological and biochemical studies aimed at distinguishing among cell types. These important properties have drawn interest to phytohaemagglutinins, and lectins purified from diverse sources have undergone considerable characterization (Sharon and Lis, 1989, 1990; Chrispeels and Raikhel, 1991). The purpose of this section is not to provide an extensive review of these proteins, but rather simply to draw attention to those that seem to play a storage role in the seed.

Table 3 identifies a number of legume seed storage lectins, and each bears substantial amino acid sequence homology to the others. Beyond sequence homologies, the lectins can be subdivided into subfamilies based upon whether they contain one or two peptide chains. The one-chain lectins are in general purified from seeds as tetramers. The two-chain phytohaemagglutinins, on the other hand, are usually purified as dimers, and each subunit in the dimer contains two peptide chains that are derived post-translationally from a common precursor. As a result of this maturation reaction, a small internal peptide is removed from the subunit precursor. The large fragments in the resulting two-chain structure are derived from the N-terminal end of the precursors known as β-chains, while the small α-fragments originate from

the C-terminal. The association between the α- and β-chains of the mature subunits are stabilized by non-covalent interactions as opposed to the situation encountered in the 11S legumin-like or 2S storage albumins. In the latter proteins, disulfide bonds seem to stabilize the interactions. The synthesis, post-translational modification, and assembly of the seed storage lectins proceed much like that of other storage proteins, as will be described more completely in Section 5.

Concanavalin A is an interesting variant in the family of legume seed storage lectins. Like the other lectins in this group, it is synthesized as a propolypeptide. The precursor undergoes post-translational proteolytic modification to yield a two-chain protein like many other lectins. During this process, however, the C-terminus of the α-chain becomes covalently attached to the N-terminus of the β-chain by a mechanism called cyclic permutation (Carrington et al., 1985; Bowles et al., 1986). The reaction is catalyzed by an asparaginyl endopeptidase found in the protein bodies of seeds (Ishii, 1994; Min and Jones, 1994). Thus, cyclic permutation results in a one-chain lectin, even though concanavalin A undergoes reactions similar to the two-chain members of the lectin group.

Two signature sequences have been described that are characteristic of seed storage lectins (Bairoch and Bucher, 1994):

Signature 1 (β-chain): [LIV]-[STAG]-V-[EQV]-[FLI]-D-[ST]

Signature 2 (α-chain): [LIV]-X-[EDQ]-[FYWKR]-V-X-[LIV]-G-[LF]-[ST]

These signatures provide a useful tool to screen for storage lectins and phylogenetically related proteins and permit each of the lectins listed in Table 4 to be identified in the database. An interesting example of a lectin identified in the database found using the signature sequence is a protein from *Lotus*. Despite sequence divergence and apparent differences in the mode of post-translational modification, this protein is clearly related to the seed lectins identified in Table 4, and may reflect ancestral progenitors of these molecules.

In addition to those lectins that play an obvious storage role, other lectins or lectin-like proteins occur in seeds. An important example are members of the ricin superfamily [ricin (Halling et al., 1985) and abrin (Wood et al., 1991; Hung et al., 1993)]. Members of this group, which are frequently referred to as ribosome inactivating proteins (RIPs), are rRNA N-glycosidases. Each possess the Shiga toxin signature shown below:

[LIVMA]-X-[LIVMSTA](2)-X-E-[AGV]-[STAL]-R-[FY]-[RKNQS]-X
-[LIVM]-[EQS]-X(2)-[LIVMF]

Unlike the typical storage lectins, the RIPs do not accumulate in sufficient amounts in seeds to contribute significantly to nutritional aspects of seedling

growth and development. Rather, they may serve in defensive roles against seed predators.

B. Lipoxygenases. Seeds accumulate lipoxygenases, large enzymes that catalyze the addition of molecular oxygen to the *cis*, *cis*-1-4-pentadienes of polyunsaturated fatty acids. Fatty acid hydroperoxides result from reactions catalyzed by these enzymes, and in soybeans the enzymes can account for 1–2 percent of the total seed protein at physiological maturity (Vernooy-Gerritsen et al., 1983). The lipoxygenases are of commercial interest in the food industry because they are responsible for the generation of both objectionable (Mustakas et al., 1969; Wolf, 1975; Hildebrand and Kito, 1984; Davies et al., 1987), and pleasant (Axelrod, 1974; Eskin et al., 1977; Galliard and Chan, 1980) aromas and flavors. The objectionable tastes originate from short chain aldehydes, principally hexanal and hexenal, that are derived from fatty acid hydroperoxides formed by the lipoxygenases. The pleasant flavors and aromas arise from aromatic aldehydes and alcohols and apparently are derived from alternative metabolic pathways. Defatted soy is also an important component in bread-making where it is responsible for bleaching carotenoid pigments to produce whiter bread. In this instance, the lipoxygenases catalyze oxidations and subsequent cross-linking of wheat glutens to enhance bread texture (Eskin et al., 1977). Interestingly, similar uses have been made of seed extracts from *V. faba* and *P. sativum* (Frazier, 1979).

Axelrod et al. (1981) described the preparation of four lipoxygenase isozymes (denoted L1, L2, L3a and L3b) from mature, dry soybeans. L3a and L3b are quite similar and, for the purposes of this discussion, will be considered to be the same. Kato et al. (1992) have described at least three new lipoxygenases that appear in cotyledons upon germination and at least two lipoxygenases that are distinct from the cotyledon enzymes have been found in the seed axis (Park et al., 1994). Each of the main cotyledonary enzymes have been purified, are m_r 95–98,000, contain a single atom of tightly bound non-haem, non-Rieske iron per molecule (Vliegenthart and Veldink, 1982; Shibata et al., 1988), and have been cloned and sequenced (Shibata et al., 1987, 1988; Yenofsky et al., 1988). The enzymes can be subdivided into two classes depending upon their substrate preferences, heat stability, and the pH optimum of the reaction they catalyze. L1 is considered prototype for Class-1 lipoxygenases. It has a pH optimum around 9.0, is heat stable and prefers fatty acids as substrate. The Class-2 enzymes, L2 and L3, have pH optima near neutrality, are heat inactivated and prefer esterified substrates (Christopher et al., 1970, 1972 a,b). Additional differences in regiospecificity (Christopher, 1972), secondary and peroxidative reactions (Garssen et al., 1971; Pistorius, 1974; Pistorius et al., 1976; Bild et al., 1977), anomalous dependence of reaction rate on substrate concentration (Christopher et al., 1972a,b), and the generation of singlet oxygen (Kanofsky and Axelrod, 1986) also exist. The enzymes differ in the proportions of 9- and 13-hydroperoxides they produce

(Christopher and Axelrod, 1971). Finally, x-ray crystal structure of L1 has been deduced (Boyington et al., 1993) and the structure of active sites in other isozymes determined (Minor et al., 1993).

Genetic null-alleles of genes encoding each of the three main soybean cotyledonary lipoxygenase isozymes have been identified (Hildebrand and Hymowitz, 1982; Kitamura et al., 1983, 1985; Davies and Nielsen, 1986). Only in the case of *lx2* has the genetic lesion responsible for the null phenotype been determined (Wang et al., 1994). From genetic segregation studies, the *Lx1* and *Lx2* gene loci were found to be linked, while the *Lx3* locus segregated independently from the other two genes. Mutagenesis of the L1L3-less double mutant was employed to obtain a triple null-line (Takamura et al., 1991), although an apparent triple null line has also been identified by searching progeny from a cross involving a L1L3-less double null and a L2-less null-line (Nielsen, unpublished data). The genetic relationships among the triple null prepared by mutation breeding and that obtained from the segragation population remain to be determined.

Significantly, the availability of the null-lines has facilitated the development of soybean cultivars in which the objectionable beany off-flavor due to hexanal has been reduced substantially (Davies et al., 1987). Removal of L2 from the seed has a dramatic effect in reducing off-flavor, and when null alleles for L2 and L3 are both combined, an even milder soymilk product can be produced. Finally, monoclonal antibodies have been generated that recognize each of the three types of isozymes (Evans et al., 1994), and these have proven beneficial in breeding efforts to select the desired plant genotypes.

Casey and colleagues have characterized lipoxygenases from *P. sativum*. Pea seeds contain two major lipoxygenases, both of which are Class-2 enzymes (Casey et al., 1985b; Ealing and Casey, 1988, 1989). In addition to these two major enzymes, however, three minor pea seed lipoxygenases are also detected (Domoney et al., 1990). As in soybean, the rate of accumulation of the enzyme is highest late in seed development. Genetic analysis of size variants of the major lipoxygenases have permitted a locus responsible for them to be located on linkage group 4, close to the *le* locus (North et al., 1989). RFLP analyses have also identified a genetic locus on linkage group 4, but other loci apparently are observed to react with the lipoxygenase probes used for these studies (Domoney et al., 1991). Therefore, in both pea and soybean, small families of genes encode seed lipoxygenases.

C. Miscellaneous Seed Proteins. A number of proteins have been reported in legumes that typically account for 5 percent or less of the total seed protein, but which are present in quantities greater than those generally associated with metabolic proteins. One such group is the narbonins, originally purified from *Vicia narbonensis* by Schlesier and Scholz (1974) and Schlesier et al. (1978). DNA-derived sequences are now available for narbonin (Nong et al., 1995) and related proteins from *Vicia pannonica* (Nong et al., 1994d), *Vicia sativa*

(Nong and Müntz, 1994a), *V. faba* (Nong et al., 1994f), *G. max* (Nong and Müntz, 1994b), and *C. ensiformis* (Nong et al., 1994e). The deduced amino acid sequences do not appear to contain a signal sequence, so it is unclear if these proteins accumulate in protein bodies or are located elsewhere in the cell. The deduced narbonin-like sequences are not obviously similar to any others described in the databases, although a weak relatedness to endo-β-N-acetylglucosaminidase H from *Streptomyces plicatus* (class II chitinase) has been proposed (Coulsen, 1994). X-ray crystallographic resolution to 1.8Å revealed that narbonin has an 8-stranded α/β-barrel structure (Hennig et al., 1992), the most common of all globular domain structures (Farber, 1993). These structures are known as TIM-barrels because of their original discovery in the triose isomerase molecule, but it remains unclear how the narbonins relate functionally to other proteins with similar three dimensional structural elements. The proposed chitinase structural background for the narbonins is of interest because there is a clear relationship between concanavalin B (Morrison et al., 1984) and chitinase (Setsuda et al., 1994). In the case of concanavalin B, however, and unlike the narbonins, a signal sequence was identified when the deduced amino acid sequences from cDNAs and actual protein sequence information were compared (Schlesier et al., 1995).

Several albumins from *P. sativum* have been identified that collectively account for as much as 10 percent of the total seed protein (Higgins et al., 1986). They are denoted as either PA1 or PA2 proteins. Two PA1 polypeptides, one m_r 6,000 and the other m_r 4,000, are isolated from mature pea seeds in 60% methanol and are both derived from a m_r 13,000 precursor. After removal of a signal sequence, the resulting m_r 11,000 protein is cleaved endoproteolytically to yield the two peptide chains. PA1 appears at about the same time during seed development as the major 11S storage proteins (Chandler et al., 1984). Amino acid and nucleotide sequence analyses are consistent with the presence of multiple, slightly diverged PA1 genes in developing pea cotyledons (Rerie et al., 1992). As mentioned previously, PA1 has some similarity to protease inhibitors and to other low-molecular-weight seed albumins (Higgins et al., 1986). The homologous protein from *P. vulgaris* may be involved in vacuolar sorting (Tanchak and Chrispeels, 1989). Because PA1 contains a relatively high concentration of cysteine (11%), it may prove useful as a vehicle to increase legume sulfur amino acid content.

The PA2 albumin consists of two distinct proteins m_r 24–26,000 in size, although their relative concentrations vary among pea genotypes (Schroeder, 1984; Rao et al., 1989). The PA2 protein has also been called PsaMA, PMA, MA, PMA-L and PMA-S (Croy et al., 1984; Harris and Croy, 1985; Rao etal., 1989). The mRNA encoding PA2 appears late in seed development, well after the onset in appearance of PA1 (Chandler et al., 1984; Rerie et al., 1992). A 230 amino acid sequence has been derived from a cloned nucleic acid sequence for PA2 and contains four imperfect repeat sequences (Higgins et al., 1987). It does not contain a signal sequence, does not undergo detectable post-

translational modification, and is considered to be accumulated in cytoplasm rather than protein bodies (Croy et al., 1984; Harris and Croy, 1985; Higgins et al., 1987). Because electrophoretic variation of the apparent molecular weights for the two PA1s occurs, it has been possible to show they are alternate products of a single genetic locus (Rao et al., 1989). The physiological role played by PA2 in the seed has not been resolved. However, unlike PA1, PA2 contains few sulfur amino acids and is not rich in amide-containing amino acids. Therefore, it does not share characteristics typical of most seed storage proteins targeted to the protein bodies. Nonetheless, discussions about PA2 occasionally appear in the seed storage protein literature.

Soybean proteins that sediment at 7 to 8S and have a basic isoelectric point were originally described by both Hu and Esen (1982) and Yamauchi et al. (1984). In each of these studies, the proteins were extracted in dilute salt solutions and were considered to be globulins. By contrast, Kagawa and Hirano (1989) described a basic 7S protein that could be released from seeds by briefly soaking seeds in warm water. On the basis of identity in their amino acid sequences, it is clear that the globulins purified by the earlier workers are the same as the 7S basic protein described by Kagawa and Hirano (1989). The protein has elevated amounts of the amide amino acids like typical seed storage proteins and can account for 5 to 10 percent of the total seed protein in some soybeans. It is isolated from the seed as an oligomer of about m_r 158,000 and is considered to be a tetramer whose subunits are about m_r 42,000 (Lilley and Nielsen, unpublished results). Like many seed storage proteins, each subunit consists of two protein chains joined by disulfide bonds. The larger of these chains has a basic isoelectric point and is m_r 27,000, while the smaller chain is about m_r 16,000. The basic protein has a higher sulfur amino acid content than either the typical 7S or 11S storage proteins, but it has fewer acidic amino acids. It is this latter property that imparts a high isoelectric point (pH 9.5) on the protein. Although directed into the lumen of the endoplasmic reticulum by a signal sequence, the protein is deposited in the middle lamella of cell walls and the plasma membranes rather than in protein storage vacuoles (Hirano et al., 1992). Conglutin χ from lupin (Derbyshire et al., 1976; Kolivas and Gayler, 1993), together with a protein from vegetative carrot tissue that is induced in response to wounding (Satoh et al., 1992) are homologous to the 7S basic protein. While the former protein from lupin is cleaved post-translationally, the one from carrot does not seem to undergo cleavage. Interestingly, these proteins also contain the motif DXGXXXLWV near the N-terminals, a motif associated with active sites of aspartyl proteases. No evidence exists, however, that it actually functions as an aspartyl protease. Based upon the induction profile of the carrot protein in response to wounding and its cellular localization, the 7S basic proteins are considered to be part of dermal tissue (Satoh et al., 1992).

An interesting seed protein of unknown function (USP) has been described in *V. faba* (Bassüner et al., 1983). Although not accumulated in significant

amounts in seeds, the abundance of mRNA encoding this protein exceeds that of mRNA encoding the major seed storage globulins. Both cDNA and genomic clones encoding the USP have been obtained and used to study some aspects of USP cell biology (Bassüner et al., 1988a,b). The USP is synthesized as a 30 kDa pre-proprotein on rough endoplasmic reticulum where it enters the cell secretory system like the storage globulins. During seed development, translation products of the USP gene are recovered as dimers whose subunits are disulfide linked and are localized in small electron dense vacuoles of the cotyledon (Bassüner et al., 1994). A search of databases with the derived USP amino acid sequence did not reveal any significant homology with other proteins, although a drought-inducible protein from *A. thaliana* does show limited similarity (Yamaguchi-Shinozaki and Shinozaki, 1993). The gene in clone *usp*30.1 contains two small introns and a 650-bp promoter region able to drive seed-specific expression of reporter genes in transgenic tobacco (Bäumlein et al., 1991). Although products from this gene have amino acid compositions more like metabolic proteins than storage proteins, and probably do not perform storage functions in the cell, its expression is seed specific and occurs at the same developmental time as the major storage globulins.

