Isolation and expression of a pea vicilin cDNA in the yeast *Saccharomyces cerevisiae*

Martin D. WATSON,*† Nigel LAMBERT,† Ashton DELAUNEY,* Jennifer N. YARWOOD,* Ron R. D. CROY,* John A. GATEHOUSE,* David J. WRIGHT† and Donald BOULTER*

*Department of Botany, Science Laboratories, University of Durham, South Road, Durham DH1 3LE, U.K., and †Molecular Science Division, A.F.R.C. Institute of Food Research, Norwich Laboratory, Colney Lane, Norwich NR4 7UA, U.K.

A cDNA clone containing the complete coding sequence for vicilin from pea (*Pisum sativum* L.) was isolated. It specifies a 50000- M_r protein that in pea is neither post-translationally processed nor glycosylated. The cDNA clone was expressed in yeast from a 2 μ m plasmid by using the yeast phosphoglycerate kinase promoter and initiator codon. The resultant fusion protein, which contains the first 16 amino acid residues of phosphoglycerate kinase in addition to the vicilin sequence, was purified and subsequently characterized. It has slightly slower mobility on SDS/polyacrylamide-gel electrophoresis than standard pea vicilin and forms a mixture of multimers, some of which resemble the native protein.

INTRODUCTION

Vicilin is one of the major storage proteins found in pea seeds and its structure and properties are now well documented (Gatehouse *et al.*, 1984; Wright, 1988). Each vicilin molecule is composed of three subunits of M_r approx. 50000 originating from a pool of highly homologous, though variable, polypeptides; at least 11 genes code for the different pea vicilin subunits (Gatehouse *et al.*, 1983; Domoney & Casey, 1985). The vicilin subunits undergo post-translational glycosylation, proteolysis and possibly deamidation to various degrees (Davey & Dudman, 1979; Gatehouse *et al.*, 1982, 1984), adding to their intrinsic heterogeneity. Analysis of vicilin under denaturing conditions by SDS/PAGE separates at least ten polypeptides ranging in M_r from 12500 to 50000. The lower- M_r components arise from the post-translational proteolysis 'nicking' of some of the 50000- M_r subunits (Gatehouse *et al.*, 1982, 1983). A number of cDNAs encoding vicilin polypeptides have been isolated and characterized (Croy *et al.*, 1982; Lycett *et al.*, 1983; Spencer *et al.*, 1983), and both nucleotide and amino acid sequence data have demonstrated extensive homologies within the pea vicilin polypeptide family, and with vicilin-like storage proteins from other species (Borotto & Dure, 1987; Wright, 1988).

A major objective of the food industry is the exploitation of the functional properties (e.g. gelation, emulsification, foaming) of legume storage proteins, e.g. soya-bean proteins (Kinsella, 1979; Wright & Bumstead, 1984). However, the intrinsic heterogeneity of storage proteins in general has severely impaired studies attempting to relate structure to functionality. The advent of genetic engineering, and the ability to express specific foreign genes in micro-organisms, possibly afford a route for preparing sufficient homogeneous material to carry out the required functionality studies. To date, phaseolin,

the vicilin-like storage protein from French bean, is the only reported example of expression of such storage proteins in micro-organisms (Cramer *et al.*, 1985), although the level of expression was low. In addition, by using this approach it is possible to modify specifically the protein's primary structure by mutagenesis *in vitro*; on the basis of vicilin structural data, such alterations can be made in a predictable manner, and thus allow the effect of specific changes in the primary sequence on the physicochemical and functional properties to be assessed.

In the present paper we describe the isolation and expression in yeast of a cDNA for a non-glycosylated non-proteolytically processed vicilin subunit by using a yeast phosphoglycerate kinase (PGK) expression vector.

MATERIALS AND METHODS

Chemicals

Tricine, lyticase and leupeptin were from Sigma Chemical Co. HA-Ultrogel (hydroxyapatite) was from LKB Produkter. Superose 6, PD-10 and 10/10 f.p.l.c. columns were from Pharmacia. Growth media were from Difco Laboratories. Peroxidase-linked goat anti-(rabbit IgG) second antibody was from Bio-Rad Laboratories. Affinity-purified rabbit anti-(pea vicilin) antibodies were purified as described previously (Croy *et al.*, 1980). Other chemical were from BDH Chemicals.

Bacterial and yeast strains

The yeast strain used was *Saccharomyces cerevisiae* MC16 (ade2–1, leu2–3, lys2–1, his4–712^{FS}, SUF2) (Beggs, 1978).

Escherichia coli GL80 (*recA*[–]) was provided by G. Lycett, and *E. coli* 910 (*recB*[–] *recC*[–]) was provided by W. J. Brammar.

