Plant Molecular Biology 2: 259–267 (1983).

Sequence interrelationships of the subunits of vicilin from pea seeds

Donald Spencer[1], Peter M. Chandler[1], Thomas J. V. Higgins[1], Adam S. Inglis[2] & Michael Rubira[2]
[1] *CSIRO, Division of Plant Industry, P.O. Box 1600, Canberra 2601, Australia*
[2] *CSIRO, Division of Protein Chemistry, 343 Royal Parade, Parkville, 3052, Victoria, Australia*

Keywords: vicilin subunits, *Pisum sativum*, post-translational processing, amino acid sequences

Abstract

Serological studies and comparison of N-terminal amino acid sequences with the amino acid sequence deduced from a cDNA clone have been used to establish the sequence relationships between the subunits of the pea seed storage protein, vicilin. Subunits smaller than $M_r \sim 50\,000$ (i.e., M_r 34 000, 30 000, 25 000, 18 000, 14 000, 13 000 and 12 000) show extensive homology with molecules within the $M_r \sim 50\,000$ group. Both the sequencing and serological data confirm earlier evidence from studies on vicilin synthesis *in vivo* and *in vitro* which indicated that the vicilin subunits smaller than $M_r \sim 50\,000$ arose by endoproteolytic cleavage of parent molecules within the $M_r \sim 50\,000$ group. Cleavage in different M_r 50 000 parent molecules containing either one or both of two susceptible processing sites accounts for the formation of all the vicilin subunits smaller than $M_r \sim 50\,000$, with the possible exception of the M_r 34 000 polypeptide. The positions of these sites in the putative parents were defined by reference to a complete amino acid sequence deduced from the sequence of DNA complementary to mRNA for one member of the $M_r \sim 50\,000$ group.

Introduction

Most legume seeds studied to date contain a major storage protein of M_r 150 000 to 180 000 (7 to 9S) (8). In some legumes, such as *Phaseolus vulgaris* (20) and *Glycine max* (21) the equivalent storage protein is made up of three major subunits in the range M_r 40 000 to 60 000, while in others, such as *Pisum sativum* (22) and *Lupinus angustifolius* (3), the subunit composition is much more complex and the size range more extensive. Vicilin, the 7S storage protein fraction of *P. sativum*, can be resolved by one dimensional sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) into ten major subunits with approximate M_r values of 75 000, 70 000, 50 000 (a group of subunits), 34 000, 30 000, 25 000, 18 000, 14 000, 13 000 and 12 000 (1). The group of $M_r \sim 50\,000$ is made up of a number of polypeptides which can be resolved into at least four components by one-dimensional SDS-PAGE (23) with a still more complex pattern revealed by isoelectric focusing (9).

Evidence presented earlier indicated strongly that the vicilin subunits with M_r less than 50 000 are not primary translation products but are derived from one or more members of the M_r 50 000 group by post-translational modification. This conclusion was based on the observations that *in vitro* translation experiments with total poly (A)-containing RNA from developing pea seeds showed no vicilin-related translation products smaller than M_r 50 000 (9, 12). In addition, pulse-chase labelling experiments with intact developing pea cotyledons showed that vicilin subunits smaller than M_r 50 000 were not detected until 3 to 12 h after the $M_r \sim 50\,000$ group had been synthesized, assembled into oligomeric form and transported intracellularly from their site of synthesis on the endoplasmic reticulum to their site of storage in the protein bodies (5).

The extensive post-translational modifications indicated by these results imply that there must be specific sequence relationships between some component(s) of the M_r 50 000 group and all the pea vicilin subunits of $M_r < 50\,000$, as well as between

some of the individual subunits of $M_r < 50\,000$. Previous work has already shown that the M_r 14 000 vicilin subunit is the glycosylated form of either the M_r 12 000 subunit or a closely related polypeptide (7). The present work was aimed at further defining the relationships between the vicilin subunits of M_r 50 000 and less, and identifying the processing sites present in the putative M_r 50 000 parent molecule(s). Serological relationships and amino acid sequences of vicilin subunits both indicate the existence of at least six categories of M_r 50 000 molecule of which at least three are parent molecules containing either one or both of two well-defined sites at which post-translational processing occurs. Processing at these two sites accounts for the formation of all but one of the major vicilin polypeptides of $M_r < 50\,000$.

