A THESIS

FOR THE DEGREE OF MASTER OF SCIENCE

COMPARATIVE ANALYSIS OF A CYSTEINE PROTEASE INHIBITOR, CYSTATIN B FROM A TELEOST *Oplegnathus fasciatus* AND TWO MOLLUSK SPECIES *Haliotis discus discus* AND *Ruditapes philippinarum*

H.K. Ajith Premachandra

Department of Marine Life Sciences

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

February 2013

COMPARATIVE ANALYSIS OF A CYSTEINE PROTEASE INHIBITOR, CYSTATIN B FROM A TELEOST *Oplegnathus fasciatus* AND TWO MOLLUSK SPECIES *Haliotis discus discus* AND *Ruditapes philippinarum*

H.K. Ajith Premachandra

(Supervised by Professor Jehee Lee)

A thesis submitted in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

February 2013

This thesis has been examined and approved by

_____

Thesis director, Choon Bok Song, Professor of Marine Life Sciences

_____

Seungheon Lee, Professor of Marine Life Sciences

_____

Jehee Lee, Professor of Marine Life Sciences

_____

Date

Department of Marine Life Sciences

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

REPUBLIC OF KOREA

CONTENTS

요약문

Cystatin 은 cysteine proteinase inhibitor 에 속하는 단백질로 항상성 유지, 염증 반응, 항체 생성, 암세포 전이, 면역반응, cathepsin-의존적 세포 사멸 등의 생체내의 생리학적 및 병원체에 관련된 다양한 생체공정에 관여하는 많은 단백질들이 이에 포함된다. 이러한 cystatin superfamily 는 살아있는 모든 생물에서 존재하는 것으로 알려져 있다. Cystatin 중에서 cystatin B 는 intracellular cysteine proteinases inhibitor 로 papain 과 같은 cysteine proteinase 에 결합하여 그 활성을 억제하는 역할을 하는 것으로 알려져 있으며, lysosome 으로부터 유출된 proteinases 로부터 보호 및 외부 침입으로부터 생물학적 방어에 관여하는 것으로 생각되고 있다. 상염색체 열성 질환인 신경성 발적의 원인인 cystatin B 의 결함은 인간에서 근간대성경련, 강직간대 발작등을 일으킨다고 알려져 있으며, 이러한 이유로 경제적으로 중요한 해양 생물에서 면역학적 관점에서의 기능적 특성 및 구조분석을 통한 Cystatin B 의 역할에 대한 이해는 매우 중요하다고 할 수 있다.

이 연구에서는 돌돔 (RbCyt B), 까막전복 (AbCyt B), 바지락 (McCyt B)으로부터 각각 cystatin B 유전자를 확인하였고, transcriptome 분석을 통하여 cDNA 서열을 확인 하였고, BAC library 를 제작하고 Cystatin 을 포함하는 BAC clone 의 Sequencing 을 통하여 genomic DNA 의 서열을 확인하였다. 돌돔과 까막전복의 genomic DNA 서열은 Spiedy 를 이용하여 exon 과 intron 및 구조를 확인하고 TF search online tool 을 이용하여 프로모터 부위의 예상되는 Transcription factor 결합부위를 분석하였다. Complete cDNA 확인 하고 단백질 암호화 부위를 확인하였다. 이를 이용하여 다른 종의 cystatin B 단백질들과 ClustalW pairwise alignment 와 sequence analysis 를 수행하였다. Cystatin B 의 orthologue 간의 관계를 확인 하기 위하여 보고된 cystatin family 의 아미노산 서열을 이용하여 계통분류학적 분석을 수행하였고, 구조 및 기능에 대한 이해를 위하여 I-TASSER sever for protein structure and function prediction 을 이영하여 3 차원 구조 modeling 을 수행하였다. 이 세 종으로부터 얻어진 cystatin 유전자의 단백질 암호화 부위를 pMAL-c2X expression vector 에 cloning 하고 재조합 단백질의 overexpression 및 정제 후

cysteine proteinase inhibitory assay 를 수행하였다. 또한 quantitative real-rime RT-PCR 을 통하여 조직특이적 발현과 면역 자극후의 발현변화를 관찰 하였다.

