Isolation and characterisation of legumin promoter sequence from chickpea (*Cicer arietinum* L.)

Ajit Kumar Shasany & Kirpa Ram Koundal*  
*National Research Centre, Plant Biotechnology, Indian Agricultural Research Institute, Pusa, New Delhi 110 012, India

Received 6 July 1999; revised 3 November 1999

Seed specific promoters are useful for expression of foreign genes in the seeds. We have isolated a *Cicer arietinum* legumin promoter from XEMBL genomic library and subcloned in pBluescript II KS (-) in Eco RV and Pst I site. The 2.762 kb Hind II Pst I fragment was sequenced completely by dideoxy chain termination method by creating a set of unidirectional deletions of the inserts in pAKKIII. The insert contains mainly upstream sequence (2240 bps) and only a part of structural gene (522 bps) sequence. The 522 bps of the structural gene shows ~80% homology with pea legumin A and this is almost the same as chickpea legumin in its sequence. The amino acid sequence derived from the part of the structural gene was similar to the chickpea 5' part of the legumin structural gene with a few variations. A 21 amino acid signal peptide was also deduced like many other legumes. Transcription start site (CAT) was located at 55 bp upstream of the initiation codon ATG. One codon downstream to ATG codon Hind III site was present. TATA box was observed at -30 position, with a consensus of CCTATAATATACC. The consensus CATGCAAG, a part of legumin box was noticed at -110 bp position. At -295 to -265 bp upstream AGGA box like sequences were observed and a 56 bp perfect repeat was located between -913 bp and -972 bp. Strong homology with pea promoter sequence near the CAT sequence was noticed which gradually decreased towards the upstream region. Thus the cloned fragment contains a full length promoter which can be utilised for expression of foreign genes in seeds of chickpea.

The seed storage proteins especially the major one like legumins (11S) from various legumes have been studied in details, including their biosynthesis and structure. The genes for these proteins are expressed to high levels in the cotyledons of legumes and the corresponding mRNAs are accumulated temporally at the middle to late phases of seed development1. As these proteins are the products of group of developmentally regulated genes, which have tissue specific expression, the level of expression changes with the response to the environment. The spatial and temporal expression of the legumin genes7 make these excellent model systems for the study of the molecular mechanism related to differential gene regulation in plants1.

Most of the knowledge on legume storage proteins and their genes comes from Pea, Soyabean, French bean and Broad bean1. The 321 nucleotides towards the upstream from the coding region of leg A, B, C of *Pisum* were analysed and found to be identical1. This part of leg A contains TATA, AGGA/CAAT boxes and sequence homologous to adenovirus enhancer. Also the legumin box containing 30 bp having consensus sequence of CATGCAAG, has been identified, which is responsible for efficient expression of legumin genes4. This box is conserved in all the legumin genes. Leg J in Pism has only TATA and legumin boxes3. Similarly TATA box at -25 to -30 bp and CAAT box at -100 bp have been identified in the 5'region of the glycumin genes. A 9 bp 5'CAACACAAT 3' sequence, which is responsible for seed specific expression is present beside the legumin box1. Sequence comparison in Glycine, *Pisum* and *Vicia* revealed the presence of similar consensus regulatory sequences4.

Identification, cloning and characterisation of the legumin promoter in chickpea is the first step to study the gene expression in a model system. Seed specificity may also be used to produce the transgenic plant resistant to *Heliothis armigera* by directing the expression of Bt δ endotoxin gene in seed, which may reduce the insect insurgency. However, the molecular biology studies in this aspect is very limited in chickpea, which is the major pulse crop of India. So
Materials and Methods

Enzymes and chemicals—Restriction enzymes, modifying enzymes, ligation kit and plasmid pBluescript II KS(-) were procured from Stratagene (USA). Nick translation kit was obtained from Du Pont (UK). Erase-a-base system kit was purchased from Promega Crop Inc. (USA) and sequenase version 2.0 was from USB (USA). 32P dCTP and 35S dATP used were the product of BARC (India).

Genomic library and probes - Chickpea (Cicer arietinum L.) cv Pusa 256, genomic library constructed in EMBL-3 a lambda replacement vector6 was used in the present investigation. Initially, inserts from pDUB6 and pDUB8 containing pea legumin cDNA inserts of 1100 bp (from C-terminal) and 952 bp (from N terminal) respectively, at Bam HI site were used to screen the genomic library.

