Identification and functional characterization of legumain in amphioxus *Branchiostoma belcheri*

Lei TENG*†, Hiroshi WADA† and Shicui ZHANG*1

*Department of Marine Biology, Ocean University of China, Qingdao 266003, People’s Republic of China, †Department of Biology, Medical College of Qingdao University, Qingdao 266071, People’s Republic of China, and ‡Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572, Japan

**Synopsis**

Legumain has been reported from diverse sources such as plants, parasites (animals) and mammals, but little is known in the lower chordates. The present study reports the first characterization of legumain cDNA from the protochordate *Branchiostoma belcheri*. The deduced 435-amino-acid-long protein is structurally characterized by the presence of a putative N-terminal signal peptide, a peptidase_C13 superfamily domain with the conserved Lys123, Gly124-Asp125 motif and catalytic dyad His153 and Cys195 and two potential Asn-glycosylation sites at Asn85 and Asn270. Phylogenetic analysis demonstrates that *B. belcheri* legumain forms an independent cluster together with ascidian legumain, and is positioned at the base of vertebrate legumains, suggesting that *B. belcheri* legumain gene may represent the archetype of vertebrate legumain genes. Both recombinant legumain expressed in yeast and endogenous legumain are able to be converted into active protein of ~37 kDa via a C-terminal autocleavage at acid pH values. The recombinant legumain efficiently degrades the legumain-specific substrate Z-Ala-Ala-Asn-MCA (benzyloxyarylcarbonyl-L-alanyl-L-alanyl-L-asparagine-4-methylcoumaryl-7-amide) at optimum pH 5.5; and the enzymatic activity can be inhibited potently by iodoacetamide and N-ethylmaleimide, partially by hen’s-egg white cystatin, but not by E-64 [trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane], PMSF and pepstatin A. In addition, legumain is expressed in vivo in a tissue-specific manner, with main expression in the hepatic caecum and hind-gut of *B. belcheri*. Altogether, these results suggest that *B. belcheri* legumain plays a role in the degradation of macromolecules in food.

**Key words:** amphioxus, asparaginyl endopeptidase, *Branchiostoma belcheri*, lancelet, legumain, protochordate

**INTRODUCTION**

Legumain (EC 3.4.22.34), also known as asparaginyl endopeptidase, is a novel class of cysteine proteinase, which was identified originally in leguminous plants, hence the name legumain [1], and later first identified in animals in the human blood fluke *Schistosoma mansoni* [2]. It has since been found in mammals [3,4], in the parasitic protozoan *Trichomonas vaginalis* [5], in the parasitic nematodes *Toxocara canis* and *Brugia malayi* [6], and *Haemonchus contortus* [7], and in the parasitic arthropods *Ixodes ricinus* [8]. *Fasciola gigantica* [9] and *Haemaphysalis longicornis* [10]. However, little information regarding legumains is available in the lower chordates.

Legumain is shown to be expressed in a number of mammalian tissues, such as the kidney, placenta, spleen, liver and testis [11]. It specifically catalyses the hydrolysis of asparaginyl bonds [12,13]. However, legumain usually exists in inactive form that can be autoactivated by incubation at acid pH [12]. Its enzymatic activity can be inhibited by chicken ovocystatin and human cystatin, but not by compound E-64 [trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane], a potent inhibitor of many cysteine peptidases of the papain family, including most of the lysosomal cathepsins [3,4,11]. Functionally, legumains from different sources have been shown to be involved in processing of other proteins, including processing of food storage proteins in plants [14], processing of peptide antigens for MHC II presentation, degradation of several biologically active proteins and

**Abbreviations used:** qCt, threshold cycle value; DIG, digoxigenin; DTT, dithiothreitol; E-64, trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; NEM, N-ethylmaleimide; Ni-NTA, Ni2+-nitrilotriacetate; ORF, open reading frame; qRT-PCR, quantitative real-time PCR; RACE, rapid amplification of cDNA ends; SMART, simple modular architecture research tool; UPM, Universal Primer A Mix; UTR, untranslated region; Z-Ala-Ala-Asn-MCA, benzyloxyarylcarbonyl-L-alanyl-L-alanyl-L-asparagine-4-methylcoumaryl-7-amide

The nucleotide sequence data reported for *Branchiostoma belcheri* legumain will appear in the DDBJ, EMBL, GenBank® and DDBJ Nucleotide Sequence Databases under accession number FJ842396.

