Aquaculture 356–357 (2012) 264–271

Contents lists available at SciVerse ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aqua-online

Cathepsin L proteases of the parasitic copepod, *Lepeophtheirus salmonis*

Elaine McCarthy [a,*], Eleanor Cunningham [a], Lorraine Copley [b], David Jackson [b], David Johnston [c], John P. Dalton [d], Grace Mulcahy [a]

[a] School of Agriculture, Food Science and Veterinary Medicine, Veterinary Sciences Building, University College Dublin, Belfield, Dublin 4, Ireland
[b] Marine Institute, Rinville, Oranmore, Co. Galway, Ireland
[c] Biomedical Parasitology Division, Department of Zoology, Natural History Museum, Cromwell Road, London, UK
[d] Institute of Parasitology, McGill University, Montreal, Canada

ARTICLE INFO

Article history:
Received 7 July 2010
Received in revised form 23 April 2012
Accepted 5 May 2012
Available online 12 May 2012

Keywords:
L. salmonis
Protease
Cathepsin L
Ectoparasite

ABSTRACT

The salmon louse, *Lepeophtheirus salmonis*, is a parasitic copepod that feeds on the mucus, skin and blood of salmonids. We describe the identification of two complete *L. salmonis* cathepsin L-like gene sequences and their molecular characterisation. *L. salmonis* cathepsin L1 (LsCL1), is 978 base pairs in length, encoding a protein of 325 amino acid residues while *L. salmonis* cathepsin L2 (LsCL2) is 1149 base pairs in length, encoding a protein of 382 amino acid residues. The predicted molecular weights of LsCL1 and LsCL2 are 35,964 Da and 42,150 Da respectively. The two proteases share only 25% identity in the primary sequences; however, the catalytic triad of cysteine, histidine and asparagine is highly conserved for both. Biochemical analysis of *L. salmonis* extracts revealed that cathepsin L has an optimum activity at pH 6.5, at 15 °C and remains stable at this temperature. Cathepsin L activity is present in all of the parasite life stages assayed, with the chalimus life stage extract exhibiting the most activity. Cathepsin L activity was also observed in the secretory/excretory products possibly indicating a role for this protease in immunoevasion and establishment of the parasite on the host.

1. Introduction

Lepeophtheirus salmonis infection is a parasitic infection that causes major economic loss in the salmonid farming industry worldwide (Costello, 2009). During infection, the copepodid stage attaches to the host and molts into the first of four chalimus stages, all of which are physically attached to their host by a frontal filament The fourth chalimus stage molts into the first of two pre-adult stages which are free moving on the host with the exception of a short time during the molt when a temporary frontal filament is formed. The second preadult stage molts into free moving adults which undergo no further molting. All developmental stages of the parasite feed on fish mucus and skin, while the pre-adult and adult stages can also ingest blood. While heavy infestations rarely occur in wild populations, sea lice infection can have a devastating effect on farmed salmon if left untreated (Schram, 1993). While the literature relating to sea lice is ever expanding and our knowledge of the parasites' biology and ecology is growing, more research is required to identify the molecules employed by these parasites to establish and maintain infection.

Cysteine proteases are members of the papain-like proteases classed as the C1 family (Bromme, 2001). All papain-like cysteine proteases consist of a signal peptide, a pro-peptide and a mature proteolytically active enzyme. Signal peptides are responsible for translocation into the endoplasmic reticulum during ribosomal protein expression. Propeptides of variable length function as a scaffold for protein folding of the catalytic domain, as a chaperone for the transport of the propeptide to the endosomal lysosomal compartment and as a high affinity reversible inhibitor preventing the premature activation of the catalytic domain (Lecaille et al., 2002). Several cathepsin-type cysteine proteases have been detected in non-lysosomal regions of eukaryotic cells and embryos, leading to speculation that the enzymes function in reactions outside lysosomes and may accumulate in different organelles (Butler et al., 2001).

