Parasitol Res (2007) 100:881–886
DOI 10.1007/s00436-006-0344-3

SHORT COMMUNICATION

A cathepsin L-like cysteine proteinase gene from the protozoan parasite, *Cryptobia salmositica*

Palmy R. R. Jesudhasan · Chung-Wei Tan · Nikos Hontzeas · Patrick T. K. Woo

Received: 15 June 2006 / Accepted: 13 September 2006 / Published online: 17 November 2006
© Springer-Verlag 2006

Abstract The present study describes the identification of a cathepsin L-like cysteine proteinase gene (*CYS*) from the hemoflagellate *Cryptobia salmositica*. Genomic DNA sequence of cysteine proteinase was obtained by genome walking using degenerate primers. Specific primers were designed to amplify the cDNA of cysteine proteinase from mRNA by rapid amplification of cDNA ends-PCR. The open reading frame of *CYS* is 1,329 bp, with 443 deduced amino acids. Based on the sequence analysis, cysteine proteinase of *C. salmositica* is similar to the cathepsin L-like cysteine proteinase of kinetoplastid parasites such as *Leishmania* spp. and *Trypanosoma* spp. The identification of *CYS* proteinase gene could help to design cysteine proteinase specific inhibitors. Further studies are required to characterize the complete genomic organization of the cysteine proteinase.

Introduction

Cryptobia spp. are flagellated protozoa that are parasitic on/in invertebrates and vertebrates. They are usually either ectoparasites on the body surface or endoparasites in the digestive tract or in the blood of fishes. *Cryptobia salmositica* Katz is a hemoflagellate that has been recorded in all species of salmonids (*Oncorhynchus* spp.) (Woo 2003, 2006) in rivers and streams on the Pacific coast of North America. In streams, the parasite is transmitted indirectly by the leech *Piscicola salmositica* (Woo 2003, 2006); however, it can also be transmitted directly between fish under certain hatchery conditions. This parasite causes cryptobiosis in susceptible salmonids, and the clinical signs include exophthalmia, abdominal distension with ascites, anorexia, and anemia. Outbreaks of the disease have been recorded in hatcheries and in fish kept in sea cages. The pathogen causes high mortality in naturally and experimentally infected salmonids (Woo 2003, 2006). An attenuated strain of *C. salmositica* has been produced, and it has remained avirulent for at least 15 years. This strain is used routinely as an experimental live vaccine. It circulates in the blood of rainbow trout, *Oncorhynchus mykiss*, for months and does not cause disease but protects fish for at least 2 years (Woo 1994, 2006).

Cysteine proteinases have been identified in a wide variety of organisms (Barrett and McDonald 1980; Bazan and Fletterick 1988; North 1982; Morihara 1974; Souza et al. 1992), including *C. salmositica* (Zuo and Woo 1997, 1998a,b). Cysteine proteinases are important in parasite development and pathogenesis (Coombs and Baxter 1984; Coombs et al. 1982; Yamakami et al. 1995). They also allow the parasite to evade the host's immune system (Alexander et al. 1998) and are virulence factors in some pathogenic organisms (Mottram et al. 1996). Hence, cysteine proteinase is a potential target of antiparasite agents and is a good candidate as an immunogen in vaccine production. Though the presence of cysteine proteinase in *C. salmositica* has been identified, its gene sequence is not known. Hence, the aim of our study was to identify the gene sequence to pave the way for further research.

The nucleotide sequence data reported in this paper are available in the GenBank™ accession number: AY713477.

P. R. R. Jesudhasan (✉) · C.-W. Tan · N. Hontzeas · P. T. K. Woo
Department of Integrative Biology, University of Guelph,
Guelph, ON N1G 2W1, Canada
e-mail: jesudhasan@poultry.tamu.edu

Present address:
P. R. R. Jesudhasan
Department of Poultry Science, 420 Kleberg Center,
Texas A & M University,
College Station TX 77843, USA

In this paper, we describe the identification of the cysteine proteinase gene sequence of *C. salmositica* using BD-Universal GenomeWalker kit, cloning, and analysis of CYS protein.

