The EMBO Journal vol.4 no.4 pp.883–889, 1985

Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome

Leslie M.Hoffman and Debra D.Donaldson

Agrigenetics Advanced Research Division, 5649 East Buckeye Road, Madison, WI 53716, USA

Communicated by H.Kössel

A λ1059 library of *Phaseolus vulgaris* cv. 'Tendergreen' DNA was screened with a cloned lectin-like cDNA. Among the phages selected was clone λB10 containing two complete lectin genes in the same orientation ~4 kb apart. The DNA sequences of the lectin genes and their flanking regions have been determined and their transcriptional initiation sites were located by S1 nuclease mapping. On the basis of the deduced amino acid sequences and compositions and the mol. wts. of their encoded glycoproteins, the genes, *dlec1* and *dlec2*, are predicted to encode erythro- and leucoagglutinating phytohemagglutinins (PHA-E and PHA-L), respectively. The mRNA coding regions of *dlec1* and *dlec2* are 90% homologous, suggesting an origin involving duplication of an ancestral gene. Both lectin genes are intronless and have at least two ATG codons in a short (11–14 bp) 5′-untranslated region. Most of their 5′-untranslated regions consist of alternating pyrimidines and purines (RY repeats). Upstream sequences are also highly conserved between *dlec1* and *dlec2*, including stretches of nine or more alternating R and Y residues. RY repeats of such length are not found within the protein coding portion of *dlec1*, *dlec2* or a *Phaseolus* lectin-like gene previously described. Overlapping double (*dlec1*) or triple (*dlec2*) polyadenylation addition signals are found and there is an unusually high degree of homology (84%) between their 3′-untranslated regions. 20-mer oligodeoxynucleotides specific for *dlec1* or *dlec2* sequences each hybridized with a 1-kb poly(A) RNA from French bean cotyledons, indicating that both genes are expressed.

Key words: multigene family/*Phaseolus vulgaris*/phytohemagglutinin/pyrimidine-purine repeats

Introduction

Lectins are proteins or glycoproteins of non-immune origin which specifically bind carbohydrates and often agglutinate cells (Goldstein *et al.*, 1980). Lectins are commonly found in the seeds of leguminous plants (Toms and Western, 1971), but their physiological role is unclear. Lesser quantities of lectins are also detected in the leaves, stems, and roots of legumes (Talbot and Etzler, 1978; Borrebaeck, 1984). Peptide sequencing has revealed the primary structure of a number of legume lectins. These include concanavalin A from jackbean (Cunningham *et al.*, 1975), lentil lectin (Foriers *et al.*, 1981) and fava bean lectin (Hopp *et al.*, 1982). In addition, Higgins *et al.* (1983) have deduced the peptide sequence of pea lectin from a nearly full length cDNA clone, and Vodkin *et al.* (1983) have determined the amino acid sequence of soybean lectin from a genomic clone.

The cloning and characterization of a *Phaseolus vulgaris* gene which expresses a lectin-like mRNA has been described (Hoffman, 1984). In this report the structures of two closely linked genes encoding PHA-E and PHA-L, respectively, are compared. The linked genes, *dlec1* and *dlec2*, are intronless, as is the lectin-like gene, and they contain highly homologous transcribed regions. Their nucleotide sequences provide the first complete peptide sequences of PHA-E and PHA-L, for which only 24 N-terminal residues were previously reported (Miller *et al.*, 1975).

Evidence is presented that both lectin genes are expressed in cotyledon tissue. *dlec1* and *dlec2* encode PHA polypeptides of very similar mol. wts., but *dlec1* contains three glycosylation sites while *dlec2* has two. Thus, the slower migration of PHA-E compared with PHA-L in an SDS-polyacrylamide gel system may be related to the presence of an additional carbohydrate side chain.

Results

Molecular cloning and sequence comparison of lectin genes

A λ1059 phage library of *P. vulgaris* DNA was screened with a ^{32}P-labeled insert of plasmid pPVL134 (Hoffman *et al.*, 1982). Twenty-seven of 5 x 10^5 recombinant phage hybridized with the cDNA probe. One of the isolates, λB10, was shown by restriction endonuclease mapping and Southern hybridization to contain two genes ~4 kb apart (Figure 1A). The genes, designated *dlec1* and *dlec2*, were subcloned as *Pst*I fragments into pBR322 to produce plasmids pdlec1 and pdlec2.

Nucleotide sequencing by the method of Maxam and Gilbert (1980) was performed on the gene coding and flanking regions of *dlec1* and *dlec2* according to the strategies shown in Figure 1B. The sequences of *dlec1* and *dlec2* (Figure 2) demonstrate that the genes are transcribed in the same direction (see Figure 1A). As in the case of a lectin-like *Phaseolus* gene (Hoffman, 1984), neither *dlec1* nor *dlec2* contains an intron.

The N-terminal sequence of red kidney bean PHA (Miller *et al.*, 1975) was used to determine the ends of the mature PHA proteins predicted by the clones. The amino acid compositions and N-terminal sequences of the proteins deduced from pdlec1 and pdlec2 correspond to PHA-E and PHA-L. The N terminus of *dlec1* mature peptide has glutamine residues in two positions where glutamate is predicted and a tyrosine instead of a serine from the published PHA-E sequence. The *dlec2* mature peptide differs from PHA-L only by having a glutamine in place of glutamate.

dlec1 contains an open reading frame for a 21-residue hydrophobic signal peptide (2200 daltons) and a 254 amino acid mature protein (27 500 daltons) (Figure 3). *dlec2* encodes a 20-amino acid signal peptide (2200 daltons) and a 252-residue mature lectin (27 300 daltons) (Figure 3). A high homology (82%) is seen between the predicted amino acid sequences of the two PHAs. The substitution of a serine residue at position 103 of *dlec1* for alanine in the corresponding position in *dlec2* creates a new glycosylation site (asparagine-x-threonine or serine) (Figure 3). Variation in the degree of glycosylation and in amino acid sequence could account for the different *in vitro* biological activities of PHA-E and PHA-L (Miller *et al.*, 1975).