Materials and methods

Peas (*P. sativum* L. cv. PI/G, a selection from cv. Greenfeast) were grown in artificially-lit cabinets with a 16 h photoperiod but at 20 °C rather than at 25 °C as used earlier (16). For the preparation of vicilin subunits used for anti-serum production and N-terminal amino acid sequencing, purified pea vicilin was fractionated by SDS-PAG electrophoresis on a preparative-scale using an acrylamide gradient of 12.5 to 25% (7, 18). After staining the gel with 1 M KCl, the regions of the gel containing the relevant subunits (M_r approximately 50 000, 34 000, 30 000, 25 000, 18 000, 13 000 and 12 000) were cut out, electroeluted (7) and the electroeluted subunits were precipitated with 60% (v/v) methanol and 40 mM acetic acid (final concentrations). The subunits were then washed in 80% (v/v) acetone, dissolved in 0.05% SDS and checked for purity by SDS-PAGE. Aliquots were then taken for the production of antisera in rabbits as described earlier (7) and for N-terminal amino acid sequencing as described below. To test the reaction of these antisera with vicilin polypeptides, extracts of salt-soluble proteins from developing pea cotyledons (21 days after flowering) were fractionated by gradient (12.5 to 25% acrylamide) SDS-PAGE and the polypeptides transferred electrophoreticallyto nitrocellulose paper (2), reacted with antisera and the antibody-antigen complexes detected with ^{125}I-labelled Protein A (24).

For N-terminal amino acid sequencing, aliquots of the subunits in 0.05% SDS, prepared as described above, were precipitated with 80% (v/v) acetone, washed with 100% acetone and dried. Amounts of 5 to 25 nm, dissolved in 98% formic acid, were taken for the amino acid sequence determinations. These were performed on two modified Edman-Begg-type sequenators, one incorporating a stainless steel spinning cup designed to suit a more efficient solvent extraction system, the other a stationary stainless steel reaction module. In the first system the conventional scoop at the top of a sequenator cup was replaced by tubing passing directly to and just clear of the centre of the bottom of the redesigned cup. The cup was stopped after the solvent extraction stages and the solvent transferred to waste or the fraction collector by the nitrogen pressure inside the bell jar (14). In the second sequenator the entire drive unit, spinning cup and reaction chamber were replaced by a small thermostatically controlled unit in which the chemical reactions and solvent extractions of the degradation cycle were made on a protein film immobilised on the base of the reaction vessel (Inglis and Sutherland, in preparation). The amino acid thiazolinones were converted in 25% trifluoroacetic acid (89 °C, 10 min) to the phenylthiohydantoins. These were identified by high performance liquid chromatography at 40 °C on a Zorbax ODS (Du Pont) column (ammonium acetate-acetonitrile gradient) using a Laboratory Data Control system equipped with a chromatography control module and monitoring at 269 and 320 nm.

Construction and characterization of vicilin M_r 50 000 cDNA clones

cDNA clones were constructed as previously described (4) except in this instance the starting poly(A)-containing RNA was derived from cotyledons harvested 15 days after flowering (DAF). At this stage of development, vicilin M_r 50 000 polypeptides are the major products of protein synthesis (17), but note that plants are now grown at 20 °C rather than 25 °C; 15 DAF at 20 °C is approximately equivalent in development to 11 DAF at 25 °C.

Clones for the vicilin M_r 50 000 polypeptides were identified by hybrid arrest and release translation, and by DNA sequencing followed by compar-

ison of the deduced amino acid sequence with the N-terminal amino acid sequence of vicilin M_r 50 000 polypeptides. Full details of the characterization of the vicilin clones will be published elsewhere (P. M. Chandler, in prep.). The insert in one plasmid (pBS15-84), encoding an entire vicilin M_r 50 000 polypeptide, was sequenced using the methods described by Maxam & Gilbert (15).

Results

Serological relationships

An extract of total salt-soluble proteins from the cotyledons of developing pea seeds was fractionated by SDS-PAGE and the polypeptides electrophoretically transferred to nitrocellulose paper. Replicate 1 cm tracks of this paper were then challenged with antiserum prepared against vicilin subunits of the M_r 50 000 group and those of M_r 30 000, 18 000, 14 000 or 13 000. Positive antigen–antibody interactions were detected with ^{125}I-labelled-Protein A (Fig. 1). Here we will consider only the reactions between antisera and subunits of M_r 50 000 and less, since *in vivo* labelling studies indicated strongly that the M_r 50 000 group of polypeptides contains the parent molecules for all the smaller subunits (5). Sequence relationships between the M_r 50 000 group and the larger vicilin subunits (M_r 75 000 and 70 000) are the subject of current investigations.

Examination of the reaction of individual antisera with the complete spectrum of seed protein polypeptides (Fig. 1) indicates that the vicilin subunits of M_r <50 000 fall into two serological families. One family (see Table 1) consists of the polypeptides of M_r ~ 34 000, 30 000, 18 000 and 13 000 and is characterized by failure to react with antiserum to M_r 14 000. The other family consists of polypeptides of M_r 25 000, 14 000 and 12 000 which fail to react with antisera to M_r 18 000 and M_r 30 000. All members of both families were related to the M_r 50 000 group. These results indicate that each serological family is related to sequences in different regions of the 'parent' M_r ~ 50 000 molecules.

