The EMBO Journal vol.5 no.12 pp.3075-3082, 1986

Molecular analysis of two phytohemagglutinin genes and their expression in *Phaseolus vulgaris* cv. Pinto, a lectin-deficient cultivar of the bean

Toni A.Voelker, Paul Staswick[1] and Maarten J.Chrispeels

Department of Biology C-016, University of California, San Diego, La Jolla, CA 92093, USA

[1]Present address: Department of Agronomy, University of Nebraska, Lincoln, NE 68583, USA

Communicated by T.A.Bickle

Phytohemagglutinin (PHA), the seed lectin of the common bean, *Phaseolus vulgaris*, is encoded by two highly homologous, tandemly linked genes, *dlec*1 and *dlec*2, which are coordinately expressed at high levels in developing cotyledons. Their respective transcripts translate into closely related polypeptides, PHA-E and PHA-L, constituents of the tetrameric lectin which accumulates at high levels in developing seeds. In the bean cultivar Pinto UI111, PHA-E is not detectable, and PHA-L accumulates at very reduced levels. To investigate the cause of the Pinto phenotype, we cloned and sequenced the two PHA genes of Pinto, called *Pdlec*1 and *Pdlec*2, and determined the abundance of their respective mRNAs in developing cotyledons. Both genes are more than 90% homologous to the normal PHA genes found in other cultivars. *Pdlec*1 carries a 1-bp frameshift mutation close to the 5′ end of its coding sequence. Only very truncated polypeptides could be made from its mRNA. The gene *Pdlec*2 encodes a polypeptide, which resembles PHA-L and its predicted amino acid sequence agrees with the available Pinto PHA amino acid sequence data. Analysis of the mRNA of developing cotyledons revealed that the *Pdlec*1 message is reduced 600-fold, and *Pdlec*2 mRNA is reduced 20-fold with respect to mRNA levels in normal cultivars. A comparison of the sequences which are upstream from the coding sequence shows that *Pdlec*2 has a 100-bp deletion compared to the other genes (*dlec*1, *dlec*2 and *Pdlec*1). This deletion which contains a large tandem repeat may be responsible for the low level of expression of *Pdlec*2. The very low expression of *Pdlec*1 is as yet unexplained .

Key words: phytohemagglutinin genes/*Phaseolus vulgaris*/*dlec*1 and 2/*Pdlec*1 and 2

Introduction

Several cultivars of legume crops have been identified in which particular proteins or polypeptides of the storage proteins or lectins are absent. In these cultivars different types of mutations have been identified: gene deletions (Ladin *et al.*, 1984), frame degeneration (Brown *et al.*, 1985), and transposition insertion (Rhodes and Vodkin, 1985). Certain cultivars of the common bean, *Phaseolus vulgaris*, are shown to be deficient in the seed lectin phytohemagglutinin (PHA). Here, we present a molecular analysis of the genes coding for PHA in the lectin deficient cultivar Pinto UI111 (hereafter called Pinto).

PHA, a lectin, is the second most abundant protein in the seeds of the common bean; it accounts for 5-10% of the total protein content of the mature seeds of most cultivars. Two tandemly linked genes, *dlec*1 and *dlec*2, encode PHA (Hoffman and Donaldson, 1985). The expression of both genes is highly regulated in developing bean cotyledons (Staswick and Chrispeels, 1984; Chappell and Chrispeels, 1986). Though they are >90% homologous, the two gene products differ in biological properties. The gene *dlec*1 codes for PHA-E, which confers erythroagglutinating activity to the tetrameric protein, whereas the gene product of *dlec*2, PHA-L, makes the tetramer leucoagglutinating and mitogenic (Miller *et al.*, 1975). Both subunits are present in comparable levels in mature beans.

Brown *et al.* (1981) found that a seed extract of the Pinto cultivar failed to agglutinate red blood cells and had very little or no PHA protein. Pusztai *et al.* (1981) showed that this cultivar contains a PHA-L-like protein at a level 50-100 times lower than the total PHA level of most other cultivars. Vitale *et al.* (1985) showed Pinto PHA to be mitogenic but not erythroagglutinating and Northern analysis revealed a 50- to 100-fold reduced abundance of PHA-specific mRNA, while levels of mRNA for other seed proteins were normal in Pinto (Staswick and Chrispeels, 1984). We set out to find the genetic structures responsible for this phenotype by analyzing the Pinto genome with respect to PHA genes. By comparing their structure with the known PHA high expression genes of the other cultivars, we hoped to identify the elements essential for the high expression phenotype. Here, we present the nucleotide sequences of two PHA genes of Pinto. One gene, called *Pdlec*2, has an open reading frame which encodes PHA-L protein, while the other gene, *Pdlec*1, is a pseudogene with a frameshift mutation early in the coding sequence. Using gene-specific probes for Northern analysis, we found strikingly different reductions of mRNA levels for each Pinto PHA gene.

