Two Lectin Genes Differentially Expressed in *Dolichos biflorus* Differ Primarily by a 116-Base Pair Sequence in Their 5' Flanking Regions*

(Received for publication, September 21, 1989)

John J. Harada, Joanne Spadoro-Tank†, John C. Maxwell§, Danny J. Schnell¶, and Mariellyn E. Ktzler∥

*From the Departments of Botany and Biochemistry & Biophysics, University of California, Davis, California 95616*

---

**Abstract**

Previous studies in our laboratory have shown that the *Dolichos biflorus* plant contains two similar lectins, a seed lectin and a stem and leaf lectin called DB58, that are present at different stages in the plant's life cycle. We have now established that each of these lectins is encoded by a separate gene by isolating these lectin genes from a library of *D. biflorus* nuclear DNA. Restriction mapping and nucleotide sequencing analyses show that the seed lectin and DB58 genes are located in the same transcriptional orientation within 3-kilobase pairs of one another. The lectin genes contain no introns and show greater than 90% nucleotide sequence identity in their protein coding and untranslated regions. This sequence similarity extends to both the 5' and 3' flanking regions of the genes; the major exception is that a 116-base pair segment located at position -215 to -100 from the transcription start site of the seed lectin gene is missing in the 5' flanking region of the DB58 gene. The possible significance of this segment with respect to the differential expression of these genes is discussed.

Lectins are widely distributed throughout the plant kingdom and have been most intensively studied in the seeds of legumes where they constitute up to 10% of the soluble protein of the cotyledons (1, 2). A number of these legume seed lectins show a high degree of similarity in amino acid sequence, suggesting a common evolutionary origin (3). The genes encoding the seed lectins are highly regulated resulting in the accumulation of the mRNAs and proteins primarily during the maturation stage of embryogeny and disappears during seedling growth; it has not been detected in the mature organs of the plant (15). By contrast, the DB58 lectin is present in the stems and leaves of developing seedlings where it is noncovalently associated with the cell walls (14). Although this lectin is present in stems and leaves throughout the vegetative life cycle, DB58 is most prevalent in tissues undergoing rapid growth (16).

In order to gain insight into the evolutionary relationship of these two closely related lectins and the mechanisms controlling their differential accumulation, seed lectin and DB58 genes were isolated and characterized. In this paper we confirm that these lectins are encoded by separate genes and show that they are closely linked in the *D. biflorus* genome. A comparison of nucleotide sequences demonstrates that the two genes and their flanking regions exhibit a high degree of similarity, with the major exception of a 116-base pair segment present only in the 5' flanking region of the seed lectin gene.

**EXPERIMENTAL PROCEDURES**

*Nucleic Acid Isolation*—Nuclear DNA was isolated from the leaves of 3-week-old *D. biflorus* plants by the procedure of Jofuku and Goldberg (17). Total RNA was isolated by the method of Taylor and Powell (18) from the stems and leaves of 3-week-old *D. biflorus* plants and from developing seeds at 20 days postanthesis. DNA from recombinant phage and plasmid clones was isolated using standard procedures (19).

*Genomic Clone Isolation*—A nuclear DNA library was constructed in the λ vector Charon 36 using size-selected DNA fragments obtained by limited digestion with Sau3AI (17). Recombinant clones containing lectin genes were identified in plaque hybridization experiments (20) with the unamplified library using the seed lectin cDNA clone as a probe (12).

*DNA Gel Blot Hybridization*—The hybridizations of nuclear DNA gel blots were done as previously described (21) with the following exceptions. First, cDNA probes were labeled with 32P by the method...
Comparison of Two D. biflorus Lectin Genes