Materials and methods

Nucleic acids preparation and GenomeWalking

Cryptobia salmositica (vaccine strain) grown in minimum essential medium (Woo and Li 1990) at stationary phase (approximately 7×10^6 parasites/ml) was used for genomic DNA extraction. Genomic DNA was isolated using a Nucleospin tissue kit (Clontech, Mountain View, CA, USA). The quality of DNA was checked on a 0.7% agarose gel. BD GenomeWalker Universal kit (Clontech) was used to identify the partial genomic DNA sequence of the *CYS* gene. Adaptor ligated genomic DNA "genomewalker libraries" of *C. salmositica* were prepared using instruction manual no. PT3042-1. These genomewalker libraries were used as templates for PCR amplification. Gene-specific degenerate primers were designed by aligning the amino acid sequences of cysteine proteinase of different species of *Trypanosoma* spp and *Leishmania* spp. using the ClustalW (1.81) program. Sequences used for the alignment were *Trypanosoma brucei* (XP_826279), *Trypanosoma congolense* (S37048) and *Trypanosoma cruzi* (AAF75546), *Leishmania major* (CAJ02287), *Leishmania mexicana* (CAA71085), *Leishmania infantum* (CAD12393), and *Leishmania donovani* (AAL09443). Conserved regions were selected from the aligned amino acid sequences and were translated into the simplest wild card codon representing all codons using the universal translation table. The translated codons were used as gene-specific primers (Table 1). Primary PCR was performed using adapter primer (AP1) supplied by the manufacturer and *CYS* gene-specific primer (GSPC1). One microliter of each library was added to an Advantage 2 Polymerase mix (Clontech) comprising of 1 µl (10 µM) of gene-specific primer, 1 µl of deoxyribonucleotide triphosphates (10 mM), 1 µl of AP1 (10 µM), 1 µl Advantage 2 Polymerase, 5.0 µl of buffer, and 40.0 µl of deionized H_2O. Parameters for PCR amplifications were 7 cycles of for 2 s at 94°C, 3 min 72°C and 32 cycles of 2 s at 94°C, 3 min at 67°C. Primary PCR product was diluted 1:50 and 1 µl of the diluted product was used as a template for nested-PCR with adaptor primer (AP2) and 1 µl (10 µM) of *CYS* gene-specific primer (GSPC2). Parameter for nest-PCR were 5 cycles of 2 s at 94°C, 3 min at 72°C and 20 cycles of 2 s at 94°C, of 3 min 67°C. The major secondary PCR products obtained were gel extracted using QIAquick gel extraction kit (Qiagen, Hilden, Germany) and sequenced and aligned with the help of Kodon. Complete genomic DNA sequence was obtained by further genomewalking (second walk) with gene-specific primers [primary PCR (GSP-F, GSP-R) and nested PCR (GSP-F1, GSP-R1)] (Table 1) designed from the partial genomic DNA sequence obtained from the initial walk. Primary and nested PCR were carried out with adaptor primers (AP1 and AP2).

Table 1 List of primers and degenerate primers used in this study

Primers	Sequence (5′–3′)	Tm* (°C)
Degenerate primers		
GSPC1	AARGGNGCNGTNACNCCNGTNAARGAY	
GSPC2	CCNTAYTGGATHATHAARAAYWSNTGG	
2nd walk		
GSP-F	GGGCGGCGACGATAACGGTCTTCATTG	69
GSP-R	CGCTCCTCGTCAGGAGACGCGTAGTTA	69
GSP-F1	ACGCCATGGGAGCCCCAACAACCG	69
GSP-R1	CACAACGGAAGAGCTGGGGTGGCTCG	70
cDNA		
GP-F	ATGAAGACCGTTATCGTCGC	56
GP-R	GAAGCCCCTAGGGAGCAG	59
5′ RACE-PCR		
GRR-P	GGCGTTGTGGCGGGTCTGGAACTCC	70
GRR-N	GTTGGGGCCGAAGGTGG	59
3′ RACE-PCR		
GRF-P	GCTGGGCAACACCGGCTACTTCGAGAA	69
GRF-N	TTCGAGAACATATGCACCAGC	57
PCR-ORF		
cys F	cggaattcATGAAGACCGTTATCGTCGC	56
cys R	ccctcgagGAAGCCCCTAGGGAGCAG	59

