

Cloning and structural analysis of a cDNA clone encoding glycinin (*Gly-1*) seed storage protein of peanut

Ashok K. Jain[#]

Plant Biotechnology Lab.
College of Engineering Sciences, Technology and Agriculture
Florida A&M University
Tallahassee, FL 32307, USA
E-mail: ashok.jain@fam.edu

Financial support: The research was supported by United States Department of Agriculture (USDA), Capacity Building Grant # 98-388 14-6944.

Keywords: arachin, *Arachis hypogaea*, cupin superfamily, glycinin, legumins, peanut, seed storage protein.

[#]**Present address:** Environmental Biotechnology Lab, Room #123, Dyson Bldg, College of Pharmacy and Pharmaceutical Sciences, Tallahassee, Florida 32307, Phone: 850 561 2788, Fax: 850 599 3731.

A cDNA clone (peanut *Gly-1*) encoding for glycinin protein was isolated from the immature seeds (from yellow-1 maturity pods) and characterized. The clone spanning a total of 1836 bp, predicted protein of 529 amino acid residues with a calculated mass of 60,447.61 Da. Peanut *Gly-1* sequence comparison shows high level of sequence homology with other two peanut glycinin (arachin) genes [*Ara h3* (95%) and *Ara h4* (94%)] and glycinin (legumin) genes of other legumes such as soybean, broad bean, French bean and pea etc., both at nucleotide (67 to 69%) and amino acid (60 to 63%) levels. The N- and C-terminals of peanut *Gly-1* are highly conserved with other glycinin genes; central region of the gene possess three variable regions, which also show conservation with other glycinin genes. Peanut glycinin-1 gene deciphers 11S type A seed storage protein. Mapping for conserved domains indicate that Peanut *Gly-1* consists of bi-cupin domain. RNA gel blot studies demonstrate that the gene expressed during embryo development that is transcriptionally activated early in embryogenesis (white pod maturity) and is repressed late in seed maturation (orange pod maturity stage). Peanut *Gly-1* does not express in other tissues like leaf, stem, root, flower, pegs or post germinating seedlings.

(Shewry and Casey, 1999).

Peanut (*Arachis hypogaea* L.), a legume with 24% protein (Cherry, 1977), is a major source of plant protein in most tropical and subtropical regions of the world. In the United States, peanuts are mostly grown in the southern states and they have significant economic impact. Peanuts are the third most important source of plant protein and provide approximately 11% of the world's protein supply. The globulin fraction of peanut seed consists of arachin (glycinin or legumin) and con-arachin (vicilin or conglycinin) (Johnson et al. 1950; Cherry et al. 1973). Of the globulins, glycinin (arachin) accounts for more than 50% of the total protein (Jones and Horn, 1930). Glycinin is superior to con-glycinin from the viewpoint of nutritional value (Millerd, 1975) as well as functional properties (Kinsella, 1978). Immunological studies also demonstrate that glycinin type seed storage peanut proteins (such as *Ara h3* and *Ara h4*) are less severe compared to their vicilin (*Ara h1*) and conglutin (*Ara h2*) type seed storage proteins (Burks et al. 1995; Shin et al. 1998; Viquez et al. 2003; Koppelman et al. 2004). It is obvious, that it is important to understand the biosynthesis, targeting and biological functions of seed proteins as a prerequisite to their rational manipulation in improving nutritional value (Shewry and Casey, 1999). However, such information is still in paucity for seed storage proteins of peanut. Therefore, molecular studies were initiated to isolate, clone and characterized peanut glycinin gene(s) and/or other important peanut seed storage protein gene/s. To this end, degenerate oligonucleotides were designed based on the conserved region of known glycinin sequences from other legumes and to amplify peanut glycinin sequences. Amplified DNA fragment was homologous to glycinin protein of soybean. Full-length cDNA clone was cloned, which encodes for glycinin (11S type A) seed storage protein. This paper

Seed proteins are an extremely and increasingly important component of nutrition for both humans and animals (Shewry and Casey, 1999). Plant proteins are the cheapest available sources of protein in many countries where few can afford meat or dairy products, hence are an important part of human diet. Supply of proteins from plant is considered an ideal source of protein to humans without the concern of cholesterol and legumes have been well documented for their high seed storage protein contents

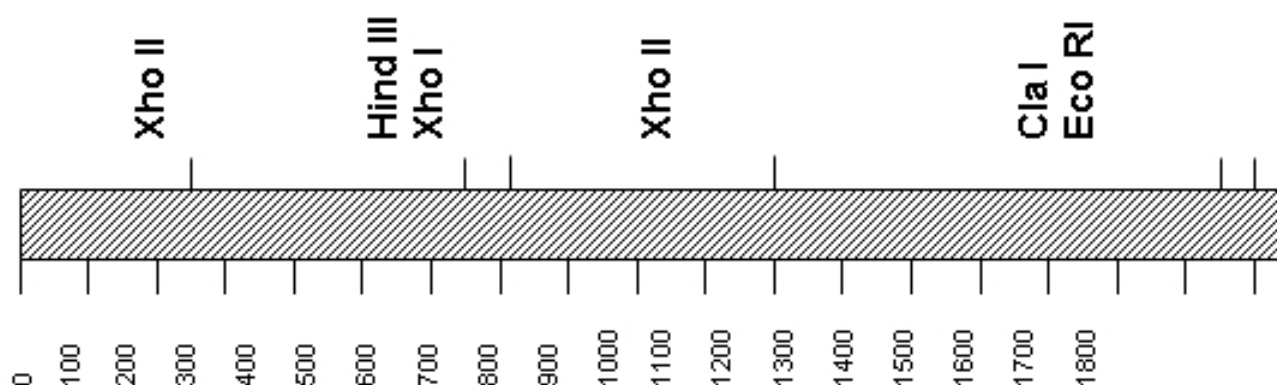


Figure 1. Restriction map of peanut (*Arachis hypogaea* L.) cDNA clone (peanut *Gly-1*) encoding glycinin gene (11S arachin type A seed storage protein) representing unique restriction sites.

describes and discusses the gene structure of peanut glycinin clone (*Gly-1*), homology with other seed storage proteins, its expression and discuss about the potential function during seed germination.

MATERIAL AND METHODS

Materials

Peanuts (*Arachis hypogaea* L. cv. Florunner) were grown in experimental plots at the University of Florida, Experimental Station, Quincy, FL following the recommended cultural practices. Randomly selected plants were dug at weekly intervals for 9 weeks between 90 and 160 days after planting; pods were harvested, washed with tap water and separated into six maturity categories (white, yellow-1, yellow-2, orange, brown and black) following the Hull-scrap method of Williams and Drexler (1981). The pods were subsequently washed with distilled water, surface sterilized with 2% NaOCl (without any wetting agent) for 5 min, thoroughly washed (six times) with sterile water and excess water was removed with dry paper towels. Pods were shelled, seed coats removed and frozen immediately in liquid nitrogen and stored at -80°C until used.

