Note

Plant Biotechnology, 17 (3), 253-257 (2000)

Isolation and structure determination of a cDNA encoding for a thiol protease from the cultured shoot primordia of *Matricaria chamomilla*

Chie KOHCHI[1], Masashi YASUDA[2], Toshifumi HIRATA[2]* and Lionel John GOAD[3]

[1] *Radioisotope Center, Hiroshima University, Higashi-Hiroshima 739-8526, Japan*
[2] *Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan*
[3] *School of Biological Sciences, Life Sciences Building, University of Liverpool, Liverpool L69 7ZB, UK*
**Corresponding author* E-mail address: thirata@sci.hiroshima-u.ac.jp

Received 4 February 2000; accepted 24 April 2000

Abstract

A cDNA encoding for a thiol protease was isolated from *Matricaria chamomilla*. The cDNA contained an open reading frame consisting of 501 amino acids, which had three active site motifs of thiol protease.

Most living organisms have developed defense reaction to protect themselves against exogenous stimulus. As one of the defense reactions, intracellular enzymes are secreted from the plant cells in response to an exogenous stimulus (Chibbar *et al.*, 1984; Izumi *et al.*, 1995). We have reported that various monoterpenoids, especially geraniol, exhibit a potent activity for the induction of apoptosis-like cell death as a defense reaction (Izumi *et al.*, 1999) in the cultured shoot primordia of *M. chamomilla* (German camomile). Recently we found that several proteases are secreted from the cultured shoot primordia of *M. chamomilla* when geraniol was administered to the cultures (Izumi *et al.*, 1996; and unpublished data). In continuation of the structure determination of proteins concerning to the defense reaction in higher plants (Kohchi *et al.*, 1999), we have now investigated the protein with protease activity in the cultured shoot primordia of *M. chamomilla* and report here the primary structure of a thiol protease.

Shoot primordia of *M. chamomilla* were cultured as described elsewhere (Hirata *et al.*, 1993; Takano *et al.*, 1991). Total cellular RNA was isolated by guanidine thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi, 1987). Contaminated polysaccharide was removed from RNA by precipitation with 2M LiCl.

cDNA cloning was performed using reverse transcription-polymerase chain reaction (RT-PCR) and its modified method, i.e., the rapid amplification of cDNA end (RACE) method (Loh *et al.*, 1989; Ohara *et al.*, 1989). The sequences of PCR primers other than Oligo dT_{12-18} (Pharmacia Biotech, Tokyo, Japan) and the schematic positions of the primers on a supposed mRNA for *M. chamomilla* thiol protease are summarized in **Table 1** and **Fig. 1**, respectively. Moloney murine leukemia virus reverse transcriptase (Promega, Madison, USA), terminal deoxy transferase (GIBCO BRL, Rockville), and KOD polymerase (TOYOBO, Osaka, Japan) were used for cDNA preparation and PCR. Each PCR product was cloned in pBluescriptSKII(-). DNA sequencing of double stranded plasmid DNAs was conducted according to the method reported previously (Sanger *et al.*, 1977; Smith *et al.*, 1986) and the standard manufacturer's protocol with a DNA sequencing kit for dye terminator cycle sequencing using an Applied Biosystems 377 DNA sequencer (Perkin Elmer Japan, Applied Biosystems Division, Chiba, Japan).

Recently, we isolated a protease from cultured shoot primordia of *M. chamomilla* as a defense-related protein against chemical stress (unpublished data). The partial amino acid sequence of the protease (PEAAHAINTG) was found to have homology with a region (IEAAHAIATG) of *Glycine max* thiol protease (TPase) (Kalinski *et al.*, 1990). Thus, in an attempt to clone TPase gene of *M. chamomilla*, we first performed RT-PCR using degenerated primers corresponding to the following amino acid sequence, PEAAHAINTG (pTP-1), QELVDCV (pTP-2), WIAKNS (pTP-3), and WGEDWG (pTP-4). The latter three amino acid sequences are those of *G. max* TPase, and they position downstream of IEAAHAIATG (**Table 1**, **Fig. 1**). As illustrated in **Fig. 1**, cDNA reverse transcribed with Oligo dT12-18 was applied to 1st