RACE rapid amplification of cDNA ends, *ORF* open reading frame

cDNA synthesis and rapid amplification of cDNA ends-PCR

Messenger RNA from stationary growth phase of *C. salmositica* was extracted using FastTrack 2.0 kit (Invitrogen, Carlsbad, CA, USA). cDNA was synthesized using Gene Amp® RNA PCR Kit (Applied Biosystems, Branchburg, NJ, USA). Messenger RNA was reversed transcribed to cDNA using an oligo (dT) primer. Gene-specific primers (GP-F and GF-R) were used to synthesize *CYS* gene cDNA. One microgram of mRNA was used as a template in the GeneRacer™ Kit (Invitrogen) following the manufacturer's instructions. To amplify the 5′ and 3′ ends of *C. salmositica* cathepsine L-like cystein proteinase gene, different primers (primary and secondary PCR) (Table 1) were designed from the genomic DNA sequence of cysteine proteinase and were used along with GeneRacer primers supplied by the manufacturer. The PCR reactions were carried out following the manufacturer's instructions. The resulting amplicons were purified and sequenced using dideoxy termination method (DNA sequencing facility,

College of Biological Science, University of Guelph, Guelph, Ontario, Canada) on an automated sequencer. To amplify the coding region of *CYS*, primers (cys F and cys R) were used (Table 1). *Eco*RI and *Xho*I restriction sites were introduced into the product for directional cloning. The amplified product was purified, cloned into pET24(a)$^{+}$ vector, and sequenced. Plasmid DNA was purified from recombinant colonies using QIAprep Spin Miniprep Kit (Qiagen), verified by restriction enzyme digestion, and sequenced.

Sequence and identity analysis

Sequence alignment was performed using Kodon (Applied Maths, Sint-Martens-Latem, Belgium). The GenBank database was searched using the blastX utility (Altschul et al. 1990). Analysis of the *C. salmositica* cysteine proteinase gene was performed using the blastX V.2.1.1 program (Altschul et al. 1997). To calculate its identity with other cathepsin L-like proteinases, complete sequences of cathep-

