Cloning and Expression of the Cathepsin F-like Cysteine Protease Gene in *Escherichia coli* and Its Characterization

Han Seung Joo[3], Kwang Bon Koo[1], Kyung In Park[1], Song Hwan Bae[4], Jong Won Yun[2], Chung Soon Chang[3], and Jang Won Choi[1]*

[1]*Department of Bioindustry, College of Life and Environment, Daegu University, Kyungbuk 712-714, Republic of Korea*
[2]*Department of Life Engineering, College of Engineering, Daegu University, Kyungbuk 712-714, Republic of Korea*
[3]*Department of Biochemistry, College of Medicine, Inha University, Inchon 400-103, Republic of Korea*
[4]*Department of Food & Biotechnology, College of Science and Engineering, Hankyong National Univercity, Ansung 456-749, Republic of Korea*

In this study, we have cloned a novel cDNA encoding for a papain-family cysteine protease from the Uni-ZAP XR cDNA library of the polychaete, *Periserrula leucophryna*. This gene was expressed in *Escherichia coli* using the T7 promoter system, and the protease was characterized after partial purification. First, the partial DNA fragment (498 bp) was amplified from the total RNA via RT-PCR using degenerated primers derived from the conserved region of cysteine protease. The full-length cDNA of cysteine protease (*PLCP*) was prepared via the screening of the Uni-ZAP XR cDNA library using the ^{32}P-labeled partial DNA fragment. As a result, the *PLCP* gene was determined to consist of a 2591 bp nucleotide sequence (CDS: 173-1024 bp) which encodes for a 283-amino acid polypeptide, which is itself composed of an 59-residue signal sequence, a 6-residue propeptide, a 218-residue mature protein, and a long 3'-noncoding region encompassing 1564 bp. The predicted molecular weights of the preproprotein and the mature protein were calculated as 31.8 kDa and 25 kDa, respectively. The results of sequence analysis and alignment revealed a significant degree of sequence similarity with other eukaryotic cysteine proteases, including the conserved catalytic triad of the Cys^{90}, His^{226}, and Asn^{250} residues which characterize the C1 family of papain-like cysteine protease. The nucleotide and amino acid sequences of the novel gene were deposited into the GenBank database under the accession numbers, AY390282 and AAR27011, respectively. The results of Northern blot analysis revealed the 2.5 kb size of the transcript and ubiquitous expression throughout the entirety of the body, head, gut, and skin, which suggested that the PLCP may be grouped within the cathepsin F-like proteases. The region encoding for the mature form of the protease was then subcloned into the pT7-7 expression vector following PCR amplification using the designed primers, including the initiation and termination codons. The recombinant cysteine proteases were generated in a range of 6.3% to 12.5% of the total cell proteins in the *E. coli* BL21(DE3) strain for 8 transformants. The results of SDS-PAGE and Western blot analysis indicated that a cysteine protease of approximately 25 kDa (mature form) was generated. The optimal pH and temperature of the enzyme were determined to be approximately 9.5 and 35°C, respectively, thereby indicating that the cysteine protease is a member of the alkaline protease group. The evaluation of substrate specificity indicated that the purified protease was more active towards Arg-X or Lys-X and did not efficiently cleave the substrates with non-polar amino acids at the P1 site. The PLCP evidenced fibrinolytic activity on the plasminogen-free fibrin plate test.

Keywords: cathepsin F-like cysteine protease, catalytic triad, fibrinolytic activity

The cysteine proteases are a group of enzymes that belong to one of the four major classes of proteolytic enzymes, and are generated by a variety of organisms, including viruses (Allaire *et al.*, 1994), bacteria, protozoa, yeasts, plants, helminths, insects, and mammals (Baker and Drenth, 1987; Polgar, 1989; Shaw, 1990; Berti and Storer, 1995). The cysteine proteases are a widespread group of enzymes that catalyze the hydrolysis of different proteins and perform a pivotal role in the degradation and turnover of intracellular proteins (Bond and Butler, 1987; Chapman *et al.*, 1997). These proteolytic enzymes can be classified into more than 20 different families, including the papain family, the calpains, streptopains,

clostripains, viral cysteine proteinases, and caspatase. The largest of these families is the papain family (Rawlings and Barrett, 1994). Thus far, 12 mammalian C1 peptidases have been identified, representing a principal component of the endosomal/lysosomal proteolytic system. According to their respective tissue distributions, two functional groups can be delineated. The first group, the cathepsins B, C, F, H, L, O, and Z evidence ubiquitous expression, and may perform a pivotal function in non-specific terminal protein degradation, whereas the members of the second functional group, which includes cathepsins J, K, L2, S, and W, evidence a tissue-restricted expression pattern (Rawlings and Barrett, 1999; Tisljar *et al.*, 1999; http://www.bi.bbsrc.ac.uk/Merops/Merops. htm). The C1 family of the papain-like cysteine proteases is characterized by an active site comprised of a cysteine, a histidine, and an asparagine residue, which constitute a

---

so-called 'catalytic triad' (Berti and Storer, 1995).

Periserrula leucophryna is one of the polychaetes which inhabits the tidal mud flats of the Korean west sea (Yellow sea). This species has been taxonomically reported as a unique genus and species (Brands, Systema Naturae 2000). *P. leucophryna* is a sedentary polychaete which inhabits burrows in mud flats, and consists of more than 300 segments. *P. leucophryna* specimens detected in tidal flats at Kwangwha Island have been measured to be 2 to 2.5 meters in length. They are also scavengers, and pick up dead material, ingest sand or mud, and extract nutrition from the particles. This species also performs essential ecological functions, and serves as a predator of small invertebrates and as food for fish and large invertebrates. As they are recognized to perform a crucial function in environmental sanitation and ecological circulation (Jeong *et al.*, 2006), proteases may also be considered to be important in this role. Accordingly, the cloning and expression of the protease gene was conducted in *E. coli,* and its functions were characterized.

In this study, we describe the isolation and characterization of the cysteine protease gene from *P. leucophryna.* First, RT-PCR was conducted using the degenerated primers corresponding to the conserved regions of the catalytic triad for the established cysteine proteases. The RT-PCR yielded a partial gene construct of the cysteine protease expressed in *P. leucophryna.* Secondly, by using the partial gene as a probe in the plaque hybridization step, the full-length cDNA clones encoding for the cysteine protease of *P. leucophryna* were obtained, characterized, and compared to the cysteine proteases previously identified in other organisms. Thirdly, in order to generate the recombinant protein, the cysteine protease gene has been cloned into the *E. coli* expression vector under the control of the T7 promoter (Kim *et al.*, 2005). The expression levels of the recombinant proteins were determined and optimized in two different strains. The recombinant proteins were partially purified and the biological activity was measured in order to assess the functions of the cysteine protease from *P. leucophryna.*

Materials and Methods

Bacterial strains, medium, and plasmids

E. coli HB101 (Lacks and Greenberg, 1977) was employed as a primary host for the transformation and propagation of plasmids. *E. coli* JM109 (Yanisch-Perron *et al.*, 1985) was utilized to verify the introduction of insert DNA via an α-complementation chromogenic test. The *E. coli* BL21 (DE3) strain was utilized for the expression of the T7 promoter system, particularly for non-toxic protein expression. *E. coli* BL21(DE3)pLysE harbors the same DE3 lysogen as the BL21(DE3)pLysS which harbors the T7 lysozyme, a T7 RNA polymerase inhibitor, which prevents leaky expression in uninduced cells, and also harbors the pLysE plasmid with the T7 lysozyme, which is useful for the expression of toxic proteins (Studier, 1991). The pGEM-Teasy vector (Promega, USA) was utilized in the cloning of the PCR products. The pT7-7 plasmid (Tabor and Richardson, 1985) harboring the T7 promoter was utilized in the construction of the expression plasmids containing the cysteine protease gene. *E. coli* XL1-Blue MRF' (F', *pro*AB, *lac*I^{q}ZΔM15, *thi,*

*rec*A, *gyr*A, *rel*A, *sup*E, Tn10) was utilized as the host strain for the construction of a cDNA library and an *E. coli* SOLR strain (F', *pro*AB, *lac*I^{q}ZΔM15, *lac, gyr*A, *rel*A, *thi, sbc*C, *rec*B, Tn5, Su⁻) was utilized for *in vivo* excision. *E. coli* cells harboring the plasmid were grown in Luria-Bertani (LB) medium (Sambrook *et al.*, 1989) supplemented with 50 μg of ampicillin/ml or 25 μg of chloramphenicol/ml.

Amplification of cysteine protease partial gene.

