

Molecular Cloning of a Protein Associated with Soybean Seed Oil Bodies That Is Similar to Thiol Proteases of the Papain Family*

(Received for publication, February 27, 1990)

Andrzej Kalinski, Jane M. Weisemann, Benjamin F. Matthews, and Eliot M. Herman

From the Plant Molecular Biology Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland 20705

A 34,000-Da protein (P34) is one of the four major soybean oil body proteins observed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of isolated organic solvent-extracted oil bodies from mature seeds. P34 is processed during seedling growth to a 32,000-Da polypeptide (P32) by the removal of an amino-terminal decapeptide (Herman, E. M., Melroy, D. L., and Buckhout, T. J. (1990) *Plant Physiol.*, in press). A soybean λ ZAP II cDNA library constructed from RNA isolated from midmaturation seeds was screened with monoclonal antibodies directed against two different epitopes of P34. The isolated cDNA clone encoding P34 contains 1,350 base pairs terminating in a poly(A)⁺ tail and an open reading frame 1,137 base pairs in length. The open reading frame includes a deduced amino acid sequence which matches 23 of 25 amino-terminal amino acids determined by automated Edman degradation of P34 and P32. The cDNA predicts a mature protein of 257 amino acids and of 28,641 Da. The open reading frame extends 5' from the known amino terminus of P34 encoding a possible precursor and signal sequence segments with a combined additional 122 amino acids. Prepro-P34 is deduced to be a polypeptide of 42,714 Da, indicating that the cDNA clone apparently encodes a polypeptide of 379 amino acids. A comparison of the nucleotide and deduced amino acid sequences in the GenBankTM Data Bank with the sequence of P34 has shown considerable sequence similarity to the thiol proteases of the papain family. Southern blot analysis of genomic DNA indicated that the P34 gene has a low copy number.

Seed oils are reserve substances that are stored in discrete organelles termed oil bodies, lipid bodies, or spherosomes which are accumulated in storage tissues such as cotyledons, endosperm, or scutellum. Proposed theories on the mechanism of assembly of oil bodies in maturing seeds are still controversial. The accumulated oil bodies are utilized as a carbon source during seedling growth. However, the mechanism by which the oil bodies are mobilized is not well understood.

Oil bodies are very simple structures, consisting of a central core of triglyceride oil surrounded by a monolayer of phos-

pholipids containing a few polypeptides (1-3). Recent work in this (4) and other laboratories (5, 6) has studied these proteins as *in situ* probes to examine regulation and developmental ontogeny. Studies involving biochemical and immunological techniques have partially characterized the major oil body proteins of soybean (24-kDa oleosin (4) maize (L3, 16-kDa oleosin (5, 6), and rapeseed (20-kDa oleosin (7). The sequence of a near full-length cDNA clone for one of these proteins, L3 in maize, has been published (5) as well as the cDNA and genomic sequence of a related maize 18-kDa oleosin (8). Our studies have focused on the ontogeny of soybean oil body proteins. The soybean oil body has four abundant proteins of 34, 24, 18, and 17 kDa (4), referred to herein as P34, 24-kDa oleosin, P18, and P17, respectively. We have found that the soybean 24-kDa oleosin has sequence similarity to the maize 16-kDa oleosin.¹ Immunochemical observations with monoclonal antibodies have shown that one of these proteins, P34, is processed to 32,000 Da (P32) during seedling growth (9). This processing is the apparent consequence of the removal of a highly basic decapeptide from the amino terminus of P34 (9). To obtain more detailed information on the structure and regulation of this protein, we have isolated an apparent full-length cDNA clone of P34 from a λ ZAP II expression vector library. Reported herein is the deduced sequence of the P34 clone which apparently encodes a polypeptide with close similarity to the thiol proteases of the papain family.

