Eur. J. Biochem. 225, 787–795 (1994)

Isolation and characterisation of arcelin-5 proteins and cDNAs

Alain GOOSSENS, Roberto GEREMIA, Guy BAUW, Marc VAN MONTAGU and Geert ANGENON
Laboratorium voor Genetica, Universiteit Gent, Belgium

(Received May 19/August 22, 1994) – EJB 94 0721/2

Arcelins are seed storage proteins present in some wild bean accessions (*Phaseolus vulgaris*). They are implicated in the resistance phenotype of these wild beans towards the Mexican bean weevil. Arcelin 5, one of six arcelin electrophoretic variants, has been characterised in detail. The purified arcelin-5 protein fraction contained two major polypeptides of 32.2 and 31.5 kDa, designated arcelin 5a and arcelin 5b, respectively, and one minor polypeptide of 30.8 kDa, designated arcelin 5c. The three polypeptides have an identical isoelectric point and are identical for their first nine N-terminal amino acids. Arcelin 5a and arcelin 5b are glycoproteins whereas arcelin 5c is not glycosylated. Native arcelin 5 has a molecular mass corresponding to a dimer form. Using amino acid sequence analysis and PCR techniques, two different arcelin-5 cDNA sequences were obtained, designated *arc5-I* and *arc5-II*. Both encode proteins of 261 amino acids with a signal peptide of 21 amino acids. The identity between the two is 99% at the DNA level and 97% at the level of the deduced amino acid sequences. The *arc5-I* and *arc5-II* cDNAs encode arcelin 5a and arcelin 5b, respectively. Sequence comparisons and protein characteristics show clearly that arcelin 5 is related to, but distinct from, other arcelin variants and lectins of *P. vulgaris*.

Most cultivars of common bean (*Phaseolus vulgaris*) contain two major seed storage proteins: phaseolin and the bean lectin phytohemagglutinin. A number of electrophoretic variants of phaseolin are known with polypeptides in the 45–51-kDa range (Brown et al., 1981a), whereas phytohemagglutinin variants have polypeptides in the 33–41-kDa range (Brown et al., 1981b). In a few wild *P. vulgaris* accessions, a third important seed storage protein fraction is found which is not present in any known cultivar. This fraction was named arcelin (Romero Andreas et al., 1986) and until now six allelic arcelin variants have been identified. These variants have electrophoretic mobilities different from either phaseolin or phytohemagglutinin. Arcelin 1 and arcelin 2 consist of several polypeptides of approximately 35 kDa (Osborn et al., 1986, 1988a; Hartweck et al., 1991). Nucleotide sequences of an arcelin-1 and an arcelin-2 cDNA were determined and have a similarity of 99.3% at the level of deduced amino acid sequences (John and Long, 1990; Osborn et al., 1988b). Arcelin 3 and arcelin 4 have similar electrophoretic patterns consisting of several polypeptides of 35–40 kDa (Osborn et al., 1986; Hartweck et al., 1991). More recently, Lioi and Bollini (1989) and Santino et al. (1991) identified a fifth arcelin variant consisting of two major polypeptides and a sixth arcelin variant made up of a major polypeptide of about 38 kDa, respectively.

The available sequence data show that these arcelin variants, phytohemagglutinin, and the bean α-amylase inhibitor are all encoded by related members of a lectin gene family. Arcelin expression is inherited as a single locus (Romero Andreas et al., 1986), unlinked to the genes encoding phaseolin but tightly linked to those encoding phytohemagglutinin and the α-amylase inhibitor (Osborn et al., 1986).

Many plant lectins probably have a defense function against insects and other predators (Chrispeels and Raikhel, 1991). Arcelins too would have such a role and are proposed to be responsible for resistance against the Mexican bean weevil (*Zabrotes subfasciatus*) (Osborn et al., 1988b). Resistance towards this pest, an important cause of post-harvest losses in cultivated common beans from tropical countries, was only found in wild *P. vulgaris* accessions containing one of the arcelin variants (Schoonhoven et al., 1983). Highest levels of resistance were found in wild accessions containing arcelin 5, followed in order by arcelin 4, arcelin 1, arcelin 2, and arcelin 3 (Cardona et al., 1990). When these accessions were used in a breeding programme, highest resistance levels were obtained from crosses with arcelin-1 or arcelin-5 parents. Arcelin-2-derived lines were intermediate whereas Sturm, A. & Chrispeels, M. J. (1986) The high mannose oligosac- Since all arcelin-3-containing lines were susceptible, this variant was not used for breeding purposes (Cardona et al., 1990).

The arcelin-5 variant, in contrast to the other arcelin variants, has not yet been further characterised. In this study, we report on the isolation and the characterisation of the arcelin-5 protein and the corresponding cDNAs.

Correspondence to M. Van Montagu, Laboratorium voor Genetica, Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

Abbreviations. *arc5-I* and *arc5-II*, gene(s) (DNA, mRNA) encoding arcelin 5; pI, isoelectric point; RACE, rapid amplification of cDNA ends.

Enzymes. Endoglycosidase H (EC 3.2.1.96); *N*-glycosidase F (EC 3.5.1.52); α-mannosidase (EC 3.2.1.24); trypsin (EC 3.4.21.4).

Note. The novel nucleotide sequence data published here have been submitted to the EMBL sequence data bank and are available under accession numbers Z36943 (*arc5-I*) and Z36970 (*arc5-II*).

MATERIALS AND METHODS

Plant materials

Three lines of *P. vulgaris* kindly provided by Dr C. Cardona (CIAT, Cali, Colombia) were used for this study: (a) G02771 is the wild accession containing arcelin 5, phytohemagglutinin, and phaseolin and showing high levels of resistance towards *Z. subfasciatus*; (b) 'Ica Pijao' is a bean cultivar that contains both phaseolin and phytohemagglutinin but no arcelin; (c) a BC1F1 line which was obtained by crossing G02771 with 'Ica Pijao', and subsequently back-crossing to 'Ica Pijao'. This line contains phaseolin, phytohemagglutinin, and arcelin 5 and is resistant to *Z. subfasciatus*.