5. Synthesis and deposition of legume storage proteins

A general picture of the processes involved during the synthesis and deposition of the legume seed storage proteins has emerged during the past decade. Nuclear genes, which will be described in more detail later in this review, encode the storage protein subunits. Messenger RNAs derived from the genes are translated on rough endoplasmic reticulum (ER), and then these proteins undergo a series of post-translational modifications as they are transported from the ER to their final site of deposition in protein bodies. The initial translation products of the genes are pre-proproteins. As the translation products emerge into the lumen of the ER, signal sequences are removed co-translationally to form proglobulins (Sengupta et al., 1981). Core, high mannose, glycosyl groups become N-linked to asparagine residues in the 7S proteins, and the intra- and interchain disulfide bonds typical of the 11S proteins are formed as the proglobulins undergo chaperone-mediated folding and assembly into trimers (Nam, 1994). Chrispeels et al. (1982a,b) have proposed that both the 7S and 11S proglobulin oligomers transit the Golgi enroute to protein bodies, although Robinson et al. (1995) have proposed a somewhat different model. In the latter model, the Golgi is by-passed by the nonglycosylated storage proteins in the later stages of seed development. The model proposed by Robinson et al. (1995) bears similarity with the model described by Galili et al. (1995) to account for the intracellular movement of proteins in cereals during seed development. While in the Golgi, the core glycosyl groups on the 7S proglobulins undergo modification to form the glycans

TABLE 5

Post-translational proteolytic modification sites in proteins from seed storage vacuoles

Many proteins in seed storage vacuoles are modified post-translationally after arriving in protein storage vacuoles. The 11S globulins, some 2S albumins and 2-chain lectins are usually cut after asparaginyl residues by an asparaginyl endopeptidase (Scott et al., 1992; Muramatsu and Fukazawa, 1993; Hara-Nishimura et al., 1995). The cleavage sites for these proteins can be denoted P_1 (N) P_1' (X), where P_1 and P_1' correspond to amino acids on the amino and carboxyl terminal sides of the peptide bond that is cleaved. Asparaginyl endopeptidase from developing soybean seeds specifically requires asparagine at P_1, what exhibits considerably less specificity at P_1 (Jung et al., 1991). The 7S proteins are also frequently processed post-translationally. In these cases, however, cleavage usually does not allow the P_1 (N) P_1' (X) rule. Proteases other than asparaginyl endopeptidases that likely reside in protein storage vacuoles are candidates to carry out these reactions.

Processing steps involving the elimination of more than one propolypeptide are presented in more detail. In those cases, the N-terminal site of the propolypeptide is marked NH_2–; the propolypeptides are labeled *pp*, and the resulting native chains are marked *chain*, *ssu* (small subunit), or *lsu* (large subunit). The cleavage sites are indicated by double dashes (–). C-terminal propeptides are shown in full length followed by -COOH. Processing of a single propeptide or a single split are indicated by double dashes (–) with the adjacent amino acids. Multiple amino acid alternatives in one position are shown in square brackets; minor or alternative cleavage sites are in round brackets.

Accession numbers in the legumin section refer to the GenBand/EMBL/DDBJ database. All other accession numbers refer to the Swiss Protein database except for numbers with an asterisk, which refer to the PIR database. All databases are accessible through the Internet at http://ftp.bio.indiana.edu.

typical of many plant proteins (Sturm et al., 1987). In developing pea cotyledons, prolegumin and prolectin have been found in clathrin-coated vesicles (Harley and Beevers, 1989; Robinson et al., 1989). Therefore, it is likely that these organelles participate in the transport of storage globulins between the *trans*-Golgi and the protein storage vacuoles. After their arrival in storage vacuoles, the globulins undergo additional assembly, although this process is not well understood. Associated with these assembly mechanisms is the post-translational modification of each subunit in 11S oligomers into the mature disulfide-linked acidic and basic polypeptides. As will be discussed later, the legume seed lectins and 2S proteins also probably undergo endoproteolytic modification in this subcellular compartment. With several exceptions, post-translational modification of the 11S proglobulins is carried out at a well conserved Asn-Gly peptide bond by a unique asparaginyl endopeptidase (Hara-Nishimura et al., 1985, 1995; Hara-Nishimura and Nishimura, 1987; Scott et al., 1992; Muramatsu and Fukazawa, 1993). The exceptions include an Asn-Asn bond in legumin from *Gingko biloba* (Arahira and Fukazawa, 1994; Häger et al., 1995) and an Asn-Asp bond in a legumin subunit from

TABLE 5

Protein and Accession Number	Species	Naturally occurring cleavage sites	Reference
LEGUMINS			
CRA1, X14312	*Arabidopsis thaliana*	RHGN–GLEE	Pang et al., 1988
CRAB, X14313	*Arabidopsis thaliana*	EIAN–GLEE	Pang et al., 1988
GLAV 1 gene, X74740	*Avena sativa*	QSFN–GLEE	Tanchak et al., 1995
cruciferin, X62120	*Brassica napus*	PQDN–GLEE	Rodin et al., 1992
cruA gene, X14555	*Brassica napus*	EEVN–GLEE	Ryan et al., 1989
cru4 gene, X57850	*Brassica napus*	EEGN–GLEE	Sjodahl et al., 1991
BnC1 gene, X59294	*Brassica napus*	TEVN–GLEE	Breen and Crouch, 1992
BnC2 gene, X59294	*Brassica napus*	TEAN–GLEE	Breen and Crouch, 1992
pPG-beta-2, M36407	*Cucurbita pepo*	ESEN–GLEE	Hayashi et al., 1988
11S globulin, Z50777	*Ephedra gerardiana*	NAGN–GFPS	Braun et al., 1995a
legumin, 11S globulin, Z50777	*Ephedra gerardiana*	NAGN–GFPS	Braun et al., 1995a
ginnacin mRNA, X75426	*Gingko biloba*	REGN–NVEE	Arahira and Fukazawa, 1994
glycinin A1aBx precursor, X02985	*Glycine max*	SRRN–GIDE	Negaro et al., 1985
glycinin subunit A-2B-1a gene, K02646	*Glycine max*	RSRN–GIDE	Marco et al., 1984
glycinin Gy3 gene, X15123	*Glycine max*	QSRN–GIDE	Cho and Nielsen, 1989
soybean mRNA for G4 subunits of glycinin, X02626	*Glycine max*	ETRN–GVEE	Hirano et al., 1985; Momma et al., 1985; Scallon et al., 1987
soybean mRNA for G4 subunits of glycinin, internal acidic chain cleavage site, X02626	*Glycine max*	EQSN–RRGS	Hirano et al., 1985; Momma et al., 1985; Scallon et al., 1987

TABLE 5 (continued)

Protein and Accession Number	Species	Naturally occurring cleavage sites	Reference
LEGUMINS (continued)			
seed storage protein subunit A3-B4, mRNA, M10962	*Glycine max*	QTRN–GVEE	Fukazawa et al., 1985
legumin, 11S globulin, Z507729	*Gnetum gnemon*	GSGN–GVAE	Braun et al., 1995a
beta-globulin, mRNA, clone C94, M16905	*Gossypium hirsutum*	SQDN–GLEE	Chlan et al., 1986
beta-globulin B, mRNA, complete cds, clone C134 M16936	*Gossypium hirsutum*	RSGN–GLEE	Chlan et al., 1986
11S storage protein (G3-D1) DNA, M28832	*Helianthus annuus*	GWSN–GVEE	Vonder Haar et al., 1988
mRNA for legumin precursor (A2), X82464	*Magnolia salicifolia*	FRLN–GLEE	Fischer, 1994
mRNA for legumin precursor (B14), X82465	*Magnolia salicifolia*	GRRN–GFEE	Fischer, 1994
glutelin, X15833	*Oryza sativa*	SRWN–GLEE	Wen et al., 1989
mRNA for glutelin, pCSW-321, X06149	*Oryza sativa*	GCPN–GLDE	Wang et al., 1987
rice glutelin subunit mRNA, 17513	*Oryza sativa*	GCSN–GLDE	Higuchi and Fukazawa, 1987
glutelin 1 (Gt22) mRNA, M28259	*Oryza sativa*	SCSN–GLDE	Okita et al., 1989
glutelin, X14393	*Oryza sativa*	TRCN–GLDE	Masumura et al., 1989
glutelin 1 (Gt1) gene, M28156	*Oryza sativa*	GCPN–GLDE	Okita et al., 1989
glutelin type I, clone pREE 103, X05662	*Oryza sativa*	GCSN–GLDE	Takaiwa et al., 1987
mRNA for legumin-like storage protein, X63192	*Picea glauca*	GDEN–GVEE	Newton, 1991
major glutelin	*Pinus pinaster*	xxxx–AVEE	Allona et al., 1992
mRNA for pine globulin-1, Z11486	*Pinus strobus*	DSEN–GVEE	Rugh and Kamalay, 1991

TABLE 5 (continued)

Protein and Accession Number	Species	Naturally occurring cleavage sites	Reference
LEGUMINS (continued)			
legumin subunit pair precursor mRNA, clones pRC2.2.4 and pRC2.11.7, J01255	*Pinus sativum*	QGDN–GLEE	Croy et al., 1983
legJ gene for minor legumin, X07014	*Pisum sativum*	ERKN–GLEE	Gatehouse et al., 1988
clone pCD32, M16890	*Pisum sativum*	GRKN–GLEE	Domoney et al., 1986a
minor legumin	*Pisum sativum*	xxxx–FLEE	March et al., 1988
legumin-like storage protein mRNA, L07484	*Pseudotsuga menziesii*	ENEN–DVEE	Leal and Misra, 1993
pgCruRsE5 gene for cruciferin, X59808	*Raphanus sativus*	PQDN–GLEET	Depigny-This et al., 1992
legumin A1, pre-pro-polypeptide, X55013	*Vicia faba*	HGDN–GLEE	Schlesier et al., 1990
legumin B, Z26489	*Vicia faba*	GGKN–GLEE	Heim et al., 1994
legumin B gene LeB4, X03677	*Vicia faba*	QGRN–GLEE	Baumlein et al., 1986
RNA for legumin A, precursor, Z46803	*Vicia narbonensis*	HGDN–GLEE	Nong et al., 1994
mRNA for legumin A, Z32835	*Vicia sativa*	QGDN–GLEE	Nong et al., 1994
mRNA for legumin B, Z32796	*Vicia sativa*	QGRN–GLEE	Nong et al., 1994
legumin, 11S globulin, Z50780	*Welwitschia mirabilis*	RMAN–GVAE	Braun et al., 1995a
Minor vicilin, α-β cleavage site, J01260*, P02856	*Pisum sativum*	R[GS]L[RK]–D[RK]RQ	Spencer et al., 1983; Boulter, 1984
Minor vicilin, β-χ cleavage site, J01260*, P02856	*Pisum sativum*	RNENEQGKEN–D[HK]EE	Spencer et al., 1983; Boulter, 1984
α and α' subunits of β-conglycinin, P13916, P11827	*Glycine max*	NLLK–VE[EK]E	Doyle et al., 1986; Sebastiani et al., 1990
globulin-1, P15590	*Zea maize*	SSED–ERER	Belanger and Kriz, 1989; Kriz, 1989

TABLE 5 (continued)

Protein and Accession Number	Species	Naturally occurring cleavage sites	Reference
2S STORAGE ALBUMINS AND RELATED PROTEINS			
storage albumin (2S), P15457	*Arabidopsis thaliana*	NH_2-SIY. . . *pp*. . . ATN–PIG. . . *ssu* . . . RSD–EFD. . . . . . *pp*. . . MEN–PQG. . . *lsu*. . . FPS–FY-COOH	Krebbers et al., 1988
storage albumin (1.8S), P04403	*Bertholletia excelsa*	NH_2–FRA. . . *pp*. . . EEN–QEE. . . *ssu*. . . EES–PYQTM–PRR(–) GME. . . *lsu*. . . CNI–SPMR-COOH	Altenbach et al., 1987; de Castro et al., 1987; Gander et al., 1991
storage albumin napin (1.7S), P01090	*Brassica napus*	NH_2–IYR. . . *pp*. . . TDS–AGP. . . *ssu* . . . GPS–WTL. . . *pp*. . . MEN–PQG . . . *lsu*. . . MPG–PSY-COOH	Crouch et al., 1983; Ericson et al., 1986
conglutin δ (2S) storage albumin, P09930, P09931, B23217*, A23617*	*Lupinus augustifolius*	EED–HALKLRGIKHVIL–RHR	Lilley and Inglis, 1986; Gayler et al., 1990
PA1, P08687	*Pisum sativum*	NH_2–ASC. . . *ssu*. . . PSG–VFLRTN–DEHP. . . *lsu*. . . PK–LLKSVSTA-COOH	Higgins et al., 1986
storage albumin, P01089	*Ricinus communis*	NH_2–FAY. . . *pp*. . . RTN–PSQ. . . *ssu*. . . PRR-SDN–QER. . . *lsu* . . . CRF-COOH	Irwin & Lord, 1990; Irwin et al., 1990

TABLE 5 (continued)

Protein and Accession Number	Species	Naturally occurring cleavage sites	Reference
SEED STORAGE LECTINS			
concanavalin A, P02866	*Canavalia ensiformis*	NH_2–STH. . . *n-term. chain*. . . DAN –VIR. . . *pp*. . . DFN(–)AAYN–ADT. . . *c-term. chain*. . . KSN–EIP-COOH	Becker et al., 1975; Reeke et al., 1975; Carrington et al., 1985; Bowles et al., 1986
storage lectin, P12306, P12307, P04122	*Lathyrus ochrus*	TYPN–ETSY	Richardson et al., 1984; Yarwood et al., 1985
lentil storage lectin, P02870	*Lens culinaris*	TYPN–VTSY	Loris et al., 1993
pea storage lectin, P02867	*Pisum sativum*	TYPN–SLEEEN–VTSY	Higgins et al., 1983a, b; Hoedemaeker et al., 1994
storage lectin favin, P02871	*Vicia faba*	LYPN–LTGY	Hopp et al., 1982

Pseudotsuga menziesii (Douglas fir) (Leal and Misra, 1993). The 11S proteins in *P. sativum* (March et al., 1988) and *Pinus pinaster* (Allona et al., 1992) contain a Phe and Ala on the N-terminals of their respective basic chains, and these also are probably preceded by Asn residues in the proproteins (Table 5). While the purified asparaginyl endopeptidase is specific for Asn residues on the N-terminal side of the broken peptide bond, considerably less specificity is exhibited for residues on the C-terminal side (Jung et al., 1991), and this could account for the variability among legumin subunits that have been documented. Thus, consistent with standard nomenclature for proteolytic enzymes, this asparaginyl endopeptidase activity can be summarized as having $P_1(N)P_1'(X)$ specificity, where P_1 represents the amino acid on the N-terminal side, and P_1' the amino acid on the C-terminal side, of the peptide bond cleaved.

Several lines of evidence indicate that the post-translational cleavage is required for transition of 11S prolegumin trimers into hexamers during assembly in protein storage vacuoles. Assembly with proglobulins synthesized in vitro and carried out by Dickinson et al. (1989) provided the first direct demonstration that cleavage was necessary for hexamer assembly. In these original experiments, re-assembly of trimers into hexamers did not occur unless the trimers were first treated with papain. Subsequent experiments by Nam (1994) showed that cleavage of proglobulin trimers by purified asparaginyl endopeptidase promotes their direct assembly into hexamers. The results of Duranti et al. (1987, 1992) also support the conclusion that cleavage is required for conversion of proglobulin trimers to hexamers of mature cleaved subunits. They described legumin oligomers in white lupin seed in which trimers consisting of uncleaved subunits were accumulated in protein bodies together with hexamers composed of mature, cleaved subunits. Finally, deletion of the conserved Asn residue at the cleavage site in a B-legumin from *V. faba* permits accumulation of the mutant protein as trimers rather than hexamers in protein bodies in seeds of transgenic tobacco (Müntz et al., 1993). Perhaps cleavage somehow results in a protein conformational change that facilitates hexamer formation. However, the mechanism by which assembly occurs remains to be determined.

As alluded to previously, it has been possible to develop an in vitro system with which to study the various steps in the assembly of seed storage protein (Dickinson et al., 1987, 1988; Lelievre et al., 1992a). In this system, proglobulin precursors are synthesized in a rabbit reticulocyte lysate using transcripts produced in either SP6 or T7 promoter driven systems. After synthesis and incubation in the reticulocyte lysate, assembly of proglobulins into oligomers can be analyzed by isopycnic sucrose density centrifugation. With this system, it has been possible to analyze both the effect of various components from ER on assembly (Nam, 1994) and the consequence of various mutations in sequences encoding the proglobulins (Dickinson et al., 1989). Elimination of ATP from the system after synthesis of subunits, but before assembly has had

time to occur, blocks formation of oligomers and substantially enhances the recovery of insoluble globulin polypeptides after their synthesis (Nam, 1994). Because chaperone-dependent folding requires ATP, and because misfolded proteins are insoluble and tend to aggregate, the observation is consistent with participation of chaperones from the reticulocyte system in folding and assembly of trimers in vitro. This finding has been extended to the mechanism that occurs in seeds where components that remain unidentified have been shown to facilitate the folding and assembly reactions. The experiments carried out by Nam (1994) also revealed that correct formation of disulfide bonds in the proglobulins is required for efficient assembly of trimers, and that hexamer formation does not occur in vitro unless post-translational cleavage of the proglobulins is carried out using an asparaginyl endopeptidase preparation isolated from developing seeds. These data implicate the participation of protein disulfide isomerase in the folding of legumin and vicilin-like storage proteins in endoplasmic reticulum. In addition to identifying factors in the export pathway important for folding and assembly, the in vitro assembly procedure could provide an opportunity to evaluate the importance of conformational changes elicited by endopeptidase digestion, and how these might relate to the assembly mechanism.