Table 1. Oligo nucleotides used for RT-PCR

Primer	Sequence (5' → 3')
3'RACE-A	TGG AAG AAT TCG CGG CCG CAG TTT TTT TTT TTT TTT TTT
RACE-C	TGG AAG AAT TCG CGG
3'RACE-E	TCG CGG CCG CAG TTT
5'TACE-D	TGG AAG AAT TCG CGG CCG CTT AAG GGG GGG GGG GGG
5'RACE-E	CG CGG CCG CTT A
pTP-1	CCI GA(AG) GCI GCI CA(CT) GCI AT(ACT) AA(CT) AC(ACGT) GG
pTP-2	CAA GAG CTC GT(AT) GAC TGT G
pTP-3	CA (TC)GA GTT CTT GA(TC) GAT CCA
pTP-4	CC CCA (AG)T(CT) T(GT)C TCC CCA (TC)GA
pTP-5	GGG TTC GGC TTA TGA TTT CC
pTP-6	ACA CCG GGG GTG TAT ATA AC
pTP-7	CGC ACC CAT AGT CAT AAG TG
pTP-8	CGT GTC CAT GTT TCC ACC AT
pTP-9	CTC TTC ATT ACT CAA GTC AGC

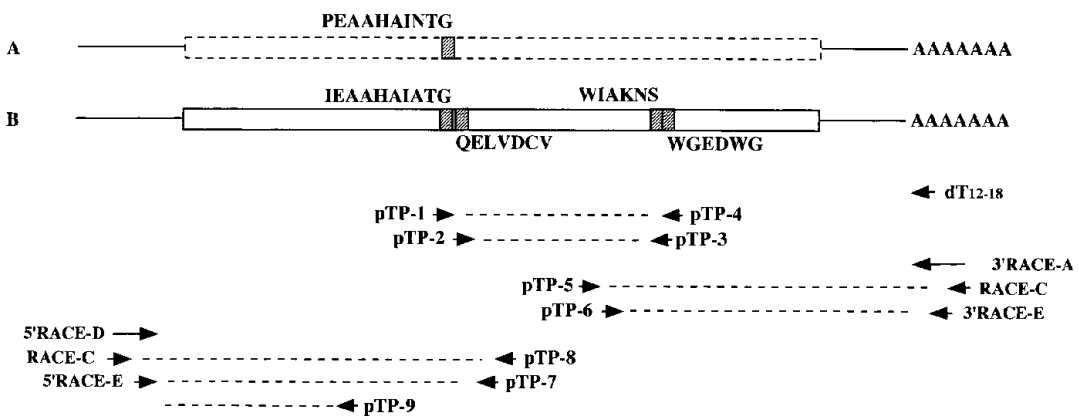

Fig. 1. The schematic positions of primers used for PCR on a supposed mRNA of *M. chamomilla* thiol protease.
A supposed mRNA and its ORF (dotted line square) for *M. chamomilla* thiol protease (A), and a mRNA and its ORF (solid line square) for *G. max* thiol protease (B) are shown together with their partial amino acid sequences (gray squares). Arrows show primers, their directions, and schematic positions corresponding to mRNAs above.

PCR using pTP-1 and pTP-4 as primers. An aliquot of the 1st PCR product was subjected to a second nested PCR using pTP-2 and pTP-3. The nested PCR gave a 410-bp DNA fragment. Sequence analysis revealed that the DNA fragment is a part of an open reading frame (ORF), in which an eukaryotic thiol protease histidine active site is contained. Thus, we named the cDNA as *ctp* regarding as a cDNA encoding for one of the thiol proteases of *M. chamomilla*.