Results

Characterization of the genomic clones of Pinto DNA

Southern hybridization analysis of Pinto DNA restricted with

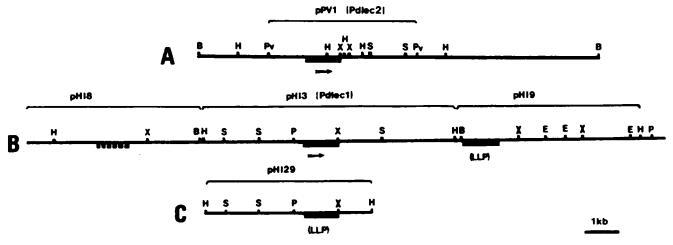
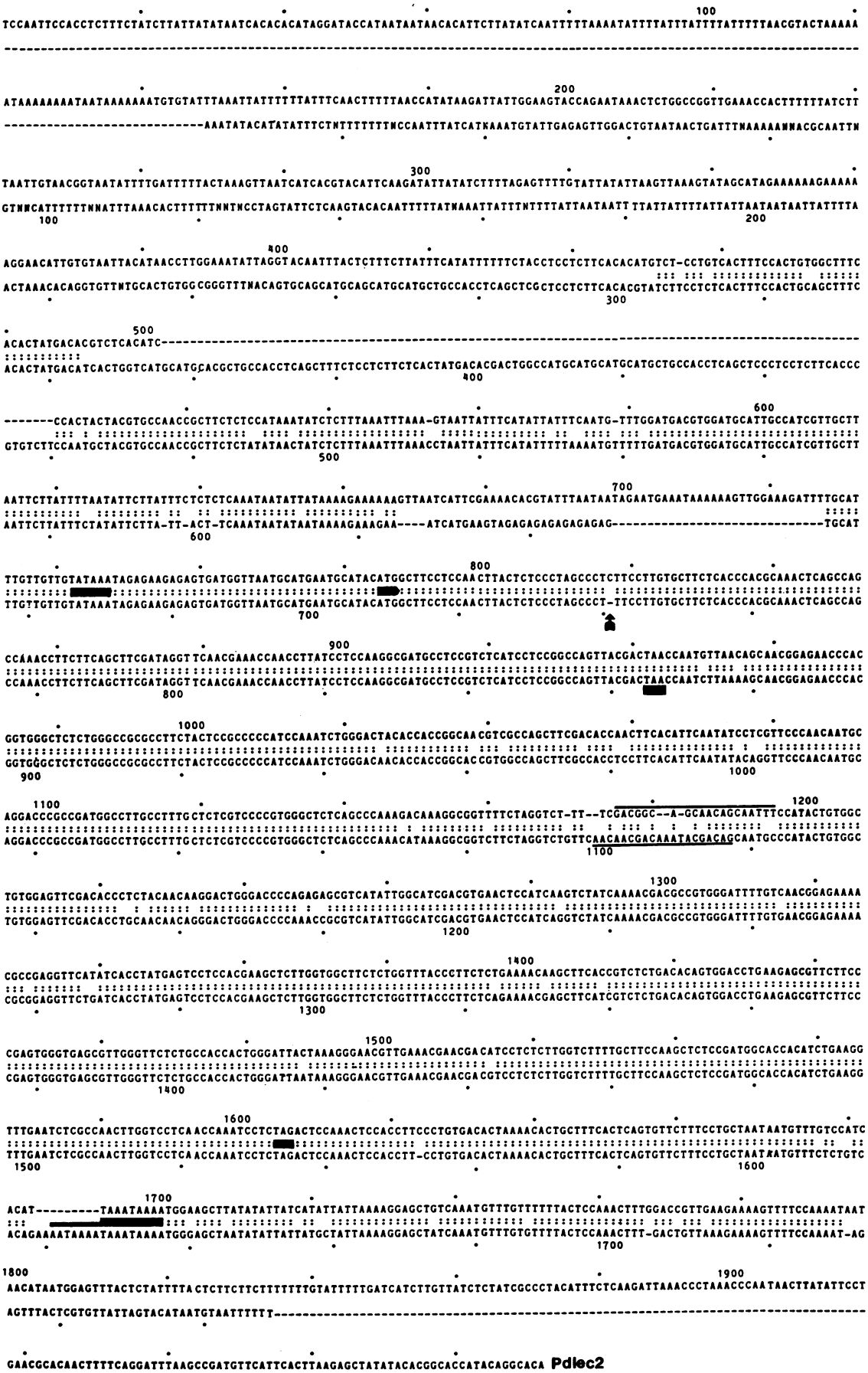


Fig. 1. Partial restriction enzyme maps of cloned genes for PHA and lectin-like protein of Pinto. Heavy lines indicate regions homologous to pSC1, a cDNA clone for PHA, or to pPVL134, a cDNA clone of a lectin-like protein (LLP). The region indicated by an interrupted broad line in pHI8 hybridizes weakly to pSC1 and pPVL134. (**A**) A 12-kb *Bam*HI fragment cloned in λ EMBL 4 containing the *Pdlec*2 gene. Arrow indicates transcription orientation. (**B**) A 19-kb fragment cloned in a λ Charon 35 phage containing *Pdlec*1. Arrow indicates transcription orientation. (**C**) Part of a 13-kb Charon 35 insert containing one possible LLP gene. Symbols are: B, *Bam*HI: H, *Hin*dIII; Pv, *Pvu*II; X, *Xba*I; S, *Sst*I; E, *Eco*RI. Brackets indicate the extent of pUC subclones.


Pdiec2
Pdiec1

several restriction enzymes, probed at high stringency with pSC1 and pSC2 (cDNA clones derived from PHA-E, respectively pHA-L mRNA; Staswick and Chrispeels, 1984) indicated that PHA-related sequences are present in the genome (Vitale *et al.*, 1985; Staswick *et al.*, 1986). By rehybridizing the filters with pPVL134, a cDNA clone specific for a lectin-like gene in *P. vulgaris* (Hoffman, 1984), we obtained several new bands, indicating that Pinto also contains genes for these lectin-like proteins. The relative mobilities of the various cross-hybridizing restriction fragments differ between Pinto and the high expression cultivar Greensleeves, indicating structural differences in their genes. It is important to note that Pinto is not a mutant of Greensleeves, but that both were independently domesticated.

We used pSC2 to identify homologous sequences in a λ library of Pinto genomic DNA. Analysis of an EMBL4 phage λ isolate containing a 12-kb *Bam*HI insert showed that one PHA gene was contained in the *Pvu*II fragment subcloned in pPV1 (Figure 1A). Additional clones with different characteristics were isolated from a Pinto total genomic library in Charon 35 phage λ. One contained an insert of 19 kb which yielded three *Hin*dIII fragments of interest. The respective restriction enzyme map is shown in Figure 1B. pSC1 hybridized most strongly to the 7.5-kb central fragment subcloned in pHI3, and pPVL134 was most homologous to the 5.3-kb fragment (pHI9). The 4.3-kb fragment (pHI8) gave a weak signal with both probes. The other isolate had an insert of 13 kb which contained a 5.0-kb *Hin*dIII fragment (pHI29) that hybridized to pPVL134. A restriction map of this fragment is shown in Figure 1C. Considerable homology of pHI29 with the 7.5-kb *Hin*dIII fragment pHI3 is evident. The apparent co-migration of the fragments with the highest homology to PHA genes in genomic digests and in the clones indicated to us that we had obtained the PHA genes of Pinto.