Tissues such as fully developed leaf, stem, root, flowers and pegs were collected between 75-85 days after germination. For post-germinating seedlings, seeds were sterilized first by washing with detergent (2% Tween 20) for 5 min; detergent was removed by washing the seeds 5-6 times with distilled water. Later seeds were treated with Clorox (0.15%) for 5 min and washed with 7-8 times with autoclaved distilled water. Sterilized seeds were germinated in lab (at 25°C) in Petri dishes on moist filter paper. Seedlings at 0, 2, 4, 6, 8, 10, 12, and 14 days after imbibitions were sampled, frozen immediately in liquid nitrogen and stored at -80°C until used.

Nucleic acid isolation

Total RNA was isolated from the seeds of different maturity pods (white to black maturity category), leaf, stem, root, flowers, pegs and post-germinating seedlings of 0 to 14 days. Tissues were homogenized in 4 M guanidinium thiocyanate, 25 mM sodium citrate (pH 7.0), 0.5% sarcosyl, 0.1 M β -mercaptoethanol, and then extracted with an equal volume of phenol: chloroform (1:1). To the aqueous phase equal volume of isopropanol was added and chilled on ice for an hour. Total RNA was recovered by centrifugation (5000 g) and was further purified at least twice using LiCl (final concentration 2 M) precipitation and dissolved in sterile distilled water. Finally, the total RNA was precipitated with 2.5 volumes of 95% EtOH in the presence of 0.1% 3M NaOAc (pH 7.0) and stored at -80°C until used for northern blot hybridization studies.

PCR amplification

Two forward (*GlyUP.1* coding for DRHQKIYNFRE; and *GlyUP.2* coding for ENEGEDKAI) and one reverse (*GlyDN.1* coding for FVPHYNLNA) degenerate oligonucleotide primers were designed and synthesized (Genosys, Biotechnologies, Inc., Texas) based on the conserved region among soybean (*Glycine max*) (Nielsen et al. 1989), and broad bean (*Vicia faba*) (Panitz et al. 1997) glycinin protein sequences. PCR amplification was performed in 50 μl reaction volume with 1 μM for each of the forward and reverse primers, 1 μl of peanut seed (yellow-1 pod maturity) cDNA library (titer 1.2×10^9 pfu per ml), 5 mM dNTP's, 1X PCR buffer [10 mM Tris HCl (pH 9.0), 50 mM KCl, 25 mM MgCl_2 and 0.1% Triton X-100] and 10 U of Taq DNA Polymerase (Promega). The template was denatured by heating to 94°C for 3 min, followed by 30 cycles of 30 seconds at 94°C , followed by combinations of annealing temperatures for 1 min (37°C or 45°C) and polymerization for 1 min 30 seconds (at 68°C or 72°C), with a single final extension of 5 min at 72°C . Total of eight combinations (including primer combination's, annealing and polymerization temperature) were used to amplify DNA sequences for peanut glycinin clone. The

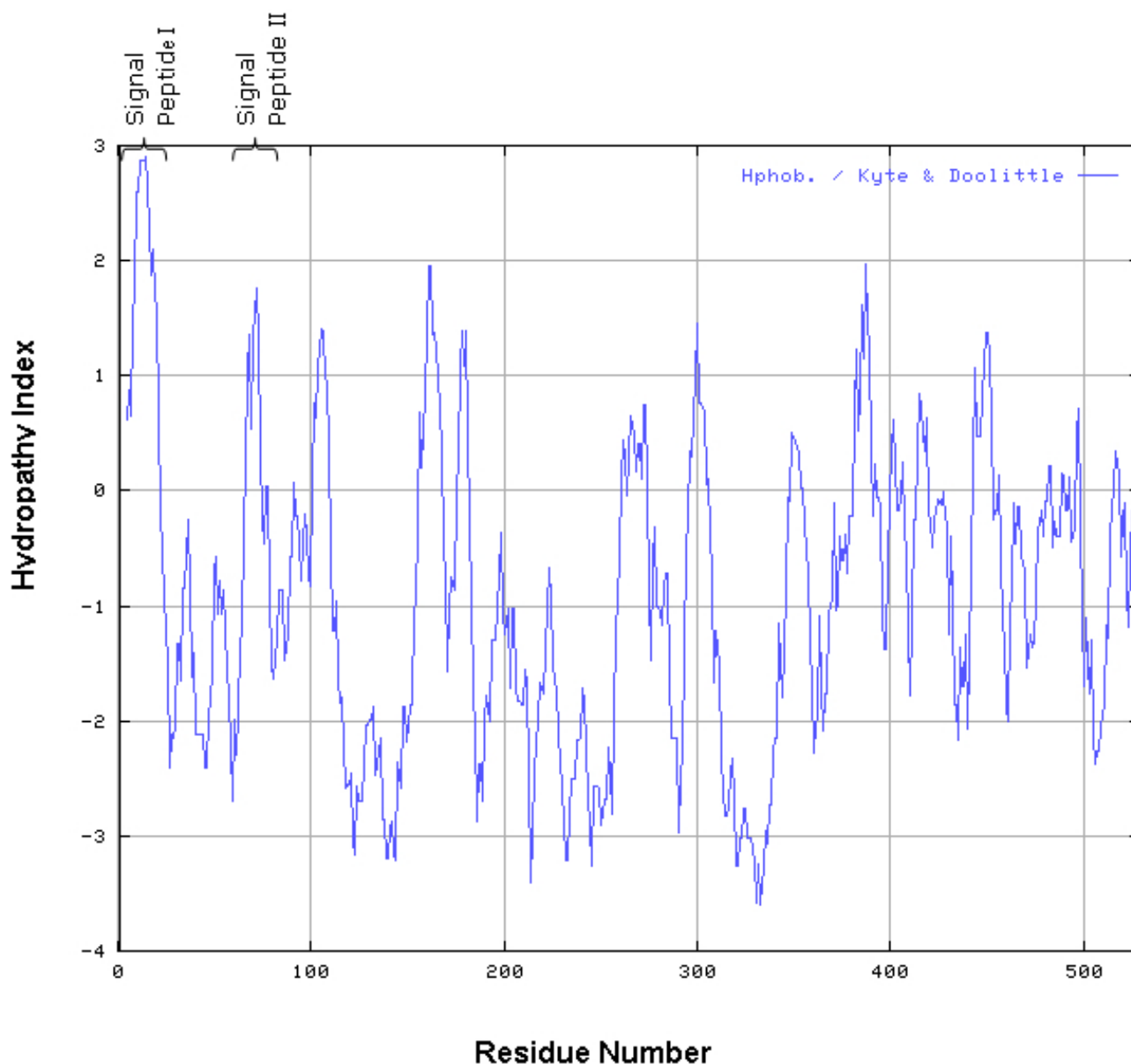


Figure 2. Hydropathy plot of peanut *Gly-1* (Kyte Doolittle, 1982). Potential transmembrane domain is labeled (Signal peptide I and II).

PCR amplified product (25 μ l) was separated by electrophoresis on a 1.0% agarose gel.

