Construction and characterization of the soybean leaf metalloproteinase cDNA

Jhang Ho Pak, Chuan Y. Liu, Jun Huangpu, John S. Graham*

Department of Biological Sciences, Bowling Green State University, Bowling Green, OH 43403, USA

Received 18 November 1996; revised version received 30 January 1997

Abstract The cloning and analysis of a cDNA clone encoding the soybean metalloproteinase obtained by polymerase chain reaction (PCR) and the rapid amplification of cDNA ends (RACE) reaction are described. The cDNA was constructed from poly(A)$^+$ RNA isolated from 15–17 day old leaves. The deduced amino acid sequence of the cDNA reveals that the plant metalloproteinase is synthesized as a preproenzyme and the proenzyme form shares a structural motif, responsible for maintenance of inactive zymogen, with the matrix metalloproteinase (e.g. collagenase) family of enzymes from vertebrate origin. Northern and Western blot analysis demonstrated that the metalloproteinase transcript and protein are under a strict developmental program in that both are expressed only in leaf tissue and in a temporal fashion. The physiological function of the metalloproteinase still remains unclear although the data suggest that the enzyme is extracellular and a portion of the mature form of the enzyme is tightly bound to the cell wall.

Key words: Metalloproteinase; Soybean; cDNA; Developmental regulation

1. Introduction

The number of reports of zinc metalloproteinases in a variety of organisms has substantially increased over the last decade. These enzymes represent an ancient family of proteins with homologues in such diverse organisms as unicellular bacteria and multicellular invertebrate and vertebrate organisms. The active site zinc binding signature has been used to classify these enzymes into many families [1]. A major division of the zinc metalloproteinases has been termed the zincins. The crystal structure of metalloproteinases from the crayfish, snake venom and human fibroblast collagenase has prompted the designation of a new superfamily termed the 'metzincins' because of the presence of a methionine residue (referred to as a Met-turn) C-terminal to the active site zinc binding motif [2,3].

Our laboratory was the first to clearly demonstrate the presence of a metalloproteinase in higher plants [4,5]. A similar enzyme has been reported from the green alga, *Chlamydomonas* [6]. This enzyme is secreted and involved in cell wall degradation to facilitate fusion of gametes. Reviews have failed to include metalloproteinases of phototrophs [1,3] within any family of metal-requiring enzymes. Although members of the zinc metalloproteinases have been implicated in such processes as protein turnover, the action of tetanus and botulism toxins, embryogenesis development, cancer and arthritis, the function of the plant enzyme is still unclear.

In this report we characterize the cDNA and analyze the tissue-specific and temporal expression of the metalloproteinase in the soybean plant. The deduced primary structure of the plant enzyme reveals that it is synthesized as a preproenzyme and is a member of a superfamily of metalloproteinases, the metzincins.

2. Materials and methods

2.1. Plant material

Soybean plants (*Glycine max* var. Williams 82) were grown in environmental growth chambers under a 16 h photoperiod. Seeds were obtained from Mid-Wood Inc., Bowling Green, OH. Appropriate tissue was collected in liquid N_2 and stored at −70°C until use.

2.2. RNA isolation and Northern hybridization

Total RNA from 1 g of plant tissue was isolated using RNAgents Total RNA Isolation System as per the manufacturer's instructions (Promega, Madison, WI). Poly A$^+$ mRNA was isolated from total RNA of 17 day old leaves (post-emergence) using the poly AT Tract mRNA Isolation System as directed (Promega). This mRNA was used to synthesize double stranded cDNA using the Riboclone cDNA Synthesis System (Promega). For Northern blot analysis, total RNA was separated on a 1.2% agarose/formaldehyde gel [7]. The gel was dried on a gel dryer and hybridization performed in situ [8]. DNA probes were labeled to a specific activity of 1×10^8 cpm/μg with [α-^{32}P]CTP using the Amersham Multiprime kit as per the manufacturer's instructions. The probe used for all hybridizations was the 306 bp initial PCR product, encoding amino acids 134–304 of SMEP1. Hybridization and post hybridization washes were performed at 55°C as previously described [7].

2.3. Protein extraction and Western blot analysis

Soluble proteins were extracted from plant material by freezing the tissue in liquid N_2 and grinding to a fine powder with a mortar and pestle. The powdered tissue was extracted in a buffer consisting of 100 mM Tris-HCl, pH 8.5, 10 mM EDTA, 2 mM $CaCl_2$, 14 μM β-mercaptoethanol. Insoluble material was removed by centrifugation at $10\,000\times g$ for 15 min. Protein concentrations were estimated by the Bio-Rad assay system according to manufacturer's procedures. SDS-PAGE was performed [9] on leaf samples followed by electroblot transfer to nitrocellulose (MSI Laboratories). Western analysis was performed as previously described [10] using antiserum (1:1000 dilution) to SMEP1 [4].