Within the first family, the serological relationships indicate that M_r 13 000 and 18 000 subunits are related to different sequences in M_r 30 000, (which in turn is related to M_r 34 000 and M_r ~ 50 000). This is shown by the fact that antiserum against M_r 13 000 and antiserum against M_r 18 000 both react with M_r 30 000, 34 000 and 50 000 (Fig. 1, lanes 1 and 3) but anti-M_r 13 000 does not react with M_r 18 000 and anti-M_r 18 000 does not react with M_r 13 000 (Fig. 1, lanes 1 and 3). Consistent with this, anti-M_r 30 000 reacts with both M_r 13 000 and 18 000 (Fig. 1, lane 4).

Antiserum against the M_r 14 000 subunit did not react with any member of the first family (M_r 34 000, 30 000, 18 000 and 13 000) but reacted

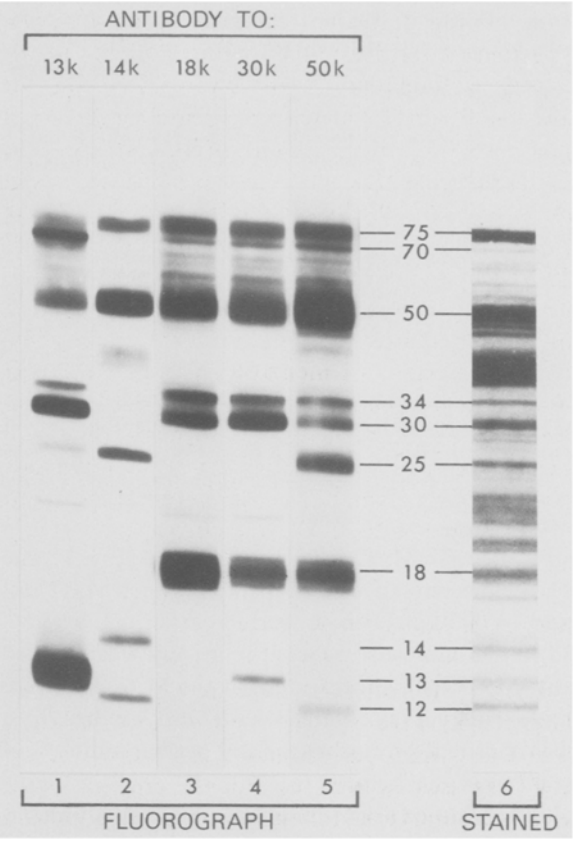

Fig. 1. Serological relationships of vicilin polypeptides. A total salt-soluble extract of cotyledons from developing pea seeds at 15 days after flowering was fractionated by SDS-PAGE. The fractionated polypeptides were transferred electrophoretically to nitrocellulose and one replicate track was stained with Amido black (lane 6). Other replicate tracks were challenged with antiserum against vicilin subunits of M_r approximately 13 000, 14 000, 18 000, 30 000 or 50 000 (lanes 1 to 5, respectively). Positive antigen–antibody reactions were detected with ^{125}I-Protein A followed by fluorography. The numbers between lanes 5 and 6 indicate the approx. $M_r \times 10^{-3}$ of the vicilin subunit in the total cotyledon extract.

Table 1. Reaction of low molecular weight vicilin polypeptides with antisera produced against selected vicilin chains.

Polypeptide	Antiserum				
	Anti M_r 13 000	Anti M_r 14 000	Anti M_r 18 000	Anti M_r 30 000	Anti M_r 50 000
34 000	++[a]	–	+++	+++	+++
30 000	+++	–	+++	+++	+++
18 000	–	–	+++	+++	+++
13 000	+++	–	–	+	–
25 000	+	+++	–	–	+++
14 000	–	+++	–	–	–
12 000	–	+++	–	–	++

[a]+++, ++, +, –, indicate strong, medium, weak and no reaction with antibody, respectively.

strongly with M_r 12 000 and M_r 25 000 (Fig. 1, lane 2). A close sequence relationship between M_r 12 000 and 14 000 has been suggested earlier (7). The M_r 25 000 subunit also reacted relatively weakly with anti-M_r 13 000, suggesting that the M_r 25 000 may have sequences in common with both families. This was confirmed by a comparison of their amino acid sequences (see below).