Sequence of Pinto PHA genes

Since differential hybridization with PHA-cDNA probes revealed the fragments pPV1 and pHI3 to be the likeliest candidates to contain PHA genes, we sequenced about 2000 bp of each. The results are shown in Figure 2. The sequence obtained from pHI3 contains a gene which is for about 1300 bp >90% homologous to the two known PHA genes of the cultivar Tendergreen (Hoffman and Donaldson, 1985), but slightly more related to *dlec*1 than to *dlec*2 (see below). It is therefore labeled Pinto *dlec*1 or *Pdlec*1. The nucleotide sequence derived from clone pPV1 again shares about 1200 bp with the known PHA genes but resembles *dlec*2 more than *dlec*1 and was therefore labelled *Pdlec*2. The *Pdlec*1 and *Pdlec*2 coding areas (Figure 2) have nearly identical sequences with one significant exception (from bp 1100 to 1300 for *Pdlec*1), where in a stretch of 20 bp no homology between the genes exists; in this region, *Pdlec*2 has a 6-bp deletion with respect to *Pdlec*1. In contrast, this stretch of 20-bp is shared completely between *Pdlec*2 and the *dlec*2 gene of the cultivar Tendergreen. *Pdlec*1 and *dlec*1 from the cultivar Tendergreen differ by only 1 bp in this area. This is one of the reasons we classified the PHA gene on pPV1 as *Pdlec*2 and the gene on pH13 on *Pdlec*1. Outside of the coding area, both genes are almost

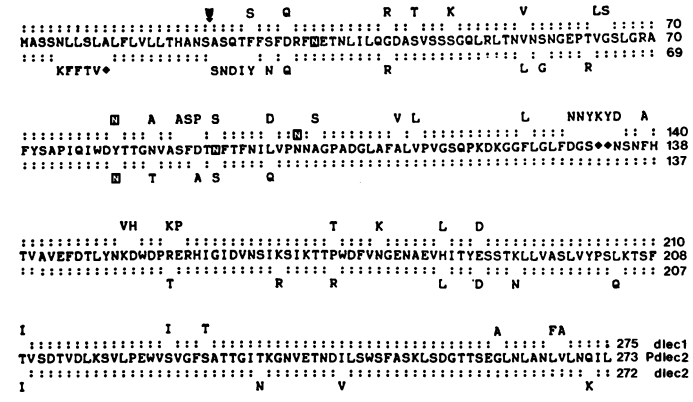


Fig. 3. A comparison of the derived amino acid sequence of *Pdlec*2 from *P. vulgaris* Pinto (middle) with the derived amino acid sequences of PHA-E (top) and PHA-L (bottom) of the cultivar Tendergreen (Hoffman and Donaldson, 1985). Double dots indicate identity with the *Pdlec*2 sequence; diamonds indicate amino acid deletions. Slim arrow: cleavage site of the signal peptide as determined by amino acid sequencing of the N-terminal ends of PHA-E and PHA-L (Miller *et al.*, 1975) and the PHA-L isolated from Pinto (Strosberg *et al.*, 1983). Asparagine residues in negative print are potential glycosylation sites.

identical for another 190 bp 3′ of the PHA protein termination codon, including triple (*Pdlec*1) or single (*Pdlec*2) poly(A) addition sites (Proudfoot and Brownlee, 1976). Upstream of the coding areas, both genes show stretches of extensive homology for several hundred basepairs. No significant homology is found further upstream where both isolates have different AT-rich sequences. TATA boxes are found 40 bp upstream of the proposed translation start codons.

The coding frame of Pdlec1 is interrupted

Taking the results from the analysis of the very homologous known PHA genes of Tendergreen as reference (Hoffman and Donaldson, 1985), a 1.1-kb-long poly(A)$^+$ mRNA can be predicted from the *Pdlec*1 gene of Pinto. None of the three possible reading frames allow the translation of its message into a complete PHA polypeptide. The PHA translational start codon, as assigned by Hoffman and Donaldson (1985), is indicated by the horizontal arrow after bp 708 of *Pdlec*1 (Figure 2). It starts a reading frame which codes for a polypeptide with an N-terminal sequence completely identical to the *dlec*1 gene product, but is out of register after 10 codons due to a single basepair deletion (see vertical arrow after bp 740 in Figure 2). This reading frame is terminated by UAA (bar below *Pdlec*1 sequence after bp 864) and the predicted truncated 50-amino acid polypeptide has no resemblance to any PHA sequence for its 40 C-terminal residues. The predicted polypeptide of 50 amino acids shares only its 10 N-terminal amino acids with PHA. Inserting a single nucleotide pair between positions 740 and 741 would restore the complete *Pdlec*1 reading frame. A polypeptide of 275 amino acids can be predicted, which would be most similar to PHA-E in its sequence (not shown). A second alternative reading frame, −1 in register relative to the proposed initiating AUG described above, is open

Fig. 2. Comparison of the nucleotide sequences of the phytohemagglutinin genes *Pdlec*1 (lower strand) and *Pdlec*2 (upper strand) of Pinto. DNA fragments from pPV1 and pHI3 (Figure 1) clones were subcloned in M13 vectors of the mp series and sequenced by the dideoxy method. Both strands were sequenced between bp 450 and 1710 for *Pdlec*2 and between bp 620 and 1540 (bp 1090−1265 excepted) for *dlec*1. Gaps have been introduced for maximal alignment of the two sequences. Consensus TATA boxes and polyadenylation sites are indicated by horizontal bars. The proposed translational start codon (ATG) and termination codon (TAG) for PHA (Hoffman and Donaldson, 1985) are shown by short horizontal bars between the sequences. The vertical arrow points to the single basepair deletion which interrupts the coding frame of *Pdlec*1 and the resulting stop codon TAA is indicated by a short horizontal bar (bp 865−867). Thin lines above and below the sequences indicate the sequences to which complementary 20-bp oligonucleotides were synthesized and used in Northern analysis as gene specific probes (see Figure 5).

