Isolation and characterization of a minor legumin and its constituent polypeptides from *Pisum sativum* (pea)

John F. MARCH,* Darryl J. C. PAPPIN,† and Rod CASEY*‡

*John Innes Institute, Colney Lane, Norwich NR4 7UH, and †Department of Biochemistry, University of Leeds, Leeds LS2 9JT, U.K.

The purification and characterization of a minor legumin species from *Pisum sativum* is described. Electrophoretic data indicate that it corresponds to a legumin subunit pair previously designated L1. The β -polypeptides of the minor legumin have a phenylalanine *N*-terminus. This is the first time that an amino acid other than glycine has been reported as the *N*-terminus of the basic polypeptides from legumin-like proteins from any plant species. Sequence analyses of the isolated α - and β -polypeptides of the minor legumin show that it does not correspond to any of the three legumin gene families that have previously been defined on the basis of DNA hybridizations and genetic analyses.

INTRODUCTION

Legumin, one of the major storage proteins of pea (*Pisum*) seeds, is an 11 S species that contains six subunit pairs, each of which consists of disulphide-linked acidic (α -, $M_r \sim 40000$) and basic (β -, $M_r \sim 20000$) polypeptides (Derbyshire *et al.*, 1976; Casey, 1979a). The pairing between different α - and β -polypeptides is specific (Domoney, 1981; Matta *et al.*, 1981), this specificity being a consequence of the synthesis of legumin as a covalently linked precursor (Croy *et al.*, 1980; Spencer & Higgins, 1980) of the form $\text{NH}_2\text{-}\alpha\text{-}\beta\text{-CO}_2\text{H}$ (Croy *et al.*, 1982). A single disulphide bond is formed between the α - and β -polypeptide sequences within the precursor (Croy *et al.*, 1980) and six of these disulphide-linked species are then combined to produce the final 11 S protein, each morphological subunit (Casey *et al.*, 1980) therefore being an α - β dimer of $M_r \sim 60000$.

A number of minor forms of *Pisum* legumin have been described. Minor species of *Pisum* legumin α -polypeptide (the so-called α^m -polypeptides) have been recognized on O'Farrell-type two-dimensional gels (O'Farrell, 1975; Casey, 1979b), the α^m -polypeptides as a group having a greater apparent M_r and a lower pI than their major counterparts (α^M -polypeptides). Matta *et al.* (1981), using a number of electrophoretic techniques, demonstrated the existence of several minor disulphide-linked α - β pairs (named L1, L2, L3 and L5) in *P. sativum*, in addition to the major (L4) subunit pairs. The distribution of subunit pairs within various molecular forms of legumin that could be separated under non-denaturing conditions led Matta *et al.* (1981) to designate subunit pairs L1–L3 as minor 'big' legumin species and subunit pairs L5 as minor 'small' legumin species.

Hybridization studies using three cDNA probes (pCD32, pCD40 and pCD43) indicated the existence of at least three classes of legumin gene in *Pisum* (Domoney & Casey, 1985). Sequencing of the three cDNAs showed that pCD40 and pCD32 were more similar to each other than to pCD43 (Domoney *et al.*, 1986a). The three cDNAs correspond to precursor molecules of different size; pCD43 corresponds to precursors of $M_r \sim 60000$,

pCD40 to precursors of M_r 63000–65000 and pCD32 to large precursors of $M_r \sim 80000$ (Domoney & Casey, 1984, 1985). The cDNAs have been used to establish the existence of polymorphisms in the sizes of the restriction fragments carrying legumin genes (Domoney *et al.*, 1986b). Segregation analysis of these fragments with respect to other markers showed that the genes corresponding to pCD43 are at a locus on chromosome 7, near *r* and therefore probably correspond to the locus for the major legumin polypeptides (Davies, 1980). Other legumin genes (those corresponding to pCD40 and at least one corresponding to pCD32) have been mapped to chromosome 1 (Domoney *et al.*, 1986b); no polypeptides, major or minor, have been mapped to chromosome 1.