Several unexplained interactions were found between antisera and individual polypeptides. Antiserum against vicilin M_r 14 000 subunit consistently reacted with a family of bands at $M_r \sim 40\,000$ which do not correspond to any known vicilin component but are coincident with the acidic subunits of legumin, the second major storage protein in pea seeds (Fig. 1, lane 2). It is possible that some legumin-related polypeptides co-migrate with vicilin M_r 14 000 on SDS-PAGE and thus were contaminating the subunit preparations used to immunize rabbits. Antibodies raised against a subunit which has been prepared by one-dimensional electrophoresis may not be mono-specific owing to the presence of other polypeptides with similar electrophoretic mobility. A second unexplained interaction was the failure of anti-M_r 50 000 serum to react with M_r 13 000 and M_r 14 000 and its rather weak reaction to M_r 12 000 (Fig. 1, lane 5). In the reciprocal test, anti-M_r 14 000 reacted very strongly with the M_r 50 000 complex (Fig. 1, lane 2) but anti-M_r 13 000 reacted relatively weakly (Fig. 1, lane 1). It is possible that the antigenic sites in those regions of the $M_r \sim 50\,000$ molecules corresponding to the sequences of M_r 13 000, 14 000 and 12 000 are not readily accessible to promote antibody formation. It should also be borne in mind that in Fig. 1 the polypeptides are being challenged with antibodies raised against mature subunits. Therefore reaction of an antibody with polypeptides of greater size than the subunit against which it was raised does not represent reaction with a precursor (assuming that processing is complete), but with a molecule similar to the precursor but which was not similarly processed. If an amino acid sequence is important in both processing and in determining antigenicity it may be poorly represented in the precursor-like polypeptides.

Amino acid sequence relationships

More precise alignment of vicilin subunit sequences to one another was obtained by sequencing the N-terminal regions of selected vicilin subunits ($M_r \sim 50\,000$ group, 30 000, 25 000, 18 000, 14 000 and 12 000). The relationship of these sequences to specific regions of their putative parent $M_r \sim 50\,000$ molecules was established by comparing them with the complete amino acid sequence of one member of the $M_r \sim 50\,000$ group. This sequence was deduced from the nucleotide sequence of a cDNA clone derived from mRNA for that M_r 50 000 component. These results are summarized in Fig. 2.

Fig. 2. Amino acid sequence relationships between vicilin subunits of $M_r \sim 50\,000$ (a group of related polypeptides), 30 000, 25 000, 18 000, 14 000, 13 000 and 12 000. The N-terminal regions of the above subunits (except M_r 13 000) were determined and compared with a 432 residue polypeptide predicted from a fully sequenced cDNA representing one member of the M_r 50 000 group. The residues in this deduced sequence are numbered relative to the N-terminus of the M_r 50 000 subunits. The full sequence of M_r 13 000 subunit reported by Hirano *et al.* (13) is included for comparison.

(a): The amino acid sequence of a member of the M_r 50 000 group deduced from the sequence of its cDNA;

(b) (c) (d) (e) (g) and (h): N-terminal sequences of the M_r 50 000 group, M_r 30 000, M_r 18 000, M_r 25 000, M_r 14 000 and M_r 12 000 subunits, respectively;

(f): The complete sequence of M_r 13 000 subunit (13).

Three spaces have been inserted in the region of the M_r 13 000, 14 000 and 12 000 sequences to maximize homology. The identification of the serine residue (S) in the M_r 12 000 sequence is uncertain.

```
                                10          20          30          40          50
(a) 50k (deduced)  SRSDPQNPFI FKSNKFQTLF ENENGHIRLL QKFDQLSKIF ENLQNYRLLE

(b) 50k            SRSDPQNPFI FKSNKFQTLF
                    RSDPQNPFI FKSNKFQTLF

(c) 30k             RSDQENPFI F-

(d) 18k             RSDQENPFI FKSNRFQTLY SN-


                                60          70          80          90          100
(a) 50k (deduced)  YKSKPHTIFL PQHTDADYIL VVLSGKAILT VLKPDDRNSF NLERGDTIKL


                                110         120         130         140         150
(a) 50k (deduced)  PAGTIAYLVN RDDNKELRVL DLAIPVNRPG QLQSFLLSGN KNQQNYLSGF


                                160         170         180         190         200
(a) 50k (deduced)  SKNILEASFN TDYEEIEKVL LEEHEKETQH RRSLKDKRQQ SQEENVIVKL
(e) 25k                                                    DKDQQ SQEENVIVK-
(f) 13k                                                    DRRQE [I/L]SNENV[I/L]VKV


                                210         220         230         240         250
(a) 50k (deduced)  SRGQIEELSK NAKSTSKKSV SSESEPFNLR SRGPIYSNEF GKFFEITPEK
(f) 13k            SR[R/E]Q[I/L]EE[I/L]SK NAKSSSRRSV SSE[S/N]GPFN[I/L]R SS[E/D/R]P[I/L]YSNNS GKFFE[I/L]TPEK


                                260         270         280         290         300
(a) 50k (deduced)  NPQLQDLDIF VNSVEIKEGS LLLPHYNSRA IL IV TVNEGK GDFELVDQRN
(f) 13k            NQQ[I/L]QD[I/L]D[I/L]F VNSVD[I/L]KEGS [III/LLL]PNYNSRA [II/LL]V[I/L]V[I/L]VVNEGK GDFE[I/L]VGQRN


                                310         320         330         340         350
(a) 50k (deduced)  ENQQEQRKED DEEEEQGEEE INKQVQNYKA KLSSGDVFVI PAGHPVAVKA
(f) 13k            ENQGKEN
(g) 14k                       DKEEEQ EEE TSKQVQL-
(h) 12k                       DKEEEQ EEE TSKQVQL-


                                360         370         380         390         400
(a) 50k (deduced)  SSNLDLLGFG INAENNQRNF LAGDEDNVIS QIQRPVKELA FPRSAQEVDR


                                410         420         430
(a) 50k (deduced)  ILENQKQSHF ADAQPQQRER GSRETRDRLS SV
```