The general features of the pathway leading to the formation and assembly of 7S and 11S oligomers are similar for the legume storage lectins and the 1.8S storage albumins. These proteins are synthesized on endoplasmic reticulum, traverse the cellular protein secretory pathway en route to protein bodies, and then are apparently modified post-translationally by proteases upon their arrival in protein bodies. Interestingly, the bonds cleaved in the two-chain lectins and concanavalin A resemble the P_1(N)P_1'(X) cleavage site of the seed storage globulins. On the basis of amino acid sequence information shown in Table 5, it can be deduced that the same kind of site is utilized for processing of some, but not all, 2S storage proteins. In those cases that are exceptions, P1(N)P_1'(X) target bonds can be identified near the sites cleaved in a few cases, and it is possible that amino or carboxyexopeptidases trim the ends of the protein chains after initial cleavage by an asparaginyl endopeptidase. Indeed, consistent with this suggestion, a ragged N-terminus has been described for the large chain of the high methionine Brazil nut protein, and C-terminal trimming has been documented in the small α-chain of pea lectin (Hoedemaeker et al., 1994). In other cases, however, the cleavage sites do not correspond to the target peptide bonds of proteases recognized to be present in protein bodies. Therefore, it is likely that proteolytic modification by enzymes in the various subcellular compartments through which exported proteins traverse will increasingly be discovered to play important roles in the assembly and deposition of seed proteins. The presence of proteolytic enzymes in compartments into which storage proteins are deposited has important practical implications for efforts to engineer improved seed quality. Engineered proteins must remain stable in the adverse environments

of the vacuole where they will be accumulated. This type of consideration may account for the instability of a number of engineered seed proteins tested in tobacco (Hoffman et al., 1987; Saalbach et al., 1988; Nielsen et al., 1995). Although a number of proteases are known to be contained in seeds, examples of which include carboxypeptidases, aminopeptidases, di- and tripeptidases, and aspartyl proteases (Shutov and Vaintraub, 1987; de Clercq et al., 1990; Bednarek and Raikhel, 1992; Hoedemaeker et al., 1994), little is known about their localization and function in the cell.

As indicated earlier, most vacuolar proteins are transported to the storage vacuole via the secretory system by as yet poorly characterized mechanisms. As in mammalian cells and yeast, secretion in plant cells occurs by a bulk-flow mechanism (Hofte and Chrispeels, 1992; see also Chrispeels, 1991b and references therein). Thus, in the absence of a positive sorting signal, proteins entering the secretory system are excreted from the cell. Elimination of the sorting signals from storage proteins results in their secretion from the cell (reviewed by Chrispeels and Raikhel, 1992).

What features of the vacuolar proteins permit retention of storage proteins in subcellular compartments during seed development? In plant cells, as in yeast or mammalian cells, proteins are retained in the ER by virtue of C-terminal HDEL or KDEL motifs. Proteins with typical ER retention signals include protein disulfide isomerase and BiP (*Bi*nding *P*rotein) (Denecke et al., 1991). That these C-terminal extensions are sufficient for targeting to ER can be demonstrated by changing the C-terminus of the vacuolar lectin phytohaemagglutinin from KL to KDEL (Herman et al., 1990). In response to this structural change, about half of the lectin is retained in the ER, whereas the remainder was accumulated in storage vacuole. Similar results were described by Wandelt et al. (1992) who modified the C-terminal of pea vicilin to include KDEL. Accumulation of the mutant vicilin in recombinant alfalfa mesophyl cells increased 100-fold due to the mutant gene. The protein was deposited in ER-derived, membrane bound, electron dense structures 0.5–1.0 micrometer in diameter and resembled the ER-derived protein bodies found in endosperm cells of maize and sorghum.

How are proteins targeted to the various locations within the secretory system? In mammalian cells, sorting to the lysosome, the equivalent of the vacuole in the storage parenchyma cells of seeds, depends upon information contained in glycans (Kornfeld and Mellman, 1989). In plants, however, the targeting information is contained within the primary structure of the protein (Chrispeels, 1991a). Three general types of mechanisms are recognized to be involved in these processes. In the first, accumulation in vacuoles depends upon information resident in a short C-terminal fragment that is removed from the propolypeptide during maturation. Examples include barley lectin and the vacuolar chitinases of cucumber and tobacco (reviewed by Raikhel and Lerner, 1991). The maturation of chitinase is an example of this mechanism. Elimination of these C-terminal propeptide sequences results in

recovery of vacuolar chitinase in the intercellular leaf fluid, whereas presence at the C-terminal of non-vacuolar proteins directs them to the vacuole. The second type of targeting mechanism is dependent upon an N-terminal propeptide. As in the case of C-terminal propeptides, the N-terminal peptide is also eliminated post-translationally during transport and maturation of the protein. Examples of proteins targeted by an N-terminal propeptide include sporamin, the storage protein in sweet potato (Matsuoka and Nakamura, 1991), and aleurain, a vacuolar thiol protease from barley (Holwerda et al., 1990). The third mechanism, in which a peptide is removed during maturation, appears to rely upon structural information contained within the protein. In these cases, the proteins apparently do not undergo detectable modification during movement to storage vacuoles. Examples of proteins targeted by this type of mechanism include the phytohaemagglutinin of *P. vulgaris* (Tague et al., 1990) and the legumin of *V. faba* (Saalbach et al., 1991). Interestingly, while the available data indicate that various methods have evolved to present targeting information to the sorting machinery in the secretory system, the environment in which the information is presented is important because protein secondary structure in the region surrounding the signal is an important factor in efficiency of targeting (Valls et al., 1990; Bednarek and Raikhel, 1992).

Information concerning the organization of storage proteins in the protein storage vacuoles is sparse. The deposition process of pea and field bean globulins has been studied by monitoring the immunogold labeling in thin sections by electron microscopy (Craig et al., 1980a,b; zurNieden et al., 1984). These studies showed that the storage proteins first appeared as small clumped deposits at the periphery of the large central vacuole of storage parenchyma cells. Microscopic examination of serial sections from developing seeds suggested that protrusions of the central vacuole became increasingly complex as seed development proceeded, and that the protuberances eventually fragmented into small, discrete, spherical vacuoles. Double immunogold labeling with gold particles of different sizes to mark 7S and 11S subunits revealed that these two proteins were co-localized within the same vacuole and distributed uniformly throughout the organelle. Because both the 7S and 11S vacuoles share many common features (Argos et al., 1985; Plietz et al., 1987; Wright, 1988; Lawrence et al., 1994; Shutov et al., 1995) and are co-located in the same vacuole, the question arises as to whether subunits from the two kinds of storage proteins can co-assemble. However, this does not seem to be the case. Oliveira (1994) demonstrated that neither monomers and trimers of the G4 proglycinin subunits, nor monomers and trimers of the α-subunit of β-conglycinin are capable of mixed assembly in vitro. If the folded conformations of the glycinin and β-conglycinin subunits produced in vitro resemble the shape of proteins formed in vivo, these results indicate that the two types of storage protein are sufficiently different so that they normally are not found together in the same oligomer.

Conclusions drawn by Colman et al. (1980) could provide insight concerning the organization of the proteins in seeds. These workers reported that small angle x-ray diffraction patterns of small crystals of cucurbitin (cucurbit 11S protein) spun into a pellet are the same as those obtained from sections of dry and wet native seed. Additionally, they reported that edestin, the crystalline storage globulin from *Cannabis*, and a crystalline globulin from tobacco seed both have the same crystal lattice as cucurbitin and, very likely, the same subunit structure. Their data, together with the observations by Oliveira described above, could indicate that small crystals consisting exclusively of 11S proteins, and others consisting exclusively of 7S proteins, are assembled independently in the protein storage vacuole during seed ontogeny, and that the two types of structures are distributed uniformly throughout the organelle. While the subcellular localization of 2S albumins and seed storage lectins and protease inhibitors are in the protein storage vacuoles, how these proteins are arranged with respect both to one another and to the major seed storage proteins is unknown. However, the considerable variation in the content of storage organelles alluded to earlier in this review would be facilitated by a random packaging mechanism of the type suggested.

6. Seed storage protein genes

An impressive number of seed protein genes have now been sequenced, and citations for many of these are contained in the information given in Tables 1, 2, and 3, and Figure 2. The coding regions of most 11S genes are interrupted three times, and the 7S genes five times, by short introns whose positions appear conserved among genes in different species. The borders of the introns conform with those reported for other eukaryotic genes, but introns in different genes vary in length and sequence. With the exception of what has been called the hypervariable region (HVR) in exon 3 of the 11S protein genes (Nielsen et al., 1989), differences among exon sequences are base substitutions, together with small insertions or deletions. Sequence homology between corresponding 11S or 7S genes range from near 40% to greater than 90% depending upon the relatedness of the species being compared. The homologies are sufficiently high to clearly indicate that all of the genes are related to a common ancestral gene irrespective of the evolutionary distance between species.

As indicated previously, large differences can occur in exon 3 of 11S proteins where insertions of more than 1500 bp can be found in the HVR. These in large part account for the significant size differences between Group-1 and Group-2 soybean 11S genes (Nielsen et al., 1989) and the homologous A and B genes in *V. faba* and *P. sativus* (Wobus et al., 1984, 1986). Interestingly, the number of introns in 11S genes is known to be variable. For example, the Group-2 (B) genes from *V. faba* generally lack the third intron (Bäumlein et

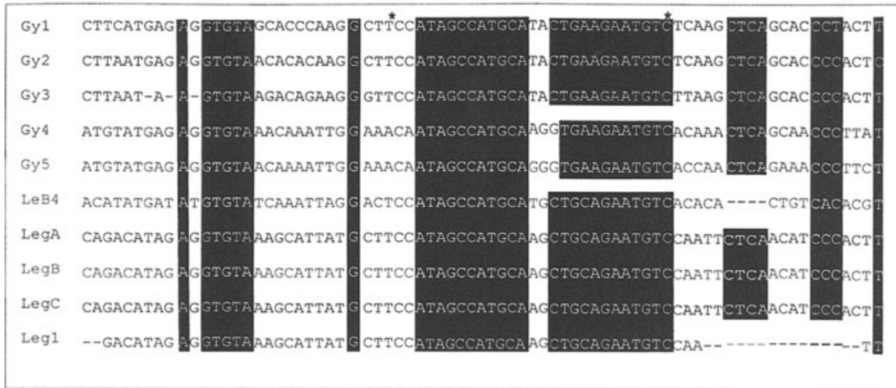
al., 1986; Heim et al., 1989, 1994), as is also the case for helianthinin genes from sunflower (vonder Haar et al., 1988). *G. biloba*, on the other hand, contains a fourth intron near the 3′ end of the gene (Häger et al., 1995). Based on the placing of introns within the 11S genes, Shutov et al. (1995) speculate that introns-1, -2, -3, and -4 of the 11S protein genes in *Gingko* correspond to introns-1, -2, -4, and -5 in the vicilin genes. Shutov et al. (1995) suggest that this, together with the substantial structural similarity among the 11S and 7S protein structures, is evidence that the ancestral gene that led to both of these proteins predates the gymnosperms.

It is well established that the genes encoding the predominant seed proteins are subjected to precise developmental regulation of both a temporal and spatial nature. Careful and extensive studies by Goldberg and his colleagues (Goldberg and Perez-Grau, 1989; Goldberg et al., 1989, 1994) have revealed many features about these regulatory circuits, and these are discussed more completely elsewhere in this volume. Briefly, however, several models that complement one another have been elaborated to help explain transcriptional regulation of seed protein genes. One is a combinatorial model for promoter function and was first elaborated by Benfey and Chua (1989, 1990) on the basis of experiments carried out with the cauliflower 35S promoter. According to this model, positive and negative acting expression elements are interspersed in the upstream regulatory sequences of plant genes, and these elements control both qualitative and quantitative features of gene expression. The second model proposed by Thomas et al. (1991) stresses the bipartite nature of plant promoters. According to Thomas, proximal promoter regions such as the TATA-like and CAAT-like elements confer seed-specific expression, while the more distal regions refine and enhance the basic expression pattern of the gene (Thomas, 1993). Examples of distal elements include A/T rich regions, ABA-responsive elements, as well as 'legumin-box' and 'vicilin-box' regulatory elements. A third model was recently described by Frisch et al. (1995). It is directed toward explaining how seed-specific transcriptional regulation might be achieved via interaction of factors with the promoter region of genes. It recognizes the requirement that histones associated with the chromatin structure must be removed in order that *trans*-acting factors can bind to the promoter. This event might be achieved in several ways; for example, by developmentally programmed events like endoduplication that might permit *trans*-acting factors to interact and thereby inhibit re-attachment of histones. Alternatively, seed-specific *trans*-acting factors might actively disrupt the chromatin structure while interacting with nucleosomal DNA to prevent nucleosome reassembly. Whatever the mechanism, the naked DNA serves as a target of general transcription factors that finally modulate and initiate gene transcription. Although promoter function remains far from being understood, each of these models is helpful when trying to understand the function of the expanding number of regulatory elements discovered in promoters of seed protein genes. These will be briefly described.

In addition to TATA elements involved in recognition of the seed protein genes by RNA polymerase II, the 11S protein genes from legumes contain a 28-bp element in the proximal promoter region that is referred to as the legumin-box (Bäumlein et al., 1986). The element contains a CATGCAT motif located at its center, and this motif is often repeated several times upstream in promoters of many 11S protein genes (Dickinson et al., 1988). The CATGCAT motifs are widely dispersed among seed protein genes and, in addition to being present in many 7S lectins and genes (Dickinson et al., 1988), are found in the 1.8S napin genes (Ericson et al., 1991), pea albumin PA1 (Higgins et al., 1986), the ricin gene (Halling et al., 1985), and the unknown seed protein (USP) from *V. faba* (Bäumlein et al., 1991; Fiedler et al., 1993). The function of the CATGCAT motif in promoters from a Gy2 glycinin gene of soybean (Lelievre et al., 1992b) and a LeB legumin gene from *V. faba* (Bäumlein et al., 1986) have been investigated by ligating normal and modified promoters to β-glucuronidase (GUS) reporter DNA, and testing promoter function by introduction of this chimeric gene into tobacco. Both studies revealed that deletion of the CATGCAT core element in the legumin-box reduced GUS expression 20- to 50-fold (Bäumlein et al., 1992; Lelievre et al., 1992b). Oliviera (1994) extended these original observations by using site-directed mutagenesis to invert the central 4 bp of the motif to CACTACT. Inversion of these nucleotides, like deletion of the element, caused a drastic reduction in GUS promoter activity. Because base pair substitutions were considered less detrimental to the spatial arrangement of nucleotides in the promoter than deletions, the results make it more likely that it is the changes to the CATGCAT motif that result in loss of promoter activity rather than conformational changes in the promoter elicited by deletion of the motif.

As shown in Figure 3a, the CATGCAT motif is not the only one conserved in DNA surrounding the legumin-box. Extensive highly conserved domains are found on both 5′ (A_GTGTA) and 3′ (TGAAGAATGTC) sides of the CATGCAT element. When the DNA nucleotides surrounding the legumin-box are arranged such that their positions in a DNA helix can be visualized, it becomes apparent that the conserved regions are more heavily clustered on one side of the helix than the other (Figure 3b). This observation could indicate that this region serves as a receptor for a *trans*-acting factor (Figure 3b). Oliveira (1994) tested the effect of changing the register of the three conserved motifs with respect to one another by the addition of additional nucleotides between the conserved sequence motifs. By moving the 5′ motif half of a turn with respect to the CATGCAT and 3′ motif, or by making a similar change between the central and 3′ motifs, a down regulation in promoter activity was observed. Interestingly, the magnitude of the down regulation differed between the two cases tested. Severe down regulation, equivalent to removal of the CATGCAT motif, occurred when the spatial relationship of the 5′ and central CATGCAT elements was perturbed. However, only a 2-fold decrease in activity took place when the spatial relationship between the

A.



B.

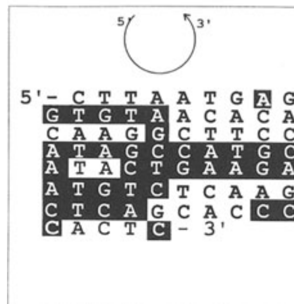


Fig. 3. Comparison of the legumin box region from ten 11S seed protein gene promoters. A: Alignment of the promoter regions. Conserved residues are marked by inverse printing. Arrows denote the boundaries of the legumin-box described by Bäumlein et al. (1986). A CATGCAT element is centrally located within the legumin box (Dickinson et al., 1988) that is an important enhancer of expression from these genes (Bäumlein et al., 1992; Lelievre et al., 1992b). The figure indicates the presence of conserved nucleotide sequences both 5′ and 3′ from the legumin box. *Vicia faba* LeB4 genes (Bäumlein et al., 1986); pea LegA, LegB, and LegC genes (Fischer and Goldberg, 1982) and group-II Gy4 and Gy5 (Scallon et al., 1985). B: Display of the legumin-box nucleotides in a helical array. Each row corresponds to one turn of the helix. Nucleotides shown in inverse print are conserved in 11S gene promoters as indicated in the A panel. The data indicate the conserved sequences, which are separated from one another when arranged in linear array, would be located in patches if the DNA is arranged in a helix in vivo. Asterisks indicate the site of four base pair insertions that alter the spatial relationships between conserved sequences and result in loss of promoter activity (Oliveira, 1994).

central and 3′ elements was altered. These data indicate that the CATGCAT motif probably acts combinatorially with flanking regulatory elements in the legumin-box. While it is tempting to speculate that the alterations perturbed the interaction of a *trans*-acting factor, other possibilities related to changes

in promoter conformation undoubtedly exist that could equally well account for the changes observed. In this regard, while Ericson et al. (1991) and Wobus et al. (1995) have reported the binding of nuclear factors to either CATGCAT-like elements or the legumin-box, respectively, we (unpublished results) and others (Riggs et al., 1989; Shirsat et al., 1990) have been unable to detect binding. Perhaps transcription complexes that consist of multiple components interact within this region of the promoter which would make it difficult to reconstruct the complex in vitro. Interestingly, the CATGCAT motifs are RY elements theoretically able to form Z-DNA structures, and this in turn may facilitate recognition and binding by transcription factors.

