Appl Biochem Biotechnol (2010) 162:1858–1871
DOI 10.1007/s12010-010-8964-6

Molecular Cloning, Expression, and Characterization of Cathepsin L from Mud Loach (*Misgurnus mizolepis*)

Sang Jung Ahn · Ji Hea Sung · Na Young Kim · A Ram Lee · Soo Jin Jeon · Jung-Soo Lee · Joong Kyun Kim · Joon Ki Chung · Hyung Ho Lee

Received: 10 November 2009 / Accepted: 5 April 2010 /
Published online: 18 April 2010

Abstract Cathepsin L is an important protease in the initiation of protein degradation and one of the most powerful endopeptidases. In this study, we cloned mud loach (*Misgurnus mizolepis*) cathepsin L (*MlCtL*) cDNA, and the pro-mature enzyme of *MlCtL* (proMlCtL) was expressed in *Escherichia coli* as a fusion protein with glutathione *S*-transferase in a pGEX-4 T-1 vector. The recombinant proMlCtL was overexpressed in *E. coli* DH5αMCR as a 62-kDa protein. Its activity was quantified by measuring the cleavage of synthetic fluorogenic peptide substrates, and the protease activity of proMlCtL was also demonstrated by gelatin zymography. Antipain and leupeptin were shown to inhibit the protease activity of proMlCtL. Our results suggest that the structural features and evolutionary relationship of the mud loach cathepsin L gene were similar to that of the other mammalian cathepsin Ls; however, the proMlCtL protein was more stable at neutral and alkaline pH. The optimum temperature for the proMlCtL enzyme was found to be 40 °C. In addition, proMlCtL activity was dependent upon the presence of several metal ions and detergents.

Keywords Cysteine protease · Cathepsin L · Glutathione *S*-transferase (GST)-fusion protein · Mud loach (*Misgurnus mizolepis*) · Zymography

S. J. Ahn · J. H. Sung · A. R. Lee · J.-S. Lee · J. K. Kim · H. H. Lee (✉)
Department of Biotechnology, Pukyong National University, Busan 608-737, Korea
e-mail: hyunghl@pknu.ac.kr

S. J. Ahn
Department of Embryology, Carnegie Institution of Washington, Baltimore, MD 21218, USA

N. Y. Kim · J. K. Chung
Department of Aquatic Life Medicine, Pukyong National University, Busan 608-737, Korea

S. J. Jeon
Department of Pathology, Stony Brook University Medical Center, Stony Brook, NY 11794-8691, USA

Introduction

Fisheries wastes are increasingly becoming one of the most serious and dangerous sources of environmental contamination. It has been estimated that more than 50% of the remaining material from total fish capture is not used as food and is discarded as processing waste or by-products [1]. An interesting alternative to simply discarding fish waste is to hydrolyze the waste to obtain fish protein hydrolysates (FPHs), which contain proteins that have desirable functional properties. FPHs have been obtained by different methods such as utilization of acids [2], organic solvents, exogenous proteases [3–5], or enzymatic autolysis [1, 6, 7]. Among them, autolytic degradation of proteins in fish and meat has received wide attention due to its applicability in modifying and increasing the quality attributes of food and in the preparation of value-added products such as protein hydrolysates, sauce, and silages [8–10]. The most abundant commercial proteases used for the hydrolysis of fish proteins are both from plant sources such as papain [1, 11] and from animal sources such as pepsin [12], chymotrypsin and trypsin [13].

During enzymatic autolysis, degradation of intracellular proteins and breakdown of extracellular proteins are mainly due to lysosomal cysteine proteinases called cathepsins, which belong to the papain superfamily (clan CA, family C1) [14, 15]. There are at least 11 different types of papain-type cysteine protease (cathepsins B, H, L, S, C, K, O, F, V, X, and W), and their classification in these groups depends on their enzymatic characteristics. Among them, cathepsin L is an important proteinase in the initiation of protein degradation and one of the most powerful endopeptidases [16]. Recently, cathepsin L has been implicated in major histocompatibility complex (MHC) class II-associated invariant chain (Ii) degradation and T-cell development [17]. Cathepsin L is capable of hydrolyzing a broad range of proteins including myosin, actin, nebulin, cytosolic proteins, collagen, and elastin [18]. Cathepsin L is a major proteinase that degrades myofibrillar proteins in antemortem or postmortem muscle of chum salmon (*Oncorhynchus keta*) [19], spotted mackerel (*Scomber japonicus*) [20], and Pacific whiting (*Merluccius productus*) [21].

Mud loach (*Misgurnus mizolepis*; Cypriniformes, Cobitidae) is a freshwater fish found mainly in Korea and China [22] and is often used as a restorative food (choo-o-tang; Korean traditional loach soup). This species has several attractive features for genetic studies such as a relatively small body size, transparent embryos, rapid embryonic development, short generation time, superior tolerance to low oxygen tension and diseases, year-round spawning under controlled conditions, and high fecundity [23, 24]. To better understand the role of cathespin L in fish and its potential applications, we attempted to identify a cathepsin L in mud loach *M. mizolepis*, using cathepsin-L-specific PCR primers. We succeeded in cloning the cDNA of cathepsin L, determined its structural and phylogenetic features, and recombinantly expressed it in *Escherichia coli*. A recombinant GST-tagged pro-mature enzyme was purified and the biochemical properties of the recombinant mud loach cathepsin L were characterized.

Materials and Methods

mRNA Isolation and Mud Loach Muscle cDNA Library Construction

Poly $(A)^+$ RNA was isolated from *M. mizolepis* muscle using the PolyATtrack® System 1,000 (Promega), according to the manufacturer's instructions. cDNA was synthesized from Poly$(A)^+$ RNA using the λ ZAP cDNA synthesis kit (Stratagene) and size fractionated into

a>0.4-kb pool. The cDNA pool was cloned into the Uni-ZAP vector containing the whole phagemid pBluescript SK sequence (Stratagene) and packed into a λ phage extract (Gigapack gold extract, Stratagene). The cDNA library was constructed with 2.75×10^6 plaque-forming units/μl.