†To whom any correspondence should be addressed (email sczhang@ouc.edu.cn).
inhibition of osteoclast formation and bone resorption in mammals [4,15–18]. Legumains in the blood-feeders *Sch. mansoni* [19,20] and *H. contortus* [7] are implicated in Hb degradation by proteases and function by activating cathepsin Bzymogens as mature enzymes with haemoglobinase activity in the gut of the parasites. However, few reports characterize legumains in the lower vertebrates exist.

Amphioxus or lancelet, a cephalochordate, is a small filter-feeding animal that lives in shallow sea water and spends most of its time buried in the sand with its rostral end protruding above the surface. It has long been regarded as the extant invertebrate most closely related to the proximate ancestor of vertebrates [21]. The study of legumain in this evolutionarily important animal will provide important implications for understanding the origin and evolution of its molecular structure and biological function. As an initial step, we report in the present study the identification, cloning and sequencing of legumain cDNA from *Branchiostoma belcheri* collected in the vicinity of Qingdao, and poly(A) products were gel-purified using a DNA gel extraction kit (Axyan). The first-strand cDNA was synthesized at 4°C overnight, and transformed into Top10 competent cells (Tiangen). The positive clones were selected and sequenced with an ABI PRISM 3730 DNA sequencer. The sequences were searched in GenBank® Nucleotide Sequence Database with BLASTx for comparative analysis.

The gene-specific primers P3 and P4 (Table 1) were used in RACE (rapid amplification of cDNA ends) reactions for the full-length cDNA cloning. Both the 3′-RACE-Ready cDNA and 5′-RACE-Ready cDNA were prepared by using a SMART (simple modular architecture research tool) RACE cDNA amplification kit (Clontech) according to the manufacturer’s instructions. The 3′-RACE reaction mixture (final volume 20 μl) contained 2 μl of 10× Advantage 2 PCR buffer, 0.4 μl of dNTPs (10 mM), 2 μl of 10× UPM (Universal Primer A Mix; Clontech; Table 1), 0.4 μl of P3 (10 μM), 13.8 μl of PCR-grade water, 0.4 μl of 50× Advantage 2 polymerase mix and 1 μl of 3′-RACE-Ready cDNA as a template. The thermal conditions for the 3′-RACE reaction were: initial denaturation at 94°C for 4 min, followed by 33 cycles at 94°C for 30 s, 65°C for 30 s, 72°C for 90 s, and then an additional extension at 72°C for 7 min. The 5′-RACE reaction mixture (final volume 20 μl) contained 2 μl of 10× Advantage 2 PCR buffer, 0.4 μl of dNTPs (10 mM), 2 μl of 10× UPM, 0.4 μl of P4 (10 μM), 13.8 μl of PCR-grade water, 0.4 μl of 50× Advantage 2 polymerase mix and 1 μl of 3′-RACE-Ready cDNA. The PCR was run under the following conditions: initial denaturation at 94°C for 4 min, followed by 33 cycles at 94°C for 30 s, 68°C for 30 s, 72°C for 90 s, and the final extension step at 72°C for 7 min. The 5′-RACE and 3′-RACE products were gel-purified, subcloned and sequenced as described above.

### MATERIALS AND METHODS

#### Cloning and sequencing of legumain cDNA

Total RNAs were extracted with TRIZol® (Invitrogen) from *B. belcheri* collected in the vicinity of Qingdao, and poly(A)+ RNA (polyadenylated RNA) was purified using the poly(A) tract mRNA isolation system II (Promega) according to the manufacturer’s instructions. The first-strand cDNA was synthesized with the reverse transcription system (Promega) using oligo d(T) primer. The fragments of *B. belcheri* legumain gene were amplified by PCR with the degenerate primer pairs, P1 and P2 (Table 1), designed based on the conserved motifs of known legumains. PCR amplification was carried out at 94°C for 4 min, followed by 31 cycles of 94°C for 45 s, 53°C for 45 s, 72°C for 90 s and a final extension step at 72°C for 7 min. The PCR products were gel-purified using a DNA gel extraction kit (Axygen), ligated into the T/A cloning vector pGEM-T Easy (Promega) and transformed into Top10 competent cells (Tiangen). The positive clones were selected and sequenced with an ABI PRISM 3730 DNA sequencer. The sequences were searched in GenBank® Nucleotide Sequence Database with BLASTx for comparative analysis.