In order to generate the partial gene fragment of the cysteine protease, the oligonucleotide primers were designed from the conserved sequences flanking the active site histidine, asparagine, and cysteine residues of the eukaryotic cysteine proteases, which are characteristic of the C1 family of papain-like cysteine proteases. The primers, 5CysF1 primer; 5'-TG YGGIWSITGYTGGGCITTYWSIRCI-3' and 3CysR1 primer; 5'-CCAISWRTTYTTIACDATCCARTAIGG-3' were synthesized for the conserved regions, [CGSCWAFSA(T)] and [PYWIVKNSW], respectively. Total cDNAs or RNA were employed as the template for the polymerase chain reaction. The DNA was amplified in a thermal cycler (Technique, USA) under the following conditions: 95°C 5 min, 45°C 1 min, 72°C 1.5 min (1 cycle); 95°C 1 min, 45°C 30 sec, 72°C 1.5 min (29 cycles); 72°C 10 min (1 cycle). The PCR products were then fractionated on 1% agarose gel and the DNA fragments were recovered and cloned into pGEM-Teasy vector (Promega, USA). The proper clones were then screened and sequenced in both directions using M13 forward and reverse primers, with an automated sequencer (Applied Biosystems, USA).

Construction and screening of cDNA library from Periserrula leucophryna

The polychaetes utilized herein (*Periserrula leucophryna*) were collected from tidal flats of the Kwanghwa Island in the Korean Yellow Sea. The total RNAs were isolated from a variety of tissues using Trizol reagent (Sigma, USA) and a QIAGEN RNA extraction kit (Hilden, Germany) or via the guanidinium/phenol method (Chomczynski and Sacchi, 1987). The extracted total RNA was then purified further in order to isolate the poly(A)⁺RNA, via oligo (dT) cellulose (Stratagene, USA) chromatography (Sambrook *et al.*, 1989). The resultant cDNA was size-fractionated via Sepharose CL-2B gel filtration. The cDNAs with sizes larger than 500 bp were ligated to the Uni-ZAP XR vector and were packaged with Gigapack III gold extracts (Stratagene, USA). The efficiency of the total cDNA library was determined via titering with serial dilutions of the final packaging reaction. The packaged phages were then amplified for one round in *E. coli* XL1-Blue MRF' cells. The full-length cDNA clones were screened via plaque hybridization with the *P. leucophryna* cDNA library. Approximately 5×10⁵ plaques from the library per plate were utilized in the screening of the full-length cDNA. The partial cysteine protease gene (498 bp) was labeled using the Prime-a-Gene labeling kit (Promega, USA), and the unincorporated nucleotides were removed using a Sephadex G-50 spun column. The resultant probe was hybridized to plaques on a colony/plaque screen filter (PerkinElmer, USA) at 42°C overnight in 2x PIPES (0.8 M NaCl and 20 mM PIPES; pH 6.5), 50% deionized

formamide, 0.5% SDS, and salmon sperm DNA (100 μg/ml). The positive phages were acquired after tertiary screening, and the pBluescript phagemid harboring the insert DNA was excised using the ExAssist™ helper phage and the *E. coli* SOLR strain, in accordance with the instructions of the manufacturer (Stratagene, USA)

Northern blot analysis

In order to prepare the total RNA blot from various tissues of *P. leucophryna*, 20 μg of total RNA from each specimen was fractionated on 1.2% agarose gel containing formaldehyde, then transferred to a Nytran-Plus membrane (Schleicher and Schuell, Germany). The blot was hybridized with ^{32}P-labeled partial cysteine protease DNA fragments at 42°C in a buffer containing 2x PIPES (0.8 M NaCl and 20 mM PIPES, pH 6.5), 50% deionized formamide, 0.5% SDS, and salmon sperm DNA (100 μg/ml). The membrane was washed with increasing precision from 2x SSC (0.3 M NaCl and 0.03 M sodium citrate) solution to a final wash of 0.1 x SSC and 0.1% SDS. RNA integrity and equal loading were confirmed via hybridization with β-actin probe.

Construction of expression vector

The sequence encoding for the predicted mature cysteine protease was amplified from the obtained clone (named *PLCP*) via PCR using an oligonucleotide primer set to which *Nde*I and *Sal*I restriction enzyme sites had been added (underlined) to assist cloning into the pT7-7 expression vector (Tabor and Richardson, 1985) in the frame. The primers for *PLCP* were as follows: 5'-GGGG<u>CATATG</u>CCACCAGCATCGTTT GACTGGAGA-3' and 5'-CCCC<u>GTCGAC</u>CTACTTGACAA TAGCAGATGT-3'. The mature form of the cysteine protease gene was amplified under the following conditions: 95°C 1 min, 45°C 1 min, 72°C 1.5 min (1 cycle); 95°C 1 min, 50°C 1 min, 72°C 1.5 min (29 cycles); 72°C 7 min (1 cycle). The PCR products were fractionated on 1% agarose gel and the DNA fragments were recovered, digested with *Nde*I and *Sal*I, and cloned into linearized pT7-7 with the same enzymes.

SDS-PAGE and western blot analysis

The recombinant expression plasmid harboring the mature form of the cysteine protease gene was transferred into *E. coli* BL21(DE3) or *E. coli* BL21(DE3)pLysE. When the cell density achieved an A$_{600}$ value of 0.5, IPTG (Isopropyl-β-D-thiogalactoside) was added to a final concentration of 1 mM. After an additional 6 h of incubation, the lysates were prepared from 1 ml of the culture. The cell pellets were suspended in 200 μl of sample buffer (0.05 M Tris-HCl, pH 6.8, 0.1 M DTT, 2% SDS, 10% glycerol, and 0.1% bromphenol blue) and boiled for 5 min at 90°C. The samples (10 μl) were analyzed via SDS-PAGE, as described by Laemmli (1970), using a mini-gel system (Bio-Rad, USA). The proteins were then transferred to PVDF membranes using a Trans-Blot SD apparatus (Bio-Rad, USA). The blots were blocked for 1 h in Tris-buffered saline (TBST) containing 5% skim milk and incubated with polyclonal rabbit anti-cysteine protease antibody (1:1,000 dilution) from PEPTRON (Daejeon, Korea) for 1 h at room temperature. After the unbound primary antibodies were washed with washing buffer (TBST and 0.1% Tween 20) three times for 15 min each, the

blots were treated with horseradish peroxidase-conjugated goat anti-rabbit antiserum diluted with TBST (1:3,000 dilution) and developed using an ECL™ Western blotting detection system (RPN2109) from Amersham Pharmacia Biotech, USA.

Refolding of the recombinant PLCP

E. coli BL21(DE3) harboring the recombinant expression plasmid was grown overnight at 37°C in LB broth supplemented with 50 μg of ampicillin/ml. The culture was diluted to 1/200 into 200 ml of the same medium and incubated. When the cell density reached an A$_{600}$ value of 0.5, IPTG (Isopropyl-β-D-thiogalactoside) was added to a final concentration of 1 mM. After an additional 6 h of incubation, the lysates were prepared from 200 ml of the culture. The cells were washed briefly in 0.1 M Tris-HCl buffer (pH 8.2), suspended in ice-cold identical buffer (5 ml/g of the wet cells), sonicated (10 cycles of 30 sec each, with cooling for 2 min between the cycles), then centrifuged for 30 min at 12,000×g at 4°C. The pellet was then washed twice in 0.1 M Tris-HCl buffer (pH 8.2) containing 2.5 M urea and 2% Triton X-100, and centrifuged for 30 min at 12,000×g and 4°C. The pellet was solubilized in 0.1 M Tris-HCl buffer (pH 8.2) containing 8 M urea, 10 mM NaCl, 2 mM EDTA, 1 mM DTT, and 0.2% SDS for 4 h at 4°C with vigorous stirring. The insoluble material was removed via 30 min of centrifugation at 20,000×g at 4°C, and the supernatant was recovered. The protein was refolded via the gradual removal of urea upon dialysis into the buffer containing 0.1 M Tris-HCl (pH 8.2) at 4°C with vigorous stirring.

Partial purification of PLCP

Polyclonal antibody was purified from the rabbit serum immunized with the internal peptide of the PLCP (KFTVG DVAVYINSSVNI) in order to prepare the immunoaffinity resin. An immunoaffinity column (1.0×5 cm) was equilibrated with 0.1 M Tris-HCl buffer (pH 8.2) containing 0.5 M NaCl. The refolded proteins were applied to the column and washed in identical buffer until the absorbance of the flow-through at 280 nm declined to almost zero. The bound protease was then eluted with 0.5% acetic acid. The flow rate was 5 ml/h and 1 ml fractions containing 0.1 volume of 1 M Tris-HCl, pH 9.0, were collected in order to accomplish the quick neutralization of the eluent. The active fractions were concentrated using Centriprep-10 Microconcentrators (Millipore, USA), divided into aliquots, and stored at -70°C until use.