In the latter stage a fragment of 0.8 kb from 5' region of chickpea legumin gene and a 0.5 kb DNA fragment extending into the structural gene was used to screen the genomic library (Fig. 1). These probes were developed in our laboratory5 and the two fragments were eluted from the low melting agarose gel after Eco RI, Hind III and Eco RV restriction of the recombinant pBluescript II SK(-).

Phage DNA isolation, restriction and Southern hybridisation—Phage DNA was isolated as described by Sambrook et al.6 The DNA samples were restricted with Sal I first and a set of restriction enzymes later and run on 0.8% agarose gel. Southern hybridisation using insert from pDUB6, pDUB8 (0.8 kb 5' probe and 0.5 kb structural gene probes) was performed6.

Subcloning in pBluescript II KS(-)—The 2.76 kb fragment generated by digesting the genomic clone F with Hind II and Psi I was extracted and purified from low melting agarose gel. Plasmid was double digested with Psi I and Eco RV, phenol: chloroform (1:1) extracted, precipitated, dried and dissolved in sterile distilled water. The ligation mixture consisted of 1 µl vector (100ng), 2 µl of fragment DNA (200 ng), 1 µl ligation buffer, 1 µl of 100 mM ATP, 1 µl of T4 DNA ligase (4 weiss unit) and 4 µl sterile water. The mixture was briefly spun and incubated at 15°C overnight. Next day competent E. coli DH 5α cells were transformed with 5 µl ligation mixture and grown in Luria agar plates containing 100 µg/ml ampicillin and coated with 2% X-gal (50 µl) and 0.1 M IPTG (10 µl). The plates were incubated at 37°C for 16 hr and the white bacterial colonies developed were checked for the presence of recombinant plasmid.

Unidirectional deletion with Exonuclease III—The subclone containing the insert was restricted with Kpn I and Hind II. Then unidirectional deletion was carried out according to manufacturers guidelines (Erase-a-base system kit, Promega). After ligation and transformation into E. coli, DH 5α, the deleted subclones were checked for the size of inserts by restricting with Bss III. In the same way unidirectional deletion was carried out from the other end of the insert.

Double stranded DNA sequencing—Plasmid DNA minipreparation was carried out according to Stephen et al.9 DNA sequencing using 35S dATP and Sequenase version 2.0 DNA sequencing kit was carried out according to the dyeoxy chain termination method10. The sequence data were submitted to EMBL Nucleotide Sequence Database (accession # Y13166).

Results and Discussion

Out of nine EMBL-3 recombinant clones screened, using cDNA inserts of pDUB6 and pDUB8, eight hybridised strongly while one clone hybridised faintly. The strong hybridisation with these pea legumin cDNA clones indicated strong homology

![Fig. 1—0.8 kb and 0.5 kb DNA inserts, obtained from pARK3 used as probe for promoter analysis](image-url)
between chickpea and pea legumin genes. DNA samples of the recombinant phages were isolated and restricted with Sal I. Two separate blots were Southern hybridised with inserts of pDUB6 and pDUB8 separately to reconfirm the presence of structural genes.

The variation in intensity of hybridisation signal between the clones may be due to the absence of full length fragment homologous to the probe. In addition, variation exists between the gene families and also within the family which can explain the difference in intensity of signal described interclass homologies within a species as about 50-60% at the nucleotide level. Whereas intergeneric comparisons of a particular group on class show 70-80% homology. These homology studies led to put the *Pisum* leg A, *Vicia legumin* A and *Glycine glycinin* genes in one group and *Pisum* leg J, *Vicia legumin* B and *Glycine glycinin* genes in another group.