Cathepsin-L-Specific Primers Design Using Multiple Alignment Analysis

Homologous cathepsin L sequences identified by BLASTX and BLASTP searches of the GenBank™ database were aligned manually using the BioEdit Sequence Alignment Editor version 5.0.9 [25]. The degenerated cathepsin L gene-specific (GS) primers were designed according to the multiple amino acid and nucleotide sequences alignment of cathepsins between different species, which were highly conserved protein sequences (HWEQWK, CGSCWAF, FEGEDVDG, and WIVKNSW). In order to identify mud loach cathepsin L, four degenerated oligonucleotides, CtL-F1 (5′-CABTGGVANCWVTG GAAGA-3′), CtL-F2 (5′-TGYGGTKCWTGCTGGRSYTT-3′), CtL-R1 (5′-CCARCTGT TYTTNAYNADCCA-3′), and CtL-R2 (5′- CCATCVACATCYTCDCCYTCAAA-3′), were employed in the amplification of cDNAs from an *M. mizolepis* muscle cDNA library.

cDNA Cloning for the Complete Coding Sequence of *MlCtL*

After the second PCR amplification of cDNA from an *M. mizolepis* muscle cDNA library using CtL-F2 and CtL-R2 primers, we obtained a 479-bp fragment. On the basis of the partial mud loach cathepsin L cDNA clone, we performed 3′RACE-PCR using the GeneRacer™ Kit (Invitrogen). Total RNA was isolated from *M. mizolepis* with the SV total RNA isolation system (Promega), in accordance with the manufacturer's instructions. Mud loach full-length cDNA was synthesized from total RNA with the GeneRacer™ kit (Invitrogen) and SuperScript™ II RT (Reverse transcriptase, Invitrogen). The 3′RACE sequence was amplified using gene-specific primers (sense primer, MCL-F1: 5′-GCCTACCCCTACTTGGGAACTGATG-3′) and GeneRacer™ primers (antisense primer, GeneRacer™ 3′primer: 5′-GCTGTCAACGATACGCTACGTAACG-3′, and GeneRacer™ 3′Nested Primer: 5′-CGCTACGTAACGGCATGACAGTG-3′). To isolate the full-length mud loach cathepsin L, PCR amplification was performed with the *M. mizolepis* muscle cDNA library using the specific primer (antisense primer, MCL-R1: 5′-GTCAA CAAATCCTGTGTCGTTGGCTGC-3′) and T3 and M13 Reverse vector primers. The primers for 5- and 3-end amplification were designed according to the DNA sequence obtained after sequencing of the partial amplified PCR product. The amplified products were subcloned into pGEM®-T Easy vector (Promega), using the *E. coli* strain DH5αMCR as the host bacterium. Plasmid DNA was prepared from *E. coli* using an AccuPrep® Plasmid DNA Extraction Kit (Bioneer). The *E. coli* clones containing the recombinants were overlaid with 100 μg/ml of ampicillin and 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) was added to the agar-gel plate and incubated at 37 °C overnight. White colonies were randomly chosen, cultivated, and stored in freezing medium at −80 °C. The plasmid DNA was extracted using an AccuPrep Plasmid Extraction Kit (Bioneer). DNA sequencing was conducted using M13 Forward/Reverse primers, with the ABI 3730xl DNA Sequencer from Macrogen Inc. (Korea). Finally, using the combination of the DNA sequences of the cDNA library screening clones and the 3′ RACE product, the full-length mud loach cathepsin L cDNA (*MlCtL*) sequence was generated.

Sequence Analysis and Phylogenetic Analysis

Nucleotide and predicted peptide sequences were analyzed using DNAsis for Windows version 2.5 (Hitachi software engineering), BioEdit Sequence Alignment Editor, and BLAST programs in the nonredundant databases of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). Multiple alignment of amino acid sequences was analyzed using the CLUSTAL W version 1.8 [26]. *MlCtL* signal sequences and putative cleavage sites were identified using the software servers listed on the ExPASy web site (http://us.expasy.org/tools/). Predictions of the pro-region cleavage sites, as well as the active sites, were based on alignment of the cathepsin protein sequences with the vertebrate orthologs. The identity and homology between protein sequences were analyzed using BioEdit Sequence Alignment Editor version 5.0.9. Phylogenetic analysis was carried out using MEGA version 3.0 [27], and the phylogenetic tree was constructed using the neighbor-joining methods.

Expression and Purification of Recombinant proMlCtL in E. coli

To prepare an expression vector suitable for production of recombinant fish cathepsin L in *E. coli*, we first generated a 963-bp DNA fragment containing the coding sequence for the *proMlCtL* by PCR amplification of the *M. mizolepis* pro-mature cathepsin L (*proMlCtL*). The primers (proMlCtL-*Eco*F, 5′-GGCCGAATTCGCACCAAGTTTAGAC-3′; proMlCtL-*Xho*R, 5′-CCGGCTCGAGCTAGACAAGTGGATAGC-3′) contained *Eco*RI and *Xho*I restriction sites, allowing cloning of the amplified DNA in a predicted orientation into pGEX-4 T-1 (Amersham Pharmacia Biotech). Recombinant plasmids (*proMlCtL/pGEX*) were transformed into the *E. coli* strain DH5αMCR. Transformed cells were grown in LB broth (100 ml) containing 100 mg/ml ampicillin at 37 °C for about 16 h, diluted 1/100 with the same medium, and grown to an A_{600} of 0.6. IPTG was then added to a final concentration of 0.4 mM, and the incubation was continued for 3 h. Cells were collected by centrifugation, washed, and resuspended in 0.2 volumes of phosphate-buffered saline (PBS). The cells were then lysed using a sonicator (Vibra cell, Sonics & Materials Inc., USA) at a setting of 40% and centrifuged at 20,000×*g* for 20 min at 4 °C. The soluble supernatant was subjected to a glutathione–Sepharose 4B column (Pharmacia Biotech Co., USA) that had been equilibrated with PBS. After washing the column with equilibration buffer, protein was eluted in elution buffer, 50 mM Tris, pH 8.0, and 10 mM reduced glutathione (Sigma). The fractions containing sufficient amounts of active enzyme were pooled, and then dialyzed and concentrated using Centricon 10 concentrators (Amicon). The purified proMlCtL protein was subjected to SDS-PAGE, Western blotting, and enzyme activity assay.

SDS-PAGE, Western Blotting, and Zymography

Electrophoresis procedures were performed as described by Laemmli [28]. All samples were denatured in buffer containing 60 mM Tris, pH 6.8, 25% glycerol, 2% SDS, 14.4 mM 2-mercaptoethanol, and 0.1% bromophenol blue; boiled for 5 min; and separated by 12% SDS-PAGE (Bio-Rad, USA). Prestained molecular weight markers (Bio-Rad, USA) were run as standards on each gel. Following electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250.

The molecular mass and purity of the proMlCtL protein were analyzed via Western blotting. Prestained molecular weight markers (Bio-Rad, USA) were run as standards. The

electrophoresed samples were transferred to nitrocellulose membranes (Schleicher & Schuell Co., USA) using a Hoefer transblotting system (Pharmacia. Co., USA). After transfer, the membrane was blocked with 3% BSA in TTBS (200 mM Tris, pH 7.0, 1.37 M NaCl, and 1% Tween 20) for 1 h at room temperature. The membrane was incubated with monoclonal anti-GST antibody (Santa Cruz Biotechnology, Inc., USA) at 4 °C overnight, rinsed and washed as before, and then incubated with phosphatase-labeled goat anti-mouse IgG antibody (1:1,000 dilution, Kirkegaard Perry Laboratories, USA) for 90 min at room temperature. The membrane was washed and rinsed as before, and the expressed proteins were visualized using an AP conjugation kit (Kirkegaard Perry Laboratories).

