

Heterologous Expression of Legumin Gene in *E. coli* Isolated from cDNA Clones of Immature Seeds of Pigeonpea (*Cajanus cajan* L.)

Subodh Gupta · Avinash Mishra · A. K. Verma ·
Ruchi Pandey · Laxmi Chand

Received: 3 March 2008 / Accepted: 7 May 2008 /
Published online: 3 July 2008
© Humana Press 2008

Abstract Proteins are one of the targets for improving the nutritional quality, and attempts are being made through manipulation of its native gene(s). Pigeonpea (*Cajanus cajan* L.) is one of the nutritionally important legumes of tropical and subtropical regions of the world, and studies of the structure of seed storage proteins and their interactions have been limited by the difficulty of isolating single-protein subunits in large amounts from a complex mixture of the seed endosperm. One way to overcome this problem is the expression of seed storage protein-encoded gene(s) in heterologous systems that have additional advantages wherein specific gene modifications can be made and the new gene constructs can quickly be expressed. Legumin protein was extracted from pigeonpea seeds of different developmental stages (5th to 25th day after flowering [DAF]) and characterized. The legumin gene (*leg*) of size 1.482 kb was screened, using the deoxygenin-labeled legumin probe, from the complementary deoxyribonucleic acid (cDNA) library, constructed from 18-day-old (DAF) immature seeds of pigeonpea and sequenced (accession no. AF3555403). The legumin gene was further characterized by DNA blotting, and its probable secondary structure was predicted using online ExPASy server. Significant Protein Data Bank (PDB) alignment of the deduced legumin protein by BLASTP was observed with proglycinin of soybean. Comparative 3D structural homology was predicted by Cn3D software, and the

S. Gupta · A. K. Verma · L. Chand
Department of Biochemistry, G. B. Pant University of Agriculture and Technology,
Pantnagar 263 145, India

A. Mishra · R. Pandey
Department of Molecular Biology and Genetic Engineering,
G. B. Pant University of Agriculture and Technology, Pantnagar 263 145, India

Present address:

S. Gupta
Department of Biochemistry, Central Institute of Fisheries Education, Versova, Mumbai, India

Present address:

A. Mishra (✉)
Molecular Biology and Biotechnology, Central Salt and Marine Chemicals Research Institute,
Bhavnagar, 364 002 Gujarat, India
e-mail: avinash@csmcri.org

legumin protein showed the 3D structure alignment and interaction homology with proglycinin chain 1FXZA (PDB no. 1FXZ). The legumin gene was subcloned in vector pET-24a driven by the bacterial promoter, and its expression was detected in *Escherichia coli* by immunoblotting using polyclonal antibodies, raised against the purified legumin protein.

Keywords Heterologous expression · Globulin · Legumin · Pigeonpea · Seed storage protein

Abbreviations

CDS complementary domain sequences
DAF days after flowering
DIG deoxygenin
SSP seed storage protein
d.b. dry-weight basis
D.W. dry weight

Introduction

Seed storage proteins (SSPs) are a set of proteins that accumulate to high levels in seeds during later stages of seed development. During seed germination, these proteins are degraded, and the resulting amino acids are utilized by developing seedlings as a nutritional source. SSPs, the major proteins of grains, are the plant proteins most abundantly consumed by humans. The accumulation of SSPs is therefore a closely regulated biosynthetic process of great agronomic and economic importance. SSPs of most dicotyledonous species are synthesized exclusively in developing embryos, and their biosynthesis represents an excellent system for the investigation of tissue-specific and developmentally controlled gene expression [1].

SSPs are classified into four groups: globulins, albumins, prolamins, and glutelins [2]. The globulins, dominantly represented in legumes, are further divided into subgroups: 11S legumin-type and 7–8S vicilin-type [3]. Legumin is a member of storage proteins (11S globulins) originally found in the family Leguminaceae [3–7] but later found in a diverse array of higher plants, including monocots [6–10].

Legumin is a histidine- and glutamine-rich protein and synthesized as a single polypeptide of 50–55 kDa. Prolegumin is generated by linking of α - and β -subunit regions via cysteine residues through disulfide bonds at the time of insertion of nascent prolegumin into the endoplasmic reticulum lumen followed by cleavage of the signal sequences. The prolegumins are assembled into trimers, transferred through vesicles into the storage vacuole, and cleaved into α and β -subunits to form a hexamer [8]. Molecular masses of α - and β -subunits are generally 30–40 and 18–20 kDa, respectively [3–5, 7, 8]. The amino acid sequences of β -chains are more homogeneous than those of the α -chain, which vary considerably in length because of different numbers of repeats in the C-terminal region [3, 7, 8].

The 11S globulins proteins are collectively called as legumins, but in many plants, trivial names are derived from the genus of the plant, for example, the 11S of soybean (*Glycine max*), Douglas fir, sunflower, and rapeseed are referred to as glycinin, psuedotsugin, helianthin, and cruciferin, respectively. In addition to the 11S globulins (legumin), there is another closely related storage protein, the 7S/8S globulin, commonly called vicilin, which forms a trimer linked by disulfide linkages [8, 11, 12].

There is a direct link between the study of gene expression in plants and the importance of manipulating the control of gene expression. Cloned cDNAs corresponding to mRNA

encoding 7S and 11S storage globulins have allowed the analysis of the primary amino acid sequence of storage proteins, the number of genes encoding proteins, expression of these genes during seed development, and isolation and determination of the structure of genes themselves. Globulins, SSPs, are encoded by gene families varying from a few to as many as 20 or more members. Group 1 glycinins of soybean are encoded by three or four genes [13], while in pea, it was estimated that there are eight genes for legumin, 11 for vicilin, and one for convicilin [14]. The oat 12S globulin gene family appears to contain six to eight members [15]; however, no study is found in literature for legumin gene isolation from pigeonpea and its heterologous expression.

Recent advances in relation to transgenic transfer of storage protein genes have opened up possibilities of examining regulatory sequences that control the developmental expression of SSPs. In most cases, genes of storage proteins have been expressed only in seed tissues and not in others. Further, researchers are needed to delineate the *cis*-acting sequences regulating the expression of genes, which would help in the hyperexpression of desired storage protein genes [16, 17]. Attention has now shifted toward the incorporation of useful genes for reducing yield losses and those for quality improvement.

The ultimate aim of the present study is to develop crop varieties that combine high quality with high yield having balanced essential amino acids. Therefore, proteins are one of the targets for improving the nutritional quality, and attempts are being made through the manipulation of its native legumin genes. Studies of the structure of SSPs, legumin, and their interactions have been limited by the difficulty of isolating subunits in large amounts from the complex mixture of the seed endosperm; besides this, it is also difficult to crystallize mature SSPs, prepared from seeds, because of the heterogeneity of molecular species [18]. One way to overcome this problem is the expression of the legumin gene in heterologous systems. These systems have the additional advantage that specific gene modifications can be made and the new gene constructs can quickly be expressed. Our previous research have proved that the *Escherichia coli* system is very efficient in expressing large amounts of individual SSPs [19], which are biologically active too, derived from a variety of eukaryotic organisms. Previously, we isolated and characterized the vicilin gene of the pigeonpea [20]; in the present study, the legumin protein was extracted from pigeonpea seeds of different developmental stages for characterization, and the legumin gene was isolated from the cDNA library, constructed from 18-day-old (days after flowering [DAF]) immature seeds of pigeonpea. Legumin gene was further characterized by DNA blotting, and its expression was detected in *E. coli* by immunoblotting using a polyclonal antibody.