Enzyme assay and optimum pH

For routine measurements, protease activity was assayed using Tos-Gly-Pro-Lys-pNA as a substrate. The reaction was conducted in a total volume of 200 μl of assay buffer (0.1 M glycine-NaOH; pH 9.5), coupled with an appropriate quantity of enzyme. Ten microliter of 5 mM substrate was added to the reaction mixture and the enzyme reaction was allowed to proceed for 10 min at 37°C. The quantity of released *para*-nitroanilide was determined at 405 nm and an absorption coefficient of 10,000 cm/M was utilized for the calculation. One unit of the enzyme was defined as the amount of enzyme necessary for the liberation of 1 nmol of *para*-nitroanilide per min under standard assay conditions (Joo *et al.*, 2002).

In order to determine the optimum pH for PLCP, the enzyme activity was determined at 37°C in the following buffer systems: 0.1 M sodium phosphate (pH 7.0-7.5); 0.1 M Tris-HCl (pH 8.0-9.0); and 0.1 M glycine-NaOH (pH 9.5-10).

Substrate specificity

The purified protease was assessed with regard to amidolytic activity toward a variety of chromogenic peptide substrates. The reaction was conducted in a total volume of 200 μl of assay buffer with 10 ng of the purified enzyme. Then, 10 μl of 5 mM stock of each substrate was added to the reaction mixture, and the reaction was allowed to proceed further at 37°C for an additional 10 min. The proteolytic activity toward Tos-GPK was utilized as a control.

Inhibition test

The purified protease was pre-incubated with several inhibitors, including E-64 and PMSF, for 30 min at room temperature, after which the remaining activity was determined under standard assay conditions. Residual activities in the presence of the inhibitors were compared with the controls, without inhibitor.

Fibrinolytic activity

The fibrinolytic activity of PLCP was assayed via the fibrin plate method as described by Astrup and Müllertz (1952). Ten mililiter of 0.5% (w/v) human plasminogen-free fibrinogen solution in 50 mM Tris-HCl buffer (pH 8.2) was poured into a petri dish (87×15 mm), to which 0.2 ml of bovine thrombin (10 NIH U/ml) was added, and the plate was incubated for 30 min at room temperature in order to form fibrin clots. After 5 mm paper discs were carefully positioned onto the fibrin plates, 10 μl of each sample was applied to the discs and incubated for 4 h at 37°C. An equal volume of plasmin solution (1 U/ml) was also incubated as a positive control, and phosphate-buffered saline (PBS) was utilized as a negative control.

Results and Discussion

Construction of Periserrula leucophryna cDNA library

Total RNA was initially extracted from the entire bodies of the *Periserrula leucophryna* specimens. DNA in the extract was undetectable, and the integrity of RNA was verified via evaluation of the agarose gel resolution pattern of the total RNA. Poly(A)$^+$RNA was isolated by two passes via oligo (dT)-cellulose chromatography, as described by Sambrook et al. (1989). The eluted fractions were then precipitated with ethanol and redissolved in DEPC-treated water. A yield of approximately 1% of the total RNA fractionated was generated, with an A_{260}/A_{280} value of 1.9. Using 5 μg of poly(A)$^+$RNA as a template, a cDNA library was synthesized in accordance with the instructions of the manufacturer (Stratagene, USA). As a result, the efficiency of the cDNA library was calculated to be 8×10^5 plaque forming units (pfu)/μg of cDNA containing approximately 97% recombinant phages. As a primary library can be unstable, the library was immediately amplified for one round in the *E. coli* XL1-Blue MRF' host strain. The titer of the amplified cDNA library was determined to be 5×10^{10} pfu/ml. Using

the ExAssist helper phage and the *E. coli* SOLR strain, an efficient excision of the pBluescript phagemid from the Uni-ZAP XR vector was conducted in order to generate various sizes of the phagemid. When the phagemids were digested with *Eco*RI and *Xho*I, insert DNAs ranging from 0.5 kb to a maximum of 3 kb were distributed throughout the library. The cDNA library was screened for the isolation of the cysteine protease gene and total cDNAs extracted from the library were utilized for PCR amplification as a template for the screening of the partial genes of cysteine protease.

Amplification of cysteine protease partial gene and Sequence analysis

In an effort to isolate the novel cysteine protease gene expressed in *Periserrula leucophryna*, RT-PCR amplification was conducted using the degenerated primers corresponding to [CGSCWAFSA(T)] and [PYWIVKNSW] of the conserved catalytic domains in the C1 family of papain-like cysteine proteases, as previously described (Sakanari et al., 1989; Gschwend et al., 1997). The partial cDNA products were acquired from the total RNA or total cDNA of *Periserrula leucophryna*. As a result of the sequencing and BLAST searches, one type of novel cysteine protease cDNA fragment was acquired, and the exact size of the partial gene was determined to be 498 bp in length. The obtained clone was designated as the 6C-1 clone. The DNA sequence incorporating the PCR primers was determined to faithfully reflect the actual DNA sequence of the amplified gene fragment. As a result of our search of the NCBI (National Center for Biotechnology Information) conserved domain database, the putative amino acids of the 6C-1 clone evidenced 76.6% homology with the papain-like cysteine protease. A comparison of the deduced amino acid sequences with those in the GenBank database indicated that the amplified cDNA fragment encoded for a distinct papain-like cysteine protease evidencing homology to the regions surrounding the conserved active site residues. Using the CLUSTAL V multiple sequence alignment program (Higgins and Sharp, 1989), the multiple alignment of the translated sequences of the cysteine protease PCR fragment, with their best match candidates, revealed a striking conservation of a number of residues and catalytic triad regions within the sequence. The partial gene fragment isolated via this method was utilized as a probe DNA for the screening of an oligo(dT)-primed *P. leucophryna* cDNA library. As a result, the full-length cDNA encoding for cysteine protease was acquired.