Abbreviations used: PGK, phosphoglycerate kinase; PAGE, polyacrylamide-gel electrophoresis; f.p.l.c., fast protein liquid chromatography; e.l.i.s.a., enzyme-linked immunosorbent assay.

† To whom all correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00722.

Plasmids

The expression vector pMA257 was kindly supplied by Dr. A. J. Kingsman and Dr. S. M. Kingsman (Tuite *et al.*, 1982). The two pea vicilin cDNAs, pDUB2 and pDUB4, have been described previously (Lycett *et al.*, 1983).

Growth conditions and general methods

All bacterial growth conditions, media, transformations and recombinant DNA techniques were as described by Maniatis *et al.* (1982). Yeast growth conditions and media were as described by Beggs (1978).

Construction of a cDNA bank

Preparation and size fractionation of cDNA and subsequent addition of *Bam*HI linkers were as described previously (Lycett *et al.*, 1983). The cDNA was ligated into pBR322 and subsequently transformed into competent *E. coli* strain 910. Vicilin-containing clones were identified by colony hybridization (Grunstein & Wallis, 1979; Maniatis *et al.*, 1982).

DNA sequencing

The pDUB9 cDNA insert was sequenced by a combination of the dideoxy 'forward-and-backwards' technique on purified restriction fragments (Maat & Smith, 1978; Seif *et al.*, 1980) and sequencing of an M13 clone of the entire 1.4 kb vicilin insert (Sanger *et al.*, 1977; Biggin *et al.*, 1983; Messing, 1983). Primers were designed from previously compiled sequence data and synthesized with the use of an Applied Biosystems 381A synthesizer. Sequence data were compiled and analysed by the use of computer programs of our own design.

Small-scale protein extracts

Transformed yeast cultures (10 ml) were grown overnight at 30 °C with aeration in yeast minimal medium supplemented with the required amino acids. The cells were pelleted by centrifugation, cooled and immediately resuspended in 100 µl of lysis buffer [1 ml of 0.2 M Tris/HCl buffer, pH 6.0, containing 0.2% SDS and 10% (w/v) sucrose, 10 µl of leupeptin (200 µg/ml), and 10 µl of phenylmethanesulphonyl fluoride (36 mg/ml in ethanol)]. An equal volume of acid-washed 450–500 µm glass beads was added, and then the cells were vortex-mixed for five bursts of 15 s each. The debris was removed by centrifugation, and the supernatant was boiled for 10 min before being loaded on to an SDS/PAGE gel.

Purification of vicilin from yeast

Step 1: cell growth and preparation of yeast soluble extract. Yeast cells (after inoculation from a 10 ml stationary-phase culture in yeast minimal medium) were grown at 25 °C in 1-litre shake-cultures in yeast minimal medium supplemented with the required amino acids. The cells were harvested by centrifugation (5000 *g* for 10 min) at late exponential phase (approx. 24 h), corresponding to an A_{600} of approx. 3.0. Cells were resuspended in 30 ml of 25 mM-potassium phosphate buffer, pH 7.5, containing 1 M-sorbitol and 0.2 mg of lyticase/ml before incubation at 25 °C without shaking. Spheroplast formation was monitored turbidimetrically by the decrease in A_{600} after treatment with SDS (Scott & Schekman,

1980). When production of spheroplasts had proceeded to at least 80% completion (approx. 1 h), the suspension was centrifuged gently at 2000 *g* for 5 min. The pellet (spheroplasts and intact cells) was homogenized, in a 50 ml Teflon/glass homogenizer on ice, in 40 ml of 50 mM-Tricine buffer, pH 8.2, containing 1 mM-phenylmethanesulphonyl fluoride and leupeptin (0.1 mg/ml). All subsequent steps were performed at 1 °C unless stated otherwise. The extract was centrifuged at 12000 *g* for 30 min, and the pellet, containing intact yeast and cell debris, was discarded. The supernatant comprised the yeast soluble extract.

Step 2: (NH₄)₂SO₄ fractionation. Solid (NH₄)₂SO₄ was added slowly with stirring to the above extract to a final concentration of 45% (relative saturation). After 15 min the extract was centrifuged at 12000 *g* for 30 min, and the pellet was discarded. The supernatant was made 95% saturated with respect to (NH₄)₂SO₄ and centrifuged as described above. The resulting supernatant was discarded, and the pellet was stored at –20 °C for further processing.