EXPERIMENTAL PROCEDURES

Materials—Seeds of soybean (*Glycine max* L. Merr. cv Century, Pando, Minsoy, PI 290136, Barc 2, PI 83945, and T 135) were obtained from Dr. Robert Yaklich (Germplasm Quality and Enhancement Laboratory) and Dr. Thomas Devine (Plant Molecular Biology Laboratory, United States Department of Agriculture, Beltsville, MD). Leaves and cotyledons from maturing seeds were obtained from greenhouse-grown plants maintained as previously described (4). The *recA*⁺ *Escherichia coli* host strain XL 1-Blue, R408 helper phage, the M13 forward primer, and some restriction enzymes were from Stratagene Cloning Systems. Additional restriction enzymes were purchased from United States Biochemical Corp. T4 polynucleotide kinase was purchased from Bethesda Research Laboratories; and DNA size markers were obtained from Pharmacia LKB Biotechnology Inc. A random-primed DNA labeling kit was obtained from Boehringer Mannheim. [α -³²P]dCTP (3000 Ci/mmol) and [γ -³²P]ATP (3000 Ci/mmol) were from Amersham Corp. and Du Pont-New England Nuclear, respectively. Double-stranded phagemid DNA was isolated as described by Holmes and Quigley (10). Guanidinium isothiocyanate, agarose, distilled phenol, and proteinase K were from Bethesda Research Laboratories. All other chemicals were of analytical-grade.

Maturing Soybean Seed cDNA Library and Antibody Screening—Total RNA was isolated from 10 g of seed tissue. The seeds were

* This work was supported in part by United States Department of Agriculture Office of Competitive Grant CRCR-86-2021 (to E. M. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J05560.

¹ A. Kalinski, J. M. Weisemann, B. F. Matthews, and E. M. Herman, submitted for publication.

ground in liquid nitrogen with a mortar and pestle, and the powder was suspended in 14 ml of 25 mM sodium citrate, pH 7.0, 4 M guanidinium isothiocyanate, 1.5% (w/v) sodium lauroylsarcosine, 100 mM β -mercaptoethanol and agitated for 60 s at room temperature. Total RNA was further extracted with phenol/chloroform (1:1), selectively precipitated with 3 M LiCl, and then redissolved in 2% potassium acetate and precipitated with ethanol (11). Two milligrams of total RNA were used for the custom construction of a cDNA by Stratagene Cloning Systems. cDNAs of 500 bp² or greater in size were cloned into the unique *Eco*RI site of λ ZAP II DNA as described (12). This library was plated in 0.7% agarose at a density of 20,000 plaques/100-mm plate on XL1-Blue cells and grown for 4 h at 42 °C. Dry nitrocellulose filters previously soaked in 10 mM isopropyl- β -D-thiogalactopyranoside were then placed on plates, and fusion protein synthesis was induced for 4.5 h at 37 °C. The filters were removed from plates; twice washed in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl (TBS) for 5 min at room temperature; and then soaked in 3% (w/v) gelatin/TBS for 1 h with gentle shaking at room temperature. Following the washing steps, the filters were incubated overnight in a 1:100 dilution of monoclonal antibody P3E1 ascites fluid (9) in 1% (w/v) gelatin/TBS containing 0.05% (v/v) Tween 20 (TBST). Following the incubation, the filters were washed in TBST and indirectly labeled with a 1:1000 dilution of anti-mouse IgG-alkaline phosphatase conjugate in 1% gelatin/TBST for 1.5 h. Alkaline phosphatase was visualized with the chromogenic substrates nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate. Positive plaques were re-screened with monoclonal antibody P4B5 ascites fluid, which is directed at a different epitope on the P34 polypeptide (9). Doubly positive plaques, immunoreactive with both monoclonal antibodies P3E1 and P4B5, were rescreened at low density with monoclonal antibody P3E1 to ensure that each positive plaque was derived from a single parent phage. *In vivo* excision and recircularization of the cDNA insert contained within the λ ZAP II vector to form pBluescript SK(-) double-stranded phagemid was as described (12).