Screening and characterization of full length cDNAs for cysteine protease

Approximately 100,000-600,000 recombinant phage plaques were evaluated to obtain the full-length cDNA of the cysteine protease using the cloned PCR fragment termed 6C-1 as a probe. Via primary, secondary, and tertiary screening, a total of 20 different clones hybridized to the 6C-1 probe were identified. Five plasmid clones from the hybridized phages were rescued and digested with *Eco*RI and *Xho*I in order to confirm the homogeneity of the insert DNA size. When the obtained clones larger than 2.0 kb were sequenced and characterized completely, heterogeneity among the different clones was detected within the 5'-coding region. The

```
TACCTGTCTGCCCATAAAGAAAAATTCACAAAAACTCCAGAAAGTAACAAACTCCAAGACAAACCATATTCCTAT    75
TTTAGGAGCAGATGATATTGATAATGCTCACTGGGGAGTGTTTGAGAACTTCAAGCTAAAGTACAACAAACAGTA   150
CAAGAACAGGGCAGAGGAAGAAATGAGGTTCAAGATATTTAGAGAGAATATGAAGAAGATTAACACACTGAATGA   225
                    M   R   F   K   I   F   R   E   N   M   K   K   I   N   T   L   N   D
CAATGAACTTGGTGATGCTGAGTATGGAGTCACTCAGTTCTCTGATTTGGCTGAGGAGGAGTTTCGTCGCTACTA   300
  N   E   L   G   D   A   E   Y   G   V   T   Q   F   S   D   L   A   E   E   E   F   R   R   Y   Y
CCTGACACCAAAGTGGGACCTCAGTCATCGACCAGATCTGGTGAGAGCAAAAATCCCAGATGTTGACCCACCAGC   375
  L   T   P   K   W   D   L   S   H   R   P   D   L   V   R   A↑K   I   P   D   V   D↑P   P   A
ATCGTTTGACTGGAGAGACCACAATGCTGTGACTCCCGTCAAGAACCAGGGAATGTGTGGGTCCTGCTGGGCCTT   450
  S   F   D   W   R   D   H   N   A   V   T   P   V   K   N   Q   G   M   C   G   S   Ĉ   W   A   F
CTCAACCACAGAAAACATTGAGGGACAGTGGGCCATTCACAGGAACAAACTGGTGTCTCTGTCTGAGCAAGAACT   525
  S   T   T   E   N   I   E   G   Q   W   A   I   H   R   N   K   L   V   S   L   S   E   Q   E   L
GGTGGACTGTGACAAGCTTGATGATGGATGTGAAGGTGGATTACCAGTTAATGCCTATGAGGAAATCATCAGATT   600
  V   D   C   D   K   L   D   D   G   C   E   G   G   L   P   V   N   A   Y   E   E   I   I   R   L
AGGAGGCCTGGAATCGGAGAAGAAGTATCCTTATGATGCTGAAGATGAGAAGTGCAAGTTCACAGTGGGAGATGT   675
  G   G   L   E   S   E   K   K   Y   P   Y   D   A   E   D   E   K   C   K   F   T   V   G   D   V
GGCAGTCTACATCAACAGTTCTGTCAACATCTCTAGTAATGAGGCCGATATGGCAGCATGGTTGTACAAGAATGG   750
  A   V   Y   I   N   S   S   V   N   I   S   S   N   E   A   D   M   A   A   W   L   Y   K   N   G
ACCCATCTCTATTGGTATCAATGCTTTTGCTATGCAGTTTTACATGGGTGGAGTGTCACATCCATTCAGTTTCCT   825
  P   I   S   I   G   I   N   A   F   A   M   Q   F   Y   M   G   G   V   S   H   P   F   S   F   L
GTGTAGTCCGGATGAACTGGATCATGGTGTTCTCATAGTGGGCTATGGCACCAAGAAGGGCTGGTTCAGTGACAG   900
  C   S   P   D   E   L   D   Ĥ   G   V   L   I   V   G   Y   G   T   K   K   G   W   F   S   D   S
TCCATACTGGATCGTGAAGAACAGCTGGGGAGCGAGCTGGGGAGTACAGGGATACTACTTGGTGTATCGTGGTGA   975
  P   Y   W   I   V   K   Ñ   S   W   G   A   S   W   G   V   Q   G   Y   Y   L   V   Y   R   G   D
CGGAGTCTGTGGTCTTAACAAGATGCCCACATCTGCTATTGTCAAGTAGTTGCCCTGCAGCTTCCCAACAATTTA  1050
  G   V   C   G   L   N   K   M   P   T   S   A   I   V   K
```

```
TTTTTGTGATGCAACGATACTCCAGATAACTTGTTATATGTCAGCTGACTGGATCAGTTGTATAAAGTTTTGACT  1125
TTTTACTTTTTTTTTTAATAGTAGCTACTAATACAATATATTATGTATGTATGTATGTTGCCTATTGGCAGATAAA  1200
ATACTCTGTATGGCTAAGTACACTAATGTGAATAATACAATTAACATATATATTCTTGTGCCTTTGGTTCACTAT  1275
ATTTGAAGGTATATGTTTCAGTCAGGTCTCTTCATAATTTGCAATAACAATTCAAGTTTAGGTTATTATTATAAA  1350
GGTTTTCTGATTATATTCTGCACATGTCACATGAAACACTAATATAGAAGTGTATGTATCTTTCTATATGGGGGG  1425
AGGTCAGATCCCACATCTTTAGGTCAGGTTCCACATCTTTAGGTCAGGTTCCACATCTTTAGGTCAGGTTCCACA  1500
TCTTTAGGTCTGGTTCCACATCTTTAGGTCTGGTTCCACATCTTTAGGTCAGGTTCCACGTCTCAAGGTCTGGTT  1575
CCACATCTTTAGGTCTGGTTCCACATCTTTAGGTCTGGTTCCACATCTTTAGGTCAGGTTCCACATCTTTAGGTC  1650
AGGTTCCACATCTTTAGGTCTGGTTCCACGTCTCTAGGTCTGGTTCCACATCTTTAGGTCTGGTTCTACATCTTT  1725
AGGTCTGGTTCCACGTCTTTAGGTCTGGTTCCACGTCTTTAGGTCTGGTTCCACATCTTTAGGTCAGGTTCCACA  1800
TCTTTAGGTCTGGTTCCACATCTTTAGGTCTGGTGCCACATCTTTAGGTCTGGTTCCACATCTTTAGGTCTGGTT  1875
CCACATCTTTAGGTCTGGTTCCACATCTTTAGGTCAGGTTCCACATCTTTAGGTCTGGTTCCATGTCTCAAGGTC  1950
TGGTTCCACATCTTTAGGTCAGGTTCCGGCCATGTCTAGGTCATTTTCCATGTCGTTAGGTCTGGTTCTACATCA  2025
TTAATCTCAGACACAAAAGCTACTTTCAGAGATCACTCACAAAAGTTGAAATATGTAGTGTGTCAAACTGTAGCA  2100
GATATATGCTAAAGATTCATAATTTAGGCCAGTGTAATGTCACAGACAAATTAATTCAAATCCATATTCCGTAAA  2175
TCATATACAGCGGTATCAAGTAATGCTGTGACACTAATCAAATAAAGGTAAAACATATACTTGTAACTCAGTCTG  2250
CACACAGCCCCCATCTGTCCATCTGGAGTTTCAGTGAGCACTTAGGAAATAGGGGGGCTGTATGCAGACTGCTTG  2325
TAACACAATAATAAATGTGTGTAAATATATACAGGTAATATGAATAACAAAGATTTAACATAATGAGTCCAGACA  2400
TATCAAACATCTTTGACCTAGCTTACAAACCCAGTAATGTCATTTCCATGTTGTAAACCACACTCTGTTGTGAAT  2475
GAATCAGAGATAATAATATTTGTCAGTCCTAAATCATATTAACAACTTATTGTGATATTTTGAAAAATATGTGTT  2550
ATTCAAGAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   2591
```

Fig. 1. Nucleotide and deduced amino acid sequences of the cysteine protease cDNA clone (*PLCP*) from *Periserrula leucophryna*. The *PLCP* gene consists of a 2591 bp nucleotide sequence (CDS: 173-1024 bp) encoding for a polypeptide of 283 amino acids, composed of a 59-residue signal sequence, a 6-residue propeptide, and a 218-residue mature protein. The arrows indicate a propeptide and a probable cleavage site of the signal peptide predicted with the signal P program, respectively (Bendtsen *et al.*, 2004). The conserved amino acid residues utilized for the synthesis of degenerate primers are underlined. The active site residues, Cys[80], His[226] and Asn[250] of the catalytic triad characteristic of the C1 family of cysteine proteases are shown in the shaded boxes. The nucleotide and the deduced amino acid sequences of the *PLCP* gene were deposited into the GenBank database under the accession numbers AY390282 and AAR27011, respectively.

full-length DNA of the cysteine protease corresponding to the 6C-1 probe DNA was designated as the *PLCP* gene. The complete cDNA sequence for the *PLCP* gene and its corresponding putative open reading frame (ORF) was characterized, and is shown in Fig. 1. The entire cDNA of the *PLCP* clone was 2,591 bp (CDS: 173-1024 bp), including an 852 bp coding nucleotide sequence, a 172 bp 5'-untranslated region, and a 1567 bp 3'-noncoding region. The arrows indicate a propeptide and a probable cleavage site of the signal peptide predicted with the signal P program, respectively