Zymography was conducted to determine cathepsin activity and the molecular mass of the recombinant proMlCtL, as previously described with slight modifications [29]. Electrophoresis was performed at 4 °C, using 12% polyacrylamide gels containing gelatin, bovine hemoglobin, azocasein, and bovine serum albumin (0.1% w/v) (Sigma). After electrophoresis at 4 °C, the SDS was removed from the gel by incubating it in 2.5% Triton X-100 for 1 h at room temperature. The gel was then incubated in 0.1 M sodium acetate buffer (pH 7.5) with 2 mM DTT for 18 h at 37 °C. Gels were stained with 0.1% Coomassie Brilliant Blue R-250 and then destained. The protease activity was detected as a clear zone against the blue background.

Enzyme Activity Assays

Cathepsin L activity was determined by evaluating the cleavage of specific fluorogenic substrates using a modified method of Barrett and Kirschke [30]. Briefly, 10 μl affinity-purified proMlCtL enzyme was preincubated at 37 °C for 2 h with 80 μl of 100 mM sodium acetate buffer at different pH values together with 2 mM DTT. The enzyme reaction was initiated by adding 10 μl of 1 mM Z-Phe-Arg-7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-AMC; Sigma) at 37 °C for 10 min. Substrate specificities were investigated using Z-Gly-Gly-Arg-AMC (Sigma), Z-Arg-Arg-AMC (Sigma), Z-Gly-Pro-Arg-AMC (Sigma), Z-Leu-Leu-Glu-AMC (Sigma), and Ala-Ala-Phe-AMC (Sigma) with 100 mM sodium acetate (pH 7.5) and 2 mM DTT. Substrates were added to a final concentration of 100 μM. The optimum temperature for the proMlCtL enzyme was determined by varying the incubation temperatures between 4 °C and 80 °C using 100 mM sodium acetate buffer (pH 7.5). The mud loach cathepsin L activity was then determined using Z-Phe-Arg-AMC as described above. The 7-amido-4-methylcoumarin (AMC) was measured using a Microplate Fluorometer (Packard Co. USA) at an excitation wavelength of 380 nm and emission wavelength of 460 nm.

Effect of Enzyme Inhibitors, Metal Ions, and Detergents

The effects of enzyme inhibitors on protease activity were studied using *trans*-epoxysuccinyl-L-leucyl-amido(4-guanidino) butane (E-64; Sigma), antipain (Sigma), leupeptin (USB Co., USA), pepstatin A (Sigma), phenylmethylsulfonyl fluoride (PMSF; Sigma), chymostatin (Sigma), ethylene diamine tetraacetic acid (EDTA; Sigma), 1,10-phenanthroline (Sigma), ethylene glycol-bis(β-aminoethyl ether)*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA; United States Biochemical Corp.), *N*-ethylmaleimide (NEM; Sigma), and aprotinin (Sigma). The effect of various metal ions (1 and 5 mM) on enzyme activity was investigated using $ZnSO_4$, $CuSO_4$, $CoCl_2$, KCl, $MgSO_4$, $CaCl_2$, and $HgCl_2$. The effects of some surfactants (Triton X-100, Tween 20, and SDS) on enzyme stability were also studied. The effects of enzyme inhibitors, metal ions, and detergents (0.01% and 0.05%) on the activity of proMlCtL protease were studied at pH 7.5 and 37 °C.

Statistical Analysis

All experiments were conducted three times. The values shown in the tables and graphical data represent the mean of three assays (±standard deviation). All analyses were performed using Microsoft Excel and SigmaPlot software package.

Results and Discussion

Molecular Cloning and Characterization of Cathepsin L cDNA from Mud Loach

Full-length mud loach cathepsin L (*MlCtL*) cDNA was obtained by combining DNA sequences of cDNA library clones and the 5′ GeneRACE PCR product. The 1,379-bp *MlCtL* cDNA contained an open reading frame of 1,014 bp from the first ATG start codon to the TAG stop codon flanked by the 28-bp 5′-untranslated region (UTR) and the 337-bp 3′-UTR. The *MlCtL* sequence was deposited in the GenBank database under the accession no. EF535003. The 1,379-bp *MlCtL* cDNA contained a putative signal peptide of 17 amino acids, which was analyzed with SignalP 3.0 Server, a 98-residue propeptide, and a 222-residue mature enzyme (Fig. 1).

Figure 2 shows alignment of *MlCtL* with sequences of other cathepsin Ls. The pre-region was composed predominantly of hydrophobic residues, which was characteristic of a signal peptide [31] and was predicted to be cleaved after phenylalanine (F) preceding a leucine (L) or alanine (A) residue (Fig. 1). Essentially all papain-like cysteine proteases are synthesized as an inactive proenzyme. Their activation occurs through proteolytic cleavage of the N-terminal fragment of the enzyme, the so-called proregion. The cleavage site between the proregion and the mature protease, as shown in Fig. 1, was identified by aligning *MlCtL* with sequences of other piscine and mammalian cathepsin Ls. The amino acid sequence of *MlCtL* contained an ERF/WNIN motif, which is highly conserved in the interspersed cathepsin L family. In addition, *MlCtL* also contained the GNFD motif in the proregion, which is generally conserved in most of the cysteine proteases in the papain superfamily [15, 32] (Fig. 2). The presence of the ERFNIN/GNFD motif in the proregion of *MlCtL* clearly indicated that this protein was related to the cathepsin L group and was separate from the cathepsin B subfamily. The propeptide of mammalian cathepsins has been shown to be essential for proper folding of the enzyme, for stabilizing its structure upon exposure to pH changes, and for microsomal and lysosomal targeting [15].

All cysteine proteases have a conserved active site consisting of a cysteine, a histidine, and an asparagine residue. The cysteine residue (Cys-25 based on papain numbering) is embedded in a highly conserved peptide sequence, CGSCWAFS. The histidine residue (His-159; papain numbering) is adjacent to small amino acid residues such as glycine or alanine and asparagine (Asn-175; papain numbering) is part of the Asn-Ser-Trp motif. Like all cysteine protease members, the protein structure encoding *MlCtL* contained all of these features.