Protein purification

Arcelin 5 was purified from seeds of the wild *P. vulgaris* accession G02771. This was done by two successive extractions of the bean flour with 10 mM NaCl, 50 mM glycine, pH 2.4 for 1 h at room temperature (10 g flour was extracted each time with 100 ml buffer). After each extraction, the mixture was centrifuged at 20000×*g* (20 min at 4 °C). The supernatants were combined and dialysed for 24 h at 4 °C against three changes of double-distilled H_2O. The precipitate (globulin fraction) was pelleted by centrifugation and in this way separated from the supernatant (albumin fraction). Arcelin 5 was further purified by chromatofocusing of the globulin fraction, resuspended in the start buffer (0.025 M ethanolamine/acetic acid pH 9.4), on a PBE94-column (Pharmacia) with a pH gradient of 9−6. Before running the sample, the column was equilibrated with start buffer. Proteins were eluted with a 10-fold dilution of Polybuffer® (Pharmacia), brought to pH 6.0 with acetic acid. The fractions having a pH of 7.5−8 contained highly purified arcelin 5.

Electrophoresis

Protein samples were prepared and separated by SDS/PAGE as described by Hames and Rickwood (1990). The molecular mass of denatured proteins was determined by comparing the mobility of proteins to those of low-molecular-mass standard proteins. Isoelectric points (pI) of the proteins were determined by using Ampholine® PAGplates (Pharmacia) and by comparing the mobility with those of IEF standards for pI calibration of native IEF gels.

Gel filtration

The molecular mass of native arcelin-5 protein was determined by filtration through a column (1.0×110 cm) of Sephacryl S-300® (Pharmacia) with a flow rate of 6.75 ml/h. Phosphate-buffered saline (NaCl/P_i: 0.14 M NaCl, 3 mM KCl, 50 mM Na_2HPO_4, 10 mM KH_2PO_4, pH 7.2) was used as the equilibration, elution, and sample-dissolving buffer.

Detection of glycoproteins

Glycosylation was detected in purified arcelin and phytohemagglutinin fractions by dansylhydrazine staining (Eckhardt et al., 1976). Glycoproteins can then be visualised and photographed under ultraviolet light.

Deglycosylation

Purified arcelin or phytohemagglutinin (used as a positive control) were chemically deglycosylated using trifluoromethanesulfonic acid as described by Edge et al. (1981). Arcelin or phytohemagglutinin were treated in the acid solution for 2.5 h at 4 °C.

Enzymic deglycosylation was performed on 50 µg purified arcelin or phytohemagglutinin (as a positive control) using three different glycosidases (all purchased from Boehringer, Mannheim). (a) Digestion with N-glycosidase F (of *Flavobacterium meningosepticum* from recombinant *Escherichia coli*) was performed by incubation at 37 °C for 20 h in 500 µl of 20 mM sodium-phosphate pH 7.2, 10 mM sodium-azide, 50 mM EDTA, 0.5% NP40 (by vol.) with 2 U enzyme. (b) Digestion with endoglycosidase H (of *Streptomyces plicatus* from recombinant *E. coli*) was performed by incubation at 37 °C for 24 h in 500 µl 100 mM sodium-acetate pH 5.8 with 10 mU enzyme. (c) Digestion with α-mannosidase (from jack bean) was performed by incubation at 37 °C for 24 h in 500 µl 50 mM sodium-citrate pH 4.6, 5 mM $ZnSO_4$ with 4 U enzyme.

Prior to digestion the proteins were denatured by heating at 100 °C for 2 min in the presence of 0.12% SDS and 50 mM dithiothreitol in a volume of 50 µl.

Protein sequencing

N-terminal protein sequencing was performed on Immobilon®-bound proteins as described by Bauw et al. (1990). Trypsin digestion was performed on Immobilon®-bound arcelin-5 proteins. The resulting peptides were subsequently eluted, separated by reverse-phase HPLC, and sequenced according to Bauw et al. (1990).

Mild acid hydrolysis

The arcelin-5 fraction was separated using SDS/PAGE and visualised with Coomassie blue staining; the bands corresponding to the arcelin-5 proteins were cut out, washed with distilled H_2O and treated for 48 h in 80% formic acid at room temperature. The cleavage mixture was then lyophilized, washed with distilled H_2O, lyophilized, and resuspended in SDS/PAGE sample buffer. The cleavage products were separated on a second SDS/PAGE and visualised by silver-staining.

Arcelin-5 proteins were also cleaved at Asp-Pro *in situ* on poly(4-vinyl-*N*-methylpyridinium iodide) glass-fiber blots as described by Bauw et al. (1988). This allowed the determination of one or two internal peptide sequences.

Hemagglutination tests

Hemagglutination tests were performed on purified arcelin and phytohemagglutinin fractions using human type A^+ erythrocytes. Purified proteins (with a starting concentration of 4 mg/ml) were serially diluted with equal volumes of NaCl/P_i and assayed for agglutinating activity on native and pronase-treated erythrocytes in wells of microtiter plates. Native and pronase-treated erythrocytes were prepared as described in Hartweck et al. (1991). Hemagglutinating activity was scored visually after a 1-h incubation at room temperature.

cDNA cloning

The first step in the cloning strategy consisted in the determination of the amino acid sequence of the N-terminus and some internal peptides from the arcelin-5 protein. On the

basis of these amino acid sequences, degenerate oligonucleotide primers were designed for carrying out reverse transcription/PCR reactions (Saiki et al., 1988). Total RNA of immature seeds of the G02771 accession (prepared as described by Jones et al., 1985) was used as a template. In this way a partial cDNA clone of 369 nucleotides, corresponding to an internal fragment of the arcelin-5 mRNA, was obtained. Based on this sequence, specific oligonucleotide primers were designed for carrying out 5′ and 3′ rapid amplification of cDNA ends (RACE) (Frohman et al., 1988); again total RNA of immature seeds of the G02771 accession was used as a template. In this way, partial cDNA fragments were obtained, corresponding to the 5′ end and the 3′ end of the arcelin-5 mRNA, both overlapping with the internal fragment obtained earlier. Both reverse transcription/PCR and RACE reactions were performed using the appropriate kits from Gibco BRL. These fragments were cloned into pGEM®-2 (Promega, Madison WI) and subsequently four independent clones of each fragment were sequenced in both directions.

Northern blot analysis

Total RNA from immature beans from the G02771 accession was analysed by Northern blot analysis as described in Sambrook et al. (1989) using an arcelin-5 cDNA probe.

