Journal of Plant Biology, September 2004, 47(3) : 216-220

Isolation and Characterization of a Root Nodule-Specific Cysteine Proteinase cDNA from Soybean

Chang Jae Oh, Hyoungseok Lee, Ho Bang Kim, and Chung Sun An*
School of Biological Sciences, Seoul National University, Seoul 151-742, Korea

We have determined that a nodule-specific cDNA clone (*GmCysP1*), obtained from a soybean root nodule-specific EST pool, encodes cysteine proteinase. Its amino acid sequence homology, as well as the conservation of typical motifs and amino acid residues involved in active site formation, shows that GmCysP1 can be classified as a legumain (C13) family cysteine proteinase, belonging to clan CD. Moreover, based on its expression patterns, *GmCysP1* is a nodule-specific cysteine proteinase gene that is possibly associated with nodule development or senescence. Our genomic Southern analysis also suggests that *GmCysP1* is a member of a multigene family. Therefore, we propose that GmCysP1 is the first to be identified as a nodule-specific and senescence-related cysteine proteinase that belongs to the legumain family from soybean.

Keywords: cysteine proteinase, legumain family, nodule development, senescence, soybean

Proteolytic activity is ubiquitous in biological systems. In plants, protein degradation is essential for growth, development, and environmental responses. Selective proteolysis provides a mechanism for protein turnover and reutilization of nitrogen for maintaining cellular homeostasis and growth. This complex process involves many enzymes. Among them, the cysteine proteinase enzymes (EC 3.3.22) are the most common endopeptidases studied because they apparently play a central role in a wide range of proteolytic functions (Ho et al., 2000). Cysteine proteinases are peptidases in which the nucleophile is the sulfhydryl group of a cysteine residue. Their peptidase domain is responsible for peptide bond hydrolysis. The peptidases can be divided into six clans (i.e., proteins that are evolutionarily related), and further subdivided into 45 families based on the architectures of their catalytic dyad or triad (Barrett and Rawlings, 2001).

Cysteine proteinases play a major role in mobilizing storage proteins during seed germination to supply young seedlings with the reduced nitrogen needed for growth (Becker et al., 1994). In addition, they have been implicated in plant developmental and physiological processes, such as leaf abscission (Wittenbach et al., 1982), and senescence (Hensel et al., 1993), as well as in responses to environmental stresses, e.g., dehydration in *Arabidopsis* and pea (Guerrero et al., 1990; Koizumi et al., 1993), mechanical wounding in tobacco (Linthorst et al., 1993), and low temperature in tomato (Schaffer and Fischer, 1988). In some species, including Chinese milk vetch (*Astragalus sinicus*) and pea (*Pisum sativum*), cysteine proteinases belonging to papain family have been separated into the nodule-specific group. And especially, that of Chinese milk vetch was expected to mediate protein turnover in the symbiosome compartment, based on its cellular localization (Kardailsky and Brewin, 1996; Naito et al., 2000; Vincent and Brewin, 2000). In the case of soybean, most of the cysteine proteinases that have been isolated from seeds are involved in the degradation of storage proteins, such as β-conglycinin, during seed development or germination. These enzymes also are members of the papain family (C1) in clan CA (Nong et al., 1995; Seo et al., 2001).

In this study, we isolated and characterized a cDNA clone encoding cysteine proteinase from a root nodule-specific EST pool for soybean. Our objective was to evaluate changes in its transcript level during nodule development.

MATERIALS AND METHODS

Plant Material and Bacterial Strain

To nodulate plants of soybean (*Glycine max* L. cv Baektae), we inoculated sprouted seeds with

*Corresponding author; fax +82-2-872-6881
e-mail ancs@snu.ac.kr

Bradyrhizobium japonicum strain USDA 110, which had previously been cultured in YEM (Vincent, 1970) and suspended in a BNM (Ehrhardt et al., 1992). The inoculated seedlings were then reared in a growth chamber at 27°C under a 16 h photoperiod, and were watered with 0.5X BNM. To analyze expression patterns during their development, nodules were harvested at various stages and stored in liquid nitrogen at –80°C.