To clone the whole ORF of ctp, we subsequently tried 3' and 5' RACE-PCR. For 3' RACE-PCR, two sense primers (pTP-5 and pTP-6) were designed based on the determined ctp sequence, and three antisense primers (3'RACE-A, RACE-C, and 3'RACE-E) having Not I adapter site were prepared (**Table 1**). cDNA reverse transcribed with 3'RACE-A was applied to 1st PCR using pTP-5 and RACE-C as primers. An aliquot of the 1st PCR product was subjected to a second nested PCR using pTP-6 and 3'RACE-E (**Fig. 1**). The nested PCR gave DNA fragments which have polyA stretch at the 3'-end and are overlap with *ctp* at 5'-end.

For 5' RACE-PCR, two antisense primers (pTP-7 and pTP-8) were designed from the determined *ctp* sequence, and three sense primers (5'RACE-D, RACE-C, and 5'RACE-E) having Not I adapter site were prepared (**Table 1**). Oligo dT_{12-18} primed single stranded cDNA was dC tailed at the 3' end and then the double stranded cDNA was synthesized using oligo dG/Not I adaptor primer (5'RACE-D). First PCR was performed using RACE-C and pTP-8 as primers. An aliquot of the 1st PCR product was subjected to a 2nd nested PCR using

5'RACE-E and pTP-7. After confirming that the nested PCR product has a cDNA overlapping with *ctp*, pTP-9 primer was newly prepared. To determine the sequence of 5' non coding region of *ctp* more longer, 5' RACE-PCR was re-performed starting with cDNA reverse transcribed with pTP-8 followed by 5'RACE-PCR using pairs of primers, RACE-C / pTP-7 and 5'RACE-E / pTP-9 (**Fig. 1**).

Sequences of PCR products were analyzed for at least 6 independent clones for each region and consensus sequences were determined, respectively. **Fig. 2** shows the combined 1732 bp-entire sequence of *ctp* and predicted amino acid sequence (GenBank accession # AF182079).

The largest ORF, which is composed of 501 amino acids, starts at nucleotide position (ntp) 44. The predicted protein from the ORF has three eukaryotic thiol protease motifs, i.e., cysteine active site at amino acid position (aap) 161-172, histidine active site at aap 305-315, and asparagine active site at aap 322-341. We designated the predicted protein as cTPase. The cTPase may not be identical to the protease which we have previously purified from oil body of *M. chamomilla*, because PEAAHAINTG sequence is not found in the deduced amino acid sequence. Instead of PEAAHAINTG, a related amino acid sequence, IESANAIATG is found at aap 176-185 in cTPase. Presumably, this region appears to cross react with the degenerated PCR primer, pTP-1, so that *ctp* was amplified by RT-PCR.

Homology search in the protein database indicated that cTPase has approximately 40% homology with TPase of *Lycopersicon esculentum* (NCBI # AJ003137) (Lers *et al.*, 1998), *Phaseolus vulgaris* (NCBI # Z99954), *Solanum tuberosum* (NCBI # AJ245924), and *Glycine max* (NCBI # J05560) (Kalinski *et al.*, 1990) as shown in **Fig. 3**. These plant TPases have a high number of N-myristoylation sites in their molecules. cTPase also has 14 possible N-myristoylation sites. This suggests that the plant TPases, including cTPase, might localize