N-terminal amino acid sequencing of the $M_r \sim$ 50 000 group of polypeptides showed the presence of two chains which, in their first 20 residues, differed only to the extent that one chain contained an extra residue (serine) at the N-terminus (Fig. 2b). This relatively simple sequence for the total $M_r \sim$ 50 000 group indicates a close relationship between members of this group. The N-terminal sequences of the subunits of M_r 30 000 and 18 000 were identical to one another for at least the first 11 residues (Fig. 2c, d) and both, in turn, were identical to the M_r 50 000 N-terminal sequence which lacked the extra serine residue (Fig. 2b). The M_r 18 000 subunit was completely homologous with the M_r 50 000 sequence for the first 19 residues and then diverged at residue 20. These results indicate that M_r 30 000 and 18 000 are derived from the N-terminal region of members of the $M_r \sim$ 50 000 group.

The $M_r \sim$ 50 000 N-terminal sequence enabled us to locate the equivalent region in an amino acid sequence deduced from the nucleotide sequence of a cDNA clone complementary to mRNA for one member of the $M_r \sim$ 50 000 group (Fig. 2a). This cDNA sequence actually extends from the leader sequence through the sequence of the mature M_r 50 000 polypeptide to the poly(A) sequence in the 3′ untranslated region. Complete details of this cDNA sequence and that of clones of related members of the $M_r \sim$ 50 000 group will be published elsewhere (P. M. Chandler, in prep.). Figure 2a shows the deduced amino acid sequence for that region of the cDNA corresponding to the mature $M_r \sim$ 50 000 polypeptide. For this purpose the residue in the deduced sequence corresponding to the analytically determined N-terminus of the $M_r \sim$ 50 000 group is called residue number one. There was complete homology between the deduced sequence and the 20 N-terminal residues determined directly for the $M_r \sim$ 50 000 group (Fig. 2a, b). The first 11 residues of M_r 30 000 were also identical to the deduced sequence (Fig. 2c), as were the first 19 residues of M_r 18 000 (Fig. 2d). There was divergence in two out of the next three residues of M_r 18 000. This lack of complete homology indicates that the cDNA does not represent the particular member of the $M_r \sim$ 50 000 group which gives rise to the M_r 18 000 subunit.

The first 14 amino acids of the N-terminal sequence of the M_r 25 000 subunit diverged by only one residue from a region of the deduced amino acid sequence which began 186 amino acids from the amino acid equivalent to the N-terminus of the $M_r \sim$ 50 000 group (Fig. 2e). This sequence was also closely, but not completely, homologous with the N-terminal region of the complete sequence of the M_r 13 000 subunit reported by Hirano *et al.* (13) and included here for comparison (Fig. 2f). Single spaces have been introduced into the deduced M_r 50 000 sequence (Fig. 2a) after residues 282 and 284 in order to maximize homology with the M_r 13 000 sequence. These comparisons of amino acid sequences confirm the close relationships between $M_r \sim$ 50 000, 30 000, 18 000 and 13 000 which were deduced from the serological interactions (Fig. 1). They are also consistent with the reaction between anti-M_r 13 000 serum and the M_r 25 000 subunit (Fig. 1, lane 1).

The N-terminal sequences of the M_r 14 000 and 12 000 subunits were identical to one another (Fig. 2g, h). This is consistent with earlier evidence that they are closely related molecules, differing mainly in that the M_r 14 000 is glycosylated (7). Although M_r 14 000 and 12 000 subunits were serologically related to M_r 25 000 (Fig. 1), there was no homology between their N-terminal amino acid sequences and that of M_r 25 000. Taken together with the fact that they were serologically unrelated to the M_r 13 000 (Fig. 1, lane 1), this suggested that they would be related to the C-terminal portion of the M_r 25 000 subunit. In fact their N-terminal sequences were related to a region beginning 311 amino acids from the N-terminus of the deduced M_r 50 000 sequence, with homology in 12 out of 16 residues (Fig. 2a, g, h). Since the complete deduced M_r 50 000 sequence consisted of 432 amino acid residues, this would place the homologous region for the M_r 14 000 and 12 000 subunits at the C-terminal end of the M_r 50 000 sequence and at the C-terminal half of the M_r 25 000 sequence (Fig. 2a, e, g, h). The degree of divergence between the N-terminal region of the M_r 14 000 and 12 000 subunits and the related region of the deduced M_r 50 000 sequence indicates that the particular M_r 50 000 polypeptides represented by the cDNA is not the parent molecule for either of these small subunits. Furthermore, the deduced M_r 50 000 sequence completely lacks a potential glycosylation site, namely, Asn-X-Ser (or Thr) (19).