Cloning and Sequence Analysis

The cDNA clone that encodes the cysteine proteinase gene was isolated from a soybean root nodule-specific EST pool (Lee et al., 2004a) and sequenced using ABI PRISM 3700TM (Applied Biosystems, USA). The resultant sequences were analyzed using the basic local alignment search tool (BLAST) program of NCBI (Altschul et al., 1990), and the ClustalW multiple alignment program (Higgins et al., 1992).

Genomic DNA Southern Blot Analysis

Genomic DNA was isolated from soybean leaves as described by Doyle and Doyle (1990). Ten micrograms of genomic DNA was digested with restriction enzymes (*Eco*RI, *Hin*dIII, or *Xba*I), fractionated on a 0.8% agarose gel, and transferred to a nylon membrane (Amersham, UK) by the capillary method (Sambrook et al., 1989). Hybridization was performed at 62°C for 24 h, using ^{32}P random-primer-labeled full-length cDNA as a probe. The hybridized blot was then washed at 62°C, with stringency gradually increasing to 0.5 × SSC / 0.1% SDS, before being exposed to X-ray film (Fuji, Japan).

RNA Gel Blot Analysis

Total RNA was extracted from various tissues as previously described (Uhde-Stone et al., 2003). Ten micrograms of total RNA was separated on a 1.2% formaldehyde agarose gel and transferred to a nylon membrane. Hybridization and membrane washing were performed under the same conditions mentioned above.

Semi-Quantitative Reverse Transcription-PCR

After treatment with RNase-free DNase (Promega, USA), 2 μg of total RNA was used as template for synthesizing first-strand cDNA with oligo dT. The following PCR primers were designed to specifically amplify the *GmCysP1*: GmCysP1F, 5'-AATCACAGAATTTTCG-CAGT-3'; *GmCysP1*R, 5'-GATGCCTGTAGTTTCCG-TAG-3'. In addition, the primers for amplifying actin -- GmactinF, 5'-CAGCATGAAAATCAAGGTGGT-3' and GmactinR, 5'-AGGGGACCTAACGGAGAAACT-3' -- were used as the qualitative control (Lee et al., 2004b). Each PCR reaction (25 cycles total) included initial denaturation at 96°C for 5 min, followed by 96°C for 50 sec, 52°C for 30 sec and 72°C for 45 sec, then ending with 5 min of final elongation at 72°C. Amplified PCR products were separated on a 1.2% agarose gel.

RESULTS AND DISCUSSION

Isolation and Characterization of the cDNA Clone Encoding Cysteine Proteinase

We have isolated a cDNA clone showing high sequence homology with the previously described cysteine proteinase genes from our soybean root nodule-specific EST pool. This new clone is named *GmCysP1* (*Glycine max Cysteine Proteinase 1*). Nucleotide sequence analysis revealed that *GmCysP1* is identical as the clone *GmVPE* (*Glycine max* vacuolar processing enzyme; GenBank accession number D28876) that was reported by Shimada et al. (1994). GmVPE was originally isolated from soybean seeds and is presumed to be converted from proproteins to its corresponding mature form. The *GmCysP1* gene encodes 495 amino acids and the predicted protein has a molecular weight of 55.2 kDa and a pI of 5.78. Its deduced amino acid sequence shares high similarities with cysteine proteinases belonging to the legumain (C13) family in the clan CD defined in the MEROPS database, ranging from 95.57% (with *Phaseolus vulgaris*) to 81.65% (*Arabidopsis thaliana*) (Fig. 1) (Rawlings et al., 2004). However, when compared with other species or soybean cysteine proteinases that belong to the papain (C1) family, GmCysP1 shows only ~40% sequence similarities. Before our study, nodule-expressed cysteine proteinases had been reported in pea, alfalfa, *Medicago truncatula*, and Chinese milk vetch, all of which are grouped into the papain family in the clan CA. Therefore, *GmCysP1* is remarkable because it is the first nodule-expressed cysteine proteinase gene known to belong to the legumain family.