2.4. PCR amplification and cloning

To generate a partial SMEP1 cDNA, two degenerate oligonucleotides derived from the primary amino acid sequence data of SMEP1 [5], were used in a polymerase chain reaction with ds cDNA synthesized with the Riboclone cDNA Synthesis System (Promega) to am-

*Corresponding author. Fax: (1) (419) 372-2024.
E-mail: jgraham@opie.bgsu.edu

Abbreviations: BSA, bovine serum albumin; pCMB, *p*-chloromercuribenzoic acid; SLAP, soybean leaf Azocoll-digesting proteinase; SMEP1, soybean metalloproteinase 1; IWF, intercellular wash fluid; nt, nucleotide; UTR, untranslated region

The nucleotide sequence data reported in this paper will appear in GenBank with the following accession number U63725.

PII S0014-5793(97)00141-5

plify a 306 bp fragment encoding the first 102 amino acids of the mature polypeptide.

The rapid amplification of cDNA ends (RACE) technique (GIBCO BRL) was used to construct the complete cDNA of SMEP1. RACE reactions were performed according to the supplier. Briefly, the 5′ RACE reaction (24 μl) was initiated using 1 μg of mRNA as template for first strand cDNA synthesis in the 5′ direction followed by dC-tailing. For the 3′ RACE reaction (20 μl) an oligo dT containing adapter primer was used for first strand cDNA synthesis. PCR was performed with 5 μl of either the 5′ or 3′ RACE reactions in a 100 μl reaction mixture containing 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 1.5 mM $MgCl_2$, 200 μM of each deoxynucleotide (dATP, dCTP, dGTP, and dTTP), 0.4 μM of each primer and 2.5 units of *Taq* Polymerase (Perkin-Elmer Cetus) in a 0.5 ml tube. Reaction mixtures were covered with 100 μl of light mineral oil, the tubes sealed, placed in an automated thermal cycler, heated to 94°C for 5 min and subjected to 30 step-cycles of 1 min at 94°C, 2 min at 60°C, and 3 min at 72°C. The final incubation at 72°C was extended for an extra 7 min and the reactions were cooled to 4°C.

PCR products were purified by extraction from 1% low melting agarose gels using the Magic PCR Purification kit (Promega). Purified PCR products were ligated into the TA plasmid (Invitrogen) and transformed into competent DH5a cells [7].

2.5. DNA sequencing and sequence analysis

The DNA sequence of both strands of denatured SMEP1 cDNAs in TA plasmids was determined by the dideoxy chain termination method [11] using [^{35}S]dATP and the Sequenase kit (US Biochemical). Commercial vector-specific and custom designed oligonucleotides were used as sequencing primers. The DNA and amino acid sequence was compiled and analyzed using MacVector software (International Biotechnologies Inc., New Haven, CT).