Cathepsin L cysteine proteases belong to the C1 family and are categorised into the Clan CA (Rawlings and Barrett, 1994; Yamaji et al., 2009). This class of protease has long been implicated in critical parasitic functions. In the parasitic ciliate *Ichthyophthirius multifiliis*, it has been suggested that cathepsin L may play a role in invasion of the fish host epidermis (Jousson et al., 2007). Cathepsin L-like proteases from the helminth parasites *Fasciola hepatica* and *Schistosoma mansoni* have been shown to be involved in activities such as invasion of tissues, feeding, immune evasion and egg shell formation (Brady et al., 2000; Dvorak et al., 2009; Smooker et al., 2000). Similarly, in the Chinese liver fluke, *Clonorchis sinensis*, cathepsin L facilitates the invasion of the cercariae into the intermediate fish host by secretion of the enzyme through the tegument of parasite thus penetrating the skin and migrating to the host muscle (Li et al., 2009). In crustaceans, cathepsin L has been purified from the gastrointestinal juice of the American lobster, *Homarus americanus* (Laycock et al., 1989). Additionally, cathepsin L cDNA has been cloned from the digestive glands, the hepatopancreas, of *H. americanus* (Laycock et al., 1991) and from the shrimps *Litopenaeus*

* Corresponding author. Tel.: +1 353 1 7166138; fax: +1 353 1 7166185.
E-mail address: elaine.mccarthy@ucd.ie (E. McCarthy).

0044-8486/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.aquaculture.2012.05.007

vannamei (Le Boulay et al., 1996) and *Metapenaeus ensis* (Hu and Leung, 2004). More recent studies have demonstrated that shrimp cathepsin L has a role in intracellular digestion within the vacuoles of cells and is secreted extracellularly into the lumen of the hepatopancreas and stomach (Hu and Leung, 2007).

To date, there has been limited characterisation of the proteases of *L. salmonis*. A number of *L. salmonis* trypsin and trypsin-like serine proteases have been identified and localised to the gut, involved in digestion (Johnson et al., 2002; Kvamme et al., 2004, 2005) and to the egg, involved in yolk degradation (Skern-Mauritzen et al., 2009). Our group has identified and biochemically characterised an *L. salmonis* cathepsin B-like cysteine protease (Cunningham et al., 2010) and localised it to the gut of the parasite (unpublished data), but there is little characterisation on other cysteine proteases of the parasite. Three reportedly complete *L. salmonis* cathepsin L-like proteins have been deposited in GenBank (GenBank accession nos. ADD24149, ADD24160 and EF490928), and a single putative incomplete cysteine protease nucleotide sequences was deposited in GenBank following an EST sequencing project (GenBank accession no. EF490847), but no further analysis on these sequences has been published. The present paper details two complete cathepsin L gene sequences and their molecular characterisation. We also show that cathepsin L activity is expressed in the different *L. salmonis* life stages and have biochemically characterised the activity in adult stages.

2. Materials and methods

2.1. Chemicals and parasites

The peptide substrate Z-Phe-Arg-NHMec was purchased from Bachem Biosciences (Germany). Protease inhibitors used included L-3-*trans*-(propylcarbamyl)oxirane-2-carbonyl)-L-isoleucyl-L-proline (CA074) which inhibits cathepsin B activity and *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane (E64) (Sigma St. Louis, Mo.) which inhibits both cathepsin L and B activity.

Adult *L. salmonis* were collected from infected salmon found on established fish farms in the West of Ireland. Female adult lice were separated and stored in RNAlater® (Ambion®, Applied Biosystems, Austin, TX 78744-1832) at −20 °C and subsequently used for library construction. Various life stages (chalimus I–IV, preadult male I–II, preadult female I–II, adult male, adult female and egg strings) were collected from infected fish and stored at −20 °C. Nauplius stages were hatched from egg strings as per Pietrak and Opitz (2004) and stored at −20 °C. Life stages of parasites processed into extracts were subsequently used for biochemical characterisation as per Section 2.5.

2.2. Construction of L. salmonis cDNA library

Approximately 120 adult female sea lice were homogenised under RNase-free conditions. Total RNA was isolated using RNeasy® Protect Midi Kit (Qiagen, CA 91355, USA) according to the manufacturer's instructions. mRNA was subsequently prepared with an Oligotex mRNA Midi kit (Qiagen). Poly-A RNA (6 μg) was used for cDNA library construction using a ZAP cDNA synthesis kit (Stratagene, CA 92037, USA). cDNA was digested with *Xho I* and resulting cDNA fragments were size fractionated on a Sepharose column and ligated into the Uni-ZAP vector. Following ligation, the whole library was packaged using Gigapack® III Gold packaging extract (Stratagene) according to the manufacturer's protocol.