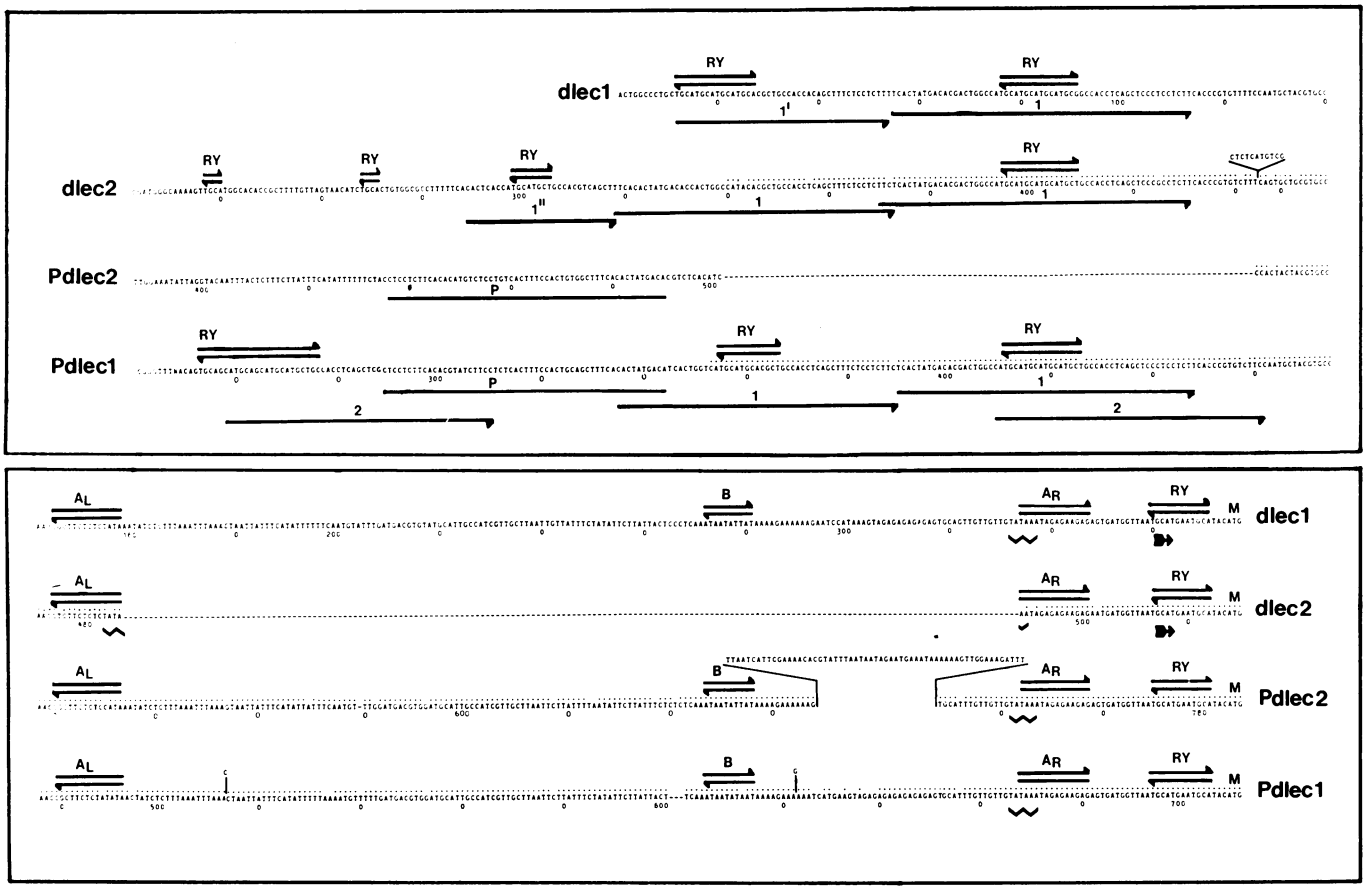


Fig. 4. Comparison of nucleotide sequences 5′ of the coding region of the PHA genes of Tendergreen and Pinto. Each of the four sequences begins on the left and in the upper box and ends with the ATG (also marked by M) translational start codon at the right in the lower box. Deletions and insertions are introduced to maximize homology (DNA sequences *dlec*1 and *dlec*2 are from Hoffman and Donaldson, 1985). All sequences are compared to *dlec*1 at the top and nucleotide identity is indicated by a dot above the symbol. Horizontal arrows in *dlec*1 and *dlec*2 indicate the transcriptional start sites as determined by S1 protection experiments of Hoffman and Donaldson, 1985. TATA boxes are enhanced by zigzag underlining. Areas of direct and inverted repeats are indicated by thin arrows and double arrows, respectively.

and a PHA-sized polypeptide can be predicted. This polypeptide does not share any resemblance with PHA at its N terminus, but it translates from the PHA reading frame after the single basepair deletion at position 740. Its only possible initiating AUG codon resides directly at the putative cap site of the *Pdlec*1 transcript (see below) and is not surrounded by sequences found to be favorable for a translation initiation [Kozak, 1986; see below and Hoffman and Donaldson (1985) for a more detailed discussion]. It is unlikely that this frame can be used efficiently *in vivo*. We, therefore, conclude that the single frameshift mutation after bp 740 causes the absence of PHA-E subunits in the Pinto cultivar even though some mRNA from this gene can be found (see below).

Pdlec2 codes for a PHA-L polypeptide

In contrast to *Pdlec*1, the reading frame of *Pdlec*2 is intact and a polypeptide of 273 amino acids very homologous to the known PHA polypeptide can be predicted (Figure 3). Several lines of evidence indicate that *Pdlec*2 is expressed in Pinto and its gene product represents the small amount of PHA found in the cotyledons of this cultivar. PHA was isolated from Pinto seeds by Pusztai *et al.* (1981), and Strosberg *et al.*, (1983) reported the sequence of the 23 amino acids at the N terminus of the Pinto PHA. A comparison of this amino acid sequence with our predicted *Pdlec*2 gene product sequence shows identity from position 22 to 44 with the exception of position 30. Our data predict an aspartic acid residue, while the amino acid sequencing shows an asparagine at this position. The first N-terminal 21 amino acids of the *Pdlec*2 gene product, not found in the mature PHA of Pinto, represent its signal peptide, which is identical to the signal peptide of PHA-E and follows the canonical sequence requirements found in signal peptides of higher eucaryotes (von Heijne, 1985). The predicted sequence of the *Pdlec*2 protein shows two putative glycosylation sites (white letter N in black square in Figure 3). The N-terminal site is positioned at residue 33, the homologous position of the glycosylation site of normal PHA-E and PHA-L shown to carry a high-mannose oligosaccharide in the mature protein (Sturm and Chrispeels, 1986). The second glycosylation site of PHA, which carries the complex oligosaccharide (Sturm and Chrispeels, 1986), is at Asn81 for PHA-E and PHA-L. This site is not available in the gene product of *Pdlec*2 due to a single basepair substitution. In contrast, a different site, 11 amino acid residues more C-terminal and not present in normal PHA-E and PHA-L, exists at position 92 of the gene product of *Pdlec*2 (Figure 3).