(Bendtsen *et al.*, 2004). The conserved amino acid residues utilized for the synthesis of the degenerate primers are underlined. The active site residues Cys[80], His[226] and Asn[250] of the catalytic triad characteristic of the C1 cysteine protease family are indicated in the shaded boxes. In the 3'-noncoding region, the ACATCTTTAGGTCAGGTTCC sequences are specifically repeated 27 times, although the function remains unknown. The *PLCP* gene encodes for a polypeptide of 283 amino acids, which are composed of a 59-residue signal sequence, a 6-residue propeptide, and a 218-residue mature
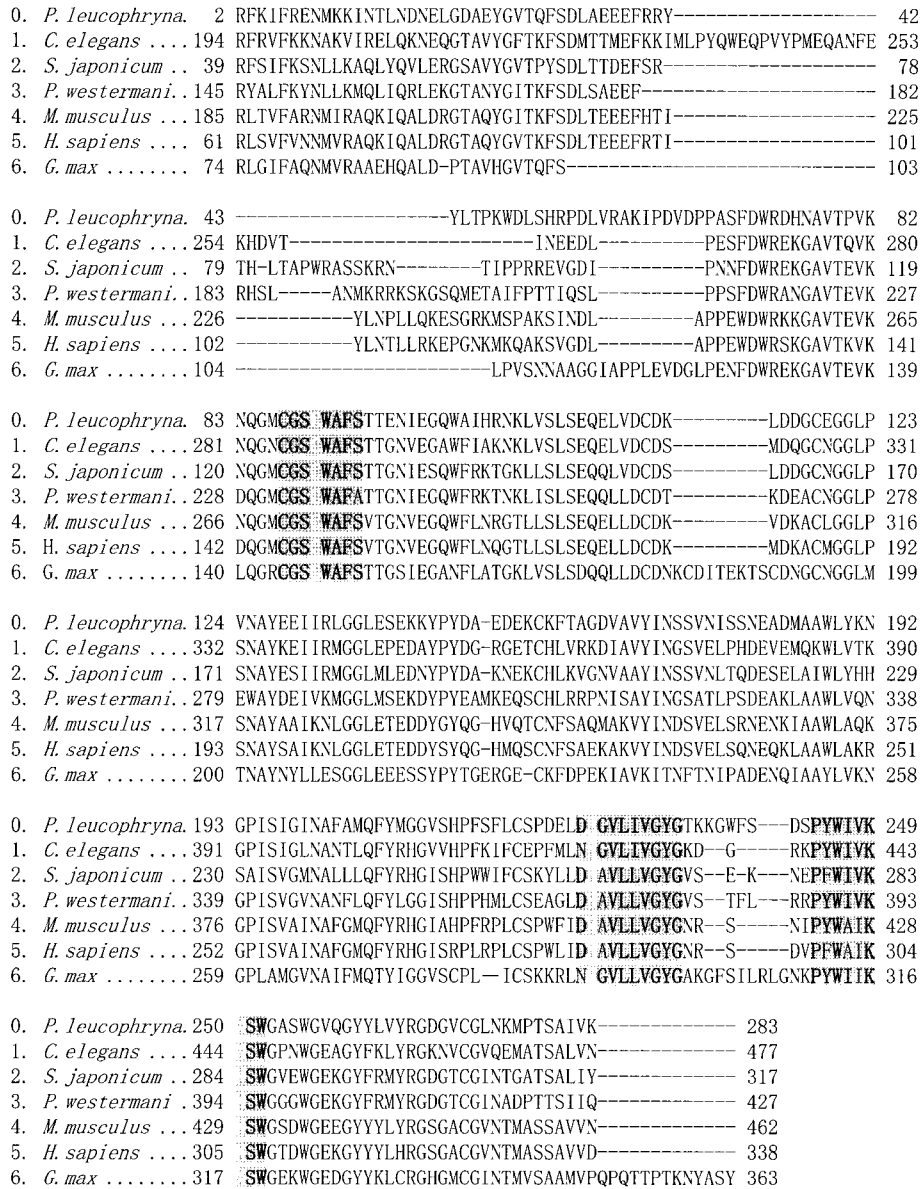
```
0.  P. leucophryna.   2 RFKIFRENMKKINTLNDNELGDAEYGVTQFSDLAEEEFRRY----------------- 42
1.  C. elegans .... 194 RFRVFKKNAKVIRELQKNEQGTAVYGFTKFSDMTTMEFKKIMLPYQWEQPVYPMEQANFE 253
2.  S. japonicum .. 39 RFSIFKSNLLKAQLYQVLERGSAVYGVTPYSDLTTDEFSR------------------- 78
3.  P. westermani. 145 RYALFKYNLLKMQLIQRLEKGTANYGITKFSDLSAEEF--------------------- 182
4.  M. musculus ...185 RLTVFARNMIRAQKIQALDRGTAQYGITKFSDLTEEEFHTI----------------- 225
5.  H. sapiens .... 61 RLSVFVNNMVRAQKIQALDRGTAQYGVTKFSDLTEEEFRTI----------------- 101
6.  G. max ........ 74 RLGIFAQNMVRAAEHQALD-PTAVHGVTQFS-------------------------- 103


0.  P. leucophryna. 43 -------------------YLTPKWDLSHRPDLVRAKIPDVDPPASFDWRDHNAVTPVK 82
1.  C. elegans ....254 KHDVT--------------------INEEDL----------PESFDWREKGAVTQVK 280
2.  S. japonicum .. 79 TH-LTAPWRASSKRN--------TIPPRREVGDI----------PNNFDWREKGAVTEVK 119
3.  P. westermani..183 RHSL-----ANMKRRKSKGSQMETAIFPTTIQSL----------PPSFDWRANGAVTEVK 227
4.  M. musculus ...226 -----------YLNPLLQKESGRKMSPAKSINDL----------APPEWDWRKKGAVTEVK 265
5.  H. sapiens ....102 -----------YLNTLLRKEPGNKMKQAKSVGDL---------APPEWDWRSKGAVTKVK 141
6.  G. max ........104 -----------------------LPVSNNAAGGIAPPLEVDGLPENFDWREKGAVTEVK 139


0.  P. leucophryna. 83 NQGMCGS WAFSTTENIEGQWAIHRNKLVSLSEQELVDCDK--------LDDGCEGGLP 123
1.  C. elegans ....281 NQGNCGS WAFSTTGNVEGAWFIAKNKLVSLSEQELVDCDS---------MDQGCNGGLP 331
2.  S. japonicum ..120 NQGMCGS WAFSTTGNIESQWFRKTGKLLSLSEQQLVDCDS---------LDDGCNGGLP 170
3.  P. westermani..228 DQGMCGS WAFATTGNIEGQWFRKTNKLISLSEQQLLDCDT---------KDEACNGGLP 278
4.  M. musculus ...266 NQGMCGS WAFSVTGNVEGQWFLNRGTLLSLSEQELLDCDK---------VDKACLGGLP 316
5.  H. sapiens ....142 DQGMCGS WAFSVTGNVEGQWFLNQGTLLSLSEQELLDCDK---------MDKACMGGLP 192
6.  G. max ........140 LQGRCGS WAFSTTGSIEGANFLATGKLVSLSDQQLLDCDNKCDITEKTSCDNGCNGGLM 199


0.  P. leucophryna.124 VNAYEEIIRLGGLESEKKYPYDA-EDEKCKFTAGDVAVYINSSVNISSNEADMAAWLYKN 192
1.  C. elegans ....332 SNAYKEIIRMGGLEPEDAYPYDG-RGETCHLVRKDIAVYINGSVELPHDEVEMQKWLVTK 390
2.  S. japonicum ..171 SNAYESIIRMGGLMLEDNYPYDA-KNEKCHLKVGNVAAYINSSVNLTQDESELAIWLYHH 229
3.  P. westermani..279 EWAYDEIVKMGGLMSEKDYPYEAMKEQSCHLRRPNISAYINGSATLPSDEAKLAAWLVQN 338
4.  M. musculus ...317 SNAYAAIKNLGGLETEDDYGYQG-HVQTCNFSAQMAKVYINDSVELSRNENKIAAWLAQK 375
5.  H. sapiens ....193 SNAYSAIKNLGGLETEDDYSYQG-HMQSCNFSAEKAKVYINDSVELSQNEQKLAAWLAKR 251
6.  G. max ........200 TNAYNYLLESGGLEEESSYPYTGERGE-CKFDPEKIAVKITNFTNIPADENQIAAYLVKN 258


0.  P. leucophryna.193 GPISIGINAFAMQFYMGGVSHPFSFLCSPDELD GVLIVGYGTKKGWFS---DSPYWIVK 249
1.  C. elegans ....391 GPISIGLNANTLQFYRHGVVHPFKIFCEPFMLN GVLIVGYGKD--G-----RKPYWIVK 443
2.  S. japonicum ..230 SAISVGMNALLLQFYRHGISHPWWIFCSKYLLD AVLLVGYGVS--E-K---NEPFWIVK 283
3.  P. westermani..339 GPISVGVNANFLQFYLGGISHPPHMLCSEAGLD AVLLVGYGVS--TFL---RRPYWIVK 393
4.  M. musculus ...376 GPISVAINAFGMQFYRHGIAHPFRPLCSPWFID AVLLVGYGNR--S-----NIPYWAIK 428
5.  H. sapiens ....252 GPISVAINAFGMQFYRHGISRPLRPLCSPWLID AVLLVGYGNR--S-----DVPFWAIK 304
6.  G. max ........259 GPLAMGVNAIFMQTYIGGVSCPL-ICSKKRLN GVLLVGYGAKGFSILRLGNKPYWIIK 316


0.  P. leucophryna.250 SWGASWGVQGYYLVYRGDGVCGLNKMPTSAIVK------------- 283
1.  C. elegans ....444 SWGPNWGEAGYFKLYRGKNVCGVQEMATSALVN------------- 477
2.  S. japonicum ..284 SWGVEWGEKGYFRMYRGDGTCGINTGATSALIY------------- 317
3.  P. westermani .394 SWGGGWGEKGYFRMYRGDGTCGINADPTTSIIQ------------- 427
4.  M. musculus ...429 SWGSDWGEEGYYYLYRGSGACGVNTMASSAVVN------------- 462
5.  H. sapiens ....305 SWGTDWGEKGYYYLHRGSGACGVNTMASSAVVD------------- 338
6.  G. max ........317 SWGEKWGEDGYYKLCRGHGMCGINTMVSAAMVPQPQTTPTKNYASY 363
```