In GmCysP1, a putative transmembrane domain is located in the N-terminal region and five motifs for hemoglobinase (i.e., another name for the legumain family) are also detected. In particular, motifs 4 and 5

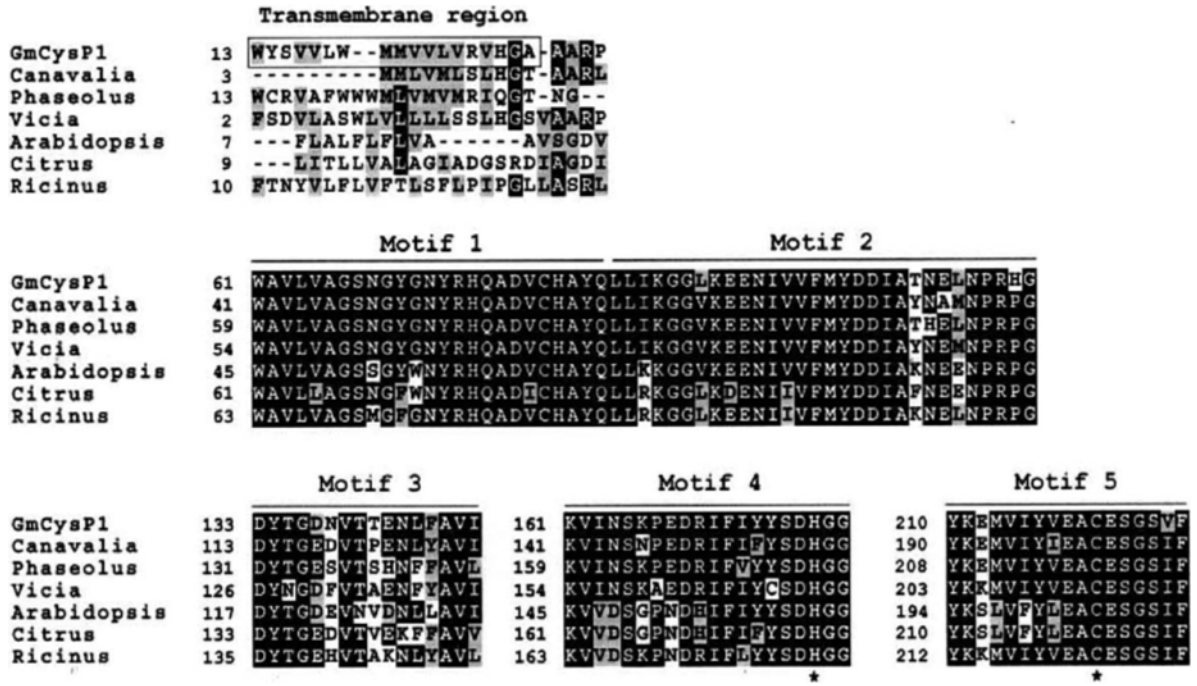

Figure 1. Multiple sequence alignment of GmCysP1 and previously reported cysteine proteinases belonging to the legumain family, using ClustalW. Putative transmembrane region of GmCysP1 is indicated by box and five hemoglobinase motifs are indicated above corresponding sequences. Asterisks (*) represent conserved residues that form an active site of cysteine proteinases Sources and GenBank accession numbers of sequences used in alignment are: *P. vulgaris* (Z99957), *Canavalia ensiformis* (D31787), *Vicia narbonensis* (Z99174), *Ricinus communis* (D17401), *A. thaliana* (NM_128154), *Citrus sinensis* (Z47793).

contain the well-conserved residues (His178, Cys_{220}) that form putative active site residues for members of that family (Chen et al., 1998). Based on these results, we propose that *GmCysP1* gene encodes a cysteine proteinase typical of the legumain family.

Expression Pattern of *GmCysP1*

We performed northern hybridization and semi-quantitative RT-PCR to investigate the expression patterns of *GmCysP1* in various soybean organs. This clone is entirely nodule-specific, with expression levels gradually increasing in proportion to the extent of nodule development (Fig. 2A). This behavior is similar to that of other cysteine proteinases linked with senescence. Transcript accumulation during nodule development was confirmed by northern hybridization (Fig. 2B). In the nodules of annual legumes such as soybean, the symbiosome membranes of senescing nodules fuse to form vacuoles containing the debris of digested bacteroids, thereby indicating the possible recovery of nitrogenous compounds (Mellor, 1989; Roth and Stacey, 1989). Consistent with those previous reports, the fact that *GmCysP1* is both nodule-specific and associated with gradually increased expression patterns suggests that GmCysP1 may be a membrane protein with functional roles such as the digestion of structural compounds of symbiosome in senescing nodules. Cellular localization analysis of GmCysP1 would provide further evidence to support this hypothesis.