Probing P34 cDNA pBluescript SK(-) with Synthetic Oligonucleotide—The degenerate mixture of oligonucleotide probes coding for the amino-terminal sequence of P34 (9) (NH₂-Lys-Lys-Met-Lys-Lys-Glu-Gln-Tyr-) was synthesized (Synthecell Corp.) (5'-AA(A/G)AA(A/G)ATGAA(A/G)AA(A/G)GA(A/G)CA(A/G)TA(T/C)-3') and 5'-end-labeled with [γ -³²P]ATP using T4 polynucleotide kinase. The labeled oligonucleotides were separated from unincorporated label on Sephadex G-50. Two micrograms of pBluescript SK(-) DNA containing P34 cDNA were digested overnight with 8 units of *Eco*RI. The digested DNA was electrophoresed in a 1% agarose gel. The DNA was then blotted onto a nitrocellulose filter and heated at 80 °C for 2 h. The filter was prehybridized for 1 h at 57 °C in 6 × SSPE, 0.5% SDS, 5 × Denhardt's solution, and 50 μ g/ml denatured salmon testes DNA. Hybridization was performed in the same solution with added 5'-end-labeled synthetic oligonucleotide (10⁷ cpm/ml) for 17 h at 57 °C. The filter was washed twice at room temperature for 20 min in 2 × SSPE, 0.1% SDS and once at 57 °C for 30 min in the same solution. The filter was air-dried, and autoradiography was done at -70 °C for 4 h.

Isolation of DNA and Southern Blot Analysis—Cellular DNA was isolated from soybean leaves as described previously (13). Eight to nine micrograms of high molecular weight DNA were digested with 75–100 units of a given restriction enzyme under conditions specified by the manufacturer and electrophoresed on a 1% agarose gel. DNA was blotted in 20 × SSC onto a nitrocellulose filter and heated at 80 °C for 2 h. The filter was prehybridized with a solution containing 5 × SSPE, 5 × Denhardt's solution, 50% formamide, 0.1% SDS, and 200 μ g/ml denatured salmon testes DNA for 3 h at 42 °C. Hybridization was performed overnight at 42 °C in the same solution containing denatured random-primed labeled P34 cDNA (8 × 10⁶ cpm/ml) prepared according to Feinberg and Vogelstein (14). Following hybridization, the filter was washed twice in 2 × SSPE, 0.1% SDS and in 0.1 × SSPE, 0.1% SDS at room temperature for 15 min. Finally, the filter was washed in 0.1 × SSPE, 0.1% SDS at 65 °C for 25 min. The nitrocellulose was exposed to Kodak X-AR film with intensifying screens at -80 °C overnight.

DNA Sequence Analysis—P34 inserts were further digested with *Bam*HI, *Hae*III, or *Taq*I and subcloned into the M13 vectors mp18 and mp19 (15). The sequencing method used was the dideoxy chain termination method of Sanger *et al.* (16). Two sets of reactions were done using DNA polymerase I (Klenow fragment). The remainder

were done using modified T7 DNA polymerase (Sequenase 2.0). The primers used were the M13 forward primer and two oligonucleotides made using an Applied Biosystems DNA synthesizer: 5'-GGCTTTGCATCTACCC-3' (corresponding to bases 697 to 682 of the insert) and 5'-GTACTTGCAAGCTCCCAAG-3' (corresponding to bases 320–338 of the insert). DNA sequences were compiled and analyzed with DNA Inspector IIe (Textco) running on a Macintosh SE/30 computer and the University of Wisconsin GCG sequence analysis package running on a VAX 8250.

RESULTS

Identification of cDNA Encoding P34—Approximately 1.2 × 10⁶ recombinants from a soybean λ ZAP II cDNA library were initially screened for isopropyl- β -D-thiogalactopyranoside-induced expression of the P34 fusion protein using anti-P34 monoclonal antibody P3E1 (9). Three recombinant λ plaques which were strongly immunoreactive were isolated. These plaques were replated and assayed for the isopropyl- β -D-thiogalactopyranoside-induced expression of the P34 fusion protein by immunolabeling with anti-P34 monoclonal antibody P4B5 (9). Monoclonal antibodies P3E1 and P4B5 are directed against distinct epitopes on the P34 polypeptide separated by at least 5 kDa. Recombinant phage immunoreactive with both monoclonal antibodies were selected for *in vivo* excision of the pBluescript SK(-) phagemid. The insert size of the selected P34 cDNA clone was determined to be ~1.3 kilobase pairs by agarose gel electrophoresis of the phagemid *Eco*RI digest (Fig. 1A). Since the amino-terminal sequence of P34 was experimentally determined (9), this sequence was used to synthesize a degenerate oligonucleotide probe. The *Eco*RI-liberated insert strongly hybridized to the probe corresponding to the first 8 amino acids of the amino terminus of P34 (Fig. 1B). Therefore, we concluded that the selected cDNA clone appeared to encode both the amino terminus of P34 and two distinct sites in the interior of the polypeptide.