Restriction fragment analysis

Total leaf DNA was prepared following a modification of the RNA preparation procedure described by Jones et al. (1985). After precipitation with lithium acetate, the DNA was recovered from the supernatant by addition of 2 vol. ethanol and centrifugation. The DNA pellet was resuspended in sterile water and precipitated again by adding 0.1 vol. 2 M sodium acetate pH 4.8 and 2 vol. ethanol. The pellet was washed with 70% ethanol and resuspended in sterile water. This DNA was analysed by Southern blot hybridisation as described in Sambrook et al. (1989) using an arcelin-5 cDNA probe or a phytohemagglutinin cDNA probe. cDNA fragments that were used for this analysis were obtained following 3′ RACE reactions performed on total RNA from immature beans from the G02771 accession. Autoradiographs were taken after washing the blots with NaCl/Cit (150 mM NaCl, 15 mM trisodium-citrate, pH 7.0), 0.1% SDS, 65 °C (low stringency), and after washing with 0.1×NaCl/Cit, 0.1% SDS, 65 °C (high stringency).

RESULTS

Protein purification

Arcelin-5 protein was identified by comparing 0.5 M NaCl extracts from seeds of arcelin-5-containing and arcelin-5-lacking accessions on SDS/PAGE (Cardona et al., 1990). Seeds of the wild G02771 accession contain three major protein fractions (Fig. 1A): phaseolin (45–49 kDa), phytohemagglutinin (32.5 kDa), and arcelin 5 (30.8–32.2 kDa). Phytohemagglutinin and arcelin 5a (see further) have almost the same mobility on SDS/PAGE. Arcelin 5 was purified from seeds from the wild G02771 accession using different solubility properties and chromatofocusing (see Materials and Methods). As the initial low-salt extract (10 mM NaCl, pH 2.4; Fig. 1, lane 1) contained a much higher arcelin/phaseolin ratio as compared to a high-salt extract (0.5 M NaCl; Fig. 1, lane 2), this low-salt extract was used for further purification of arcelin 5. After dialysis of the extract against double-distilled H_2O, arcelin proteins were present in both the globulin fraction (i.e. the precipitate; Fig. 1, lane 4) and albumin fraction (i.e. the supernatant; Fig. 1, lane 3) but were much more enriched in the former. The globulin fraction seemed to contain mainly arcelin 5 as judged from SDS/PAGE. However, isoelectric focusing revealed that this fraction still contains phytohemagglutinin as a major impurity (see Fig. 1 B, lane 4). Since native arcelin 5 and phytohemagglutinin differ greatly in pI (7.7 and 5.3; Fig. 1 B, lanes 5 and 6, respectively), the globulin fraction was submitted to chromatofocusing with a pH gradient of 9–6 in order to remove phytohemagglutinin and other impurities. After chromatofocusing, highly purified arcelin 5 was obtained in the fractions having a pH between 7.5–8 (Fig. 1, lane 5). With this purification procedure, 1 g bean flour yields approximately 15 mg highly purified arcelin-5 proteins.

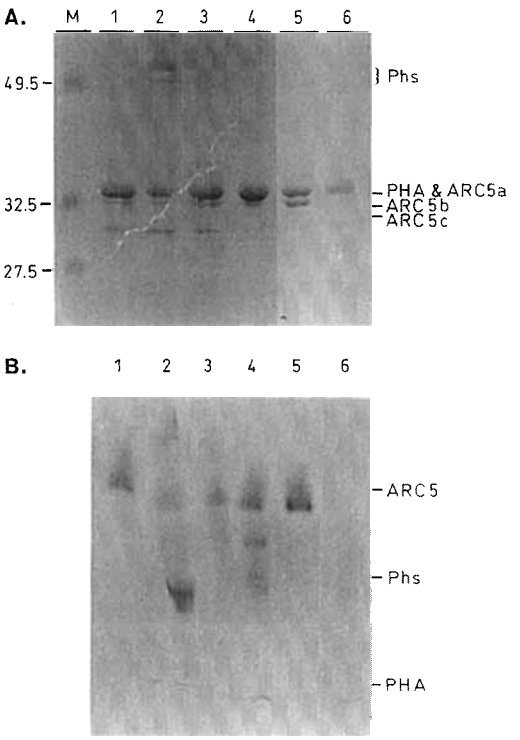

Fig. 1. Purification of the arcelin-5 protein. Crude, partially purified and purified seed proteins of G02771 separated by (A) SDS/PAGE, visualised with Coomassie blue or (B) IEF, visualised with Coomassie blue. Samples shown are crude seed proteins extracted with 10 mM NaCl, pH 2.4 (lane 1) or 0.5 M NaCl (lane 2), partially purified proteins, albumins (lane 3) or globulins (lane 4) and purified proteins, arcelin 5 (ARC5, lane 5) or phytohemagglutinin (PHA, lane 6) obtained after chromatofocusing. Phs = phaseolin. Lane M contains marker proteins (molecular mass indicated on the left in kDa).

Since the chromatofocusing column was run with a pH 9–6 gradient and phytohemagglutinin has an pI below 6, it remained bound to the column. A fraction containing mainly phytohemagglutinin (32.5 kDa) could be obtained upon washing the chromatofocusing column with 1 M NaCl (Fig. 1, lane 6).

Characterisation of arcelin-5 protein

The arcelin-5 protein fraction consists of two major polypeptides (32.2 kDa and 31.5 kDa) and one minor