Materials and Methods

Plant Material

Mature seeds of pigeonpea (*Cajanus cajan* L.) var. UPAS-120 were obtained from the Department of Plant Breeding, G.B. Pant University of Agriculture and Technology, Pantnagar, India, and the variety H 82-1 was taken from Indian Institute of Pulse Research, Kanpur, India. The moisture content was determined by taking 1.0 g seeds of variety UPAS-120 from 5, 10, 15, 20, and 25 DAF, using the method of Iwabuchi and Yamauchi [21].

Seed Storage Protein Fractionation and Analysis

Total SSPs of mature seed were isolated using the method of Wright and Boulter [4] and estimated by the biuret method [22]. The extracted SSPs were precipitated with ammonium sulfate and separated in two fractions, soluble albumin and insoluble globulins, by dissolving in sodium citrate buffer (pH 4.7). Both fractions were lyophilized and quantified, and globulins were again fractionated on Sephadex G-50 column using the zonal isoelectric precipitation method of Shutov and Vaintraub [23]. The procedure of Laemmli [24] was used to check the protein-banding pattern, total number of polypeptides, and molecular weight of each fraction. Slab gel electrophoresis of albumins and globulins: vicilins, and legumins of pigeonpea var. UPAS-120 and H 82-1 was performed. The albumins and globulins of both the varieties were run on 10% polyacrylamide gel electrophoresis (PAGE) and sodium dodecyl sulfate (SDS) PAGE [25], and the gels were analyzed by Molecular Imager System (Bio-Rad, USA). The molecular weights were calculated, and the banding pattern was analyzed using Quantity One 1-D analysis software (Bio-Rad).

Preparation of Polyclonal Antibodies and Immunological Study

Polyclonal antibodies against legumin 11S polypeptide was raised in a New Zealand hyperimmune white rabbit using the method of Lee et al. [26]. The cross-reactivity was checked for polyclonal antibodies by dot-blot analysis and indirect enzyme-linked immunosorbent assay (ELISA). Dot blot is an efficient technique for the detection of proteins as it is a very rapid technique and also semiquantitative in nature, while indirect ELISA is a quantitative as well as qualitative method of protein detection.

cDNA Library and Cloning of Legumin Gene

Total RNA was isolated from 18-day-old (DAF) immature seeds of pigeonpea using the guanidium thiocyanate method [27] and electrophoresed in denatured 1.5% agarose gel [25, 28], prepared in 3-(*N*-morpholino)propanesulfonic acid buffer (0.1 M 3-(*N*-morpholino)propanesulfonic acid, 40 mM sodium acetate, and 5 mM ethylenediamine tetraacetic acid, pH 7.0) and quantified (spectrophotometer, optical density, OD₂₆₀). mRNA was purified using an oligo dT column (New England Biolabs, UK) and subjected for the construction of a representative cDNA library of SSP genes by using the cDNA synthesis system (Promega, USA). The resulting cDNA was cloned into a vector (λ gt-11).

The cloned cDNA was packaged into bacteriophage heads using the Ready-to-go packaging extract (Pharmacia Biotech, UK). The extract had been prepared from *E. coli* strain Y1090 R⁻M⁺ lysogen in which the prophage carried a *cos* mutation. The *cos* mutation is a deletion in the *cos* site, which prevented the endogenous prophage from being packaged while exogenous recombinant DNA is efficiently packaged. After packaging cDNA, cloned in the λ gt-11 vector, the recombinant and nonrecombinant clones were checked by the selection of blue/white plaques with isopropyl- β -D-thiogalactopyranoside (IPTG)/X-gal.

Screening of cDNA Library for Legumin Gene

The nonradioactive deoxygenin (DIG)-labeled legumin cDNA probe (PRC 924), a gift from R. Casey, was used to screen the legumin gene. In situ plaque hybridization was carried out

using a N⁺ Nylon membrane and DIG DNA detection kit (Boehringer Mannheim, Germany), following the manufacturer's protocol.

Subcloning and Sequencing of Legumin Gene

Lambda phage (recombinant λ gt-11 clones) was isolated from all positive legumin plaques [25] and subjected to polymerase chain reaction (PCR; initial denaturation at 94°C for 5 min, 40 cycles of denaturation 94°C for 2 min, annealing 55°C for 2 min and extension at 72°C for 2 min) using specific λ gt-11 forward (5'-GGT GGC GAC GAC TCC TGG AGC CCG-3') and reverse (5'-TTG ACA CCA GAC CAA CTG GTA ATG-3') primers, to determine the insert size. PCR amplicons were purified (Qiagen, USA) and subjected to *Eco*RI restriction digestion (as an insert is flanking with this site that was introduced with adapters during cDNA synthesis and cloning). The restriction-digested insert was gel eluted and purified (Qiagen) and ligated to dephosphorylated pUC18 vector arms. The ligation mix was transformed in the *E. coli* DH 5 α strain, and positive colons were preliminarily screened by blue/white selection with IPTG/X-gal analysis, followed by secondary screening with PCR (both colony and plasmid) and restriction digestion; thereafter, putative legumin clones were sequenced by an automated DNA sequencer. Sequences of clones were subjected for the BLAST and studied using online available bioinformatics tools and submitted to the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) gene data bank.

Characterization of Legumin Gene (DNA Blotting)

Plasmid DNA from positive legumin clones was isolated and digested with the *Eco*RI restriction endonuclease enzyme to excise the insert. The intact plasmid and digested insert were electrophoresed on 1.2% agarose gel, and the DNA was transferred to a N⁺ Nylon membrane. After transferring, it was hybridized with nonradioactive DIG-labeled legumin probes using the DIG DNA detection kit (Boehringer Mannheim, Germany) following the manufacturer's protocol.

Legumin Protein Analysis

Legumin protein analysis was done using computational biology. Amino acid sequences were deduced from nucleotide sequences and imported to ExPASy server (<http://www.expasy.org/tools/>) for primary analysis and to P-val and PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>) for secondary structure prediction and analysis. Legumin amino acid sequences were also subjected to NCBI BLAST to compare with the Protein Data Bank (PDB) database. The PDB database was explored for the comparative study of the legumin protein domain and residue interactions with others homologous domains using the online Cn3D4.1 software (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>).