DNA sequencing and sequence analysis

The resulting PCR product was purified from the gel using GENECLAN Kit (BIO 101, Vista, CA) and cloned into pBlueScript II KS⁺ (Stratagene, La Jolla, CA) at Eco RV site and designated as pKS *Gly1-350*. The cloned PCR fragment and purified cDNA clone(s) were sequenced on double stranded DNA using a dye-terminator cycle sequencing kit (PE Applied Biosystems) and an automated capillary DNA sequencer (Genetic Analyzer ABI 310, PE Applied Biosystems).

All sequencing data was analyzed using LASERGENE

software (DNASTAR, Madison, WI). Searches for information and homology of nucleotide and amino acid sequences were carried out using homology search with BLAST against the sequences in the NCBI GenBank and EMBL DNA databases.

Construction of cDNA Library

Poly (A)⁺ RNA was obtained from total RNA of seeds from yellow-1 pod maturity, by two successive runs on oligo (dT)-cellulose spin column following the manufacturer's protocol (Clontech, Palo Alto, CA). First strand synthesis was performed in a total volume of 10 ml. The reaction mixture consisted of 2 μ g mRNA, 1 μ l of 1 μ M oligo dT (dT15) primer, 200 units of MMLV reverse transcriptase, 2 μ l of 5 mM dNTP and 2 μ l of 5X first

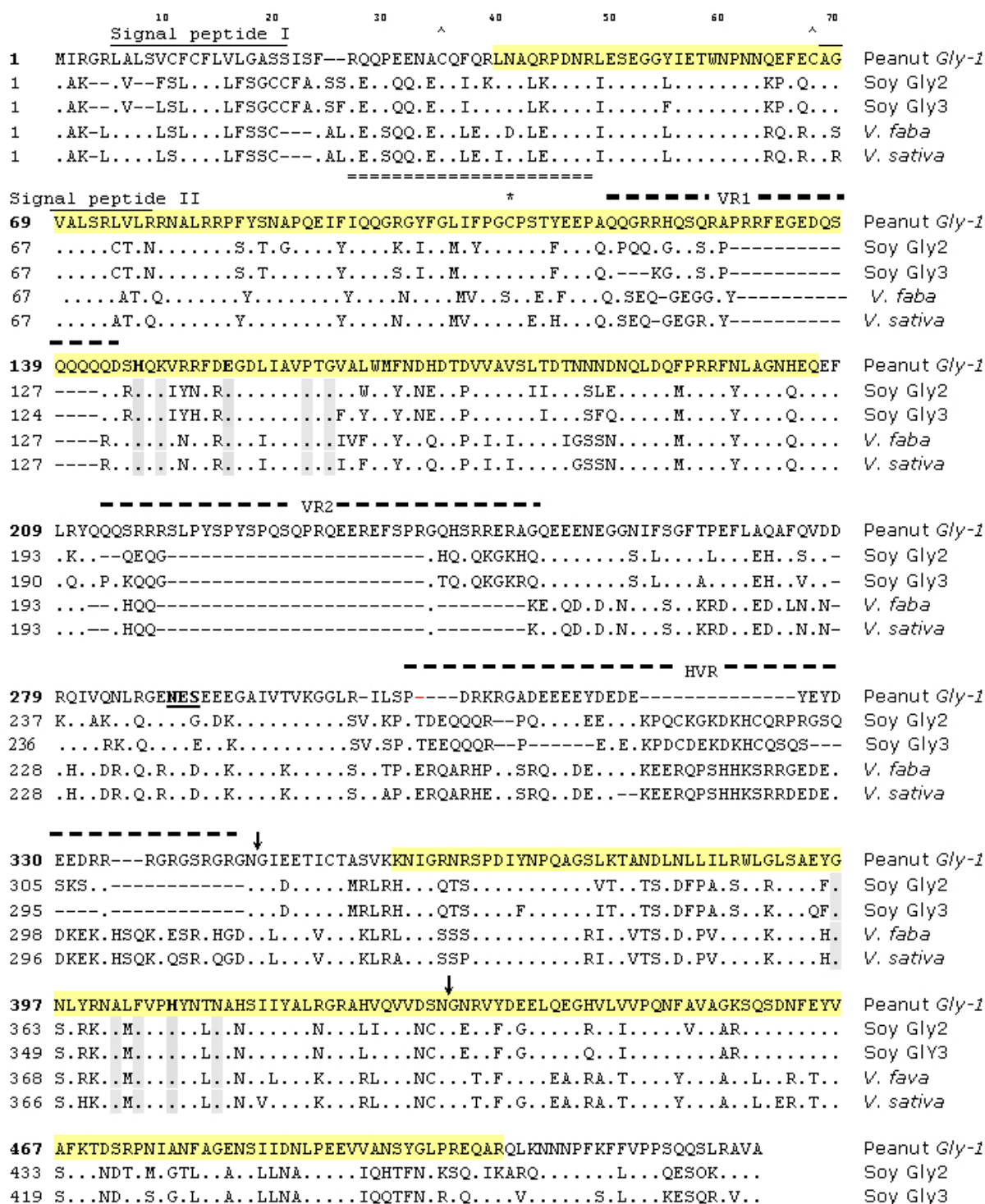


Figure 3. Comparisons of deduced amino acid sequence of peanut glycinin peanut *Gly-1* (Acc #AF125192, gi|9864776) with sequences from soybean Gy2 (gi|18636) and Gy3 gi|18638) subunit; *V. sativa* legumin A (gi|483448) and *V. faba* legumin A2 gi|100101) comparing conserved residues and binding motifs.

Numbering on the left indicates the sequence of corresponding residue.
Gaps have been introduced to maximize alignments.
Dashes indicate gaps introduced to maximize alignment.
Dots indicate identical amino acid residues to peanut *Gly-1* clone.
Cysteine residues marked by an asterisk (*) involved in interchain linkages, whereas those marked by a cap (^) identify conserved residues possibly involved in interchain linkages.
The signal peptide (SP) are shown by a straight (____) line, cleavage site NG indicated by the vertical arrowhead (residue 344/345 and residue 431/432);
Glycosylation signals (NXS) are in bold and underlined.
Regions homologous to variable regions I and II (VR1, VR2) and to the hypervariable region (HVR) known from other glycinin is indicated by broken (----) line.
The consensus cupin signature residues are in highlighted, the metal-binding active site ligands of two histidines and a glutamate are in bold.
Doubly underlined residues indicate conserved a-subunit region of typical glycinin/legumin gene family.

Cloning and structural analysis of a cDNA clone encoding glycinin (*Gly-1*) seed storage protein of peanut

strand buffer [250 mM Tris (pH 8.3), 30 mM MgCl₂, 375 mM KCl]. Double stranded cDNA was synthesized in the presence of DNA polymerase I (9.0 U/μl). Following the addition of Eco RI linkers the cDNA was ligated into the Uni-ZAP XR vector and packed *in vitro* with GigapackIII Gold Cloning Kit as recommended by the manufacturer (Stratagene, La Jolla, CA). The primary library was amplified in SM buffer [0.05 M Tris-HCl (pH 7.5), 0.58% NaCl, 0.2% MgSO₄ and 0.01% gelatin] according to the manufacturer's instructions and the aliquots of the amplified library was stored in 7% DMSO (v/v) at -80°C.