### Table 1 Sequences of the primers used in the present study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′→3′)</th>
<th>Sequence information</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (sense)</td>
<td>AAYTACMGRCACCCAGGC</td>
<td>Legumain cDNA fragment primer</td>
</tr>
<tr>
<td>P2 (antisense)</td>
<td>AACCCTGTCTCCACACAG</td>
<td>Legumain cDNA fragment primer</td>
</tr>
<tr>
<td>P3 (sense)</td>
<td>CCCTCCCTCGTTACAGAGGTCCCGC</td>
<td>3′-RACE primer</td>
</tr>
<tr>
<td>P4 (antisense)</td>
<td>CGTGTCGACCAGGGCTGCTGGTGC</td>
<td>5′-RACE primer</td>
</tr>
<tr>
<td>UPM</td>
<td>CTAATACGATCTACCATAGGGGC</td>
<td>UPM</td>
</tr>
<tr>
<td>P5 (sense)</td>
<td>ATACTCAGAAAAAGATTCCCCGCGATTTTGGGA</td>
<td>Recombinant primer</td>
</tr>
<tr>
<td>P6 (antisense)</td>
<td>CTAGCTAGACGCGAGCGCTGGCGTCATTGGCAG</td>
<td>Recombinant primer</td>
</tr>
<tr>
<td>P7 (sense)</td>
<td>GCCCGGCCGATGGCGTC</td>
<td>Real-time primer</td>
</tr>
<tr>
<td>P8 (antisense)</td>
<td>GTCCCGGCCTGCGTCCAGTGC</td>
<td>Real-time primer</td>
</tr>
<tr>
<td>P9 (sense)</td>
<td>GAGACCTCACCACGCCAGGC</td>
<td>β-Actin primer</td>
</tr>
<tr>
<td>P10 (antisense)</td>
<td>CGTCAGTCCAGCCAGTCCAGACC</td>
<td>β-Actin primer</td>
</tr>
</tbody>
</table>
were predicted with the NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc). Multiple protein sequences were aligned using the MegAlign program by the ClustalW method in DNASTAR software package. The phylogenetic tree was constructed using the neighbour-joining method in PHYLIP 3.67 software package using 1000 bootstrap replicates.

Expression of legumain in Pichia pastoris

The complete region encoding B. belcheri legumain without the predicted signal peptide was amplified by PCR using the specific primers P5 and P6 (Table 1). The reaction was carried out under the following conditions: initial denaturation at 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 90 s and then an additional extension at 72 °C for 7 min. The purified PCR product was digested with XhoI and Xbal and then cloned into the plasmid expression vector pPICZαA (Invitrogen) previously cut with the same restriction enzymes. The identity of the insert was verified by sequencing, and the plasmid was designated pPICZαA/legumian.

The constructed plasmid pPICZαA/legumain was linearized with Sall and transformed into the competent cells of P. pastoris X33 by the electroporation method according to the manufacturer’s instructions (Invitrogen). The cells were then spread on YPD (1% yeast extract, 2% peptone, 2% dextrose, 2% agar, 1 M sorbitol and 100 μg/ml zeocin) plates to select positive clones.