Small-scale Expression of Legumin Gene in *E. coli*

The complementary domain sequence (CDS) of the legumin gene was amplified from pUC18 cloned legumin cDNA, using primer combination (F: 5'-CTC TCA TAT GAG AGT ACA GGC ACA GC-3' and R: 5'-CTC TCT CGA GTC AGC ATA GTT TTG TG-3'). The *Nde*I (CATATG) and *Xho*I (CTCGAG) restriction sites were incorporated in the forward and reverse primers, respectively, for the directional and in-frame cloning. The PCR

reactions were heated at 94 °C for 2 min to denature the DNA, cycled 35 times at 94 °C for 1 min, 62 °C for 1 min, 72 °C for 1.5 min, and finally incubated at 72 °C for 7 min. PCR products were purified, digested with *Nde*I and *Xho*I, ligated into the expression vector pET-24a (Novagen, USA), and transformed to *E. coli* BL21 (DE3) host cells. A single bacterial colony, containing the recombinant legumin gene in pET-24a, was inoculated into 50 ml 2YT medium containing 50 µg ml⁻¹ kanamycin and 100 mM IPTG, and incubated at 37 °C with gentle shaking until the OD₆₀₀ was approximately 0.6. The cells from 50 ml of the culture were collected by centrifugation at 10,000×g for 1 min and redissolved in lysis buffer (0.1 M MgCl₂, 4% [w/v] SDS, 10% [v/v] glycerol, 5% [v/v] 2-mercaptoethanol, 0.01% [w/v], bromophenol blue, 100 mM Tris-HCl, pH 6.8) [19, 29], and the proteins were analyzed by SDS-PAGE and protein gel blotting using the polyclonal antibody. Legumin and purified globulin proteins isolated from pigeonpea were used as the control.

Result and Discussion

SSPs are found only in seeds, and their accumulation in legume seeds shows sigmoidal kinetics, closely following dry weight (D.W.) gain and is associated with the cell expansion phase of seed development [30]. Detailed analysis of seed protein accumulation has shown that there are usually differential rates of accumulation for each of the storage proteins during seed development [31, 32]. Studies have shown that there is a continuous decrease in moisture content in the developing seeds, which is necessary for the inactivation of metabolic enzymes. In the present study, moisture content was determined by taking 1.0 g seeds of variety UPAS-120 5, 10, 15, 20, and 25 DAF, and it was found to be 90.9%, 80.7%, 78.7%, 73.02%, and 57.51%, respectively, while the moisture content of mature seeds was observed to be 9.1%. This is achieved by lowering the water content; hence, dry and mature seeds have very low amounts of water [33].

Seed Storage Proteins Fractionation and Analysis

The nutritional quality of pigeonpea in terms of its chemical constituents, proteins, amino acids, and digestibility has received increasing attention. Protein quality is of prime importance in pigeonpea products used for human food [34]. The total seed protein content in the two varieties of pigeonpea UPAS-120 and H82-1 were analyzed by using the biuret method [22] and was found to be 29.27% (on dry-weight basis [d.b.]) and 23.15% (on d.b.), respectively (Table 1). Hulse [35] found that the protein content of pigeonpea seed samples ranged between 18.5% and 26.3%, while Singh and Jambunathan [36] found that protein content of 43 commonly cultivated varieties of pigeonpea ranged from 21.1% to 28.1% for dhal samples. Mehta et al. [37] have also reported that the protein content in legume seeds in the range of 20–40% on d.b. The amounts of storage proteins extracted with phosphate buffer were found to be 25.78% in UPAS-120 and 21.78% in H 82-1 (D.W.), which accounts for 88.08% and 94.08% of total proteins, respectively.

The total extracted proteins were subjected to 0–100% ammonium sulfate precipitations. Only 88–94% of the total protein was extracted, of which 79–91% (of extracted protein) were precipitated by ammonium sulfate. The major fractions, globulins and albumins, were separated by sodium citrate buffer (pH 4.7) and found to be 62–63% and 37% of the precipitated proteins, respectively. The ratio of globulins to albumins in mature seeds was approximately 2:1 (Table 1). Almost the same type of results were also obtained in a quantitative characterization of seeds from 59 pea (*Pisum sativum* L.) lines and relative taxa

Table 1 Amount of SSPs extracted from pigeonpea (*Cajanus cajan* L.) seeds variety UPAS-120 and H 82-1.

	UPAS-120				H82-1			
	Protein mg/g (FW)	Percent protein (FW)	Protein mg/g (DW)	Percent protein (DW)	Protein mg/g (FW)	Percent protein (FW)	Protein mg/g (DW)	Percent protein (DW)
Total protein	270.04	27.00	292.71	29.27	207.56	20.76	231.50	23.15
Extracted protein	234.18	23.42	257.82	25.78 (88.08)	195.24	19.52	217.82	21.78 (94.08)
After ammonium sulfate precipitation	213.30	21.33	234.71	23.47 (91.04)	155.67	15.57	173.62	17.36 (80.07)
Globulin	125.00	12.50	137.53	13.75 (58.58)	90.00	9.00	100.41	10.04 (57.83)
Albumin	75.26	7.53	82.74	8.27 (35.23)	53.16	5.32	59.15	5.92 (34.01)

Each value is the mean of three replications. Values in parentheses indicate the percent fraction of total.

FW Fresh weight, DW dry weight

of wide geographical origin, concerning its starch and protein contents and globulin composition [38].

Globulins and albumins of developing seeds, collected at 5, 10, 15, 20, 25 DAF, were also determined. In the phases of seed development, cell division occurred, and intermediates were built up [39]. The rate of protein synthesis remains low with very little proteins being synthesized initially. During later stages and maturation of seeds, food reserved in the form of lipids, carbohydrates, and proteins was laid down in the cotyledons. Ultrastructural studies of *Phaseolus vulgaris* [40], *Vicia faba* [41], and *P. sativum* [42] showed the pattern of multivacuolation in the cotyledon cells. The albumin fraction continuously increased from 5th to 25th day (DAF), and globulin proteins started synthesizing at later stages from the tenth day, and the maximum amount was observed at the 20th day; after that, its synthesis decreased gradually. The same pattern was also observed in several other legumes crops [39, 43–45].

Globulin fractions from mature seeds of pigeonpea varieties UPAS-120 and H82-1 were further fractionated on a Sephadex G-50 column and eluted with citrate buffer (pH 4.7). The elution profile produced two peaks in both varieties: the first was soluble and second insoluble globulins, i.e., vicilin (7S) and legumin (11S), respectively, corresponding to pH 4.7. Wright and Boulter [4] also found third storage proteins named convicilin [46], during zonal isoelectric precipitation of pea globulins. They proposed that convicilin was found to be associated with legumin and vicilin protein fraction at ionic strengths of 0.15 and 0.25 M (or higher), respectively. Croy et al. [46] eluted convicilin by zonal isoelectric precipitation separately from vicilin and legumin by a linear salt gradient at pH 4.7 and found a third overlapping peak between peaks I and II. The convicilin was eluted from the interphase of these peaks by running on a Sephadex G-100 column. Fractions containing convicilin were pooled and applied directly to a column of hydroxyapatite and eluted with a linear concentration gradient of potassium phosphate buffer (pH 8.0). Wright and Boulter [4] have also reported that the zonal isoelectric precipitation is an extremely effective method for obtaining crude legumin and vicilin (about 95% pure) in a single step. In the present study, the interphase, containing convicilin fraction, was discarded, and the remaining two fractions, vicilin and legumin, were collected through ammonium sulfate precipitation. The ration of legumin to vicilin in two varieties of pigeonpea was 39:59 (UPAS-120) and 39.5:59.5 (H 82-1).