Sequence of cDNA and Deduced Amino Sequence of P34—The partial restriction map and sequence strategy are illustrated in Fig. 2. The entire 1,350-bp sequence and the deduced amino acid sequence of the largest open reading frame which includes the experimentally determined amino-terminal sequence (9) are shown in Fig. 3. Three potential ATG start codons 5' from the known amino terminus located at nucleotides 3, 234, and 360 are included in this open reading frame. The nucleotide sequence surrounding the third ATG codon (CAAAATGGC) is very similar to the consensus sequence for translation initiation in plants (AACAATGGC) (17, 18). The third ATG codon at bp 360 is 9 bp upstream from the

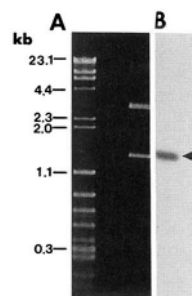
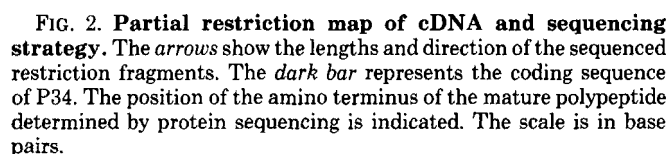


FIG. 1. *Eco*RI release of P34 insert from pBluescript SK(-) phagemid DNA (A) and Southern hybridization (B) with mixture of synthetic oligonucleotides corresponding to amino terminus of P34. Two micrograms of pBluescript SK(-) DNA containing P34 cDNA were digested with *Eco*RI, separated on a 1% agarose gel, transferred to nitrocellulose, and hybridized with a mixture of ³²P end-labeled synthetic oligonucleotides coding for the amino terminus of P34 (Lys-Lys-Met-Lys-Lys-Glu-Gln-Tyr). The arrowhead denotes the position of the cDNA insert. kb, kilobase pairs.

² The abbreviations used are: bp, base pair(s); SDS, sodium dodecyl sulfate.

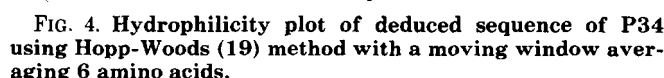
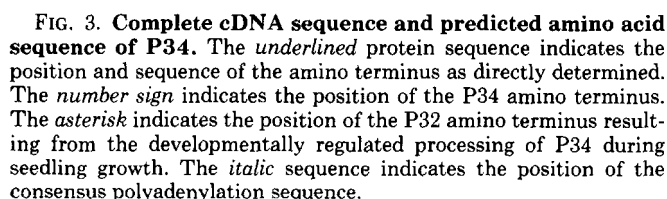


Confirmation that the cDNA clone encodes a polypeptide corresponding to P34 was obtained by comparing the deduced amino acid sequence with the 25 amino-terminal residues experimentally determined for P34 (9). The deduced sequence from bp 369 to 443 shows exact identity for 23 of the 25 amino acids. The calculated amino acid composition of the open reading frame from bp 360 to 1139 is in good agreement with the experimentally determined amino acid composition of the purified protein (Table I).

The hydrophilicity of the deduced amino acid sequence was plotted using the algorithm of Hopp and Woods (19) and is shown in Fig. 4. The Hopp-Woods plot also shows that no particular structural feature dominates the internal region of P34.

A homology search of the P34 cDNA was run with the

Comparison of cDNA-derived amino acid composition of P34 for the open reading frame at bp 369-1139 with experimental data determined from the purified polypeptide

^a ND, not determined.