Discussion

Earlier evidence indicated that vicilin subunits of M_r less than ~ 50 000, i.e. M_r 34 000, 30 000, 25 000, 18 000, 14 000 and 12 000, all arose by post-translational modification of one or more members of the M_r ~ 50 000 group of subunits (5, 9, 12). At the same time it is important to note that only a portion of the M_r ~ 50 000 group is modified in this way; since subunits of M_r ~ 50 000 are the most abundant size class in mature vicilin. The present studies on serological relatedness and amino acid sequences of these subunits confirm these conclusions and permit the construction of a scheme of processing steps which lead to the formation of the smaller subunits from their putative parent molecules in the M_r ~ 50 000 group. These are summarized in Fig. 3. For convenience, the M_r 50 000 group is represented as a single molecule 432 amino acids long. This is based on the length of the open reading frame in a cDNA clone representing one member of the M_r ~ 50 000 group, taking the amino acid corresponding to the N-terminus of M_r ~ 50 000 group (Fig. 2b) as the first residue. The position of the latter was obtained directly by sequencing the N-terminal region of the total M_r ~ 50 000 group from mature vicilin (Fig. 2b).

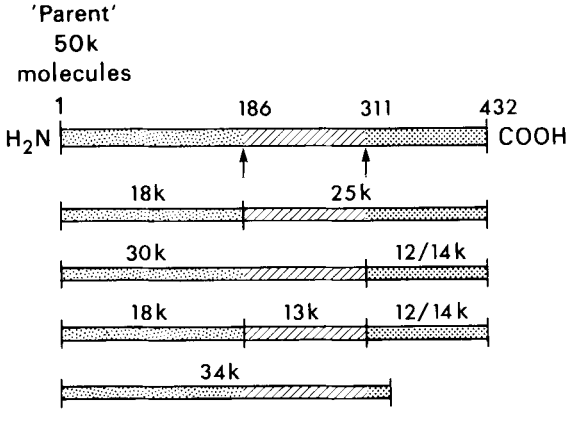

Fig. 3. A scheme representing the sequence interrelationships of the vicilin subunits. The numbers on the upper bar represent amino acid residues in the sequence deduced from a cDNA of one member of the M_r 50 000 group. Arrows indicate potential sites of endoproteolytic cleavage within members of the M_r 50 000 group. Processing of a range of parent molecules at either one or both of these sites yields the smaller subunits as indicated in the lower bars. Subunits of M_r 50 000, 34 000, 25 000, 18 000, 14 000, 13 000 and 12 000 are designated as 50k, 34k etc.

The results indicate that the subunits of M_r 30 000, 25 000, 18 000, 14 000, 13 000 and 12 000 arise from endoproteolytic cleavage at either one or both of two processing sites in different parent M_r 50 000 molecules. The sum of the apparent molecular weights of the component polypeptides exceeds M_r 50 000. This implies that there must be extensive overlap between sequences of members within each family. However, it is not known whether processing always occurs wherever there is a processing site. These sites are at or about position 185/186 and 310/311 in the deduced M_r 50 000 sequence. Processing only at position 185/186 yields the M_r 18 000 and 25 000 subunits; processing only at position 310/311 yields the M_r 30 000 and M_r 12 000 or 14 000 subunits; processing at both these positions yields the M_r 18 000, 13 000, and the 12 000 or 14 000 subunits (Fig. 3). The M_r 30 000 and 18 000 subunits are derived from the N-terminus of their parent M_r ~ 50 000 molecules. We have no sequence data for the M_r 34 000 subunit but serological data indicate that it is a large N-terminal fragment of an M_r 50 000 polypeptide because it shows close serological relatedness to the M_r 30 000 subunit and consequently with the smaller subunits of M_r 18 000 and 13 000 (Fig. 1). This assignment of processing sites to specific residues is done with reference to only one member of the M_r ~ 50 000 group (Fig. 2a). The actual sites may vary slightly when the true parent molecules are identified and sequenced.

While this manuscript was in preparation, Gatehouse *et al.* (10) reported results which are consistent with the above findings and conclusions. On the basis of tryptic peptide profiles and amino acid sequencing of some of the vicilin subunits and comparisons with a partial sequence of a cDNA complementary to an M_r 50 000 polypeptide, they concluded that there were up to two processing sites in precursor M_r 50 000 molecules. Their ordering of the position of the smaller subunits relative to the precursors is also in agreement with our findings.