One positive clone was inoculated into 100 ml of BMGY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 4 × 10^{-5} % biotin and 1% glycerol) and grew at 28 °C until the culture achieved a D_{600} (attenuance at 600 nm) of 2–6. The cells were harvested by centrifuging at 2000 g for 10 min at room temperature (25 °C), resuspended in 500 ml of BMGY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 4 × 10^{-3} % biotin and 0.5% methanol) and cultured at 28 °C. To induce expression, methanol was added every 24 h to a final concentration of 0.5% for two successive days. The culture was centrifuged at 10000 g for 20 min at 4 °C. Subsequently, solid (NH_4)_2SO_4 was added to the supernatant to a final concentration of 75% saturation. After stirring at 4 °C overnight, the suspension was centrifuged at 10000 g for 20 min at 4 °C. The precipitate was suspended in dialysis buffer (20 mM PBS with 500 mM NaCl, pH 7.4) and dialysed against the same buffer, which was changed three to four times, until traces of (NH_4)_2SO_4 were removed. The dialysed sample was pooled, filtered through a 0.45 μm Millipore filter and loaded on to an Ni-NTA (Ni^{2+}-nitrilotriacetate) resin column (Amersham). The column was washed with the washing buffer (20 mM PBS containing 500 mM NaCl and 20 mM imidazole, pH 7.4) and eluted with the elution buffer (20 mM PBS containing 500 mM NaCl and 250 mM imidazole, pH 7.4). The purity of the eluted samples was analysed by SDS/PAGE (12% gel) as described by Laemmli [22], and stained with Coomassie Brilliant Blue R-250. To remove imidazole, the eluted samples with higher purity were dialysed against 0.05% CHAPS buffer with 1 mM EDTA for 24 h at 4 °C. The purified recombinant legumain was aliquoted and stored at −70 °C until used.

The protein concentrations were determined by the method of Bradford [23] using BSA as a standard.

Preparation of antibody

The purified recombinant legumain was used for raising antibody in a rabbit. Approx. 400 mg of the legumain was emulsified with Freund’s complete adjuvant and injected subcutaneously at multiple sites of the rabbit. Two booster injections of 200 mg antigen mixed with Freund’s incomplete adjuvant were administered subcutaneously at intervals of 2 weeks. At 8 days after the final booster, the blood was collected and the serum was prepared. Antibody titre in the serum was determined by the dot-blot assay. The antiserum was aliquoted and stored at −70 °C.

Western blotting

The different tissues including the hind-gut, hepatic caecum, gill, muscle, notochord, ovary and testis were dissected out of B. belcheri, and homogenized in 0.05% CHAPS buffer with 1 mM EDTA on ice. The homogenates were centrifuged at 10000 g at 4 °C for 10 min, and the supernatants were collected. The supernatants were subjected to SDS/PAGE (12% gel). Western blotting was carried out as described by Towbin et al. [24] with slight modifications. In brief, the gels were washed for 5 min in 15.6 mM Tris/HCl (pH 8.3) containing 120 mM glycine and 20% (v/v) methanol, and the proteins on the gels were blotted on to a nitrocellulose membrane (Hybond; Amersham Biosciences). The blotted membranes were incubated first in 20 mM PBS containing 3% defatted milk powder at 30 °C for 2 h and then in the rabbit antiserum diluted 1:300 with 20 mM PBS containing 0.1% Tween 20 for 2 h. After washing with 20 mM PBS, the membranes were incubated in HRP (horseradish peroxidase)-labelled goat anti-rabbit IgG (Zhongshan) diluted 1:1000 at 30 °C for 2 h. The bands were visualized using 0.06% DAB (diaminobenzidine) and 0.03% H_2O_2.

Assay for recombinant legumain activity

To determine the optimum pH for autocatalytic activation, aliquots of 5 μl of the recombinant legumain (0.5 mg/ml) were mixed with 20 μl of 0.05% CHAPS buffer [containing 2 mM DTT (dithiothreitol) and 1 mM EDTA] with pH gradient 3–12. The mixtures were incubated at 25 °C overnight, and assayed by SDS/PAGE.

The activity of recombinant legumain was assayed by cleavage of the synthetic fluorescent substrate Z-Ala-Ala-Asn-MCA (benzyloxycarbonyl-l-alanyl-l-alanyl-l-asparagine-4-methyl-coumaryl-7-amide; Peptide Institute, Osaka, Japan) specific for legumain, and all the assays were performed in triplicate. Briefly, aliquots of 5 μl of 0.5 mg/ml of the recombinant legumain were mixed with 118 μl of 0.05% CHAPS buffer (pH 5.5 and 7 respectively) containing 2 mM DTT and 1 mM EDTA and preincubated at 25 °C for 0, 1, 2, 4, 8 and 16 h respectively. After the addition of 2 μl of 10 mM Z-Ala-Ala-Asn-MCA into the mixtures, they were incubated at 25 °C for 20 min and the enzymatic activities were measured using a microplate reader (Tecan Genios Plus) at 360 and 465 nm respectively. One unit of activity was defined as the activity releasing 1 μM of MCA per minute.
To test the sensitivity of the recombinant legumain to specific inhibitors, the endopeptidase inhibitors used were iodoacetamide (2 mM; Sigma), NEM (N-ethylmaleimide; 2 mM; Sigma), E-64 (100 μM; Sigma), hen’s-egg white cystatin (3 μM; Sigma) and pepstatin A (100 μM; Sigma) and PMSF (2 mM; Amresco). The inhibition assays of enzyme activities were carried out as described above except that the inhibitors were included in the reaction mixtures.