For further characterization, all fractions were pooled, lyophilized, and subjected to 10% PAGE and SDS-PAGE. A similar globulin-banding pattern was observed in both varieties, but many differences were observed in albumins, which revealed that globulin proteins are much conserved proteins than albumins. Two fractions of globulin, legumin, and vicilin showed much homology among the varieties; however, there is heterogeneity between legumin and vicilin in the solubility and banding pattern on SDS-PAGE. These results clearly revealed that both the proteins have originated from different groups of genes [47], even though evolution study [48] confirmed that globulin proteins, mainly legumin, were originated by a single cluster of genes from *Archaea* bacteria to plant and found more homology among them. The globulin pattern remained quantitatively almost the same throughout the development.

PAGE analysis of globulins of variety UPAS-120 and H 82-1 showed differences in some low- and high-molecular-weight bands, viz., 93.12-, 90.76-, and 84.51-kDa bands were present in UPAS-120 but absent in var. H82-1. Similarly, low-molecular-weight bands of 44.23 and 26.72 kDa were present in UPAS-120 only, while 105.84-, 42.64-, and 35.32-kDa bands were present in H 82-1 only. All other bands were the same in both varieties. SDS-PAGE for the globulin of UPAS-120 had 56.08-, 30.95-, and 15.75-kDa bands, and

the globulin of H 82-1 had 97.98-, 96.43-, and 14.65-kDa bands different than each other, while the remaining bands were the same.

Globulins showed a difference in banding patterns, when electrophoresed in the presence and absence of β -mercaptoethanol. Bands of 93.12, 90.47, 83.54, 57.01, 44.23, and 32.77 kDa of UPAS-120 globulins disappeared, and some new bands became sharp (42.68, 38.07, 35.01, 33.45, 31.95, 29.29, 22.82, 21.84, 20.72, 17.47, 16.59, 15.75, 8.58, 4.32, and 1.52 kDa) in a 1D scan of the gel; similarly, in H 82-1, 105.84- and 54.11-kDa bands disappeared, and some new bands appeared (38.62, 35.28, 23.76, 23.05, 22.10, 21.69, 21.01, 14.65, 9.98, 5.08, and 1.88 kDa) in the presence of β -mercaptoethanol, other bands were unchanged. An almost similar result was also observed in pea globulin composition, when the protein was electrophoresed with and without β -mercaptoethanol [38].

Mehta et al. [37] revealed extensive α - and β -polypeptide heterogeneity in dissociated legumins. These proteins mainly composed of two subunits, two acidic α -subunits (40 kDa) and six different basic β -subunits (20 kDa). These 20- and 40-kDa subunits are cross-linked by disulfide bonds [49]. Gupta et al. [50] found 19 polypeptides ranging from 13.8 to 92.26 kDa in lentil (*Lens culinaris* M), while Sital and Narang [51] observed three subunits, with molecular weights 16, 50, and 56 kDa for mung bean (*Vigna radiata* L.). In the present study, the globulin fraction of peak II (legumin, 11s) proteins showed nine bands with the same banding pattern in both varieties. These studies revealed that globulin genes are highly conserved and belong to multiple gene family [48, 52, 53].

Immunological Study

Globulin fractions were again fractionated by zonal isoelectric precipitation into legumin (11S) and vicilin (7S) proteins. Polyclonal antibodies were raised against purified protein legumin 11S, and the titer was found to be approximately 1:10,000. Cross-reactivity was also checked for polyclonal antibodies by dot-blot analysis, and it was found that legumin antisera showed no cross-reactivity with vicilin and albumins. Legumin protein contents were determined by the indirect ELISA method, and it was found that the seed harvested at 5, 10, 15, 20, and 25 DAF and mature seeds contained 48.16, 54.23, 58.13, 62.98, 62.16, and 60.58 $\mu\text{g}/\text{mg}$ legumin protein, respectively.

The ratio of vicilin to legumin was also calculated in developing seeds of pigeonpea from the 5th to 25th day (DAF) by the same indirect ELISA, and it was much higher at the early maturation phase, and then it decreased at a later stage. This was a typical curious observation that vicilin started to synthesize first and reached at a maximum level within 15 to 20 days and legumin reached at a maximum level at 20 days after flowering. Similar results were also observed in several other legume seeds [51, 54]. These studies revealed the nature of legumin and vicilin genes; both genes were originated separately during evolution and show subunit heterogeneity among them, but the developmental study of the seeds revealed that both types of genes expressed at 15 to 20 days after flowering.

cDNA Library and Cloning of Legumin Gene

mRNA synthesis for storage proteins must be initiated very early during the development [1, 55–57], with a slow rate [58], and vicilin mRNA level decreases much earlier than that of legumin [59]. Saha and Koundal [60] used 18-DAF seeds of chickpea for the construction of cDNA library. In the present study, it was observed that 18–20-DAF legumin proteins are synthesized with a maximum rate, and a previous study [20] revealed that the isolation of legumin (11S) gene from cDNA clones requires the cDNA library of immature

seeds of pigeonpea at 20 DAF. Total RNA/mRNA was isolated 20 DAF and was subjected to cDNA library preparation [1]. The library was preliminary screened with a blue/white color selection. The total number of plaques in the library was 2,268, out of which 399 were nonrecombinant (blue), while the remaining 1,869 (82.40%) were recombinant (white plaques).

Domoney and Casey [14] obtained 120 recombinant clones when used largest cDNA fraction, fractionated on sucrose gradient. Ishibashi and Minamikawa [61] and Fujino et al. [62] both obtained about 20,000 phages from 2 μg of poly A⁺ RNA, isolated from cotyledons of mature seeds using $\lambda\text{gt-10}$ vectors, while Saha and Koundal [60] obtained 1.5×10^6 PFU/ μg of RNA, and the library contained nearly 40% recombinant plaques as found by blue–white color selection using IPTG/X-gal. After size exclusion column purification through Sephacryl S-300 or other gradient methods, it was found that small-size cDNA clones (~300 kb) were present in large numbers, but they did not have a complete sequence to represent a polypeptide or proteins. Therefore, in the present study, all cDNAs smaller than ~300 kb were discarded, and only larger cDNAs were picked to reduce the number of clones in the library but increase the chances of large-size cDNA genes, which may have a complete sequence of the polypeptide. A total of 1.35×10^4 clones were obtained from 5 μg of mRNA, which indicate that the cDNA library must have all major SSP genes.