Computer-assisted alignment of *dlec1* and *dlec2* sequences

© IRL Press Limited, Oxford, England.

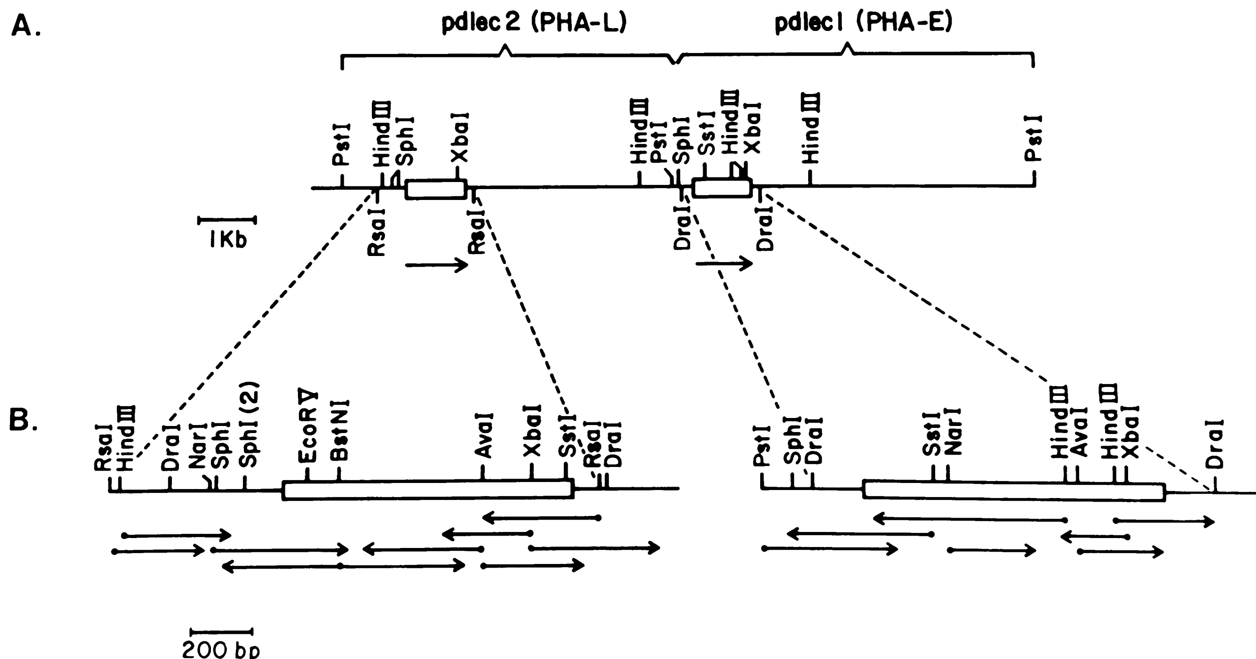

Fig. 1. Restriction maps of cloned lectin genes and their sequencing strategies. (A) The organization of *dlec1* and *dlec2* genes in phage λB10. Arrows denote the direction of transcription. (B) The sequencing strategies for the *Pst*I fragments subcloned into pBR322 to create pdlec1 and pdlec2. The arrows denote the extent and direction of sequence reading of each fragment.

demonstrates a 90% homology between their mRNA and their protein coding regions (Figure 2). Unexpectedly, the 3'-non-coding sequences are 84% conserved, with the homology rapidly diminishing ~20 bp 3' of the polyadenylation addition sites. When several gaps are introduced, a high degree of conservation is seen between 5'-flanking regions. Consensus eucaryotic promoter elements (TATA boxes) are found 22–27 bases upstream from the transcriptional start sites (Figure 2). A region resembling the canonical 'CCAAT' element is found 88 bp 5' to the capping site of *dlec1*, but the significance of such sequences to plant gene expression is not known.

Approximately 95 bp downstream from the protein terminator codons lie double (*dlec1*) or triple (*dlec2*) overlapping polyadenylation addition signals, denoted by asterisks in Figure 2. These sequences are precisely homologous to the AAUAAA consensus for eucaryotic messengers proposed by Proudfoot and Brownlee (1976).

Several direct and inverted repeats are found in each gene region (Figure 2). Of particular interest are the sequences from 10–29 and 74–93 of *dlec1*, which contain both inverted and direct repeats in the form of alternating pyrimidine and purine (R and Y) residues. Upstream from the protein coding domain of *dlec2* are two types of direct repeats, one of which (Figure 2, labeled a) occurs three times and follows a region of 9–17 alternating RYs in each case. A 17-base inverted repeat includes part of the 'TATA' box of *dlec2* (Figure 2, labeled 1). Both *Phaseolus* PHA genes and the soybean seed lectin gene contain the sequence ACCTAT flanked by inverted repeats in similar positions (Figure 2, labeled 2; Vodkin *et al.*, 1983). The sequence ACCTAT is the apparent site of insertion element integration into the soybean *Le1* gene.