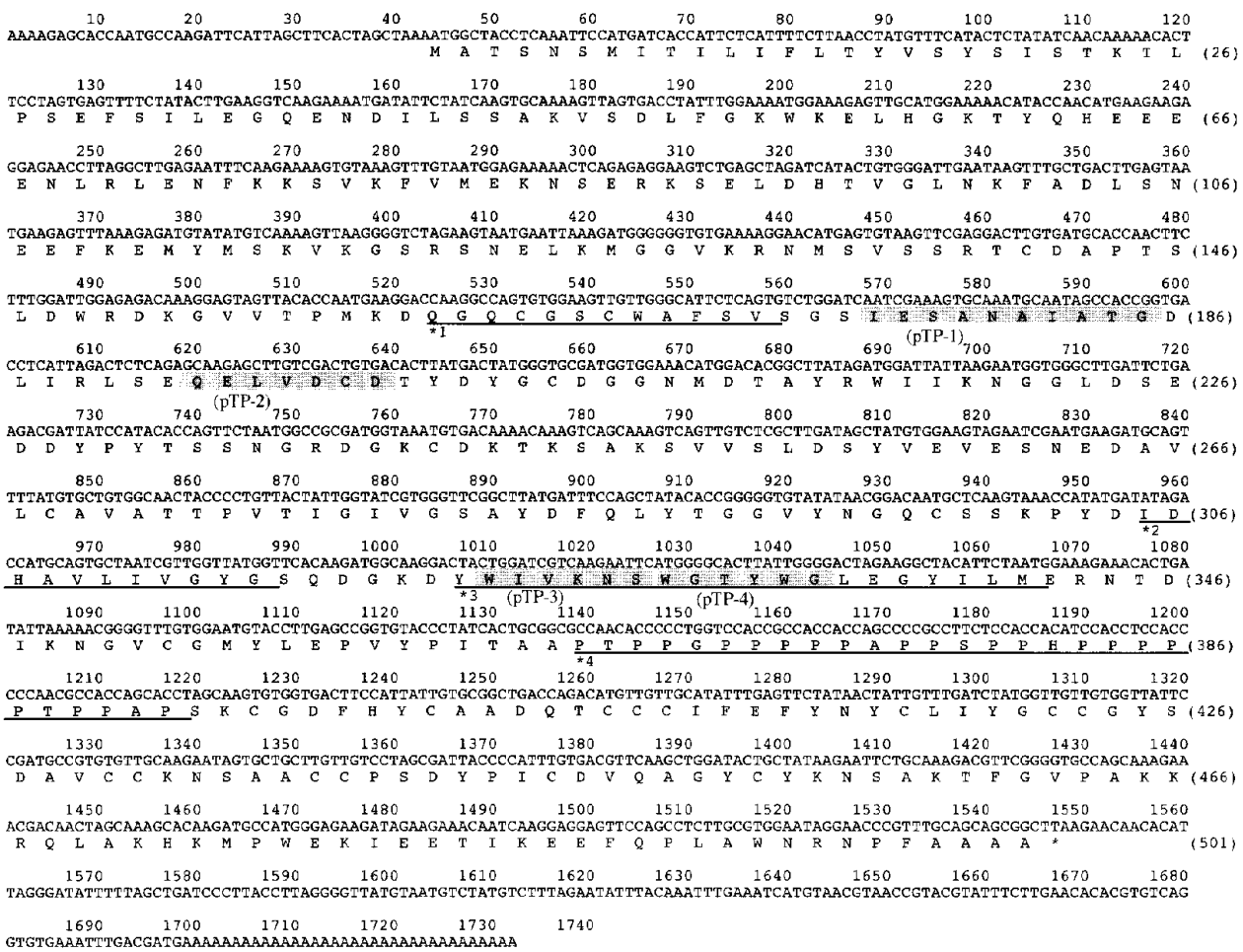

Fig. 2. The nucleotide sequence of *ctp* encoding *M. chamomilla* TPase (cTPase).
The predicted amino acid sequence is shown below the DNA sequence. Asterisks show motifs as follows; *1 Eukaryotic thiol protease cysteine active site, *2 Eukaryotic thiol protease histidine active site, *3 Eukaryotic thiol protease asparagine active site, *4 proline rich region. Shadows represent the possible amino acid loci cross-reacting with pTP-1, 2, 3, and 4.