Preparation of radiolabelled probe

Amplified PCR fragment was excised from pKS *Gly1-350* using Sal I and Eco RI restriction enzymes. Fragment was purified using GENECLAN Kit (BIO 101, Vista, CA). The double strand DNA (100 ng) was denatured by heating at 95°C for 5 min and radiolabelled with α³²P dATP following the Prime-a-Gene Labeling System (Promega, Madison, WI). The unincorporated nucleotides were removed using Spin Column 30 (Sigma Chemical Co., St. Louis, MO) and radiolabelled probe was as a probe to screen cDNA library. For northern and southern blot hybridization, 1.6 kb Eco RI fragment was excised from the cDNA clone, radiolabelled (as described above) and used as probe.

Library screening, Isolation of cDNA clone and sequence comparison

The cDNA library was screened in duplicate by plaque hybridization using 350 bp radiolabelled fragment as probe (Sambrook et al. 1989). Three putative positive clones were purified and cDNA clones were designated as λ*peanut-1*, λ*peanut-2* and λ*peanut-3*. All the three clones were excised *in vitro* into pBlueScript SK phagemid following the *in vivo* excision procedure as recommended by the manufacturer (Stratagene, La Jolla, CA) and designated as pSK *peanut1*, pSK *peanut2* and pSK *peanut3*. All the three clones were digested to excise the insert and to determine the insert size through agarose gel electrophoresis. pSK *peanut1*, -2 and -3 were sequenced on double stranded DNA using a dye-terminator cycle sequencing method and by primer walking. Sequences for the other homologous genes were retrieved by the National Center for Biotechnology Information BLAST network service. DNA sequence analysis such as Restriction Map, ORF region nucleotide and amino acid multiple sequence alignment was performed using LASERGENE software (DNASTAR, Madison, WI). After sequence homology and sequence comparison of pSK *peanut1*, -2 and -3, the biggest clone pSK *peanut1* (1.8 kb) was used for further analysis and designated as peanut *Gly1*. The predicted protein data of peanut *Gly1* cDNA clone was used for domain analysis

with the Conserved Domain Database (Marchler-Bauer et al. 2003).

Northern blot analysis

RNA concentrations were determined spectrophotometrically. Ten microgram (μg) of total RNA from seeds (white to black pod maturity category), fully developed leaf, stem, root, flowers, pegs and post-germinating seedlings (0 to 14 days) was separated on a denaturing agarose gel and blotted onto a nitrocellulose membrane (MSI, Westborough, MA) following methods as described (Sambrook et al. 1989). The membrane was UV cross-linked using the GS Gene Linker (Bio-Rad Laboratories, CA). Membrane pieces were pre-hybridized at 42°C in 5 X SSC, 50% formamide, 5 X Denhardt's solution [50 X stock solution consisting of 1% Ficoll 400, 1% polyvinylpyrrolidone, 1% BSA (Albumin Bovine Fraction IV)] in distilled water, 0.5% SDS and 100 μg/ml denatured calf thymus DNA for 16 hrs. The 1.6 kb Eco RI fragment of peanut *Gly1* was excised, radiolabelled and used as probe. Hybridization was performed at 42°C for overnight. Blots were washed twice at 65°C in a solution containing 1X SSC and 0.1% SDS for 30 min. Hybridized membrane pieces were exposed to Kodak-X-OMAT AR film for 18 hrs with an intensifying screen.

RESULTS

Isolation of peanut glycinin DNA sequences

Amino acid sequence of glycinin seed storage protein of soybean and broad bean was used to design and synthesize degenerate oligonucleotide primers. PCR was performed using combinations of degenerate oligonucleotide primers (*GlyUP.1*, *GlyUP.2* and *GlyDN.1*), annealing and polymerization temperatures and cDNA library of immature peanut seeds as template. No PCR product was detected from the primer combination of *GlyUP.1* and *GlyDN.1*. One distinct PCR product (350 bp) was detected from the primer combination *GlyUP.2* and *GlyDN.1* at 45°C annealing and 72°C polymerization temperature. PCR was repeated three times, no other product was amplified. The PCR amplified fragment of 350 bp was of expected size based on the position of *GlyUP.2* and *GlyDN.1* primers on amino acid sequence of glycinin protein. As such no other product was amplified, consequently 350 bp PCR fragment was cloned and sequenced. BLAST information revealed that it shares a high degree of nucleotide homology with glycinin and legumin 11S seed storage protein.

Molecular characterization of peanut glycinin gene (peanut *Gly-1*)

Three clones were isolated and purified from the immature seed cDNA library. These clones contain inserts of 1.8,



Figure 4. Graphical overview of peanut *Gly-1* representing location of conserved barrel domain of bi-cupin motif. The N-terminal motif consists of 169 residues (38 through 206; E value $4e-13$) and the C-terminal motif consists of 145 residues (362 through 506; E value $5e-25$). Numbers indicate amino acid residue.

1.05 and 0.75 kb. All cDNA clones were sequenced and compared for the homology. On sequence analysis it was found that they do not differ, the 1.8 kb clone was found to contain a complete gene as judged by comparison to other genes and designated as peanut *Gly-1* (peanut glycinin-1; NCBI accession # AF125192, gi|9864776) clone. [Figure 1](#) shows the restriction map of the peanut glycinin clone spanning a total of 1836 bp. The cDNA lacks unique restriction sites for Bam HI, Pst I, Sac I, Sal I, Sma I and Xba I. The unique restriction sites present are Cla I, Eco RI, Xho I and Xho II ([Figure 1](#)). The predicted protein has 529 amino acid residues, with a calculated Mr of 60,447.61 Da and an isoelectric point of 5.489. The hydrophathy profile of the protein, deduced by the algorithm of Kyte and Doolittle(1982), shows the presence of two hydrophobic regions located within the first 76 residues (residues 6 to 21 and 67 to 76), each of which is long enough to span a membrane bilayer (regions 1 and 2 in [Figure 2](#)). The protein contains two potential asparagine-linked glycosylation sites (Asn-X-Ser/Thr) at position 289 and 362. Nevertheless, no data is presently available on the glycosylation of other known plant glycinin genes. The 1.05 and 0.75 kb cDNA clones that were initially selected and were found to represent shorter versions of peanut *Gly-1* cDNA clone. The isolation of multiple independent cDNAs that represent the same mRNA suggests that the corresponding mRNA is abundant.

