Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.)

John A. GATEHOUSE, Grantley W. LYCETT, Ashton J. DELAUNEY, Ronald R. D. CROY and Donald BOULTER

Department of Botany, University of Durham, Durham DH1 3LE, U.K.

(Received 10 January 1983/Accepted 10 March 1983)

Amino acid sequence data from vicilin of pea (*Pisum sativum* L.) were compared with predicted sequences from complementary DNA species. The sites of potential post-translational proteolytic cleavage of vicilin precursor polypeptides were located in polar regions of the polypeptide, at acidic or amide residues. Proteolysis did not take place in precursors containing a functionally distinct sequence:

neutral residue–hydrophobic residue–basic residue

at the cleavage site. Differences between the genomic sequences encoding vicilin thus specify proteolytic cleavage of vicilin precursor polypeptides.

Vicilin is a major component of the storage proteins of pea (*Pisum sativum* L.) seeds. It is a heterogeneous protein of M_r approx. 150 000–170 000 that has been suggested to contain three subunits of M_r approx. 50 000 (Gatehouse *et al.*, 1981). Some of these subunits contain ‘nicks’ in the polypeptide chains caused by specific post-translational proteolysis, leading to the observation on dissociating gel electrophoresis that the protein contains polypeptides of M_r less than 50 000 (33 000, 19 000, 16 000, 13 500 and 12 500 as major components). A model for the derivation of these subunit fragments from 50 000- M_r polypeptide precursors has been presented (Gatehouse *et al.*, 1982) and is supported by data demonstrating immunological relationships between protein fragments (Davey *et al.*, 1981). The proteolysis of vicilin precursors that takes place could either be the result of incomplete cleavage of essentially identical 50 000- M_r precursors at certain regions of the polypeptide to generate the observed fragments whilst leaving a significant proportion of precursors intact, or could be due to the synthesis of a set of similar precursors containing different amino acid sequences which determine that some precursors are proteolytically cleaved whilst others are not. In the latter case the pea genome would contain a number of vicilin genes differing in their coding sequences at the potential cleavage sites, whereas in the former case, although different vicilin genes might be present, sequence differences at the potential cleavage sites would not necessarily be expected.

Abbreviations used: cDNA, complementary DNA; h.p.l.c., high-pressure liquid chromatography; $1 \times \text{SSC}$, 0.15 M-NaCl/0.015 M-sodium citrate buffer, pH 7.0.

Partial proteolytic cleavage of vicilin precursors is not supported by previous observations (Gatehouse *et al.*, 1981) that one of the vicilin approx.-50 000- M_r precursors (referred to as ‘47 000- M_r precursor’) is completely cleaved by post-translational proteolysis, but cannot be eliminated.

The present paper reports nucleotide and amino acid sequence data from vicilin cDNA species and polypeptides from the regions of potential proteolytic cleavage.

Experimental

The isolation and characterization of cDNA clones pDUB2 and pDUB4 (previously referred to as pRC 2.2.1 and 2.2.10 respectively), containing inserts of approx. 900 and 300 base-pairs in pBR322, have been described previously (Croy *et al.*, 1982).

The inserts from cDNA plasmids were excised by restriction with *Bam*HI endonuclease (BRL Ltd., Cambridge, U.K.), end-labelled with [γ - ^{32}P]ATP (Amersham International) (Maxam & Gilbert, 1980), and purified by agarose-gel electrophoresis. They were isolated by electrophoresis on to DEAE-paper (DE 81; Whatman) followed by washing, elution and ethanol precipitation (Dretzen *et al.*, 1981). The insert from pDUB4 was used as a probe in a colony-screening experiment (Grunstein & Wallis, 1979) on a bank of cDNA clones produced from mRNA isolated from developing pea cotyledons as previously described (Croy *et al.*, 1982; Evans *et al.*, 1980). pDUB7 was selected by a positive hybridization to the probe, and was subsequently characterized as an insert of approx. 1150 base-pairs in pBR322. cDNA species were char-

acterized by restriction mapping (results not shown) and were sequenced by the method of Seif *et al.* (1980).