Analysis of translation products of mRNA selected by lectin genomic clones

To confirm that our genomic clones were homologous to PHA mRNAs we used them to hybrid select poly(A) RNA from cotyledons, leaves and roots. Selected RNA was translated in an *in vitro* wheat germ system and the products were subjected to SDS-polyacrylamide gel electrophoresis and fluorography (Laemmli, 1970). Total translation products of cotyledon poly(A) RNA and brome mosaic virus (BMV) RNA were co-electrophoresed as migration standards (Figure 4A and B, lanes 1 and 2). *In vitro* PHA polypeptides in the standard lanes are marked with dots, and have mol. wts. of ~31 and 29 kd when compared with the migrations of BMV translation products. *dlec1* selected cotyledon mRNAs encoding proteins for both size classes of PHA (lane 3), but did not select detectable translatable RNA from leaf or root (lanes 4 and 5). *dlec2* selected a similar mRNA species from cotyledon RNA (Figure 4B, lane 3) and a much smaller quantity of an RNA from leaf which translates a 35-kd polypeptide (lane 4). In contrast, pλlec5.7, encoding a lectin-like gene (Hoffman, 1984), predominantly selects cotyledon mRNA that translates into a 27 kd-protein (Figure 4B, lane 6). pλlec5.7 also selects lower amounts of mRNAs of the major PHAs. These data demonstrate that pdlec1 and pdlec2 selectively hybridize cotyledon mRNAs for PHAs, and that pλlec5.7 encodes the mRNA of a PHA-like protein.

S1 nuclease mapping of transcriptional start sites of two lectin genes

The Weaver and Weissman (1979) modification of the S1 nuclease mapping procedure of Berk and Sharp (1977) was used to locate the transcription initiation sites of *dlec1* and *dlec2* . Restriction endonuclease sites for 5' end-labeling were ~200 bp from the initiation AUG codons of each clone. The results (Figure 5) indicate that there are several protected fragments for each gene. The majority of transcript initiation occurs only 11–14 bp upstream from the initiation AUG in a region encoding the same sequence in each gene (Figure 5A and B). The minor protected fragment bands may be due to multiple RNA capping positions or to end nibbling of the heteroduplex by S1 nuclease (Hoffman, 1984). A reproducibly lower quantity of S1-protected fragments

```
DLEC2 TTTTGTTAATAAAATGAAAGTTAATATTTATGGATTTATTTGAACATATTTACAAGTTTTTGTTAATAAATTTATTTACAGAAGTGTTAAATTTGAAAAT  100

      GTGTTAAATGAGAGTTAATATTTATGGATATAATATTACTTTATTTAAAAATAAAATGAATTTATATTAAAATATGATTTGGTAAAAGATATTATAAAGA  200

                                                                          DLEC1 ACTGGCCCTGCTGCATGC   18
      GGAACTTGGATGAGAAAGACGACGATGGGCAAAAGTTGCATGGCACACCGCTTTTGTTAGTAACATCTGCACTGTGGCGCCTTTTTCACACTCACCATGC  300

      ATG.................................CATGCAGCTGCCACCACAGCTTTCTCCTCTTTTCACTATGACACGACTGGCCATGCATGC   82
      ATGCTGCCACGTCAGCTTTCACACTATGACACCACTGGCCATACACGTTGCCACCTCAGCTTTCTCCTCTTCTCACTATGACACGACTGGCCATGCATGC  400

      ATGCATGCGGCCACCTCAGCTCCCTCCTCTTCACCCGTGTTTT...........CCAATGCTACGTGCCAACCGCTTCTCTCTATAAATATCTCTTTAAA  171
      ATGCATGCTGCCACCTCAGCTCCCGCCTCTTCACCCGTGTCTTTCTCTCATGTCGCAGTGCTGCGTGCCAACCTCTTCTCTCTATAAAT..........  489

      TTTAAACTAATTATTTCATATTTTTTCAATGTATTTGATGACGTGTATGCATTGCCATCGTTGCTTAATTGTTATTTCTATATTCTTATTACTCCCTCAA  271
      ....................................................................................................

      ATAATATTATAAAAGAAAAAAGAATCCATAAAGTAGAGAGAGAGAGTGCAGTTGTTGTTGTATAAATAGAGAAGAGAGTGATGGTTAATGCATGAATGCA  371
      ......................................................AGAGAGAAGAGAATGATGGTTAATGCATGAATGCA  524

      TACATGGCTTCCTCCAACTTACTCTCCCTAGCCCTCTTCCTTGTGCTTCTCACCCACGCAAACTCAGCCAGCCAAACCTCCTTCAGCTTCCAAAGGTTCA  471
      TACATGGCTTCCTCCAAGT...TCTTCACTGTCCTCTTCCTTGTGCTTCTCACCCACGCAAACTCAAGCAACGATATCTACTTCAACTTCCAAAGGTTCA  621

      ACGAAACCAACCTTATCCTCCAACGCGATGCCACCGTCTCATCCAAAGGCCAGTTACGACTAACCAATGTTAATGACAACGGAGAACCCACGTTGAGCTC  571
      ACGAAACCAACCTTATCCTCCAACGCGATGCCTCCGTCTCATCCTCCGGCCAGTTACGACTAACCAATCTTAATGGCAACGGAGAACCCAGGGTGGGCTC  721

      TCTGGGCCGTGCCTTCTACTCCGCCCCCATCCAAATCTGGGACAACACCACCGGCGCCGTGGCC.GCTTCGCCAACCTCCTTCACATTCAATATCGACGT  670
      TCTGGGCCGCGCCTTCTACTCCGCCCCCATCCAAATCTGGGACAACACCACCGGCACCGTGGCCAGCTTCGCC.ACCTCCTTCACATTCAATATACAGGT  820

      TCCCAACAATTCAGGACCCGCCGATGGCCTTGCCTTTGTTCTCCTCCCCGTGGGCTCTCAGCCCAAAGACAAAGGCGGTCTTCTAGGTCTGTTCAACAAC  770
      TCCCAACAATGCAGGACCCGCCGATGGACTTGCCTTTGCTCTCGTCCCCGTGGGCTCTCAGCCCAAAGACAAAGGGGGTTTTCTAGGTCTTTTCGACGGC  920

      TACAAATACGACAGCAATGCCCATACTGTGGCTGTGGAGTTCGACACCCTCTACAACGTTCACTGGGACCCCAAACCGCGTCATATTGGCATCGACGTGA  870
      AGCA......ACAGCAATTTCCATACTGTGGCTGTGGAGTTCGACACCCTCTACAACAAGGACTGGGACCCCACAGAGCGTCATATTGGCATCGACGTGA 1014

      ACTCCATCAAGTCTATCAAAACGACGACGTGGGATTTTGTCAAAGGAGAAAACGCGGAGGTTCTGATCACCTATGACTCCTCCACGAAGCTCTTGGTGGC  970
      ACTCCATCAGGTCTATCAAAACGACGCGGTGGGATTTTGTGAACGGAGAAAACGCCGAGGTTCTGATCACCTATGACTCCTCCACGAATCTCTTGGTGGC 1114

      TTCTCTGGTTTACCCTTCTCTGAAAACAAGCTTCATCGTCTCTGACACAGTGGACCTGAAGAGCGTTCTTCCCGAGTGGGTGATCGTTGGGTTCACTGCC 1070
      TTCTCTGGTTTACCCTTCTCAGAAAACGAGCTTCATCGTCTCTGACACAGTGGACCTGAAGAGCGTTCTTCCCGAGTGGGTGAGCGTTGGGTTCTCTGCC 1214

      ACCACTGGGATTACTAAAGGGAACGTTGAAACGAACGACATCCTCTCTTGGTCTTTTGCTTCCAAGCTCTCCGATGGCACCACATCTGAAGCTTTGAATC 1170
      ACAACTGGGATTAATAAAGGGAACGTTGAAACGAACGACGTCCTCTCTTGGTCTTTTGCTTCCAAGCTCTCCGATGGCACCACATCTGAAGGTTTGAATC 1314

      TTGCCAACTTCGCCCTCAACCAAATCCTCTAGACT.CAATCTCCACCTTCCCTGTGACACTAAATCACTGCTTACACTCAGTGTTCTTTCCTGCTAATAA 1269
      TCGCCAACTTGGTCCTCAACAAAATCCTCTAGACTCCAAACTCCACCTTCCCTGTGACACTAAAACACTGCTTTCACTCAGTGTTCTTTCCTGCTAATAA 1414

                                    *    *   *
      TGTTTCTCTTTGTCTGTCACAGAAAATAAAATA...AAATAGGGTGATGATAGCTTAC.......................................... 1324
      TGTTTC......TCTGTCACAGAAAATAAAATAAATAAAATGGGAGCTCATATATTA.TTATACTATTAAAAGGAAGCTGTCAAATGTTTGTGTTTTACT 1507

      ..........................................................ACTCAGTGTTCTTTGCTCTACTTCCTTAATTCCTTAA..... 1361
      CCAAACTTTGACTGTTAAAGAAAAGTTTTCCAAAATAGAGTTTACACGCGTTATTAGTACACAATG..................TAATTTTTTAAAAAGT 1589

      ..............................TGAAAATATAATAAAAAATAATAATTAAAATATT................ 1395
      TTTATAGTAATGAAAAGGATTATTCCAAGTTAAAGTATAACACA.......GATTAAAGTATTAACCATTAGAAAAAAA 1661
```