Phylogenic and Sequence Analysis of Mud Loach Cathepsin L

By comparing the *MlCtL* with other cathepsins, we found that *MlCtL* had a high degree of identity with other piscine cathepsin Ls (90–82%). Also, *MlCtL* had a high degree of identity with other amphibian and mammalian cathepsin Ls (75–62%). However, *MlCtL* had a lower degree of identity (8–19%) with human cathepsin A, B, C, D, and Z. A

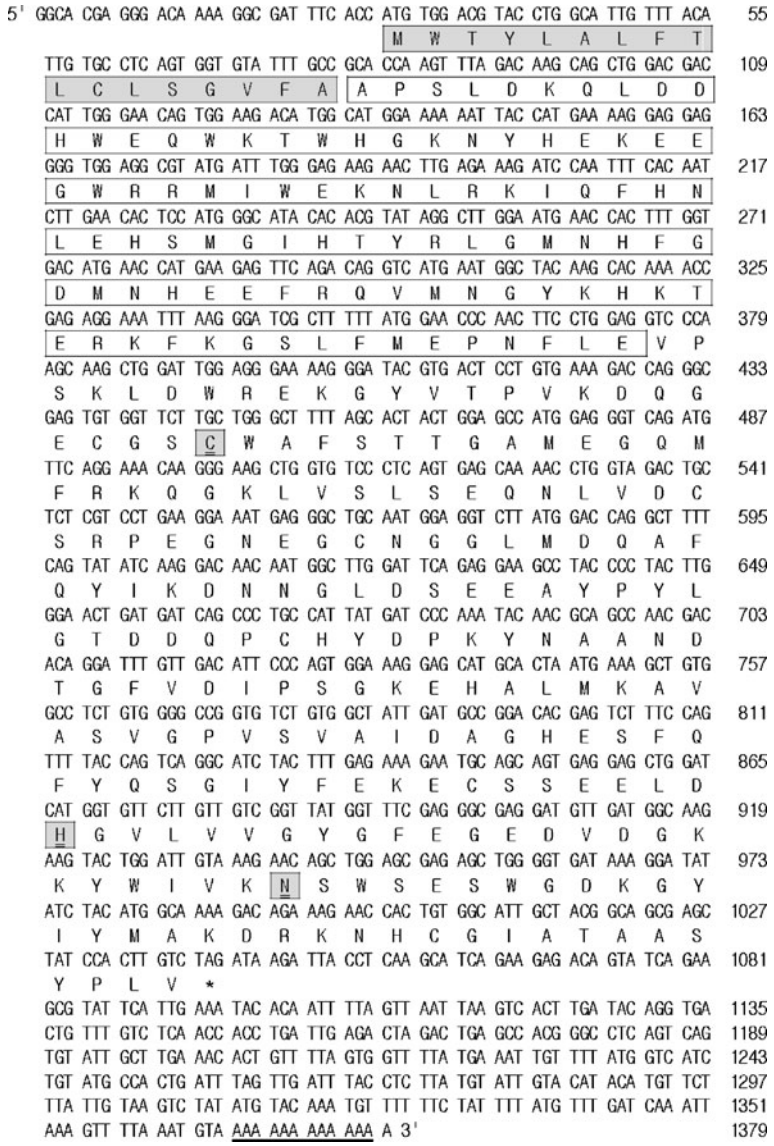

Fig. 1 Nucleotide and deduced amino acid sequence of mud loach Cathepsin L cDNA (*MlCtL*). The *shaded box* and the *open box* in the amino acid sequence indicate the putative signal peptide (pre) and the pro peptides of *MlCtL*, respectively. The active-site triad residues, Cys-25, His-164, and Asn-180, are indicated by the *double underline* and *shaded box*. The *asterisk* (*) at the end of the amino acid sequences shows the stop codon and the poly (A) tail is *underlined*

phylogenetic tree analysis based on the amino acid sequences of mud loach cathepsin L and other members of the cathepsin family showed that *MlCtL* clustered with piscine cathepsin Ls (Fig. 3). Based on a comprehensive phylogenetic analysis, the enzymes of the Family C1 peptidases (i.e., the papain superfamily of cysteine proteases) could be divided into two main evolutionary branches, A and B [33]. Branch A includes cathepsins B, C, and Z, while branch B includes the cathepsin L-like enzymes, a group consisting of papain,