Peanut *Gly-1* is homologous with other glycinin/legumin genes

The BLAST results of peanut glycinin cDNA sequences show that sequences have high levels of similarity with other homologous glycinin or legumin seed storage genes. Peanut *Gly-1* shows high homology with two other peanut glycinin clones - *Ara h3* (95%) and *Ara h4* (94%). Peanut *Gly-1* nucleotide sequence shows 67 and 69% identity with soybean glycinin gene (*Gly 3* and *Gly 2* respectively), whereas it has 60 to 63% identical and 34 to 48% conserved residues using Hein with weighted method (Hein, 1990). At present, the primary sequence of glycinin gene is known for a variety of legumes including soybean (Nielsen et al. 1989), broad bean (Panitz et al. 1997) and Pea (Casey and Domoney, 1999). From these data it becomes evident that all glycinin or glycinin like protein genes, characterized so far, share a high level of similarity in the N-terminal and C-terminal region compared to the central region of the protein. In this respect, the amino and C-terminal part of the peanut glycinin gene (residue 1-120 and 369-553) also have high level of similarity to the

previously reported glycinin sequences ([Figure 3](#)).

Peanut glycinin-1 deciphers 11S globulin type A plant seed storage protein and has bi-cupin domains

The alignment of peanut glycinin-1 sequence with those of legumin and glycinin proteins from other legumes shows that there is close similarity between these sequences ([Figure 3](#)). The sequence comparison and homology shows presence of a conserved a-subunit region in the N-terminal region, which infer that peanut *Gly-1* decode type-A legumin or glycinin protein and deciphers 11S globulin seed storage protein. The amino acids located in the identical position reveal interesting features concerning the similarities among the protein ([Figure 3](#)). The alignments shown in the [Figure 3](#) reveal three regions of variability in the legumin-like subunits. One occurs at the carboxyl ends of the acidic chains in the subunits, and is especially variable. This region has been referred to as the hypervariable region (HVR) (Argos et al. 1985). The second region of variability was first observed by Simon et al. (1985), and referred to as variable region 1 (VR1). It is located between positions 118 to 144 in [Figure 3](#). The VR1 region of peanut *Gly-1* has extended runs of glutamine and arginine that are not observed in other legumin-like proteins. Finally, the third region of internal heterogeneity, variable region 2 (VR2), spans amino acid position 212 to 251 of the alignments shown in [Figure 3](#). The composition of VR2 also tends to be rich in arginine and glutamine, a characteristic of the other variable regions ([Figure 3](#)).

The predicted amino acid sequence was searched for conserved domain that aligned peanut *Gly-1* sequence with that of recently designated cupin superfamily of proteins (Dunwell, 1998), which is among the most functionally diverse proteins. These proteins are named on the basis of the conserved b-barrel fold ('cupa' is the Latin term for a small barrel), peanut *Gly-1* have bi-cupin domains that comprises two conserved motifs ([Figure 4](#)), each corresponding to two b-strands separated by a less conserved region. The N-terminal motif consists of 169 residues (38 through 206) while the C-terminal motif consist of 145 residues (362 through 506). Both the N-terminal and C-terminal motifs of peanut *Gly-1* have high levels of similarity with cupin domain. According to the list of conserved domain database (CDD) peanut *Gly-1* clone belongs to 00190 *Pfam* family of 11S plant seed storage

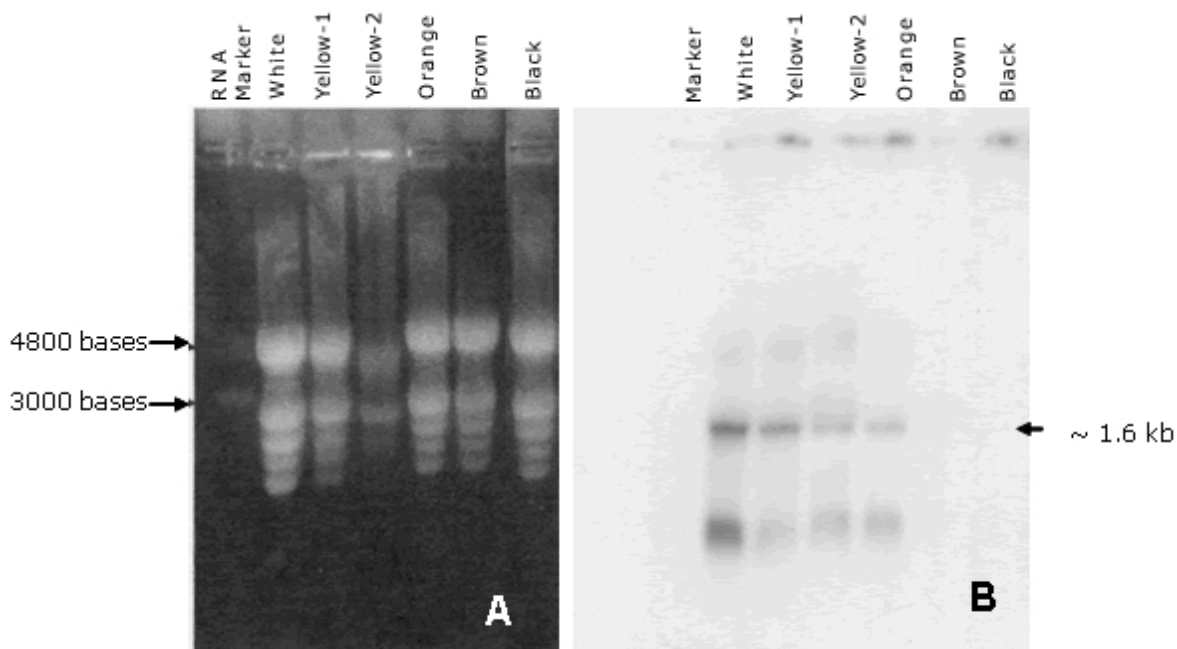


Figure 5. Northern blot analysis.

(A) Ten micrograms of total RNA from seeds (white to black pod maturity category) was separated on a denaturing agarose gel, blotted onto nitrocellulose membrane, and hybridized with *Gly-1* labeled probe.

(B) *Gly-1* expression in developing seeds, a band of approximately 1.6 kb size represent *Gly-1* transcript. The smaller band on the blot is the truncated version of *Gly-1* transcript.

protein. The phylogenetic and functional studies based on sequence alignment and conserved domains of known sequences reveal that most of them contain a single metal-binding site, which has been characterized by two conserved histidines and a glutamate in the first motif, together with a third conserved histidine in the second motif (Dunwell et al. 2001; Khuri et al. 2001). It is known that these conserved residues act as ligands for active-site metal, which has been shown to be a single manganese atom for barley germin (Requena and Bornemann, 1999; Woo et al. 2000) and for plant germin-like proteins (GLPs, Carter and Thornburg, 1999). Peanut *Gly-1* does not contain the second histidine residues in the conserved active site of first motif instead it has a lysine. The presence of positively charged lysine in first motif region might provide the same function as histidine except altering the binding characteristics of the active site. Multiple alignment of the protein sequence presented in this paper (Figure 3) reveal the presence of a lysine in place of second histidine in the first motif that confirms the conservation of lysine with other legumin/glycinin genes. Such sequence variations have also been observed in some *Glycine max* ESTs clones (gi|5759981, gi|6725711, and gi|6566187). This infers that these proteins may change its function and may not bind a manganese atom, or that they may bind it loosely or strongly and only under certain pH condition. However, the conservation in lysine residue in the legumin/glycinin genes among legumes like soybean, field bean and broad bean indicates close phylogenetic

relationship among these legumes and evolution of functional diversity in the cupin motif from other organisms. However, the evolutionary origin of bi-cupins, has yet to be resolved (Dunwell et al. 2001; Khuri et al. 2001).