To examine the effects of endopeptidase inhibitors on autocatalytic activation, 1 μl of iodoacetamide (2 mM), NEM (2 mM), E-64 (100 μM), hen’s-egg white cystatin (3 μM), pepstatin A (100 μM) and PMSF (2 mM) were mixed with 4 μl of recombinant legumain (0.5 mg/ml) in 20 μl of 0.05% CHAPS (pH 4) containing 2 mM DTT and 1 mM EDTA. The mixtures were incubated at 25°C for 16 h, and subjected to SDS/PAGE (12% gel).

**Assay of endogenous legumain activity**

The tissue extracts were prepared from the hind-gut, hepatic caecum, gill, muscle, notochord, ovary and testis by homogenizing in 0.05% CHAPS buffer with 1 mM EDTA on ice and centrifuging at 10 000 g at 4°C for 10 min. The supernatants were collected and their protein concentrations were determined. The legumain activities in the supernatants were monitored as above at indicated pH.

**Assays for expression of legumain gene in different tissues**

To examine the expression of legumain gene in *B. belcheri*, Northern blotting and qRT-PCR (quantitative real-time PCR) were performed. Total RNAs were prepared with TRIzol® (Gibco) from the whole *B. belcheri* as well as the different tissues hind-gut, hepatic caecum, gill, muscle, notochord, ovary and testis. For Northern blotting, an aliquot of 5 μg of RNAs of each tissue and the whole animal was electrophoresed and blotted on to a nylon membrane (Roche). The DIG (digoxigenin)-labelled riboprobes of approx. 750 bp were synthesized in *vitro* from the linearized plasmid DNA with *B. belcheri* legumain ORF (open reading frame) by following the DIG-UTP supplier’s instructions (Roche). The hybridization was performed as described by Fan et al. [25].

For qRT-PCR, the total RNAs were digested with RQ1 RNase-free DNase (Promega) to eliminate genomic contamination, and the cDNAs were synthesized with a reverse transcription system (Promega) using oligo (dT) primer and used as a template. Two primers, P7 and P8 (Table 1), specific for *B. belcheri* legumain cDNA, were used to amplify a product of 141 bp. The β-actin gene was chosen as the reference for internal standardization. Two β-actin primers, P9 and P10 (Table 1), were used to amplify a β-actin gene fragment of 97 bp. The qRT-PCR amplifications were carried out in triplicate in a total volume of 20 μl of reaction mixture containing 10 μl of 2 × SYBR® Premix Ex Taq™ (Takara), 0.4 μl of ROX Reference Dye II (50×), 1 μl of the 1:5 diluted cDNA, 0.2 μl each of P7 and P8 (or P9 and P10) primers (20 μM) and 8.2 μl of PCR-grade water. The qRT-PCR program was as follows: denaturation at 95°C for 10 s, followed by 40 cycles of 95°C for 5 s/60°C for 15 s/72°C for 35 s. Dissociation analysis of amplification products was performed at the end of each PCR reaction to confirm that a single PCR product was amplified and detected. After the PCR programme, the data were analysed with 7500 System SDS software v 1.4.0 (Applied Biosystems) and quantified with the Ct (threshold cycle value) method (2−ΔΔCt method) based on Ct values for both *B. belcheri* legumain and β-actin genes in order to calculate the relative mRNA expression level.

Statistical analysis was performed using SPSS 13.0 for Windows. All the data were given in terms of relative mRNA and expressed as means ± S.D. The data obtained from qRT-PCR were subjected to one-way ANOVA followed by the Dunnett two-sided test to determine differences in the mean values among the different tissues. *P* < 0.05 was considered statistically significant.