3. Results

3.1. Construction of SMEP1 cDNA

A strategy utilizing the PCR was designed to characterize the mRNA encoding the soybean leaf metalloproteinase, SMEP1. We previously reported on the purification of the mature form of SMEP1 [4] and the complete chemical se-

```
                                                                  M   T   L      3
AC CAC CCT TCC AAC TTT ATC TTT TCT CTA TAG CTC CGT CCA AAC ATG ACT CTC     53

 R   N   H   Q   E   L   L   V   A   L   A   T   L   Y   F   L   A   T     21
CGC AAC CAC CAA GAG CTC TTG GTT GCT CTT GCA ACT CTA TAT TTT CTT GCC ACC    107

 S   L   P   S   V   S   A ⇓ H   G   P   Y   A   W   D   G   E   A   T     39
TCA CTC CCT TCA GTT TCA GCT CAT GGC CCA TAT GCA TGG GAT GGG GAG GCC ACA    161

 Y   K   F   T   T   Y   H   P   G   Q   N   Y   K   G   L   S   N   V     57
TAT AAA TTC ACT ACT TAC CAC CCT GGC CAA AAC TAC AAA GGT TTA TCC AAT GTC    215

 K   N   Y   F   H   H   L   G   Y   I   P   N   A   P   H   F   D   D     75
AAA AAC TAC TTC CAC CAT CTC GGC TAC ATC CCC AAT GCA CCA CAC TTC GAC GAC    269

 N   F   D   D   T   L   V   S   A   I   K   T   Y   Q   K   N   Y   N     93
AAC TTC GAT GAC ACC CTC GTA TCT GCC ATC AAA ACC TAC CAA AAG AAT TAC AAC    323
     *
 L   N   V   T   G   K   F   D   I   N   T   L   K   Q   I   M   T   P    111
CTC AAC GTC ACC GGC AAG TTC GAC ATC AAC ACT CTT AAA CAA ATC ATG ACA CCC    377
                                                *
 R   C   G   V   P   D   I   I   I   N   T   N   K   T   T   S   F   G    129
CGG TGT GGC GTC CCC GAC ATA ATA ATC AAC ACA AAC AAA ACC ACA TCG TTT GGC    431

 M   I   S   D ↓ Y   T   F   F   K   D   M   P   R   W   Q   A   G   T    147
ATG ATC TCG GAC TAC ACG TTT TTC AAG GAC ATG CCG CGG TGG CAA GCT GGA ACC    485

 T   Q   L   T   Y   A   F   S   P   E   P   R   L   D   D   T   F   K    165
ACA CAA CTC ACC TAC GCT TTC TCC CCG GAG CCA AGA CTT GAT GAC ACT TTC AAA    539

 S   A   I   A   R   A   F   S   K   W   T   P   V   V   N   I   A   F    183
AGC GCG ATT GCA AGG GCC TTC AGC AAG TGG ACC CCA GTG GTG AAC ATC GCG TTC    593

 Q   E   T   T   S   Y   E   T   A   N   I   K   I   L   F   A   S   K    201
CAG GAG ACG ACG TCG TAT GAA ACA GCC AAC ATT AAG ATT CTT TTC GCG AGT AAG    647

 N   H   G   D   P   Y   P   F   D   G   P   G   G   I   L   G   H   A    219
AAC CAC GGT GAT CCG TAT CCT TTT GAT GGT CCA GGT GGG ATA TTG GGC CAT GCA    701

 F   A   P   T   D   G   R   C   H   F   D   A   D   E   Y   W   V   A    237
TTC GCT CCC ACT GAT GGG AGG TGC CAC TTT GAT GCC GAC GAA TAT TGG GTG GCG    755

 S   G   D   V   T   K   S   P   V   T   S   A   F   D   L   E   S   V    255
TCT GGC GAT GTC ACC AAA TCG CCG GTG ACA AGT GCG TTT GAC CTT GAA TCT GTG    809

 A   V   H   E   I   G   H   L   L   G   L   G   H   S   S   D   L   R    273
GCA GTG CAC GAG ATC GGG CAC TTG CTC GGA TTA GGC CAC TCG TCG GAC CTA AGA    863

 A.....I.....M.....Y   P   S   I   P   P   R   T   R   K   V   N   L   A   Q    291
GCC ATC ATG TAT CCT TCT ATA CCA CCT CGA ACT AGG AAG GTG AAT CTA GCG CAA    917

 D   D   I   D   G   I   R   K   L   Y   G   I   N   P                    305
GAT GAT ATA GAC GGG ATT CGA AAG CTC TAT GGT ATC AAC CCC TAA TTA TTA TAG    971
CTA GGT GTT GTT TGT ACT AGA ACT AGA GTT TTA TTT CGG TAA TTT GCA TTA AAG   1025
GAT TTT GAA AGT TAT CGA GGT GAA ATT ATT CTA ATT CAA GAA CAA CAA AAA TAG   1079
CAT GTA AGA TAT TGT ATT TTA GTT TTC TCT TAA TTA AGA ATA AAC CAG CTG ATA   1133
CTA GGT GAT GCA TAT ATG ATA ATG AAT AAA TGT GAC AAA TTT ACC TTA AAA AAA   1187
```

Fig. 1. Nucleotide and deduced amino acid sequence of the SMEP1 cDNA. Open (⇓) and solid (↓) arrows denote the putative signal peptide and propeptide cleavage sites, respectively. Asterisks denote potential *N*-glycosylation sites. Underlined sequence represents the HEXXH motif. Double underlined regions represent polyadenylation consensus sequences. A dotted line represents putative Met-turn. Italicized single amino acids letters represent putative second Zn^{2+} or Ca^{2+} ligands (see text).

SMEP1	D	I	N	T	L	K	Q	I	M	T	-	P	R	**C**	G	V	P	D	I
Collagenase	D	A	E	T	L	K	V	-	M	K	Q	P	R	**C**	G	V	P	D	V
Stromelysin	D	S	D	T	L	E	V	-	M	R	K	P	R	**C**	G	V	P	D	V
Gelatinase	D	Q	N	T	I	E	T	-	M	R	K	P	R	**C**	G	N	P	D	A
GLE(Chlamy)	D	S	V	T	G	Y	P	-	M	D	D	P	R	**C**	N	V	P	R	A
HE	D	A	D	T	A	E	L	-	L	S	T	P	R	**C**	G	V	P	D	V