Fig. 2. Amino acid sequence alignment of cysteine proteases from different species. The shaded boxes indicate the conserved regions of the catalytic triad. Polychaeta (*Periserrula leucophryna*), Accession no. AAR27011; Nematode (*Caenorhabditis elegans*), NP_505215; Parasite (*Schistosoma japonicum*), AAM44058; Lung fluke (*Paragonimus westermani*), AAF21461; House mouse (*Mus musculus*), NP_063914; Human (*Homo sapiens*), AAC78838; Soybean (*Glycine max*), AAD46920.

protein. The predicted molecular weights of the preproprotein and mature protein were calculated as 31.9 kDa (283 amino acids and pI 4.92) and 24.2 kDa (218 amino acids and pI 4.79), respectively. Via a search of the SMART database, the deduced amino acid sequences of the cysteine protease from *Periserrula leucophryna* was shown to be a papain-like cysteine protease. As can be observed in Fig. 2, a BLAST homology search evidenced 58% identity (for amino acid sequences compared) with *Caenorhabditis elegans* (nematode, GenBank accession no. NP505215), 58% with *Schistosoma japonicum* (parasite, Liu *et al.*, 2006), 53% with *Paragonimus westermani* (lung fluke, Park *et al.*, 2001), 53% with *Homo sapiens* (human, Somoza, 2002), 54% with *Mus musculus*

(house mouse, Deussing *et al.*, 2000) and 51% with *Glycine max* (soybean, Nong *et al.*, 1995). The PLCP protein evidences a long preproregion of 65 amino acids, including cathepsin B and L (60-100 amino acids) and human cathepsin F (251 amino acids) (Park *et al.*, 2001), thereby suggesting that PLCP may belong to the cathepsin F subgroup of enzymes, of which *P. westermani*, *Mus musculus*, and human cathepsin F are members. Finally, the *PLCP* gene was identified as a papain-like cysteine protease. Sequence analysis and alignment showed significant sequence similarity to other eukaryotic cysteine proteases, and the conserved catalytic triad of the Cys^{90}, His^{226}, and Asn^{250} residues of the deduced amino acids are indicative of the C1 family of papain-like cysteine proteases.

The nucleotide and deduced amino acid sequences of the *PLCP* gene were deposited in the GenBank database under the accession numbers AY390282 and AAR27011, respectively.

Northern blot analysis

In order to verify the tissue-specific expression of the *P. leucophryna* cysteine protease gene at the transcriptional level, total RNA was prepared from the whole body, head, gut, and skin as described in the Materials and Methods section and analyzed via Northern blotting, using the labeled partial cysteine protease gene as a probe. The quantities of total RNAs loaded were quantitated using an β-actin probe of 1.8 kb, which is constitutively expressed in all types of tissues, and were then compared with 28S ribosomal RNA. As can be seen in Fig. 3, a single positive band approximating 2.5 kb for the cysteine protease mRNA was observed specifically in all of the tested tissues, including the whole body, head, gut, and skin samples. The transcription level of the cysteine protease mRNA was higher in the head preparation than in any of the other tissues, despite the equivalent levels of 28S ribosomal RNA detected in all samples. Cathepsins B, C, F, H, L, O, and Z among the C1 family of papain-like cysteine proteases were determined to exhibit ubiquitous expression, whereas the expressions of cathepsins J, K, L2, S, and W were shown to be restricted to specific tissues or cell types (Rawlings and Barrett, 1999; Tisljar *et al.*, 1999). Accordingly, the cysteine protease from *Periserula leucophryna* may be grouped within the cathepsin F-like proteases by virtue of its mRNA expression pattern and sequence analysis, thereby indicating that this ubiquitously-expressed protease may also perform an essential function in nonspecific terminal protein degradation (Barrett, 1992). The reason for the elevated mRNA expression in the head of *Periserula leucophryna* remains to be determined.

Construction of expression vector containing the mature form of cysteine protease

In order to express the cysteine protease gene using the T7 promoter system in *E. coli*, the region of the mature form

Fig. 3. Northern blot analysis of total RNAs from various tissues of *P. leucophryna*. The blot containing 20 μg of total RNA per lane was hybridized with ^{32}p-labeled cysteine protease and the β-actin probe DNAs. The quantities of total RNAs loaded were quantitated with β-actin probes and were indirectly compared with 28S rRNA. W : Whole body; H : Head; G : Gut; S: Skin.

was amplified using two primers flanking each end of the cysteine protease gene. The forward primer harbors the initiation codon (AUG) and the *Nde*I restriction site, whereas the reverse primer harbors the termination codon and the *Sal*I site. The restriction enzyme sites were introduced in order to ease the process of cloning to the pT7-7 vector (Tabor and Richardson, 1985). The DNA fragment encoding for the mature form of cysteine protease (intracellular protein) was amplified from the *PLCP* gene using ampli-Taq Gold polymerase (Perkin-Elmer, USA) and a specific primer set using different concentrations of template DNA. The 660 bp fragment was amplified and cloned into pGEM-Teasy vector (data not shown). The putative clones were then screened via size selection and restriction enzyme digestion. After the sequences of the clones were confirmed, the proper clone harboring the mature form of the cysteine protease gene was designated the 5Cys clone, and employed in the construction of the expression vector after the amplification of the plasmid DNA in *E. coli*.

In order to introduce the amplified 5Cys clone into the pT7-7 expression vector, the vector was digested with *Nde*I and *Sal*I and the linear DNA of 2.47 kb was eluted from the gels. The 5Cys clone, which harbors the mature form of the *PLCP* gene, was digested with *Nde*I and *Sal*I, and a 660 bp DNA fragment was eluted. The resultant fragment was ligated to pT7-7 (2.47 kb), digested with the same restriction enzymes, and transferred into *E. coli* JM109. The transformants were grown in LB medium supplemented with ampicillin (50 μg/ml) and the plasmid DNAs were isolated. The putative clones were then screened principally by size selection, digested with *Nde*I and *Sal*I, and finally verified via DNA sequencing. As a result, all of the clones evidenced the correct SD sequence (AAGGAG), initiation codon, and 5'-coding regions. The resultant clone was designated the pTCys plasmid and transferred to the *E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysE strains, respectively. The expression patterns for each transformant with the identical expression vector were then evaluated with different hosts [*E. coli* BL21(DE3) and *E. coli* BL21(DE3)pLysE], media, and time intervals.
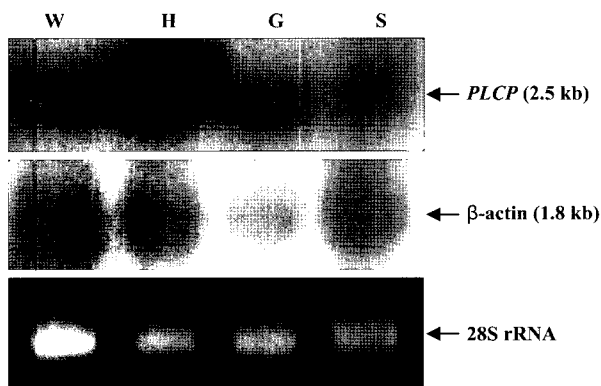
SDS-PAGE and western blot analysis