The cDNA clones were amplified in *E. coli* LE 392 host cells and legumin clones screened by the nonradioactive DIG-labeled cDNA probe of legumin (PRC 924, gifted from R. Casey). All possible positive clones were reamplified and plated for secondary screening with the same probe. Finally, after the third screening, five positive legumin clones were obtained, and all were subjected to PCR amplification using specific $\lambda\text{gt-11}$ forward and reverse primers followed by restriction digested with *EcoRI* to find out the insert size [63]. The putative legumin insert of size 1.5 kb was ligated to the dephosphorylated pUC 18 vector and transformed to *E. coli* DH 5 α . Positive clones were preliminarily screened by blue/white selection with IPTG/X-gal analysis, followed by secondary screening with PCR (both colony and plasmid) and restriction digestion, and thereafter sequenced with automated DNA sequencer. Sequences of clones were subjected for BLAST and studied using online available bioinformatics tools and submitted to NCBI gene data bank (accession number AF 3555403).

Pairwise homology of nucleotide sequences of isolated legumin gene with other known legumin genes was determined by the BLASTN 2.2.16 program of NCBI [64]. Close homology was observed with *P. sativum* leg A class precursor gene. The overall homology of the isolated legumin gene was observed to be 61.76% with cDNA library genes and 62.35% with genomic library genes. The size of the legumin gene was 1.482 kb, which gave an idea that the clone may contain a complete expressed gene sequence (legumin mRNAs). Saha and Koundal [60] obtained cDNA ranged between 0.7 and 1.8 kb, while Domoney and Casey [14] observed legumin, vicilin, and convicilin seed storage genes of sizes ranging between 0.6 and 2.3 kb. Hager et al. [53] isolated cDNA clones of sizes 1.4 to 1.8 kb, and Gijis [65] isolated a legumin-like storage protein gene from flax (*Linum usitatissimum* L.) with the size of ~1.0 kb. Ealing and Casey [66] reported two cDNA clones of pea (*P. sativum*) seed lipoxygenase of size 3 kb; however, Marraccini et al. [67] isolated a 1-kb cDNA clone of the complete 11S SSP gene of *Coffea arabica*. The NCBI homology BLAST search, comparative homology, and in silico studies also indicated that isolated legumin cDNA contained a 5' untranslated region, signal peptide, an open reading frame (ORF; CDS), as well as a 3' poly-A tail. It is, therefore, concluded that the cDNA clone of the SSP gene legumin (*leg*) of size 1,482 b is full length and contained complete expressed gene sequences with CDS.

Characterization of Legumin Gene (DNA blotting)

Plasmid DNA from positive legumin clones was isolated and insert was excised by *EcoRI* restriction digestion. The intact plasmid and insert was electrophoresed on 1.2% agarose gel, and the DNA was transferred to a N⁺ Nylon membrane. After transferring, it was hybridized with nonradioactive DIG-labeled legumin probes. Both intact and digested DNA gave positive signals, which proved that legumin cDNA was cloned in the pUC18 vector (Fig. 1).

Legumin Protein Analysis

Legumin gene of 1,482 bp contained an untranslated 5' leader sequence (86 bp) and an ORF of 1,032 bp, coding for 343 amino acids. In silico translation and the primary structure of the legumin gene was done using online bioinformatics tools (ExpPASy Server based—ProtParam [68]). Theoretical *pI*, aliphatic index, and grand average of hydropathicity of the legumin protein (comprising of 343 amino acids with Mr 39,495.9 Da), were found to be 9.70, 84.20, and -0.333, respectively. The extinction coefficient of the deduced legumin protein was found to be 51,045 (assuming all Cys residues appear as half-cystines) and 50,420 M⁻¹ cm⁻¹ (assuming no Cys residues appear as half-cystines) at 280 nm measured in water. The instability index was computed to be 58.92 and classified the protein as unstable. The half-life of the proteins was estimated to be 30 (in mammalian reticulocytes, in vitro), greater than 20 (yeast, in vivo), and greater than 10 h (*E. coli*, in vivo). The secondary structure of the legumin protein (Fig. 2) was predicted using PSIPRED protein structure prediction server [69, 70], and the significant PDB alignment of the legumin protein (Fig. 3) by BLASTP-2.2.16 [64, 71] was observed with the crystal structure of proglycinin chain A (A1aB1b homotrimer) of soybean (PDB no. 1FXZ [18]).

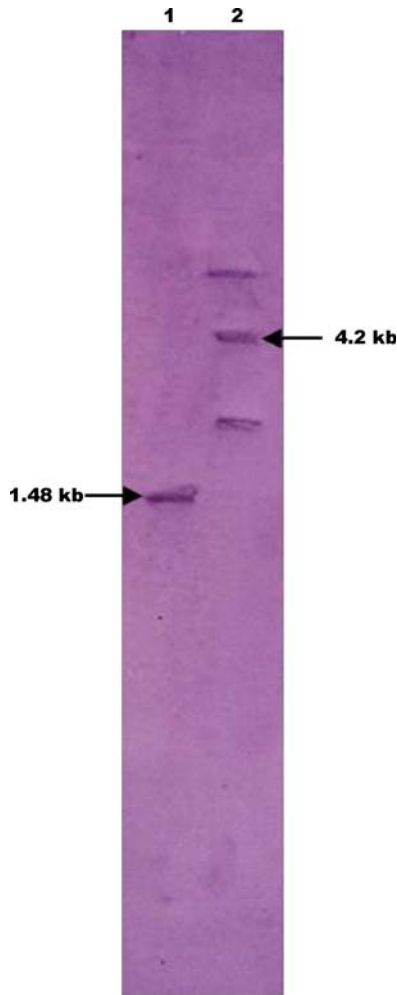
Legumin protein sequences (accession no. AF355403) were imported in Cn3D software and compared with the 3D structure of soybean (*Glycine max*) SSP, proglycinin (PDB no. 1FXZA, MMDB Id 17263). Proglycinin is a homotrimer of 1FXZ_A, 1FXZ_B, and 1FXZ_C and each containing two domains d1 and d2. The legumin protein showed homology with 1FXZA-d1. The alignment of the legumin protein (Fig. 4) to the 3D structure of the proglycinin protein revealed the right ORF of the nucleotide sequences encoding the mature active proteins.

Small-scale Expression of Legumin Gene in *E. coli*

The ORF, which encoded a functional common buckwheat legumin-like protein of 195 amino acids, was expressed in the host strain *E. coli* BL21 (DE3), induced by IPTG; thereafter, recombinant proteins were analyzed by protein gel blot [72]. The production of recombinant 11S globulins in vitro and in vivo may facilitate the elucidation of various steps in their biogenesis, assembly, and deposition. The production of recombinant globulins in *E. coli* overcomes the limitation of isolating globulin fractions (legumin, vicilin, etc.) in large amounts from the complex mixture of the seed endosperm to study the structure of proteins and their interactions.