Fig. 1 Nucleotide sequence of the cDNA encoding *C. salmositica* cathepsin L-like (*CYS*) and its deduced amino acid sequence. The pyrimidine-rich nucleotide sequences of the N-terminal are *blocked* in *bold*. The signal peptide is *underlined*. The *arrow* indicates the N terminus. The fourteen Cys (C) residues forming disulphide bridges are in *bold, italics, and double underlined*. The conserved catalytic site residues [Cys138 (C), His277 (H), and Asn297 (N)] are *black boxed*. Amino acid residues forming the ERFNIN (ERFN) and GNFD (GNFD) motifs and that determining the substrate specificity Met (M) are *grey shaded*, the Gln (Q) residue of the oxyanion hole is *white lettered* and *grey shaded*. The GCNGG (GCNGG) motif is *boxed*. The *asterisk* marks the stop codon. The putative polyadenalation signal (AATAAA) is *blocked in bold and dotted underlined*. GenBank accession no. AY713477

```
                                tgcatatgtgtcaggaaatggcgcagtgcacaaaca
cacattttccggggtgtggatttttttttcaagtaaataaaaacaaccgcaactaagtga
atgaagaccgttatcgtcgccgcccttctgatggtgtgtaacgccatgggagccccaaca   60
M  K  T  V  I  V  A  A  L  L  M  V  C  N  A  M  G  A  P  T     20
Accgaggtcctgttcggcaacttcaaggccgcacatgcccgtaactacgcgtctcctgac  120
T  E  V  L  F  G  N  F  K  A  A  H  A  R  N  Y  A  S  P  D     40
Gaggagcgcaagcgctttgagatattcgccgggaacatgaagaaggccgcagtgctgaac  180
E  E  R  K  R  F  E  I  F  A  G  N  M  K  K  A  A  V  L  N     60
Cgcaaaaaccccatggccaccttcggccccaacgagtttgccgacatgacttccgaggag  240
R  K  N  P  M  A  T  F  G  P  N  E  F  A  D  M  T  S  E  E     80
Ttccagacccgccacaacgccgctcgccactacgctgctgcgaaggcccgcccacccaag  300
F  Q  T  R  H  N  A  A  R  H  Y  A  A  A  K  A  R  P  P  K    100
                                       ↓
Aacaccaagacatttactgcagaggagattaaggccgcggtcggccagcagatcgactgg  360
N  T  K  T  F  T  A  E  E  I  K  A  A  V  G  Q  Q  I  D  W    120
cgtttgaaaggagccgtgaccccagtgaagaatcaaggcgcttgcggttcctgctggagc  420
R  L  K  G  A  V  T  P  V  K  N  Q  G  A  C  G  S  C  W  S    140
Ttctctacaactggaaacatcgaaggacagcacgccattgccaccggtcagctagtggct  480
F  S  T  T  G  N  I  E  G  Q  H  A  I  A  T  G  Q  L  V  A    160
Gtttctgagcaggaacttgtttcctgtgaccccattgacgacgggtgcaatggcggtctc  540
V  S  E  Q  E  L  V  S  C  D  P  I  D  D  G  C  N  G  G  L    180
Atggacaacgccttcggatggctaatctctgcccacaagggccaaatcgcaaccgaggca  600
M  D  N  A  F  G  W  L  I  S  A  H  K  G  Q  I  A  T  E  A    200
Aactacccatacgtcagcggcaacggcattgtccccgcgtgctcatccagccctgagtcc  660
N  Y  P  Y  V  S  G  N  G  I  V  P  A  C  S  S  S  P  E  S    220
Aagcccgtcggagccaccatcagtgcctttcaggacatcgccagaaccgaggaggacatg  720
K  P  V  G  A  T  I  S  A  F  Q  D  I  A  R  T  E  E  D  M    240
Gcagcctttgtatttaagcatggaccactgtctattggagtggatgcctcgacctggcag  780
A  A  F  V  F  K  H  G  P  L  S  I  G  V  D  A  S  T  W  Q    260
Tcctacgcgggaggcatcatgtcctactgcccccaagatcagatcgaccacggtgtccta  840
S  Y  A  G  G  I  M  S  Y  C  P  Q  D  Q  I  D  H  G  V  L    280
Atcgtgggcttcgatgacaccgcctccaccccttactggatcataaagaactcctggaca  900
I  V  G  F  D  D  T  A  S  T  P  Y  W  I  I  K  N  S  W  T    300
Gccaactggggagaggagggctacatccgcgttgctaagggcagcaaccagtgtggtctg  960
A  N  W  G  E  E  G  Y  I  R  V  A  K  G  S  N  Q  C  G  L    320
Acgagccaccccagctcttccgttgtgggtaacagtcccagtccaaccccagcccccacc 1020
T  S  H  P  S  S  S  V  V  G  N  S  P  S  P  T  P  A  P  T    340
Acccccgggtccggctccctcatccagatgtactgcttcgacgacaagtgctcaaatggc 1080
T  P  G  S  G  S  L  I  Q  M  Y  C  F  D  D  K  C  S  N  G    360
Tgccgcaagaacaccttacccctgcacacgtgccttccattgaatggtggaggctctgca 1140
C  R  K  N  T  L  P  L  H  T  C  L  P  L  N  G  G  G  S  A    380
Attgcatcctgcaatcccatccaggtgatcctgtccatataccagaccatcgactgcacc 1200
I  A  S  C  N  P  I  Q  V  I  L  S  I  Y  Q  T  I  D  C  T    400
Ggcccttcccagcccaacgccatgtccctgaaccagtgtctgctgggcaacaccggctac 1260
G  P  S  Q  P  N  A  M  S  L  N  Q  C  L  L  G  N  T  G  Y    420
Ttcgagaacatatgcaccagcaacaccaacaccgccatgccaaagggtctcctgctccct 1320
F  E  N  I  C  T  S  N  T  N  T  A  M  P  K  G  L  L  L  P    440
aggggcttctaatgctccactccaggtgaaacgagaattaggggaagacggggaaataaa 1380

R  G  F  *                                                    443
Taaatatttatatatgagcaataaaaaaaaaaaaaaaa
```

sin L-like proteinases were obtained from the National Center for Biotechnology Information web site. Multiple sequence alignment was carried out using ClustalW programs. The deduced protein was analyzed using http://www.expasy.org/tools. The nucleotide and amino acid sequences were assembled using ExPASy translate tool (hosted by Canadian Bioinformatics Resource, Canada).