Fig. 2. Nucleotide sequences of *dlec1* and *dlec2*. Gaps have been introduced to maximize sequence homology. Potential promoter elements are doubly overlined and RY repeats are in cross-hatched areas. A bracket is indicated above the capping region determined by S1 nuclease mapping. Numbered horizontal arrows denote inverted repeats for *dlec1* (upper sequence) and *dlec2* (lower sequence). Lettered horizontal arrows show direct repeats for *dlec1* (upper sequence) and *dlec2* (lower sequence). Vertical dotted lines flank the mature lectin peptide coding region and the initiation ATG codon is denoted by a vertical arrow. Wavy lines are shown above sequences from which complementary 20-mer oligonucleotides were derived. Polyadenylation signals are marked with asterisks. Potential transposable element insertion sites are overlined with a broken line.

results from the hybridization of end-labeled *dlec2* to cotyledon mRNA. The *T*ms of the heteroduplexes involving *dlec1* or *dlec2* are nearly identical, indicating that the hybridization conditions are of equal stringency for both genes. The results suggest a lower steady-state level of *dlec2* than *dlec1* mRNA in bean cotyledons.

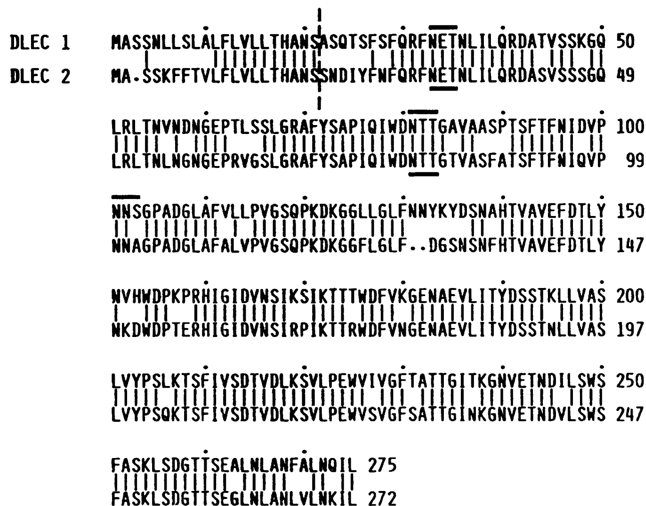

Fig. 3. Comparison of PHA-E (pdlec1) and PHA-L (pdlec2) polypeptides. Gaps have been introduced to maximize sequence homology. Glycosylation sites are underlined or overlined and a vertical dotted line separates the signal peptide from the mature proteins.