**RESULTS**

**Characteristics and phylogeny**

The full-length cDNA of *B. belcheri* legumain was 1637 bp long, which contained an ORF of 1308 bp, a 5'-UTR (5'-untranslated region) of 60 bp and a 3'-UTR of 269 bp with a polyadenylated tail. The ORF encoded a protein of 435 amino acids with a predicted molecular mass of approx. 48.9 kDa and a pI of 5.1. The deduced protein (Figure 1) had a single peptidase_C13 superfamily domain located at the residues 34–293, with the conserved KGD (Lys123-Gly124-Asp125) motif and the catalytic dyad His173 and Cys195, which are both characteristic of legumains. It also possessed two potential Asn-glycosylation sites at Asn87 and Asn270, and all the highly conserved cysteine residues at the residual positions 226, 386, 398, 418 and 435. SignalP software analysis revealed the presence of a signal peptide with a predicted cleavage site between Gly15 and Phe16. Alignment analysis (Figure 1) showed that *B. belcheri* legumain shared 36.6, 45.9, 46.1, 49.8, 49.4, 47.1 and 48.5% identity with the legumains of *Sch. mansoni*, *H. longicornis*, *Ciona intestinalis*, *Danio rerio*, *Xenopus laevis*, *Bos taurus* and *Homo sapiens* respectively. The phylogenetic tree constructed using the sequences of representative legumains including that of *B. belcheri* demonstrated that *B. belcheri* legumain formed an independent group together with *C. intestinalis* legumain (Figure 2), which is positioned at the base of vertebrate cluster, suggesting that *B. belcheri* legumain gene may be the archetype of vertebrate legumain genes.

**Characteristics and enzymatic activities of recombinant legumain**

An expression vector including the entire region of *B. belcheri* legumain without the signal peptide and a 3' additional His15 tag of pPICZαA was constructed and transformed into *P. pastoris* X33, and this resulted in the addition of Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu-Asn-Ser-Ala-Val-Asp-(His)n to the C-terminus of the recombinant protein. The recombinant protein was purified by affinity chromatography on an Ni-NTA resin column, and...
analysed by SDS/PAGE (12% gel), followed by staining with Coomassie Brilliant Blue R-250, which demonstrated the presence of a single protein band of ∼50 kDa (Figure 3A). The predicted molecular mass calculated from *B. belcheri* legumain cDNA excluding the signal peptide was 47.2 kDa, and the His$_6$ tag at the C-terminus in the expression vector contributed 2.5 kDa to the size of recombinant protein, thus leading to increase of the molecular mass of recombinant legumain to ∼50 kDa.

Human and ixodid tick prolegumains have been shown to be converted into mature active legumains through C-terminal autocleavage under acid conditions [10,12]. We found that *B. belcheri* recombinant legumain was also degraded at acid...
Figure 2  Phylogenetic tree constructed by the neighbour-joining method with PHYLIP 3.67 software package
Bootstrapped majority consensus values on 1000 replicates are indicated at each branch point in percent. The scale bar represents branch length (number of amino acids substitutions per 100 residues). Rice legumain was used as the outgroup. Accession numbers of legumain sequences obtained from GenBank® are: Oryza sativa (BAA84650); Sch. mansoni (CAJ45481); B. malayi (XP_001901514); H. longicornis (BAF51711); T. laevis (XP_001125596); C. intestinalis (XP_002130826); D. rerio (NP_999924); X. laevis (NP_001079911); S. salar (ACN10644); D. rerio (NP_999924); T. nigroviridis (CAG13252); Paralichthys olivaceus (ABM88796); Sus scrofa (NP_776526); Canis familiaris (XP_537355); Rattus norvegicus (AAH87708); Mus musculus (NP_035305); Pongo abelii (NP_001126789); Macaca fascicularis (Q4R4T8); H. sapiens (BAA09530). The sequence obtained in the present study is marked in boldface.