In the absence of genetic linkage data for minor legumin polypeptides, an alternative means of relating polypeptides to cDNA clones (and cloned genes) is through sequence analyses, ideally within one genotype. This paper describes: the isolation of a minor legumin species, which is distinct from the minor species containing α^m -polypeptides, from *P. sativum* cv. 'Birte'; determination of the *N*-terminal 58 and 70 amino acids of its α - and β -polypeptides, respectively; comparison of these sequences with the homologous regions of the amino acid sequences derived from cDNA sequences (also from cv. 'Birte') and genomic DNA sequences corresponding to the genes on *P. sativum* chromosomes 1 and 7 (Lycett *et al.*, 1984; Domoney *et al.*, 1986a,b; Gatehouse *et al.*, 1988). The results indicate that the minor legumin cannot be the product of any of the legumin gene sequences so far identified in *Pisum* and is the product of a fourth class of *Pisum* legumin gene.

EXPERIMENTAL

Materials

Sephadex G-200 (40–120 μm) was from Pharmacia, Milton Keynes, U.K., hydroxyapatite (Bio-Gel HTP) from Bio-Rad Laboratories, Watford, U.K., and DEAE-Trisacryl M from LKB Instruments, Croydon, U.K. Molecular-mass markers for SDS gels were SDS-6H and

Abbreviation used: DTT, dithiothreitol.

‡ To whom correspondence should be addressed.

SDS-7 from Sigma, Poole, Dorset, U.K. Aristar urea was from BDH Chemicals, Poole, Dorset, U.K. All other chemicals were as described by Casey (1979a).

Purification of minor legumin

Ground sieved meal of dry mature seeds of *P. sativum* cv. 'Birte' was extracted in 0.5 M-NaCl/0.1 M-sodium phosphate buffer/0.1 mM-dithiothreitol (DTT)/0.02 % (w/v) NaN_3 , pH 8.0, and the material precipitating between 40 and 60 % saturation of ammonium sulphate at 5 °C (initial protein concentration $\sim 10 \text{ mg} \cdot \text{ml}^{-1}$) was prepared as before (Casey, 1979a).

The ammonium sulphate pellet was dissolved in 0.1 M-sodium phosphate buffer/0.1 mM-DTT/0.02 % (w/v) NaN_3 , pH 8.0 (57 ml/100 g of starting material) and applied to a column (5 cm $\times$ 79 cm) of Sephadex G-200 in the same buffer. The column was developed at $36 \text{ ml} \cdot \text{h}^{-1}$ at 20 °C, the eluate was monitored at 280 nm and fractions were collected at 15 min intervals (9 ml). Fractions were analysed by SDS-gel electrophoresis (see later) and protein was recovered from those fractions that were enriched in the minor legumin (see Fig. 1a) by precipitation with 90 %-saturated ammonium sulphate at 5 °C.

The material from the G-200 column was dissolved in 0.1 M-sodium phosphate buffer/0.1 mM-DTT/0.02 % (w/v) NaN_3 , pH 8.0 (15 ml/100 g of starting material), dialysed exhaustively against the same buffer at 5 °C and loaded onto a column (2.6 cm $\times$ 26 cm) of hydroxyapatite that had been equilibrated with the same buffer at 20 °C. After elution of unbound material, the column was developed with a linear gradient from 0.1–0.5 M-sodium phosphate buffer/0.1 mM-DTT/0.02 % (w/v) NaN_3 , pH 8.0 at 20 °C and $50 \text{ ml} \cdot \text{h}^{-1}$. The total gradient volume was 1 litre and fractions (12.5 ml) were collected at 15 min intervals. The fractions were examined by electrophoresis and protein was recovered from those rich in the minor legumin (see Fig. 1b).

Isolation of minor legumin polypeptides

Minor legumin from hydroxyapatite was dissolved to a final concn. of $10 \text{ mg} \cdot \text{ml}^{-1}$ in 8 M-urea/0.05 M-HCl/0.1 M-DTT, pH 8.6 at 20 °C (adjusted with Tris), incubated for 1 h at 20 °C, dialysed against two 250 ml portions of 8 M-urea/0.05 M-HCl/0.1 mM-DTT, pH 8.6 at 20 °C (adjusted with Tris) and applied to a column (1.5 cm $\times$ 24 cm) of DEAE-Trisacryl M in the dialysis buffer at 20 °C. Unbound material was eluted at $60 \text{ ml} \cdot \text{h}^{-1}$ with the same solution and fractions (5 ml) were collected at 5 min intervals. The column was then developed with a linear NaCl gradient (0–0.3 M) in 8 M-urea/0.05 M-HCl/0.1 mM-DTT, pH 8.6 (adjusted with Tris) at 20 °C and $36 \text{ ml} \cdot \text{h}^{-1}$. The total gradient volume was 600 ml and fractions (5.4 ml) were collected at 9 min intervals. Selected fractions corresponding to α - and β -polypeptides (see Fig. 1c) were dialysed exhaustively against water at 5 °C and freeze-dried.