Step 3: HA-Ultrogel chromatography. The following experiments were performed at room temperature. Portions of the 45–95%-satn.-(NH₄)₂SO₄ pellet were dissolved in 2 ml of 150 mM-sodium phosphate buffer, pH 7.5, containing 1 mM-phenylmethanesulphonyl fluoride, and quickly desalted on a PD-10 gel-filtration column equilibrated in the same buffer. The resulting sample was loaded, with the aid of a 10 ml Superloop, on to an HA-Ultrogel column packed into a 10/10 glass f.p.l.c. column, interfaced to an f.p.l.c. system with a UV1 u.v. monitor and a 280 nm filter as a detection system (D. J. Wright, B. K. Newby & N. Lambert, unpublished work). The column was pre-equilibrated in the above phosphate buffer, and unbound protein was eluted with this buffer. Bound material (vicilin) was eluted with a linear gradient of increasing sodium phosphate concentration (see Fig. 7). The flow rate throughout was 1 ml/min, and all buffers used in conjunction with the f.p.l.c. were filtered through 0.22 µm-pore-size filters before use. The purified vicilin was concentrated by ultrafiltration and stored at –20 °C.

Purification of vicilin from peas

The vicilin fraction from peas (*Pisum sativum* L. cultivar Feltham First) was isolated and purified according to the method of Lambert *et al.* (1986).

Assessment of subunit assembly by gel filtration-f.p.l.c.

Native pea vicilin and vicilin expressed in yeast (0.5 mg of protein in a volume of 50 µl or 100 µl) were analysed on a Superose 6 gel-filtration column attached to an f.p.l.c. system (Lambert *et al.*, 1986). The flow rate was 0.4 ml/min and the solvent system was 50 mM-Tris/HCl buffer, pH 8.2, containing 200 mM-NaCl. The detection system and pretreatment of buffers were as described above.

SDS/PAGE and Western blotting

Electrophoresis and preparation of samples under reducing conditions were as described in Bacon *et al.* (1987).

Western blotting was performed on Coomassie Blue-stained and unstained gels pre-equilibrated in 25 mM-

Tris/192 mM-glycine buffer, pH 8.2, containing 20% (v/v) methanol as described by Towbin *et al.* (1979). Nitrocellulose papers were developed by using the method described in the Amersham manual, with affinity-purified rabbit anti-(pea vicilin) antibodies and a horseradish-peroxidase-conjugated second antibody with chloronaphthol/H₂O₂ as substrate.

Immunoassays

E.l.i.s.a. assays were performed in 96-well microtitre plates (NUNC, Immunoplate 1F) according to the method of Voller *et al.* (1979). Wells were coated freshly with 300 μ l of pea vicilin (1 μ g/ml). Alkaline-phosphatase-conjugated second antibody and nitrophenyl phosphate were used as a detection system. The assay was linear over the range 0.05–3 μ g of pea vicilin/ml.

Determination of protein

Protein concentrations were measured by using the method of Lowry *et al.* (1951) with bovine serum albumin as a standard.

RESULTS

Isolation of a full-length vicilin cDNA

A cDNA bank has been prepared from polyadenylated mRNA purified from developing pea cotyledons. In an attempt to isolate full-length clones, the total cDNA was size-fractionated on agarose gels before the addition of *Bam*HI linkers and cloning into pBR322 (Lycett *et al.*, 1983). Vicilin-cDNA-containing recombinants were identified by colony hybridization with the use of the vicilin cDNA inserts from pDUB2 and pDUB4 (Lycett *et al.*, 1983). Neither of these previously identified cDNAs is full-length. One particular clone that strongly hybridized to pDUB2 and pDUB4, namely pDUB9, was analysed in detail and found to have a 1430 bp insert. This clone was completely sequenced and found to encode the entire amino acid sequence of a mature 50000-*M_r* vicilin. The restriction map of the pDUB9 insert is shown in Fig. 1.

The DNA sequence was determined by using the technique of dideoxy-induced termination of nick translation developed by Maat & Smith (1978). DNA restriction fragments were purified and sequenced as shown in Fig. 2. The complete nucleotide sequence and its deduced amino acid sequence are shown in Fig. 3. Inspection of the DNA sequence reveals that the 5' end codes for three amino acid residues, valine, serine and serine, before the *N*-terminal end found in the mature protein. These are from the 15-amino acid-residue presumptive signal peptide (Lycett *et al.*, 1983).