Using the T7 promoter/T7 RNA polymerase system, the cysteine protease gene was expressed. The expression of T7 RNA polymerase was induced via the addition of IPTG and the subsequent transcription of the gene under the T7 promoter was initiated. For more precise control, the host harboring the *pLys*E gene was used. This gene generates the T7 lysozyme, a natural inhibitor of T7 RNA polymerase, and thus reduces the transcription of target genes in the uninduced cells. *E. coli* BL21(DE3) or *E. coli* BL21(DE3) pLysE strains harboring the pT7Cys vector were induced at different time intervals, and the total cell lysates for each strain were analyzed via SDS-PAGE and Western blotting. As is shown in Fig. 4, all strains harboring the expression vector generated cysteine protease, whereas no protein was detected in the strain harboring the pT7-7 vector. The expression efficiencies of cysteine protease in the *E. coli* BL21 (DE3) strain were superior to that identified in the *E. coli* BL21(DE3)pLysE strain during 6 h of induction (data not
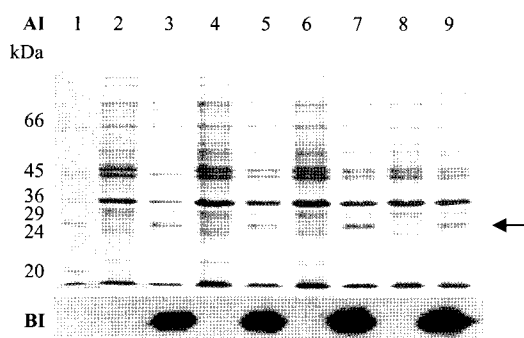
Fig. 4. SDS-PAGE and Western blot analysis of the cysteine protease. All strains harboring expression vectors were induced in LB medium over 6 h by 1 mM IPTG. The expression patterns of pT7Cys-7 transformant in *E. coli* BL21(DE3) by time intervals (2, 4, 6, 23 h) (AI). The proteins were fractionated in 12% SDS-PAGE and analyzed with synthetic polyclonal antibody. The cysteine protease was quantitated with an AlphaImager 1220 analysis system. Western blot analysis (Panel BI).
AI; Lane1: MW marker, 2: transformant pT7Cys-7/*E. coli* BL21 (DE3) (C, 2 h), 3: pT7Cys-7 (I, 2 h, 7%), 4: pT7Cys-7 (C, 4 h), 5: pT7Cys-7 (I, 4 h, 8.5%), 6: pT7Cys-7 (C, 6 h), 7: pT7Cys-7 (I, 6 h, 13.5%), 8: pT7Cys-7 (C, 23 h), 9: pT7Cys-7 (I, 23 h, 12%) (C) and (I) represent the uninduced and induced conditions. The arrows indicate the position of the cysteine protease.
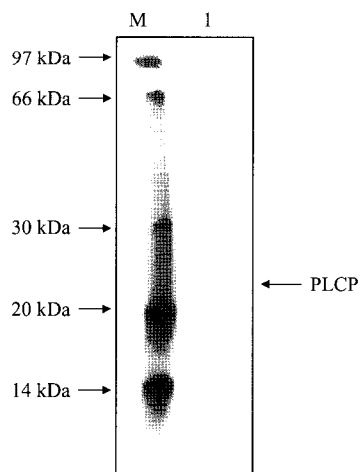
Fig. 5. SDS-PAGE of the partial purified PLCP protein. M, low molecular weight marker; lane 1, partially purified PLCP

shown), thereby indicating that the cysteine protease may be non-toxic to the growth of the strain (without T7 lysozyme), due to a leaky phenomenon occurring during the early logarithmic phase. Alternatively, the protein may not be degraded by the proteases present in cellular environments. Accordingly, it is normally worthwhile to assess several different vector/host combinations in order to generate the best possible protein yields. We compared the expression levels after 6 h of induction in *E. coli* BL21(DE3) where different eight transformants evenly harbored the same expression vector (pT7Cys). The cysteine proteases were in a range of 6.3% to 12.5% of total cell proteins in the *E. coli* BL21(DE3)

strain (data not shown). Among them, the *E. coli* BL21(DE3) strain (pT7Cys-7 transformant) evidenced elevated expression levels (12.5%), thereby indicating that each transformant originating from the same expression vector evidenced different patterns of expression. Also, the expression levels were slightly higher in the LB medium than in the M9 minimal medium (data not shown). In the time-interval experiments for the *E. coli* BL21(DE3) strain (pT7Cys-7 transformant), the expression time is directly positively related to the intensities of the bands of interest. That is, the expression level was highest at 6 h of induction time. However, in the case of 23 h of induction, the expression levels were slightly reduced. It is generally believed that the rate of protein degradation is increased by starvation of a required nutrient, or that the expressed proteins may be detrimental to cell growth. In Western blot analysis, the proteins were clearly detected using synthetic polyclonal antibody.

Although it is difficult to speculate regarding the physiological functions of the cysteine proteases at present, the characterization of the over-expressed gene product will generate biochemical information regarding the functions of this enzyme, as well as the roles that they may perform in the physiology of this organism, *Periserrula leucophryna*. Accordingly, the *E. coli* BL21(DE3) strain harboring the pT7Cys expression vector (pT7Cys-7 transformant), which evidenced the highest level of expression of the cysteine protease, was cultured, after which the recombinant cysteine protease was partially purified and characterized.

Characterization of the cysteine protease

E. coli BL21(DE3) harboring the pT7Cys plasmid (pT7Cys-7 transformant) was grown overnight at 37°C in 5 ml of LB broth supplemented with 50 μg of ampicillin/ml. The culture was then diluted 1/200 into 200 ml of the same medium, incubated, and induced. After an additional 6 h of incubation, the lysates were prepared from 200 ml of the culture. The pellet solubilization and refolding were conducted as described in the Materials and Methods section. The protein was purified using an immunoaffinity column prepared with the polyclonal antibody for the internal peptide of PLCP (KFT VGDVAVYINSSVNI). The active fractions were concentrated, divided into aliquots, and stored at -70°C until use. The partially purified PLCP was fractionated on 12% SDS-PAGE (Fig. 5). The protein was approximately 25 kDa in size. As can be observed in Fig. 6, the optimum pH (Panel A) and temperature (Panel B) of the enzyme was determined to be approximately pH 9.5 and 35°C, respectively. Accordingly, the cysteine protease was determined to be a member of the group of alkaline proteases, as the high pH optimum is a common characteristic of alkaline proteases. The enzyme activity was determined to be significantly reduced under neutral conditions, evidencing approximately 28% of its maximal activity at neutral pH. The substrate specificity of the PLCP was evaluated using a variety of synthetic substrates. As is shown in Table 1, Tos-GPK, which is known to be a specific substrate for plasmin, was the most susceptible substrate to the enzyme. Another substrate for plasmin, including Val-Leu-Lys-pNA (S-2251) was also specifically hydrolyzed by the purified protease. However, the enzyme did not efficiently hydrolyze substrates harboring non-polar amino acids
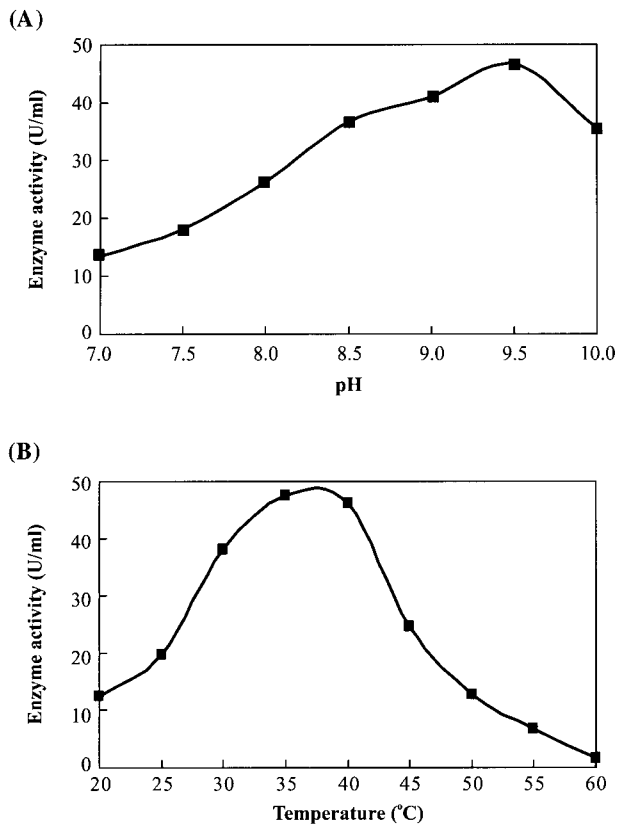
(A)

(B)

Fig. 6. Optimum pH (Panel A). The purified protease was incubated in various buffers with different pH(s) ranging from 7 to 10 and the enzyme activities were assessed (unit/ml).; Optimum temperature (Panel B). The purified protease was incubated at various temperatures ranging from 20 to 60°C under standard assay conditions.