The properties predicted from the DNA sequence of *Pdlec*2 are in agreement with the observations of Vitale *et al.* (1985) that Pinto PHA is synthesized on polysomes attached to the rough ER and co-translationally glycosylated with two oligosaccharide chains per polypeptide. Furthermore, Pinto PHA co-migrates in SDS−polyacrylamide electrophoresis with PHA-L (Pusztai *et al.*, 1981; Staswick and Chrispeels, 1984; Vitale *et al.*, 1985). Pin-

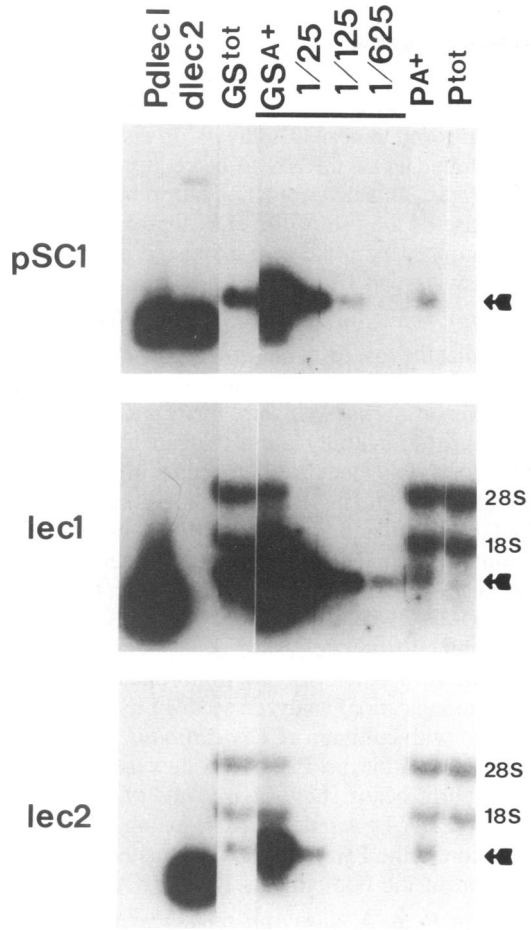


Fig. 5. Comparison of PHA mRNA levels in developing cotyledons of Greensleeves and Pinto. Total cytoplasmic RNA was isolated from developing cotyledons of Greensleeves and Pinto (size stages 3–5 according to Chappell and Chrispeels, 1986) as described in Materials and methods and size fractionated on formaldehyde agarose gels, transferred to nitrocellulose and blotted sequentially with different probes, see below. Probes: **pSC1**, the nick-translated probe pSC1 (Staswick and Chrispeels, 1984) hybridizing to all known PHA genes was used. For hybridization conditions and identity of probe, see Materials and methods. **lec1**, A 20-base oligonucleotide lec1 specific for *dlec*1 mRNA (complete match) and *Pdlec*1 mRNA (1 bp mismatch at position 7). See Materials and methods and Hoffman and Donaldson (1985) for sequence. **lec2**, A 20-base oligonucleotide lec2, specific for *dlec*2 and *Pdlec*2 message, complete matching with either mRNA predicted. See Materials and methods for exact sequence. Lanes: **Pdlec1**, a 900-bp artificial *Pdlec*1 mRNA. A cloned DNA fragment spanning bp 650–1535 of *Pdlec*1 (see Figure 2) was used as template for PS6 runoff transcription. **dlec2**, An 870-bp artificial *dlec*2 mRNA. The sense strand of a cloned DNA fragment containing the entire coding area of *dlec*2 (Voelker *et al.*, 1986) was transcribed using SP6 RNA polymerase. **GS^{tot}**, total cytoplasmic RNA from Greensleeves (5 μg). **Gs^{A+}**, poly(A)$^+$ selected RNA from Greensleeves (5 μg). **1/25**, GS^{A+} RNA, quantity 1/25 of GS^{A+} lane. **1/125**, GS^{A+} RNA, quantity 1/125 of GS^{A+} lane. **1/625**, GS^{A+} RNA, quantity 1/625 of GS^{A+} lane. **P^{A+}**, poly(A)$^+$ selected RNA from Pinto (5 μg). **P^{tot}**, total cytoplasmic RNA from Pinto (5 μg).

to PHA does not agglutinate erythrocytes like PHA-E and is an active mitogen (Vitale *et al.*, 1985), both properties known to be conferred to PHA by its L subunit (Miller *et al.*, 1975). In summary, the small amount of PHA found in Pinto is of the subunit L type and the evidence indicates that the *Pdlec*2 gene we have sequenced codes for this protein. By comparing the amino acid sequences of PHA-E, PHA-L and Pinto PHA-L, it is possible to identify tentatively the amino acid sequences involved in conferring their differing biological activity. PHA-E and PHA-L differ from each other at the immediate N terminus, in an area between residues 86 and 95, and are quite different around residue 132 (Figure 3). Pinto PHA-L shares its N-terminal sequence with the PHA-E and differs from PGA-E and PHA-L in the area between positions 86 and 95. However, it resembles the sequence of PHA-L completely around residue 132. Assuming that the different biological activities of the E and L subunits of PHA residue in the amino acid sequence, this leaves the region around bp 132 to be the likeliest candidate involved directly in the erythroagglutinating versus mitogenic/leucoagglutinating activities of the different PHA subunits.

Comparison of the 5′ upstream and 3′ downstream nucleotide sequences of the PHA genes

The 5′ upstream and 3′ downstream nucleotide sequence of *Pdlec*1 and *Pdlec*2 are very homolgous to the respective genes *dlec*1 and *dlec*2 of Tendergreen, including the signals known to be important for correct transcriptional start and termination. Downstream 3′ of the PHA coding frames, all four genes share sequences for another 100 bp including multiple or single (*Pdlec*2) copies of the conserved poly(A) addition site (see Figure 2; and Hoffman and Donaldson, 1985). Immediately after the poly(A) addition site, *dlec*1 diverges from the other three sequences completely, *Pdlec*2 diverges from *Pdlec*1 and *dlec*2 90 bp further downstream, and *Pdlec*1 and *dlec*2 stay homologous for the rest of the known sequence (130 bp). All four genes have the canonical YGTGTTYY sequence 30–60 bp downstream of the poly(A) addition signal. This sequence has been shown to be important for correct termination to produce a functional mRNA in mammalian cells (McLauchlan *et al.*, 1985). It is not known whether this signal is important in plants. Figure 4 compares the regions upstream of the coding frames of all four sequenced PHA genes. Up to a position of −40 relative to the translation initiation codon, *Pdlec*1 and *Pdlec*2 are identical to *dlec*1, including the TATA box and the transcriptional start site determined by Hoffman and Donaldson (1985). Immediately upstream from the TATA box, the Pinto genes follow the *dlec*1 sequence where *dlec*2 is missing 176 bp. Since these 176 bp are surrounded by a nearly perfect 28-bp inverted repeat ('A_L-A_R', in Figure 4) in all four PHA genes and in the lectin-like gene, we believe that the presence of this inverted repeat may be related to the formation of the deletion.