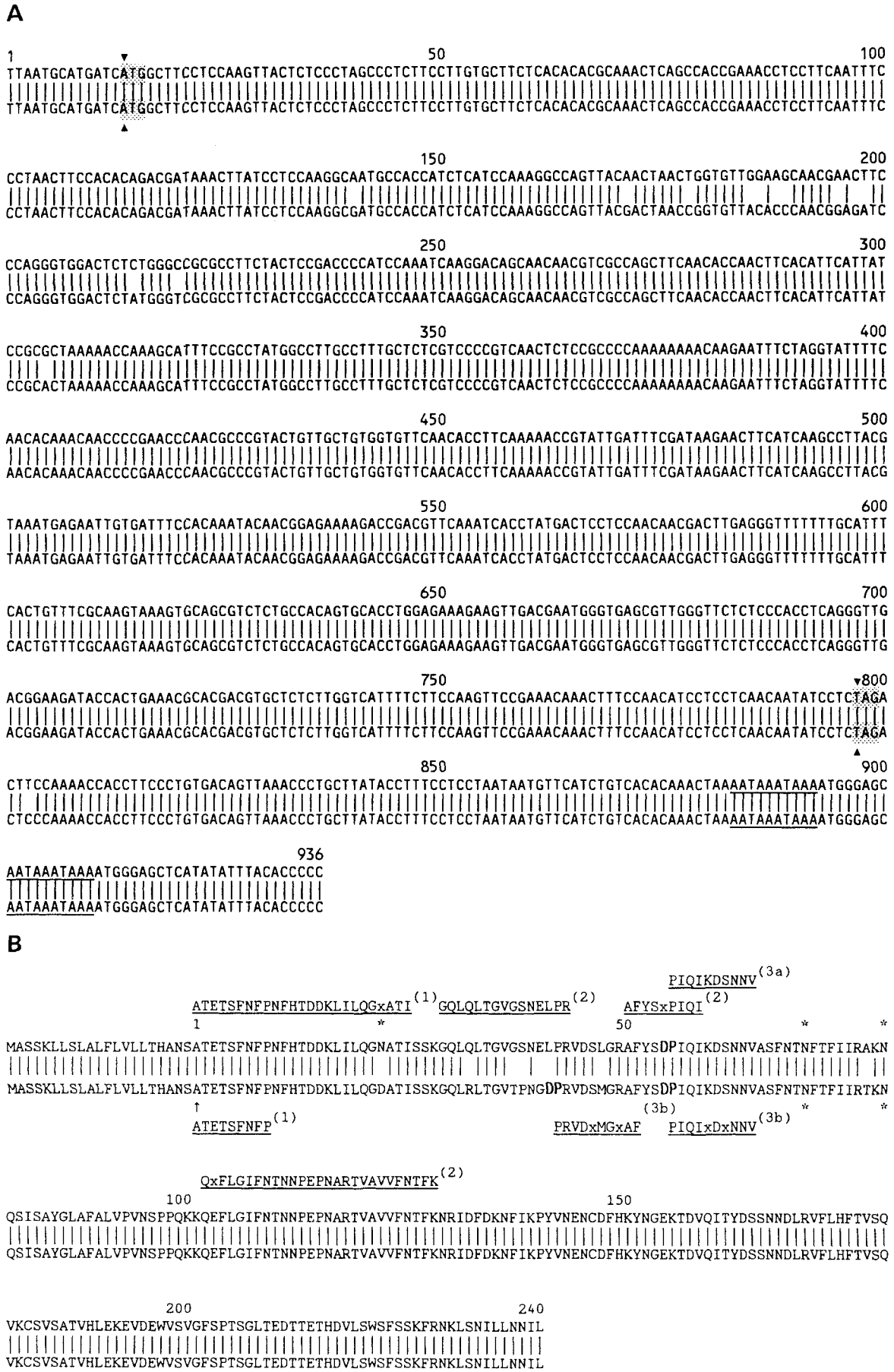

Fig. 2. Sequence analysis of arcelin-5 cDNAs. (A) cDNA sequence of the two different arcelin-5 copies *arc5-I* (top) and *arc5-II* (bottom). ▼▲ = start or stop codon, AATAAA = polyadenylation signal. (B) Deduced protein sequence of the two different arcelin-5 cDNA copies *arc5-I* (top) and *arc5-II* (bottom); numbering on top starts from the N-terminus of the mature protein. ↑ = N-terminus of the mature protein; * = potential N-glycosylation site; **DP** = Asp-Pro cleavage site. Peptide sequences (underlined) written above or below the complete protein sequences deduced from *arc5-I* and *arc5-II* were determined by protein sequencing of arcelin 5a and 5b, respectively. Numbers at the end of the peptide sequences indicate the method by which they were determined: (1) N-terminal sequencing; (2) after trypsin digest; (3a) after Asp-Pro cleavage of arcelin 5a; and (3b) after Asp-Pro cleavage of arcelin 5b.

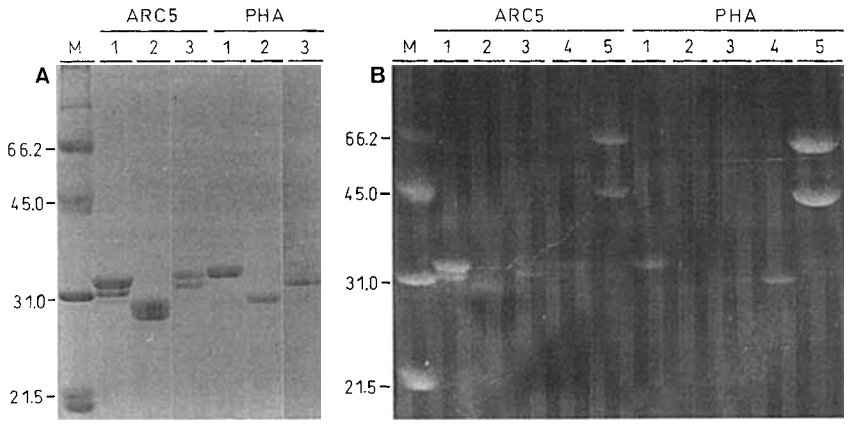

Fig. 3. Analysis of glycosylation of arcelin-5 (ARC5) and phytohemagglutinin (PHA) polypeptides. (A) SDS/PAGE stained for total protein (with Coomassie blue); (B) SDS/PAGE stained for carbohydrate (with dansylhydrazine). Samples shown are denatured proteins (lane 1), chemically deglycosylated proteins (lane 2), N-glycosidase-F-treated proteins (lane 3), endoglycosidase-H-treated proteins (lane 4), α-mannosidase-treated proteins (lane 5). The major bands of 65 kDa and 45 kDa that are visible in lane 5 represent the α-mannosidase. Lane M contains marker proteins (molecular mass indicated on the left in kDa).

polypeptide (30.8 kDa), all with the same pI (Fig. 1A; lane 5). They are designated arcelin 5a, 5b and 5c, respectively. The N-terminal sequence (25 amino acid residues of 5a, 9 amino acid residues of 5b and 5c; Fig. 2B) of each of these polypeptides was determined and showed that all three have the same N-terminal sequences. After trypsin digestion of arcelin 5a the sequence of four internal peptides could be determined (Fig. 2B). These sequences could be aligned with the deduced amino acid sequences from the arcelin-1, phytohemagglutinin-E, phytohemagglutinin-L, and the α-amylase inhibitor genes. Highest similarity was obtained with arcelin 1.

The pI of native arcelin is approximately 7.7 (Fig. 1B, lane 5). The molecular mass of the native protein was determined by gel filtration. Arcelin 5 eluted as a single peak of 74 kDa, most likely corresponding to a dimer of polypeptide subunits. If proteins of this peak were separated on SDS/PAGE, the three arcelin-5 polypeptides were visible (data not shown) indicating that all three are used in dimer formation.