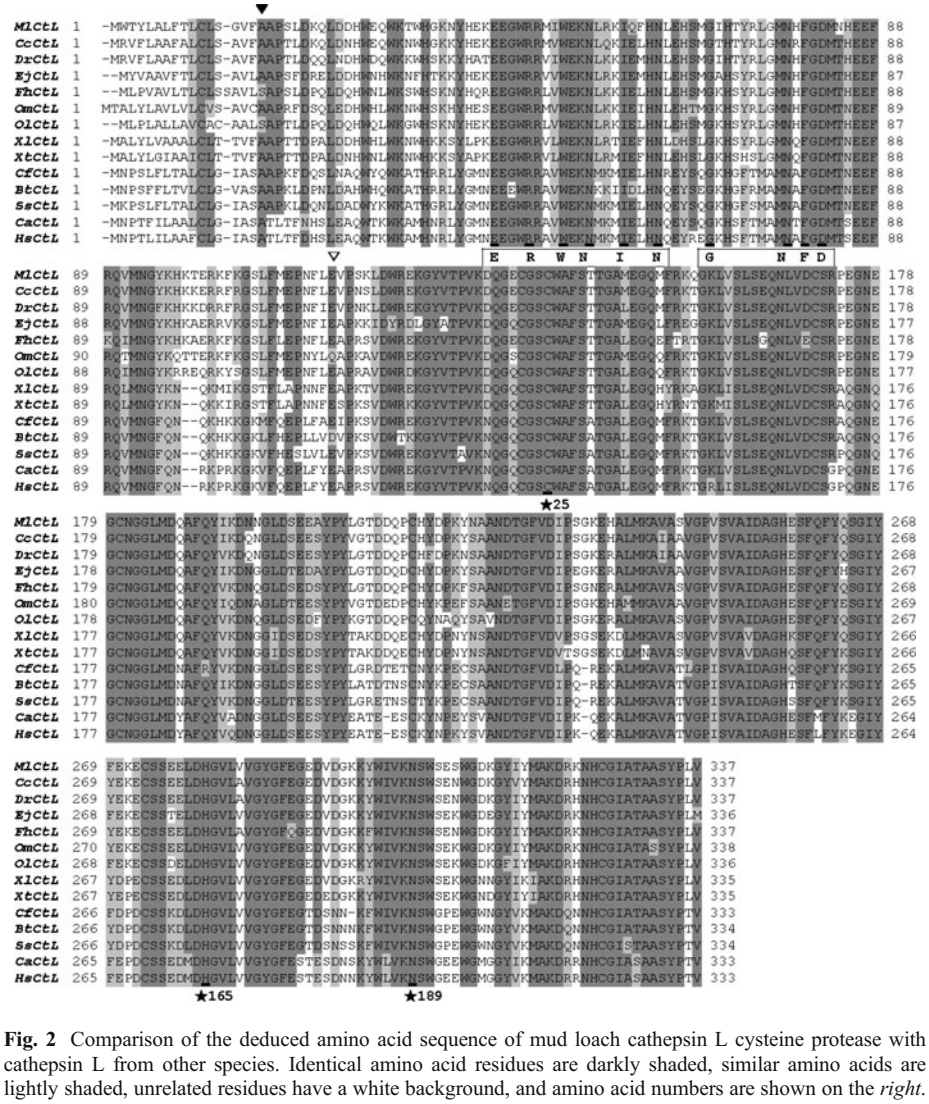

Fig. 2 Comparison of the deduced amino acid sequence of mud loach cathepsin L cysteine protease with cathepsin L from other species. Identical amino acid residues are darkly shaded, similar amino acids are lightly shaded, unrelated residues have a white background, and amino acid numbers are shown on the *right*. The conserved inter-spaced motif in the proregion, Glu, X3, Arg, X2, (Ile/Val), Phe, X2, Asn, X3, Ile, X3, Asn ('ERF/WNIN'; named after the single letter code for amino acids; X is any amino acid) and the GNFD motifs are indicated on bottom of the alignment. A *filled arrowhead* indicates a putative signal cleavage site in cathepsin Ls and an *open arrowhead* indicates a putative mature enzyme in cathepsin Ls. The cysteine residue of *MlCtL* is embedded in a highly conserved peptide sequence, CGSCWAFS. The cysteine proteinase catalytic triad residues (C, H, and N) are shown in the *asterisk* (*). The alignment was created using the BioEdit Sequence Alignment Editor version 5.0.9 using the Clustal W algorithm. Aligned amino acid sequences extracted from the GenBank/EMBL/DDBJ (accession number in parentheses) are *CcCtL*, *C. carpio* cathepsin L (BAD08618); *DrCtL*, *Danio rerio* cathepsin L (NP_997749); *EjCtL*, *Engraulis japonicus* (Japanese anchovy) cathepsin L (BAC16538); *FhCtL*, *Fundulus heteroclitus* cathepsin L (AAO64471); *OmCtL*, *Oncorhynchus mykiss* (rainbow trout) cathepsin L (AAK69706); *OlCtL*, *Oryzias latipes* (Japanese medaka) cathepsin L (BAD27581); *XlCtL*, *Xenopus laevis* (African clawed frog) cathepsin L (AAH80004); *XtCtL*, *Xenopus tropicalis* (Silurana tropicalis) cathepsin L (AAH74718); *CfCtL*, *Canis familiaris* (dog) cathepsin L (CAC08809); *BtCtL*, *Bos taurus* (cow) cathepsin L (AAI02313); *SsCtL*, *Sus scrofa* (pig) cathepsin L (CAC44793); *CaCtL*, *Cercopithecus aethiops* (African green monkey) cathepsin L (AAG35605); *HsCtL*, *Homo sapiens* cathepsin L (NP_666023)

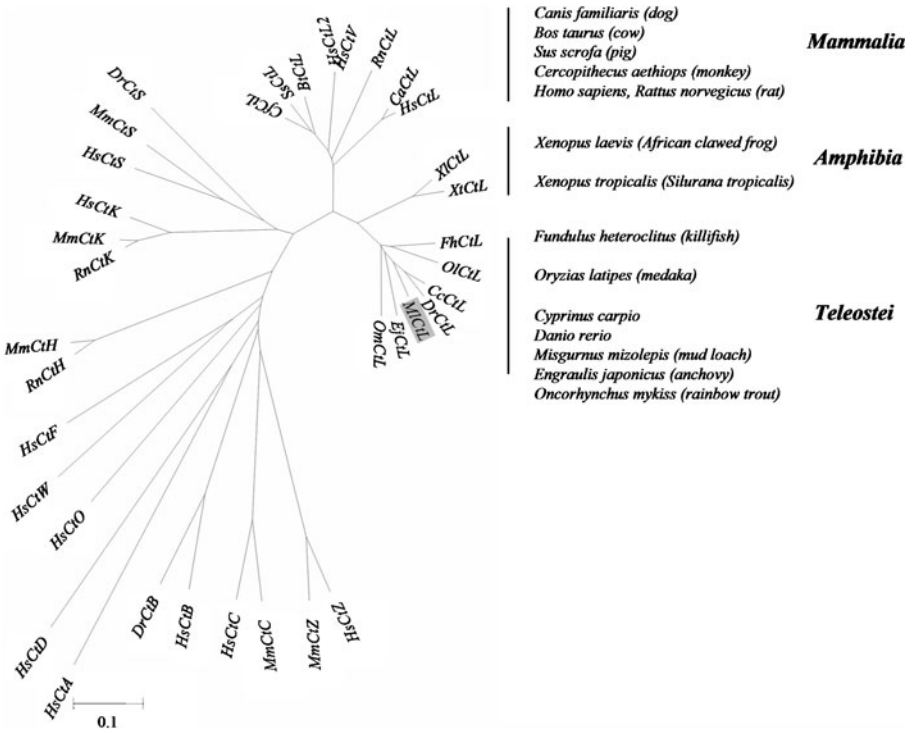

Fig. 3 Phylogenetic relationships of *MlCtL* among representative eukaryotic groups based on the cathepsin genes. In this neighbor-joining phylogram, all individuals are represented and the branches are based on the number of inferred substitutions as indicated by the bar. *MlCtL* is darkly shaded. Abbreviated amino acid sequences and GenBank accession numbers not previously listed are as follows (accession number in parentheses): *HsCtL2*, *H. sapiens* cathepsin L2 (CAI15053); *HsCtV*, *H. sapiens* cathepsin V (BAA25909); *RnCtL*, *Rattus norvegicus* cathepsin L (AAH63175); *DrCtS*, *D. rerio* cathepsin S (AAH95788); *MmCtS*, *Mus musculus* cathepsin S (CAA05360); *HsCtS*, *H. sapiens* cathepsin S (NP_004070); *HsCtK*, *H. sapiens* cathepsin K (NP_000387); *MmCtK*, *M. musculus* cathepsin K (NP_031828); *RnCtK*, *R. norvegicus* cathepsin K (NP_113748); *MmCtH*, *M. musculus* cathepsin H (NP_031827); *RnCtH*, *R. norvegicus* cathepsin H (AAH85352); *HsCtF*, *H. sapiens* cathepsin F (Q9UBX1); *HsCtO*, *H. sapiens* cathepsin O (NP_001325); *HsCtD*, *H. sapiens* cathepsin D (AAP35556); *HsCtA*, *H. sapiens* cathepsin A (P10619); *DrCtB*, *D. rerio* cathepsin B (NP_998501); *HsCtB*, *H. sapiens* cathepsin B (AAH10240); *HsCtC*, *H. sapiens* cathepsin C (AAQ08887); *MmCtC*, *M. musculus* cathepsin C (NP_034112); *MmCtZ*, *M. musculus* cathepsin Z (NP_071720); *HsCtZ*, *H. sapiens* cathepsin Z (AAC39839)