In summary, all signals known to be important for the correct initiation and termination of transcription are present in both Pinto PHA genes. In addition, putative transcripts from *Pdlec*1 and *Pdlec*2 would be so similar to the PHA mRNA found in high-expression cultivars that Pinto PHA mRNA would have the same stability and primary translational efficiency as PHA mRNA from normal cultivars.

Pdlec1 and Pdlec 2 mRNAs are found in developing cotyledons at differently reduced levels

The nick-translated cDNA probe pSC1 and two gene-specific oligonucleotides were used to detect Pinto PHA transcripts in developing cotyledons. Figure 5, top section, shows the analysis obtained with the nick-translated cDNA probe, pSC1 (Staswick and Chrispeels, 1984), which hybridizes to all PHA transcripts equally under the stringency applied here. The first two lanes represent bacteriophage SP6 sense transcripts from cloned *Pdlec*1 and *dlec*2 genes and both hybridize equally well. A 1.1-kb sized mRNA is detected in Greensleeves total cotyledon RNA (lane GS^{tot}) which is present at about 40-fold higher concentration in a poly(A)$^+$ selected preparation (lane GS^{A+}). In contrast, no signal is seen at the same exposure when Pinto total cytoplasmic RNA is loaded (see Figure 5, lane P^{tot}). After poly(A)$^+$ selec-

tion, a weak signal representing 1% of the Greensleeves level is found in developing cotyledons of Pinto (lane P^{A+}). Since the poly(A)$^+$ selection enriched poly(A)$^+$ mRNA of Pinto to a comparable extent (estimated 40-fold as tested by probing the same blot subsequently with a phaseolin clone, see Materials and methods, data not shown), we confirm earlier work (Staswick and Chrispeels, 1984) that the total PHA mRNA in Pinto developing cotyledons is reduced ~100-fold relative to Greensleeves.

A pair of 20-base oligonucleotides, called lec1 and lec2, was used as probes to distinguish between the lec1 and lec2 transcripts of both cultivars. They hybridize to a region in the PHA coding frames where both types of genes differ completely (see Figure 2, and Hoffman and Donaldson, 1985). The middle panel in Figure 5 shows the results obtained with the lec1 probe. It hybridizes strongly with the *Pdlec*1 *in vitro* transcript (lane *Pdlec*1), and the *dlec*2 SP6 transcript is not detected (lane *dlec*2). Comparing RNA from Greensleeves and Pinto, this lec1 specific probe detects a poly(A)$^+$ transcript in Pinto, which is present in only about 1/600 of the level of *dlec*1 message found in developing cotyledons of Greensleeves. Results obtained with the lec2 probe (the left two lanes show its complete specificity for lec2 mRNA) are shown in the bottom panel of Figure 5. This probe hybridizes to poly(A)$^+$ RNA in Pinto at a 20-fold reduced level when compared to Greensleeves. As indicated in Figure 5, 18s and 28s ribosomal RNA, predominant in all RNA preparations (not shown), gave signals with both probes under our hybridization conditions. Since all PHA mRNAs were expected to have higher electrophoresis mobility (Figure 5, pSC1), evaluation of our PHA data is not interfered. In summary, *Pdlec*1 message is reduced ~600-fold versus *dlec*1 of Greensleeves and *Pdlec*2 ~20-fold versus *dlec*2 of Greensleeves. Comparing the PHA mRNA levels measured in both cultivars with the nick-translated probe pSC1 and with the oligonucleotide lec2, allows the alignment of all four transcript levels on one scale. Since the *Pdlec*2 transcripts represent most of Pinto PHA message and *Pdlec*1 mRNA can be neglected, the nick-translated pSC1 probe measures *Pdlec*2 mRNA in Pinto, while in Greensleeves it measures *dlec*1 and *dlec*2 mRNA combined. The ratio of *Pdlec*2 versus *dlec*2 mRNA levels can be obtained with the selective oligonucleotide lec2. Since the ratios are 1:100 compared to 1:20, we can assume that in Greensleeves the *dlec*1 mRNA level is ~3- to 4-fold greater than the *dlec*2 mRNA level for midmaturation cotyledons. Accordingly, we observe with Western blots of Greensleeves cotyledon total protein preparations higher levels of PHA-E polypeptide over PHA-L with anti-PHA antibodies (data not shown). The observed relative levels of mRNA transcribed from the four homologous genes can therefore be set on one scale and the ratio *dlec1:dlec2:Pdlec2:Pdlec1* found to be 600:200:10:1. Assuming that the stabilities of all four messages are similar (see above), we postulate that these ratios reflect the different transcription rates to a first approximation.

Discussion

DNA and transcription analysis of the Pinto PHA genes

Pinto contains two genes with PHA coding areas both >90% homologous to the *dlec*1 and *dlec*2 alleles found in normal cultivars (Hoffman and Donaldson, 1985; and unpublished observations from this laboratory). One Pinto PHA gene, *Pdlec*1, is a pseudogene; its coding frame is interrupted by a single basepair deletion. This explains the complete absence of PHA-E polypeptide from Pinto beans. The properties of the predicted polypeptide of *Pdlec*2 agree with all observations obtained with PHA isolated from Pinto. Accordingly, we find that *Pdlec*2 mRNA is reduced to the same extent as its translation product; therefore, the low level of PHA-L in Pinto is caused by lower levels of its mRNA compared to normal cultivars. In disagreement, Horowitz (1985) found PHA mRNA to be reduced only 10-fold in the Pinto cultivar. This determination has to be interpreted with caution because of the use of the cDNA probe pVL134 which is derived from the lectin-like gene expressed in normal amounts in Pinto (Staswick and Chrispeels, 1984).

Are both Pinto genes low-level expression alleles?

We propose that the low levels of mRNA for *Pdlec*1 and *Pdlec*2 in developing cotyledons are due to lower rates of gene transcription in Pinto versus normal cultivars. First, mRNAs of both Pinto PHA genes are structurally almost identical to their normal counterparts and polyadenylated. It seems unlikely, therefore, that Pinto mRNA would be less stable and more rapidly degraded. Second, Chappell and Chrispeels (1986), using nuclear run-off experiments, found barely detectable rates of PHA gene transcription in Pinto cotyledons compared to the cultivar Greensleeves.

Evidence gained from a classical genetic approach favors *cis* elements to be responsible for the Pinto phenotype. A.Pusztai (personal communication) analyzed seeds of hybrids obtained by crossing Pinto with cultivars of *P. acutifolius* which have large amounts of PHA-E but no PHA-L in their seeds. In all cases tested, the hybrid beans show high levels of PHA-E and low levels of Pinto PHA-L. A putative *trans*-acting factor driving the high expression of the PHA-E gene obviously does not enhance the expression of the homologous Pinto PHA-L gene *in vivo*.