Carbohydrate staining with dansylhydrazine showed that arcelin 5a and 5b are glycoproteins while 5c is not glycosylated (Fig. 3B, lane 1). Chemical deglycosylation of the arcelin-5 protein yields two polypeptides of almost equal size (approximately 28–29 kDa; Fig. 3A, lane 2). The chemically deglycosylated arcelin-5 polypeptides showed no carbohydrate staining with dansylhydrazine (Fig. 3B, lane 2). However, even when the proteins were denatured in the presence of SDS and dithiothreitol prior to digestion, enzymic deglycosylation does not result in a shift in molecular mass of the arcelin-5 polypeptides (Fig. 3A) nor in a change of carbohydrate staining (Fig. 3B). This is the case for (a) N-glycosidase-F, which cleaves all types of asparagine-bound N-glycans except when a fucose is attached α1→3 to the asparagine-linked *N*-acetylglucosamine residue (Tretter et al., 1991); (b) endoglycosidase H, which cleaves high-mannose glycan chains but not modified glycan chains; and (c) the exoglycosidase α-mannosidase, which removes terminal mannose residues. In contrast, phytohemagglutinin polypeptides purified from the G02771 accession decrease in molecular mass after digestion with these three glycosidases but still have a higher mass than after chemical cleavage and still show carbohydrate staining with dansylhydrazine (Fig. 3B). Taking into account the specificity of the different glycosidases, we conclude (a) that phytohemagglutinin from the G02771 accession is glycosylated in the same way as phytohemagglutinin from the *P. vulgaris* cv. Greensleeves, which contains one high-mannose oligosaccharide and one modified, fucosylated oligosaccharide (Vitale et al., 1984; Sturm and Chrispeels, 1986) and (b) that the glycan chains attached to arcelin 5a and 5b are most probably of the modified, fucosylated type.

Hemagglutination tests

Purified arcelin-5 and phytohemagglutinin fractions from the G02771 accession were tested for hemagglutinating activity with human erythrocytes. Arcelin 5 did not agglutinate native erythrocytes (a maximum concentration of 4.5 mg/ml was tested) whereas phytohemagglutinin agglutinated native erythrocytes at a minimum concentration of 7.8 μg/ml. However, arcelin 5 showed agglutinating activity with pronase-treated erythrocytes at a minimum concentration of 500 μg/ml whereas for phytohemagglutinin, a minimum concentration of 1 μg/ml was required to agglutinate pronase-treated erythrocytes.

Arcelin-5 cDNA clones

Based on the peptide sequences of arcelin 5a, primers were designed which were used to carry out reverse transcription/PCR and 5′ and 3′ RACE reactions on total RNA of immature seeds of the G02771 accession as a template (see Materials and Methods). In this way overlapping partial cDNA fragments, corresponding to the 5′ end (375 nucleotides), an internal fragment (369 nucleotides), and the 3′ end (592 nucleotides) of the arcelin-5 mRNA, were obtained. After cloning and sequencing of the resulting fragments, complete cDNA sequences were obtained which revealed the presence of two classes of clones, designated *arc5-I* and *arc5-II*, both with a length of 936 nucleotides [without the poly(A) tail; Fig. 2A]. The identity between the two copies is 98.3% at the DNA level. There are 14 nucleotide differences between the two clones, 8 of which result in a difference at the amino acid level. These cDNAs contain three potential translation initiation sites near the 5′ end. The first and second AUG codons probably do not function as initiators because

they are situated very close to the cap site (Kozak, 1989) and are surrounded by sequences unfavourable for translation initiation. Only the third AUG codon is flanked by consensus initiator sequences (ACAA**ATG**GCT; Joshi et al., 1987a) and is most probably the functional start codon. Identical situations were also found with the arcelin-1, the bean α-amylase inhibitor, and phytohemagglutinin genes, which all have short leader sequences containing several AUG codons. All of the start codons in these leader sequences are in suboptimum context and are not followed by terminator codons before the proposed/observed functional AUG is reached (Hoffman, 1984; Hoffman and Donaldson, 1985; Osborn et al., 1988b). Translation starting at the third AUG codon of arcelin-5 transcripts and ending at the UAG stop codon at position 797, which is also situated in a consensus sequence for termination of translation (Angenon et al., 1990), results in a coding sequence of 261 codons. The arcelin-5 3′ cDNA clones all end at the same position which is preceded by four consensus polyadenylation signals (AATAAA; Joshi et al., 1987b) between position 883 and 910. Both classes of clones encode proteins of 261 amino acids with signal peptides of 21 amino acids (Fig. 2B). The identity between the two copies at the level of the deduced amino acid sequences is 96.9%. The presence of the signal peptide was deduced upon comparison of the nucleotide-derived amino acid sequence and the N-terminal amino acid sequence of the arcelin-5 proteins. This signal peptide is also recognized as such using the method developed by von Heijne (1986) to identify a N-terminal secretory signal sequence and to predict the site of cleavage between the signal peptide and the mature protein. This signal peptide is very similar to and has the same length as those of arcelin-1, bean α-amylase inhibitor, and phytohemagglutinin proteins.

Transcript size

Northern blot analysis of total RNA from immature seeds of the G02771 accession confirmed the presence of arcelin-5 transcripts with a size of approximately 1000 nucleotides (data not shown).

Correlation between the two arcelin-5 cDNA sequences and the two major arcelin-5 polypeptides

The peptide sequences from the arcelin-5a protein exactly match the deduced amino acid sequence of *arc5-I* (Fig. 2B) which shows that *arc5-I* encodes the arcelin-5a protein. To confirm this and to determine whether *arc5-II* encodes the arcelin-5b protein, a mild hydrolysis with formic acid was performed on both the arcelin-5 proteins. This method cleaves proteins between aspartic acid (D) and proline (P) residues (Landon, 1977). The deduced amino acid sequences of *arc5-I* and *arc5-II* contain one and two DP sites, respectively (Fig. 2B). Separation of the formic-acid-treated proteins on SDS/PAGE showed that the cleavage pattern obtained for arcelin 5a and 5b correspond to those expected for *arc5-I* and *arc5-II*, respectively (data not shown). Sequence analysis of the generated peptides was also in agreement with this interpretation (Fig. 2B). Mild acid hydrolysis was also performed on the minor arcelin-5c polypeptide and sequence analysis showed the presence of two DP sites at the same position as in the amino acid sequence deduced from *arc5-II*.