Fig. 2. Sequence homology of the putative autolytic activation sites in the propolypeptide domains of SMEP1 and members of the MMP family. Collagenase from human fibroblast [22]; human stromelysin [23]; human gelatinase [21]; GLE (gamete lytic enzyme) from *Chlamydomonas reinhardtii* [6]; HE (hatching enzyme) from sea urchin [16]. Dashes denote gaps introduced to maximize sequence alignment. The cysteine residue that binds to the active site zinc ion is indicated by a bold letter.

quencing of the polypeptide [5]. This information was used to synthesize a set of degenerate oligonucleotide primers to amplify a portion of the SMEP1 mRNA. Double stranded cDNA, from Poly A^+ mRNA isolated from 17 day old leaves, was used as template in the *Taq* polymerase-catalyzed PCR. Amplification products were analyzed by agarose gel electrophoresis and subcloned into the TA plasmid. A 306 bp PCR product was obtained and subcloned into the TA plasmid and designated our initial construct, pJHP1. Nucleotide sequencing of JHP1 showed that the insert encoded amino acids 1–102 of the mature SMEP1 polypeptide as predicted. There were no ambiguities when the deduced amino acid sequence of the PCR product was compared to that of the chemically sequenced polypeptide.

The nucleotide sequence of pJHP1 provided the necessary information to synthesize two sets of oligonucleotides to serve as primers in the PCR to amplify nucleotide sequences 5′ and 3′ to those encoding the N-terminal region of the mature SMEP1 polypeptide. Amplification products were analyzed by agarose gel electrophoresis and subcloned into the TA plasmid. A 443 bp amplification product (pJHP2) was sequenced and found to encode N-terminal mature SMEP1 amino acids and 133 5′ amino acids not present in the mature form of the enzyme. Also, a 46 base 5′ UTR was present upstream of the initiation ATG codon. Another set of oligonucleotides, were used in conjunction with the oligo dT-adapter primer (Gibco BRL) to amplify sequences 3′ to the C-terminal region of SMEP1. Again, amplification products were analyzed by agarose gel electrophoresis and subcloned into the TA plasmid. A 474 bp insert (pJHP3) was found to encode C-terminal amino acids present in mature SMEP1 and a 93 base 3′ UTR. The nucleotide information from the various amplification products was used to synthesize the full length cDNA, pJHP4.

3.2. Primary structure of the soybean metalloproteinase

The nucleotide sequence and deduced primary structure of the full length cDNA and the preproprotein SMEP1 polypeptide, respectively, is presented in Fig. 1. The 1196 nucleotide sequence contains a single large open reading frame (915 nucleotides) that presumably begins at nucleotide position 44. The sequences surrounding this ATG, AAACATGAC, meet the requirement for the start codon consensus sequence in plants, AACAATGGC, where the most critical nucleotide at position −3 is conserved [12]. The open reading frame encodes a 305 amino acid polypeptide of calculated molecular mass 34 042 Da. The nascent SMEP1 polypeptide contains an N-terminal putative signal peptide with a central hydrophobic core and predicted cleavage site between amino acids Ala_{28} and His_{29} [13]. We believe that the signal peptide serves to direct the proteinase to the extracellular compartment of the leaf. We base this proposal on our previous finding that the specific activity of SMEP1 is highest within the IWF [4] and the observation that it remained bound to isolated leaf cell walls after high salt washes [14].

The remainder of the polypeptide represents proSMEP 1 of 31 000 Da. The suggestion that the 31 kDa protein represents the proenzyme is consistent with our previous finding [5] that the N-terminal amino acid of the mature SMEP1 is Tyr_{134} and the observation that members of the matrix metalloproteinase family are synthesized in a prepro form with the proenzyme representing an inactive zymogen [2,15]. Examination of the amino acid sequence of the pro region reveals a stretch of 17 amino acids at positions 101–117 (Fig. 2) that display sequence homology to the putative autolytic activation site in the propolypeptide domains of members of the matrix metalloproteinase family present in such diverse sources as *Chlamydomonas*, sea urchin and humans [6,15,16]. The conserved sequence Pro-Arg-Cys-Gly-Val-Pro has been shown to play a critical role in the conversion from inactive proenzyme to active enzyme with multiple cleavage sites within the region shown. Autolytic conversion has been shown to occur in several of these enzymes but has not been explicitly shown to occur in SMEP1 by experiment. The Cys_{113} residue is thought to serve as a ligand to the active site zinc ion in the inactive proenzyme (Fig. 2; [6,17]).

The deduced amino acid sequence of the SMEP1 cDNA also reveals the presence of a C-terminal proline residue not present in the mature SMEP1 polypeptide. We suspect that the hydrolysis of this residue by limited carboxypeptidase activity represents an additional post-translational modification of the nascent polypeptide. The nucleotide sequence contains two putative polyadenylation (AATAAA) signals beginning at nucleotides 1117 and 1157 (Fig. 1) with the latter motif 17 nt upstream of the poly(A) track and probably serving as the functional sequence.