Isolation of Lectin Genes—To determine unambiguously whether separate seed lectin and DB58 genes are present in the genome, they were isolated from a library of D. biflorus nuclear DNA (see "Experimental Procedures"). Seventeen recombinant phage containing overlapping regions of the genome were identified with the seed lectin cDNA clone at a hybridization criterion that permitted cross-reaction of the probe with both genes (data not shown). Fig. 2 summarizes the results of restriction mapping, DNA gel blot, and nucleotide sequence studies which indicated that distinct genes encoding seed lectin and DB58 are located in a region of the genome defined by two representative phages, λDb1 and λDb5. DNA gel blot hybridization experiments with phage DNAs confirmed results obtained with nuclear DNA (see Fig. 1); the 5.0-kb HindIII fragment in λDb1 and the 8.9-kb HindIII fragment in both λDb1 and λDb5 hybridized preferentially with the seed lectin and DB58 cDNA clones, respectively (data not shown). The data also provided evidence that the cloned DNAs were not grossly altered. The results support the conclusion that seed lectin and DB58 are encoded by separate genes and that the two genes are closely linked.

Nucleotide Sequence of Lectin Genes—The nucleotide sequences of the seed lectin and DB58 genes are presented in Fig. 3. Comparison of these nucleotide sequences with previously determined mRNA sequences (12, 13) confirmed the identities of the two lectin genes and established their relative locations and transcriptional orientations in this region of the genome (see Fig. 2).

Each gene contains a single open reading frame specifying a 22-amino acid residue signal peptide and a 253-amino acid residue mature polypeptide that is not interrupted by intervening sequences. The absence of introns is characteristic of other plant lectin genes (30-33). In their protein coding regions, the two genes are 94.3 and 87.6% identical in sequence at the nucleotide and amino acid levels, respectively. Comparison of the cloned mRNA and gene sequences and sequencing of portions of the cDNA clones revealed errors in the previously published mRNA sequences (12, 13) that are detailed in the legend to Fig. 3.

RESULTS

Analysis of the Seed Lectin/DB58 Gene Family—Previous analyses of seed lectin and DB58 cDNA clones showed that the two mRNAs are similar but not identical, sharing 92% nucleotide sequence similarity (12, 13, 29). To determine whether these two lectins are encoded by different genes, the number of seed lectin/DB58 genes in the D. biflorus genome and their similarity to the lectin mRNAs were examined. Seed lectin and DB58 cDNA clones were hybridized separately with leaf nuclear DNA digested with restriction endonucleases containing the intergenic region of M13, pUC118 (25), and pUC128 (a gift from S. Tamaki, University of California, Riverside), respectively. Deletion derivatives of the clones were generated using exonuclease III digestion procedures (24). DNA was sequenced by the dideoxy-nucleotide chain termination method (25). In several cases oligonucleotide primers were synthesized and used in sequencing reactions. Nucleotide sequences presented in this paper were determined for both strands of the sequences. The analyses were performed with the Pustell Sequence Analysis Computer program (26) and the Genetics Computer Group Programs, University of Wisconsin (27).

Primer Extension Analysis—The transcriptional initiation sites of the seed lectin and DB58 genes were determined in primer extension experiments (28) with developing seed and leaf stem mRNA, respectively. The primer was a synthetic oligonucleotide complementary to sequences from position +123 to +141 (relative to the translation initiation site) of both genes.

---

1 The abbreviations used are: kb, kilobase pair; bp, base pair.
Fig. 3. Nucleotide sequences of seed lectin (SL) and DB58 (DB) genes and their flanking regions. Gaps have been introduced to maximize homology. The transcription initiation site is indicated with an asterisk and the nucleotides of the protein coding regions are grouped in triplets. The TATA box and polyadenylation sites are overlined. Previously published sequences for seed lectin and DB58 cDNA clones were found to be incorrect. Resequencing of the cDNA clones show that the cloned mRNA sequences correspond precisely to the gene sequences. Translation of these corrected nucleotide sequences results in changes in the previously reported seed lectin amino acid sequence at positions 39, 41, 42, 45, 177, and 232, which become isoleucine, threonine, proline, leucine, alanine, and lysine, respectively, and changes at positions 43, 44, 60, 205, and 246, which become serines. Positions 214 and 215 of the previously reported DB58 sequence are changed to phenylalanines.