Vicilin polypeptides of *M_r* less than 47000 are produced by post-translational cleavage of 47000–50000-*M_r* precursor molecules. There are two potential cleavage points, which occur at residues 184 and 309 in processed vicilin molecules, and these are shown as b and c in Fig. 3. Processing may occur at none, either or both of these sites (Gatehouse *et al.*, 1984). pDUB9 codes for a vicilin that cannot be processed at either of these sites since it lacks appropriately positioned asparagine residues at which cleavage would occur. Some vicilin polypeptides are glycosylated, but the substitution of lysine for asparagine at amino acid 348 (Fig. 3) in pDUB9 indicates

Fig. 1. Restriction map of the vicilin cDNA insert in pDUB9

Distances are in base-pairs. α , β and γ represent the three domains of vicilin produced upon post-translational proteolysis. Key: B, *Bam*HI; P, *Pst*I; H2, *Hind*III; G, *Bgl*II; N, *Bst*NI; X, *Xba*I; A, *Acc*I; H3, *Hind*III.

Fig. 2. Sequencing strategy of the vicilin cDNA insert in pDUB9

Sequence determination from positions indicated by black circles was by the 'forwards-and-backwards' technique. Determination from positions indicated by vertical bars was by sequencing of an M13 clone of the entire 1.4 kb vicilin cDNA fragment. The arrows represent the direction and extent of sequence information obtained from a particular experiment. Key: A, *Acc*I; B, *Bam*HI; C, *Hinc*II; F, *Hinf*I; G, *Bgl*II; H, *Hind*III; P, *Pst*I; X, *Xba*I.

that the vicilin coded for is unglycosylated. The 3' end of the cDNA has 129 bp of untranslated sequence after the chain termination codon (TGA), but lacks the poly(A) tail found on the mRNA species from which it is derived.

Expression of vicilin in yeast (*S. cerevisiae*)

The 1.4 kb vicilin cDNA fragment was cloned into the yeast expression vector pMA257, as outlined in Fig. 4. pMA257 is a derivative of pMA91 (Tuite *et al.*, 1982), which carries the yeast *Leu2* gene as a selectable marker, the yeast 2 μ m origin of replication, *Amp^R* and *ori* from pBR322, and the promoter and 5' coding sequences from the yeast *PGK* (phosphoglycerate kinase) gene. The 1.4 kb vicilin cDNA was ligated into the single *Bam*HI site in pMA257, which is 16 amino acid residues after the *PGK* initiation codon. The orientation of the vicilin fragment within pMA257 was confirmed by digestion with *Hind*III (Fig. 4). The *PGK*-vicilin *Hind*III fragment is larger (2.5 kb) in the correct expression orientation (pDUB2018) than in the reverse orientation (1.9 kb in pDUB2017). Insertion of the vicilin fragment in pMA257 in the correct orientation creates an in-frame fusion, shown in Fig. 5, with 16 amino acid residues from *PGK*, three from the *Bam*HI linker, followed by three from the putative vicilin signal peptide and then the mature vicilin sequence.