3.3. Expression of the metalloproteinase during development

An earlier experiment designed to approximate, by Western blot analysis, the temporal expression of the metalloproteinase during leaf development has been previously documented [4,18] and shows that EDTA-sensitive Azocoll-digesting activity (SMEP1) is detectable only in mature (> 10 day old) leaves. The accumulation of the mature (active) SMEP1 correlates with the previous activity measurements in that the appearance of an immunoreactive band begins to accumulate in 7–10 day old leaves and remains fairly constant during the

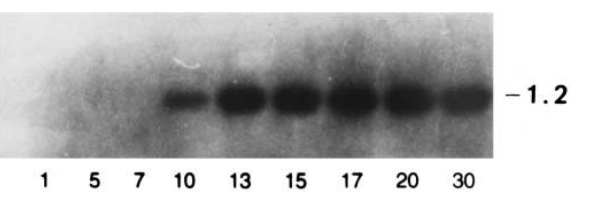

Fig. 3. Accumulation of SMEP1 mRNA during leaf development. Northern blot analysis of 10 μg of total RNA extracted from leaves and hybridized with ^{32}P-labeled SMEP1 cDNA insert. The ages (in days) of primary leaves used in the experiment are below each lane. Molecular size marker is indicated on the right (in kilobase pairs).

30 day time course. As shown by Northern blot analysis in Fig. 3, steady state levels of the 1.2 kb SMEP1 transcripts were only detected in leaves beginning at approx. 10 days old and remain at fairly constant levels for the remainder of the period examined. Thus, the relative SMEP1 transcript levels are directly correlated with the levels of the SMEP1 polypeptide (18). The data are consistent with the proposal that preproSMEP1 and proSMEP1 are short-lived intermediates which are not detectable in leaf tissue by Western analysis.

We previously examined the expression of EDTA-sensitive Azocoll-digesting activity in various tissues of a 37 day old soybean plant and showed that activity was limited to mature leaf tissue [4]. Fig. 4A represents a Western blot of soluble proteins extracted from tissues of a 21 day old plant and shows that SMEP1 is present only in primary, first and second trifoliate leaves all of which are greater than 10 days old (lanes 1–3), but is absent in the fourth trifoliate leaf which has just emerged (lane 4). Stem and root tissue older than 10 days (lanes 5–7) failed to show any significant SMEP1 accumulation. Northern blot analysis (Fig. 4B) of total RNA isolated from identical samples used in Fig. 4A showed that SMEP1 transcripts are only detectable in those tissues that contain the polypeptide. These data show that SMEP1 accumulation is regulated at the transcriptional level and in a tissue-specific manner. However, the possibility that regulation may be exerted at the level of RNA stability can not be ruled out. We examined cell wall bound proteins for the presence of SMEP1 to determine if the polypeptide present in leaves had any affinity for this cellular fraction. The data show that at least a portion of SMEP1 does bind strongly to cell walls [18].

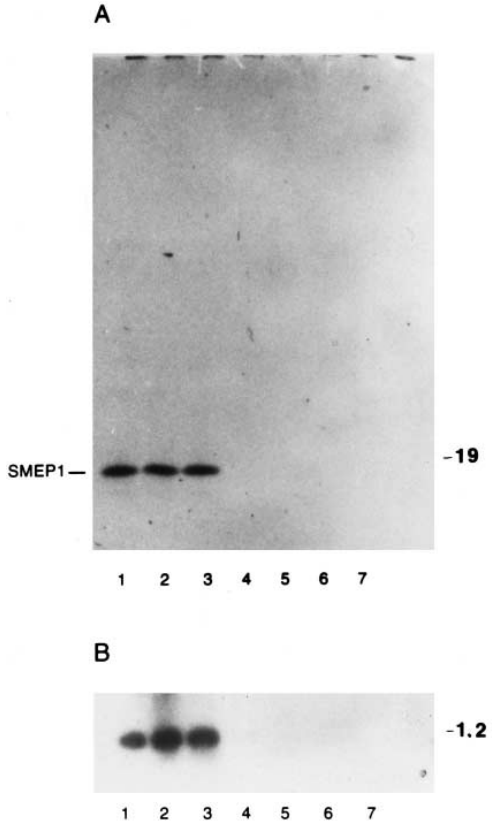

Fig. 4. Tissue specific and temporal expression of SMEP1. Various tissues were harvested from a 21 day old soybean plant. (A) Western blot analysis of 10 μg of total soluble proteins extracted from different tissues were probed with SMEP1 antisera (1:1000 dilution). Molecular mass is indicated on right (in kDa). (B) Northern blot analysis of 10 μg of total RNA extracted from different tissues were hybridized with ^{32}P-labeled SMEP1 cDNA insert. Lanes: 1, primary leaves; 2, first trifoliate leaves; 3, second trifoliate leaves; 4, third trifoliate leaves (buds); 5, upper stem; 6, lower stem; 7, root. Molecular size marker is indicated on the right (in kilobase pairs).