2.3. L. salmonis cDNA library sequencing

In vivo mass excision was performed on the library according to manufacturer's recommendations. Plasmids were purified from resulting white colonies and glycerol stocks prepared. Over 1900 expressed sequence tags (ESTs) were sequenced commercially in one direction using the M13 reverse primer (5′-CATGGTCATAGCTGTTTCC-3′). Vector sequences were trimmed using the Sequencher program and BLAST analysis was performed on the resulting sequences against the GenBank database. The partial *L. salmonis* cathepsin L2 gene was identified by this method.

2.4. PCR amplification and sequencing of L. salmonis cathepsin L1 and L2 genes and phylogenetic analysis

Analysis of the GenBank database revealed four *L. salmonis* cathepsin L sequences in the public databases (GenBank accession nos. EF490928, EF490847, ADD24149 and ADD24160). *L. salmonis* cathepsin L1 was so named as it was most similar to the previously defined copepod crustacean *Caligus clemensi* cathepsin L1 (GenBank accession no. ACO15375). *L. salmonis* cathepsin L2 was named to continue this numerical sequence.

Based on alignments carried out on the GenBank accession no. EF490928, it was apparent that this sequence was incomplete. In order to complete this *L. salmonis* cathepsin L1, (LsCL1), rapid amplification of cDNA ends (RACE) was performed (First choice® RLM-RACE kit, Ambicon). Gene specific RACE reverse primers (outer primer: 5′-GAAATATTGTCCCTCTACAGAG-3′; inner primer: 5′AACCTTTGGCTAATCCAGT-3′) were designed based on the nucleotide sequence available and used to amplify the 5′ end of the gene, thus confirming a nucleotide insertion in the published sequence. In order to confirm a deletion present in the published sequenced, the 3′ end of the gene was amplified using the forward primer (5′-CTTATTGCTAATAAAGGTATAGCTAC-3′) and reverse primer (5′-TTAAATGACGGGATATGAAGCCAGTG-3′). The resultant cDNA fragments were ligated into the pGem-T Easy® vector (Promega, Madison, WI 53711, USA) for sequencing.

A partial sequence for LsCL2 was obtained following the sequencing of the *L. salmonis* cDNA library (Section 2.3). RACE was used to amplify the missing 5′ end of LsCL2 (outer primer: 5′-ACCAGTCACAGATGAGGCAT-3′; inner primer: 5′-GGCACCCTTACATCC-3′). cDNA fragments amplified by this method were sequenced and found to be identical to the published sequence (GenBank accession no. EF490847) albeit missing its 5′ nucleotide sequence.

Using the sequence data above, a full length LsCL2 gene was amplified from *L. salmonis* adult female cDNA using the forward primer (5′ATGAAAATTCATTTCAAGATG-3′) and the reverse primer (5′TTACTCAAATGTGGCTCC-3′) and ligated into the pGem-T Easy® vector for sequencing. Each PCR reaction contained 1X PCR Buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 1 U DNA taq polymerase, 10 pmol primer and cDNA template (0.1 μg/μl) cycled as follows: initial denaturation of 1 cycle of 95 °C for 3 min, 35 cycles of denaturation at 95 °C for 1 min, primer annealing at 50 °C for 1 min, extension at 72 °C for 1 min and a final extension of 1 cycle of 72 °C for 7 min.

The ExPASy Molecular Biological Server (http://ca.expasy.org/) (Gasteiger et al., 2005) provided the tools by which to analyse the secondary structure of the encoded LsCL1 and LsCL2 proteins. The program Prot Param available on ExPASy was used to predict the molecular weight and theoretical pI of the elucidated *L. salmonis* cathepsin L protein sequences. Glycosylation sites were predicted using the NetNGly 1.0 program available on ExPASy. The prediction of signal peptide cleavage sites was carried out using the SignalP program available on ExPASy.

Multiple alignments were performed on the UniProt Server (http://www.uniprot.org/) using the ClustalW algorithm (Thompson et al., 1994). Positioning of relevant cathepsin L conserved sites in Fig. 1 were determined following an alignment of the *L. salmonis* cathepsin L sequences (GenBank accession numbers: HM439291, HM439290, ADD24160 and ADD24149) and the closest matches of the elucidated LsCL1 sequence (*C. clemensi* cathepsin L1 precursor, *C. clemensi* cathepsin L precursor, *Caligus rogercresseyi* cathepsin L precursor and *L. vannamei* cathepsin L available at GenBank accession