Peanut glycinin-1 gene expresses during early seed development

In order to examine the expression of *Gly-1* transcripts in plants, total RNA from leaf, stem, root, flower, developing peg, developing seeds and post germinating seeds was analyzed by northern blot using labeled probe derived from the Eco RI fragment of peanut *Gly1* clone. As expected, the northern blotting detected *Gly-1* mRNA band of approximately 1.6 kb. Northern blot results show high level of *Gly-1* expression in seeds at immature pod stage (white) and decayed at the brown pod maturity stage (Figure 5). High levels of *Gly-1* mRNAs transcripts were detected in immature embryos at white pod maturity stage (15 to 20 days after pod formation), which started declining at yellow-1 pod maturity stage (25 to 30 days old pods). *Gly-1* expression continues to decline during yellow-2 (about 35 days old pods) and orange pod maturity stage (40 days old pods). Finally, *Gly-1* expression decayed to non-detectable levels at brown and black pod maturity stage (48 days or older pods) (Figure 5). Peanut *Gly-1* expression was not detected in mature plant leaves, roots, stems, flower and pegs (Figure 5). Northern blots of post germinating seedlings (0 to 14 days) also demonstrate that peanut *Gly-1*

does not express.

Several glycinin genes are present in the peanut genome

The cDNA clone of peanut *Gly-1* does not have recognition sites for Bam HI, Pst I, Sal I and Xba I restriction enzymes. Several unique restriction enzymes including Bam HI, Pst I, Sal I and Xba I were used to digest the genomic DNA in DNA gel blot analysis study. It is possible that genomic DNA region that transcribes peanut *Gly-1* mRNA might be interrupted by one or more intron because at least three introns are present in the soybean glycinin gene family (Nielsen et al. 1989). If introns are present in the genomic region of peanut *Gly-1*, it is possible that these interrupted regions might have additional restriction sites, and it might generate more than one band upon digestion with Bam HI, Pst I, Sal I or Xba I restriction enzyme. The blot showed one or two dark bands and presence of some lighter bands under high stringency wash conditions. Presence of several lighter bands in the DNA gel blot under high stringency wash condition suggests that the peanut genome contains additional glycinin (arachin) gene(s). Earlier studies with peanut proteins have demonstrated that peanut glycinin is composed of multiple subunits (Krishna and Mitra, 1987). Two other glycinin genes *Ara h3* (Rabjohn et al. 1999) and *Ara h4* (Kleber-Janke et al. 1999) from peanut have been isolated and characterized for their allergic property. In legumes, the 11S seed storage (legumins/glycinins) genes are the major storage proteins, presence of multiple subunits for these proteins is a common feature (Argos et al. 1985; Shewry et al. 1995) and an individual gene encodes for one subunit (Nielsen et al. 1989). Hence, a small gene family encodes the 11S seed storage proteins (Cho et al. 1989; Nielsen et al. 1989). Lighter bands in the peanut gel blot analysis suggest the presence of additional gene/s in the peanut genome with some degree of homology with peanut *Gly-1*. Isolation and characterization of peanut genomic clones for other glycinin (arachin) genes will be helpful in understanding the complexity of the gene family, peanut genome and specific role of each glycinin subunit.

DISCUSSION

Comparison of 11S globulin (glycinin or legumin type) seed storage protein genes suggests that there are at least two protein subfamilies, A and B, which are probably of widespread occurrence (Wobus et al. 1986). Both subfamilies share sequence homology and some structural similarities; however, they have divergence that gave rise to two groups that occurred prior to speciation of soybean, pea and broad bean (Nielsen et al. 1989). Three cDNA clones for peanut glycinin protein (*Gly-1*, *Ara h3* and *Ara h4*) were isolated and submitted to the NCBI gene bank around the same time. Designation this gene as *Gly-1* was due to the fact that the PCR product isolated earlier showed high sequence homology with only other glycinin sequences. Other two genes are - *Ara h3* isolated at University of

Arkansas (Rabjohn et al. 1999) and *Ara h4* isolated from Borstel, Germany (Kleber-Janke et al. 1999) has shown IgE mediated allergic reaction (Kleber-Janke et al. 1999; Rabjohn et al. 1999). A detailed understanding of the protein structure and diversity is an important prerequisite for attempts to manipulate quality because it indicates the extent to which the structure of the proteins can be manipulated without affecting their biological properties (Shewry and Casey, 1999).

The characteristic of mature 11S subunits is conserved asparagines-glycine amino acids that are located at the site in the precursor proteins where cleavage occurs to form the acidic and basic chains. It is worth noting, that asparagine-glycine peptides are unique in that they are capable of forming a cyclic imide (Bornstein and Balian, 1977). It is possible that such a cyclic structure could be involved in the proteolytic cleavage mechanism. It has been shown that a single cysteine covalently links the acidic and basic polypeptide chains in soybean glycinin gene family (Nielsen et al. 1989). This is also likely to be the case with peanut *Gly-1* because these cysteines are conserved (Figure 3, position 109 and 373). Only two other cysteine residues in the glycinin subunits are strictly conserved, which are also conserved in peanut *Gly-1* clone (Figure 3, positions 33 and 66) indicate that these same cysteine residues are conserved in other legumin-like subunits as well. The conservation of these two cysteines suggests that they could be involved in an intrachain disulfide linkage (Staswick et al. 1984). The cysteine residues involved in the disulfide bond that link the acidic and basic chains of the glycinin subunits occur within 10 amino acids of each region. This suggests that these sequences are in close proximity in the folded protein.

The cupin superfamily of proteins in plants has been found to share a small number of globally conserved residues with 7S (vicilin) and 11S (legumin) seed storage proteins (Baumlein et al. 1995) and are generally known as germin-like proteins (GLPs). It has been found that some are expressed at critical developmental stages such as embryogenesis (Domon et al. 1995; Neutelings, 1998) or floral induction (Heintzen et al. 1994; Staiger et al. 1999), and many are induced by a range of stresses, either biotic, such as infection with powdery mildew (Thordahl-Christensen et al. 1997; Schweizer et al. 1999), or abiotic, such as exposure to salt (Hurkman and Tanaka, 1996), aluminum (Hamel et al. 1998), or high temperatures (Vallelian-Bindschedler et al. 1998). Present study demonstrates that peanut *Gly-1* is specific to embryogenesis and expressed during early embryo development.