GmCysP1 in the Genome of *G. max*

In many plant species, including *Arabidopsis* and maize, cysteine proteinase genes are members of a multigene family (Pechan et al., 1999). To determine the copy number of *GmCysP1* gene in the soybean genome, we performed Southern blot analysis using a full-length insert of the *GmCysP1* clone as probe. When genomic DNA was digested with three restriction enzymes, multiple hybridization signals were detected in all three lanes. Although *GmCysP1* has three *Hin*dIII restriction sites in its sequence, the hybridization signals were more numerous than expected (Fig. 3). Therefore, those multiple signals demonstrate that *GmCysP1* is a member of a multigene family. Further investigation, e.g., promoter analysis of the *GmCysP1* genomic clone and *in situ* hybridization in the nodules, will facilitate a better

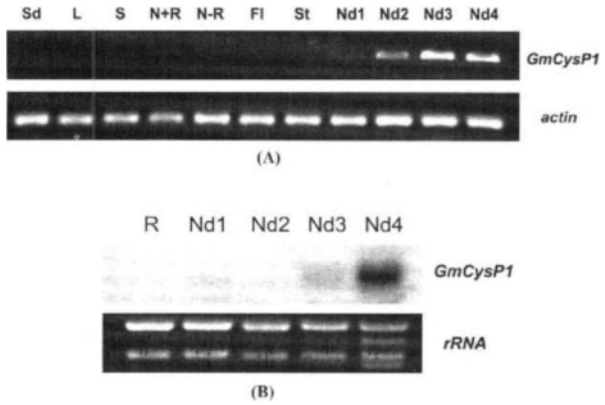

Figure 2. Expression pattern of *GmCysP1* gene. **A.** Semi-quantitative RT-PCR analysis for different tissues, with actin transcripts amplified as qualitative control. Sd, seedling; L, leaves; S, stem; N+R, nitrogen-supplied soil-grown root; N-R, nitrogen-deficient soil-grown root; Fl, flower; St, shoot tip; Nd1, root at 4 DAI (days after inoculation); Nd2, nodule at 2 WAI (weeks after inoculation); Nd3, nodule at 5 WAI; Nd4, nodule at 8 WAI. **B.** RNA gel blot analysis of *GmCysP1* during nodule developmental stages. R, uninoculated root.

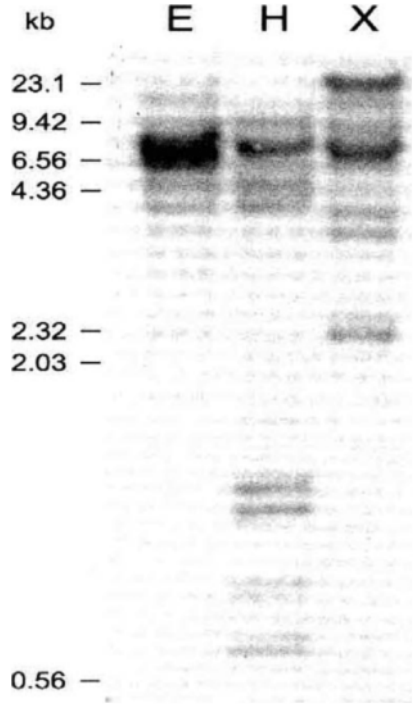

Figure 3. Genomic Southern analysis of *GmCysP1* gene. DNA blot was probed with ^{32}P-labeled full-length *GmCysP1* clone. Molecular mass markers are shown in kb on left of autoradiogram. E, *Eco*RI; H, *Hin*dIII; X, *Xba*I.

understanding of the function and regulation of cysteine proteinase genes in the soybean.

ACKNOWLEDGMENTS

This work was supported by a grant from the Crop Functional Genomics Center (Project No. M101KG010001-03K0701-01810, Korea) to C. S. An. C. J. Oh and H. Lee were supported by the Brain Korea 21 Project of the Ministry of Education, Korea.