Figure 3  SDS/PAGE of recombinant legumain and autocatalytic products
(A) SDS/PAGE of recombinant legumain. Lane M, molecular mass standards; lane 1 and 2, supernatant of methanol-induced positive P. pastoris X33 containing pPICZαA/legumain (two different clones); lane 3, supernatant of methanol-induced negative P. pastoris X33 containing pPICZαA. (B) SDS/PAGE of autocatalytic products after recombinant legumains were incubated overnight at 25°C in a buffer with pH gradient 3–12. Every pH value is marked on the top of its corresponding lane.
Legumain in amphioxus

Figure 4 Hydrolytic efficiency of recombinant legumain toward the legumain-specific fluorogenic substrate Z-Ala-Ala-Asn-MCA detected at pH 5.5 and 7 at different incubated times

pH values, mainly yielding a protein of \( \sim 37 \text{kDa} \) (Figure 3B), while the degradation did not occur under alkaline conditions. On the other hand, the enzymatic activity assays showed that *B. belcheri* recombinant legumain was able to efficiently hydrolyse the legumain-specific fluorogenic substrate Z-Ala-Ala-Asn-MCA at pH 5.5 in a time-dependent manner, and the specific activity reached \( 7.9 \times 10^{-4} \mu \text{M/min per mg} \) at 16 h after incubation (Figure 4). In contrast, it exhibited little enzyme activity at pH 7 even after prolonged incubation. These results indicated that *B. belcheri* legumain was also capable of activation via autocleavage of the C-terminal extension, resulting in the conversion into the mature active enzyme. The autocleavage sites of tick and human legumains were demonstrated to be at Asn\(^{364-365}\) and Asn\(^{323-324}\), which are both located in the most hydrophilic regions [12,26,27]. Consistently, there existed also an Asn\(^{347}\) in the most hydrophilic region (Figure 1) in *B. belcheri* legumain predicted using ProtScale tool (http://www.expasy.org/cgi-bin/protscale.pl), implicating that the autocleavage may be at the site Asn\(^{347-348}\) and the autocleavage at this site causes the removal of a \( \sim 10 \text{kDa} \) C-terminal peptide from the precursor. This was apparently supported by the fact that the incubation of *B. belcheri* recombinant legumain at acid pH values resulted in the production of a \( \sim 37 \text{kDa} \) protein (Figure 3B). These strongly suggest that *B. belcheri* legumain was activated by the autocleavage at Asn\(^{347-348}\).

The enzymatic activity inhibition assays showed that the thiol-blocking reagents iodoacetamide and NEM potently blocked the enzyme activity, whereas the papain-like cysteine protease (family C1) inhibitor E-64, the serine protease inhibitor PMSF and the aspartic protease inhibitor pepstatin A had little inhibitory effects on the enzyme activity (Table 2). Notably, hen’s-egg white cystatin, an unusual potent inhibitor of legumains [3], displayed less significant inhibitory activity towards *B. belcheri* legumain (58% inhibition by 3 \( \mu \text{M cystatin} \)). SDS/PAGE revealed that *B. belcheri* recombinant legumain was not autocatalytic in the presence of iodoacetamide and NEM (Figure 5, lanes 3 and 8) or at alkaline pH (Figure 5, lane 2), and was only slightly degraded.

**Table 2** Inhibition of recombinant legumain activity toward specific fluorogenic substrates by different protease inhibitors

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Concentration</th>
<th>Inhibition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine protease inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iodoacetamide</td>
<td>2 mM</td>
<td>90</td>
</tr>
<tr>
<td>NEM</td>
<td>2 mM</td>
<td>88</td>
</tr>
<tr>
<td>E-64</td>
<td>100 ( \mu \text{M} )</td>
<td>7</td>
</tr>
<tr>
<td>Hen’s egg white cystatin</td>
<td>3 ( \mu \text{M} )</td>
<td>58</td>
</tr>
<tr>
<td>Serine protease inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMSF</td>
<td>2 mM</td>
<td>11</td>
</tr>
<tr>
<td>Aspartic protease inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pepstatin A</td>
<td>100 ( \mu \text{M} )</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 5 SDS/PAGE of autocatalytic products after legumains were incubated overnight with different protease inhibitors at 25°C at pH 4

Every pH and inhibitor is marked on the top of the Figure.

![Figure 5](attachment:image_url)
in the presence of hen’s-egg white cystatin (Figure 5, lane 6), but it was heavily cleaved in the presence of E-64, pepstatin A and PMSF (Figure 5, lanes 4, 5 and 7). It was apparent that iodoacetamide, NEM and hen’s-egg white cystatin were all capable of suppressing the autocleavage of \( B. \ belcheri \) legumain.