Host cells, harboring the recombinant legumin gene, were grown in the presence of IPTG, and total bacterial protein was isolated. Previously purified globulin, legumin protein (from pigeonpea seed), and total bacterial protein was electrophoresed (SDS-PAGE), and proteins were transferred on to the nitrocellulose membrane. The nitrocellulose membrane was first exposed with antilegumin antisera or primary antibodies (1:500 dilution) for antigen–antibody reaction and then subjected to second antibody conjugated with alkaline

Fig. 1 DNA blot analysis of the legumin gene cloned in pUC18 vector, using a nonradioactive DIG-labeled cDNA probe of legumin (PRC 924, gifted from R. Casey). *Lane 1*—*Eco*RI-digested *leg* insert of size 1.48 kb; *lane 2*—undigested vector pUC18 (size 4.2 kb) harboring the *leg* gene



phosphatase enzyme. The membrane was developed with the substrate of alkaline phosphatase enzyme (NBT/BCIP). Blue color bands appeared in response to the primary legumin antibody reaction. Globulin and legumin proteins produced several bands with sizes of approximately 62.22, 42.31, 37.28, 34.43, 32.56, 24.72, 23.19, 20.87, and 18.22 kDa; however, signals were strong in globulin. The cloned legumin gene expressed in *E. coli* host cells also produced several bands with sizes of approximately 60.48, 40.83, 35.00, 34.19, 32.31, 22.81, 18.00, and 10.71 kDa (cluster form). Two bands (35.00 and 34.19 kDa) were observed in the total bacterial protein of recombinant cells (Fig. 5) compared to host cells (nonrecombinant), which were the product of legumin genes of the cDNA insert in the plasmid, and others may be the product of host cells and may be having homology to legumin proteins as it was clearly known that the legumin protein evolved from bacteria to plant [48].

The present study may be the first report of the heterologous expression of the legumin gene of pulses (pigeonpea) in *E. coli*. There are few reports of the heterologous expression of the legumin gene, viz., the expression of cDNA encoding proglycinin subunits of

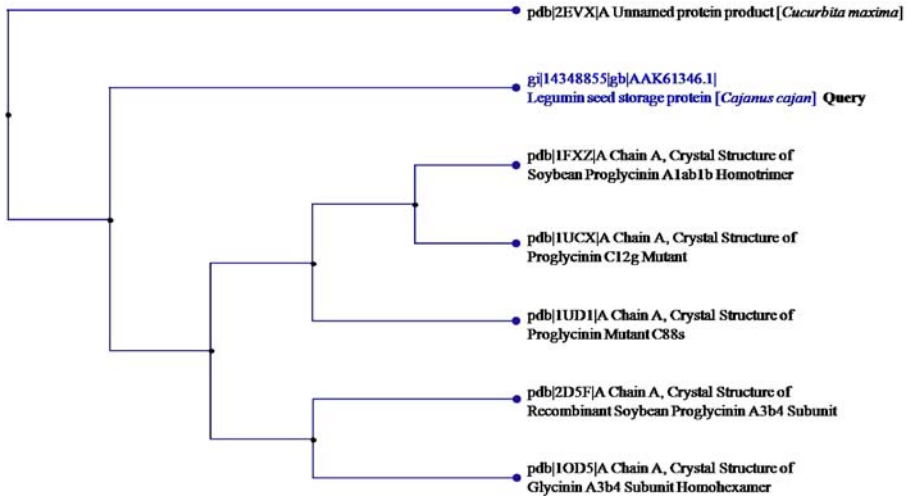


Fig. 3 BLAST tree showing the alignment of the legumin protein with PDB databases, generated by the fast minimum evolution method of BLASTP-2.2.16 [64, 71]

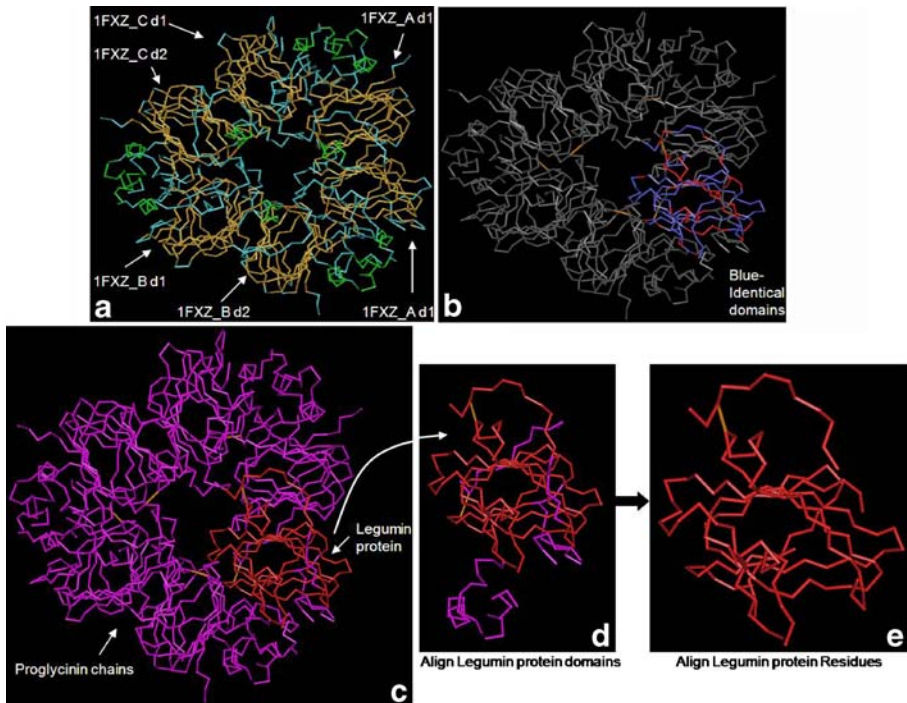
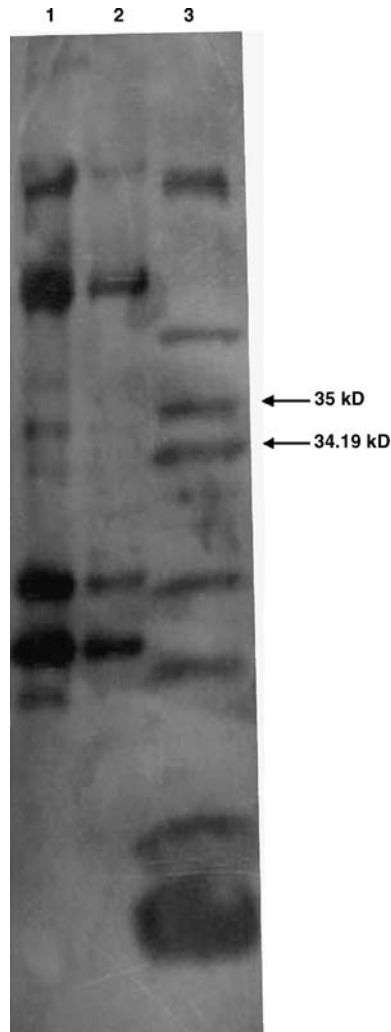