Results

Identification and characterization of cathepsin L-like cysteine proteinase gene

Partial cysteine proteinase genomic DNA sequence of *C. salmositica* was amplified using degenerate primers, along with BD GenomeWalker adaptor primers. Complete genomic DNA sequence was obtained by further genome walking using gene-specific primers (GSP-F, GSP-R, GSP-F1, and GSP-R1).

The isolated mRNA was reverse transcribed into cDNA by PCR using gene-specific primers. Rapid amplification of cDNA ends-PCR product of *CYS* gene encoding a cathepsin L-like proteinase with a 5′ noncoding region of 136 nucleotides and a 3′ noncoding region of 42 nucleotides downstream of the stop codon TAA and upstream of the polyadenylation signal AATAAA was obtained. The open reading frame of the *CYS* gene sequence was confirmed by direct amplification of genomic DNA of the vaccine strain using primers (cys F and cys R) designed from the 5′ and 3′ ends of the *CYS* gene. The genomic

Fig. 2 Alignment of the deduced amino acid sequence of *C. samositica* cathepsin L-like cysteine proteinase with different cathepsin L-like cysteine protinases of *Trypanosoma* and *Leishmania*. Identical and similar amino acid residues between the *C. salmositica* cathepsin L-like proteinase and others are indicated beneath the alignment sequences by *asterisks* and *dots*, respectively. The alignment was optimized by introducing gaps using ClustalW (1.83) multiple sequence alignment program. The *boxed* amino acid residue is located at the end of the predicted signal peptide according to calculations using SignalP 3.0 (CBS) program from the ExPASy proteomics tools. The conserved catalytic site residues Cys (C), His (H), and Asn (N) are *black boxed* and indicated by a *sharp sign* above the alignment; conserved Cys (C) residues and the ERFNIN-like and GNFD motifs, the oxyanion hole forming Gln (Q) residues and amino acid residues involved in determining enzyme substrate specificity, are *bold*, *grey shaded* and indicated by C, ERFN, and GNFD and above the alignments, respectively. The structurally important GCNGG motif is *boxed* and indicated by GCNGG. The GenBank accession numbers for the analyzed sequences are *L. major* (AAB48120), *L. mexicana* (CAA78443) *T. brucei* (CAA38238), *T. congolense* (CAA81061), and *T. cruzi* (AAA30181)

```
                             C                                  E
C. sal  ---------MKTVIVAALLMVCNAMGAP-------TTEVLFGNFKAAHARNYASPDEERK 44
T. bru  MPRTEMVRFVRLPVVLLAMAACLASVALGSLHVEESLEMRFAAFKKKYGKVYKDAKEEAF 60
L. maj  MATSRAALCAVAVVCVVLAAACAPARAI---YVGTPAAALFEEFKRTYQRAYGTLTEEQQ 57
L. don  MATSRAALCAVAVVCVVLAAACAPARAI---YVGTPAAALFEEFKRTYRRAYGTLAEEQQ 57
                      :     .* .  *        .    *  **  : : *    **