The three variable regions in the 11S proteins may be of importance in efforts to improve seed nutritional quality by protein engineering. The naturally occurring variability in the regions implies that they do not perform critical functions in assembly of the subunits into oligomers. Therefore, it suggests that the changes engineered into the variable regions, and in particular the HVR, would be less

likely to adversely affect the assembly of the subunits than changes in areas where the molecules have been more highly conserved during evolution. Deletion and insertions studies in the glycinin HVR of soybean G4 subunit support this strategy because no detectable effect was reported (Dickinson, 1988).

The 11S legumins/glycinins are the major storage proteins not only in legumes but also in many dicots and some cereals (Shewry and Casey, 1999). The mature proteins are an oligomer of six similar (hexameric) subunits (Badley et al. 1975; Shewry et al. 1995). Peanut contains multiple seed storage proteins some of the arachins (glycinins) are identified as allergens, which are responsible for triggering IgE mediated allergic reaction (Rabjohn et al. 1999; Kleber-Janke et al. 1999; Viquez et al. 2003; Koppleman et al. 2004). High sequence homology of peanut *Gly-1* with *Ara h3* and *Ara h4* suggest that peanut *Gly-1* also carry corresponding four IgE binding regions (R1 consists of amino acid residue 42 to 77; R2 consists of amino acid residue 157 to 177; R3 consists of amino acid residue 250 to 288 and R4 consists of amino acid residue 290 to 347 of the peanut *Gly-1* protein). Understanding the importance of these IgE binding regions for the plant development and health will allow the researchers to mutate and/or alter the IgE binding region to make peanuts non-allergic. These strategies of protein engineering will aid designing hypoallergenic transgenic peanut without losing their relevance to plant growth and development (Tabe and Higgins, 1998; Viquez et al. 2003).

ACKNOWLEDGMENTS

The author is grateful to NCBI for creating an invaluable BLAST resource and North Florida Research and Education Center, University of Florida, Quincy for peanut material.

REFERENCES

ARGOS, P.; NARAYANA, S.V.L. and NIELSEN, N.C. Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO Journal*, 1985, vol. 4, p. 1111-1117.

BADLEY, R.A.; ATKINSON, D.; HAUSER, H.; OLDANI, D.; GREEN, J.P. and STUBBS, J.M. The structure, physical and chemical properties of the soybean protein glycinin. *Biochimica et Biophysica Acta*, 1975, vol. 412, p. 214-228.

BAUMLEIN, H.; BRAUN, H.; KAKHOVSKAYA, I.A. and SHUTOV, A.D. Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. *Journal of Molecular Evolution*, 1995, vol. 41, no. 6, p. 1070-1075.

BORNSTEIN, P. and BALIAN, G. Cleavage at Asn-Gly bonds with hydroxylamine. *Methods in Enzymology*, 1977, vol. 47, p. 132-145.

BURKS, A.W.; COCKRELL, G.; STANLEY, J.S.; HELM, R.M. and BANNON, G.A. Recombinant peanut allergen *Ara h I* expression and IgE binding in patients with peanut hypersensitivity. *Journal of Clinical Investigation*, October 1995, vol. 96, no. 4, p. 1715-1721.

CARTER, C. and THORNBURG, R.W. Germin-like proteins: structure, phylogeny and function. *Journal of Plant Biology*, 1999, vol. 42, p. 97-108.

CASEY, R. and DOMONEY, C. Pea globulins. In: SHEWRY, P.R. and CASEY, R. eds. *Seed Proteins*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, p. 171-208. ISBN 0412815702.

CHERRY, J.P. Potential sources of peanut seed proteins and oil in the genus *Arachis*. *Journal of Agriculture and Food Chemistry*, 1977, vol. 25, p. 188-193.

CHERRY, J.P.; DECHARY, J.M. and ORY, R.L. Gel electrophoretic analysis of peanut proteins and enzymes. 1. Characterization of DEAE-cellulose separated fractions. *Journal of Agriculture and Food Chemistry*, 1973, vol. 21, p. 652-655.

CHO, T.J.; DAVIES, C.S. and NIELSEN, N.C. Inheritance and organization of glycinin genes in Soybean. *The Plant Cell*, 1989, vol. 1, p. 329-337.

DICKINSON, C.D. Assembly properties of glycinin subunits: Development of a novel *in vitro* assembly assay. Ph.D. thesis, Purdue University, West Lafayette, IN, 1988.

DOMON, J.M.; DUMAS, B.; LAINÉ, E.; MEYER, Y.; DAVID, A. and DAVID, H. Three glycosylated polypeptides secreted by several embryogenic cell lines of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer. *Plant Physiology*, 1995, vol. 108, no. 1, p. 141-148.

DUNWELL, J.M. Cupins: a new superfamily of functionally-diverse proteins that include germins and plant seed storage proteins. *Biotechnology and Genetic Engineering Reviews*, 1998, vol. 15, p. 1-32.

DUNWELL, J.M.; CULHAM, A.; CARTER, C.E.; SOSA-AGUIRRE, C.R. and GOODENOUGH, P.W. Evolution of functional diversity in the cupin superfamily. *Trends in Biochemical Science*, 2001, vol. 26, p. 740-746.

HAMEL, F.; BRETON, C. and HOUDE, M. Isolation and characterization of wheat aluminum-regulated genes: possible involvement of aluminum as a pathogenesis response elicitor. *Planta*, 1998, vol. 205, no. 4, p. 531-538.

HEIN, J. Unified approach to alignment and phylogenies. *Methods in Enzymology*, Academic Press, Inc. San Diego, CA. 1990, vol. 183, p. 626-645.

HEINTZEN, C.; FISCHER, R.; MELZER, S.; KAPPELER, K.; APEL, K. and STAIGER, D. Circadian oscillations of a transcript encoding a germin-like protein that is associated with cell walls in young leaves of the long-day plant *Sinapis alba* L. *Plant Physiology*, 1994, vol. 106, no. 3, p. 905-915.