Electrophoretic techniques

Two-dimensional isofocusing/SDS-gel electrophoresis and one-dimensional SDS-gel electrophoresis [15 % (w/v) acrylamide] were as previously described (Casey, 1979a). pH gradients were measured by extracting slices of gel from the first dimension in degassed water (Matta *et al.*, 1981). If gel slices were extracted in urea before measuring the pH gradient (Casey, 1979a), the apparent

Fig. 1. Chromatographic purification of the *P. sativum* minor legumin and its constituent polypeptides

(a) Gel filtration on Sephadex G-200 of the 40–60 % saturated ammonium sulphate pellet. (b) Ion-exchange chromatography on hydroxyapatite of the material from (a). (c) Ion-exchange chromatography on DEAE-Trisacryl M in 8 M-urea of the material from (b). —, Absorbance at 280 nm; ---, [phosphate] (b) or [NaCl] (c). The cross-hatched areas represent the fractions taken for further purification (a) and (b) or characterization (c). The arrow in (c) indicates a change in fraction size and flow-rate. See text for further details.

pI was increased by about one pH unit in the region where the α -polypeptides focused (see also Bull *et al.*, 1964; Casey, 1979b).

Two-dimensional non-reducing/reducing SDS-gel electrophoresis was as described by Matta *et al.* (1981) except that the first dimension gel was 12 % (w/v) acrylamide, the second dimension 15 % (w/v) acrylamide, and the first dimension (2 mm $\times$ 12.5 cm) was performed in tubes for 16 h at $3 \text{ V} \cdot \text{cm}^{-1}$. The positions of the minor legumin polypeptides were identified on two-dimensional gels of whole-seed extracts by adding known amounts of the purified legumin to whole-seed extracts.

Gels were stained for carbohydrate using periodic acid/Schiff reagent as described by Kapitany & Zebrowski (1973).

Structural characterization of minor legumin polypeptides

For automated N-terminal sequence analyses, the S-carboxymethyl- α - and β -polypeptides (0.5 mg) were dissolved in $50 \mu\text{l}$ of 0.2 M- NaHCO_3 /0.25 % (w/v) SDS, coupled to 15 mg of *p*-phenylenedi-isothiocyanate glass (17 nm pore size, 200–400 mesh) (Wachter *et al.*, 1973) at 56 °C for 60 min under nitrogen and then washed with water and methanol [containing 0.5 % (v/v) n-propyl-

Fig. 2. Gel electrophoretic analyses of whole protein or purified minor legumin polypeptides from *P. sativum* cv. 'Birte'

(a) 15 % (w/v) SDS gel. Lane 1, whole protein extract equivalent to 0.5 mg of dry seed meal; lane 2, minor legumin ($\sim 50 \mu\text{g}$); lane 3, minor legumin α -polypeptide ($\sim 15 \mu\text{g}$); lane 4, minor legumin β -polypeptide ($\sim 15 \mu\text{g}$). The open arrows indicate the positions of the major polypeptides of legumin within the whole protein pattern. (b) Two-dimensional non-reducing 12 % (w/v)/reducing 15 % (w/v) SDS gel. The equivalent of 1 mg of dry seed meal was loaded. The closed and open arrows indicate the polypeptides of the minor and major legumin subunit pairs respectively. (c) As (b) except minor legumin ($24 \mu\text{g}$). (d) Two-dimensional isofocusing/SDS gel. The equivalent of 2 mg of dry seed meal was loaded. The boxed polypeptides indicated by closed arrows are the α - and β -polypeptides of the minor legumin and the α^m -polypeptides (Casey, 1979b); the open arrow indicates the major legumin α -polypeptides. (e) As (d) except minor legumin ($20 \mu\text{g}$) and the boxes indicate the polypeptides that were selected (see Fig. 1c) for the sequence analyses reported in Fig. 3. All the gels were stained with Coomassie Brilliant Blue G-250; the α -polypeptides stain more intensely than equivalent molar amounts of β -polypeptides. The numbers at the side of each section indicate the positions adopted by marker proteins of known M_r . I.e.f., isoelectric focusing.