Hybridization of gene-specific oligonucleotides to lectin mRNAs

To demonstrate the transcription of *dlec1* and *dlec2* in cotyledon tissue it was necessary to obtain nucleic acid probes specific for each gene. The 3′-untranslated regions are 84% homologous and not useful as specific probes. Therefore, oligodeoxynucleotides corresponding to the coding regions shown in Figure 2 were synthesized. After 5′ end-labeling, the synthetic 20-mers were hybridized with Northern blots of cotyledon poly(A) RNA. The results indicate that each gene-specific 20-mer hybridized with a 1.0-kb poly(A) RNA species, the expected size for lectin mRNA (data not shown; Hoffman *et al.*, 1982). The specificity of the probes was demonstrated in control experiments in which the 20-mers were hybridized to Southern blots of three lectin or lectin-like gene subclones. The *dlec1* probe hybridized only with its respective gene, and the *dlec2* 20-mer cross-hybridized only to a very minor extent with *dlec1* (data not shown). These results, in addition to the S1 nuclease protection studies (Figure 5), indicate that the lectin genes contained by pdlec1 and pdlec2 are actively transcribed in cotyledon tissue.

Discussion

Comparison of pdlec1 and pdlec2 with a French bean lectin-like gene and with other plant multigene families

French bean lectins are encoded by a small multigene family of ~4–6 members (Hoffman, 1984). A related lectin-like gene, λlec5.7, has been isolated and shown to hyridize preferentially to a cotyledon mRNA species which translates into a 27-kd pro-

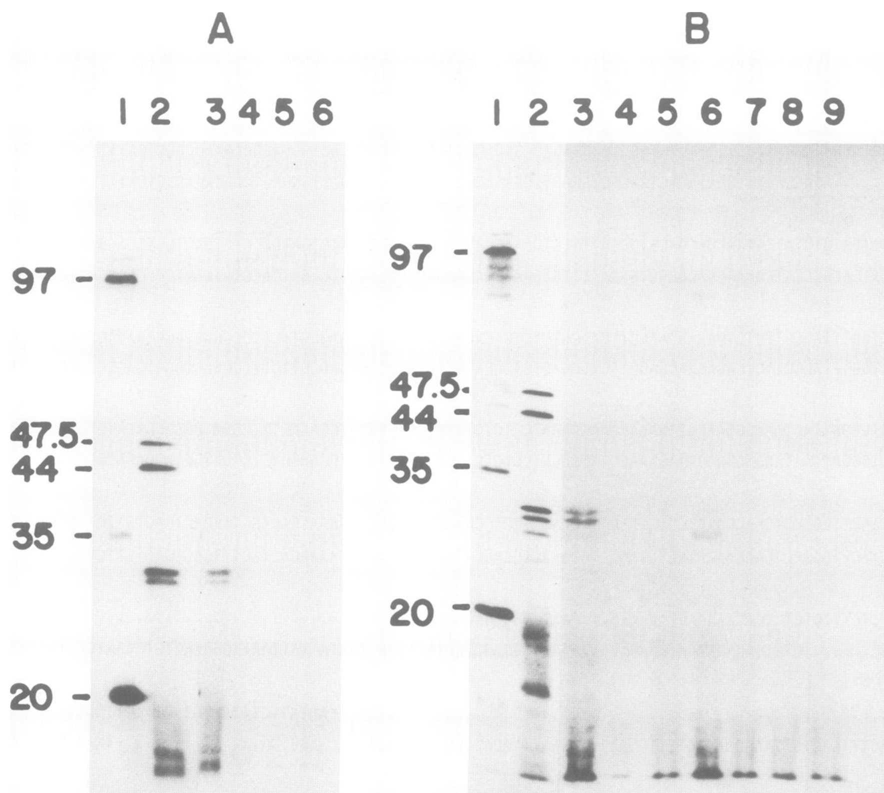

Fig. 4. *In vitro* translation products from mRNA hybrid-selected by pdlec1 and pdlec2. Linearized plasmid DNAs were bound to nitrocellulose filters which were hybridized with poly(A) RNA from *P. vulgaris* cotyledons, leaves or roots. After a succession of washes, hybridized RNA was eluted from the filters and translated in a wheat germ *in vitro* system. Translation products were electrophoresed in an SDS-polyacrylamide gel which was fluorographed. Dots indicate the position of PHA translation products. **(A)** Translation products selected by pdlec1 from poly(A) RNA of cotyledon **(lane 3)**, leaf **(lane 4)** and root **(lane 5)**. **Lanes 1** and **2** contain translation products of BMV RNAs 1–4 and total poly(A) cotyledon RNA, respectively. **Lane 6** shows endogenous translation products of the wheat germ system. **(B)** Translation products selected by pdlec2 **(lanes 3–5)** and pλlec5.7 **(lanes 6–8)** (Hoffman, 1984) from poly(A) RNA of cotyledon **(lanes 3** and **6)**, leaf **(lanes 4** and **7)** and root **(lanes 5** and **8)**. **Lanes 1** and **2** contain translation products of BMV RNAs 1–4 and total poly(A) cotyledon RNA, respectively. In **lane 9** are endogenous translation products of the wheat germ system.