Table 1. Substrate specificity of the purified cysteine protease

Substrate	% activity
S-2222 (BZ-Ile-Glu(γ-OR)-Gly-Arg)	3.3
S-2238 (Phe-Pip-Arg)	5.7
S-2251 (Val-Leu-Lys)	28.4
S-2266 (Val-Leu-Arg)	25.9
S-2302 (Pro-Phe-Arg)	12.7
S-2390 (Val-Phe-Lys)	18.8
S-2403 (pyroGlu-Phe-Lys)	9.8
S-2423 (Ac-Ile-Glu-Gly-Arg)	6.0
S-2765 (Z-D-Arg-Gly-Arg)	22.3
Suc-Ala-Ala-Ala	0.3
Suc-Ala-Ala-Val	0.3
Tos-Gly-Pro-Lys	100.0

at the cleavage sites, including Suc-Ala-Ala-Ala-pNA and Suc-Ala-Ala-val-pNA. According to the substrate specificity results, it was shown that the purified protease was more active towards Arg-X or Lys-X, and did not efficiently cleave substrates harboring non-polar amino acids at the P1 site.
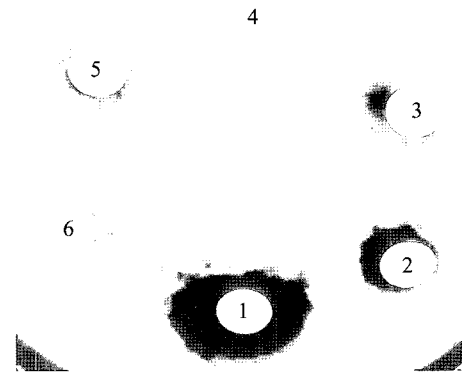

Fig. 7. Fibrinolytic activity of the partial purified PLCP on a plasminogen-free fibrin plate. 10 μl of each sample was applied to the disc after the 5 mm-paper disc was carefully positioned on the fibrin plate and incubated for 4 h at 37°C. 1: Plasmin (10 mU), 2: PLCP (1 μg), 3: PLCP (0.2 μg), 4: Trypsin (1 μg), 5: Trypsin (2 μg), 6: PBS

Table 2. Effects of various inhibitors on the activity of the purified cysteine protease

Inhibitors	Remaining activity (%)
None	100.0
Bestatin	88.3
Cystatin	29.1
PMSF	92.2
E-64	22.3
EDTA	94.3

When the fibrin as a natural substrate was utilized to determine fibrinolytic activity, PLCP evidenced a clear zone similar to that of the plasmin utilized as a positive control on the plasminogen-free fibrin plates (Fig. 7), thereby indicating that the PLCP is a protease with fibrinolytic activity. The effects of the inhibitors on the purified protease were also assessed using Tos-GPK as a substrate. The inhibitory effects on the activity of the purified protease are shown in Table 2. PLCP was sensitive to E-64 (L-carboxy-trans-2,3-epoxypropyl-leucylamido(4-guanidino) butane), which is known to be a typical cysteine protease inhibitor. The enzyme activity was efficiently inhibited by cystatin, a cysteine protease inhibitor, whereas bestatin, EDTA, and PMSF evidenced minimal inhibitory effects on the enzyme. According to these results, PLCP can be classified as a cysteine protease.

In this study, a new cysteine protease gene, designated *PLCP*, was cloned from *Periserrula leucophryna*. The sequence analysis and Northern blot data identify this as a cathepsin F subgroup enzyme. The optimum pH of the protease is approximately 9.5, and it was determined to be alkaline protease. Also, the protease specifically hydrolyzed the substrates of plasmin. However, studies regarding the optimization of the expression of the cysteine protease gene, purification by mass production, and the biological activities of the recombinant cysteine protease, will be required to make the detailed elucidation of the physiological function of the enzyme possible.

References

Allaire, M., M.M. Chernaia, B.A. Malcolm, and M.N. James. 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369, 72-76.

Astrup, T. and S. Mullertz. 1952. The fibrin plate method for the estimation of fibrinolytic activity. *Arch. Biochem. Biophys.* 40, 346-351.

Baker, E.N. and J. Drenth. 1987. The thiol proteases: Structure and mechanism, p. 313-368. Biological Macromolecules and Assemblies, 3rd (ed.). Wiley, New York, USA.

Barrett, A.J. 1992. Cellular proteolysis. An overview. *Ann. NY Acad. Sci.* 674, 1-15.

Barrett, A.J. 1994. Classification of peptidases. *Methods Enzymol.* 244, 1-15.

Bendtsen, J.D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783-795.

Berti, P.J. and A.C. Storer. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.* 246, 273-283.

Bond, J.S. and P.E. Butler. 1987. Intracellular Proteases. *Annu. Rev. Biochem.* 56, 333-364.

Brands, S.J. (comp.) 1989-2002. Systema Naturae 2000. Amsterdam, The Netherlands. [http://sn2000.taxonomy.nl/]

Chapman, H.A., R.J. Riese, and G.P. Shi. 1997. Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* 59, 63-88.

Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156-159.

Deussing, J., K. Tisljar, A. Papazoglou, and C. Peters. 2000. Mouse cathepsin F: cDNA cloning, genomic organization and chromosomal assignment of the gene. *Gene.* 251, 165-173

Gschwend, T.P., S.R. Krueger, S.V. Kozlov, D.P. Wolfer, and P. Sonderegger. 1997. Neurotrypsin, a novel multidomain serine protease expressed in the nervous system. *Molecul. Cellul. Neurosci.* 9, 207-219.

Higgins, D.G. and P.M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Bioinformatics* 5, 151-153.

Jeong, B.R., S.M. Chung, N.J. Baek, K.B. Koo, H.S. Baik, H.S. Joo, C.S. Chang, and J.W. Choi. 2006. Characterization, cloning and expression of the ferritin gene from the Korean polychaete, *Perisernula leucophryna*. *J. Microbiol.* 44, 54-63.

Joo, H.S., G.C. Park, W.R. Cho, E.S. Tak, S.R. Paik, and C.S. Chang. 2002. Purification and characterization of a prothrombin-activating protease from *Nephila clavata*. *Toxicon.* 40, 289-296.

Kim, H.G., T.N. Phan, T.S. Jang, M.J. Koh, and S.W. Kim. 2005. Characterization of *Methylophaga* sp. Strain SK1 cytochrome c$_L$ expressed in *Escherichia coli*. *J. Microbiol.* 43, 499-502.

Lacks, S. and B. Greenberg. 1977. Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* 114, 225-232.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Liu, F., J. Lu, W. Hu, S.Y. Wang, S.J Cui, M. Chi, Q. Yan, X.R. Wang, H.D. Song, X.N. Xu, J.J. Wang, X.L. Zhang, X. Zhang, Z.Q. Wang, C.L.Xue, P.J. Brindley, D.P. McManus, P.Y. Yang, Z. Feng, Z. Chen, and Z.G. Han. 2006. New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog.* 2, E29.

Nong, V.H., C. Becker, and K. Muntz. 1995. cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. *Biochim. Biophys. Acta.* 1261, 435-438.

Park, H., K.M. Hong, J.A. Sakanari, J.H. Choi, S.K. Park, K.Y. Kim, H.A. Hwang, M.K. Paik, K.J. Yun, C.H. Shin, J.B. Lee, J.S. Ryu, and D.Y. Min. 2001. *Paragonimus westermani*: Cloning of a cathepsin F-like cysteine proteinase from the adult worm. *Exp. Parasitol.* 98, 223-227.

Polgar, L. 1989. Mechanisms of Protease Action. *In* Polgar L., (ed.). CRC Press, Boca Raton, FL, USA.

Rawlings, N.D. and A.J. Barrett. 1994. Families of cysteine peptidases. *Meth. Enzymol.* 244, 461-486.

Rawlings, N.D. and A.J. Barrett. 1999. MEROPS: the peptide database. *Nucleic Acids Res.* 27, 325-331.

Sakanari, J.A., C.E. Staunton, A.E. Eakin, C.S. Craik, and J.H. McKerrow. 1989. Serine proteases from nematode and protozoan parasites: Isolation of sequence homologs using generic molecular probe. *Proc. Natl. Acad. Sci. USA* 86, 4863-4867.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *In* Molecular cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Shaw, E. 1990. Cysteinyl proteinases and their selective inactivation, p. 271-347. *In* A. Meister, Editor, Advances in Enzymology, Wiley, New York, USA.

Somoza, J.R., J.T. Palmer, and J.D. Ho. 2002. The crystal structure of human cathepsin F and its implications for the development of novel immunomodulators. *J. Mol. Biol.* 322, 559-568

Studier, F.W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37-44.

Tabor, S. and C.C. Richardson. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* 82, 1074-1078.

Tisljar, K., J. Deussing, and C. Peters. 1999. Cathepsin J, a novel murine cysteine protease of the papain family with a placenta-estricted expression. *FEBS Lett.* 459, 299-304.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene.* 33, 103-19