- HURKMAN, W.J. and TANAKA, C.K. Effect of salt stress on germin gene expression in barley roots. *Plant Physiology*, 1996, vol. 110, no. 3, p. 971-977.
- JOHNSON, P.; SHOOTER, E.M. and RIDEAL, E.H. The globulins of the groundnut. II Electrophoretic examination of the Arachin system. *Biochimica et Biophysica Acta*, 1950, vol. 5, p. 376-395.
- JONES, D.B. and HORN, M.H. The properties of Arachin and Con-arachin and the proportionate occurrence in the peanut. *Journal of Agricultural Research*, 1930, vol. 40, p. 672-682.
- KHURI, S.; BAKKER F.T. and DUNWELL, F.T. Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. *Molecular Biology and Evolution*, 2001, vol. 18, no. 4, p. 593-605.
- KINSELLA, J.E. Functional properties of soybean proteins. *Journal of the American Oil Chemist's Society*, 1978, vol. 56, p. 242-258.
- KLEBER-JANKE, T.; CRAMERI, R.; APPENZELLER, U.; SCHLAAK, M. and BECKER W.M. Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *International Archives of Allergy and Immunology*, 1999, vol. 119, no. 4, p. 265-274.
- KOPPELMAN, S.J.; WENSING, M., ERTMANN, M.; KNULST, A.C. and KNOL, E.F. Relevance of *Ara h1*, *Ara h2* and *Ara h3* in peanut-allergic patients, as determined by immunoglobulin E western blotting, basophil-histamine release and intracutaneous testing: *Ara h2* is the most important peanut allergen. *Clinical and Experimental Allergy*, April 2004, vol. 34, no. 4, p. 583-590.
- KRISHNA, T.G. and MITRA, R. Arachin polymorphism in groundnut (*Arachis hypogaea*). *Phytochemistry*, 1987, vol. 26, p. 897-902.
- KYTE, J. and DOOLITTLE, R.F. A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology*, 1982, no. 1, vol. 157, p. 105-132.
- MARCHLER-BAUER A.; ANDERSON, J.B.; DEWEESE-SCOTT, C.; FEDOROVA, N.D.; GEER, L.Y.; HE, S.; HURWITZ, D.I.; JACKSON, J.D.; JACOBS, A.R.; LANCZYCKI, C.J.; LIEBERT, C.A.; LIU, C.; MADEJ, T.; MARCHLER, G.H.; MAZUMDER, R.; NIKOLSKAYA, A.N.; PANCHENKO, A.R.; RAO, B.S.; SHOEMAKER, B.A.; SIMONYAN, V.; SONG, J.S.; THIESSEN, P.A.; VASUDEVAN, S.; WANG, Y.; YAMASHITA, R.A.; YIN, J.J. and BRYANT, S.H. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Research*, 2003, vol. 31, no. 1, p. 383-387.
- MILLER, A. Biochemistry of legume seed proteins. *Annual Review Plant Physiology*, 1975, vol. 26, p. 53-72.
- NEUTELINGS, G.; DOMON, J.M.; MEMBRE, N.; BERNIER, F.; MEYER, Y.; DAVID, A. and DAVID, H. Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine. *Plant Molecular Biology*, 1998, vol. 38, no. 6, p. 1179-1190.
- NIELSEN, N.C.; DICKINSON, C.D.; CHO, T-JU; THANH, V.H.; SCALLON, B.J.; FISCHER, R.L.; SIMS, T.L.; DREWS, G.N. and GOLDBERG, R.B. Characterization of the Glycinin gene family in Soybean. *The Plant Cell*, 1989, vol. 1, no 4, p. 313-328.
- PANITZ, R.; MANTEUFFEL, R.; BAUMLEIN, H. and WOBUS, U. Biphasic expression of a *Vicia faba* legumin B gene in developing seeds of transgenic tobacco. *Journal Plant Physiology*, 1997, vol. 150, no. 1-2, p. 115-126.
- RABJOHN, P.; HELM, E.M.; STANLEY J.S.; WEST C.M.; SAMPSON, H.A.; BURKS, A.W. and BANNON, G.A. Molecular cloning and epitope analysis of the peanut allergen, *Ara h3*. *Journal of Clinical Investigation*, February 1999, vol. 103, no. 4, p. 535-542.
- REQUENA, L. and BORNEMANN, S. Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. *Biochemical Journal*, 1999, vol. 343, no. 1, p. 185-190.
- SAMBROOK, J.; FRITSCH, E.F. and MANIATIS, T. Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY. 1989. ISBN 0 87969 309 6.
- SCHWEIZER, P.; CHRISTOFFEL, A. and DUDLER, R. Transient expression of members of the germin-like gene family in epidermal cells of wheat confers disease resistance. *Plant Journal*, 1999, vol. 20, no. 5, p. 541-552.
- SHEWRY, P.R. and CASEY, R. Seed proteins. Dordrecht: Kluwer Press, 1999. 883 p. ISBN 0412815702.
- SHEWRY, P.R.; NAPIER, J.A. and TATHAM, A.S. Seed storage proteins: Structure and biosynthesis. *The Plant Cell*, 1995, vol. 7, no. 7, p. 945-958.
- SHIN, D.S.; COMPADRE, C.M.; MALEKI, S.J.; KOPPER R. A.; SAMPSON, H.; HUANG, S.K.; BURKS, A.W. and BANNON G.A. Biochemical and structural analysis of the IgE binding sites on *Ara h1*, an abundant and highly allergenic peanut protein. *Journal of Biological Chemistry*, May 1998, vol. 273, no. 22, p. 13753-13759.
- SIMON, A.E.; TENBARGE, K.M.; SCOFIELD, S.R.; FINKELSTEIN, R.R. and CROUCH, M.L. Nucleotide sequence of a cDNA clone from Brassica napus 12S storage protein shows homology with legumin from *Pisum sativum*. *Plant Molecular Biology*, 1985, vol. 5, p. 191-201.
- STAIGER, D.; APEL, K. and TREPP, G. The *Atger3* promoter confers circadian clock-regulated transcription with peak expression at the beginning of the night. *Plant Molecular Biology*, 1999, vol. 40, no. 5, p. 873-882.
- STASWICK, P.E.; HERMODSON, M.A. and NIELSON, N.C. The amino acid sequence of the A2B1a subunit of glycinin. *Journal of Biological Chemistry*, 1984, vol. 259, no. 21, p. 13431-13435.
- TABE, L. and HIGGINS, T.J.V. Engineering plant protein composition for improved nutrition. *Trends in Plant*

Science, 1998, vol. 3, no. 7, p. 282-286.

THORDAHL-CHRISTENSEN, H.; ZHANG, Z.; WEI, Y. and COLLINGE, D.B. Subcellular localization of H₂O₂ in plants: H₂O₂ accumulation in papillae and hypersensitive response during powdery-mildew interaction. *Plant Journal*, 1997, vol. 11, no. 6, p. 1187-1194.

VALLELIAN-BINDSCHEDLER, L.; MOSINGER, E.; METRAUX, J.P. and SCHWEIZER, P. Structure, expression and localization of a germin-like protein in barley (*Hordeum vulgare* L.) that is insolubilized in stressed leaves. *Plant Molecular Biology*, 1998, vol. 37, no. 2, p. 297-308.

VIQUEZ, O.M.; KONAN, K.N. and DODO, H.W. Structure and organization of the genomic clone of a major peanut allergen gene, *Ara h1*. *Molecular Immunology*, December 2003, vol. 40, no. 9, p. 565-571.

WILLIAMS, J.E. and DREXLER, J.S.A. A non-destructive method for determining peanut pod maturity. *Peanut Science*, 1981, vol. 8, p. 134.

WOBUS, U.; BAUMLEIN, H.; BASSUNER, R.; HEIM, H.; JUNG, R.; MUNTZ, K.; PUCHEL, M.; SAALBACH, G. and WESCHKE, W. Characteristics of two types of legumin genes in the field bean genome as revealed by cDNA analysis. *FEBS Letter*, 1986, vol. 201, p. 74-80.

WOO, E.J.; DUNWELL, J.M.; GOODENOUGH, P.W.; MARVIER, A.C. and PICKERSGILL, R.W. Barley germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nature Structural Biology*, 2000, vol. 7, p. 1036-1040.