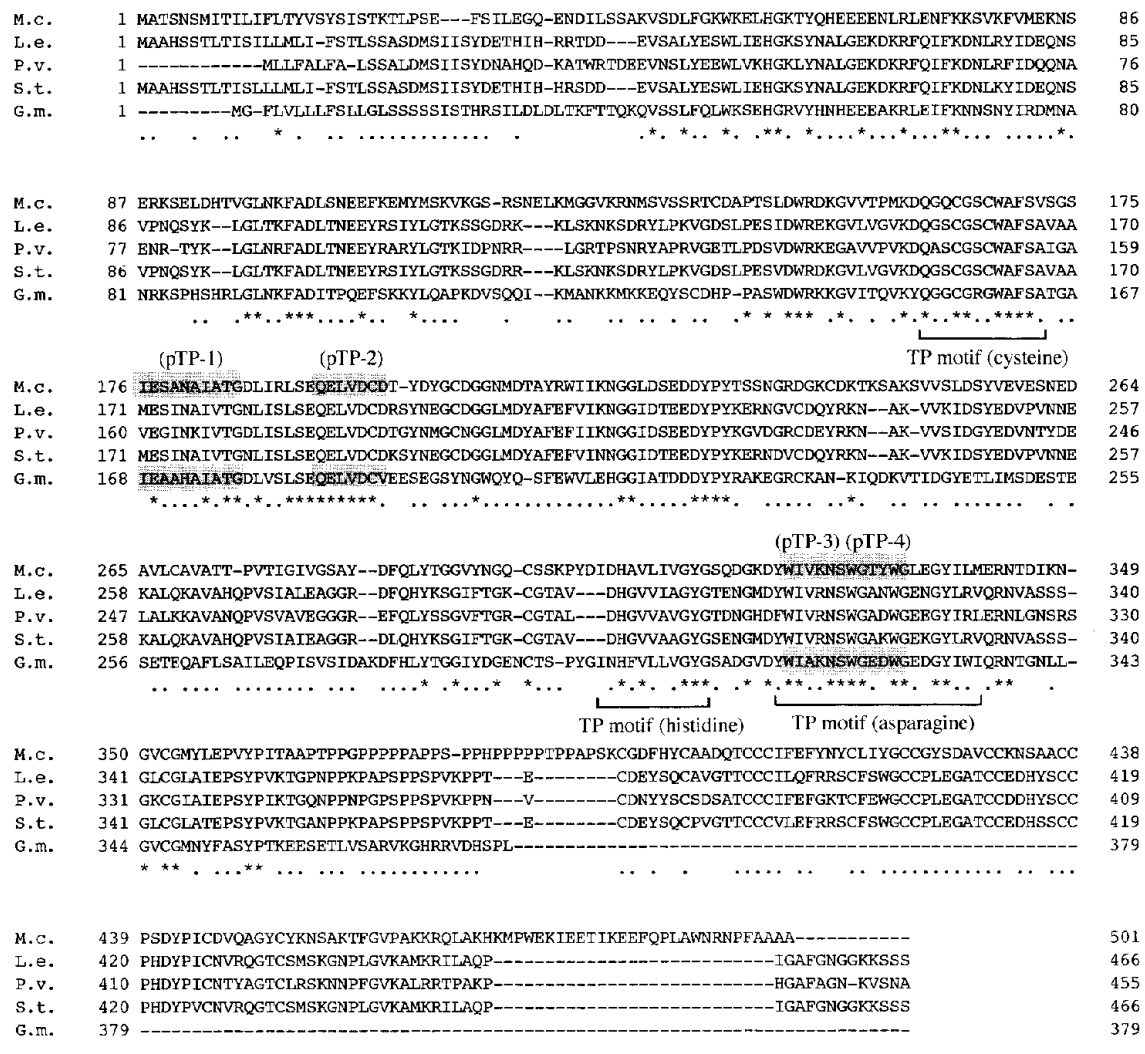

Fig. 3. Comparison of amino acid sequences of thiol protease among several plants. M.c.: *Matricaria chamomilla*, L.e.: *Lycopersicon esculentum* (NCBI # AJ003137) (Lers et al., 1998), P.v.: *Phaseolus vulgaris* (NCBI # Z99954), S.t.: *Solanum tuberosum* (NCBI # AJ245924), G.m.: *Glycine max* (NCBI # J05560) (Kalinski *et al.*, 1990). Asterisks denote identical amino acid residues through all thiol proteases and dots denote common amino acid residues to three of five thiol proteases. Dashes represent gaps. The thiol protease active sites are shown below the amino acid sequences. Shadows represent the possible amino acid loci cross-reacting with pTP-1, 2, 3, and 4.