Fig. 4 Structural alignment of the legumin protein of pigeonpea with proglycinin homotrimer (PDB no. 1FXZ, MMBD-Id 17263). **a** Cn3D structure of proglycinin (homotrimer, showing all chains and domains, [23]); **b** superimposition of legumin protomer on proglycinin, showing identical domains (blue in color); **c** structural alignment of legumin protein with proglycinin chain 1FXZ_A, domain-1; **d** legumin protein domains and **e** legumin protein residues, aligned with proglycinin

Fig. 5 Immunoblotting of the legumin protein. Lane 1—total globulin protein, lane 2—legumin protein, fractionated from total SSP of pigeonpea, and lane 3—total protein isolated from log phase of recombinant *E. coli*, harboring the pET-24a expression vector with the *leg* gene



soybean in *E. coli* [73] and expression of legumin-like protein, FA02 β -subunit, of common buckwheat (*Fagopyrum esculentum*) [62], although it is often counted as cereals, where enough studies have been done on heterologous expression of seed storage genes in *E. coli*. The ability to produce the legumin gene in heterologous systems, as well as to assemble them into oligomeric forms, will allow the study of these proteins and analysis of their structural functional relationships.

Conclusion

The low sulfur content of legume seeds is the main nutritional limitation; hence, the isolation of the gene coding for high-sulfur legumin is an important step toward crop improvement. The *leg-3* gene, isolated from chickpea, contained five methionine and six cysteine amino acid residue-coding sequences [74]. Leg A protein isolated from *P. sativum*

also contained six methionine and cysteine amino acid residues [75] (NCBI accession no. AJ132614). In other legumin proteins, methionine and cysteine residues were present, but their number may vary from 2 to 10. The legumin protein has inter- and intradisulfide bonding between acidic and basic polypeptides; hence, the number of cysteine residues is increased. The isolated legumin gene contains a good number of sulfur containing amino acids (10 methionine and 11 cysteine), and it is efficiently expressed in the heterologous system (*E. coli*). This gene thus can be transferred in multicopies for the overexpression in pigeonpea as well as in other legume crops so that sulfur amino acids may be increased and the improvement in nutritional quality of legumes may be achieved. The expression of the SSP genes in *E. coli* also allows the expression in higher plants without modifications and further increases the certainty of expression in transgenics. This knowledge may open new avenues for the expression of other important SSPs and the improvement of legume quality by genetic engineering technology.

References

- Milislavjević, M. D., Timotijević, G. S., Radović, S. R., Brkljačić, J. M., Konstantinović, M. M., & Maksimović, V. R. (2004). *Journal of Agricultural and Food Chemistry*, *52*, 5258–5262.
- Mandal, S., & Mandal, R. K. (2000). *Current Science*, *79*, 576–589. <http://www.ias.ac.in/currsci/sep102000/576.pdf>.
- Shewry, P. R., Napier, J. A., & Tatham, S. A. (1995). *Plant Cell*, *7*, 945–956.
- Wright, D. J., & Boulter, D. (1974). *Biochemical Journal*, *141*, 413–418.
- Saalbach, G., Jung, R., Kunze, G., Saalbach, I., Adler, K., & Muntz, K. (1991). *The Plant Cell*, *3*, 695–708.
- Shutov, A. D., Kakhovskaya, I. A., Braun, H., Baumlein, H., & Muntz, K. (1995). *Journal of Molecular Evolution*, *41*, 1057–1069.
- Shewry, P. R., & Casey, R. (1999). *Seed proteins*. Dordrecht: Kluwer.
- Muntz, K. (1996). *Journal of Experimental Botany*, *298*, 605–622.
- Katsube, T., Kurisaka, N., Ogawa, M., Maruyama, N., Ohtsuka, R., Utsumi, S., et al. (1999). *Plant Physiology*, *120*, 1063–1074.
- Tai, S. S. K., Wu, L. S., Chen, E. C., & Tzen, J. T. C. (1999). *Journal of Agricultural and Food Chemistry*, *47*, 4932–4938.
- Tai, S. S. K., Lee, T. T. T., Tsai, C. C. Y., Yiu, T. J., & Tzen, J. T. C. (2001). *Plant Physiology & Biochemistry*, *39*, 981–992.
- Itoh, T., Garcia, R. N., Adachi, M., Maruyama, Y., Tecson-Mendoza, E. M., Mikami, B., et al. (2006). *Acta Crystallographica. Section D, Biological Crystallography*, *62*, 824–832.
- Fischer, R. L., & Goldberg, R. B. (1982). *Cell*, *29*, 651–660.
- Domoney, C., & Casey, R. (1984). *European Journal of Biochemistry*, *139*, 321–327.
- Shutov, M. A., Afonso, C., Davies, E., Chesnut, R. S., & Larkins, B. A. (1988). *Plant Physiology*, *87*, 698–704.
- Le, B. H., Wagmaister, J. A., Kawashima, T., Bui, A. Q., Harada, J. J., & Goldberg, R. B. (2007). *Plant Physiology*, *144*, 562–574.
- Mishra, A., Tomar, A., Bansal, S., Khanna, V. K., & Garg, G. K. (2008). *Molecular Biology Reports*, *35*, 81–88.
- Adachi, M., Takenaka, Y., Gidamis, A. B., Mikami, B., & Utsumi, S. (2001). *Journal of Molecular Biology*, *305*, 291–305.
- Pandey, R., Mishra, A., & Garg, G. K. (2008). *Molecular Biology Reports*, *35*, 153–162.
- Gupta, S., Verma, A. K., Chand, , & Chand, L. (2006). *Indian Journal of Biotechnology*, *5*, 32–36. http://www.niscair.res.in/ScienceCommunication/ResearchJournals/rejour/ijbt/ijbt2k6/ijbt_jan06.asp#p32.
- Iwabuchi, S., & Yamauchi, F. (1987). *Journal of Agricultural and Food Chemistry*, *35*, 200–205.
- Layne, E. (1957). *Methods in Enzymology*, *3*, 447–455.
- Shutov, A. D., & Vaintraub, I. A. (1965). *Ukrainskij Biokhimičeskij Žurnal*, *37*, 177–181.
- Laemmli, U. K. (1970). *Nature*, *227*, 680–685.
- Sambrook, T., Fritsh, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. New York: Cold Spring Harbour.