To confirm the presence of structural gene in the Sal I fragmented clones (Fig. 2a) were again Southern hybridised with 0.5 kb and 0.8 kb probes described earlier (Fig. 1). The clones B,D,F,I hybridised with 0.5 kb probe containing structural gene from N terminal (Fig. 2b), where as, clones D and F hybridised with 0.8 kb legumin promoter probe (Fig. 2c). This result substantiated the presence of promoter and structural gene in the insert of phage clone D and F and the phage clones B and I contain either a part or full structural gene. From the restriction and hybridisation pattern, assuming D and F are same, clone D was selected for further analysis. The phage clone was digested with a set of restriction enzymes singly and in combination and the hybridisation patterns were studied (Fig. 3a). A fragment of about 2.76 kb showed hybridisation (Fig. 3b, Lane 9) with 0.8 kb probe. This is the only fragment showing the signal from the restriction of clone D with Pst I and Hind II when the other blot was hybridised with 0.5 kb probe the same 2.76 kb fragment showed the signal along with a 0.8 kb fragment (Fig. 3c). The hybridisation signal for 2.76 kb fragment was stronger in comparison to the other minor fragment. Therefore it was concluded that the 2.76 kb fragment contains the 5' upstream promoter region along with the part of structural gene. The structural gene part seemed to be less than 0.5 kb as indicated by an additional faint band in the autoradiogram.

**Subcloning 2.76 kb fragment containing promoter sequence**—The genomic clone D was restricted with Pst I and Hind II to excise out the fragment. The desired fragment was eluted from low melting agarose gel and

---

**Fig. 2a**—Restriction pattern of λEMBL genomic clones with SalI. DNA samples from positive clones identified in tertiary screening were restricted with SalI, run in agarose gel and two separate blots were prepared. One of the blot was hybridized with 0.8 kb and the other with 0.5 kb labelled DNA probes. (A) Lane M Marker λ DNA Hind III digests. Lane A to I Genomic clones restricted with Sal I. (B) Autoradiogram of the blot A hybridized with 0.8 kb probe. (C) Autoradiogram of the blot A hybridized with 0.5 kb probe.
Fig. 3—Restriction pattern of \( \lambda \) EMBL-3 genomic clone D restricted with different enzymes run in 0.8% agarose gel and southern blot hybridization. (A) Electrophoretic pattern of restricted genomic clones Lane N, Marker \( \lambda \) DNA restricted with Hind III and EcoRI Lane 1-9 clone DNA restricted with EcoRV Hind II, Hind III + EcoRI, Dra I, Xba I, BSP 1061 PS+1 and Pst I + Hind II respectively. (B) Autoradiogram of A hybridized with 0.8 kb probe. (C) Autoradiogram of A hybridized with 0.5 kb probe

ligated to pBluescript II KS(-) restricted with Pst I and Eco RV. The E. coli DH 5α cells were transformed, plated on 2% Xgal (50 μl) and 0.1 M IPTG (10 μl) plate containing ampicillin and white recombinant colonies were isolated. Plasmids from these clones were isolated, restricted with Pst I and Sal I, analysed in 0.8% agarose gel (Fig. 4a) and Southern hybridised with 0.8 kb promoter probe (Fig. 4b). As seen in Fig. 4 all four clones hybridised with the probe indicating the presence of promoter sequence.

**Unidirectional deletion by Exonuclease III**—As described in the materials and methods the subclone

Fig. 4—Agarose gel electrophoresis of subclone pAKK III DNA digested with Pst I and Sal I. (A) Lane N, Marker \( \lambda \) DNA, Hind III and EcoRI digest Lane 1, pBluescript II KS(-) Lane 2-5, pAKK III clone DNA. (B) Autoradiogram of A hybridized with 0.8 kb probe

Fig. 5—Agarose gel electrophoresis of unidirectional deletion of pAKKIII DNA with exo-nuclease III and S1 nuclease. Lane M, Marker \( \lambda \)DNA Hind III digest. Lane N, Marker \( \lambda \)DNA Hind III and EcoRI, Lane 1-9 Deleted subclones of pAKKK III, digest restricted with Bss III
was digested with Kpn I and Hind II. Kpn I protects the vector from exonuclease III digestion where as Hind II produces blunt terminus at the insert end which is susceptible to exonuclease III. So a set of overlapping deletions were created for dideoxy sequencing (Fig. 5).

Sequencing of 2.76 kb fragment (promoter)—A total of 2762 bases of the insert cloned in pBluescript II KS(−) were sequenced by sequencing and overlapping the deleted sequences. The complimentary strand was also sequenced to confirm the sequence data. The complete nucleotide sequence data (Fig. 6) and the restriction map derived from the sequence data (Fig. 7, pAKK III) depicts the consensus of the promoter and the amino acid sequences of a part of the structural gene.