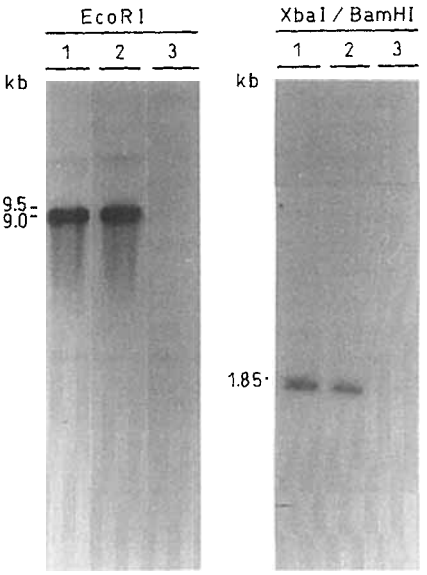

Fig. 4. Autoradiograph from a Southern blot after washing with 0.1 × NaCl/Cit, 0.1% SDS at 65°C. Total leaf DNAs cut with *Eco*RI (sizes in kb shown on the left) or *XbaI*/*Bam*HI (sizes in kb shown on the left) of G02771 (lane 1), BC1F1 (lane 2), and 'Ica Pijao' (lane 3) are shown. Lanes 1 and 2 of the *Eco*RI digest contain two bands close to each other of 9.0 and 9.5 kb.

Restriction fragment analysis

Southern blot analysis of the DNA from the G02771 accession indicated the presence of two different fragments hybridising with an arcelin-5 cDNA probe. Autoradiographs taken after low-stringency washing of the blots showed several fragments resulting from hybridisation of the arcelin-5 cDNA probe with arcelin-5 genes and genes encoding related proteins (data not shown). However, if autoradiographs were taken after high-stringency washing, only one or two fragments (depending on the restriction enzyme used) remained visible (Fig. 4). Some of the other bands hybridising to the arcelin-5 cDNA probe under low-stringency conditions correspond to phytohemagglutinin genes as they are visible after hybridisation with a phytohemagglutinin cDNA probe under high-stringency conditions (data not shown). If the arcelin-5 cDNA probe was used, most of the restriction enzymes (*Bam*HI, *Sac*I, *Sty*I, and *Xba*I) yielded one fragment (Fig. 4). However, one restriction enzyme, namely *Eco*RI, yielded two fragments of 9.0 and 9.5 kb (Fig. 4). This indicates that there are two arcelin-5 genomic regions, very similar to each other and most probably correlating with the two cDNA copies, *arc5-I* and *arc5-II*. These findings are valid for both the G02771 wild accession and the BC1F1 line (Fig. 4). On the other hand, the genome of 'Ica Pijao' does not contain these two similar arcelin-5 fragments (Fig. 4) which indicates that the presence of these two fragments is strictly correlated with the presence of the arcelin-5 proteins.

DISCUSSION

The purpose of this study was to isolate and characterise arcelin-5 proteins and corresponding cDNAs from seeds of a wild *P. vulgaris* accession with the aim of determining similarity and relationships of arcelin 5 with other arcelins and with lectins.

The arcelin variants 1, 2, 3 and 4, which had been characterised previously (Hartweck et al., 1991), all consist of sev-

Table 1. Molecular masses and isoelectric points of purified arcelins and phytohemagglutinin. Native protein molecular mass was determined by gel filtration; denatured and deglycosylated protein molecular masses were determined by SDS/PAGE (see Materials and Methods); n.d., not determined. Data for arcelin 1d, arcelin 1t, and phytohemagglutinin were obtained from Osborn et al. (1988a) and for arcelins 2, 3, and 4 from Hartweck et al. (1991).

Protein	Molecular mass of			pI
	native	denatured	deglycosylated	
	kDa			
Arcelin 1d	80.9	35.8–37.4	29.9	6.5
Arcelin 1t	159.6	36.4	n.d.	6.9
Arcelin 2	76.7	34.3	29.8	6.2
Arcelin 3	197.0	36.9	n.d.	n.d.
Arcelin 4	198.0	36.9	29.8	6.6
Arcelin 5	74.0	30.8–32.2	28.0–29.0	7.7
Phytohemagglutinin	146.1	30.6–38.5	28.9–30.1	5.2–5.4

Table 2. Sequence similarity at the nucleotide and protein level between arcelin 5 and the other members of the lectin gene family of *P. vulgaris*. Arc, arcelin; PHA-E, phytohemagglutinin E; PHA-L, phytohemagglutinin L; αAI, α-amylase inhibitor.

Protein	Similarity at nucleotide (protein) level with					
	Arc5-II	Arc1	Arc2	PHA-E	PHA-L	αAI
	%					
Arc5-I	98.3 (96.9)	81.9 (63.2)	81.9 (62.8)	73.9 (52.9)	73.5 (55.2)	77.4 (57.0)
Arc5-II		82.0 (63.2)	82.0 (62.8)	74.4 (53.6)	73.9 (55.6)	77.8 (57.8)
Arc1			99.4 (98.5)	77.1 (57.4)	79.5 (58.9)	81.6 (61.1)
Arc2				77.1 (57.4)	79.3 (59.2)	81.6 (60.7)
PHA-E					90.8 (86.1)	77.0 (54.5)
PHA-L						76.5 (59.8)

eral polypeptides. This was also found for the arcelin-5 variant, which consists of three strongly related polypeptides (5a, 5b, and 5c) with an identical N-terminal sequence and isoelectric point. They have molecular masses in the range of 32.2–30.8 kDa, which is quite distinct from those of the other arcelin variants (34–37 kDa; Table 1). Native arcelin-5 protein has a molecular mass corresponding to that of a dimer form which was also found for the arcelin variants arcelin 1d and arcelin 2, whereas the other arcelin variants (arcelin 1t, arcelin 3, and arcelin 4) and phytohemagglutinin are composed of tetramer proteins (Table 1).