cathepsin L, cathepsin S, cathepsin K, cathepsin H, and cathepsin F. *MlCtL* was more closely related to the cathepsin L subfamily (L, S, K, and H) than to the cathepsin B, Z, A, C, and D subfamilies. Although piscine cathepsin Ls have a structure that is similar to amphibian and mammalian cathepsin Ls, they differ from Mammalia, Amphibia, and Teleostei cathepsin Ls. It is possible that piscine cathepsin L evolved early and independently into the ancestral genome of Teleostei.

Enzymatic Characterization of the Recombinant proMlCtL

The *MlCtL* (*proMlCtL*) gene was expressed in *E. coli* as a fusion protein with the 26-kDa glutathione *S*-transferase (GST). From the SDS-PAGE analysis, the recombinant fusion protein had a molecular weight of 62 kDa, which was close to expected size (Fig. 4a). The

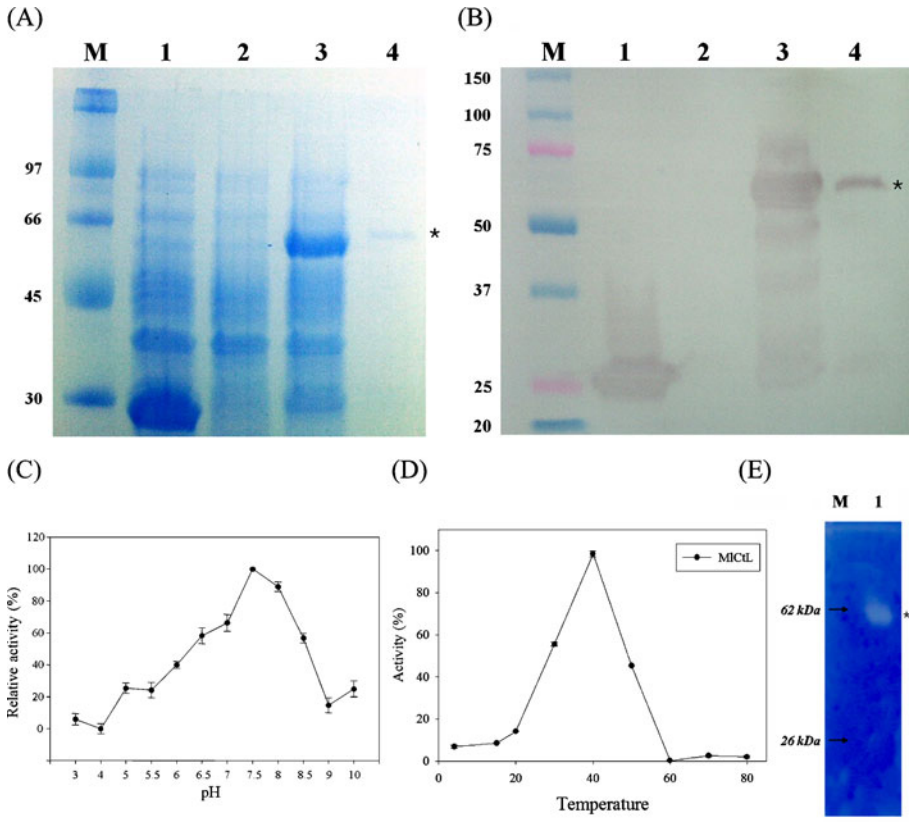

Fig. 4 a Expression and purification of recombinant proMlCtL fused with GST expressed in *E. coli* and Western blotting. Protein samples were separated by SDS-PAGE (12%) and visualized by Coomassie R-250 blue staining and destaining. *Arrows* indicate GST fused with proMlCtL (62 kDa). *Lane M*, standard size marker; *lane 1*, overexpressed GST protein (37 °C); *lane 2*, non-induced proMlCtL protein; *lane 3*, overexpressed GST-fused proMlCtL (37 °C); *lane 4*, glutathione–Sepharose 4B affinity column-purified proMlCtL. b *M*, prestained protein size marker; *lane 1*, expressed GST protein (37 °C) reacted with monoclonal anti-GST antibody (positive control); *lane 2*, non-induced proMlCtL protein (negative control); *lane 3*, overexpressed GST-fused proMlCtL (37 °C); *lane 4*, glutathione–Sepharose 4B affinity column-purified proMlCtL. c pH dependency of proMlCtL. The purified proMlCtL was preincubated at 37 °C in buffers of varying pH for 2 h. d Determination of temperature optimum of proMlCtL. e Gelatin zymography of proMlCtL. *Lane M*, purified GST fusion protein (4 μg); *lane 1*, GST-affinity column-purified proMlCtL

overproduced soluble GST-fusion protein (proMlCtL) was then purified by glutathione–Sepharose 4B column chromatography and the proMlCtL fusion protein band had a high purity and the correct size by SDS-PAGE and Western blot analysis (Fig. 4). The activity of the recombinant proMlCtL was enzymatically characterized using fluorogenic and protein substrates. The purified proMlCtL activity was quantified by measuring the cleavage of a synthetic fluorogenic peptide substrate, Z-FR-AMC. In previous reports, Z-FR derivatives were shown to be efficiently hydrolyzed by both cathepsin B and L, whereas Z-RR derivatives were found to be efficiently hydrolyzed by only cathepsin B [30], although to a lower degree compared with Z-FR-AMC hydrolysis. The proMlCtL enzyme was active over a wide pH range (pH 6–10) with optimal activity at pH 7.5 (pH 7.2, cytosolic pH); however, its activity was greatly decreased at pH 5 (lysosomal pH) and below (Fig. 4c). The effect of

Table 1 Substrate specificity of mud loach cathepsin L.