        R   F  N                G N F D
C. sal  RFEIFAGNMKKAAVLNRKNPMATFGPNEFADMTSEEFQTRH-NAARHYAAAKARPPKNTK 103
T. bru  RFRAFEENMEQAKIQAAANPYATFGVTPFSDMTREEFRARYRNGASYFAAAQKRLR---K 117
L. maj  RLANFERNLELMREHQARNPHARFGITKFFDLSEAEFAARYLNGAAYFAAAKQHAGQHYR 117
L. don  RLANFERNLELMREHQARNPHARFGITKFFDLSEAEFAARYLNGAAYFAAAKQHAGQHYR 117
        *:  *  *::        ** * ** . * *::  ** :*: *.* ::***: :     :

                                       C  #
C. sal  TFTAEEIKAAVGQQIDWRLKGAVTPVKNQGACGSCWSFSTTGNIEGQHAIATGQLVAVSE 163
T. bru  TVNVT--TGRAPAAVDWREKGAVTPVKDQGQCGSCWAFSTIGNIEGQWQVAGNPLVSLSE 175
L. maj  KARAD--LSAVPDAVDWREKGAVTPVKNQGACGSCWAFSAVGNIESQWAVAGHKLVRLSE 175
L. don  KARAD--LSAVPDAVDWREKGAVTPVKNQGACGSCWAFSAVGNIESQWARVGHGLVSLSE 175
        .  .    . .   :*** ********:** *****:**: ****.*   .   ** :**

             C      C                                     C
C. sal  QELVSCDPIDDGCNGGLMDNAFGWLISAHKGQIATEANYPYVSGNGIVPACSSSPESKPV 223
T. bru  QMLVSCDTIDSGCNGGLMDNAFNWIVNSNGGNVFTEASYPYVSGNGEQPQCQMNG--HEI 233
L. maj  QQLVSCDHVDNGCGGGLMLQAFEWVLRNMNGTVFTEKSYPYVSGNGDVPECSNSSE-LAP 234
L. don  QQLVSCDDKDNGCNGGLMLQAFEWLLRHMYGIVFTEKSYPYTSGNGDVAECLNSSK-LVP 234
        * *****  *.**.**** :** *::    * : ** .***.****  . *  .

                                                      C      #
C. sal  GATISAFQDIARTEEDMAAFVFKHGPLSIGVDASTWQSYAGGIMSYCPQDQIDHGVLIVG 283
T. bru  GAAITDHVDLPQDEDAIAAYLAENGPLAIAVDATSFMDYNGGILTSCTSEQLDHGVLLVG 293
L. maj  GARIDGYVSMESSERVMAAWLAKNGPISIAVDASSFMSYHSGVLTSCIGEQLNHGVLLVG 294
L. don  GAQIDGYVMIPSNETVMAAWLAENGPIAIAVDASSFMSYQSGVLTSCAGDALNHGVLLVG 294
        ** *  .  :   *  :**:: ::**::*.***::  .* .*::: *  : ::****:**

                     #                    C
C. sal  FDDTASTPYWIIKNSWTANWGEEGYIRVAKGSNQCGLTSHPSSSVVGNSPSPTP----AP 339
T. bru  YNDNSNPPYWIIKNSWSNMWGEDGYIRIEKGTNQCLMN-QAVSSAVVGGPTPPP----PP 348
L. maj  YNMTGEVPYWVIKNSWGEDWGEKGYVRVTMGVNACLLTGYPVSVHVSQSPTPGPNTTTTT 354
L. don  YNKTGGVPYWVIKNSWGEDWGEKGYVRVAMGKNACLLSEYPVSAHVPRSLTPGPGTESEE 354
         : ..  ***:*****   ***.**:*:  * * * :.  . *  *  . :* *

                    C    C   C         C            C
C. sal  TTPGSGSLIQMYCFDDKCSNGCRKNTLPLHTCLPLNGGGSAIASCNPIQVILSIY-QTID 398
T. bru  PPPPSATFTQDFCEGKGCTKGCSHATFPTGECVQTTGVGSVIATCGASNLTQIIYPLSRS 408
L. maj  HAPKRVTVKQITCTDYFCRKGCKTTVIPTKECLPNGAGGSFQMECGDHQVLKLTY-TSMN 413
L. don  RAPKRVTVEQVMCTDMYCREGCKKSLLTANVCYKNGGGGSSMTKCGPKKVLMCSY-SNPH 413
         .*   :. *  * .  * :**    :.   *    . **    *.  ::    *  .

        C             C           C       S2
C. sal  CTGPSQPNAMSLNQCLLGNTGYFENICTSNTNTAMPKGLLLPRGF 443
T. bru  CSGPSVPITVPLDKCIPILIGSVEYHCSTNPPTKAAR--LVPHQ- 450
L. maj  CTGEAKYTVTREGKCGISWSGSSKSICQYV---------------- 443
L. don  CFGPGLCLETPDGKCAPYFLGSIMNTCQYT---------------- 443
        * * .       .:*     *     *
```