pDUB2017 and pDUB2018 were transformed into yeast strain MC16. Transformants were grown on yeast minimal medium, and small-scale total protein extracts

```

10                               60
TTTCTTCTAGGTCTGATCCTCAAAATCCTTTTATCTTCAAGTCTAACAAGTTTCAAACCTCTTTTGAG
* S S R S D P Q N P F I F K S N K F Q T L F E
      a
70                               130
AATGAAATGGGCACATTGCACTTCTGCAGAAATTTGACCAACGTTCTAAAATTTTCGAGAATCTACAA
N E N G H I R L L Q K F D Q R S K I F E N L Q
      PstI
140                               200
AACTACCGTCTTTTGGAATATAAGTCCAAACCTCACACAATATTTCTTCCACAGCACACCGATGCCGAT
N Y R L L E Y K S K P H T I F L P Q H T D A D
      66
210                               270
TACATCCTTGTTGTACTCAGTGGAAGCTATACTCACAGTGTGAAACCGATGATAGAAACTCCTTC
Y I L V V L S G K A I L T V L K P D D R N S F
      89
280                               340
AACCTTGAGCGCGGAGATACGATAAACTTCTGCTGGCACAATTGCTTATTTGGTTAAACAGAGATGAC
N L E R G D T I K L P A G T I A Y L V N R D D
      HindIII
350                               410
AACGAGGAGCTTAGAGTATTAGATCTCGCCATTCCCGTAAATAGACCTGGCCAACCTTCAGTCTTCTTA
N E E L R V L D L A I P V N R P G Q L Q S F L
      BglII BstNI
420                               480
TTGCTGGAAATCAAAACCAACAACTACTTATCTGGGTTCAAGTAAGAACATTTAGAGGCTTCCTTC
L S G N Q N Q Q N Y L S G F S K N I L E A S F
      XbaI
490                               550
AATACTGATTATGAAGAGATAGAAAAGGTTCTTTTGAAGAGCATGAGAAAGAGACACAACACAGAAGA
N T D Y E E I E K V L L E E H E K E T Q H R R
      181
560                               620
AGCCTTAAGGATAAGAGGCAGCAAGTCAAGAAGAGAATGTAATAGTAAAATTATCAAGGGGACAAATT
S L K D K R Q Q S Q E E N V I V K L S R G Q I
      b
630                               690
GAGGAATTGAGTAAAAATGCAAAGTCTACTCCAAAAAAGTGTTCCTCTGAATCTGAACCATTCAC
E E L S K N A K S T S K K S V S S E S E P F N
      AccI
700                               750
TTGAGAAGTCGCGGTCTATCTATTCCAACGAGTTTGGAATAATCTTTGAAATCACCCAGAGAAAAAT
L R S R G P I Y S N E F G K F F E I T P E K N
      250
770                               820
CCACAGCTTCAAGACTTGGATATATTGTCAATTCTGTAGAGATTAAGGAGGGATCTTTATTGTTGCCA
P Q L Q D L D I F V N S V E I K E G S L L L P
      273
830                               890
CACTACAATTCAAGGGCCATAGTAATAGTAACATTAACGAAGGAAAAGGAGATTTTGAACCTGTGGGT
H Y N S R A I V I V T V N E G K G D F E L V G
      HindIII
900                               960
CAAAGAAATGAAAACCAACAAGAGCAGAGAAAAGAAGATGACGAGGAAGAGGAACAAGGAGAGAAGAG
Q R N E N Q Q E Q R K E D D E E E E Q G E E E
      c
970                               1030
ATAAATAAACAAGTGCAAAATTACAAAGCTAAATTATCTTCAGGAGATGTTTTGTGATTCCAGCAGGC
I N K Q V Q N Y K A K L S S G D V F V I P A G
      319
1040                               1100
CATCCAGTTGCCCTAAAGCTTCTCAAACTCTTGATTGCTTGGGTTGGTATTAATGCTGAGAACAAAT
H P V A L K A S S N L D L L G F G I N A E N N
      d
1110                               1170
CAGAGGAACCTTCTTGCAGGCGATGAGGATAATGTGATTAGTCAGGTACAGCGACCAAGTGAAGAGCTT
Q R N F L A G D E D N V I S Q V Q R P V K E L
      365
1180 BstNI                               1240
GCATTCCCTGGATCAGCTCAAGAGGTTGATAGGATACTAGAGAATCAGAAACAATCCCACTTTGCAGAT
A F P G S A Q E V D R I L E N Q K Q S H F A D
      388
1250                               1310
GCTCAACCTCAACAAAGGAGAGAGGAAGTCGTGAAACAAGAGATCGTCTATCTTCAGTTTGAATGTT
A Q P Q Q R E R G S R E T R D R L S S V
      411
1320                               1370
TCTTAATGAGTGGACAAAATACTATGTATGTATGCTATCAAGAGATATATCTCACGGGGAGCAATGAAT
      431
1390                               1433
AAAACAATGTTATCTTATAACTATAATTATATATCCACTTTTCTACTATGAATA

```

Fig. 3. DNA sequence of the vicilin cDNA insert in pDUB9

a indicates the position of the *N*-terminal end of the mature protein. b and c indicate the sites of specific proteolytic cleavage found in other vicilin polypeptides.

Fig. 4. Cloning of the vicilin cDNA insert into pMA257

Fig. 5. DNA sequence and predicted amino acid sequence of the fusion between PGK and vicilin

were made. These extracts were analysed on SDS/PAGE gels followed by Western blotting with polyclonal anti-(pea vicilin) antibodies. Only pDUB2018 showed any immunoreactive protein with the vicilin antibody (results not shown). The presumptive vicilin in pDUB2018 transformed yeast was then isolated and subjected to preliminary characterization.

Purification and characterization of vicilin from yeast

Yeast cells from a 1-litre culture were broken by enzymic cell-wall digestion followed by hypo-osmotic disruption of the resulting spheroplasts. This represents a milder more-controllable method of cell disruption compared with others (Schatz, 1979), and presents no problems (procedural or safety) for future scale-up. The inclusion of proteinase inhibitors and the need to isolate vicilin quickly were essential to avoid severe degradation of the final product. Fig. 6 shows SDS/PAGE and Western blots of the various fractions obtained during purification.