**Endogenous legumain activity and autocleavage**

Total proteins were extracted from the different tissues of \( B. \ belcheri \) in order to assay the endogenous legumain activity. The enzymatic activity assays showed that the legumain was mainly present in the hepatic caecum and hind-gut (Figure 6A), with the highest activities of \( 6.0 \times 10^{-4} \) and \( 5.2 \times 10^{-4} \) \( \mu \)M/min per mg respectively. It was also observed that the legumain in the hepatic caecum was able to hydrolyse the substrate Z-Ala-Ala-Asn-MCA over a wide range of pH values (Figure 6B), with the optimum pH at 5.5, consistent with the results obtained for recombinant legumain (Figure 3B).

The rabbit antiserum against the purified legumain with a titre of 1:1500 was obtained. Western blotting exhibited that the rabbit antiserum reacted with the recombinant legumain, forming a single band of \( \sim 50 \) kDa (Figure 7). The antiserum was also reactive with the extracts of the hepatic caecum and hind-gut, all yielding two bands with molecular masses of \( \sim 47 \) and \( \sim 37 \) kDa (Figure 7), but no positive signals were seen in the extracts of gill, muscle, notochord, ovary and testis, agreeing with the results of enzymatic activity assays. These results also indicated that the enzymatic activation occurred in the endogenous legumain.

**Legumain expression**

Northern blotting revealed that legumain transcripts were 1600 bp in size, and they were abundant in the hepatic caecum and hind-gut, and at a very low level in the ovary and testis, agreeing with the results of enzymatic activity assays. These results also indicated that the enzymatic activation occurred in the endogenous legumain.
DISCUSSION

Asparaginyl endopeptidase/legumain has been reported from diverse sources, ranging from plants to parasites (animals) to mammals. The present study describes the identification, expression and biochemical characterization of legumain in the amphioxus *B. belcheri*. To our knowledge, this is the first such data from a protochordate. The deduced 435-amino-acid-long protein displays features in common with those in humans [3], arthropod [8] and schistosomes [28]; it is structurally characterized by the presence of a putative N-terminal signal peptide of 15 amino acids, a peptidase_C13 superfamily domain with the conserved KGD (Lys123-Gly124-Asp125) motif and the catalytic dyad His153 and Cys195, two potential Asn-glycosylation sites at Asn65 and Asn270 and five highly conserved cysteine residues at the positions 226, 386, 398, 418 and 435. Phylogenetic analysis demonstrates that *B. belcheri* legumain is grouped with *C. intestinalis* legumain, forming an independent cluster, and is positioned at the base of vertebrate legumains, suggesting that *B. belcheri* legumain gene may represent the archetype of vertebrate legumain genes.

Like human and ixodid tick prolegumains [10,12], *B. belcheri* legumain is converted into active form via C-terminal autocleavage. The predicted molecular mass calculated from *B. belcheri* legumain cDNA excluding the signal peptide was 47.2 kDa; at acid pH, *B. belcheri* recombinant legumain is degraded, yielding a main band of protein of ∼37 kDa. The autocleavage sites of human and tick legumains are at Asn364–365 and Asn232–234, located in the most hydrophilic regions [12,26,27]. There is also Asn147 in the most hydrophilic region of *B. belcheri* legumain, and autocleavage at this site exactly causes the removal of a ∼10 kDa C-terminal peptide from the precursor. This further supports that the removal of C-terminal peptide from prolegumain is required for legumain activity.

The biochemical properties of *B. belcheri* legumain expressed in *P. pastoris* agree with the data reported for mammalian [3,11,29], arthropod [8], ixodid tick [10], schistosomes [28] and plants [30]. Recombinant *B. belcheri* legumain is acid-active, with an optimum pH for the hydrolysis of the substrate Z-Ala-Ala-Asn-MCA at 5.5, and it loses activity at neutral to alkaline pH. This is in contrast with the optimum pH values of *Ala-Asn-MCA* at 5.5, and it loses activity at neutral to alkaline pH. This is in contrast with the optimum pH values of *Sch. mansoni* [19,20] and *H. contortus* [7].

In summary, the present study reports for the first time the characterization of legumain in lower chordates, and the presence of legumain in the hepatic caecum and hind-gut of *B. belcheri* suggests a function for this enzyme associated with the degradation of food and/or activation of cysteine proteases.

REFERENCES