To determine whether these polypeptides are encoded by different genes and are not just glycosylation variants of one gene product, cDNA sequences were determined. Two different arcelin-5 cDNA sequences were obtained, designated *arc5-I* and *arc5-II*. At the nucleotide level, the arcelin-5 coding sequence is very similar (73–82% identity; Table 2) to those of the genes encoding arcelin 1 and 2 (John and Long, 1990; Osborn et al., 1988b), phytohemagglutinin E and phytohemagglutinin L (Hoffman and Donaldson, 1985) and the bean α-amylase inhibitor (Hoffman, 1984). The sequence similarity at the amino acid level is 53% to 63% (Table 2). These comparisons show very clearly that arcelin 5 is related to the other arcelin variants and lectins of *P. vulgaris* but they also point out that arcelin 5 represents a new type of arcelin. Based on comparison of N-terminal sequences (Hartweck et al., 1991), arcelin 5 seems to be more related to arcelin 4 and the arcelin-1 tetramer forms than to arcelin 2 and the arcelin-1 dimer forms. Arcelins and the other members of the *P. vulgaris* lectin family have very similar DNA sequences but yet they encode proteins with different functions in the seeds; i.e. a lectin (phytohemagglutinin), an enzyme inhibitor (α-amylase inhibitor), and a putative insect resistance factor (arcelin). This is also reflected by the relatively low sequence similarity at the amino acid level (Table 2).

Southern blot analysis using an arcelin-5 cDNA probe revealed the presence of two similar arcelin-5 fragments, which probably correlate with the two cDNA copies, and confirmed the similarity with other lectins from *P. vulgaris*. The G02771 accession appears to contain a family of genes encoding arcelin 5, phytohemagglutinin and other related proteins, perhaps an α-amylase inhibitor. Identical situations can be expected for the lines containing the other arcelin variants. The SARC1 line (Osborn et al., 1988b) for example, a back-cross line of *P. vulgaris* expressing the arcelin-1 variant, contains at least genes encoding phytohemagglutinin, the arcelin-1 dimer form, and two different arcelin-1 tetramer forms (Hartweck et al., 1991).

Primary sequences of legume lectins exhibit remarkable similarities, with a significant number of invariant amino acid residues. Most of these are involved in carbohydrate binding or in metal binding (Ca^{2+} and Mn^{2+}) which, in turn, is necessary for binding of carbohydrates (Sharon and Lis, 1990). Sequence comparison shows that, in the case of arcelin 5, a number of these residues have been substituted while others have been deleted. To determine whether arcelin-5 protein still has lectin activity (i.e. carbohydrate bind-

ing activity), hemagglutination tests on purified arcelin-5 protein were performed. No hemagglutinating activity was observed with native erythrocytes whereas treatment of erythrocytes with pronase, which exposes potential binding sites, resulted in weak agglutinating activity of arcelin 5: the minimum concentration of arcelin 5 required for agglutination of pronase-treated erythrocytes was 500-fold that of phytohemagglutinin. Therefore, we cannot completely exclude the possibility that the observed agglutination is caused by phytohemagglutinin contamination. These results, together with the sequence comparison, indicate that arcelin-5 protein has no or very low carbohydrate-binding activity. It is also possible that arcelin 5 recognizes other sugar moieties not present on human erythrocytes.

A correlation between the different arcelin-5 polypeptides and the different cDNA sequences could be established by mild acid hydrolysis. This protein cleavage method revealed that arcelin 5a and 5b are encoded by *arc5-I* and *arc5-II*, respectively. The molecular mass of the polypeptides derived from the deduced amino acid sequences of *arc5-I* and *arc5-II* are almost identical (26.9 and 27.0 kDa, respectively). Therefore, the difference in molecular mass between the two major arcelin-5 proteins (as seen on SDS/PAGE) is likely to be due to post-translational modifications. Since arcelin 5a and 5b are both glycoproteins, the degree of glycosylation could be such a factor. Indeed, both proteins have a different number of potential glycosylation sites (NXT or NXS) as could be determined from the deduced amino acid sequences of *arc5-I* and *arc5-II* (Fig. 2B). Moreover, preliminary results obtained with electrospray mass spectrometry indicate the presence of two glycan chains on arcelin 5a, one glycan chain on arcelin 5b and no glycan chains on arcelin 5c (A. G. & B. De Vreese, unpublished results).

Mild acid hydrolysis of arcelin 5c yielded identical cleavage products as were found for arcelin 5b, excluding the possibility that it is encoded by *arc5-I* and indicating that it may be encoded by *arc5-II*. As carbohydrate-staining showed that arcelin 5c was not glycosylated, it could represent a non-glycosylated isoform of the arcelin 5b polypeptide. Alternatively, arcelin 5c could be encoded by a third copy of the arcelin-5 gene which is expressed at a much lower rate and was, therefore, not found by cDNA cloning.

A further question was whether arcelin 5a and 5b are allelic forms. This is not the case since individual F1 seeds obtained from crosses between 'Ica Pijao' and the G02771 accession all contain either no arcelin 5 or both 5a and 5b (C. Cardona, personal communication).

Glycosylation analysis of arcelin-5 proteins showed that glycan chains attached to arcelin 5a and 5b are most probably of the modified, fucosylated type, indicating that, in contrast to phytohemagglutinin and phaseolin, all oligosaccharide chains of arcelin-5 protein are accessible to modifying enzymes during the processing in the Golgi apparatus (Vitale et al., 1984; Faye et al., 1986; Sturm et al., 1987). This analysis also indicates that arcelin-5 polypeptides probably migrate more slowly on SDS/PAGE than expected from their real molecular mass. Assuming that arcelin 5a, 5b and 5c contain two, one and no glycan chains, respectively, one would expect polypeptides of approximately 29.3, 28.2, and 27.0 kDa, respectively. Such anomalous behaviour was also found for phytohemagglutinin; for example, phytohemagglutinin E of cv. Greensleeves should have a molecular mass of 30.6 kDa (taking into account the presence of one high-mannose and one modified oligosaccharide chain; Vitale et al., 1984) whereas it appears to be 34 kDa on SDS/PAGE (Brown et al., 1981b).

Arcelins are thought to be responsible for resistance against *Zabrotes subfasciatus*. Varying levels of resistance against this pest were found for the different arcelin variants, of which arcelin 1 and arcelin 5 seem to be the most promising in conferring insect resistance. These differences in resistance could be due in part to their biochemical properties but the relative quantities of these proteins in the seed also may be an important factor (Osborn et al., 1988b; Cardona et al., 1990). Further research will be conducted on the involvement of the arcelin-5 proteins in resistance towards *Z. subfasciatus*.

We thank Dr C. Cardona (CIAT, Cali, Colombia) for providing the *P. vulgaris* lines used in this study. This work is supported by a grant from the *Algemeen Bestuur van de Ontwikkelingssamenwerking*. A. G. and G. B. are indebted to the *Instituut ter Aanmoediging van het Wetenschappelijk Onderzoek in Nijverheid en Landbouw* and *Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie* for a predoctoral and a postdoctoral fellowship, respectively.