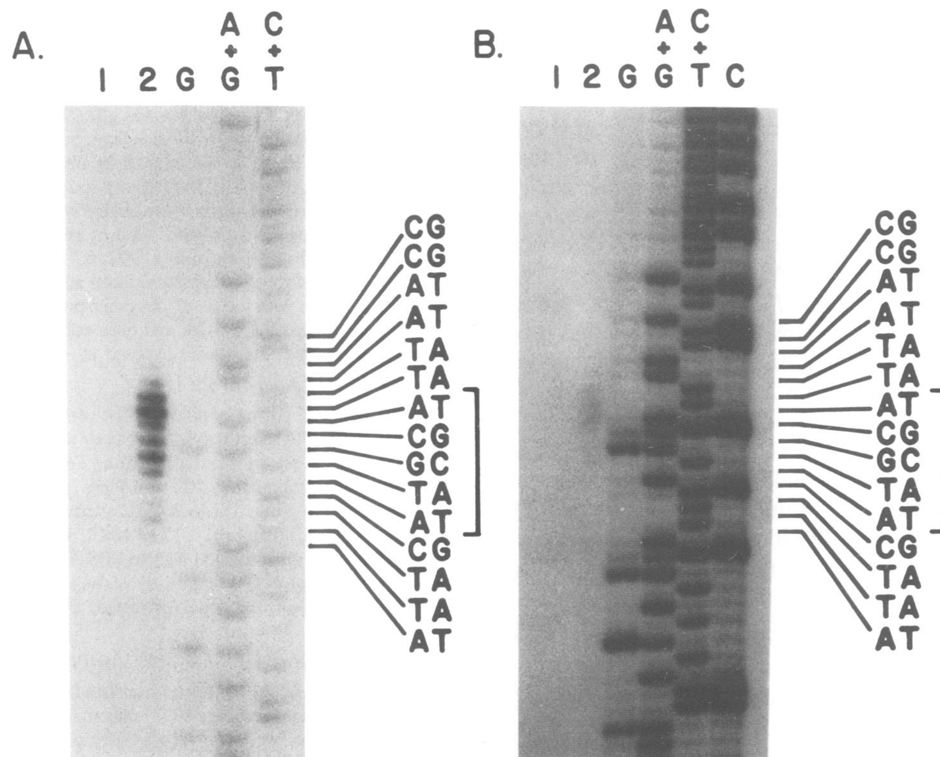

Fig. 5. S1 nuclease mapping of transcription initiation sites. 5′ end-labeled DNA probes were hybridized with RNAs and treated with S1 nuclease as described in Materials and methods. S1-protected fragments were electrophoresed in a 6% polyacrylamide sequencing gel alongside the sequence ladder of the same DNA fragments. Nucleotide sequences from the ladders are to the right of each figure, with the non-coding (+) strand on the right. **(A)** *Sst*I-*Rsa*I fragment of pdlec1 hybridized with 5 μg each tRNA **(lane 1)** or poly(A) cotyledon RNA **(lane 2)**. **(B)** *Bst*NI-*Sph*I fragment of pdlec2 hybridized with 5 μg each tRNA **(lane 1)** or poly(A) cotyledon RNA **(lane 2)**. Brackets denote major capping regions.

tein *in vitro* (Figure 4). The mRNA coding sequences of the 27-kd gene and *dlec1* are only 66.5% homologous, but their alignment increases to 82% if gaps are excluded (data not shown). Comparison of the 27-kd lectin-like protein with the PHA-E encoded by *dlec1* reveals a 43% peptide sequence homology (data not shown). Strong conservation of the 5′-flanking and transcribed regions suggests that *dlec1* and *dlec2* arose through a relatively recent duplication of a common ancestral PHA-like gene. The 27-kd gene may be related to the PHA genes through a much earlier duplication event.

The chromosomal proximity of *dlec1* and *dlec2* is in agreement with the studies of Brown *et al.* (1981) who demonstrated genetic linkage of seed lectins in French bean by analyzing the two-dimensional gel patterns of lectins from the progeny of crosses. This report demonstrates that members of seed protein multigene families in *Phaseolus* may be found in clusters similar to the *Glycine max* Kunitz trypsin inhibitor genes (Goldberg, 1983).

Members of several plant nuclear multigene families have been characterized at the DNA sequence level, including those encoding leghemoglobins (Brisson and Verma, 1982; Wiborg *et al.*, 1983), zein (Pedersen *et al.*, 1982), the small subunit of ribulosebisphosphate carboxylase (Coruzzi *et al.*, 1983, 1984), phaseolin Slightom *et al.*, 1983), and actin (Shah *et al.*, 1983). However, only in one case (Wiborg *et al.*, 1983) has there been a comparison of 5′-flanking sequences of several members of a gene family. Five leghemoglobin genes exhibited a degree of homology from the cap site to 120 bp upstream which was similar to that of the coding sequences (Wiborg *et al.*, 1983). If three gaps in the alignment are omitted from the region extending 177 bp upstream from the cap sites of *dlec1* and *dlec2*, there is a 93% conservation of sequence (Figure 2), comparable with the 90% homology of their coding sequences.

Potential transcription regulatory elements of lectin genes

It is notable that the transcription initiation sites of four legume lectin genes are contained within or closely flanked by RY repeats, including the common sequence ATGCAT (Figure 2; Hoffman, 1984; Vodkin *et al.*, 1983). RY repeats may promote the formation of Z-DNA under certain conditions of superhelical density (Nordheim and Rich, 1983). Kilpatrick *et al.* (1984) have recently shown that DNA near the junction of RY repeats and non-RY repeat regions may be selectively sensitive to S1 nuclease. We postulate that Z-DNA formation near the cap site of lectin genes might increase the accessibility of the transcription initiation region to RNA polymerase.

RY repeats up to 18 bp long are found in the 5′-flanking domains of *dlec1* and *dlec2*, the 27-kd gene and the soybean *Le1* gene, but not within protein coding regions. They are also present within inverted repeats 5′ to the wheat and pea genes encoding the small subunit of ribulosebisphosphate carboxylase (Rubisco) (Coruzzi *et al.*, 1984). The repeats of the Rubisco gene upstream region are adjacent to sequences with homology to eucaryotic and viral enhancer elements (Weiher *et al.*, 1983). No enhancer consensus sequences are found flanking the lectin genes *dlec1* and *dlec2*.