Assuming all eligible sites are processed, our data suggest that the M_r 50 000 group must consist of at least six classes of molecule. Since M_r 50 000 subunits are a major component of fully processed vicilin, there is at least one member of this group which is not processed. Earlier work (1) indicates that there are at least two sub-classes within this category of M_r 50 000 molecule, namely, glycosylated and

non-glycosylated. Of the M_r 50 000 components which are post-translationally processed, one class yields M_r 18 000 and 25 000; another yields M_r 18 000, 13 000 and 14 000 or 12 000; another is processed only at position 311 to yield M_r 30 000 and 14 000 or 12 000. There is presumably a sixth category of $M_r \sim 50\,000$ molecule which is processed near position 311, possibly from a larger parent molecule, to yield the M_r 34 000 subunit.

The processing scheme outlined in Fig. 3 accounts for essentially the entire length of the particular $M_r \sim 50\,000$ molecule from which the cDNA clone was derived. The M_r of this molecule calculated from the deduced amino acid sequence is 49 559. Assuming no major losses of amino acid residues during processing, and using the deduced M_r 50 000 sequence as a guide to the amino acid composition of the smaller subunits as outlined in Fig. 3, one can calculate maximum molecular weights for each of the subunits which arise by post-translational cleavage. Thus, the subunits referred to as M_r 30 000, 25 000, 18 000, 13 000 and 12 000 on the basis of their relative mobility on SDS-PAGE could have maximum molecular weights of 36 011, 28 033, 21 526, 14 485 and 13 548, respectively. With one exception, the precise length of the subunits smaller than $M_r \sim 50\,000$ is not known. Hirano *et al.* (13) sequenced the M_r 13 000 subunit and estimated that its molecular weight is approximately 14 000. A comparison of the C-terminus of M_r 13 000 (Fig. 2f) with the adjacent N-terminus of the M_r 12 000 and 14 000 (Fig. 2g, h) and the corresponding region of the deduced sequence reported here (residues 186 to 310, Fig. 2a) indicates that at least three amino acid residues may be lost from the C-terminus of the M_r 13 000 subunit during processing. The loss of these three residues accounts for the difference between our calculated maximum value (14 485) and the value estimated by protein sequencing (13). The maximum molecular weights for other subunits listed above differ more substantially from their M_r values, suggesting either that they undergo more extensive loss of residues during processing or that they behave anomalously on SDS-PAGE. There is also evidence for post-translational removal of four amino acids from the carboxyl-terminus of the $M_r \sim 6\,000$ subunit of pea seed lectin, another protein which is stored and post-translationally processed in the protein bodies (11).

The present study does much to resolve the great complexity seen in pea vicilin subunits. The subunits of $M_r < 50\,000$ arise post-translationally (5) by cleavage of a range of related but not identical M_r 50 000 molecules at either one or both of two potential processing sites. The simplest hypothesis is that the $M_r \sim 50\,000$ group represents a multigene family whose members have diverged sufficiently to generate the different combinations of processing sites revealed by this work. Some members of the family have no processing sites and these constitute the M_r 50 000 group of mature vicilin. Since the processing events occur after the 'parent' molecules have been assembled into oligomers and transported to the protein bodies (5), the processing sites must occur in regions which remain exposed to the action of protease(s) after polypeptide chain folding and assembly. It should be noted that the type and extent of post-translational processing is highly reproducible and inheritable (23). In some other legumes, such as soybean and French bean, this extensive post-translational cleavage of the proteins equivalent to pea vicilin does not occur (see Introduction).

The serological studies (Fig. 1) reported here revealed common antigenic determinants between the larger vicilin subunits of M_r 75 000 and 70 000 and the $M_r \sim 50\,000$ group. However, there is evidence from *in vivo* and *in vitro* experiments which indicates that these larger subunits are distinct gene products. *In vivo* labelling experiments show that they are regulated independently of the $M_r \sim 50\,000$ group during development (18) and products corresponding to $M_r \sim 50\,000$, 70 000 and 75 000 are all formed by *in vitro* translation of total RNA from developing cotyledons (12).

The post-translational, endoproteolytic cleavage of vicilin $M_r \sim 50\,000$ polypeptides described here can be compared and contrasted with the post-translational processing of pea legumin and lectin. Legumin is fully processed from $M_r \sim 60\,000$ precursors to yield $M_r \sim 40\,000$ and 20 000 polypeptides (6, 17). Pea lectin is processed from a $M_r \sim 23\,000$ precursor to yield two polypeptides also (M_r 17 000 and 6 000) but in this case only a small amount of the parent molecule can be found in mature seeds (11). On the other hand, although vicilin processing involves at least two endoproteolytic cleavages to yield an array of small products, a large proportion of the high molecular weight

molecules ($M_r \sim 50\,000$) remain uncleaved. From sequencing data on several cDNAs, we have evidence of significant amino acid sequence divergence among members of the $M_r \sim 50\,000$ group (P. M. Chandler, in prep.).