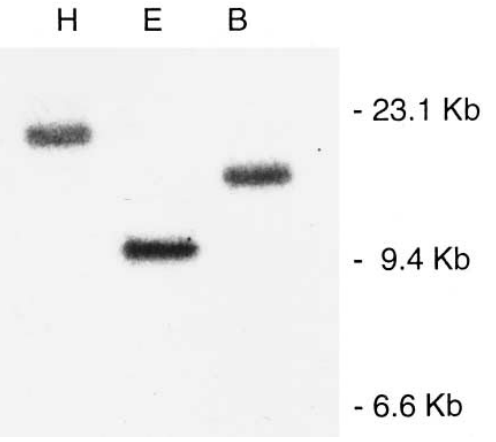

Fig. 5. Southern blot analysis of soybean genomic DNA. Soybean nuclear DNA (10 μg) was digested to completion with various restriction enzymes and probed with ^{32}P-labeled SMEP1 cDNA insert. H, *Hin*dIII; E, *Eco*RI; B, *Bam*HI.

3.4. Genomic organization of the metalloproteinase gene

To determine the complexity of the *smep1* gene within the soybean genome the pJHP 4 clone was used as a probe to hybridize with genomic DNA restricted to completion with *Eco*RI, *Hin*dIII, or *Bam*HI. There are none of these restriction sites are within the full length SMEP1 cDNA. A single strong hybridizing band is observed on the Southern blot with each enzyme digestion (Fig. 5) suggesting that *smep1* is a single copy gene. An experiment using the PCR with genomic DNA to detect for the presence of intronic regions within the *smep1* gene showed that the gene lacks introns due to the production of a 1.2 kb PCR product (data not shown).

4. Discussion

It was previously described that a Zn^{+2} proteinase was purified from the mature leaves of the soybean plant and was designated SMEP1 (soybean metalloproteinase 1). Chemical sequencing of the entire mature form of the SMEP1 polypeptide revealed that the plant enzyme shared several biochemical features with the matrix metalloproteinase family of enzymes from vertebrate origin [5]. These features include a conserved zinc binding motif, an identical cleavage site within a synthetic collagenase-specific substrate, and similar inhibition constants when complexed with the tissue inhibitor of metalloproteinases [15].

In order to gain further information about the structure and regulation of SMEP1, a full length SMEP1 cDNA was obtained by an alternative approach which utilized primary amino acid sequence information and the PCR to construct several partial cDNA clones and finally a full-length clone. The characterization of the cloned cDNA of SMEP1 provides

the first reported sequence of a preproenzyme member of the matrix metalloproteinase family from higher plants. A cDNA of 1196 nucleotides contains a single large open reading frame (915 nt), and 5′- and 3′-untranslated regions, the latter followed by a poly A tail. The predicted primary structure of SMEP1 is composed of three domains; a 28 amino acid endoplasmic reticulum signal sequence which presumably serves to direct the secretion of the protein to the extracellular space, a 105 amino acid propolypeptide that may be responsible for repressing the activity of SMEP1, and a 172 amino acid mature polypeptide that is the active form of SMEP1. Of particular interest is the conserved sequence motif (Pro-Arg-Cys-Gly-Val-Pro) present within the pro region of SMEP1 (Fig. 2) that is likely critical for conversion of MMP zymogens into active enzymes. By analogy with MMP family members [17,20], the Cys_{113} residue in the latent enzyme may be coordinated to the active site zinc ion, thus preventing activity. The dissociation of Cys_{113} from the zinc ion may result in the autolytic cleavage of the zymogen and conversion of the latent proenzyme to the active form. It should be noted that the location of the autolytic motif is 17 amino acids N-terminal to the presumed cleavage site in proSMEP1 while the motif is only several amino acids N-terminal to putative cleavage sites in previously studied members of the MMP family (Fig. 1; [6,19,21–23]). The preproform of SMEP1 has a predicted molecular mass of 35 kDa. The pro enzyme has a predicted molecular mass of 31 kDa. This value could be an underestimate since there are two potential asparagine linked glycosylation sites within this region.