REFERENCES

Angenon, G., Van Montagu, M. & Depicker, A. (1990) Analysis of the stop codon context in plant nuclear genes, *FEBS Lett. 271*, 144–146.

Bauw, G., Van den Bulcke, M., Van Damme, J., Puype, M., Van Montagu, M. & Vandekerckhove, J. (1988) Protein electroblotting on polybase-coated glass fiber and polyvinylidene difluoride membranes: an evaluation, *J. Protein Chem. 7*, 194–196.

Bauw, G., Holm Rasmussen, H., Van den Bulcke, M., Van Damme, J., Puype, M., Gesser, B., Celis, J. E. & Vandekerckhove, J. (1990) Two-dimensional gel electrophoresis, protein electroblotting and microsequencing: a direct link between proteins and genes, *Electrophoresis 11*, 528–536.

Brown, J. W. S., Ma, Y., Bliss, F. A. & Hall, T. C. (1981a) Genetic variation in the subunits of globulin-1 storage protein of French bean, *Theor. Appl. Genet. 59*, 83–88.

Brown, J. W. S., Osborn, T. C., Bliss, F. A. & Hall, T. C. (1981b) Genetic variation in the subunits of globulin-2 and albumin seed proteins of French bean, *Theor. Appl. Genet. 60*, 245–250.

Cardona, C., Kornegay, J., Posso, C. E., Morales, F. & Ramirez, H. (1990) Comparative value of four arcelin variants in the development of dry bean lines resistant to the Mexican bean weevil, *Entomol. Exp. Appl. 56*, 197–206.

Chrispeels, M. J. & Raikhel, N. V. (1991) Lectins, lectin genes, and their role in plant defense, *Plant Cell 3*, 1–9.

Eckhardt, A. E., Hayes, C. E. & Goldstein, I. J. (1976) A sensitive fluorescent method for the detection of glycoproteins in polyacrylamide gels, *Anal. Biochem. 73*, 192–197.

Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E. Jr & Weber, P. (1981) Deglycosylation of glycoproteins by trifluoromethanesulfonic acid, *Anal. Biochem. 118*, 131–137.

Faye, L., Sturm, A., Bollini, R., Vitale, A. & Chrispeels, M. J. (1986) The position of the oligosaccharide side-chains of phytohemagglutinin and their accessibility to glycosidase determines their subsequent processing in the Golgi, *Eur. J. Biochem. 158*, 655–661.

Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer, *Proc. Natl Acad. Sci. USA 85*, 8998–9002.

Hames, B. D. & Rickwood, D. (1990) *Gel electrophoresis of proteins – a practical approach*, 2nd edn, Oxford University Press, Oxford.

Hartweck, L. M., Vogelzang, R. D. & Osborn, T. C. (1991) Characterization and comparison of arcelin seed protein variants from common bean, *Plant Physiol. 97*, 204–211.

Hoffman, L. M. (1984) Structure of a chromosomal *Phaseolus vulgaris* lectin gene and its transcript, *J. Mol. Appl. Genet. 2*, 447–453.

Hoffman, L. M. & Donaldson, D. D. (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome, *EMBO J. 4*, 883–889.

John, M. E. & Long, C. M. (1990) Sequence analysis of arcelin 2, a lectin-like plant protein, *Gene 86*, 171–176.

Jones, J. D. G., Dunsmuir, P. & Bedbrook, J. (1985) High level expression of introduced chimaeric genes in regenerated transformed plants, *EMBO J. 4*, 2411–2418.

Joshi, C. P. (1987a) An inspection of the domain between putative TATA box and translation start site in 79 plant genes, *Nucleic Acids Res. 15*, 6643-6653.

Joshi, C. P. (1987b) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis, *Nucleic Acids Res. 15*, 9627–9640.

Kozak, M. (1989) The scanning model for translation: an update, *J. Cell Biol. 108*, 229–241.

Landon, M. (1977) Cleavage of Asp-Pro bands, *Methods Enzymol. 47*, 145–149.

Lioi, L. & Bollini, R. (1989) Identification of a new arcelin variant in wild bean seeds, *Bean Improvement Cooperative 32*, 28.

Osborn, T. C., Blake, T., Gepts, P. & Bliss, F. A. (1986) Bean arcelin. 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L., *Theor. Appl. Genet. 71*, 847–855.

Osborn, T. C., Burow, M. & Bliss, F. A. (1988a) Purification and characterization of arcelin seed protein from common bean, *Plant Physiol. 86*, 399–405.

Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C. & Bliss, F. A. (1988b) Insecticidal activity and lectin homology of arcelin seed protein, *Science 240*, 207–210.

Romero Andreas, J., Yandell, B. S. & Bliss, F. A. (1986) Bean arcelin. 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition, *Theor. Appl. Genet. 72*, 123–128.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science 239*, 487–491.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning, a laboratory manual*, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Santino, A., Valsasina, B., Lioi, L., Vitale, A. & Bollini, R. (1991) Bean (*Phaseolus vulgaris* L.) seed lectins: a novel electrophoretic variant of arcelin, *Plant Physiol. (Life Sci. Adv.) 10*, 7–11.

Schoonhoven, A. V., Cardona, C. & Valor, J. (1983) Resistance to the bean weevil and the Mexican bean weevil (Coleoptera: Bruchidae) in noncultivated common bean accessions, *J. Econ. Entomol. 76*, 1255–1259.

Sharon, N. & Lis, H. (1990) Legume lectins – a large family of homologous proteins, *FASEB J. 4*, 3198–3208.

Sturm, A., Van Kuik, J. A., Vliegenthart, J. F. G. & Chrispeels, M. J. (1987) Structure, position, and biosynthesis of the high-mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin, *J. Biol. Chem. 262*, 13392–13403.

Sturm, A. & Chrispeels, M. J. (1986) The high mannose oligosaccharide of phytohemagglutinin is attached to asparagine 12 and the modified oligosaccharide to asparagine 60, *Plant Physiol. 81*, 320–322.

Tretter, V., Altmann, F. & März, L. (1991) Peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached $\alpha 1 \rightarrow 3$ to the asparagine-linked *N*-acetylglucosamine residue, *Eur. J. Biochem. 199*, 647–652.

Vitale, A., Warner, T. G. & Chrispeels, M. J. (1984) *Phaseolus vulgaris* phytohemagglutinin contains high-mannose and modified oligosaccharide chains, *Planta 160*, 256–263.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites, *Nucleic Acids Res. 14*, 4683–4690.