GenBank™ Data Bank. Nucleic acid sequence homology was found with members of the papain family of thiol proteases. The highest homology scores were observed with full-length clones of the prestalk thiol protease (20), developmentally regulated thiol proteases I and II of *Dictyostelium discoideum* (21), and actinidin from kiwi fruit (22, 23). Other members of the thiol protease family including papain (24), aleurain (25), and cathepsin (26–29) also were observed to have high homology scores. High homology was also observed with two partial-length cDNA clones isolated from plants: C14, the chilling-induced protease of tomato fruit (30), and COT44 from *Brassica* seeds (31). Table II shows a compilation of the percentages of identical and similar amino acids for diverse full-length and partial cDNA clones of thiol proteases. Gap alignment using the GCG software package was used to compare the sequence similarities of P34 with the other papain superfamily proteases in the GenBank™ Data Bank. The best alignment of the preprosegment of P34 was observed with the *Dictyostelium* prestalk protease (20) and cysteine protease I (21). However, the best alignment of the mature P34 sequence was observed with several plant proteases including actinidin (23), tomato C14 (30), and *Brassica* COT44 (31).

The P34 cDNA contains the 2 amino acids which are positioned to be possible reactive nucleophiles (Cys³⁴ and His¹⁷⁹) as well as the other amino acids which may surround the active site (Gln³¹, Asn¹⁷⁸, Asn¹⁹⁹, Ser²⁰⁰, and Trp²⁰¹). However, Cys³⁴, which is a potential active-site residue, aligns with a cysteine that constitutes one-half of a disulfide bridge in the other thiol proteases (Table III). The residue which most closely aligns with the active site in P34 is glycine 37. Whether cysteine 34, which aligns only a few amino acids away from the catalytic cysteine, is close enough to function as a potential active-site nucleophile will be resolved only when we determine whether P34 has catalytic activity. P34 is unique among the papain superfamily members which have been described in replacing a glycine for the cysteine at position 37. Three disulfide bridges are found in most members of the papain superfamily; of these, two of the cysteine pairs are present in P34 (Cys⁶⁸-Cys¹⁰⁸ and Cys¹⁷¹-Cys²²⁴). However, one of the disulfide bridges found in the other thiol proteases is present as cysteine 34 and asparagine 77 in P34. Therefore, it is likely that cysteine 34 is present as a free thiol. One additional cysteine at position 10 in the *in vivo* form of P34 has no equivalent in the other described members of the

TABLE III

Alignment of active-site amino acids of diverse thiol proteases

The alignment of the residues surrounding the active-site thiol (*) and 1 of the cysteines which is part of a disulfide bridge (#) is shown. Note that P34 is unique among the known sequences of this family of proteins in that it lacks the cysteine which is aligned as the active-site thiol.

<i>Dictyostelium</i> CPI ^a	QGQCGSCWSF
<i>Dictyostelium</i> CPII	QGQCGSCWSF
<i>Dictyostelium</i> PSC	QGQCGSCWSF
Human cathepsin B	QSQCGSCWAF
Rat cathepsin H	QAQCGSCWTF
Actinidin	QGECGGCWAF
Papain	QGACGSCWTF
Aleurain	QAHCSCWTF
C14 (tomato)	QGSCGSCWAF
P34	QGCGGSCWAF
	# *

^a CPI and CPII, cysteine proteases I and II, respectively; PSC, prestalk clone.

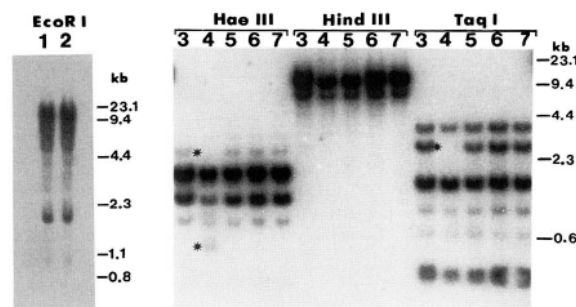


FIG. 5. Southern blot analysis of soybean genomic DNA with P34 cDNA probe. Genomic DNAs (8–9 µg/lane) isolated from cv Century (lane 1), Pando (lane 2), Minsoy (lane 3), PI290136 (lane 4), Barc 2 (lane 5), PI 83945 (lane 6), and T 135 (lane 7) were digested with the indicated restriction enzymes and subjected to blot hybridization using the ³²P-labeled cDNA. Differences in hybridization pattern are indicated (asterisks). kb, kilobase pairs.

papain superfamily. There are a number of glycine molecules, termed type II, which are important for conformation in thiol proteases. Of the 9 type II glycine residues present in actinidin, 6 are conserved in aligned locations in P34. The close similarity of P34 to other members of the papain superfamily, particularly in the regions of the active-site and glycine and disulfide residues important for conformation, suggests that P34 cDNA encodes a thiol protease.