Comparison of Pinto 5′ upstream sequences with other allelic lectin genes

In Figure 4, upstream sequences of all four PHA gene sequences are aligned according to *dlec*1; the arrow at the lower right marks the start of transcription as determined by Hoffman and Donaldson (1985). Both Pinto genes are identical to *dlec*1 from the TATA site located −28 from the transcriptional start site through the beginning of the coding frame, sharing a 5-bp spaced, repeated palindromic TGCA sequence at the 5′ end of their putative transcripts. This structure is found completely conserved at the same position in the lectin-like gene of *P. vulgaris* (Hoffman, 1984), and the *le*1 gene of the soybean lectin (Vodkin *et al.*, 1983). It may play an important role at the 5′ terminus of lectin mRNAs. In an alternative interpretatation, this sequence could be part of a tissue-specific promoter of these genes and be essential for transcription: for example, Morelli *et al.* (1985) have shown that plant promoter sequences conferring light inducibility can be located 3′ of the TATA box.

Upstream from the TATA box, all four PHA genes diverge to different degrees (Figure 4). The first 176-bp 5′ of the TATA box in *dlec*1 are deleted in *dlec*2 without obvious effect on the level of transcription. *Pdlec*2, which has a transcription activity reduced by a factor of 60 compared to *dlec*1, follows the *dlec*1 sequence very closely up to −240 relative to the transcription start site. Farther upstream, a 100-bp sequence shared by all three other PHA genes is deleted (Figure 4, upper box). This 100-bp sequence consists of an almost perfect direct repeat of 60 bp with each copy carrying an inverted repeat ('RY', Figure 4) consisting of multiples of a TGCA motif. Each of the RY structures could form Z-DNA of 18−20 bp in length. In our current working hypothesis concerning the Pinto PHA genes, we assign this 100-bp stretch of DNA to be an enhancer-like element. *Pdlec*2

devoid of this element is, therefore, expressed at a 60-fold reduced level. This element is present in *Pdlec*1, the mRNA of which is found in very reduced (600-fold) levels. The 5′ sequence of *Pdlec*1 is almost 100% in accordance to *dlec*1 as far as *dlec*1 sequence is available, including the inverted repeat at −360 to −240. All other structures found in the high expression allele are present; the RY containing direct repeat is present three times in *Pdlec*1. Why, then, is there such a low level of transcription? Three possibilities come to mind. First, even though the mRNA resembles the known, stable structures almost completely, it could be broken down rapidly. We believe this to be unlikely, as discussed above. Secondly, elements even further upstream which are responsible for high expression could be missing in *Pdlec*1. Indirect evidence from other legume seed storage protein genes expressed in transgenic plants position the elements for tissue specific up-regulation relatively close to the transcriptional start (Doyle *et al.*, 1986; R.Goldberg, personal communication). A third possible cause for the low levels of *Pdlec*1 mRNA could be down-regulation of its gene by silencing. Silencers have been found in yeast (Brand *et al.*, 1985) and mammals (Laiminis *et al.*, 1986). Introducing *Pdlec*1 and *Pdlec*2 into transgenic tobacco may allow us to test the different hypotheses about the control of their regulation.

Materials and methods

Plants

Seeds of the common bean, *P. vulgaris*, were used to grow plants in the greenhouse. Three different cultivars were used. Pinto UI111 is a cultivar with very little PHA, while Tendergreen and Greensleeves are cultivars that have large amounts of PHA and have a similar tandem organization of *dlec*1 and *dlec*2. The plasmid pPVL134 was a gift from L.Hoffman of the Agrigenetics Corporation (Madison, WI, USA).

DNA isolations

DNA was isolated from young leaves of greenhouse-grown plants of Greensleeves and Pinto UI111 by grinding 10 g of fresh tissue to a powder in liquid nitrogen in a mortar and pestle. The powder was added to 20 ml of ice-cold buffer A (0.33 M sucrose, 100 mM Tris−HCl, pH 8.5, 60 mM KCl, 30 mM $MgCl_2$, 2 mM EDTA, 1% w/v iodoacetic acid) and ground to a fine slurry. After addition of 20 ml buffer A, the material was filtered through nylon mesh and centrifuged for 10 min at 1100 × *g*. The pellet was resuspended in 10 ml buffer A and 0.4 ml of 10% Triton X-100, incubated for 5 min on ice and centrifuged for 5 min at 600 × *g*. The resuspension was repeated and the resulting pellet suspended in 10 ml of 20 mM Tris, pH 7.5, 10 mM EDTA, 1% N-lauryl sarcosyl (sodium salt). Cesium chloride (10 g) and 1 ml ethidium bromide (10 mg/ml) were added and the mixture was centrifuged for 24 h at 40 000 r.p.m., 20°C, in a Beckman Ti50 rotor.

Genomic cloning

*Bam*HI-digested DNA from Pinto was size-fractionated on a 0.7% preparative agarose gel. DNA ranging in size from ~10 to 20 kb was cloned into phage λ vector EMBL 4 according to procedures supplied by Promega Biotech. Recombinants were screened by plaque life hybridization to PHA cDNA clone pSC1 (Staswick and Chrispeels, 1984). Additional clones were selected in a similar manner from a total genomic Pinto library in Charon 35 phage λ obtained from J.Horowitz (Department of Biology, University of California, Los Angeles).

DNA sequencing

DNA fragments cloned in bacteriophage M13 mp vectors (Yanisch-Perron *et al.*, 1985) were sequenced using the dideoxy method of Sanger *et al.* (1977) as modified in the Molecular Biology Cloning and Sequencing Handbook of Amersham. The sources of the DNA fragments were either a *Bal*31 deletion series of fragments cloned into pUC vectors or restriction enzyme subfragments. Computer analysis of the sequence data and comparison with the sequences of *dlec*1 and *dlec*2 (Hoffman and Donaldson, 1985) was done on the University computer using the programs of Staden (1982).

Oligonucleotides

The 20-base oligonucleotide lec1 (5′ CTGTCGTATTTGTAGTTGTT) was made by our DNA synthesis facility. The 20-base oligonucleotide lec2 (5′ AAATTGCTGTTGCTGCCGTC) was a gift from L.Hoffman, Agrigenetics Corp. (for synthesis see Hoffman and Donaldson, 1985). After kinasing with [γ-^{32}P]ATP (Maniatis *et al.*, 1982), the probes were purified by acrylamide gel electrophoresis. Specific activities were about 2 × 10^8 c.p.m./μg).