In addition to the catalytic zinc binding domain, a putative calcium binding site is found in the primary structure of SMEP1. The calcium ion appears to stabilize the tertiary structure of collagenase [19]. The amino acid sequence, Asp-Xaa-Asp-Xaa-Asp-Gly-$(Xaa)_2$-Asp-$(Xaa)_2$-Asp, is known to provide ligands to the calcium ion, where the five aspartates can be coordinated through oxygen atoms. The glycine residue is also conserved in all calcium binding proteins [24]. Of the five aspartic acid residues involved in calcium binding, three are found to be conserved in the C-terminal region (Fig. 1, pos. 293–296) of the mature SMEP1 protein including a conserved glycine.

A second zinc ion binding site as well as a calcium ion binding are known to be important for maintaining the structural integrity of human fibroblast collagenase [19]. Four potential second zinc binding site ligands are also present in the SMEP1 polypeptide (Fig. 1, His_{203}, Asp_{205}, His_{218}, and His_{228}). Sequence alignment [5] of the catalytic domains of the MMPs has shown that these four ligands are completely conserved within the enzyme family, suggesting that the presence of two zinc ions is a common structural feature of the MMP family of enzymes.

Finally, an additional site of interest includes a conserved 'Met-turn' sequence (Ala-Leu-Met-Tyr) C-terminal to the active site zinc binding motif. The X-ray crystal structure of metalloproteinases from the crayfish, snake venom [2], and human fibroblast collagenase [19] have revealed the presence of a short loop region that places the sulfur atom of a conserved methionine residue near (5 Å) the active site zinc ion which serves to provide structural integrity. A highly homologous sequence five residues behind the third histidine ligand (position 268) is present in the mature SMEP1 protein (Ala-Ile-Met-Tyr, pos. 274–277). It is thus appropriate that the SMEP1 protein be categorized into the new superfamily of zinc peptidases termed the metzincins [1,3].

The expression of the MMPs in mammalian cells is under strict transcriptional control during development and connective tissue remodeling [15]. Two recent reports expand the MMP family to include members isolated from non-vertebrate sources. In the sea urchin embryo, a collagenase-like hatching enzyme is secreted at the blastula stage of development to degrade the protective envelope. The expression of this enzyme is developmentally regulated and is under transcriptional control [16]. Another metalloproteinase was found in the green algae, *Chlamydomonas reinhardtii*. The enzyme has been named the gamete lytic enzyme and mediates digestion of proteins within the cell wall of two mating-type gametes prior to cell fusion. The transcript level of this proteinase is increased during growth and gametic differentiation [6].

The data from the Northern and Western blot analyses indicate that the expression of SMEP1 is tightly regulated in a developmental program and in a tissue specific manner. The temporal pattern of protein accumulation parallels that of the mRNA, indicating that the expression of the SMEP1 gene appears to be under transcriptional control. SMEP1 begins to accumulate approx. 10 days after leaf emergence and remains at steady levels until leaves become senescent. Also, the presence of SMEP1 is limited to mature leaf tissue and at least a portion of the enzyme is bound to the cell wall [18].

MMPs in vertebrates are known to be coded by a multigene family [25]. In contrast to collagenase genes, Southern blot analysis of soybean genomic DNA demonstrates that the MMP relative (SMEP1) in a higher plant exists as a single copy gene. It would be interesting to know to what extent homologous counterparts(s) are present in other plants. We have already cloned and sequenced SMEP1 from *Arabidopsis thaliana* (data not shown) and found 52% amino acid identity between the higher plant enzymes.

The gene structure has been determined for a number of the MMPs: human MMP-1 and MMP-2 [26], rat MMP-3 and MMP-10 [27], and rabbit MMP-1 [28]. Unlike the *sempl* gene which lacks introns, all the aforementioned genes consist of 10 exons and 9 introns except MMP-2 which has three additional exons corresponding to the fibronectin-like domain.

Controlled production and activity of MMPs play an important role for normal development of tissue, extracellular matrix architecture, and cell migration. These enzymes can be involved in pathological processes such as joint destruction in rheumatoid and osteoarthritis, tumor invasion, and periodontitis [29]. Since SMEP1 is present only in mature leaves, the enzyme may play an important role such as tissue modeling, which must occur during leaf expansion. It is believed that the formation of intercellular space in the mature leaf requires the partial separation of cells following wall breakdown [30]. It could be argued that the accumulation of SMEP1 occurs during late stages of the expansion process, when most of the expansion of leaves has already occurred, and thus may not be involved in these events. The fact of the matter is that the physiological function of this proteinase in leaf tissue remains unclear. Alternatively, an attractive, albeit speculative hypothesis, is that the enzyme functions to serve a defensive role in plant leaves. The enzyme may, upon feeding by insects, serve to digest collagen-like proteins within the midgut lining of insect pests, thus disrupting the normal digestive physiology

of the pest. Also, the possibility that SMEP1 may possess antimicrobial activity cannot be ruled out at this time.