Phaseolus lectin genes have multiple ATG codons in their 5′-untranslated regions

Each of three lectin or lectin-like genes analyzed has at least two ATG codons in a short (14 – 16 bp) 5′-untranslated region

(Figure 2; Hoffman, 1984). *dlec1* and *dlec2* have ATG codons out of phase with lectin-coding sequences located three nucleotides from the cap site. Eucaryotic ribosomes can recognize an AUG three nucleotides from the cap site, but in such cases translation initiation is not restricted to the 5′-proximal site (Kelley *et al.*, 1982). Kozak (1981, 1983) has postulated that the sequence context of AUG codons is important for their ability to initiate. In particular, the first AUG of a mRNA is translated if it resides in the consensus sequence ANNAUGN or GNNAUGR. The second and third ATG codons (TGAATGC and TACATGG, respectively) of *dlec1* and *dlec2* are not flanked by consensus initiator sequences. However, the third and proper initiator region is similar to that of at least 11 functional initiators (Kozak, 1983). Thus, the first ATG of PHA genes is too close to the cap site to function as an initiator, the second ATG is surrounded by unfavorable sequences, and the third methionine codon, flanked by appropriate nucleotides, is the initiator of translation. *In vitro* translation of lectin mRNA hybrid selected by *dlec1* and *dlec2* and the 27-kd gene indicates that their respective transcripts are translated despite multiple upstream AUGs (Figure 4).

Phaseolus lectin genes contain potential sites for transposable element insertion

In several cultivars of soybean the presence of a 3.4-kb insertion element in the the *Le1* gene is associated with loss of expression of seed lectin (Vodkin *et al.*, 1983). Several *Phaseolus* cultivars contain little or no PHA, and one of them, Great Northern, appears by Southern blot analysis to possess a lectin gene copy number similar to Tendergreen (unpublished data). The three Tendergreen PHA or PHA-like genes as well as a pea lectin cDNA contain sequences identical to the insertion site of a mobile element in the *Le1* soybean gene (Figure 2; Hoffman, 1984; Higgins *et al.*, 1983; Vodkin *et al.*, 1983). In each case the sequence ACCTAT is closely flanked by imperfect inverted repeats. It is possible that such a structural feature is present in each *Phaseolus* lectin gene and that it serves to target transposable elements. The probability of multiple insertional inactivation events in the lectin genes of a lectinless cultivar is low, but the presence of conserved 'hotspots' might increase transposition frequency (Halling and Kleckner, 1982).

Materials and methods

Isolation of Phaseolus cotyledon poly(A) RNA

Total RNA was isolated from mid-maturation cotyledons of *P. vulgaris* cv. 'Tendergreen' by the method of Murray *et al.* (1983). Poly(A)-containing RNA was isolated by poly(U)-Sephadex chromatography as previously described (Murray *et al.*, 1981).

S1 nuclease mapping of dlec1 and dlec2

Determinations of transcription initiation sites for *dlec1* and *dlec2* mRNAs were performed by the method of Weaver and Weissmann (1979). For *dlec1*, a 840-bp *SstI-RsaI* fragment labeled at the 5′ end of the *SstI* site was denatured and hybridized with 5 μg of Tendergreen cotyledon total poly(A) RNA or tRNA for 3 h at 52°C. For *dlec2*, a 408-bp *BstNI-SphI* fragment labeled at the 5′ end of the *BstNI* site was denatured and hybridized under identical conditions. After hybridization each sample was treated with S1 nuclease, phenol extracted, ethanol precipitated and analyzed on sequencing gels as previously described (Hoffman, 1984).

DNA sequence analysis

DNA sequencing was performed by a modification of the method of Maxam and Gilbert (1980) as previously described (Hoffman *et al.*, 1982). Sequence data was analyzed by programs of the University of Wisconsin Genetics Computer Group.

Oligonucleotide synthesis

DNA oligomers of 20 nucleotides each, with sequences uniquely homologous to portions of either *dlec1* or *dlec2*, were synthesized on an Applied Biosystems 370A Automatic DNA Synthesizer and purified by h.p.l.c.

Northern blot hybridization

Poly(A) RNA from Tendergreen cotyledons was electrophoresed in a 1.2% agarose/2.2 M formaldehyde gel and transferred to nitrocellulose filters with 20 x SSC (1 x SSC = 0.15 M NaCl/0.015 M Na-citrate) (Hoffman *et al.*, 1981). Filters were pre-hybridized in 6 x SSC, 0.1% SDS, and Denhardt's solution (1 x Denhardt's solution = 0.02% each ficol, bovine serum albumin and polyvinyl pyrrolidone) for at least 3 h. Overnight hybridizations were performed in the same solution with the addition of 200 000 c.p.m./ml of end-labeled *dlec1* or *dlec2* oligomer. Pre-hybridization and hybridization were at 25°C for experiments using *dlec1* and 42°C for *dlec2*. Washes after hybridization with the *dlec1* oligomer probe were for 30 min each, in 6 x SSC, 0.1% SDS and Denhardt's solution, performed twice at room temperature, once at 42°C and once at 55°C. Washes after hybridization with the *dlec2* oligomer probe were for 30 min each, twice at 42°C in 6 x SSC, 0.1% SDS, and once each at 42°C and 55°C in 2 x SSC, 0.1% SDS. Washed filters were exposed at −80°C with Kodak XAR-5 film.

Hybrid selection of poly(A) RNAs

20 μg of denatured pdlec1 or pdlec2 DNA was bound to a 3 mm circle of nitrocellulose and used to select poly(A) RNAs by the method of Parnes *et al.* (1981). Hybrid selection was performed using 10 μg Tendergreen poly(A) RNA in 100 μl of 50% deionized formamide, 100 mM Pipes (pH 6.4), 0.1% SDS, 0.6 M NaCl, 1 mM EDTA, 1 mg/ml poly(A) at 45°C overnight. Filters were washed 10 times in 10 mM Tris-HCl (pH 7.6), 0.15 M NaCl, 1 mM EDTA, 0.5% SDS at 65°C then 5 times in the same solution without SDS. Poly(A) RNA was eluted by boiling for 1 min in 300 μl of 0.06 μg/ml tRNA and snap freezing in dry ice/ethanol. The poly(A) RNA was phenol extracted, chloroform extracted and ethanol precipitated.