In the case of the Gy2 glycinin gene and many other seed gene promoters, one or more additional CATGCAT motifs are present in the promoter, often in the distal regions of the promoter. Therefore, it was of interest to determine CATGCAT motifs distal to the legumin-box Gy2 promoter activity. Evidence reported by Oliveira (1994) clearly showed that neither elimination nor base pair replacements in CATGCAT motifs distal to the legumin-box in the Gy2 glycinin promoter caused a significant change in the level of GUS expression when the central CATGCAT in the legumin-box remained intact. When the distal motif was intact but the core CATGCAT sequence in the legumin-box was defective (Oliveira, 1994), GUS expression remained low. In this case, motifs located elsewhere in the promoter appear unable to substitute for the one located centrally in the legumin-box. Thus, the presence of the CATGCAT motif in the absence of other combinatorial elements apparently does not exhibit enhancer activity.

Insight about a possible function of the CATGCAT motif can be obtained from studies about the α' subunit gene promoter of β-conglycinin. The –140 to +13 proximal upstream regulatory sequences (URS) of this gene are sufficient for seed-specific expression. Within this region, two CATGCAT elements and binding sites for nuclear factors SEF3 and SEF4 have been identified (Allen et al., 1989; Lessard et al., 1993). The effects of mutations in these elements were tested by attaching the proximal URS from the α'-subunit gene promoter to a *uidA* reporter sequence (Fujiwara and Beachy, 1994). Mutation of the CATGCAT motifs, as well as those that bind SEF3 and SEF4, had little impact on expression from the α' gene promoter. When the α' gene promoter was ligated upstream from a core 35S promoter, however, mutations in the CATGCAT element abolished the seed specific enhancer activity of the α' gene promoter and caused expression in leaves instead of seeds. These results suggest that the CATGCAT motif plays a role in the control of seed-specific expression by this particular α' gene promoter construction, but how it interacts with other factors to accomplish seed specific regulation remains unclear. Although enhancement of gene expression is often associated with the presence of the CATGCAT motif, this is not always the case. Mutation of this motif in the proximal promoter of the USP gene from *V. faba* leads to enhancement of expression rather than inhibition. Thus, it is

the environment in which the motif is located within the promoter and the influence of mutations on the spatial organization of bases in that environment, that apparently determines the phenotype observed.

The legumin-box and the CATGCAT motifs in legume genes probably do not perform analogous roles in all 11S storage protein genes from some nonlegumes. Thomas and his colleagues have described the promoter for a helianthinin gene (Bogue et al., 1990), which encodes an 11S storage protein gene in sunflower. A region with less than 40 percent homology to the legumin-box can be identified around 200 bp 5′ to the transcription initiation site, but this region does not contain a putative CATGCAT core element. Nonetheless, an imperfect ATGCATG motif is located about –130 bp in the promoter. In the case of the helianthinin genes, W/S elements with the consensus sequence WGATST are implicated in regulation of expression. Four such elements can be found within about the first 110 bp in the promoter. Their disruption by point mutations abolishes promoter activity in embryonic tissue as judged by expression from a GUS reporter sequence. The point mutations also abolish the binding of nuclear proteins to the W/S motifs and reduce the activity of the promoter in transgenic tobacco plants (Thomas, 1993). Perhaps the W/S motifs act in concert with other proximal elements, and an important aspect about their modification is the accompanying alteration of the topology of the promoter complex. Interestingly, while W/S motifs can be found in the glycinin 11S promoters, data are unavailable to associate them with the regulation of activity in these genes.

DNA sequences homologous to the endosperm-box (E-box) core element ACGT of maize zein genes (Schmidt et al., 1992) have been identified in promoters for helianthinin, phaseolin, and the USP (Fiedler et al., 1993; Thomas, 1993) from *V. faba*. The E-box refers to the binding site for the basic zipper (bZIP) nuclear protein synthesized by a gene at the *opaque2* (*o2*) regulatory locus of maize. This bZIP protein and a second, OHP1, bind to the *o2* motif in promoters of zein genes as either homo- or heterodimers (Pysh et al., 1993). A motif in the 7S phaseolin promoter that is similar to the E-box sequence apparently binds recombinant basic *o2* protein (Thomas, 1993). The proximal promoter regions of the helianthinin promoter (–116 to +24) contain two putative E-boxes, but mutations directed to these motifs do not affect binding of nuclear proteins. They are considered unlikely to interact with bZIP proteins of the helix-loop-helix type (Thomas, 1993). E-boxes are not located exclusively in the proximal promoter domains. They can also be found in distal regions of the 7S phaseolin promoter, and Kawagoe and Murai (1992) suggested that such an element (CANNTG) binds specifically to seed nuclear proteins. Like other promoter elements, the E-boxes are thought to exert their function combinatorially. For example, the core E-box motifs have been implicated in ABA-regulation in wheat Em and rice rab16 genes.

The 7S protein genes of legumes contain a conserved sequence in the proximal promoter called the 'vicilin-box' (Gatehouse et al., 1986; Bown et

al., 1988; Higgins et al., 1988), but its function in the regulation of these genes is not well resolved. The α' gene promoter in soybean also contains the short sequence motif AA/GCCCA, which is repeated five times within approximately the first 270 bp of the 7S gene encoding the α'-subunit of β-conglycinin. Analysis of *cis*-acting regulatory elements in the promoter region of this gene have implicated this motif as being involved in gene regulation (Chen et al., 1986, 1988, 1989). These workers have shown that an intact TATA box by itself is incapable of promoting a measurable level of gene transcription in transgenic petunia plants, an observation consistent with other promoter deletion experiments. When the promoter is attached to a reporter sequence, transcriptional activity is detectable when the promoter contains bases up to position –159 but is stimulated 16- and 20-fold when the 5′ flanking sequences are extended to –208 and –257, respectively. The sequence motif AA/GCCCA is repeated once in the –159, four times in the –208 deletion, and five times in the –257 deletion gene constructions. The number of motifs in the deletion gene constructions parallels a step-wise enhancement of expression. In vitro DNA binding assays, together with a series of mutations to alter AA/GCCCA motifs in the DNA fragments, have provided additional evidence that these are recognition sequences for transcriptional binding factors (Allen et al., 1989). A *trans*-acting factor designated SEF3 binds to DNA fragments from the proximal α'-subunit gene promoter that contain these motifs. Consistent with its role in regulating transcription, SEF3 is detected only in embryonic soybean tissues and increases in parallel with expression of the α'-subunit gene.

Additional DNA regulatory elements are found in the distal regions of seed protein promoters that modulate levels of gene expression. The A/T-rich domains, which bind prevalent, ubiquitous nuclear proteins, are examples that have been identified in most, and perhaps all, seed protein genes. These regions have been studied extensively in the 7S phaseolin gene in which they serve as an enhancer (Bustos et al., 1989a,b, 1991a,b,c) and, when fused to the cauliflower 35S minimal promoter, yield high levels of root specific expression. Because the A/T-rich phaseolin promoter acts cooperatively with the root-specific enhancer of the cauliflower 35S minimal promoter, it is considered a class-b enhancer (Fromenthal et al., 1988). A/T-rich enhancer elements have also been identified in sunflower and soybean 11S protein gene promoters (Jordano et al., 1989; Lelievre et al., 1992b; Itoh et al., 1994).

Pederson et al. (1991) showed that, in some cases, the nuclear proteins that bind to the A/T-rich domains are high mobility group chromosomal proteins. This raised the possibility that some A/T rich motifs associated with highly expressed seed protein genes could function as scaffold attachment sites (Bonifer et al., 1990). Scaffold attachment regions, or SARs, are typically located in noncoding regions of DNA and function as binding sites between chromatin and the nuclear matrix. Such attachment regions are considered to be involved in chromatin organization, a concept required to explain the

observation that genes are located at specific positions within the nucleus in animal cells (Manuelidis and Borden, 1988). The SARs (300–1000 bp in length) are typically located in noncoding DNA, are A/T rich, and isolate DNA loops from the influences of flanking DNA. Because proteins involved in the transcription process (topoisomerase II, RNA polymerase II, helicase) are associated with the nuclear matrix, genes located near scaffold attachment sites are favorably positioned for transcription and exhibit high levels of expression that are independent of their location in the genome (Bonifer et al., 1990). In this regard, van der Geest et al. (1994) recently demonstrated that SARs were present on either side of the β-phaseolin gene in clone λ 177.4 (Sun et al., 1981; Murai et al., 1983). This observation accounted for the high amounts of expression and the low plant-to-plant variation noted by Sengupta-Gopalan et al. (1985) when the gene in this clone was introduced into tobacco. The A/T-rich SAR domains located on either side of the β-phaseolin gene are distinct from the A/T-rich enhancer elements in these genes that were described by Bustos et al. (1989b). Those A/T-rich enhancer sequences are located between the flanking SAR regions and respond to insertion into different environments of genomic DNA by plant-to-plant variation of expression levels when tested in transgenic plants. The discovery and use of SARs has important practical implications. Inclusion of these regions in engineered genes may be useful to ensure optimal levels of expression of engineered genes.

In addition to the positive regulatory elements discussed so far, ones that exert negative influences on expression have also been observed in seed protein promoters. Their presence is generally identified by an increased promoter activity upon their removal or inactivation. The $(CA)_n$ motif is one example that has been reported (Goldberg, 1986; Lessard et al., 1993; Vellanoweth and Okita, 1993). This motif seems to down-regulate the effect of more proximal seed elements, but its inhibitory effect can be reversed by distally located enhancer sequences (Lessard et al., 1993). A second example is described by Bustos et al. (1991b), who reported that the motif AGAAC/AA occurs frequently in negative regulatory sequences in the 7S phaseolin gene. Interestingly, this motif may interact with a nuclear factor AG-1 (Kawagoe and Murai, 1992). Finally, the CATGCAT motif, which, as described earlier, functions as an enhancer in many genes, seems to cause down-regulation of expression from the USP gene promoter (Fiedler et al., 1993). Determination of the molecular mechanisms by which the various positive and negative promoter elements operate, and the consequence of conformational changes in promoter structure elicited by their addition and removal, should provide interesting topics for future research.

7. Concluding remarks

Much of the driving force behind the study of seed proteins is rationalized as part of an effort to increase seed quality in crops of agronomic importance. While this is certainly true, the study of these proteins has also served as an important vehicle to increase our basic understanding of plant cell biology. Although additional descriptive data about seed proteins from other plant species can be expected in the future, increased emphasis will be placed on understanding the structural biology of these macromolecules. The focus will change from understanding the structures of the molecules to determining how the various parts interact with each other and contribute to the function of the protein. This information would appear critical to efficient attempts to alter proteins so that they more effectively meet the functional and nutritional requirements demanded of them in food and fiber systems. Of equal importance, however, are two other issues. One concerns understanding the regulatory circuits that control expression of the seed protein genes, and the second concerns an increased awareness about the biological events that surround the synthesis and deposition of these proteins. Although a number of potential gene regulatory elements have been discovered that seemingly affect developmental activity, precise knowledge about how they exert their effect, what factors interact with them and in what manner, remain largely unknown. Although hints appear in the literature that the three dimensional organization of the genetic material in the cell nucleus plays an important role in the control of gene expression and cell development, much of the effort in this area of research remains of a conceptual nature. The investigative tools presently available make it difficult to relate the linear array of nucleotides in gene regulatory regions to the three-dimensional space they occupy and to visualize how effectors interact with the genes to control expression. Similarly, although we know that most seed storage proteins enter the secretory pathway and are eventually localized in specialized storage vacuoles, only rudimentary knowledge exists concerning the events that occur during this process. What chaperones act on the proteins to guide their folding and assembly in the endoplasmic reticulum, and then what factors act to guide the proteins through the elaborate series of specialized membrane-bound compartments that comprise the secretory pathway? Thus, for the foreseeable future, a resolution of the practical goal of improving seed quality will continue to depend upon research about the basic biology of seed storage proteins and the regulation of the genes that encode them.

References

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

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

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

Ampe, C., Van Damme, J., de Castro, L.A.B., Sampaio, M.J.A.M., Van Montagu, M., and Vandekerckhove, J. (1986) The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (Bertholletia excelsa H.B.K.). Eur J Biochem 159: 597–604.

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

Ashton, F.M. (1976) Mobilization of storage proteins of seeds. Annu Rev Plant Physiol 27: 95–117.

Axelrod, B. (1974) Lipoxygenases. In: Whitaker, J.R. (ed) Advances in Chemistry Series, No. 136, pp. 324–248, American Chemistry Society, Washington, DC.

Axelrod, B., Cheesbrough, T.M., and Laasko, S. (1981) Lipoxygenase from soybeans. In: Lowenstein, J.M. (ed) Methods in Enzymology Vol. 71, pp. 441–451, Academic Press, New York.

Bairoch, A., and Bucher, P. (1994) PROSITE: Recent developments. Nucleic Acids Res 22: 3583–3589.

Bassüner, R., Bäumlein, H., Becker, C., Evans, I.M., Hillmer, S., Müntz, K., and Vorgias, C.E. (1994) Genotype-dependent expression and subcellular localization of a non-storage seed globulin. Abstract of 6th Int Seed Protein Symp., 'Molecular and Cellular Mechanisms of Seed Formation and Deposition'.

Bassüner, R., Bäumlein, H., Huth, A., Jung, R., Wobus, U., Rapoport, T.A., Saalbach, G., and Müntz, K. (1988a) Abundant embryonic mRNA in field bean (*Vicia faba* L.) codes for a new class of seed proteins: cDNA cloning and characterization of the primary translation product. Plant Mol Biol 11: 321–334.

Bassüner, R., Bäumlein, H., Jung, R., Saalbach, G., Müntz, K., and Wobus, U. (1988b) Analysis of complementary and genomic DNA coding for a highly abundant class of mRNA in developing field bean seeds (*Vicia faba* L. var. minor cv. Fribo). Biochem Physiol Pflanz 183: 225–231.

Bassüner, R., Manteuffel, R., Müntz, K., Puchel, M., and Schmidt, P. (1983) Analysis of in vivo and in vitro globulin formation during cotyledon development of field beans (*Vicia faba* L. var. minor). Biochem Physiol Pflanz 178: 665–684.

Bäumlein, H., Boerjan, W., Nagy, I., Bassüner, R., Van Montagu, M., Inze, D., and Wobus, U. (1991) A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and Arabidopsis plants. Mol Gen Genet 225: 459–467.

Bäumlein, H., Nagy, I., Villarroel, R., Inze, D., and Wobus, U. (1992) Cis-analysis of a seed protein gene promoter: The conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J 2: 233–239.

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

Beach, L.R., Spencer, D., Randall, P.J., and Higgins, T.J.V. (1985) Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. Nucleic Acids Res 13: 999–1013.

Becker, J.W., Reeke Jr., G.N., Wang, J.L., Cunningham, B.A., and Edelman, G.M. (1975) The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. J Biol Chem 250: 1513–1524.

Bednarek, S., and Raikel, N.V. (1992) Intracellular trafficking of secretory proteins. Plant Mol Biol 20: 133–150.

Belanger, F.C., and Kriz, A.L. (1989) Molecular characterization of the major maize embryo globulin encoded by the Glb1 gene. Plant Physiol 91: 636–643.

Benfey, P.N., and Chua, N.H. (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250: 959–966.

Benfey, P.N., and Chua, N.H. (1989) Regulated genes in transgenic plants. Science 244: 174–181.

Bild, G.S., Ramadoss, C.S., and Axelrod, B. (1977) Multiple dioxygenation by lipoxygenase of lipids containing all-cis–1,4,7-octatriene moieties. Arch Biochem Biophys 184: 36–41.

Bogue, M.A., vonder Haar, R.A., Nuccio, M.L., Griffing, L.R., and Thomas, T.L. (1990) Developmentally regulated expression of a sunflower 11S seed protein gene transgenic tobacco. Mol Gen Genet 222: 49–57.

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

Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990) Dynamic chromatin: The regulatory domain organization of eukaryotic gene loci. J Cell Biochem 47: 99–108.

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

Boulter, D. (1984) Cloning of pea storage protein genes [Vicilin and legumin, *Pisum sativum*]. Philos Trans R Soc Lond B 304: 323–332.

Boulter, D., and Derbyshire, E. (1971) Taxonomic aspects of the structure of legume proteins. In: Harborne, J.B., Boulter, D., and Turner, B.L. (eds) Chemotaxonomy of the Leguminosae, pp. 285–308, Academic Press, London.

Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R., and Burgess, J. (1986) Posttranslational processing of concanavalin A precursors in Jackbean cotyledons. J Cell Biol 102: 1284–1297.

Bown, D., Ellis, T.H.N., and Gatehouse, J.A. (1988) The sequence of a gene encoding convicilin from pea (*Pisum sativum* L.) shows that convicilin differs from vicilin by an insertion near the N- terminus. Biochem J 251: 717–726.

Boyington, J.C., Gaffney, B.J., and Amzel, L.M. (1993) The three-dimensional structure of an arachidonic acid 5-lipoxygenase. Science 260: 1482–1486.