Substrates	Concentration (μM)	Activity (%)
Z-Phe-Arg-AMC (FR)	100	100
Z-Arg-Arg-AMC (RR)	100	88.77±1.45
Z-Gly-Pro-Arg-AMC (GPR)	100	28.61±1.00
Z-Leu-Leu-Glu-AMC (LLE)	100	76.66±5.75
Ala-Ala-Phe-AMC (AAF)	100	28.66±0.99
Z-Gly-Gly-Arg-AMC (GGR)	100	ND

ND not detected

temperature on proMlCtL activity was shown in Fig. 4d, and the optimum temperature for the proMlCtL enzyme was found to be 40 °C. On gelatin-zymography, the purified proMlCtL was capable of hydrolyzing 0.1% gelatin at pH 7.5 (Fig. 4e). Generally, mammalian lysosomal cysteine proteases are relatively unstable at neutral pH, but the recombinant *MlCtL* protein was more active and remained stable at neutral and alkaline pH. In contrast, the optimal pH values of cathepsin L purified from different fish species, such as chum salmon (*O. keta*) [34, 35], carp (*Cyprinus carpio*) [36, 37], silver carp (*Hypophthalmichthys molitrix*) [38], anchovy (*Engraulis japonica*) [39] and arrowtooth flounder [40] have been reported to be between 5.0 and 6.5. In addition, the optimal pH for human procathepsin L was reported to be below pH 5.5 [41]. However, mouse cathepsin P [42] and rat, dog, and monkey cathepsin S [43, 44] were reported to have pH activity profiles similar to *MlCtL*.

The substrate specificity of the recombinant proMlCtL was assayed using the following fluorogenic peptides: Z-Arg-Arg-AMC (Z-RR-AMC) for cathepsin B activity, Z-Leu-Leu-Glu-AMC (Z-LLE-AMC) for peptidyl glutamyl peptide hydrolase (PGPH) activity, Z-Gly-

Table 2 Inhibition of proMlCtL by various proteinase inhibitors.

Inhibitors	Specificity	Concentration (mM)	Inhibition (%)[a]
Control	–	–	0
E-64	All cysteine proteinases	0.1	21.19±3.83
		0.2	34.95±2.05
		0.5	60.10±3.91
Leupeptin	Cysteine/trypsin-like serine proteinases	0.1	95.85±1.42
Pepstatin A	Aspartic proteinases	0.1	21.54±2.46
PMSF	Serine proteinases	0.1	10.60±1.17
EDTA	Metallo-proteinases	0.1	10.81±1.82
EGTA	Metallo-proteinases	0.1	40.55±2.45
1,10-phenanthroline	Metallo-proteinases	0.1	36.60±1.50
Chymostatin	Serine/cysteine proteinases	0.1	21.29±1.34
Antipain	Serine/cysteine proteinases	0.1	97.26±0.82
NEM	Serine proteinases	0.1	40.84±1.91
Aprotinin	Serine proteinases	0.1	55.22±1.81

The purified proMlCtL was preincubated with the indicated inhibitors and assayed for residual activity using synthetic fluorogenic peptide substrate (Z-FR-AMC)

[a] Averages of triplicates ± SD

Gly-Arg-AMC (GGR) for proteasomal trypsin-like activity, Z-Gly-Pro-Arg-AMC (GPR) for granzyme/kallikrein activity, and Ala-Ala-Phe-AMC (AAF-AMC) for serine protease/chymotrypsin-like activity. In these assays, this enzyme showed little or no activity on Z-GGR-AMC and AAF-AMC. However, measurable effects on the hydrolysis of both the cathepsin B-like (Z-RR-AMC) and PGPH activity (Z-LLE-AMC) substrates (77% or more compared to Z-FR-AMC) were observed. Several other fluorogenic substrates (Z-GPR-AMC, AAF-AMC, and Z-GGR-AMC) were hydrolyzed at a low rate (29% or less compared to Z-FR-AMC) (Table 1).

Although the enzymatic activity of proMlCtL was greatly reduced or completely blocked by the cysteine protease inhibitors antipain and leupeptin (Table 2), E-64, which was the strongest cysteine protease inhibitor, only weakly inhibited the activity of GST-fused piscine cathepsin L in a dose-dependent manner. However, it should be noted that the in vitro substrate specificity, pH dependency, and enzyme inhibition of the recombinant enzyme fused to GST may not reflect its in vivo activity due to a conformational difference [45].

Many metal ions and detergents often influence the activity of proteases. As shown in Table 3, the proMlCtL enzyme was partially inactivated by $ZnSO_4$, $CuSO_4$, $CoCl_2$, and $HgCl_2$ and potentially activated by $MgSO_4$, $CaCl_2$, Brij 35, Triton X-100, Tween 20, and SDS. However, in the presence of a strong anionic surfactant (SDS) at a concentration of 0.01% (w/v), the protein always exhibited a higher activity than that observed in the control assay when no metal ions and detergents were present. Surprisingly, the activity of proMlCtL was strongly inhibited by E-64 in the presence of 0.01% (w/v) SDS (data not shown). This finding is consistent with the effects of SDS on the conformation and hemolytic activity of *Stichodactyla helianthus* toxins [46].

In this study, we cloned and expressed the cathepsin L gene from *M. mizolepis* in *E. coli* and purified the GST-fused pro-mature cathepsin L enzyme from the soluble fraction. Although the sequences and purifications of identical genes from other fish have been reported, the present study is the first to report on the recombinant expression of the gene

Table 3 Effect of metal ions and various detergents on proMlCtL activity.

Agent	Relative activity (%)		Residual activity (%)	
	1 mM	5 mM	1 mM	5 mM
Control	100		30.69±2.51	30.69±2.03
$ZnSO_4$	40.40±11.61	14.75±5.88	10.67±3.05	3.91±1.40
$CuSO_4$	35.68±13.09	ND	12.92±5.48	ND
$CoCl_2$	70.16±2.18	8.01±12.87	21.22±1.04	3.60±3.85
KCl	141.16±13.47	103.61±18.16	42.66±3.90	29.72±3.59
$MgSO_4$	248.62±13.87	295.31±12.60	74.54±2.90	90.74±7.36
$CaCl_2$	202.43±39.06	171.58±8.21	57.49±8.65	51.68±1.37
$HgCl_2$	37.32±9.56	269.70±27.68	12.22±3.48	79.42±2.12
	0.01%	0.05%	0.01%	0.05%
Brij 35	329.61±36.87	296.12±12.01	95.64±4.64	90.57±6.74
Triton X-100	164.43±16.14	208.82±28.07	48.29±2.16	61.01±5.40
Tween 20	337.68±24.18	335.67±24.83	100	100
SDS	201.87±16.02	ND	59.56±0.71	ND

ND not detected

and characterization of its enzymatic properties. Our results suggest that the sequence and evolutionary relationship of the mud loach cathepsin L gene were similar to that of other cathepsin Ls; however, the recombinant MlCtL protein was more stable between pH 6.5 and 8.5 than other cathepsin Ls, and it was strongly activated by several metal ions ($MgSO_4$ and $CaCl_2$) and detergents (Brij 35, Triton X-100, Tween 20, and SDS). Also, the gelatinolytic activity was determined by gelatin zymography. Enzyme activity was shown to be strongly inhibited by antipain and leupeptin but not inhibited by EDTA and PMSF, while a slight inhibition was observed with NEM, aprotinin, EGTA, and 1,10-phenanthroline. The results presented in this study not only provide important information on the activity of mud loach cathepsin L but also indicate that this enzyme holds promise for use as a functional FPH by the fish by-product processing industry and in waste-degradation processing at fisheries.