Pea genomic DNA was prepared from leaves and developing cotyledons by extraction and CsCl gradient centrifugation as described by Graham (1978). Samples were digested to completion with restriction endonucleases and the resultant fragments were separated by agarose-gel electrophoresis, transferred to nitrocellulose paper, hybridized at 65°C in $5 \times \text{SSC}$ with the inserts from plasmids pDUB2 and pDUB4, and nick-translated to a specific radioactivity of $\geq 2 \times 10^7 \text{ c.p.m./}\mu\text{g}$ by using reagents and $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ from Amersham International (Amersham International, 1981). After hybridization to the nitrocellulose filters were washed according to the protocol of Southern (1979) to a stringency of $0.1 \times \text{SSC}$ at 65°C. The filters were dried and autoradiographed at -80°C with an intensifying screen.

Purification of vicilin and its component polypeptides, digestion conditions and reverse-phase h.p.l.c. separation and manual amino acid microsequence analysis by the methods of Chang *et al.* (1978) of peptides produced from them by proteolytic cleavage were as previously described (Hirano *et al.*, 1982).

Results and discussion

Two cDNA species encoding vicilin polypeptides formed the basis of the nucleotide sequence data reported in the present paper (see Figs. 1 and 2 for restriction maps and sequencing details). Clone pDUB2 (previously referred to as pRC 2.2.1) has been shown to hybrid-select mRNA species coding for a vicilin 50000- M_r precursor (Croy *et al.*, 1982) that was not shown to undergo post-translational proteolysis *in vivo* (Gatehouse *et al.*, 1981). A second cDNA clone, pDUB7, was selected from a cDNA bank by specific hybridization to pDUB4 (previously referred to as 'pRC 2.2.10'), which has been shown to hybrid-select mRNA species coding for a vicilin approx.-50000- M_r precursor ('47000 M_r ') (Croy *et al.*, 1982) that is proteolytically cleaves *in vivo* to give a 33000- M_r polypeptide as one product (Gatehouse *et al.*, 1981). pDUB4 was shown to be identical in sequence with a part of pDUB7, and the sequences of these cDNA clones were significantly different from that of pDUB2. Full sequence data for cDNA species pDUB2 and pDUB7 have been published (Lycett *et al.*, 1983).

Amino acid-sequence data were derived from

Fig. 1. Restriction map of, and sequencing strategy for, pDUB2 and pDUB7 cDNA inserts

The scale at the top shows the regions coding for the α -, β - and γ -subunits and their relationship and M_r values ($\alpha + \beta = 33000$; $\alpha + \beta + \gamma = 50000$). The arrows show the direction and length of sequence obtained from individual sequencing gels. The restriction map shows sites for only those enzymes used in the sequencing procedure. Abbreviations for restriction endonucleases are: L, *Bam*HI linker sequence; X, *Xba*I; B, *Bgl*II; N, *Bst*NI; E, *Bst*EII; F, *Hin*FI; C, *Hinc*II; H, *Hae*III; also kb, kilobases.

Sequences predicted by cDNA species pDUB2 and pDUB7 are compared with sequences derived from vicilin polypeptides. The one-letter notation for amino acids is given in Table 1. L/I indicates leucine or isoleucine (not distinguishable by t.l.c. residue identification method). The $\beta(1-3,500)$ -subunit sequence is taken from Hirano *et al.* (1982). Residues joined by solid horizontal lines (—) with vertical bars representing their beginnings and ends. Broken lines (----) represent inferred residues; other residues were determined by single N-terminal analyses. $\Leftarrow$ represents a direct C-terminal determination (Hirano *et al.*, 1982). A 20–50nmol portion of each peptide was sequenced; peptides: $\rightarrow$, tryptic; $\rightarrow$, *S. aureus* V8 proteolytic; $\rightarrow$, chymotryptic.