Braun, H., Horstmann, C., and Bäumlein, H. (1995a) Legumins of the Gnetatae: Characterization and evolutionary relationships of the legumin encoding cDNA from Welwitschia, Gnetum and Ephedra. GenBank/EMBL/DDBJ Database Accession Number Z50780.

Braun, H., Horstmann, C., and Bäumlein, H. (1995b) A vicilin is the main seed storage protein from *Zamia furfuracea*. Homology to other vicilins and a sucrose binding protein from *Glycine max*. GenBank/EMBL/DDBJ Database Accession Number Z50791.

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

Brown, D., Levasseur, M., Croy, R.R.D., Boulter, D., and Gatehouse, J.A. (1985) Sequence of a pseudogene in the legumin family of pea (*Pisum sativum* L.). Nucleic Acids Res 13: 4527–4538.

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

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

Burrow, M.D., Ludden, P.W., and Bliss, F.A. (1993) Suppression of phaseolin and lectin in seeds of common bean, *Phaseolus vulgaris* L.: Increased accumulation of 54 kDa polypeptides is not associated with higher seed methionine concentrations. Mol Gen Genet 241: 431–439.

Bustos, M.M., Battraw, M.J., Kalkan, F.A., and Hall, T.C. (1991a) Transient gene expression in electroporated bean cotyledon protoplasts. Plant Mol Biol Rep ISPMB 9: 322–332.

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

Bustos, M.M., Guiltinan, M.J., Cyr, R.J., Ahdoot, D., and Fosket, D.E. (1989a) Light regulation of β-tubulin gene expression during internode development in soybean (*Glycine max* [L.] Merr.). Plant Physiol 91: 1157–1161.

Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan, F.A., and Hall, T.C. (1989b) Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich,

cis-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell 1: 839–853.
Bustos, M.M., Kalkan, F.A., VandenBosch, K.A., and Hall, T.C. (1991c) Differential accumulation of four phaseolin glycoforms in transgenic tobacco. Plant Mol Biol 16: 381–395.
Carrington, D.M., Auffret, A., and Hanke, D.E. (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. Nature 313: 64–67.
Casey, R. (1979) Genetic variability in the structure of the α-subunits of legumin from Pisum - a two dimensional gel electrophoretic study. Heredity 43: 265–272.
Casey, R.R., and Davies, D.R. (1993) Peas: Genetics, molecular biology and biotechnology. Vol. ix. Wallingford, Oxon, UK: CAB International.
Casey, R., Domoney, C., and Ellis, N. (1986) Legume storage proteins and their genes. Oxf Surv Plant Mol Cell Biol 3: 1–95.
Casey, R., Domoney, C., Ellis, N., and Turner, S. (1988) The structure, expression and arrangement of legumin genes in peas. Biochem Physiol Pflanz 183: 173–180.
Casey, R., Domoney, C., Ellis, T.H.N., and Castleton, J. (1985a) The isolation, analysis and expression of specific pea seed, storage protein genes. Adv Agric Biotechnol 104–105.
Casey, R., Domoney, C., and Nielsen, N.C. (1985b) Isolation of a cDNA clone for pea (*Pisum sativum*) seed lipoxygenase. Biochem J 232: 79–85.
Casey, R., Domoney, C., and Smith, A.M. (1993) Biochemistry and molecular biology of seed products. In: Casey, R. and Davies, D.R. (eds) Peas: Genetics, Molecular Biology and Biotechnology, pp. 121–163, CAB International, Oxon, UK.
Casey, R., Domoney, C., and Stanley, J. (1984) Convicilin mRNA from pea (*Pisum sativum* L.) has sequence homology with other legume 7S storage protein m RNA species. Biochem J 224: 661–666.
Ceriotti, A., Pedrazzini, E., Fabbrini, M.S., Zoppe, M., Bollini, R., and Vitale, A. (1991) Expression of wild-type and mutated vacuolar storage protein phaseolin in Xenopus oocytes reveals relationship between assembly and intra-cellular transport. Eur J Biochem 202: 959–968.
Ceriotti, A., Pedrazzini, E., Bielli, A., Giovinazzo, G., Bollini, R., and Vitale, A. (1995) Assembly and intracellular transport of Phaseolin, the major storage protein of *Phaseolus vulgaris* L. J Plant Physiol 145: 648–653.
Chandler, P.M., Higgins, T.J.V., Randall, P.J., and Spencer, D. (1983) Regulation of legumin levels in developing pea seeds under conditions of sulfur deficiency. Rates of legumin synthesis and levels of legumin mRNA *Pisum sativum*. Plant Physiol 71: 47–54.
Chandler, P.M., Spencer, D., Randall, P.J., and Higgins, T.J.V. (1984) Influence of sulfur nutrition on developmental patterns of some major pea seed proteins and their mRNAs. Plant Physiol 75: 651–657.
Chen, Z.-L., Naito, S., Nakamura, I., and Beachy, R.N. (1989) Regulated expression of genes encoding soybean β-conglycinin in transgenic plants. Dev Genet 10: 112–122.
Chen, Z.-L., Pan, N.-S., and Beachy, R.N. (1988) A DNA sequence element that confirs seed specific enhancement to a constitutive promoter. EMBO J 7: 297–302.
Chen, Z.-L., Schuler, M.A., and Beachy, R.N. (1986) Functional analysis of regulatory elements in a plant embryo specific gene. Proc Natl Acad Sci USA 83: 8560–8564.
Chlan, C.A., Pyle, J.B., Legocki, A.B., and Dure III, L. (1986) Developmental biochemistry of cottonseed embryogenesis and germination XVIII. cDNA and amino acid sequences of members of the storage protein families. Plant Mol Biol 7: 475–489.
Cho, T.J., Davies, C.S., Fischer, R.L., Turner, N.E., Goldberg, R.B., and Nielsen, N.C. (1989a) Molecular characterization of an aberrant allele for the Gy3 glycinin gene: A chromosomal rearrangement. Plant Cell 1: 339–350.
Cho, T.J., Davies, C.S., and Nielsen, N.C. (1989b) Inheritance and organization of glycinin genes in soybean. Plant Cell 1: 329–337.
Cho, T.J., and Nielsen, N.C. (1989) The glycinin Gy3 gene from soybean. Nucleic Acids Res 17: 4388.
Chrispeels, M.J. (1991a) Protein sorting in the secretory system of plant cells. Int Rev Cyt 125: 1–45.

Chrispeels, M.J. (1991b) Sorting of proteins in the secretory system. Annu Rev Plant Physiol Plant Mol Biol 42: 21–53.

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

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

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

Chrispeels, M.J., and Raikhel, N.V. (1992) Short peptide domains target proteins to plant vacuoles. Cell 68: 613–616.

Christopher, J.P. (1972) Isoenzymes of soybean lipoxygenase: Isolation and partial characterization. Doctoral dissertation, Purdue University, West Lafayette, IN 47907.

Christopher, J.P., Pistorius, E.K., and Axelrod, B. (1972a) Isolation of a third isoenzyme of soybean lipoxygenase. Biochim Biophys Acta 284: 54–62.

Christopher, J.P., Pistorius, E.K., Regnier, F.E., and Axelrod, B. (1972b) Factors influencing the positional specificity of soybean lipoxygenase. Biochim Biophys Acta 289: 82–87.

Christopher, J., and Axelrod, B. (1971) On the different positioned specificities of peroxidation of linoleate shown by two isozymes of soybean lipoxygenase. Biochem Biophys Res Commun 44: 731–736.

Christopher, J., Pistorius, E., and Axelrod, B. (1970) Isolation of an isozyme of soybean lipoxygenase. Biochim Biophys Acta 198: 12–19.

Coates, J.B., Medeiros, J.S., Thanh, V.H., and Nielsen, N.C. (1985) Characterization of the subunits of β-conglycinin. Arch Biochem Biophys 243: 184–194.

Colman, P.M., Suzuki, E., and van Donkelaar, A. (1980) Structure of cucurbitin: Subunit symmetry and organization in situ. Eur J Biochem 103: 585–588.

Coulson, A.F.W. (1994) A proposed structure for 'Family 18' chitinases: A possible function for narbonin. FEBS Lett 354: 41–44.

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

Craig, S., Goodchild, D.J., and Millerd, C. (1980a) Structural aspects of protein accumulation in developing pea (*Pisum sativum*) cotyledons: II. 3-dimensional reconstructions of vacuoles and protein bodies from serial sections. Aust J Plant Physiol 7: 329–338.

Craig, S., Millerd, A., and Goodchild, D.J. (1980b) Structural aspects of protein accumulation in developing pea cotyledons. III. Immunocytochemical localization of legumin and vicilin using antibodies shown to be specific by the enzyme linked immunosorbent assay. Aust J Plant Physiol 7: 339.

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

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

Croy, R.R.D., Hoque, M.S., Gatehouse, J.A., and Boulter, D. (1984) The major albumin proteins from pea (*Pisum sativum* L.). Purification and some properties. Biochem J 218: 795–803.

Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N., and Boulter, D. (1983) Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295: 76–79.

Dalgarrondo, M., Raymond, J., and Azanza, J.-L. (1984) Sunflower seed proteins: Characterization and subunit composition of the globulin fraction. J Exp Bot 35: 1618–1628.

Danielsson, C.E. (1949) Seed globulins of the Gramineae and Leguminoseae. Biochem J 44: 387–400.

Davies, C.S., Coates, J.B., and Nielsen, N.C. (1985) Inheritance and biochemical analysis of four electrophoretic variants of β-conglycinin from soybean. Theor Appl Genet 71: 351–358.

Davies, C.S., and Nielsen, N.C. (1986) Genetic analysis of a null-allele for lipoxygenase–2 in soybean. Crop Sci 26: 460–463.
Davies, C.S., Nielsen, S.S., and Nielsen, N.C. (1987) Flavor improvement of soybean preparations by genetic removal of lipoxygenase–2. J Am Oil Chem Soc 64: 1428–1433.
Davies, D.R. (1980) The ra-locus and legumin synthesis in *Pisum sativum*. Biochem Genet 18: 1207–1219.
de Castro, L.A.B., Lacerda, Z., Aramayo, R.A., Sampaio, M.J.A.M., and Gander, E.S. (1987) Evidence for a precursor molecule of Brazil nut 2S seed proteins from biosynthesis and cDNA analysis. Mol Gen Genet 206: 338–343.
de Clercq, A., Vandewiele, M., de Rycke, R., Van Damme, J., Van Montagu, M., Krebbers, E., and Vandekerckhove, J. (1990) Expression and processing of an Arabidopsis 2S albumin in transgenic tobacco. Plant Physiol 92: 899–907.
Denecke, J., Goldman, M.H.S., Demolder, J., Seurinch, J., and Botterman, J. (1991) The tobacco luminal protein is encoded by a small multigene family. Plant Cell 3: 1025–1035.
Depigny-This, D., Raynal, M., Aspart, L., Delseny, M., and Grellet, F. (1992) The cruciferin gene family in radish. Plant Mol Biol 20: 467–479.
Derbyshire, E., Wright, D.J., and Boulter, D. (1976) Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 15: 3–24.
Dickinson, C.D., Evans, R.P., and Nielsen, N.C. (1988) RY repeats are conserved in the 5′-flanking regions of legume and seed-protein genes. Nucleic Acids Res 16: 371.
Dickinson, C.D., Floener, L.A., Lilley, G.G., and Nielsen, N.C. (1987) Self-assembly of proglycinin and hybrid proglycinin synthesized in vitro from cDNA. Proc Natl Acad Sci USA 84: 5525–5529.
Dickinson, C.D., Hussein, E.H.A., and Nielsen, N.C. (1989) Role of posttranslational cleavage in glycinin assembly. Plant Cell 1: 459–469.
Diers, B.W., Beilinson, V., Nielsen, N.C., and Shoemaker, R.C. (1994) Genetic mapping of the Gy4 and Gy5 glycinin genes in soybean and the analysis of a variant of Gy4. Theor Appl Genet 89: 297–304.
Domoney, C., Barker, D., and Casey, R. (1986a) The complete deduced amino acid sequences of legumin β-polypeptides from different genetic loci in Pisum. Plant Mol Biol 7: 467–474.
Domoney, C., and Casey, R. (1983) Cloning and characterization of complementary DNA for convicilin, a major seed storage protein in *Pisum sativum* L. [Peas]. Planta 159: 446–453.
Domoney, C., and Casey, R. (1985) Measurement of gene number for seed storage proteins in Pisum. Nucleic Acids Res 13: 687–699.
Domoney, C., and Casey, R. (1984) Storage protein precursor polypeptides in cotyledons of *Pisum sativum* L. Identification of, and isolation of a cDNA clone for, an 80000-M_r legumin-related polypeptide [Peas]. Eur J Biochem 139: 321–327.
Domoney, C., Casey, R., Turner, L., and Ellis, N. (1991) Pisum lipoxygenase genes. Theor Appl Genet 81: 800–805.
Domoney, C., Ellis, T.H.N., and Davies, D.R. (1986b) Organization and mapping of legumin genes in Pisum. Mol Gen Genet 202: 280–285.
Domoney, C., Firmin, J.L., Sidebottom, C., Ealing, P.M., Slabas, A., and Casey, R. (1990) Lipoxygenase heterogeneity in *Pisum sativum*. Planta 181: 35–43.
Doyle, J.J., Schuler, M.A., Godette, W.D., Zenger, V., Beachy, R.N., and Slightom, J.L. (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. J Biol Chem 261: 9228–9238.
Duranti, M., Gatehouse, J.A., Boulter, D., and Cerletti, P. (1987) In vitro proteolytic processing of pea and jack bean storage proteins by an endopeptidase from lupin seeds. Phytochemistry 26: 627–631.
Duranti, M., Guerrieri, N., Cerletti, P., and Vecchio, G. (1992) The legumin from white lapin seed. Identity of the subunit, assembly and proteolysis. Eur J Biochem 206: 941–947.
Dure, L., and Chlan, C.A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination. Plant Physiol 68: 180–186.

Ealing, P.M., and Casey, R. (1989) The cDNA cloning of a pea (*Pisum sativum*) seed lipoxygenase. Sequence comparisons of the two major pea seed lipoxygenase isoforms. Biochem J 264: 929–932.

Ealing, P.M., and Casey, R. (1988) The complete amino acid sequence of a pea (*Pisum sativum*) seed lipoxygenase predicted from a near full-length cDNA. Biochem J 253: 915–918.

Ellis, T.H.N., Domoney, C., Castleton, J., Cleary, W., and Davis, D.R. (1986) Vicilin genes of Pisum. Mol Gen Genet 205: 164–169.

Ergland, D.R., Brown, J.W.S., Casey, R., and Hall, T.C. (1983) The storage proteins of *Phaseolus vulgaris* L., *Vicia faba* L. and *Pisum sativum* L. In: Gottschalk, W., and Muller, H.P. (eds) Seed Proteins: Biochemistry, Genetics, Nutritive Value, pp. 355–375, W. Junk Publisher, Boston.

Ericson, M.L., Muren, E., Gustavsson, H.O., Josefsson, L.G., and Rask, L. (1991) Analysis of the promoter region of napin genes from *Brassica napus* demonstrates binding of nuclear protein in vitro to a conserved sequence motif. Eur J Biochem 197: 741–746.

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

Eskin, N.A.M., Grossman, S., and Pinsky, A. (1977) Biochemistry of lipoxygenase in relation to food quality. Crit Rev in Food Sci Nutr 9: 1–40.

Evans, D.E., Nyquist, W.E., Santini, J.B., Bretting, P., and Nielsen, N.C. (1994) Immunological identification of seed lipoxygenase genotypes in soybean. Crop Sci 34: 1529–1537.

Fabijanski, S., Chang, S.-C., Dukiandjiev, S., Bahramian, M.B., Ferrara, P., and Altosaar, I. (1988) The nucleotide sequence of a cDNA for a major prolamin (avenin) in oat (*Avena sativa* L. cultivar Hinoat) which reveals homology with oat globulin. Biochem Physiol Pflanzen 183: 143–152.

Farber, G.K. (1993) An alpha/beta-barrel full of evolutionary trouble. Curr Opin Struct Biol 3: 409–412.

Fiedler, U., Filistein, R., Wobus, U., and Bäumlein, H. (1993) A complex ensemble of cis-regulatory elements controls the expression of a *Vicia faba* non-storage seed protein gene. Plant Mol Biol 22: 669–679.

Fischer, H. (1994) Direct submission. GenBank/EMBL/DDBJ Database Accession Number X82464.

Fischer, H. (1994) Direct submission. GenBank/EMBL/DDBJ Database Accession Number X82465.

Fischer, R.L., and Goldberg, R.B. (1982) Structure and flanking regions of soybean seed protein genes. Cell 29: 651–660.

Frazier, P.J. (1979) Lipoxygenase action and lipid binding in breadmaking. Bakers Digest 53: 8–29.

Frisch, D.A., van der Geest, A.H.M., Dias, K., and Hall, T.C. (1995) Chromosomal integration is required for spatial regulation of expression from the β-phaseolin promoter. Plant J 7: 503–512.

Fromenthal, C., Kanno, M., Nomiyama, H., and Chambon, P. (1988) Cooperativity and hierarchical levels of functional organization in the SV40 enhancer. Cell 54: 943–953.

Fuchs, J., Joos, S., Lichter, P., and Schubert, I. (1994) Localization of vicilin genes on field bean chromosome II by fluorescent in situ hybridization. J Hered 85: 487–488.