amine]. The coupled polypeptides were sequenced as described by Laursen (1971) using a modified Anachem sequencer. Anilinothiazolinone amino acids were converted for 17 min at 70°C under nitrogen in $35 \mu\text{l}$ of 35 % (v/v) aqueous trifluoroacetic acid and the resulting phenylthiohydantoin amino acids identified by reverse-phase h.p.l.c. on a Brownlee Spheri-5 C_{18} column ($220 \text{ mm} \times 2.1 \text{ mm}$) using a method adapted from Zimmerman *et al.* (1977). Phenylthiohydantoin amino acids were quantified at 269 nm using a Hewlett-Packard HP7994A analytical work station. Serine and threonine residues were confirmed by the detection of their dehydro derivatives at 313 nm.

RESULTS AND DISCUSSION

The minor legumin was purified by an initial ammonium sulphate fractionation followed by gel filtra-

tion and ion-exchange chromatography, as shown in Figs. 1(a) and 1(b). The SDS-gel electrophoresis pattern of the purified minor legumin is shown in Fig. 2(a), lane 2. The apparent M_r of the α - and β -polypeptides of the minor legumin were estimated, by SDS-gel electrophoresis, to be $\sim 43\,500$ (α) and $\sim 20\,000$ (β). The minor legumin α -polypeptides are appreciably larger than the major legumin α -polypeptides, whereas the minor legumin β -polypeptides are apparently smaller than their major counterparts (Fig. 2a, lanes 1 and 2). Figs. 2(a), lane 1, 2(b) and 2(d) give an indication of the relatively small proportion of the total legumin represented by the minor legumin; in view of the widely disparate amounts of the minor and major legumins present in a total seed extract, no attempt has been made to quantify their relative proportions by densitometry of SDS gels.

Periodate/Schiff staining of one-dimensional SDS gels containing the minor legumin and equivalent amounts of *Pisum* vicilin gave a strong positive response for vicilin

		α -Polypeptide											Identity (%)	
		5	10	15	20	25	30	35	40	45	50	55		
Minor legumin		RQLPRQPSELYECQLDTHAREPDNRYDSEAGFTETWDXTXPDLCAGVSVLKLTINP												(100)
<i>LegA</i>		LREQPQQ-NECQLERLDALEPDNRIESEGGLIETWNPNNKQFRACAGVALSRATLQR												47
<i>LegJ</i>		TSSEFDRNLQCQLDSINALEPDHRVSEAGLTETWNPNHPEAKCAGVSLIRRTIDP												51

		β -Polypeptide														Identity (%)	
		5	10	15	20	25	30	35	40	45	50	55	60	65	70		
Minor legumin		FLEETVCTKLKHLHEDLAGSSQADVFNPRAGRITSVNSLTLPVLKLLHLSAQWVKLYKNGIFMFWNLNANS															(100)
pCD43		GLEETVCTAKLRLNIGPSSSPDIYNPEAGRIKTVTSLDLPVLRWLKLSAEHGSLSHKAMFVPHYNLNANS															59
pCD40		GLEETICSAKIRENIADAAGADLYNPRAGRITANSLTLPVLRYLRLSAEYVRLYRNGIYAPHWNINANS															61
pCD32		GLEETICSAKIRENIARPSRGDLNYSAGRITSVNSLTLPILRNRLRLSAEYVLLYRNGIYAPHWNINANS															59

Fig. 3. Comparison of the *N*-terminal sequences of the α - and β -polypeptide from the *P. sativum* minor legumin with other *Pisum* legumin polypeptides predicted from DNA sequences