Western-blot analysis revealed that the expressed vicilin resided solely in the soluble fraction of the yeast extract, with no immunoreactive polypeptides located in the particulate (insoluble) fraction or in the culture medium (results not shown). Typically the soluble extract from a 1-litre culture contained about 200 mg of protein. This was subjected to a 45–95%-relative-saturation-(NH₄)₂SO₄ fractionation. Western blots showed that the discarded 45%-saturation-(NH₄)₂SO₄ pellet contained no immunoreactive material (Fig. 6b, track 2).

Portions of the 95%-saturation-(NH₄)₂SO₄ pellet were rapidly desalted on a PD-10 column (this method being preferred to lengthy dialysis) and then fractionated on a hydroxyapatite column. Under the conditions employed, vicilin was the only component in the extract to bind to the column, and was eluted with a phosphate gradient (see Fig. 7). The purified vicilin was shown to be homogeneous by SDS/PAGE (Fig. 6a, tracks 7–11), migrating as a single species with an *M_r* of approx. 53 000. No specific proteolytic processing was observed,

Fig. 6. (a) SDS/PAGE analysis and (b) Western-blot analysis of various fractions from the purification of vicilin from yeast

(a) SDS/PAGE analysis. Samples were analysed in the presence of 2-mercaptoethanol as indicated in the Materials and methods section. Track 1, M_r standard proteins (10 μ g of each); track 2, yeast soluble extract (100 μ g); track 3, 0–45% satn.-(NH₄)₂SO₄ pellet (40 μ g); track 4, 45–95% satn.-(NH₄)₂SO₄ pellet (70 μ g); track 5, vicilin from pea (50 μ g); track 6, unbound fraction from the HA-Ultrogel column (20 μ g); tracks 7–11, fractions taken across the bound peak from the HA-Ultrogel column. (b) Western-blot analysis. The SDS/PAGE gel was processed as described in the Materials and methods section. Tracks 1 and 3, yeast soluble extract (100 μ g); track 2, 0–45% satn.-(NH₄)₂SO₄ pellet (40 μ g); track 4, 45–95% satn.-(NH₄)₂SO₄ pellet (70 μ g); tracks 5 and 9, vicilin from pea (50 μ g); track 6, unbound fraction from the HA-Ultrogel column (30 μ g); tracks 7, 8, 10 and 11, fractions taken across the bound peak from the HA-Ultrogel column.

in contrast with the vicilin isolated from pea (Fig. 6a, track 5). Both SDS/PAGE and Western-blot analysis reveal a minor band of slightly higher M_r ; this could be a translation artifact from the hybrid gene, but we have no completely satisfactory explanation as to what it may be. Typically approx. 10 mg of vicilin (as determined by direct protein assay) was obtained from a 1-litre culture, corresponding to approx. 5% of the yeast soluble protein. Assessment of the level of expression by e.l.i.s.a.,

Fig. 7. Purification of expressed vicilin by HA-Ultrogel column chromatography

See the Materials and methods section for full details. —, A_{280} ; ----, concentration of sodium phosphate.

Fig. 8. Elution profile of vicilin on a Superose 6 gel-filtration-f.p.l.c. column

See the Materials and methods section for full details. (a) Vicilin fraction from pea. (b) Vicilin expressed in yeast.

however, gave values of 0.1–0.5% of yeast soluble extract, with anti-(pea vicilin) antibodies and pea vicilin as a standard.

Analysis of the expressed vicilin by gel filtration (Fig. 8) clearly demonstrated that, although some protein had apparently assembled correctly into trimers and was

co-eluted with authentic pea vicilin, the majority of the polypeptides had not assembled properly, with monomers, dimers and hexamers being clearly visible. Solution of the problem of correct reassembly will clearly require investigation.

DISCUSSION

Although the cDNA species cloned in pDUB9 contains the complete coding sequence of a mature vicilin polypeptide, the cDNA is not full length, since it lacks most of the signal sequence at its 5' end, and at least the poly(A) tail at its 3' end. Two previously isolated and sequenced cDNAs carried by pDUB4 and pDUB7 are presumed by sequence analysis to be parts of the same mRNA (Lycett *et al.*, 1983). Together these clones represent almost all of a 47000- M_r vicilin precursor, though lacking approx. 70–80 bases of 3' coding region. Another vicilin cDNA clone, pDUB2, has also been sequenced. This clone codes for a 50000- M_r subunit, but lacks the first third of the 5' coding sequences. pDUB2 also codes for part of the 3' untranslated region. Sequence analysis of the cDNA insert in pDUB9 shows that it is highly homologous to that in pDUB2 and thus encodes a 50000- M_r vicilin subunit. It shows less homology (approx. 85%) to the 47000- M_r cDNA clones. pDUB9 codes for the entire amino acid sequence of the mature 50000- M_r vicilin, plus three amino acid residues of the signal peptide predicted to exist from the sequence of pDUB7. pDUB9 also contains 38 bases more of the 3' untranslated region than pDUB2. Analysis of the sequences of pDUB2 and pDUB9 shows clearly that they are transcribed from different vicilin genes. The differences are few and mostly conservative. It may be noted that the C-terminal end predicted from pDUB9 is different from that of pDUB2 (Lycett *et al.*, 1983). Both pDUB9 and pDUB2 lack the Asn-Ala-Ser putative glycosylation site found in the 47000- M_r derivatives; pDUB9 has Lys-Ala-Ser and pDUB2 Arg-Ala-Ser at this position.