Received March 4, 2004; accepted June 8, 2004.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Barrett AJ, Rawlings ND (2001) Evolutionary lines of cysteine peptidases. Biol Chem 382: 727-733

Becker C, Fisher J, Nong VH, Munitz K (1994) PCR cloning and expression analysis of cDNAs encoding cysteine proteinases from germinating seeds of *Vicia sativa* L. Plant Mol Biol 26: 1207-1212

Chen JM, Rawlings ND, Stevens RA, Barrett AJ (1998) Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. FEBS Lett 441: 361-365

Doyle JJ, Doyle JI (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15

Ehrhardt DW, Atkinson EM, Long SR (1992) Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* Nod factors. Science 256: 998-1000

Guerrero FD, Jones JT, Mullet JE (1990) Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes. Plant Mol Biol 15: 11-26

Hensel LL, Grbic V, Baumgarten DA, Bleecker AB (1993) Developmental and age-related processes that influence the longevity and senescence of photosynthetic tissues in Arabidopsis. Plant Cell 5: 553-564

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal V: Improved software for multiple sequence alignment. Comput Appl Biosci 8: 189-191

Ho SL, Tong WF, Yu SM (2000) Multiple mode regulation of a cysteine proteinase gene expression in rice. Plant Physiol 122: 57-66

Kardailsky IV, Brewin NJ (1996) Expression of cysteine protease genes in pea nodule development and senescence. Mol Plant Microbe Interact 9: 689-695

Koizumi M, Yamaguchi-Shinozaki K, Tsuji H, Shinozaki K (1993) Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. Gene 129: 175-82

Lee H, Hur CG, Oh CJ, Kim HB, Park S, An CS (2004a) Analysis of the root nodule-enhanced transcriptome from soybean. Mol Cells 18: 53-62

Lee H, Kim H, An CS (2004b) Cloning and expression anal-

ysis of 2-on-2 hemoglobin from soybean. J Plant Biol 47: 92-98

Linthorst HJ, van der Does C, Brederode FT, Bol JF (1993) Circadian expression and induction by wounding of tobacco genes for cysteine proteinase. Plant Mol Biol 21: 685-694

Mellor RB (1989) Bacteroids in the Rhizobium-legume symbiosis inhabit a plant internal lytic compartment: Implications for other microbial endosymbioses. J Exp Bot 40: 831-839

Naito Y, Fujie M, Usami S, Murooka Y, Yamada T (2000) The involvement of a cysteine proteinase in the nodule development in Chinese milk vetch infected with *Mesorhizobium huakuii* subsp. rengei. Plant Physiol 124: 1087-1096

Nong VH, Becker C, Muntz K (1995) cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. Biochim Biophys Acta 1261: 435-438

Pechan T, Jiang B, Steckler D, Ye L, Lin L, Luthe DS, Williams WP (1999) Characterization of three distinct cDNA clones encoding cysteine proteinases from maize (*Zea mays* L.) callus. Plant Mol Biol 40: 111-119

Rawlings ND, Tolle DP, Barrett AJ (2004) MEROPS: The peptidase database. Nucleic Acids Res 32 Database issue: D160-164

Roth LE, Stacey G (1989) Bacterium release into host cells of nitrogen-fixing soybean nodules: The symbiosome membrane comes from three sources. Eur J Cell Biol 46: 12-23

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York

Schaffer MA, Fischer RL (1988) Analysis of mRNA that accumulate in response to low temperature identifies a thiol protease gene in tomato. Plant Physiol 87: 431-436

Seo S, Tan-Wilson A, Wilson KA (2001) Protease C2, a cysteine endopeptidase involved in the continuing mobilization of soybean beta-conglycinin seed proteins. Biochim Biophys Acta 1545: 192-206

Shimada T, Hiraiwa N, Nishimura M, Hara-Nishimura I (1994) Vacuolar processing enzyme of soybean that converts proproteins to the corresponding mature forms. Plant Cell Physiol 35: 713-718

Uhde-Stone C, Zinn KE, Ramirez-Yanez M, Li A, Vance CP, Allan DL (2003) Nylon filter arrays reveal differential gene expression in proteoid roots of white lupin in response to phosphorus deficiency. Plant Physiol 131: 1064-1079

Vincent JL, Brewin NJ (2000) Immunolocalization of a cysteine protease in vacuoles, vesicles, and symbiosomes of pea nodule cells. Plant Physiol 123: 521-530

Vincent JM (1970) A Manual for the Practical Study of Root-Nodule Bacteria. I.B.P. Handbook No 15. Blackwell, Oxford

Wittenbach VA, Lin W, Hebert R (1982) Vacuolar localization of proteases and degradation of chloroplasts in mesophyll protoplasts from senescing primary wheat leaves. Plant Physiol 69: 98-102