The open reading frame of P34 may be extended 5' from the known amino terminus. The in-frame ATG codon located at bp 3 is not in good agreement with the plant consensus initiation codon. However, the deduced amino acid sequence from this ATG codon at bp 3 to the third ATG codon at bp 360 encodes a putative prosequence of an additional 122 amino acids. Extensive amino acid similarity is observed between this deduced precursor and the precursors of other thiol proteases (21, 22, 24–26, 30). This suggests that if translation begins at bp 3, then there is an additional 122-amino acid precursor to the P34 polypeptide. We are presently conducting pulse-chase experiments to examine this possibility.

Southern Blot Analysis—The number of P34 coding sequences and DNA polymorphisms in or around the P34 gene(s) was examined on Southern blots of genomic DNA. Genomic DNA isolated from soybean cultivars was digested with *Eco*RI, *Hae*III, *Hind*III, and *Taq*I; and the resulting fragments were analyzed for hybridization to P34 cDNA

TABLE II

Similarity of P34 to the other members of the papain superfamily which have been reported as full-length or partial sequences in order of identical amino acid score

Protease	Similarity	Identical	Refs.
	%	%	
Full-length clones			
<i>Dictyostelium</i> PSC ^a	56	40	20
<i>Dictyostelium</i> CPII	56	40	21
Actinidin	62	39	23
Papain	56	38	24
Rat cathepsin L	57	36	29
<i>Dictyostelium</i> CPI	55	35	21
Aleurain	56	34	25
Human cathepsin B	51	26	27
Mouse cathepsin B	48	25	28
Partial clones			
Tomato C14	65	44	27
<i>Brassica</i> COT44	60	42	28

^a PSC, prestalk clone; CPI and CPII, cysteine proteases I and II, respectively.

under stringent conditions (Fig. 5). Two hybridization bands were detected in *Hind*III digests, and seven bands were detected in *Eco*RI digests. Neither of these enzymes has restriction sites within the P34 cDNA. We conclude that P34 is coded for by a family of genes with a low copy number. *Hae*III and *Taq*I have recognition sites within the cDNA sequence, and digestion of genomic DNA produced four to six bands of varying intensities. The hybridization pattern of *Hae*III- and *Taq*I-digested DNAs isolated from the soybean PI290136 cultivar showed restriction fragment polymorphisms at 2.0, 2.3, and 0.6 kilobase pairs. It is not yet clear whether the polymorphic bands represent a significant change in the total amount of homologous sequences in the genome of this soybean cultivar and/or a redistribution of closely related sequences among the various bands. Nevertheless, a P34 cDNA probe can be used to differentiate this cultivar from Minsoy, Barc 2, PI 83945, and T 135 cultivars.

DISCUSSION

We have isolated and characterized a cDNA clone encoding the P34 polypeptide associated with soybean oil bodies (4, 9). The deduced amino acid sequence of P34 cDNA suggests that this clone encodes a thiol protease. A comparison with the sequence in the GenBank™ Data Bank clearly shows that P34 is most similar to the developmentally regulated thiol proteases of *Dictyostelium* (20, 21), actinidin from kiwi fruit (22, 23), and, at slightly lower similarity, aleurain (the hormone-induced protease of barley aleurone) (25), papain (24), and cathepsin (26–29). Why P34 is more similar to the *Dictyostelium* prestalk protease and thiol protease II than to several of the plant thiol proteases is unknown. Many of the amino acid residues mediating the proteolytic activity and surrounding the active site are conserved in this family of proteases as well as in P34. Other structural features involved in the conformation of the other members of this thiol protease family, the cysteines used for disulfide bonds (32) and the type II glycines (33), are conserved to a large degree in P34. However, we have not yet demonstrated that P34 has proteolytic activity. Current work in this laboratory is directed toward that goal.