Fuchs, J., and Schubert, I. (1995) In situ localization of seed protein genes on field bean chromosomes. Chromosome Res 3: 94–100.

Fujiwara, T., and Beachy, R.N. (1994) Tissue-specific and temporal regulation of a β-conglycinin gene: roles of the RY repeat and other cis-acting elements. Plant Mol Biol 24: 261–272.

Fukazawa, C., Momma, T., Hirano, H., Harada, K., and Udaka, K. (1985) Glycinin A3B4 mRNA. Cloning and sequencing of double-stranded cDNA complementary to a soybean storage protein. J Biol Chem 260: 6234–6239.

Galili, G., Altschuler, Y., Levanony, H., Giorini-Silfen, S., Shimoni, Y., Shani, N., and Karchi, H. (1995) Assembly and transport of wheat storage proteins. J Plant Physiol 145: 626–631.

Galliard, T., and Chan, H.W.S. (1980) Lipoxygenases. In: Stumpf, P.K. and Conn, E.E. (eds) The Biochemistry of Plants, Vol. 4, pp. 132–162, Academic Press, New York.

Gander, E.S., Holmstroem, K.-O., De Paiva, G.R., De Castro, L.A.B., Carneiro, M., and Grossi de Sá, M.-F. (1991) Isolation, characterization and expression of a gene coding for a 2S albumin from *Bertholletia excelsa* (Brazil nut). Plant Mol Biol 16: 437–448.

Garcia-Maroto F., Marana, C., Mena, M., Garcia-Olmedo, F., and Carbonero, P. (1990) Cloning of complementary DNA and chromosomal location of genes encoding the three types of subunits of the wheat tetrameric inhibitor of insect α-amylase. Plant Mol Biol 14: 845–854.

Garssen, G.J., Vliegenthart, J.F.G., and Boldingh, J. (1971) An anaerobic reaction between lipoxygenase, linoleic acid and its hydroperoxides. Biochem J 122: 327–332.

Gatehouse, J.A., Bown, D., Gilroy, J., Levasseur, M., Castleton, J., and Ellis, T.H.N. (1988) Two genes encoding minor legumin polypeptides in pea (*Pisum sativum*). Characterization and complete sequence of the LegJ gene. Biochem J 250: 15–24.

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

Gatehouse, J.A., Evans, I.M., Croy, R.R.D., and Boulter, D. (1986) Differential expression of genes during legume seed development. Philos Trans R Soc Lond B 314: 367–384.

Gatehouse, J.A., Lycett, G.W., Delauney, A.J., Croy, R.R.D., and Boulter, D. (1983) Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.). Biochem J 212: 427–432.

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

Gayler, K.R., Kolivas, S., Macfarlane, A.J., Lilley, G.G., Baldi, M., Blagrove, R.J., and Johnson, E.D. (1990) Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin delta, a sulphur-rich protein from *Lupinus angustifolius*. Plant Mol Biol 15: 879–893.

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

Goldberg, R.B. (1986) Regulation of plant gene expression. Phil Trans Roy Soc Lond B 314: 343–353.

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

Goldberg, R.B., De Paiva, G., and Yadegari, R. (1994) Plant embryogenesis: zygote to seed. Science 266: 605–614.

Goldberg, R.B., and Perez-Grau, L. (1989) Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell 1: 1095–1109.

Grimes, H.D., Overvoorde, P.J., Ripp, K., Franceschi, V.R., and Hitz, W.D. (1992) A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. Plant Cell 4: 1561–1574.

Häger, K.P., Braun, H., Czihal, A., Müller, B., and Bäumlein, H. (1995) Evolution of seed storage protein genes: Legumin genes of *Ginkgo biloba*. J Mol Evol, in press.

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

Halling, K.C., Halling, A.C., Murray, E.E., Ladin, B.F., Houston, L.L., and Weaver, R.F. (1985) Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucleic Acids Res 13: 8019–8033.

Hara, I., Wada, K., Wakabayshi, S., and Matsubara, H. (1976) Pumpkin (Cucubita sp.) seed globulin I. Purification, characterization and subunit structure. Plant Cell 17: 799–814.

Hara-Nishimura, I., and Nishimura, M. (1987) Proglobulin processing enzymes in vacuoles isolated from developing pumpkin cotyledons. Plant Physiol 85: 440–445.

Hara-Nishimura, I., Nishimura, M., and Akazawa, T. (1985) Biosynthesis and intracellular transport of 11S globulin in developing pumpkin cotyledons. Plant Physiol 77: 747–752.

Hara-Nishimura, I., Shimada, T., Hiraiwa, N., and Nishimura, M. (1995) Vacuolar processing enzyme responsible for maturation of seed proteins. J Plant Physiol 145: 632–640.
Harada, J.J., Barker, S.J., and Goldberg, R.B. (1989) Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell 1: 415–425.
Harley, S.M., and Beevers, L. (1989) Coated vesicles are involved in the transport of storage proteins during seed development in *Pisum sativum*. Plant Physiol 91: 674–678.
Harris, N., and Croy, R.R.D. (1985) The major albumin protein from *pea (Pisum sativum* L.). Localisation by immunocytochemistry. Planta 165: 522–526.
Hayashi, M., Mori, H., Nishimura, M., Akazawa, T., and Hara-Nishimura, I. (1988) Nucleotide sequence of cloned cDNA for pumpkin 11S globulin β subunit. Eur J Biochem 172: 627–632.
Heath, J.D., Weldon, R., Monnot, C., and Meinke, D.W. (1986) Analysis of storage proteins in normal and aborted seeds from embryo-lethal mutants of *Arabidopsis thaliana*. Planta 169: 304–312.
Heim, U., Bäumlein, H., and Wobus, U. (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.
Heim, U., Schubert, R., Bäumlein, H., and Wobus, U. (1989) The legumin gene family: structure and evolutionary implications of *Vicia faba* B-type genes and pseudogenes. Plant Mol Biol 13: 653–663.
Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Höhne, W.E., and Wilson, K.S. (1992) A TIM barrel protein without enzymatic activity? Crystal structure of narbonin at 1.8Å resolution. FEBS Lett 306: 80–84.
Herman, E.M., Tague, B.W., Hoffman, L.M., Kjemtrap, S.E., and Chrispeels, M.J. (1990) Retention of phytohaemagglutinin with carboxyterminal tetrapeptide KDEL in the nuclear envelope and the endoplasmic reticulum. Planta 182: 305–312.
Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244.
Higgins, D.G., and Sharp, P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appl Biosci 5: 151–153.
Higgins, T.J.V. (1984) Synthesis and regulation of major proteins in seeds. Annu Rev Plant Physiol 35: 191–221.
Higgins, T.J.V., Beach, L.R., Spencer, D., Chandler, P.M., Randall, R.J., Blagrove, R.J., Kortt, A.P., and Guthrie, R.E. (1987) cDNA and protein sequence of a major pea seed albumin (PA 2: m_r approx 26,000). Plant Mol Biol 8: 37–45.
Higgins, T.J.V., Chandler, P.M., Randall, P.J., Spencer, D., Beach, L.R., Blagrove, R.J., Kortt, A.A., and Inglis, A.S. (1986) Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. J Biol Chem 261: 11124–11130.
Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C., and Spencer, D. (1983a) The biosynthesis and primary structure of pea seed lectin. J Biol Chem 258: 9544–9549.
Higgins, T.J.V., Chrispeels, M.J., Chandler, P.M., and Spencer, D. (1983b) Intracellular sites of synthesis and processing of lectin in developing pea cotyledons *Pisum sativum*. J Biol Chem 258: 9550–9552.
Higgins, T.J.V., Newbigin, E.J., Spencer, D., Llewellyn, D.J., and Craig, S. (1988) The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. Plant Mol Biol 11: 683–695.
Higuchi, W., and Fukazawa, C. (1987) A rice glutelin and a soybean glycinin have evolved from a common ancestral gene. Gene 55: 245–253.
Hildebrand, D.F., and Hymowitz, T. (1982) Inheritance of lipoxygenase–1 activity in soybean seeds. Crop Sci 22: 851–853.
Hildebrand, D.F., and Kito, M. (1984) Role of lipoxygenases in soybean seed protein quality. J Agric Food Chem 32: 815–819.

Hirano, H., Fukazawa, C., and Harada, K. (1984) The complete amino acid sequence of the A3 subunit of the glycinin seed storage protein of the soybean (*Glycine max* (L.) Merrill). J Biol Chem 259: 14371–14377.
Hirano, H., Fukazawa, C., and Harada, K. (1985) The primary structures of the A4 and A5 subunits are highly homologous to that of the A3 subunit in the glycinin seed storage protein of soybean. FEBS Lett 181: 124–128.
Hirano, H., Kagawa, H., and Okubo, K. (1992) Characterization of proteins released from legume seeds in hot water. Phytochemistry 31: 731–735.
Hirano, H., Gatehouse, J.A., and Boulter, D. (1982) The complete amino acid sequence of a subunit of the vicilin seed storage protein of pea (*Pisum sativum* L.). FEBS Lett 145: 99–102.
Hoedemaeker, F.J., Richardson, M., Diaz, C.L., de Pater, B.S., and Kijne, J.W. (1994) Pea (*Pisum sativum* L.) seed isolectins 1 and 2 and pea root lectin result from carboxypeptidase-like processing of a single gene product. Plant Mol Biol 24: 75–81.
Hoffman, L.M., Donaldson, D.D., Bookland, R., Rashka, K., and Herman, E.M. (1987) Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seeds. EMBO J 6: 3213–3221.
Hofte, H., and Chrispeels, M.J. (1992) Protein sorting to the vacuolar membrane. Plant Cell 4: 995–1004.
Holwerda, B.C., Galvin, N.J., Baranski, T.J., and Rogers, J.C. (1990) In vitro processing of aleurain, a barley vacuolar thiol protease. Plant Cell 2: 1091–1106.
Hopp, T.P., Hemperly, J.J., and Cunningham, B.A. (1982) Amino acid sequence and variant forms of favin, a lectin from *Vicia faba*. J Biol Chem 257: 4473–4483.
Horstmann, C. (1983) Specific subunit pairs of legumin from *Vicia faba* [Broadbean seeds]. Phytochemistry 22: 1861–1866.
Horstmann, C., Schlesier, B., Otto, A., Kostka, S., and Müntz, K. (1993) Polymorphism of legumin subunits from field beans (*Vicia faba* L. var. minor) and its relation to the corresponding multigene family. Theor Appl Genet 86: 867–874.
Hu, B., and Esen, A. (1982) Heterogeneity of soybean proteins: two-dimensional electrophoretic maps of three solubility fractions. J Agric Food Chem 30: 21–25.
Hung, C.H., Lee, M.C., Lee, T.C., and Lin, J.Y. (1993) Primary structure of three distinct isoabrins determined by cDNA sequencing. Conservation and significance. J Mol Biol 229: 263–267.
Hurkman, W.J., Lane, B.G., and Tanaka, C.K. (1994) Nucleotide sequence of a transcript encoding a germin-like protein that is present in salt-stressed barley (*Hordeum vulgare* L.) roots. Plant Physiol 104: 803–804.
Irwin, S.D., and Lord, J.M. (1990) Nucleotide sequence of a *Ricinus communis* 2S albumin precursor gene. Nucleic Acids Res 18: 5890.
Irwin, S.D., Keen, J.N., Findlay, J.B.C., and Lord, J.M. (1990) The *Ricinus communis* 2S albumin precursor: A single preprotein may be processed into two different heterodimeric storage proteins. Mol Gen Genet 222: 400–408.
Ishii, S.I. (1994) Legumain: asparaginyl endopeptidase. Meth Enzymol 244: 604–615.
Itoh, Y., Kitamura, Y., and Fukazawa, C. (1994) The glycinin box: a soybean embryo factor binding motif within the quantitative regulatory region of the 11S seed storage globulin promoter. Mol Gen Genet 243: 353–357.
Jensen, U., and Berthold, H. (1989) Legumin-like proteins in gymnosperms. Phytochemistry 28: 1389–1394.
Jordano, J., Almoguera, C., and Thomas, T.L. (1989) A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. Plant Cell 1: 855–866.
Josefsson, L.-G., Lenman, M., Ericson, M.L., and Rask, L. (1987) Structure of a gene encoding the 1.7 S storage protein, Napin, from *Brassica napus*. J Biol Chem 262: 12196–12201.
Jung, R., Scott, M.P., and Nielsen, N.C. (1991) The sequence specificity of the maturation protease responsible for the post-translational processing of the 11S seed storage protein

from soybean. Molecular Biology of Plant Growth and Development. Proceedings of the ISPMB 3rd International Congress, p. 763.

Kagawa, H., and Hirano, H. (1989) Sequence of a cDNA encoding soybean basic 7S globulin. Nucleic Acids Res 17: 8868.

Kanofsky, J.R., and Axelrod, B. (1986) Singlet oxygen production by soybean lipoxygenase isozymes. J Biol Chem 21: 1099–1104.

Kato, T., Ohta, H., Tanaka, K., and Shibata, D. (1992) Appearance of new lipoxygenases in soybean cotyledons after germination and evidence for expression of a major new lipoxygenase gene. Plant Physiol 98: 324–330.

Kawagoe, Y., and Murai, N. (1992) Four distinct nuclear proteins recognize in vitro the proximal promoter of the bean seed storage protein β-phaseolin gene conferring spatial and temporal control. Plant J 2: 927–936.

Khan, R.I., Gatehouse, J.A., and Boulter, D. (1980) The seed proteins of cowpea (*Vigna unguiculata* L. walp.). J Exp Bot 31: 1599–1611.

Kim, C.H., and Choi, Y.D. (1989) Molecular cloning of a cDNA encoding the precursor to the glycinin A2B1a subunit of soybean. Korean Biochem J 22: 233–241.

Kitamura, K. (1993) Breeding trials for improving the food processing quality of soybeans. Trends Food Sci Technol 4: 64–67.

Kitamura, K., Davies, C.S., Kaizuma, N., and Nielsen, N.C. (1983) Genetic analysis of a null-allele for lipoxygenase–3 in soybean seeds. Crop Sci 23: 924–927.

Kitamura, K., Davies, C.S., and Nielsen, N.C. (1984) Inheritance of alleles for Gy1 and Gy4 storage protein genes in soybean. Theor Appl Genet 68: 253–257.

Kitamura, K., Kimagai, T., and Kikuchi, A. (1985) Inheritance of lipoxygenase–2 and genetic relationships among genes for lipoxygenase–1, –2, and –3 isozymes in soybean seeds. Japanese J Breeding 35: 413–420.

Kitamura, Y., Arahira, M., and Itoh, Y. (1990) The complete nucleotide sequence of soybean glycinin A2B1a gene spanning to another glycinin gene A1aB1b. Nucleic Acids Res 18: 4245.

Ko, T.-P., Ng, J.D., Day, J., Greenwood, A., and McPherson, A. (1993a) Determination of three crystal structures of canavalin by molecular replacement. Acta Crystallogr, D 49: 478–489.

Ko, T. P., Ng, J.D., and McPherson, A. (1993b) The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). Plant Physiol 101: 729–744.

Kolivas, S., and Gayler, K.R. (1993) Structure of the cDNA coding for conglutin gamma, a sulphur-rich protein from *Lupinus angustifolius*. Plant Mol Biol 21: 397–401.

Kornfeld, S., and Mellman, I. (1989) The biogenesis of liposomes. Annu Rev Cell Biol 5: 483–525.

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

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

Kriz, A.L. (1989) Characterization of embryo globulins encoded by the main Glb genes. Biochem Genet 27: 239–251.

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

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

Lane, B.G., Bernier, F., Dratewka Kos, E., Shafai, R., Kennedy, T.D., Pyne, C., Munro, J.R., Vaughan, T., Walters, D., and Altomare, F. (1991) Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain Physarum spherulins. J Biol Chem 266: 10461–10469.

Lane, B.G., Dunwell, J.M., Ray, J.A., Schmitt, M.R., and Cuming, A.C. (1993) Germin, a protein marker of early plant development, is an oxalate oxidase. J Biol Chem 268: 12239–12242.

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

Laroche, M., Aspart, L., Delseny, M., and Penon, P. (1984) Characterization of radish (*Raphanus sativus*) storage proteins. Plant Physiol 74: 487–493.

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

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

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

Lelievre, J.M., Dickinson, C.D., Dickinson, L.A., and Nielsen, N.C. (1992a) Synthesis and assembly of soybean β-conglycinin in vitro. Plant Mol Biol 18: 259–274.

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

Lessard, P.A., Allen, R.D., Fujiwara, T., and Beachy, R.N. (1993) Upstream regulatory sequences from two β-conglycinin genes. Plant Mol Biol 22: 873–885.

Lilley, G.G., and Inglis, A.S. (1986) Amino acid sequence of conglutin δ, a sulfur-rich seed protein of *Lupinus angustifolius* L. Sequence homology with the C-III α-amylase inhibitor from wheat. FEBS Lett 195.

Liu, X., Maeda, S., Hu, Z., Aiuchi, T., Nakaya, K., and Kurihara, Y. (1993) Purification, complete amino acid sequence and structural characterization of the heat-stable sweet protein, mabinlin II. Eur J Biochem 211: 281–287.