There are three uncertainties in the *Pisum* minor legumin α -polypeptide sequence and seven in the β -polypeptide (indicated by dots). Homologies are therefore calculated as the percentage of identical amino acids in the definitive assignments using the alignments shown. The cDNA plasmids (pCD32, pCD40 and pCD43) used for the β -polypeptide sequence comparisons hybridize to three different *Pisum* legumin gene classes (Domoney & Casey, 1985) and have been described in detail elsewhere (Domoney *et al.*, 1986a; see also Introduction). None of these plasmids include sequences corresponding to the *N*-terminals of the α -polypeptides and thus sequences of *Pisum* genes corresponding to pCD40 (*legJ*; Gatehouse *et al.*, 1988) and pCD43 (*legA*; Lycett *et al.*, 1984) have been used for comparison with the minor legumin α -polypeptide. *Pisum* genes corresponding to pCD32 have yet to be isolated. — = gap introduced to optimize alignment.

polypeptides but no reaction with the minor legumin polypeptides, suggesting that the minor legumin, like the major *Pisum* legumin (Casey, 1979a; Gatehouse *et al.*, 1980) contains little, if any, periodate-oxidizable, covalently bound carbohydrate.

In non-reduced form, the SDS-denatured minor-legumin subunit pairs have an apparent M_r slightly greater than those of the majority of legumin subunit pairs (Fig. 2b). The positions adopted by the minor legumin α - and β -polypeptides on a non-reducing/reducing two-dimensional SDS gel (indicated in Fig. 2b by the closed arrows) suggest that it may represent the subunit pair L1 or L2 described by Matta *et al.* (1981) for another *Pisum* genotype. On a two-dimensional isofocusing/SDS gel (Figs. 2d and 2e) the minor legumin α -polypeptides at $M_r \sim 43\,500$ (indicated by a closed arrow in Fig. 2d) are heterogeneous with respect to charge and adopt a position consistent with their having a higher pI than the majority of the legumin α -polypeptides (indicated in Fig. 2d by an open arrow). They do not correspond to the α^m -polypeptides described by Casey (1979b), which have a lower pI (see Fig. 2d), and their relatively high pI is consistent with their derivation from the L1, rather than the L2, subunit pair (Matta *et al.*, 1981). The pI of the minor legumin α -polypeptides was 5.6–6.0 and that of the major legumin α -polypeptides 5.4–5.6. While there is considerable genetic variation in the apparent pI of *Pisum* legumin polypeptides (Casey, 1979b), we conclude that the minor legumin α -polypeptides correspond to the L1 subunit pair (Matta *et al.*, 1981).

The minor legumin β -polypeptides of $M_r \sim 20\,000$ are represented on standard isofocusing/SDS gels (Figs. 2d and 2e) and have an apparent pI of 6.0–6.3. This is in contrast with the β -polypeptides of the major legumin, the pI values of which are sufficiently high that they do not appear on the type of two-dimensional gel shown in Figs. 2(d) and 2(e) (Casey, 1979b; Croy *et al.*, 1979; Krishna *et al.*, 1979; Gatehouse *et al.*, 1980). Such a relatively low pI for the minor legumin β -polypeptides is

again consistent with the minor legumin being equivalent to the L1 subunit pair, which has β -polypeptides of pI 6.22–6.35 (Matta *et al.*, 1981).

The α - and β -polypeptides were separated by ion-exchange chromatography in 8 M-urea and the predominant peak taken from each polypeptide class (Fig. 1c) in an attempt to simplify subsequent sequence analyses. Two-dimensional gel analysis (Fig. 2e), however, showed that the β -polypeptides selected for sequencing were virtually identical to those in the unfractionated minor legumin. The selection from the α -polypeptide peak was for three or four of the less-negatively-charged polypeptides (Fig. 2e) from the minor legumin preparation. Thus in both cases a mixture of polypeptides was subjected to sequence analysis, but no appreciable heterogeneity was observed during sequencing, suggesting that the polypeptides within each group have highly homologous sequences.

Analysis of the *N*-terminal amino acids of the purified minor legumin α - and β -polypeptides showed the α -polypeptides to have *N*-terminal arginine and the β -polypeptides to have *N*-terminal phenylalanine. These are both unusual for legumin-like proteins and in the case of the β -polypeptide this is the first report of an *N*-terminus which is not glycine.