The SMEP1 transcript represents an interesting gene for further study because its expression is synchronized with the mature stages of leaf development. The precise nature of the developmental cue(s) regulating SMEP1 transcript levels remains to be determined. Promoter analysis may provide information about the genetic mechanism of this mode of gene expression. Further studies will also be required to determine in which cell type, or at which stage of cell development or differentiation, SMEP1 accumulates. In this respect, in situ hybridization studies to determine ECM location will be important. Finally, the identification of endogenous physiological substrate(s) will aid in our attempt to understand the function of SMEP1 in vivo.

Acknowledgements: This work was supported by grants from Glaxo Research Institute, Research Triangle Park, NC, and the US Department of Agriculture Competitive Research Grants Office No. 9203124 to (J.S.G.).

References

[1] Hooper, N.M. (1994) FEBS Lett. 354, 1–6.
[2] Bode, W., Franz-Xaver, G.-R. and Stockler, W. (1993) FEBS Lett. 331, 134–140.
[3] Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F.-X., McKay, D.B. and Bode W (1995) Protein Sci. 4, 823–840.
[4] Graham, J.S., Xiong, J. and Gillikin, J.W. (1991) Plant Physiol 97, 786–792.
[5] McGeehan, G., Burkhart, W., Anderegg, R., Becherer, D., Gillikin, J.W. and Graham, J.S. (1992) Plant Physiol. 99, 1179–1183.
[6] Kinoshita, T., Fukuzawa, H., Shimada, T., Saito, T. and Matsuda Y. (1992) Proc. Natl. Acad. Sci. USA 89, 4693–4697.
[7] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[8] Ehtesham, N.Z. and Hasnain, S.E. (1991) Biotech 11, 718–721.
[9] Laemmli, U.K. (1970) Nature 227, 680–685.
[10] Towbin, H., Staehelin, T. and Gordon J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350–4354.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) EMBO J. 6, 43–48.
[13] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[14] Huangpu, J., Pak, J.H., Burkhart, W., Graham, M.C., Rickle, S.A., Liu, C.Y. and Graham, J.S. (1996) Plant Physiol. 112, 862.
[15] Woessner, J.F. Jr. (1991) FASEB J. 5, 2145–2154.
[16] Lepage, T. and Gache, C (1990) EMBO J. 9, 3003–3012.
[17] Van Wart, H.E. and Birkedal-Hansen, H. (1990) Proc. Natl. Acad. Sci. USA 87, 5578–5582.
[18] Huangpu, J., Pak, J.H., Graham, M.C., Rickle, S.A. and Graham, J.S. (1996) Biochem. Biophys. Res. Commun. 228, 1–6.
[19] Lovejoy, B., Cleasby, A., Hassell, A.M., Longly, K., Luther, M.A., Weigi, D., McGeehan, G., McElroy, A.B., Drewry, D., Lambert, M.H. and Jordan, S.R. (1994) Science 263, 375–377.
[20] Springman, E.R, Angleton, E.L., Birkedal-Hansen, H. and Van Wart, H.E. (1990) Proc. Natl. Acad. Sci. USA 87, 364–368.
[21] Collier, I.E., Wilhelm, S.M., Eisen, A.Z., Marmer, B.L., Grant., G.A., Seltzer, J.L., Kronberger, A., He, C., Bauer, E.A. and Goldberg, G.I. (1988) J. Biol. Chem. 263, 6679–6587.
[22] Goldberg, G.I., Wilhelm, S.M., Kronberger, A., Bauer, E.A., Grant, G.A. and Eisen, A.Z. (1986) J. Biol. Chem. 261, 6600–6605.
[23] Whitham, S.E., Murphy, G., Angel, P., Rahmsdorf, H.-J., Smith, B.J., Lyons, A., Harris, T.J.R., Reynolds, J.J. and Herrlich, P. (1986) Biochem. J. 240, 913–916.
[24] Lawler, J. and Hynes R.O. (1986) J. Cell. Biol. 103, 1635–1648.
[25] Murphy, G.P., Murphy, G. and Reynolds, J.J. (1991) FEBS Lett. 289, 4–7.
[26] Huhtala, P., Chow, L.T. and Tryggvason, K. (1990) J. Biol. Chem. 265, 11077–11082.
[27] Breathnach, R., Matrisian, L.M., Gesnel, M.C., Staub, A. and Leroy, P. (1987) Nucleic Acid Res. 15, 1139–1151.
[28] Fini, M.E., Plucinska, I.M., Mayer, A.S., Gross, R.H. and Brinckerhoff, C.E. (1987) Biochemistry 26, 6156–6165.
[29] Matrisian, L.M (1990) Trends Genet. 6, 121–125.
[30] Dale, J.E. (1988) Annu. Rev. Plant Physiol. Plant. Mol. Biol 39, 267–295.