peptides obtained from purified vicilin polypeptides by cleavage with trypsin, chymotrypsin and *Staphylococcus aureus* V8 proteinase (for details, see Fig. 2). The peptides were arranged into a partially complete sequence for the vicilin precursors by amino acid-sequence overlap or by comparison with sequences predicted by the cDNA species. The predicted and determined amino acid sequences in one region of potential post-translational proteolytic cleavage of vicilin are shown in Fig. 2. The vicilin precursor is divided into three sections, α , β and γ , separated by two sites of potential proteolytic cleavage, $(\alpha:\beta)$ and $(\beta:\gamma)$. The derivation of the vicilin polypeptides from the precursors was described previously (Gatehouse *et al.*, 1982) and is repeated in Fig. 1. If pDUB2 encodes a vicilin precursor that is not cleaved at either of the potential processing sites, $(\alpha:\beta)$ or $(\beta:\gamma)$ (i.e. the product of this precursor is $\alpha + \beta + \gamma$, a 50 000- M_r polypeptide) while pDUB4/7 encodes a vicilin precursor that is cleaved at the $(\beta:\gamma)$ potential processing site but not the $(\alpha:\beta)$ site (i.e. the products of this precursor are $\alpha + \beta$, the 33 000- M_r polypeptide, and γ , a 16 000- M_r polypeptide), the sequences predicted by these cDNA species would be expected to differ at the $(\beta:\gamma)$ potential processing site but be the same at the $(\alpha:\beta)$ potential processing site. The locations of these sites can be established by the *N*- and *C*-terminal sequences of the different vicilin polypeptides.

The $(\beta:\gamma)$ processing site shown in Fig. 2 was determined from the *C*-terminus of the β (13 500- M_r)-subunit (Hirano *et al.*, 1982) and the *N*-terminus of the γ (16 000 M_r)-subunit. When aligned for optimal homology, on the cDNA from pDUB7, the two termini are contiguous. In addition, the *C*-terminus of the $\alpha + \beta$ (33 000- M_r)-subunit is the same as that of the β -subunit. The sequence predicted by pDUB2 diverges from that of pDUB7 in the four amino acids preceding the processing site (in the *N*- to *C*-terminus direction) and remains at least partly divergent for 12 amino acids following this site, including an extra four amino acids in the sequence. The sequences of the $\alpha + \beta$ -, β - and γ -subunits (i.e. those that are cleaved) parallel that predicted by pDUB7 closely, and are clearly divergent from that predicted by pDUB2. Significantly, the sequence of a peptide isolated from the $\alpha + \beta + \gamma$ (50 000- M_r)-subunit corresponding to the eight amino acids immediately preceding the potential processing site parallels that predicted by pDUB2, i.e., specifying that cleavage does not occur.

The different sequences in pDUB2 and pDUB7 following the potential processing site both consist of a stretch of highly polar amino acids (9–13 residues), predominantly acidic residues (over 65% glutamate), and thus appear to be functionally similar. However, the sequence divergence in the

four amino residues immediately preceding the potential processing site, Gly-Lys-Glu-Asn in the cleaved polypeptides and Gln-Gly-Leu-Arg in the uncleaved polypeptides, appears to be the basis for whether proteolytic cleavage takes place or not. In addition, the cleavage site lies in a region with a very high proportion of polar charged amino acid residues, which suggests a location on the outside of the protein molecule, where it would be accessible to proteinases, since the post-translational cleavage is known to occur after protein assembly is complete (Gatehouse *et al.*, 1981). This type of processing appears to be different from that suggested for legumin (Croy *et al.*, 1982) on the model of insulin and other animal secretory proteins (Steiner *et al.*, 1980), which postulates proteolytic cleavage at adjacent paired basic amino acid residues.

The amino acid sequence in the region of the potential $(\alpha:\beta)$ processing site, assuming the site to be at the *N*-terminus of the β (13 500- M_r)-subunit was previously published (Hirano *et al.*, 1982; Gatehouse *et al.*, 1982). The potential $(\alpha:\beta)$ cleavage site, like the $(\beta:\gamma)$ site, lies in a region containing a high proportion of polar amino acid residues. Although the two cDNA species pDUB2 and pDUB7 show sequence differences in this region, the predicted sequences are functionally very similar. There are no acidic residues immediately preceding the potential cleavage site; instead, the sequences predicted at this region (Gly-Leu-Arg from the nucleotide sequence GGCCCTTAGG for pDUB2 and Ser-Leu-Lys from the nucleotide sequence AGTCTTAAG for pDUB7) are very similar to the sequence predicted by pDUB2 at the $(\beta:\gamma)$ potential processing site (Gly-Leu-Arg), which specifies no cleavage. Protein sequences matching the cDNA sequences have been obtained from $\alpha + \beta$ (33 000- M_r)-subunits, which are uncleaved at this site.