Comparison of the *N*-terminal amino acid sequences of the minor legumin α - and β -polypeptides (Fig. 3) with the corresponding regions of the acidic and basic polypeptides predicted from pea legumin cDNA plasmids (Domoney *et al.*, 1986a) or genes (Lycett *et al.*, 1984; Gatehouse *et al.*, 1988) shows that the minor legumin sequence differs appreciably from all of the known pea legumin sequences.

pCD43 hybridizes to the four or five members of the *legA* class under stringent conditions (Domoney & Casey, 1985). Thus the minor legumin, showing only 59% identity with 70 of the 80 β -polypeptide amino acids encoded by pCD43, is not a product of the *legA* class of genes. Similar considerations apply to the gene classes represented by pCD40 and pCD32 (Domoney &

Casey, 1985). The minor legumin polypeptides characterized here therefore represent the products of a fourth class of *Pisum* legumin gene, distinct from those previously documented. The minor legumin, which seems to be the equivalent of the L1 subunit pair described by Matta *et al.* (1981) for another genotype, appears to be as different from other classes of legumin in the same genotype as previously characterized classes are from each other.

This work was supported by the Agricultural and Food Research Council via a grant-in-aid to the John Innes Institute and by the CEC Biomolecular Engineering Programme under contract GBI-4-113-UK. We thank Dr C. Domoney for useful comments on the manuscript.

REFERENCES

- Bull, H. B., Breese, K., Ferguson, G. L. & Swenson, C. A. (1964) *Arch. Biochem. Biophys.* **104**, 297–304
- Casey, R. (1979*a*) *Biochem. J.* **177**, 509–520
- Casey, R. (1979*b*) *Heredity* **43**, 265–272
- Casey, R., Hills, G. J. & Shaw, P. J. (1980) *John Innes Annu. Rep.* 1980, pp. 20–21
- Croy, R. R. D., Derbyshire, E., Krishna, T. G. & Boulter, D. (1979) *New Phytol.* **83**, 29–35
- Croy, R. R. D., Gatehouse, J. A., Evans, I. M. & Boulter, D. (1980) *Planta* **148**, 49–56
- Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N. & Boulter, D. (1982) *Nature (London)* **295**, 76–79
- Davies, D. R. (1980) *Biochem. Genet.* **18**, 1207–1219
- Derbyshire, E., Wright, D. J. & Boulter, D. (1976) *Phytochemistry* **15**, 3–24
- Domoney, C. (1981) Ph.D. Thesis, University of East Anglia
- Domoney, C. & Casey, R. (1984) *Eur. J. Biochem.* **139**, 321–327
- Domoney, C. & Casey, R. (1985) *Nucleic Acids Res.* **13**, 687–699
- Domoney, C., Barker, D. & Casey, R. (1986*a*) *Plant Mol. Biol.* **7**, 467–474
- Domoney, C., Ellis, T. H. N. & Davies, D. R. (1986*b*) *MGG Mol. Gen. Genet.* **202**, 280–285
- Gatehouse, J. A., Croy, R. R. D. & Boulter, D. (1980) *Biochem. J.* **185**, 497–503
- Gatehouse, J. A., Bown, D., Levasseur, M., Castleton, J. & Ellis, T. H. N. (1988) *Biochem. J.* **250**, 15–24
- Kapitany, R. A. & Zebrowski, E. J. (1973) *Anal. Biochem.* **56**, 361–369
- Krishna, T. G., Croy, R. R. D. & Boulter, D. (1979) *Phytochemistry* **18**, 1879–1880
- Laursen, R. A. (1971) *Eur. J. Biochem.* **20**, 89–102
- Lycett, G. W., Croy, R. R. D., Shirsat, A. H. & Boulter, D. (1984) *Nucleic Acids Res.* **12**, 4493–4506
- Matta, N. K., Gatehouse, J. A. & Boulter, D. (1981) *J. Exp. Bot.* **32**, 1295–1307
- O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021
- Spencer, D. & Higgins, T. J. V. (1980) *Biochem. Int.* **1**, 502–509
- Wachter, E., Machleidt, W., Hofner, H. & Otto, J. (1973) *FEBS Lett.* **35**, 97–102
- Zimmerman, C. L., Appella, E. & Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573

Received 31 July 1987/13 October 1987; accepted 18 November 1987