The vicilin cDNA in pDUB9 was successfully expressed in yeast by using the PGK expression vector pMA257. The encoded protein was localized in the cytosol and migrated as a single species in SDS/PAGE with an M_r of approx. 53000, significantly higher than the 50000- M_r subunit of pea vicilin. This higher M_r value is most probably due to the additional 22 amino acid residues (approx. 3000 M_r) at the N-terminal end encoded by the fusion constructs. In the pea seed some vicilin subunits are specifically proteolytically processed or 'nicked', generating several bands when analysed on SDS/PAGE. When the expressed vicilin was isolated from yeast in the absence of proteinase inhibitors several other bands of M_r values less than 53000 were observed on Western blots (results not shown); however, the band pattern was not reproducible, varying quite markedly from preparation to preparation, indicating that this represents non-specific proteolysis by the abundant yeast proteinases. The expressed vicilin polypeptide would not be predicted to be cleaved at the specific proteolytic processing sites found in other vicilin polypeptides.

Gel-filtration analysis demonstrated that the expressed vicilin had not for the most part assembled correctly into trimers. This could be because the yeast cytosol does not afford a suitable environment for assembly, or that the modified N-terminus precludes correct assembly. Such

difficulties in expressing foreign proteins in yeast have been reported by others (Smith *et al.*, 1985). It may, however, be possible to reassociate and correctly reassemble the expressed vicilin after complete dissociation and denaturation with urea, since such experiments have proven successful with legumin-like storage protein subunits (Utsumi & Mori, 1983; Nakamura *et al.*, 1985).

The level of vicilin expression was approx. 5% of soluble protein as determined by direct protein assay. This is significantly higher than that obtained for the expression of phaseolin in yeast (0.01–0.03% of total cellular protein) by Cramer *et al.* (1985). However, phaseolin was transcribed and translated under the control of its own natural promoter, whereas we have employed the powerful yeast PGK promoter; PGK forms as much as 1% of total cell protein on glucose-grown cells (Holland & Holland, 1978). The parent plasmid of pMA257 which carries the entire PGK gene produces protein levels of 50–80% (Mellor *et al.*, 1985). The yield of vicilin produced by pDUB2018 is an order of magnitude less, but no different from other heterologous proteins produced from the PGK promoter (Tuite *et al.*, 1982). Several groups have investigated this phenomenon without providing a satisfactory explanation (Chen *et al.*, 1984; Mellor *et al.*, 1985). It should be noted that many yeast genes require a specific transcription termination signal (Zaret & Sherman, 1982). The production of the PGK-gene-based fusion vectors removes all of the 3' coding and non-coding regions plus the remaining yeast sequences downstream of the PGK gene (Dobson *et al.*, 1982; Tuite *et al.*, 1982). The 3' untranslated region of the vicilin cDNA insert within pDUB9 is extremely A+T-rich and contains an AATAAA putative polyadenylation signal (Lycett *et al.*, 1983). Whether this vicilin sequence can act as a transcription terminator in yeast is not known, but it has elements of homology with the consensus yeast transcription terminator TAG...TATGT...TTT proposed by Zaret & Sherman (1982). Termination of transcription was postulated to occur in the phaseolin 3' untranslated regions expressed in yeast (Cramer *et al.*, 1985). If transcription does not terminate within the plant cDNA sequence, then it must terminate at some point within the pBR322 sequences of pMA257.

By exploring other growth media it should be possible to increase further the amounts of vicilin produced. For example, preliminary studies with complete media gave a higher cell density at the corresponding growth phase, resulting in a soluble extract containing approx. 350 mg of protein from a 1-litre culture (i.e. a 75% increase in protein compared with cells grown on minimal media). However, even at the present level of expression, once the problem of reassembly has been surmounted, sufficient material will be available to carry out a rigorous physicochemical characterization, including assessment of functional properties.