Acknowledgments This research was supported by a project grant (YSG-RE0602) from Yeongnam Sea Grant, Korea.

References

1. Kristinsson, H. G., & Rasco, B. A. (2000). *Critical Reviews in Food Science and Nutrition, 40*, 43–81.
2. Espe, M., Haaland, H., Njaa, L. R., & Raa, J. (1992). *Food Chemistry, 44*, 195–200.
3. Baca, D. R., Peña-Vera, M. T., & Díaz-Castañeda, M. (1991). *Journal of Food Science, 56*, 309–314.
4. Ritchie, A. H., & Mackie, I. M. (1992). *Animal Feed Science and Technology, 7*, 125–133.
5. Guérard, F., Dufossé, L., De La Broise, D., & Binet, A. (2001). *Journal of Molecular Catalysis. B, Enzymatic, 11*, 1051–1059.
6. Cassia, R. O., Martone, C. B., & Sánchez, J. J. (2000). *Latin American Applied Research, 30*, 241–244.
7. Martone, C. B., Borla, O. P., & Sánchez, J. J. (2005). *Bioresource Technology, 96*, 383–387.
8. Ledward, D. A., & Lawrie, R. A. (1984). *Journal of Chemical Technology & Biotechnology, 34B*, 223–228.
9. Rao, J. R., Mahendrakar, N. S., Chakrabarty, N. M., & Raghavan, S. L. (1996). *Seafood Export Journal, 27*, 17–23.
10. Jamdar, S. N., & Harikumar, P. (2005). *Bioresource Technology, 96*, 1276–1284.
11. Aspmo, S. I., Horn, S. J., & Eijsink, V. G. H. (2005). *Process Biochemistry, 40*, 1957–1966.
12. Guérard, F., Guimas, L., & Binet, A. (2002). *Journal of Molecular Catalysis. B, Enzymatic, 19–20*, 489–498.
13. Mackie, I. M. (1982). *Animal Feed Science and Technology, 7*, 113–124.
14. Katunuma, N., & Kominami, E. (1987). *Reviews of Physiology Biochemistry & Pharmacology, 108*, 1–20.
15. Turk, B., Turk, D., & Turk, V. (2000). *Biochimica et Biophysica Acta, 1477*, 98–111.
16. Kominami, E., Ueno, T., Muno, D., & Katunuma, N. (1991). *FEBS Letters, 287*, 189–192.
17. Villadangos, J. A., & Ploegh, H. L. (2000). *Immunity, 12*, 233–239.
18. Kirschke, H., & Barrett, A. J. (1987). Academic Press, pp. 193–238.
19. Yamashita, M., & Konagaya, S. (1990). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 96*, 247–252.
20. Lee, J. J., Chen, H. C., & Jiang, S. T. (1993). *Bioscience, Biotechnology, and Biochemistry, 57*, 1470–1476.
21. Masaki, T., Shimomukai, M., Miyauchi, Y., Ono, S., Tuchiya, T., Mastuda, T., et al. (1993). *Nippon Suisan Gakkaishi, 59*, 683–690.
22. Liu, Y., Song, L., Li, X., & Liu, T. (2002). *Toxicon, 40*, 395–399.
23. Nam, Y. K., Noh, C. H., & Kim, D. S. (1999). *Aquaculture, 172*, 229–245.
24. Nam, Y. K., Cho, Y. S., Cho, H. J., & Kim, D. S. (2002). *Aquaculture, 209*, 257–270.
25. Hall, T. A. (1999). *Nucleic Acids Symposium, 41*, 95–98.
26. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). *Nucleic Acids Research, 22*, 4673–4680.
27. Kumar, S., Tamura, K., & Nei, M. (2004). *Briefings in Bioinformatics, 5*, 150–163.
28. Laemmli, U. K. (1970). *Nature, 227*, 680–685.

29. Lim, S. U., Seo, J. S., Kim, M. S., Ahn, S. J., Jeong, H. D., Kim, K. H., et al. (2005). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 142*, 283–292.
30. Barrett, A. J., & Kirschke, H. (1981). *Methods in Enzymology, 80*, 535–561.
31. Von Heijne, G. (1983). *European Journal of Biochemistry, 133*, 17–21.
32. Vernet, T., Berti, P. J., Montigny, C., Musil, R., Tessier, D. C., Ménard, R., et al. (1995). *The Journal of Biological Chemistry, 270*, 10838–10846.
33. Tingaud-Sequeira, A., & Cerdà, J. (2007). *Gene, 386*, 98–106.
34. Yamashita, M., & Konagaya, S. (1990). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 95*, 149–152.
35. Yamashita, M., & Konagaya, S. (1990). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 96*, 247–252.
36. Aranishi, F., Ogata, H., Hara, K., Osatomi, K., & Ishihara, T. (1997). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 118*, 531–537.
37. Tsunemoto, K., Osatomi, K., Nozaki, Y., Hara, K., & Ishihara, T. (2004). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 137*, 107–114.
38. Liu, H., Yin, L., Zhang, N., Li, S., & Ma, C. (2006). *Journal of Agricultural and Food Chemistry, 54*, 9584–9591.
39. Heu, M. S., Kim, H. R., Cho, D. M., Godber, J. S., & Pyeun, J. H. (1997). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 118*, 523–529.
40. Visessanguan, W., Benjakul, S., & An, H. (2003). *Comparative Biochemistry and Physiology. Part B: Biochemistry & Molecular Biology, 134*, 477–487.
41. Nomura, T., & Fujisawa, Y. (1997). *Biochemical and Biophysical Research Communications, 230*, 143–146.
42. Mason, R. W., Bergman, C. A., Lu, G., Frenck Holbrook, J., & Sol-Church, K. (2004). *The Biochemical Journal, 378*, 657–663.
43. Petanceska, S., & Devi, L. (1992). *The Journal of Biological Chemistry, 267*, 26038–26043.
44. Baker, S. M., Karlsson, L., & Thurmond, R. L. (2003). *Protein Expression and Purification, 28*, 93–101.
45. Xu, Z. J., Nakajima, M., Suzuki, Y., & Yamaguchi, I. (2002). *Plant Physiology, 129*, 1285–1295.
46. Lanio, M. E., Alvarez, C., Pazos, F., Martinez, D., Martínez, Y., Casallanovo, F., et al. (2003). *Toxicon, 41*, 65–70.