In vitro translation of hybrid selected poly(A) RNAs

Hybrid-selected poly(A) RNA was translated *in vitro* in a wheat germ cell-free extract (Davies and Kaesberg, 1973) containing [^{3}H]leucine as the label. After translation, proteins were analyzed by electrophoresis on 15% polyacrylamide gels containing SDS according to the method of Laemmli (1970). Gels were fluorographed after treatment with EN^{3}HANCE as specified by the manufacturer.

Materials

Restriction endonucleases were purchased from Promega Biotec, BRL, or New England Biolabs. T4 polynucleotide kinase was from P-L Biochemicals; calf intestinal alkaline phosphatase from Boehringer-Mannheim; S1 nuclease was from Sigma, and [γ-^{32}P]ATP, [^{3}H]leucine and EN^{3}HANCE were obtained from New England Nuclear. Poly(U)-Sephadex was from BRL and BA-85 nitrocellulose filter was from Schleicher and Schuell.

Acknowledgements

We thank Dr Jerry Slightom for providing a λ1059 library of French bean DNA and acknowledge Kenneth Idler and David Thompson for expert technical assistance in DNA sequencing. This is manuscript no. 33 from Agrigenetics Advanced Research Division.

References

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Borrebaeck,C.A.K. (1984) *Planta*, **161**, 223-228.
Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055-4059.
Brown,J.W.S., Bliss,F.A. and Hall,T.C. (1981) *Theor. Appl. Genet.*, **60**, 251-259.
Coruzzi,G., Broglie,R., Cashmore,A.R. and Chua,N.-H. (1983) *J. Biol. Chem.*, **258**, 1399-1402.
Coruzzi,G., Broglie,R., Edwards,C. and Chua,N.-H. (1984) *EMBO J.*, **3**, 1671-1679.
Cunningham,B.A., Wang,J.L., Waxdal,M.J. and Edelman,G.M. (1975) *J. Biol. Chem.*, **250**, 1503-1512.
Davies,J.W. and Kaesberg,P. (1973) *J. Virol.*, **12**, 1434-1441.
Foriers,A., Lebrun,E., VanRapenbusch,R., deNeve,R. and Strosberg,A.D. (1981) *J. Biol. Chem.*, **256**, 5550-5560.
Goldberg,R.B. (1983) in Owens,L.D. (ed.), *Genetic Engineering: Applications to Agriculture, Beltsville Symposium*, Vol. **7**, Rowman and Allanheld, Totowa, NJ, pp. 137-150.
Goldstein,I.J., Hughes,R.C., Monsigny,M., Osawa,T. and Sharon,N. (1980) *Nature*, **285**, 66.
Halling,S.M. and Kleckner,N. (1982) *Cell*, **28**, 155-163.
Higgins,T.J.V., Chandler,P.M., Zurawski,G., Button,S.C. and Spencer,D. (1983) *J. Biol. Chem.*, **258**, 9544-9549.
Hoffman,L.M. (1984) *J. Mol. Appl. Genet.*, **2**, 447-453.
Hoffman,L.M., Fritsch,M.K. and Gorski,J. (1981) *J. Biol. Chem.*, **256**, 2597-2600.
Hoffman,L.M., Ma,Y. and Barker,R.F. (1982) *Nucleic Acids Res.*, **10**, 7819-7828.

Hopp,T.P., Hemperly,J.J. and Cunningham,B.A. (1982) *J. Biol. Chem.*, **257**, 4473-4483.
Kelley,D., Coleclough,C. and Perry,R.P. (1982) *Cell*, **29**, 681-689.
Kilpatrick,M.W., Klysik,J., Singleton,C.K., Zanling,D.A., Jovin,T.W., Hanan, L.H., Erlanger,B.F. and Wells,R.D. (1984) *J. Biol. Chem.*, **259**, 7268-7274.
Kozak,M. (1981) *Nucleic Acids Res.*, **9**, 5233-5252.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1-45.
Laemmli,U.K. (1970) *Nature*, **227**, 680-685.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Miller,J.B., Hsu,R., Heinrikson,R. and Yachnin,S. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 1388-1391.
Murray,M.G., Peters,D.L. and Thompson,W.F. (1981) *J. Mol. Evol.*, **17**, 31-42.
Murray,M.G., Hoffman,L.M. and Jarvis,N.P. (1983) *Plant Mol. Biol.*, **2**, 075-084.
Nordheim,A. and Rich,A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1821-1825.
Parnes,J.R., Valan,B., Felsenfeld,A., Ramanathan,L., Ferrini,U., Appela,E. and Seidman,J.D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2253-2257.
Pedersen,K., Devereux,J., Wilson,D.R., Shelton,E. and Larkins,B.A. (1982) *Cell*, **29**, 1015-1026.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Shah,D.M., Hightower,R.C. and Meagher,R.B. (1983) *J. Mol. Appl. Genet.*, **2**, 111-126.
Slightom,J.L., Sun,S.M. and Hall,T.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1897-1901.
Talbot,C.F. and Etzler,M.E. (1978) *Biochemistry (Wash.)*, **17**, 1471-1479.
Toms,G.C. and Western,A. (1971) *Chemotaxonomy of the Leguminosae*, published by Academic Press, NY.
Vodkin,L.O., Rhodes,P.R. and Goldberg,R.B. (1983) *Cell*, **34**, 1023-1031.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.
Weiher,H., Konig,M. and Gruss,P. (1983) *Science (Wash.)*, **219**, 626-631.
Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E.O., Paludan,K. and Marcker,K.A. (1983) *EMBO J.*, **2**, 449-452.

Received on 17 October 1984; revised on 1 February 1985