This work was supported in part by an A.F.R.C. LINK grant to the Department of Botany, University of Durham, and the Institute of Food Research, Norwich. A. J. D. was supported by a research studentship from Genzyme Biochemicals Ltd. We thank Dr. A. J. Kingsman and Dr. S. M. Kingsman for plasmid pMA257 and for advice on heterologous gene expression in yeast.

REFERENCES

- Bacon, J. R., Lambert, N., Phalp, M., Plumb, G. W. & Wright, D. J. (1987) *Anal. Biochem.* **160**, 202–210
- Beggs, J. D. (1978) *Nature (London)* **275**, 104–109
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Borrito, K. & Dure, L. (1987) *Plant Mol. Biol.* **8**, 113–131
- Chen, C. Y., Oppermann, H. & Hitzeman, R. A. (1984) *Nucleic Acids Res.* **12**, 8951–8970
- Cramer, J. H., Lea, K. & Slightom, J. L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 334–338
- Croy, R. R. D., Gatehouse, J. A., Evans, I. M. & Boulter, D. (1980) *Planta* **148**, 57–63
- Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N. & Boulter, D. (1982) *Nature (London)* **295**, 76–79
- Davey, R. A. & Dudman, W. F. (1979) *Aust. J. Plant Physiol.* **6**, 435–447
- Dobson, M. J., Tuite, M. F., Roberts, N. A., Kingsman, A. J. & Kingsman, S. M. (1982) *Nucleic Acids Res.* **10**, 2625–2637
- Domoney, C. & Casey, R. (1985) *Nucleic Acids Res.* **13**, 687–699
- Gatehouse, J. A., Lycett, G. W., Croy, R. R. D. & Boulter, D. (1982) *Biochem. J.* **207**, 629–632
- Gatehouse, J. A., Lycett, G. W., Delauney, A. J., Croy, R. R. D. & Boulter, D. (1983) *Biochem. J.* **212**, 427–432
- Gatehouse, J. A., Croy, R. R. D. & Boulter, D. (1984) *CRC Crit. Rev. Plant Sci.* **1**, 287–314
- Grunstein, M. & Wallis, J. (1979) *Methods Enzymol.* **68**, 379–389
- Holland, M. J. & Holland, J. P. (1978) *Biochemistry* **17**, 4900–4907
- Kinsella, J. E. (1979) *J. Am. Oil Chem. Soc.* **56**, 242–257
- Lambert, N., Chambers, S. J., Phalp, M. & Wright, D. J. (1986) *Biochem. Soc. Trans.* **14**, 1186–1188
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- Lycett, G. W., Delauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R. R. D. & Boulter, D. (1983) *Nucleic Acids Res.* **11**, 2367–2380
- Maat, J. & Smith, A. J. H. (1978) *Nucleic Acids Res.* **5**, 4537–4546
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Mellor, J., Dobson, M. J., Roberts, N. A., Kingsman, A. J. & Kingsman, S. M. (1985) *Gene* **33**, 215–226
- Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
- Nakamura, T., Utsumi, S. & Mori, T. (1985) *Agric. Biol. Chem.* **49**, 2733–2740
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Schatz, G. (1979) *Methods Enzymol.* **56**, 40–50
- Scott, J. H. & Schekman, R. (1980) *J. Bacteriol.* **142**, 414–423
- Seif, I., Khoury, G. & Dhar, R. (1980) *Nucleic Acids Res.* **8**, 2225–2239
- Smith, R. A., Duncan, M. J. & Moir, D. T. (1985) *Science* **229**, 1219–1224
- Spencer, D., Chandler, P. M., Higgins, T. J. V., Inglis, A. S. & Rubira, M. (1983) *Plant Mol. Biol.* **2**, 259–267
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Tuite, M. F., Dobson, M. J., Roberts, N. A., King, R. M., Burke, D. C., Kingsman, S. M. & Kingsman, A. J. (1982) *EMBO J.* **1**, 603–608
- Utsumi, S. & Mori, T. (1983) *J. Biochem. (Tokyo)* **94**, 2001–2008
- Voller, A., Bidwell, D. E. & Bartlett, A. (1979) *The Enzyme Linked Immunosorbent Assay (ELISA): A Guide with Abstracts of Microplate Applications*, pp. 1–125, Flowline Press, St. Peter Port
- Wright, D. J. (1988) in *Developments in Food Proteins* (Hudson, B. J. F., ed.), vol. 5, Elsevier Applied Science Publishers, Barking, in the press
- Wright, D. J. & Bumstead, M. R. (1984) *Philos. Trans. R. Soc. London B* **304**, 381–393
- Zaret, K. S. & Sherman, F. (1982) *Cell* **28**, 563–573

Received 11 May 1987/9 November 1987; accepted 6 January 1988