Acknowledgements

We thank Zufi Ariffin for her help in sequencing the cDNA, Edward Newbigin for technical assistance and Dr P. R. Whitfeld for his comments during preparation of this manuscript.

References

1. Badenoch-Jones J, Spencer D, Higgins TGV, Millerd A: The role of glycosylation in storage-protein synthesis in developing pea seeds. Planta 153:201–209, 1981.
2. Bittner M, Kupferer P, Morris CE: Electrophoretic transfer of proteins and nucleic acids from slab gels to diazobenzyloxymethyl cellulose or nitrocellulose sheets. Anal Biochem 102:359–471, 1980.
3. Blagrove RJ, Gillespie JM: Isolation, purification and characterization of the seed globulins of Lupinus angustifolius. Aust J Plant Physiol 2:13–27, 1975.
4. Chandler PM, Higgins TJV, Randall PJ, Spencer D: Regulation of legumin levels in developing pea seeds under conditions of sulfur deficiency: rates of legumin synthesis and levels of legumin mRNA. Plant Physiol 71:47–54, 1983.
5. Chrispeels MJ, Higgins TJV, Spencer D: Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. J Cell Biol 93: 306–313, 1982.
6. Croy PRD, Gatehouse JA, Evans DM, Boulter D: Characterization of the storage protein subunits synthesized in vitro by polyribosomes and RNA from developing pea (Pisum sativum L.) I. Legumin. Planta 148:49–56, 1980.
7. Davey RA, Higgins TJV, Spencer D: Homologies between two small subunits of vicilin from Pisum sativum. Biochem Internat 3:595–602, 1981.
8. Derbyshire E, Wright DJ, Boulter D: Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 15:3–24, 1976.
9. Gatehouse JA, Croy RRD, Morton H, Tyler M, Boulter D: Characterisation and subunit structures of the vicilin storage proteins of pea (Pisum sativum L.) Eur J Biochem 118: 627–633, 1981.
10. Gatehouse JA, Lycett GW, Croy RRD, Boulter D: The post-translational proteolysis of the subunits of vicilin from pea (Pisum sativum L.). Biochem J 207:629–632, 1982.
11. Higgins TJV, Chandler PM, Zurawski G, Button SC, Spencer D: The biosynthesis and primary structure of pea seed lectin. J Biol Chem 258:9544–9549, 1983.
12. Higgins TJV, Spencer D: Precursor forms of pea vicilin subunits. Modifications by microsomal membranes during cell-free translation. Plant Physiol 67:205–211, 1981.
13. Hirano H, Gatehouse JA, Boulter D: The complete amino acid sequence of a subunit of the vicilin seed storage protein of pea (Pisum sativum L.). FEBS Lett 145:99–102, 1982.
14. Inglis AS: Modified extraction procedures in spinning-cup sequenators. In: Hirs CWH, Timasheff SN (eds) Methods in Enzymology, 91. Academic Press, New York, 1983. pp 443–450.
15. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds) Methods in Enzymology, 65. Academic Press, New York, 1980, pp 499–560.
16. Millerd A, Spencer D: Changes in RNA-synthesizing activiy and template activity in nuclei from cotyledons of developing pea seeds. Aust J Plant Physiol 1:331–341, 1974.
17. Spencer D, Higgins TJV: The biosynthesis of legumin in maturing pea seeds. Biochem Internat 1:502–509, 1980.
18. Spencer D, Higgins TJV, Button SC, Davey RA: Pulse-labelling studies on protein synthesis in developing pea seeds and evidence of a precursor form of legumin small subunit. Plant Physiol 66:510–515, 1980.
19. Struck DK, Lennarz WJ, Brew K: Primary structural requirements for the enzymatic formation of the N-glycosidic bond in glycoproteins. J Biol Chem 253:5786–5794, 1978.
20. Sun SM, McLeester RC, Bliss FA, Hall TC: Reversible and irreversible dissociation of globulins from Phaseolus vulgaris seed. J Biol Chem 249:2118–2120, 1974.
21. Thanh VH, Shibasaki K: Major proteins of soybean seeds. Subunit structure of β-conglycinin. J Agric Food Chem 26: 692–695, 1978.
22. Thomson JA, Schroeder HE, Dudman WE: Cotyledonary storage proteins in Pisum sativum. I. Molecular heterogeneity. Aust J Plant Physiol 5:263–279, 1978.
23. Thomson JA, Schroeder HE, Tassie AM: Cotyledonary storage proteins in Pisum sativum. V. Further studies on molecular heterogeneity in the vicilin series of holoproteins. Aust J Plant Physiol 7:271–282, 1980.
24. Towbin H, Staehlin T, Gordon J: Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76:4350–4354, 1979.

Received 24 May 1983; accepted 28 July 1983.