The results presented here clearly imply the presence of more than one vicilin gene in the pea genome, with sequence divergences throughout their coding sequences. They also imply a sequence specificity of post-translational proteolytic cleavage determined by the genomic coding sequences. The pea genome contains a number of sequences homologous with pDUB2, as shown in Fig. 3, where the cDNA insert from pDUB2 is used as a probe on size-fractionated endonuclease-*Eco*RI-digested pea genomic DNA. Four distinct *Eco*RI-digested fragments hybridizing to pDUB2 are detectable, ranging in size from approx. 4 to 9 kilobase-pairs. When the intensity of hybridization is compared with that given with known amounts of linearized pDUB2 (Young *et al.*, 1981), these fragments are indicated to be present in one copy (estimated sizes 4.5, 7.0, 8.4 kilobase-pairs) or two copies (estimated size 7.5 kilobase-pairs) on the genome. Similar experiments with other restriction enzymes (results not shown)

Fig. 3. Hybridization of cDNA probes prepared by nick-translation of the inserts from pDUB2 (a) and pDUB4 (b) to pea genomic DNA restricted with endonuclease *EcoRI* (enzyme in >10-fold excess)

Tracks labelled 5 and 10 contain 5 and 10 μ g of DNA respectively. The intensity of hybridization is compared with that given by 2.6, 5.2, 10.4 and 25.6 pg of endonuclease-*EcoRI*-treated pDUB2 (a) and pDUB4 (b), corresponding approx. to 0.5, 1, 2 and 5 gene copies/haploid genome per 5 μ g of genomic DNA respectively.

also showed the presence of three to six genomic fragments present in one to two copies hybridizing with pDUB2. When pea genomic DNA restricted with *EcoRI* was probed with pDUB4, a single *EcoRI*-digested fragment, estimated size 7.8 kilobase-pairs, hybridized with this cDNA species with an intensity indicating it was present in two to three copies in the genome (Fig. 3). The results of these hybridization experiments thus provide evidence for the presence of several vicilin genes of differing sequences in the pea genome.

We thank Mr. John Gilroy for carrying out protein sequencing, and Mrs. Margaret Richardson, Mr. Dave Bown and Mrs. Philippa Brown for technical assistance.

References

- Amersham International (1981) *Nick Translational Kit Pamphlet P1/86/81/4*
- Chang, Y. Y., Brauer, D. & Wittmann-Liebold, B. (1978) *FEBS Lett.* **93**, 205–214
- Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N. & Boulter, D. (1982) *Nature (London)* **295**, 76–79
- Davey, R. A., Higgins, T. J. V. & Spencer, D. (1981) *Biochem. Int.* **3**, 595–602
- Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298
- Evans, I. M., Croy, R. R. D., Brown, P. & Boulter, D. (1980) *Biochim. Biophys. Acta* **610**, 81–95
- Gatehouse, J. A., Croy, R. R. D., Morton, H., Tyler, M. & Boulter, D. (1981) *Eur. J. Biochem.* **118**, 627–633

- Gatehouse, J. A., Lycett, G. W., Croy, R. R. D. & Boulter, D. (1982) *Biochem. J.* **207**, 629–632
- Graham, D. E. (1978) *Anal. Biochem.* **85**, 609–613
- Grunstein, M. & Wallis, J. (1979) *Methods Enzymol.* **68**, 379–389
- Hirano, H., Gatehouse, J. A. & Boulter, D. (1982) *FEBS Lett.* **145**, 99–102
- Lycett, G. W., Delauney, A. J., Gatehouse, J. A., Croy, R. R. D. & Boulter, D. (1983) *Nucleic Acids Res.* in the press
- Seif, I., Khoury, G. & Dhar, R. (1980) *Nucleic Acids Res.* **8**, 2225–2240
- Southern, E. M. (1979) *Methods Enzymol.* **68**, 152–176
- Steiner, D. F., Quinn, P. S., Chen, S. J., Marsh, J. & Trager, H. S. (1980) *Ann. N.Y. Acad. Sci.* **343**, 1–16
- Young, R. A., Hagenbuchle, O. & Schibler, U. (1981) *Cell* **23**, 451–459