Two Hind III sites were found at 34 and 2283 nucleotide position from the 5' end of the sequence. On further analysis of the nucleotides near the Hind III site, (at 2283 position), an initiation codon (ATG) was detected (at 2277 position). The sequence was compared with the available sequences of the legumin genes and 5' upstream sequences of different legA gene family of various crop plants like pea14,15, Le A and Le B4 family of Vicia faba16,4 and glycinin group I genes of soyabean17.

Lyecet et al.3 reported a Hind III site near the start codon in pea legumin A gene. Accordingly sequence near Hind III site was compared with the leg A sequence of pea legumin. Based on the result ATG sequence and Hind III site near to it were numbered at +36 and +42 positions, while the position of Hind III site farther to it was numbered as -2203 nucleotide upstream.

Structural gene—The sequence of the legumin A gene of chickpea has been worked out previously8 and the intron/exon splice junctions have been determined comparing with the boundary consensus sequences18 and the nucleotide sequences of leg A gene of pea14. By comparing the initial nucleotides from ATG codon of the 2.762 kb insert in the subclone pAKK III, with that of the pea leg A, the presence of 21 amino acid signal sequence was observed (Fig. 6) which shows typical features, starting with a positively charged amino acid, β turn creating residue and a signal peptide recognition site8,19,20. Legumin storage proteins in pea are produced as precursor polypeptides which are cleaved at the cleavage site of signal peptide as well as at the α-β polypeptide processing site21,22. Similarly, in chickpea legumin protein was predicted to be 496 amino acid long including 21 amino acid transit peptide and the MW of the protein was expected to be 53 KD with α and β peptides having molecular weight of 33 KD and 20 KD, respectively8.

In pea comparing the gene sequences with the sequences of several legumin cDNAs14 the presence of three intervening sequences of 88 bp, 88bp, and 99 bp long were reported. The first two intervening sequences are present within the sequence encoding the α-subunit and the last one is present in the sequence encoding the β-subunit. In chickpea8 we observed the presence of three short introns of 89, 74 and 229 bp long. These were located by comparing pea leg A and soyabean GI gene; and by comparing the intron-exon boundary sequences with consensus splice site sequences18. These introns were observed at the homologous position as in pea which indicates the phylogenetic relationship of these two species. All the splice site sequences obey GT/AG rule and show additional homology to the donor and acceptor conserved sequences derived from plant genes rather than animal genes23. In the 2.762 kb fragment about 90 bp intervening sequence was observed at homologous position as IVS 1 of pea. When compared to the intervening sequences, the 90 bp intervening sequence was obtained at the same position. An extra T, at the stretch of 8 T increased the number to 90 bases from 89 as described earlier (Fig. 6). About 522 bases from the transcription start site was present in the clone. This shows about ~ 80% homology with the pea legumin sequences. Structural sequences of same position with chickpea legumin has been described earlier8. It was similar except at some places like at +89 and +115 position G bases are present instead of C, at +154 position G is present in place of A; at +274 position T is deleted; at +303 position a G is inserted and finally a 'T' insertion inside the intron at +354 position (Fig. 8). In Fig. 8 the part of structural gene present in the subclone pAKK III has been compared with the part of structural legumin gene of chickpea8. As predicted earlier 2.76 kb fragment includes less than 0.5 kb of the structural gene. About 479 bases are present from the Hind III site of the structural gene which is also near the ATG initiation codon.

Promoter sequences—Similarity in legumin gene and upstream regulatory sequences of various plants like pea, soybean and broad bean led to the belief that
Fig. 6—Promoter DNA sequence and downstream structural gene sequence of the 2.76 kb insert. The deduced amino acid sequence and the intron sequence (Dark boxes) are presented. Consensus sequences like TATA box, legumin box, AGGA boxes, Adeno & SV40 consensus sequences, 56 bp unidirect repeat, CAT, ATG sequences are underlined.
the chickpea regulatory sequences will definitely have the consensus sequences as those found in the other species. The consensus sequence for the cap site in this gene was found to be comparable and in agreement with the LeB4, G1 and Leg A genes. The Cap site was found to be 35 bp upstream to the start codon ATG (Fig. 6). Similarly a search of the sequence for the canonical subsequences observed at the 5' end of most of the genes from animal source and legumin genes revealed a TATA box at position -30. The consensus sequence is homologous to the TATA box of higher plants 17. The TATA consensus was found to be 5' CCCTATAAATAACC 3' compared to 5' CTCTATAAATTACC 3' in pea leg A gene. Another upstream consensus found in animal is CAAT box, the homology in some plants are either poor or not found at all. So Messing et al. 23 proposed a new consensus called AGGA box. In the sequence under study at -161 position GAAAG sequence was found and a stretch of AG rich sequence was observed between -265 to -295 nucleotide position (Fig. 6).