The sequence of P34 is particularly interesting in the putative proenzyme region. Inspection of the deduced sequence between the first and third possible start codons shows a high degree of sequence similarity between the precursor sequences of several thiol proteases and that of P34. This similarity is particularly striking in the most highly conserved residues 5' from the third ATG codon, indicating that there is considerable sequence similarity between the precursor of P34 and the precursors of *Dictyostelium* prestalk protease (20) and thiol protease II (21).

A free cysteine is required for catalytic activity of these proteases. The catalytic cysteine is one of the most highly conserved amino acid residues in the members of the thiol protease family. However, P34 is different from all of the previously analyzed members of the thiol protease family because it lacks the presumptive catalytic cysteine, which is present as glycine 37. Another cysteine at position 34 in P34 aligns as part of a disulfide bond in other thiol proteases. However, the cysteine which should be the other half of the disulfide bond is present as an asparagine in P34. Whether this could free the closely adjacent cysteine 34 to substitute for the catalytic cysteine is unknown.

The thiol proteases of the papain family function in diverse physiological roles. Cathepsin members of this family are used as lysosomal enzymes mediating protein degradation. It is also assumed, but not yet demonstrated, that the *Dictyoste-*

lium proteases are used to degrade intracellular proteins during development. Plant thiol proteases such as bromelain and papain are likely to be localized in the vacuole, a compartment which is the functional equivalent of the animal cell lysosome (34). Other members of the thiol protease family are secreted from the cell, for example the barley aleurone protease aleurain (25), which is utilized in the mobilization of extracellular reserve proteins. Other thiol proteases appear to be expressed only in response to stress, such as the chilling-induced tomato fruit C14 protease (30). A few of the thiol proteases have been documented as functioning in very specific roles in processing proinsulin (35) and proapolipoprotein A-II (36).

P34 appears to be associated with soybean seed oil bodies. The proteins which bind the oil bodies compose about 2–3% of the total seed protein, of which the major protein (24-kDa oleosin) is catabolized during seedling growth.³ This laboratory has previously shown (9) that P34 is processed to a 32,000-Da polypeptide (P32) during seedling growth. This processing is the consequence of the removal of a highly basic amino-terminal decapeptide from P34 (9). The resulting polypeptide has an amino terminus which aligns with the amino terminus of the kiwi fruit protease actinidin (22). P34 appears to be a dimer in soybean seeds as the consequence of a disulfide bridge (9). The sequence data presented here suggest that the only cysteine available to constitute a disulfide bridge is cysteine 10 in the mature protein. This cysteine has no equivalent in any other previously sequenced thiol protease cDNAs. Cysteine 10 is removed from P34 during seedling growth, producing a 32,000-Da polypeptide termed P32 (9). Whether this may result in a dimer to monomer transition of P34 is currently under investigation. One of our present projects is directed at testing whether this processing results in the activation of proteolytic activity on the surface of oil bodies. Recent observations (37) of cathepsin expressed in bacteria and purified from inclusion bodies has shown that the processing of procathepsin to cathepsin is a consequence of self-catalysis. We are currently examining isolated soybean oil bodies to test if P34 may be converted to P32 *in vitro*. It is of great interest to us that the thiol protease most similar to P34 is the *Dictyostelium* protease (21), which apparently functions in developmentally regulated autodigestion. In this regard, we are also interested in the processing of proapolipoprotein A-II by a thiol protease (36). We have cloned the most abundant oil body protein, 24-kDa oleosin.¹ The sequence of this clone indicates that the polypeptide has some of the structural features typical of mammalian serum apolipoproteins. If a developmentally regulated protease activation does occur during the course of seedling growth, then this may provide a partial mechanism for the process of oil body mobilization and the digestion of oil body proteins.