Fig. 3 shows that the seed lectin and DB58 genes also exhibit extensive similarity in their 5′ and 3′ untranslated regions. The transcription initiation site, as determined by primer extension analysis (see "Experimental Procedures"), corresponds to the 5′ terminus of the seed lectin cDNA clone (12). Sequences surrounding the transcription start site (TGCATCACC) are similar to the corresponding region of the soybean Le1 lectin gene (TGCATCACA) (29) and to a proposed consensus sequence for plant genes ((C/T)TCATCA) (34). The 37-bp 5′ untranslated regions of both genes (position +1 to +37 relative to the transcription initiation site) are identical in sequence. The 3′ untranslated regions of the seed lectin and DB58 genes extend 140 and 145 bp, respectively, from the translation termination codons to the polyadenylation sites of the cDNA clones. The two genes share 94.5% sequence identity in this region if gaps are introduced to permit maximal sequence alignment, and they contain overlapping consensus polyadenylation sequences at the same relative positions in the genes (positions 956 and 968, respectively). Overlapping polyadenylation sequences have also been found in the Phaseolus vulgaris dedc1, dedc2, and Pfil lectin genes (32, 33). The seed lectin gene contains an additional polyadenylation sequence 116 bp downstream of its protein coding region.

Sequence Similarities in Regions Flanking Lectin Genes—Computer-assisted alignment of nucleotide sequences, summarized in Fig. 3, indicates that the regions flanking the two lectin genes show extensive similarity if one prominent insertion and several minor gaps are introduced to maximize sequence alignment. The 160- and 185-bp regions downstream of the seed lectin and DB58 genes, respectively, share 90% sequence identity. In order to align the 5′ flanking regions, it was necessary to assume an insertion was present upstream of the seed lectin gene at positions -215 to -100. Without this insertion, nucleotides between positions -818 and the transcription start site of the seed lectin gene can be aligned with 92% sequence similarity to the corresponding region of the DB58 gene. This insertion of 116 bp in the seed lectin 5′ flanking region contains one copy of a repeat (AGCCCA) that is similar to a sequence (A(A/G/C)CCCA) proposed to be involved in the regulation of a gene encoding the α'-subunit of the soybean seed storage protein, P-conglycinin (35, 36). The flanking regions of both lectin genes upstream of this insertion are A/T rich (81%), as has been reported for the 5′ flanking regions of the pea (37) and P vulgaris seed lectin genes (32, 33).

Additional consensus sequences thought to play a role in gene transcription were identified in the 5′ flanking regions.

Comparison of Two D. biflorus Lectin Genes

4999
sites. Two other potential TATA box sequences are found at positions -282 and -764 of the seed lectin gene and at positions -168 and -643 of the DB58 gene. It is of interest that the pea lectin gene also has three TATA box sequences located at approximately positions -23, -56, and -272 from the transcription start site (37). The animal core enhancer sequence GTGG(A/T)(A/T)(A/T)G is found at position -1003 of the seed lectin flanking sequence but not in the DB58 flanking region. Conversely, a CAACAC sequence element similar to a (A/C/T)AACACA(A/C)/(A/C/T) consensus sequence present in the 5' flanking regions of many seed protein genes (38) is found at position -522 of the DB58 gene flanking sequence but not in the seed lectin gene flanking region. Sequences similar to the soybean lectin core sequence ATTT(A/T)AAT, found to interact with a DNA-binding protein from embryos (39), are found at positions -260 and -298 of the seed lectin gene 5' flanking region and at positions -847, -432, and -174 of the DB58 gene flanking region. A comparison of the 5' flanking region of the seed lectin gene with regions upstream of the soybean (30) and P. vulgaris (32) lectin genes showed little overall similarity 5' of the TATA boxes. Neither of the 5' flanking regions was found to contain the CATTGCATG motif that has been identified in the upstream regions of several seed storage protein genes (40) and postulated to play a role in modulating P. vulgaris lectin gene expression (41).