Lõnnerdahl, B., and Janson, J.-C. (1972) Studies on Brassica seed proteins. 1. The low molecular weight proteins in rapeseed. Isolation and characterization. Biochim Biophys Acta 278: 175–183.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., and Wyns, L. (1993) Crystal structure determination and refinement at 2.3Å resolution of the lentil lectin. Biochemistry 32: 8772–8781.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, D., and Boulter, D. (1985) The 5′-flanking regions of three pea legumin genes: Comparison of the DNA sequences. Nucleic Acids Res 13: 6733–6743.

Lycett, G.W., Delauney, A.J., Zhao, W.M., Gatehouse, J.A., and Croy, R.R.D. (1984) Two cDNA clones coding for the legumin protein of *Pisum sativum* L. contain sequence repeats [Peas, storage proteins]. Plant Mol Biol 3: 91–96.

Macas, J., Weschke, W., Bäumlein, H., Pich, U., Houben, A., Wobus, U., and Schubert, I. (1993) Localization of vicilin genes via polymerase chain reaction on microisolated field bean chromosomes. Plant J 3: 883–886.

Mahmoud, S.H., and Gatehouse, J.A. (1984) Inheritance and mapping of vicilin storage protein genes in *Pisum sativum* L. Heredity 53: 185–191.

Manuelidis, L., and Borden, J. (1988) Reproducible compartmentalization of individual chromosome domains in human CNA cells revealed by in situ hybridization and three-dimensional reconstruction. Chromosome 96: 397–410.

March, J.F., Pappin, D.J.C., and Casey, R. (1988) Isolation and characterization of a minor legumin and its constituent polypeptides from *Pisum sativum* (pea). Biochem J 250: 911–915.

Marco, Y.A., Thanh, V.H., Tumer, N.E., Scallon, B.J., and Nielsen, N.C. (1984) Cloning and structural analysis of DNA encoding an A2B1a subunit of glycinin. J Biol Chem 259: 13436–13441.

Masumura, T., Kidzu, K., Sugiyama, Y., Mitsukawa, N., Hibino, T., Tanaka, K., and Fujii, S. (1989) Nucleotide sequence of a cDNA encoding a major rice glutelin. Plant Mol Biol 12: 723–725.

Matsuoka, M., and Nakamura, K. (1991) Propeptide of a precursor to a plant vacuolar protein required for vacuolar targetting. Plant Cell 2: 941–950.

Matta, N.K., and Gatehouse, J.A. (1982) Inheritance and mapping of storage proteins in *Pisum sativum* L. Heredity 48: 383–392.

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

Matta, N.K., Gatehouse, J.A., and Boulter, D. (1981b) The structure of legumin of *Vicia faba* - A reappraisal. J Exp Bot 32: 183–197.

Medeiros, J.S. (1982) Characterization of the subunits of β-conglycinin, and application of the enzyme-linked immunosorbent assay (ELISA) to the determination of the contents of β-conglycinin and glycinin in soybean (*Glycine max*) seeds. Ph.D. thesis, Purdue University, West Lafayette, IN 47907.

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

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

Min, W., and Jones, D.H. (1994) In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. Struct Biol 1: 502–504.

Minor, W., Steczko, J., Bolin, J.T., Otwinowski, Z., and Axelrod, B. (1993) Crystallographic determination of the active site iron and its ligands in soybean lipoxygenase L–1. Biochemistry 32: 6320–6323.

Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K., and Fukazawa, C. (1985a) Glycinin A5A4B3 mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean. Eur J Biochem 149: 491–496.

Momma, T., Negoro, T., Udaka, K., and Fukazawa, C. (1985b) A complete cDNA coding for the sequence of glycinin A2B1a subunit precursor. FEBS Lett 188: 117–122.

Monsalve, R.I., Gonzalez de la Pena, M.A., Menendez-Arias, L., Lopez-Otin, C., Villalba, M., and Rodriguez, R. (1993) Characterization of a new oriental mustard (*Brassica juncea*) allergen, Bra j IE: Detection of an allergenic epitope. Biochem J 293: 625–632.

Moreira, M.A., Hermodson, M.A., Larkins, B.A., and Nielsen, N.C. (1981) Comparison of the primary structure of the acidic polypeptides of glycinin. Arch Biochem Biophys 210: 633–642.

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

Morrison, R., DeLozier, G., Robinson, L., and McPherson, A. (1984) Biochemical and x-ray diffraction analysis of concanavalin B crystals from Jack bean. Plant Physiol 76: 175–183.

Müntz, K., Jung, R., and Saalbach, G. (1993) Synthesis, processing and targetting of legume seed proteins. In: Shewry, P.R. and Stobart, K. (eds) Seed Storage Compounds, Biosynthesis, Interactions and Manipulation, pp. 128–146, Clarendon Press, Oxford.

Murai, N., Sutton, D.W., Murray, M.G., Slightom, J.L., Merlo, D.J., Reichert, N.A., Sengupta-Gopalan, C., Stock, C.A., Barker, R.F., Kemp, J.D., and Hall, T.C. (1983) Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors. Science 222: 476–482.

Muramatsu, M., and Fukazawa, C. (1993) A high-order structure of plant storage proprotein allows its second conversion by an asparagine-specific cysteine protease, a novel proteolytic enzyme. Eur J Biochem 215: 123–132.

Mustakas, L.C., Albrecht, W.J., McGlee, J.E., Black, L.T., Bookwalter, G.N., and Griffin, J.J. (1969) Lipoxidase deactivation to improve stability, odor and flavor of full-fat soy flours. J Am Oil Chem Soc 46: 623–626.

Nam, Y.-W. (1994) In vitro synthesis and assembly of legume seed storage globulins. Doctoral dissertation, Purdue University, West Lafayette, IN 47907.

Negoro, T., Momma, T., and Fukazawa, C. (1985) A cDNA clone encoding a glycinin A1a subunit precursor of soybean. Nucleic Acids Res 13: 6719–6731.
Newton, C.H. (1991a) Direct submission. GenBank/EMBL/DDBJ Database Accession Number X63192.
Newton, C.H. (1991b) P. glauca mRNA for 2S-like storage protein. GenBank/EMBL/DDBJ Database Accession Number X63193.
Newton, C.H., Flinn, B.S., and Sutton, B.C. (1992) Vicilin-like seed storage proteins in the gymnosperm interior spruce (*Picea glauca*/engelmanii). Plant Mol Biol 20: 315–322.
Ng, J.D., Ko, T.P., and McPherson, A. (1993) Cloning, expression, and crystallization of jack bean (*Canavalia ensiformis*) canavalin. Plant Physiol 101: 713–728.
Nielsen, N.C. (1984) The chemistry of legume storage proteins [Glycinin subunits from soybeans]. Philos Trans R Soc Lond B 304: 287–296.
Nielsen, N.C. (1989) Soybean Gy1 gene for glycinin subunit G1. NIH Gene Database Accession Number X15121.
Nielsen, N.C. (1995) Soybean seed composition. In: Verma, D.P.S. and Shoemaker, R. (eds) Soybean: Genetics, Molecular Biology and Biotechnology, CAB International, Wallingford, UK, in press.
Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., and Goldberg, R.B. (1989) Characterization of the glycinin gene family in soybean. Plant Cell 1: 313–328.
Nielsen, N.C., Jung, R., Nam, Y.-W., Beaman, T.W., Oliveira, L.O., and Bassüner, R. (1995) Synthesis and assembly of 11S globulins. J Plant Physiol 145: 641–647.
Nielsen, N.C., Scott, M.P., and Lago, W.J.P. (1990) Assembly properties of modified subunits in the glycinin subunit family. In: Hermann, R., and Larkins, B.A. (eds) NATO Advanced Study Institute on Plant Mol Biol, Schloss Elmau, Germany, pp. 635–640, Plenum Press, New York.
Nirasawa, S., Nishino, T., Katahira, M., Uesugi, S., Hu, Z., and Kurihara, Y. (1994) Structure of heat-stable and unstable homologues of the sweet protein mabinlin. The difference in the heat stability is due to replacement of a single amino acid. Eur J Biochem 223: 989–995.
Nong, V., Becker, C., and Müntz, K. (1994b) Cloning and heterologous expression of cDNAs encoding legumins of vetch (*Vicia sativa*) seeds. GenBank/EMBL/DDBJ Database Accession Number Z32835.
Nong, V., Becker, C., and Müntz, K. (1994a) Cloning and heterologous expression of cDNAs encoding legumins of vetch (*Vicia sativa*) seeds. GenBank/EMBL/DDBJ Database Accession Number Z32796.
Nong, V., Becker, C., and Müntz, K. (1994c) PCR cloning of legumin cDNA from *Vicia narbonensis*. GenBank/EMBL/DDBJ Database Accession Number Z46803.
Nong, V., and Müntz, K. (1994a) A genomic sequence encoding putative narbonin from *Vicia sativa*. GenBank/EMBL/DDJB Database Accession Number Z46835.
Nong, V., and Müntz, K. (1994b) The narbonin gene from soybean. GenBank/EMBL/DDJB Database Accession Number Z46825.
Nong, V., Schlesier, B., Bassüner, R., Horstmann, C., Kraft, R., and Müntz, K. (1994d) Cloning and characterization of the gene encoding a narbonin-like seed storage 2S globulin from *Vicia pannonica*. GenBank/EMBL/DDJB Database Accession Number Z25534.
Nong, V., Schlesier, B., and Müntz, K. (1994e) A cDNA encoding putative narbonin from *Canavalia ensiformis*. GenBank/EMBL/DDJB Database Accession Number Z46802.
Nong, V., Schlesier, B., and Müntz, K. (1994f) The narbonin gene from *Vicia faba* L. GenBank/EMBL/DDJB Database Accession Number Z46827.
Nong, V.H., Schlesier, B., Bassüner, R., Repik, A., Horstmann, C., and Müntz, K. (1995) Narbonin, a novel 2S protein from *Vicia narbonensis* L. seeds: cDNA, gene structure and developmentally regulated formation. Plant Mol Biol, 28: 61–72.
Nordlee, J.A., Taylor, S.L., Townsend, J.A., and Thomas, L.A. (1994) High methionine Brazil nut protein binds human IGE. J Allergy Clin Immunol 93: 209, part 2, abstract #277.
North, H., Casey, R., and Domoney, C. (1989) Inheritance and mapping of seed lipoxygenase polypeptides in Pisum. Theor Appl Genet 77: 805–808.

Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R., and Krishnan, H.B. (1989) Structure and expression of the rice glutelin multigene family. J Biol Chem 264: 12573–12581.

Oliveira, L.O. (1994) Studies on modulation of the expression of gy2 glycinin gene and on assembly properties of modified β-conglycinin subunits. Ph.D. Thesis, Purdue University, West Lafayette, IN.

Osborne, T.B. (1924), The Vegetable Proteins (2nd ed.). Longman, Green and Co, New York.

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

Park, T.K., Holland, M.A., Laskey, J.G., and Polacco, J.C. (1994) Germination-associated lipoxygenase transcripts persist in maturing soybean plants and are induced by jasmonate. Plant Sci 95: 109–117.

Patel, S., Cudney, R., and McPherson, A. (1994) Crystallographic characterization and molecular symmetry of edestin, a legumin from hemp. J Mol Biol 235: 361–363.

Paul, W., Hodge, R., Smartt, S., Draper, J., and Scott, R. (1992) The isolation and characterisation of the tapetum-specific *Arabidopsis thaliana* A9 gene. Plant Mol Biol 19: 611–622.

Pederson, T.J., Arwood, L.J., Spiker, S., Guiltinan, M.J., and Thompson, W.F. (1991) High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. Plant Mol Biol 16: 95–104.

Pistorius, E.K. (1974) Studies on isoenzymes of soybean lipoxygenase. Doctoral dissertation, Purdue University, West Lafayette, IN 47907.

Pistorius, E K., Axelrod, B., and Palmer, G. (1976) Evidence for participation of iron in lipoxygenase reaction from optical and electron spin resonance studies. J Biol Chem 251: 7144–7148.

Plietz, P., Damaschun, G., Muller, J.J., and Schwenke, K.-D. (1983) The structure of 11S globulins from sunflower and rapeseed. A small-angle x-ray scattering study. Eur J Biochem 130: 315–320.

Plietz, P., Drescher, B., and Damaschun, G. (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.

Pysh, L.D., Aukerman, M.J., and Schmidt, R.J. (1993) OHPI: A maize basic domain/leucine zipper protein that interacts with Opaque–2. Plant Cell 5: 227–236.

Rahman, S., Jolly, C.J., Skerritt, J.H., and Wallosheck, A. (1994) Cloning of a wheat 15-kDa grain softness protein (GSP). GSP is a mixture of puroindoline-like polypeptides. Eur J. Biochem 223: 917–925.

Raikhel, N.V., and Lerner, D.R. (1991) Expression and regulation of lectin genes in cereals and rice. Dev Genet 12: 255–260.

Rao, R., Costa, A., Croy, R.R.D., Boulter, D., and Gatehouse, J.A. (1989) Variation in polypeptides of the major albumin protein of pea (*Pisum sativum* L.): Inheritance and molecular analysis. Mol Gen Genet 219: 277–281.

Reeke Jr., G.N., Becker, J.W., and Edelman, G.M. (1975) The covalent and three-dimensional structure of concanavalin A. IV: Atomic coordinates, hydrogen bonding, and quaternary structure. J Biol Chem 250: 1525–1547.

Rerie, W.G., Newbigin, E.J., and Higgins, T.J.V. (1992) Genes encoding seed globulins in legumes. Adv Plant Cell Biochem Biotechnol 1: 53–104.

Rerie, W.G., Whitecross, M.I., and Higgins, T.J.V. (1990) Nucleotide sequence of one A-type legumin gene from pea. Nucleic Acids Res 18: 655.

Rerie, W.G., Whitecross, M., and Higgins, T.J.V. (1991) Developmental and environmental regulation of pea legumin genes in transgenic tobacco. Mol Gen Genet 225: 148–157.

Richardson, M., Rouge, P., Sousa-Cavada, B., and Yarwood, A. (1984) The amino acid sequences of the $\alpha 1$ and $\alpha 2$ subunits of the isolectins from seeds of *Lathyrus ochrus* (L) DC. FEBS Lett 175: 76–81.

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

Robinson, D.G., Balusek, K., and Freundt, H. (1989) Legumin antibodies recognize polypeptides in coated vesicles isolated from developing pea cotyledons. Protoplasma 150: 79–82.

Robinson, D.G., Hoh, B., Hinz, G., and Jeong, B.-K. (1995) One vacuole or two vacuoles: Do protein storage vacuoles arise de novo during pea cotyledon development? J Plant Physiol 145: 654–664.

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

Rodriguez-Palenzuela P., Royo, J., Gomez, L., Sanchez-Monge R., Salcedo, G., Molina, C.J.L., Garcia-Olmedo F., and Carbonero, P. (1989) The gene for trypsin inhibitor cme is regulated in trans by the lys 3a locus in the endosperm of barley (*Hordeum vulgare* L.). Mol Gen Genet 219: 474–479.

Roedin, J., and Rask, L. (1990) Characterization of matteuccin, the 2.2S storage proteins of the ostrich fern. Evolutionary relationship to angiosperm seed storage proteins. Eur J Biochem 192: 101–107.

Romero, J., Sun, S.M., McLeester, R.C., Bliss, F.A., and Hall, T.C. (1975) Heritable variation in a polypeptide subunit of the major storage protein of the bean, *Phaseolus vulgaris* L. Plant Physiol 56: 776–779.

Rugh, C.L., and Kamalay, J.C. (1991) Legumin mRNAs from *Pinus strobus* L. GenBank/EMBL/DDBJ Database Accession Number Z11486 .

Ryan, A.J., Royal, C.L., Hutchinson, J., and Shaw, C.H. (1989) Genomic sequence of a 12S seed storage protein from oilseed rape (*Brassica napus* c.v. jet neuf). Nucleic Acids Res 17: 3584.

Saalbach, G., Jung, R., Kunze, G., Saalbach, I., Adler, K., and Müntz, K. (1991) Different legumin protein domains act as vacuolar targeting signals. Plant Cell 7: 695–708.

Saalbach, G., Jung, R., Saalbach, I., and Müntz, K. (1988) Construction of storage protein genes with increased number of methionine codons and their use in transformation experiments. Biochem Physiol Pflanz 183: 211–218.

Saez-Vasquez, J. (1995) Differential expression of germin-like proteins in *A. thaliana*. GenBank/EMBL/DDBJ Database Accession Number U21743.

Sammour, R.H., Gatehouse, J.A., Gilroy, J., and Boulter, D. (1984) The homology of the major storage protein of jack bean (*Canavalia ensiformis*) to pea vicilin and its separation from α-mannosidase [*Pisum sativum*]. Planta 161: 61–70.

Sano, M., and Kawashima, N. (1983) Isolation and partial characterization of the major seed protein from *Nicotiana tabacum*, and accumulation during development. Agric Biol Chem 47: 1305–1310.

Satoh, S., Sturm, A., Fujii, T., and Chrispeels, M.J. (1992) cDNA cloning of an extracellular dermal glycoprotein of carrot and its expression in response to wounding. Planta 188: 432–438.

Scallon, B.J., Dickinson, C.D., and Nielsen, N.C. (1987) Characterization of a null-allele for the Gy4 glycinin gene from soybean. Mol Gen Genet 208: 107–113.