REFERENCES

1. Yatsu, L. Y., and Jacks, T. J. (1972) *Plant Physiol.* **49**, 937–943
2. Slack, C. R., Bertaud, W. S., Shaw, B. D., Holland, R., Browne, J., and Wright, H. (1980) *Biochem. J.* **190**, 551–561
3. Huang, A. H. C. (1985) in *Modern Methods of Plant Analysis* (Linskins, H. F., and Jackson, J. F., eds) Vol. I, pp. 145–151, Springer-Verlag, Berlin
4. Herman, E. M. (1987) *Planta (Berl.)* **172**, 336–345
5. Vance, V. B., and Huang, A. H. C. (1987) *J. Biol. Chem.* **262**, 11275–11279
6. Vance, V. B., and Huang, A. H. C. (1988) *J. Biol. Chem.* **263**, 1476–1481
7. Murphy, D. J., Cummings, I., and Kang, A. S. (1989) *Biochem. J.* **258**, 285–293
8. Qu, R., and Huang, A. H. C. (1990) *J. Biol. Chem.* **265**, 2238–2243

³ D. L. Melroy and E. M. Herman, unpublished observations.

9. Herman, E. M., Melroy, D. L., and Buckhout, T. J. (1990) *Plant Physiol.*, in press
10. Holmes, D. S., and Quigley, M. (1981) *Anal. Biochem.* **114**, 193-197
11. Wadsworth, G. J., Redinbaugh, M. G., and Scandalios, J. G. (1988) *Anal. Biochem.* **172**, 279-283
12. Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583-7600
13. Kalinski, A., Chandra, G. R., and Muthukrishnan, S. (1986) *J. Biol. Chem.* **261**, 11393-11397
14. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13
15. Norrander, J., Kempe, T., and Messing, J. (1983) *Gene (Amst.)* **26**, 101-106
16. Sanger, F., Nickler, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467
17. Heidecker, G., and Messing, J. (1986) *Annu. Rev. Plant Physiol.* **37**, 439-466
18. Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., and Scheele, G. A. (1987) *EMBO J.* **6**, 43-48
19. Hopp, T. P., and Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3824-3828
20. Datta, S., and Firtel, R. A. (1987) *Mol. Cell. Biol.* **7**, 149-159
21. Pears, C. J., Mahbubam, H. M., and Williams, J. G. (1985) *Nucleic Acids Res.* **13**, 8853-8866
22. Praekelt, U. M., McKee, R. A., and Smith, H. (1988) *Plant Mol. Biol.* **10**, 193-202
23. Podivinsky, E., Forster, R. L., and Gardner, R. C. (1989) *Nucleic Acids Res.* **17**, 8363
24. Cohen, L. W., Coghlan, V. M., and Dihel, L. C. (1986) *Gene (Amst.)* **48**, 219-227
25. Rogers, J. C., Dean, D., and Heck, G. R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6512-6516
26. Faust, P. L., Kornfeld, S., and Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 4910-4914
27. Chan, S. J., San Segundo, B., McCormick, M. B., and Steiner, D. F. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 7721-7725
28. Portnoy, D. A., Erickson, A. H., Kochan, J., Ravetch, J. V., and Unkeless, J. (1986) *J. Biol. Chem.* **261**, 14697-14703
29. Ishidoh, K., Towatari, T., Imajoh, S., Kawasaki, H., Kominami, E., Katunuma, N., and Suzuki, K. (1987) *FEBS Lett.* **223**, 69-73
30. Schaffer, M. A., and Fischer, R. L. (1988) *Plant Physiol.* **87**, 431-436
31. Dietrich, R. A., Maslyar, D. J., Heupel, R. C., and Harada, J. J. (1989) *Plant Cell* **1**, 73-80
32. Baker, E. N. (1980) *J. Mol. Biol.* **141**, 441-484
33. Kamphuis, I. G., Drenth, J., and Baker, E. N. (1985) *J. Mol. Biol.* **182**, 317-329
34. Van der Wilden, W., Herman, E. M., and Chrispeels, M. J. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 428-432
35. Docherty, K., Carroll, R. J., and Steiner, D. F. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4613-4617
36. Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. M., and Strauss, A. W. (1984) *J. Biol. Chem.* **259**, 15556-15563
37. Smith, S. M., and Gottesman, M. M. (1989) *J. Biol. Chem.* **264**, 20487-20495