DNA sequence of *CYS* obtained was identical to the cDNA sequence. It does not have introns and was assigned GenBank accession number AY713477. The open reading frame of 1,329 nucleotides, beginning with the presumed initial ATG codon, encoded a deduced protein of 443 amino acids (Fig. 1) with a predicted molecular weight of 49.1 kDa. There was a putative signal peptide cleavage site between the positions 17 and 18 in the amino acid sequence (preregion). The proregion contained 113 amino acid residues and the predicted mature enzyme had 330 amino acid residues. The estimated molecular weight of the mature protein was 36.63 kDa. The N terminus of the protein was composed of a putative hydrophobic signal peptide with a predicted cleavage site at Gly17. The mature protein had a GCNGG at position 175–179, and the active triad formed by Cys138, His277, and Asn297 (Figs. 1 and 2). The mature protein contained 14 cysteine residues in positions 135, 169, 176, 214, 270, 318, 352, 357, 361, 371, 384, 398, 413, and 425 for the formation of the seven disulfide bridges. The ERFNIN-like and GNFD motifs, both characterizing L-like cathepsin, were found at positions 41–52 and 69–75, respectively, of the pre–propeptide.

Multiple sequence analysis of the predicted amino acid sequence of the *C. salmositica* cysteine proteinase gene has 38–42% identity to cathepsin L-like enzymes of *Leishmania* and *Trypanosoma*, which belong to the same order Kinetoplastida. *Cryptobia salmositica* CYS has 42% identity to that of *T. brucei* (CAA38238), *T. congolense* (CAA81061), and *T. cruzi* (AAA30181) (Fig. 2).

Discussion

This study identified a cysteine proteinase gene sequence in the genome of *C. salmositica*. Based on the analysis of structurally conserved regions and the presence of motifs, it is considered to be a cathepsin L-like proteinase and shows similarity to the cathepsins of other kinetoplastid parasites such as *Leishmania* spp. and *Trypanosoma* spp. (Fig. 2), as it has both ERFN and GNFD motifs, which characterize cathepsin L-like proteinases. The cathepsin L-like proteinase precursors of *L. major* and *L. donovani* possess a signal peptide of 27 amino acids, that of *T. brucei* possesses one of 29 amino acids (Sakanari et al. 1997), and the predicted signal peptide in the preregion of *C. salmositica* cathepsin L-like proteinase contains 17 amino acids. It is composed predominantly of hydrophobic amino acid residues (von Heijne 1983). The proregion of *C. salmositica* cathepsin L-like sequence contains a region that shares high sequence similarity with cathepsin L sequences of *L. major*, *L. mexicana*, *T. brucei*, *T. congolense*, and *T. cruzi*. A region in the propeptide portion of all mammalian cathepsin H and L cysteine proteases has the "ERFNIN" motif ($EX_3RX_3FX_2NX_3IXN$) (Karrer et al. 1993), but the multiple alignment of the trypanosomatid cathepsin L-like sequences revealed that they all contained four of the conserved amino acids ($EX_3RX_3FX_2N$) (Figs. 1 and 2). The "ERFN" motif sequence ($EX_3RX_3FX_2N$) present in *C. salmositica* is similar to that of other trypanosomatid cathepsin L-like sequences. As in the case of *L. major* (Sakanari et al. 1997), cysteine proteinase in *C. salmositica* may function to inactivate its own mature enzyme, as was found by Taylor et al. 1995, in the proregions of papain and papaya proteinase IV. Hence, specific inhibitors could be designed based upon the proregion of the *C. salmositica* cysteine proteinase to control cryptobiosis in salmon.

Mature cathepsin L-like proteinases contain conserved Cys, His, and Asn in their catalytic center, six cysteine residues involved in disulfide bonds, a conserved GCNGG motif, an oxyanion hole at the Gln residue, and the substrate-specific site (Karrer et al. 1993; Rawlings and Barrett 1994; Sajid and McKerrow 2002). The role(s) of the cathepsin L-like CYS protein needs further elucidation for us to better understand its complete function and genomic organization.