SP6 RNA

An 841-bp *Xho*I fragment containing the entire coding frame of *dlec*2 (Voelker *et al.*, 1986) was inserted into the expression vector SP64 (Promega Biotech, Melton *et al.*, 1984) and an 870-bp sense strand run-off transcript synthesized with SP6 RNA polymerase *in vitro*. A *Pdlec*1 run-off sense strand transcript was produced in essentially the same way but a 900-bp DNA fragment including the coding region of *Pdlec*1 (~bp 650−1535, Figure 2, construction not shown) was used as template.

RNA isolation

Total cytoplasmic RNA used for gel electrophoresis was isolated as described in Staswick and Chrispeels (1984). Poly(A)$^+$ RNA was selected by chromatography of the total cytoplasmic RNA on oligo(dT) cellulose (Maniatis *et al.*, 1982).

Hybridization procedures

RNA−DNA hybridizations were carried out on nitrocellulose filters which had been blotted from 1% agarose gels containing formaldehyde (Maniatis *et al.*, 1982). For nick-translated probes, prehybridization and hybridization of the dried filters were carried out at 37°C in 33% formamide, 5 × SSPE, 1 × Denhardt solution, 100 μg/ml salmon sperm DNA, 0.2% SDS. Nick-translated probe labelled with ^{32}P (~2 × 10^7 c.p.m./μg) was added to a concentration of 0.5−1 × 10^6 c.p.m./ml and incubated overnight. Nick-translation probes used were a PHA (*dlec*1) cDNA clone pSC1 (Staswick and Chrispeels, 1984) and a phaseolin genomic clone (p3.8.177.4, Slightom *et al.*, 1983) provided by J.Slightom of Agrigenetics Corp. Following hybridization, filters were washed twice in 2 × SSC/0.1% SDS at 55°C and twice in 0.2 × SSC/0.1% SDS at the same temperature. For oligonucleotide probes, nitrocellulose filters were prehybridized in a solution of 6 × SSC, 1 × Denhardt solution, 0.1% SDS and 100 μg/ml tRNA at 30°C (lec1) or 37°C (lec2) for 3 h. Overnight hybridizations were carried out under the same conditions at a concentration of 5 × 10^5 c.p.m./ml of ^{32}P-kinased oligomer. Final washes after the hybridizations were 60 min in 6 × SSC at 42°C for lec1 and 60 min at 55°C for lec2. Filters were air dried and subjected to autoradiography at −70°C using intensifying screens.

Acknowledgements

This work was partially supported by the United States Department of Agriculture (Competitive Research Grants Office) and the National Science Foundation (Metabolic Biology). T.A.V. was a recipient of a Max Kade Foundation fellowship. We thank Dr A.Vitale for the preparation of a genomic library of *P. vulgaris* cv. Greensleeves, and Dr J.Horowitz for her gift of a genomic library of *P. vulgaris* cv. Pinto. We thank Roberta Bradford and Betsy Chappell for their technical assistance and Dr Joe Chappell for many interesting discussions.

References

Brown,D., Levasseur,M., Croy,R., Boulter,D. and Gatheouse,J. (1985) *Nucleic Acids Res.*, **13**, 4527−4538.

Brand,A., Breeden,L., Abraham,J., Sternglanz,R. and Nasmyth,K. (1985) *Cell*, **41**, 41−48.

Brown,J.W.S., Bliss,F.A. and Hall,T.C. (1981) *Theor. Appl. Genet.*, **60**, 251−259.

Chappell,J. and Chrispeels,M.J. (1986) *Plant Physiol.*, **81**, 50−54.

Doyle,J.J., Schuler,M.A., Godette,W.D., Zenger,V. and Beachy,R.N. (1986) *J. Biol. Chem.*, **261**, 9228−9238.

Hoffman,L.M. (1984) *J. Mol. Appl. Genet.*, **2**, 447−453.

Hoffman,L.M. and Donaldson,D.D. (1985) *EMBO J.*, **4**, 883−889.

Horowitz,J. (1985) *Mol. Gen. Genet.*, **198**, 482−485.

Kozak,M. (1986) *Cell*, **44**, 283−292.

Ladin,B., Doyle,J. and Beachy,R. (1984) *J. Mol. Appl. Genet.*, **2**, 372−380.

Laiminis,L., Koenig,M. and Khoury,G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3151−3155.

Maniatis,T., Fritsch,E. and Sambrook,J. (1982) *Molecular Cloning*. Cold Spring Harbor Laboratory, NY.

McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347−1368.

Melton,D.A., Krieg,P., Rebagliati,M., Maniatis,T., Zinn,K. and Green,M. (1984) *Nucleic Acids Res.*, **12**, 7035−7056.

Miller,J.B., Hsu,R., Heinrikson,R. and Yachnin,S. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 1388−1391.

Morelli,G., Nagy,F., Frayley,S. and Chua,N. (1985) *Nature*, **315**, 200−204.

Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211−214.

Pusztai,A., Grant,G. and Stewart,J.C. (1981) *Biochim. Biophys. Acta*, **671**, 146−154.

Rhodes,P. and Vodkin,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 493–497.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Slightom,J., Sun,S. and Hall,T. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1897–1901.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951–2961; 4731–4751.
Staswick,P., Chappell,J,. Voelker,T., Vitale,A. and Chrispeels,M.J. (1986) In Shannon,L.M. and Chrispeels,M.J. (eds), *Molecular Biology of Seed Storage Proteins and Lectins*. American Society for Plant Physiology, Rockeville, MD.
Staswick,P. and Chrispeels,M.J. (1984) *J. Mol. Appl. Genet.*, **2**, 525–535.
Strosberg,D.A., Lauwereys,M. and Foriers,A. (1983) In *Chemical Taxonomy, Molecular Biology and Function of Plant Lectins. Molecular Evolution of Legume Lectins*. Alan R.Liss, New York, pp. 7–20.
Sturm,A. and Chrispeels,M.J. (1986) *Plant Physiol.*, **81**, 320–322.
Vitale,A., Ceriotti,A. and Bollini,R. (1985) *Planta*, **166**, 201–207.
Vodkin,L.O., Rhodes,P.R. and Goldberg,R.B. (1983) *Cell*, **34**, 1023–1031.
Voelker,T., Florkiewicz,R. and Chrispeels,M.J. (1986) *Eur. J. Cell. Biol.* in press.
von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on 20 August 1986; revised on 17 September 1986