in membrane or lipid layer locus. Although cTPase would be a different protease from that we previously isolated from *M. chamomilla* (Izumi *et al.*, 1996), it could be one of the membrane- or oil body-associated proteases. As a characteristic feature of cTPase, a proline rich region is found at aap 366-392, but the role of the proline stretch is not clear. A number of possible target sites for casein kinase II, protein kinase C, and tyrosine kinase are found in cTPase, suggesting that the activity of cTPase might be modified by phosphorylation.

Further investigations are needed for clarifying the function of cTPase in the shoot primordia of *M. chamomilla*.

Acknowledgements

The present work was in part supported by Grant-in-Aids for Scientific Research No. 09480142 from the Ministry of Education, Science and Culture of Japan, and for the UK-Japan Collaboration Research Project (1997-1998) of Japan Society for the Promotion of Science (JSPS).

References

Chibbar, R. N., Cella, D., Albani, D., Huystee, R. B. V., 1984. The growth and peroxidase synthesis of two carrot cell lines. J. Exp. Botany, **35:** 1846-1852.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem., **162:** 156-

159.

Hirata, T., Izumi, S., Akita, K., Fukuda, N., Hirashima, T., Taniguchi, K., Nishimori, C., 1993. Formation of oil bodies in cultured shoot primordia of *Matricaria chamomilla*. Plant Tissue Culture Letters, **10:** 289-292.

Izumi, S., Takashima, O., Fukuda, N., Hirata, T., 1996. Mr 33K oil body associated protein in cultured shoot primordia of *Matricaria chamomilla*. Phytochemistry, **42:** 309-312.

Izumi, S., Takashima, O., Hirata, T., 1999. Geraniol is a potent inducer of apoptosis-like cell death in the cultured shoot primordia of Matricaria chamomilla. Biochem. Biophys. Res. Commun., **259:** 519-22.

Izumi, S., Yamamoto, Y., Hirata, T., 1995. Secretion of an esterase from the cultured suspention cells of *Marchantia polymorpha*. Phytochemistry, **38:** 831-833.

Kalinski, A., Weisemann, J. M., Matthews, B. F., Herman, E. M., 1990. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. J. Biol. Chem., **265:** 13843-13848.

Kohchi, C., Yasuda, M., Hirata, T., 1999. Isolation of a cDNA encoding for a carboxypeptidase, having leucine zipper structure at the N-terminal region, from the cultured shoot primordia of *Matricaria chamomilla*. Plant Biotech., **16:** 409-412.

Lers, A., burd, S., Sonego, L., Khalchitski, A., Lomaniec, E., 1998. Nucleotide sequence of a full-length C14 cDNA clone encoding for cysteine protease CYP1 from tomato. Plant Physiol., **116:** 1193-1193.

Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier L.L., Davis, M. M., 1989. Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain. Sci., **243:** 217-220

Ohara, O., Dorit, R. L., Gilbert, W., 1989. One-sided polymerase chain reaction: the amplification of cDNA. Proc. Natl. Acad. Sci. U.S.A., **86:** 5673-5677.

Sanger, F., Nicklen, S., Coulson, A. R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A., **74:** 5463-5467.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., Hood, L. E., 1986. Fluorescence detection in automated DNA sequence analysis. Nature., **321:** 674-679.

Takano, H., Hirano, M., Taniguchi, K., Tanaka, R., Kondo, K., 1991. Rapid clonal propagation of *Matricaria chamomilla* by tissue-cultured shoot primordia. Japan. J. Breed., **41:** 421-426.