Acknowledgements This study was supported by grants from the Natural Sciences and Engineering Research Council of Canada. We thank Ms. Mandi Elridge for the continuous supply of the *C. salmositica* and Ms. Angela Hollis (DNA Facility, College of Biological Science) for her cooperation in sequencing the DNA.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Alexander J, Coombs GH, Mottram JC (1998) *Leishmania mexicana* cysteine proteinase-deficient mutants have attenuated virulence for mice and potentiate a Th1 response. J Immunol 161:6794–6801

Barrett AJ, McDonald JK (1980) Mammalian proteinases: a glossary and bibliography, vol 1. Endopeptidases, 2nd edn. Academic, London, pp 182–187

Bazan JF, Fletterick RJ (1988) Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc Natl Acad Sci USA 85:7872–7876

Coombs GH, Baxter J (1984) Inhibition of *Leishmania* amastigote growth by antipain and leupeptin. Ann Trop Med Parasitol 78:21–24

Coombs GH, Hart DT, Capaldo J (1982) Proteinase inhibitors as antileishmanial agents. Trans R Soc Trop Med Hyg 76:660–663

Karrer KM, Peiffert SL, Ditomas ME (1993) Two distinct gene subfamilies within the family of cysteine protease genes. Proc Natl Acad Sci USA 90:3063–3067

Morihara K (1974) Comparative specificity of microbial proteinases. Adv Enzymol 41:179–143

Mottram JC, Souza AE, Hutchison JE, Carter R, Frame MJ, Coombs GH (1996) Evidence from disruption of the lmcpb gene array of *Leishmania mexicana* that cysteine proteinases are virulence factors. Proc Natl Acad Sci USA 93:6008–6013

North MJ (1982) Comparative biochemistry of the proteinases of eukaryotic microorganisms. Microbiol Rev 46:308–340

Rawlings ND, Barrett AJ (1994) Families of cysteine peptidases. Methods Enzymol 244:461–486

Sajid M, McKerrow JH (2002) Cysteine proteases of parasitic organisms. Mol Biochem Parasitol 120:1–21

Sakanari JA, Nadler SA, Chan VJ, Engel JC, Leptak C, Bouvier J (1997) *Leishmania major*: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. Exp Parasitol 85:63–76

Souza AE, Waugh S, Coombs GH, Mottram JC (1992) Characterization of a multi-copy gene for a major stage-specific cysteine proteinase of *Leishmania mexicana*. FEBS Lett 311:124–127

Taylor MA, Baker KC, Briggs GS, Connerton IF, Cummings NJ, Pratt KA, Revell DF, Freedman RB, Goodenough PW (1995) Recombinant pro-regions from papain and papaya proteinase IV-are selective high affinity inhibitors of the mature papaya enzymes. Protein Eng 8:59–62

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21

Woo PTK (1994) Parasitic flagellates in fishes. In: Kreier JP, Baker JR (eds) Parasitic protozoa, vol 8, 2nd edn. Academic, New York, pp 1–80

Woo PTK (2003) *Cryptobia* (*Trypanoplasma*) *salmositica* and salmonid cryptobiosis. J Fish Dis 26:627–626

Woo PTK (2006) Diplomonadida (Phylum: Parabasalia), Kinetoplastea (Phylum Euglenozoa). In: Woo PTK (ed) Fish diseases and disorders, vol 1: protozoan and metazoan infections, 2nd edn. CABI, Wallingford, pp 46–114

Woo PTK, Li S (1990) In vitro attenuation of *Cryptobia salmositica* and its use as a live vaccine against cryptobiosis in *Oncorhynchus mykiss*. J Parasitol 76:752–755

Yamakami K, Hamajima F, Akao S, Tadakuma T (1995) Purification and characterization of acid cysteine protease from metacercariae of the mammalian trematode parasite *Paragonimus westermani*. Eur J Biochem 233:490–497

Zuo X, Woo PTK (1997) Proteases in pathogenic and nonpathogenic hemoflagellates, *Cryptobia* spp. (Kinetoplastida) of fishes. Dis Aquat Org 29:57–65

Zuo X, Woo PTK (1998a) Characterization of purified metallo- and cysteine proteases from the pathogenic haemoflagellate, *Cryptobia salmositica* Katz 1951. Parasitol Res 84:492–498

Zuo X, Woo PTK (1998b) In vitro secretion of metalloprotease by the pathogenic haemoflagellate *Cryptobia salmositica* Katz 1951 and stimulation of the protease production by collagen. J Fish Dis 21:249–55