**DISCUSSION**

**Seed Lectin and DB58 Lectin Are Encoded by Separate but Similar Genes**—Previous studies from our laboratory on the protein structure and biosynthesis of the D. biflorus seed lectin and DB58 suggested that these two lectins are encoded by separate genes (9, 12, 13, 29). The isolation and nucleotide sequencing of the cloned lectin genes confirm this conclusion. Currently, the number of seed lectin and DB58 genes present per haploid genome cannot be determined unambiguously because the size of the D. biflorus genome is not known. However, no restriction site polymorphisms were detected in 17 overlapping genomic clones containing the lectin genes. If multiple copies of each gene exist, they must be highly conserved in both their coding and flanking regions.

Several other immunochemically related lectin and lectin-like proteins in D. biflorus have been isolated and characterized (9, 10). However, we have not been able to detect genes representing these other lectins using either the seed lectin or DB58 probes. Thus, it appears that the seed lectin and DB58 are encoded by a subclass of highly related lectin genes and that other members of the lectin family have diverged significantly.

**Seed Lectin and DB58 Are Related Evolutionarily**—Restriction mapping and nucleotide sequencing studies showed that the seed lectin and DB58 genes are located within 3-kb of one another (Fig. 2). The linkage of these genes and their similar transcripational orientation in the genome suggests that they may have arisen by a gene duplication event and that subsequent divergence in one of the genes or its flanking region accounts for the different expression pattern. The diec1 and diec2 lectin genes of P. vulgaris are organized similarly in the bean genome (32); in contrast to the two D. biflorus lectin genes, both diec1 and diec2 are expressed in seeds (42, 43). These results suggest that the progenitor gene in D. biflorus is the seed lectin gene. The duplication of lectin genes is not confined to leguminous plants. Sequence similarity between the Ricinus communis agglutinin and ricin suggested that these two seed lectin genes may have arisen by duplication (44, 45). Indeed, internal sequence homologies and conformational studies of the β chain of ricin indicate that this chain, itself, is a product of gene duplication (46).

The seed lectin gene of soybean does not appear to be closely linked with other lectin genes, but rather is flanked by several non-lectin genes (47). In this regard, the nucleotide sequence of the region between the two D. biflorus lectins does not contain extended open reading frames. The presence of two closely linked lectin genes in D. biflorus and P. vulgaris, but not in soybean, may indicate that duplication of the lectin gene occurred subsequent to the divergence of soybean from the other two legumes, even though all three species are in the same legume tribe. Alternatively, it is possible that a duplicated soybean lectin gene was dispersed to another region of the genome. A second soybean lectin gene has been identified although its sequence relationship to the seed lectin gene has not been established (5).

**The 5' Flanking Regions of Lectin Genes Differ by the Presence of a 116-bp Sequence**—The extensive similarity between the flanking regions of the lectin gene extends 818 base pairs upstream and 185 base pairs downstream of the seed lectin gene (Fig. 3). These 5' and 3' flanking regions show 92 and 90% sequence identity, respectively. The major difference between the genes and their flanking regions is a 116-bp segment that is found at position 215 to 100 in the 5' flanking region of the seed lectin gene and is missing from the flanking region of the DB58 gene. This segment is located 68 base pairs upstream of the TATA box.

The presence of the 116-bp segment in the region flanking the seed lectin gene raises the question as to whether it may play a role in regulating its expression. It is of interest that this 116-base pair segment contains a GC-rich sequence similar to a repeat thought to be important for the expression of a storage protein gene in developing seeds (35, 36).

The differential expression of the D. biflorus lectin genes (1), coupled with the limited differences in the sequences of their flanking regions, provide an excellent system for use in studies of gene regulation. Studies are now in progress to determine whether seed lectin and DB58 genes with altered 5' flanking regions are correctly expressed in transgenic plants.

**Acknowledgments**—We thank Judith B. Murphy and Dr. Hae-Young Kim for their technical assistance with the resequencing of the lectin cDNAs.

**REFERENCES**


*J. C. Maxwell and M. E. Etzler, unpublished results.*
Comparison of Two D. biflorus Lectin Genes

Two lectin genes differentially expressed in Dolichos biflorus differ primarily by a 116-base pair sequence in their 5' flanking regions.

J J Harada, J Spadoro-Tank, J C Maxwell, D J Schnell and M E Etzler