Scallon, B., Thanh, V.H., Floener, L.A., and Nielsen, N.C. (1985) Identification and characterization of DNA clones encoding group-II glycinin subunits. Theor Appl Genet 70: 510–519.

Schlesier, B., Bassüner, R., Van Hai, N., and Müntz, K. (1990) The cDNA derived primary structure of two distinct legumin A subunit precursors from field bean (*Vicia faba* L.). Nucleic Acids Res 18: 7146.

Schlesier, B., Manteuffel, R., Rudolph, A., and Behlke, J. (1978) Studies on seed globulins from legumes. VII. Narbonin, a 2S globulin from *Vicia narbonensis* L. Narlonne vetch. Biochem Physiol Pflanzen 173: 420–428.

Schlesier, B., Nong, V., Horstmann, C., and Hennig, M. (1995) Sequence analysis of concanavalin B from *Canavalia ensiformis* reveals homology to chitinases. GenBank/EMBL/DDBJ Database Accession Number X83426.

Schlesier, B., and Scholz, G. (1974) Studies on seed globulins from legumes. II. A crystalline protein from the globulin fraction of *Vicia narbonensis* L. Biochem Physiol Pflanzen 166: 367–369.

Schmidt, R.J., Ketudat, M., Aukerman, M.J., and Hoschek, G. (1992) Opaque–2 is a transcriptional activator that recognizes a specific target site in 22-KD zein genes. Plant Cell 4: 689–700.

Scholz, G., Manteuffel, R., Müntz, K., and Rudolph, A. (1983) Low molecular-weight polypeptides of vicilin from *Vicia faba* and products of proteolytic breakdown. Eur J Biochem 132: 103–107.

Schroeder, H.E. (1984) Major albumins of Pisum cotyledons. J Sci Food Agric 35: 191–198.

Schuler, M.A., Doyle, J.J., and Beachy, R.N. (1983) Nucleotide homologies between the glycosylated seed storage proteins of *G. max* and *P. vulgaris*. Plant Mol Biol 2: 119–127.

Schuler, M.A., Schmitt, E.S., and Beachy, R.N. (1982) Closely related families of genes code for the α and α' subunits of the soybean 7S storage protein complex. Nucleic Acids Res 10: 8225–8243.

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

Scott, M.P., Jung, R., Müntz, K., and Nielsen, N.C. (1992) A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. Proc Natl Acad Sci USA 89: 658–662.

Sebastiani, F.L., Farrell, L.B., Schuler, M.A., and Beachy, R.N. (1990) Complete sequence of a cDNA of α-subunit of soybean β-conglycinins. Plant Mol Biol 15: 197–201.

Sengupta, C., Deluca, V., Bailey, D.S., and Verma, D.P.S. (1981) Post-translational processing of 7S and 11S components of soybean storage proteins. Plant Mol Biol 1: 19–34.

Sengupta-Gopalan, C., Richert, N.A., Barker, R.F., Hall, T.C., and Kemp, J.D. (1985) Developmentally regulated expression of the bean β-phaseolin gene in tobacco seed. Proc Natl Acad Sci USA 82: 3320–3324.

Setsuda, J.E., Larson, S., Greenwood, A., Day, J., & McPherson, A. (1994) The cloning and structural refinement of concanavalin B from the Jack bean plant, *Canavalia ensiformis*. Abstracts of the 4th International Congress of Plant Mol Biol (Abstract #1190). Intl Soc Plant Mol Biol.

Sharon, N., and Lis, H. (1989) Lectins. London: Chapman and Hall.

Sharon, N., and Lis, H. (1990) Legume lectins - a large family of homologous proteins. FASEB J 4: 3198–3208.

Shewry, P.R., Sayanova, O., Tatham, A.S., Tamas, L., Turner, M., Richard, G., Hickman, D., Fido, R., Halford, N.G., Greenfield, J., Grimwade, B., Thomson, N., Miles, M., Freedman, R., and Napier, J. (1995) Structure, assembly and targeting of wheat storage proteins. J Plant Physiol 145: 620–625.

Shibata, D., Steczko, J., Dixon, J. E., Andrews, P.C., Hermodson, M., and Axelrod, B. (1988) Primary structure of soybean lipoxygenase L–2. J Biol Chem 263: 6816–6821.

Shibata, D., Steczko, J., Dixon, J.E., Hermodson, M., Yazdanparast, R., and Axelrod, B. (1987) Primary structure of soybean lipoxygenase–1. J Biol Chem 262: 10080–10085.

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

Shorrosh, B.S., Wen, L., Zen, K.C., Huang, J.K., Pan, J.S., Hermodson, M.A., Tanaka, K., Muthukrishnan, S., and Reeck, G.R. (1992) A novel cereal storage protein: Molecular genetics of the 19 kDa globulin of rice. Plant Mol Biol 18: 151–154.

Shutov, A.D., Kakhovskaya, I.A., Braun, H., Bäumlein, H., and Müntz, K. (1995) Legumin and vicilin-like seed storage proteins: Evidence for a common single-domain ancestral gene. J Mol Evol, in press.

Shutov, A.D., and Vaintraub, I. (1987) Degradation of storage proteins in germinating seeds. Phytochem 26: 1557–1566.
Sjodahl, S., Rodin, J., and Rask, L. (1991) Characterization of the 12S globulin complex of *Brassica napus*. Evolutionary relationship to other 11–12S storage globulins. Eur J Biochem 196: 617–621.
Slightom, J.C., Sun, S.M., and Hall, T.C. (1983) Complete nucleotide sequence of a French bean storage protein gene: Phaseolin. Proc Natl Acad Sci USA 80: 1897–1901.
Spencer, D., Chandler, P.M., Higgins, T.J.V., Inglis, A.S., and Rubira, M. (1983) Sequence interrelationships of the subunits of vicilin from pea seeds [*Pisum sativum*]. Plant Mol Biol 2: 259–267.
Staswick, P.E., Hermodson, M.A., and Nielsen, N.C. (1984a) The amino acid sequence of the A2B1a subunit of glycinin. J Biol Chem 259: 13424–13430.
Staswick, P.E., Hermodson, M.A., and Nielsen, N.C. (1981) Identification of the acidic and basic subunit complexes of glycinin. J Biol Chem 256: 8752–8755.
Staswick, P.E., Hermodson, M.A., and Nielsen, N.C. (1984b) Identification of the cystines which link the acidic and basic components of the glycinin subunits. J Biol Chem 259: 13431–13435.
Staswick, P.E., and Nielsen, N.C. (1983) Characterization of a soybean cultivar lacking certain glycinin subunits. Arch Biochem Biophys 223: 1–8.
Sturm, A., Johnson, K.D., Szumilo, T., Elbein, A.D., and Chrispeels, M.J. (1987) Subcellular localization of glycosidases and glycosyl transferases involved in the processing of n-linked oligosaccharides. Plant Physiol 85: 741–745.
Sun, S.M., McLeester, R.C., Bliss, F.A., and Hall, T.C. (1974) Reversible and irreversible dissociation of globulins from *Phaseolus vulgaris* seed. J Biol Chem 249: 2118–2121.
Sun, S.M., Slightom, J.L., and Hall, T.C. (1981) Intervening sequences in a plant gene-comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature 289: 37–41.
Sun, S.S.M., Altenbach, S.B., and Leung, F.W. (1987) Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.). Eur J Biochem 162: 477–483.
Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S., and Ikenaka, T. (1993) Crystallographic refinement of Bowman-Birk type protease inhibitor A-II from peanut (*Arachis hypogaea*) at 2.3 Å resolution. J Mol Biol 234: 722–734.
Svendsen, I.B., Nicolova, D., Goshev, I., and Genov, N. (1994) Primary structure, spectroscopic and inhibiting properties of a two-chain trypsin inhibitor from seeds of charlock (*Sinapis arvensis* L.), a member of the napin protein family. Int J Pept Protein Res 43: 425–430.
Sykes, G.E., and Gayler, K.R. (1981) Detection and characterization of a new β-conglycinin from soybean seeds. Arch Biochem Biophys 210: 525–530.
Tague, B.W., Dickinson, C.D., and Chrispeels, M.J. (1990) A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. Plant Cell 2: 533–546.
Takaiwa, F., Kikuchi, S., and Oono, K. (1987) A rice glutelin gene family - A major type of glutelin mRNAs can be divided into two classes. Mol Gen Genet 208: 15–22.
Takamura, H., Kitamura, K., and Kito, M. (1991) Inhibition by lipoxygenase–3 of n-hexanal generation in soybeans. FEBS Lett 292: 42–44.
Tanchak, M.A., and Chrispeels, M.J. (1989) Crosslinking of microsomal proteins identifies P–9000, a protein that is co-transported with phaseolin and phytohaemagglutinin in bean cotyledons. Planta 179: 279–287.
Tanchak, M.A., Giband, M., Potier, B., Schernthaner, J.P., Dukiandjiev, S., and Altosaar, I. (1995) Genomic clones encoding 11S globulins in oats (*Avena sativa* L.). Genome, in press.
Thanh, V.H., and Shibasaki, K. (1976) Heterogeneity of β-conglycinin. Biochim Biophys Acta 439: 326–338.
Thanh, V.H., and Shibasaki, K. (1978a) Major proteins of soybean seeds. Reconstitution of β-conglycinin from its subunits. J Agric Food Chem 26: 695–698.

Thanh, V.H., and Shibasaki, K. (1978b) Major proteins of soybean seeds. Subunit structure of β-conglycinin. J Agric Food Chem 26: 692–695.
Thomas, T.L. (1993) Gene expression during plant embryogenesis and germination: an overview. Plant Cell 5: 1401–1410.
Thomas, T.L., Vivekananda, J., and Bogue, M.A. (1991) ABA regulation of gene expression in embryos and mature plants. In: Davies, W.J. and Jones, H.G. (eds) Abscisic Acid: Physiology and Biochemistry, pp. 125–135, Bios Scientific Publishers, Oxford.
Thompson, A.J., Brown, D., Yaish, S., and Gatehouse, J.A. (1991) Differential expression of seed storage protein genes in the pea legJ subfamily; sequence of gene legK. Biochemie Und Physiologie Der Pflanzen 187: 1–12.
Thompson, J.A., and Schroeder, H.E. (1978) Cotyledonary storage proteins in *Pisum sativum*. II. Hereditary variation in components of the legumin and vicilin fractions. Aust J Plant Physiol 5: 281–294.
Thompson, J.A., Schroeder, H.E., and Tassie, A.M. (1980) Cotyledonary storage proteins in *Pisum sativum*. II. Further studies on molecular heterogeneity in the vicilin series of holoproteins. Aust J Plant Physiol 7: 271–282.
Turner, L., Hellens, R.P., Lee, D., and Ellis, T.H.N. (1993) Genetic aspects of the organization of legumin genes in pea. Plant Mol Biol 22: 101–112.
Turner, S.R., Barratt, D.H.P., and Casey, R. (1990) The effect of different alleles at the r-locus on the synthesis of seed storage proteins in *Pisum sativum*. Plant Mol Biol 14: 793–803.
Utsumi, S., Gidamis, A.B., Mikami, B., and Kito, M. (1993) Crystallization and prelimary x-ray crystallographic analysis of the soybean proglycinin expressed in *Escherichia coli*. J Mol Biol 233: 177–178.
Valls, L.A., Winther, J.R., and Stevens, T.H. (1990) Yeast carboxypeptidase Y vacuolar targeting signal is defined by four propeptide amino acids. J Cell Biol 111: 361–368.
van der Geest, A.H.M., Hall, G.E. Jr., Spiker, S., and Hall, T.C. (1994) The β-phaseolin gene is flanked by matrix attachment regions. Plant J 6: 413–423.
Vellanoweth, R.L., and Okita, T.W. (1993) Regulation of expression of wheat and rice seed storage protein genes. In: Verma, D.P.S. (ed) Control of Plant Gene Expression, pp. 377–392, CRC Press, Boca Raton, FL.
Vernooy-Gerritsen, M., Bos, A.L.M., Veldink, G.A., and Vliegenthart, J.F.G. (1983) Localization of lipoxygenase –1 and –2 in germinating soybean seeds by an indirect immunofluorescence technique. Plant Physiol 73: 262–267.
Vitale, A., and Bollini, R. (1995) Legume storage proteins. In: Kigel, J. and Galili, G. (eds) Seed Development and Germination, pp. 73–102, Marcel Dekker, New York.
Vliegenthart, J.F.G., and Veldink, G.A. (1982) Lipoxygenases. In: Pryor, W.A (ed) Free Radicals in Biology Vol. V, pp. 29–64, Academic Press, New York.
vonder Haar, R.A., Allen, R.D., Cohen, E.A., Nessler, C.L., and Thomas, T.L. (1988) Organization of the sunflower 11S storage protein gene family. Gene 74: 433–443.
Wallace, N.H., and Kriz, A.L. (1991) Nucleotide sequence of a cDNA clone corresponding to the maize globulin–2 gene. Plant Physiol 95: 973–975.
Wandelt, C.I., Rafiqul, M., Khan, I., Craig, S., Schroeder, H.E., Spencer, D., and Higgins, T.J.V. (1992) Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenics plants. Plant J 2: 181–192.
Wang, C.S., Shastri, K., Wen, L., Huang, J.K., Sonthayanon, B., Muthukrishnan, S., and Reeck, G.R. (1987) Heterogeneity in cDNA clones encoding rice glutelin. FEBS Lett 222: 135–138.
Wang, W.H., Takano, T., Shibata, D., Kitamura, K., and Takeda, G. (1994) Molecular basis of a null mutation in soybean lipoxygenase–2: Substitution of glutamine for an iron-ligand histidine. Proc Natl Acad Sci USA 91: 5828–5832.
Wen, L., Huang, J.K., Johnson, B.H., and Reeck, G.R. (1989) Nucleotide sequence of a cDNA that encodes a rice glutelin. Nucleic Acids Res 17: 9490.
Weschke, W., Bassüner, R., Hai, N.V., Czihal, A., Bäumlein, H., and Wobus, U. (1988) The structure of a *Vicia faba* vicilin gene. Biochem Physiol Pflanz 183: 233–242.

Wobus, U., Bäumlein, H., Bassüner, R., Grafe, R., Jung, R., Müntz, K., Saalbach, G., and Weschke, W. (1984) Cloning and characterizing *Vicia faba* seed storage protein genes. Kulturpflanze (Bln.) 32: 117–120.

Wobus, U., Bäumlein, H., Bassüner, R., Heim, U., Jung, R., Müntz, K., Saalbach, G., and Weschke, W. (1986) Characteristics of two types of legumin genes in the field bean (*Vicia faba* L. var. minor) genome as revealed by cDNA analysis. FEBS Lett 201: 74–80.

Wobus, U., Borisjuk, L., Panitz, R., Manteuffel, R., Bäumlein, H., Wohlfahrt, T., Heim, U., Weber, H., Miséra, S., and Weschke, W. (1995) Control of seed storage protein gene expression: New aspects on an old problem. J Plant Physiol 145: 592–599.

Wolf, J.W. (1975) Lipoxygenase and flavor of soybean protein products. J Agric Food Chem 23: 136–141.

Wolf, W.J., and Briggs, D.R. (1958) Studies on the cold-insoluble fraction of the water-extractable soybean proteins. ii. Factors influencing conformational changes in the 11S component. Arch Biochem Biophys 76: 377–393.

Wood, K.A., Lord, J.M., Wawrzynczak, E.J., and Piatak, M. (1991) Preproabrin: genomic cloning, characterisation and the expression of the A-chain in *Escherichia coli*. Eur J Biochem 198: 723–732.

Wright, D.J. (1988) The seed globulins. II. Dev Food Proteins 6: 119–178.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1993) The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. Mol Gen Genet 238: 17–25.

Yamauchi, F., Sato, K., and Yamagishi, T. (1984) Isolation and partial characterization of a salt-extractable globulin from soybean seeds [*Glycine max*]. Agric Biol Chem 48: 645–650.

Yamauchi, F., Sato, M., Sato, W., Kamata, Y., and Shibasaki, K. (1981) Isolation and identification of a new type of β-conglycinin in soybean globulins. Agric Biol Chem 45: 2863–2868.

Yarwood, A., Richardson, M., Sousa-Cavada, B., and Rouge, P. (1985) The complete amino acid sequences of β1- and β2- subunits of the isolectins LoL1 and Lol1 from seeds of *Lathyrus ochrus* (L.) DC. FEBS Lett 184: 104–109.

Yenofsky, R.L., Fine, M., and Liu, C. (1988) Isolation and characterization of a soybean (*Glycine max*) lipoxygenase–3 gene. Mol Gen Genet 211: 215–222.

Youle, R.J., and Huang, A.C. (1981) Occurrence of a low molecular weight and high cysteine containing albumin storage proteins in oil-seeds of diverse species. Am J Bot 68: 44–48.

Youle, R.J., and Huang, A.H.C. (1979) Albumin storage proteins and allergens in cottonseed. J Agric Food Chem 27: 500–503.

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

Youle, R.J., and Huang, A.H.C. (1978b) Identification of the castor bean allergens as the albumin storage proteins in the protein bodies of castor bean. Plant Physiol 61: 1040–1042.

zurNieden, U., Manteuffel, R., Weber, E., and Newman, D. (1984) Dictyosomes participate in the intercellular pathway of storage proteins in developing *Vicia faba* cotyledons. Eur J Cell Biol 34: 9–17.