The sequence motif responsible for expression of legumin gene in seed of different crop species are conserved. This conserved sequence of 28 bp are found between -89 to -116 for leg A gene of pea and at -90 to -117 position for G1 gene of soybean. The conserved sequence for pea legumin is given as "TCCATAAGCCATGCAAGCTGAGAATGTC "24. In the present fragment the sequence found at -89 to -117 nucleotide position "TCCATAAGCCATGCAAGCTGAGAATGTC" (Fig. 6). Except two nucleotides variation all the bases were same to that of pea legumin. The most conserved sequence "CATGCAAG" is found in this region. These sequences in lectin and other seed protein genes are listed and called as RY repeats 5. These consensus are responsible for expression in the seed 25 and termed as legumin box.

When about 700 nucleotide sequences from Hind III site (at +42 bp position) were aligned (DNASIS programme) with the leg A upstream sequence of pea in the computer about 65% homology was observed. The comparison of 700 bases upstream from Hind III

---

Fig. 7—Restriction map of pAKK III. Dashed area indicates the fragment sequence while dark part indicates the vector sequence indicating colE1 origin and ampicillin resistance marker.
site of pea and chickpea promoter has been shown in Fig. 9. The homology was perfect near the Hind III site, where the translation starts, which then gradually decreased towards upstream. Other strong homologies were observed in the legumin box (as described earlier), from -117 to -149 bp having similarity with the SV 40 consensus and Adeno virus enhancer sequence. All the consensus sequences including AGGA box like sequences at about -281 bp position (Fig. 6). As shown in Fig. 7 and Fig. 9 the promoter portion can be separated out (-2.25 kb) from pAKKII and cloned in Ti plasmid derived vector for tissue specific gene expression. Based on homology with plant promoters and the presence of consensus sequences like TATA, AGGA and legumin boxes the 2762 bp sequence was confirmed having a promoter sequence including the enhancer like elements in the upstream.

Study of footprinting and gene activity due to mutations in these control regions will be useful. Model plant system could be created by producing transgenic Tobacco and Petunia with different deletion constructs of the promoter. A number of storage protein genes have been transferred to tobacco and Petunia and shown to be tissue specific and developmentally regulated. The regulatory motifs within the upstream sequence as in glycinin A2B1 gene were found out by footprinting experiment involving the seed embryos factors. Lessard et al. also identified the regulatory sequences of B-conglycinin gene of soyabean. The gene containing 159 nucleotides of the 5' flanking sequence was expressed at a low level in immature Petunia. When the gene was flanked by 257

![Fig. 8](image-url)
nucleotides of the upstream region the level of expression increased at least 20 fold. Analysis of the sequence spacing the nucleotides 159 to 257 revealed four repeats of a 7 base pair GC rich sequence (AGCCACA), this play an important part in determining the level of expression of the gene in transgenic plants. The functional significance of such sequences needs to be established by mutational analysis in gene transfer methods. Such approaches will bring about the most rapid progress towards understanding plant development in molecular terms.

As for majority of the transformation experiments and those cases where the transgenic plants have been produced, the promoter used has been a derivative of CaMV 35S. In monocots, the promoters used has been either a derivative of actin or ubiquitin. Most of the promoter to a very large extent are constitutive in nature. In order to ensure the expression of the introduced genes, in a given tissue at a specific time, tissue specific promoter such as the above described one will be of immense help for chickpea improvement which is a major crop of India.

Acknowledgement

The authors are grateful to the Project Director, NRCPB for providing the facilities and Prof Donald Boulter, Univ. of Durham, UK for providing the legumin cDNA clones. Thanks are also due to Mandaokar, Lekha, Sangha, Rekha, and Gupta for providing help. The financial assistance from ICAR, CSIR, New Delhi in the form of fellowship to AKS is also acknowledged.

References