26. Lee, C. S., Yu, S. C., & Kapkim, W. (1992). *Journal of Plant Biology*, 35, 99–106. http://www.bosk.or.kr/english/sub05_03.asp?open=4.
27. Chomczynski, P., & Sacchi, N. (1987). *Analytical Biochemistry*, 162, 156–159.
28. Lebraeh, H., Dimond, D., Wozney, J. M., & Boedtke, H. (1977). *Biochemistry*, 16, 4743–4748.
29. Chen, M., & Christen, P. (1997). *Analytical Biochemistry*, 246, 263–264.
30. Higgins, T. J. V. (1984). *Annual Review of Plant Physiology*, 35, 191–221.
31. Sengupta, C., Deluca, V., Baily, D. S., & Verma, D. P. S. (1981). *Plant Molecular Biology*, 1, 19–34.
32. Yamagata, H., Sugimoto, T., Tanaka, K., & Kasai, Z. (1982). *Plant Physiology*, 70, 1094–1100.
33. Boulter, D. (1981). *Advances in Botanical Research*, 9, 1–31.
34. Salunkhe, D. K., Chavan, J. K., & Kadam, S. S. (1986). *Critical Reviews in Food Science and Nutrition*, 23, 103–145.
35. Hulse, J. H. (1977). In J. H. Hulse, K. O. Rachie, & L. W. Billingsley (Eds.), *Nutritional standards and methods of evaluation for food legume breeders* pp. 88–100. Ottawa, Canada: International Development Research Centre.
36. Singh, U., & Jambunathan, R. (1981). *Journal of the Science of Food and Agriculture*, 32, 705–710.
37. Mehta, S. L., Lodha, M. L., & Sane, P. V. (1989). *Recent advances in plant Biochemistry*. New Delhi, India: ICAR.
38. Tzitzikas, E. N., Vincken, J. P., de-Groot, J., Gruppen, H., & Visser, R. G. (2006). *Journal of Agricultural and Food Chemistry*, 54, 425–433.
39. Boulter, D., & Davis, O. J. (1968). *New Phytologist*, 67, 935–946.
40. Opik, H. (1968). *Journal of Experimental Botany*, 19, 64–76.
41. Bariarty, L. G., Coult, D. A., & Boulter, D. (1969). *Journal of Experimental Botany*, 20, 358–372.
42. Bain, J. M., & Mercer, F. V. (1996). *Australian Journal of Biological Sciences*, 19, 49–64.
43. Craig, S., Millerd, A., & Goodchild, D. J. (1980). *Australian Journal of Plant Physiology*, 7, 339–351. doi:10.1071/PP9800339. <http://www.publish.csiro.au/nid/102/paper/PP9800339.htm>.
44. Singh, U., Jambunathan, R., & Narayanan, A. (1980). *Phytochemistry*, 19, 1291–1295. doi:10.1016/0031-9422(80)80162-2.
45. Craig, S., & Millerd, A. (1981). *Protoplasma*, 105, 333–339.
46. Croy, R. R. D., Gatehouse, J. A., Evans, I. M., & Boulter, D. (1980). *Planta (Berl.)*, 148, 49–56.
47. Domoney, C., & Casey, R. (1985). *Nucleic Acids Research*, 13, 687–699.
48. Shutov, A. D., Blattner, F. R., & Baumlein, H. (1999). *Trends in Genetics*, 15, 348–349.
49. Casey, R. (1979). *Biochemical Journal*, 177, 509–520.
50. Gupta, R., & Dhillon, S. (1993). *Annals of Biology (Ludhiana, India)*, 9, 71–78.
51. Sital, J. S., & Narang, A. S. (1994). *Proceedings of the National Academy of Sciences, India*, 64, 189–194.
52. Heim, U., Schubert, R., Bäumlein, H., & Wobus, U. (1989). *Plant Molecular Biology*, 13, 653–663.
53. Hager, K. P., Muller, B., Wind, C., Erbach, S., & Fischer, H. (1996). *FEBS Letters*, 387, 94–98.
54. Yang, L. J., Barratt, D. H. P., Domoney, C. L., Hedley, C. L., & Wang, T. L. (1990). *Journal of Experimental Botany*, 41, 283–288.
55. Gatehouse, J. A., Lycett, G. W., Croy, R. R. D., & Boulter, D. (1982). *Biochemical Journal*, 207, 629–632.
56. Chandler, P. M., Spencer, D., Randall, P. J., & Higgins, T. J. (1984). *Plant Physiology*, 75, 651–657.
57. Darmency, M. A., Abdel-gawwad, M. R., Conejero, G., Verdeil, J. L., & Thompson, R. (2005). *Journal of Experimental Botany*, 56, 2019–2028.
58. Pramanik, S. K., Williams, A. L., & Bewley, J. D. (2007). *Cellular and Molecular Biology*, 53, 64–73.
59. Beach, L. R., Spencer, D., Randall, P. J., & Higgins, T. J. V. (1985). *Nucleic Acids Research*, 13, 999–1013.
60. Saha, S., & Koundal, K. R. (1998). *Journal of Bio-science*, 23, 213–223.
61. Ishibashi, N., & Minamikawa, T. (1990). *Plant and Cell Physiology*, 31, 39–44. <http://pcp.oxfordjournals.org/cgi/content/abstract/31/1/39>.
62. Fujino, K., Funatsuki, H., Inada, M., Shimono, Y., & Kikuta, Y. (2001). *Journal of Agricultural and Food Chemistry*, 49, 1825–1829.
63. Saha, S., Kausal, R., & Koundal, K. R. (1996). *Indian Journal of Experimental Biology*, 34, 1019–1025.
64. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). *Nucleic Acids Research*, 25, 3389–3402.
65. Gijs, V. R. (2000). In serving the plant science community, *Plant Biology—2000*, The American Soc. of Plant Physiology, Abs. 1019.
66. Ealing, P. M., & Casey, R. (1989). *Biochemical Journal*, 264, 929–932.
67. Marraccine, P., Deshayes, A., Petiard, V., & Regers, W. J. (1999). *Plant Physiology and Biochemistry (Paris)*, 37, 273–282.

68. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., et al. (2005). In J. M. Walker (Ed.), *The proteomics protocols handbook* pp. 571–607. Clifton, NJ: Humana.
69. McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). *Bioinformatics*, 16, 404–405.
70. Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., & Jones, D. T. (2005). *Nucleic Acids Research*, 33(Web Server issue), W36–W38.
71. Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., et al. (2001). *Nucleic Acids Research*, 29, 2994–3005.
72. Wang, Z., Wang, L., Chang, W., Li, Y., Zhang, Z., Wieslander, G., et al. (2006). *Bioscience, Biotechnology, and Biochemistry*, 70, 1195–1199.
73. Utsumi, S., Kim, C. S., Sato, T., & Kito, M. (1988). *Gene*, 71, 349–358.
74. Mandaokar, A. D., & Koundal, K. R. (2000). *Journal of Plant Biochemistry and Biotechnology*, 9, 9–12.
75. Domoney, C., Barker, D., & Casey, R. (1986). *Plant Molecular Biology*, 7, 467–474.