RbCyt B 와 AbCyt B 의 genomic sequence 는 프로모터 부위를 포함하여 약 3.85 Kb 와 8.4 Kb 를 나타내었다. RbCyt B 와 AbCyt B 의 Genomic 구조 분석은 다른 종의 cystatin 에서 나타나는 특징처럼 3 개의 exon 과 2 개의 intron 으로 구성되어 있는 것을 확인하였다. RbCyt B 와 AbCyt B 모두의 프로모터 부위에서 외부스트레스나 병원체 등에 의하여 활성화 되는 여러 전사인자 결합 부위를 찾을 수 있었다. RbCyt B 의 cDNA 전체 서열 길이는 614 bp, AbCyt B 는 1967 bp 그리고 McCyt B 는 509 bp 로 각각 300 bp, 303 bp, 297 bp 의 open reading frame (ORF)를 가지고 있어 각각 100, 101, 그리고 99 개의 아미노산 서열을 암호화 하고 있다. 세 개의 모든 Cyst B 분자들은 type I cystatin 그룹에 속하는 시스테인 분해효소 억제자의 특징인 Glycine 잔기, QxVxG motif 와 PW motif 가 잘 보존된 conserved cystatin-like 도메인을 포함하고 있으며, 약 11 kDa 정도의 분자량을 가질 것으로 예측되었다. 또한 이 세 종으로부터 확인된 3 개의 단백질 모두에서 type I cystatin 에서 보여지는 전형적인 특징이 모두 관찰되었다. RbCyt B 는 European sea bass 와 가장 높은 identity 와 similarity 를 나타내었으며, AbCyt B 와 McCyt B 는 각각 human 과 pacific oyster 와 가장 높은 identity 를 나타내었다. 계통분류학적 분석결과 이 연구에서 확인하고자 하는 RbCyt B, AbCyt B, McCyt B 들은 다른 cystatin B 단백질들과 함께 그룹을 형성하는 결과를 나타내었으며, 돌돔의 cystatin B 는 경골어류의 cystatin B 와 그룹을 형성하였고, 까막전복과 바지락의 Cystatin B 는 연체동물 군과 함께 그룹을 형성하였다. 3 차원 구조 모형화를 통하여 Cystatin B 의 stefin family proteinases 의 전형적인 특징을 확인하였다. 특히 첫 번째 hairpin loop 의 penta peptide conserved motif (QxVxG)와 PW motif, N-terminal 부위의 glycine 잔기는 papain-like cysteine protease 의 active site 에 highly compatible 한 wedged-shaped edge 형태를 나타내었고, 이는 구조적, 기능적으로 cystatin B 에 속함을 입증하고 있다. Cystatin B 의 단백질 암호화 부위 서열을 pMAL-c2X expression vector 에 클로닝하여 MBP 와 fusion 된 단백질을 발현시키고 생화확적 기능분석을 하였다. 돌돔,

까막전복 바지락의 세가지 재조합 단백질들은 azo-casein 의 가수분해에 대한 papain 의 proteolytic 활성에 대하여 눈에 띄는 protease inhibitory activity 를 보였다. Papain 과 cystatin B 가 1:1 로 존재하는 상태에서 papain activity 는 재조합 cystatin B 에 의하여 80% 이상의 저해 활성을 나타내었다.

돌돔, 까막전복, 바지락으로부터 확인한 Cystatin B 는 모든 조직에서 조직 특이적 발현이 관찰되었다. 돌돔에서는 다른 조직에 비하여 간, 비장, 아가미에서 유의적으로 많은 발현을 나타내었으며, 전복과 바지락에서는 hemocyte, 아가미, mantle 그리고 소화관에서 높은 발현을 나타내었다. 이는 외부와 자주 접촉하는 기관으로 미생물과의 잦은 접촉으로 인한 면역반응 조절에 따른 것으로 생각되었다.

하지만 박테리아와 pathogen associated molecules (PAMs)를 통한 면역 자극 후에는 면역조직에서 유의적인 발현 증가를 나타내었다. 돌돔의 신장과 비장에서는 *Edwardsiella tarda* 의 감염에 따라 유의적인 발현 증가를 나타내었다. *Vibrio parahaemolyticus* 를 처리한 전복과 *Vibrio tapetis* 를 처리한 바지락에서 모두 아가미와 hemocyte 에서 mRNA 의 발현이 증가하는 것을 나타내었으나, 바지락에서는 LPS 를 처리하였을 때 gill 과 hemocyte 모두에서 발현량이 상대적으로 감소하는 결과를 나타내었다. Poly I:C 를 처리한 바지락에서는 Cystatin B 의 mRNA 가 up-regulation 과 down-regulation 이 반복적으로 일어나는 복합적인 발현양상을 나타내었다.

결론적으로, 이 연구에서는 돌돔, 까막전복, 바지락의 세가지 종으로부터 획득한 cystatin B 분자의 비교분석에 의한 접근을 통하여 일반적인 구조와 기능적 특성이 진화적으로 잘 보존되어 있음을 확인 하였다. 이런 이유로, 이 연구는 면역학적 관점에서 경골어류와 해양 무척추동물의 세포생물학적인 cystatin B 의 특성에 대한 이해를 하는데 매우 가치 있는 자료로 사용될 수 있을 것이다.

SUMMARY

Cystatins are a large family of cysteine proteinase inhibitors which are involved in vast array of physiological and pathological processes where the cysteine proteinases participate in, including protein homeostasis, inflammatory responses, antigen processing, metastasis, immune responses, and cathepsin dependent apoptosis. The cystatin superfamily is consisted of the members representing all living organisms studied to date. Among cystatins, cystatin B is an intracellular cysteine proteinase inhibitor which is a tightly-binding reversible inhibitor of papain-like cysteine proteases. Cystatin B is also thought to play a role in protection against the proteinases leaking from lysosomes, and in biological defense system against invaders. Defects in cystatin B cause progressive myoclonic epilepsy type 1 (EPM1), which is an autosomal recessive disorder characterized by severe, stimulus-sensitive myoclonus, and tonic–clonic seizures in human. Hence, it is important to understand the role of cystatin B in economically important marine species also with respect to its structural and functional properties in immunological perspectives.

In the present study, three cystatin B genes were isolated from three different marine species, rock bream (*Oplegnathus fasciatus*), disk abalone (*Haliotis discus discus*) and Manila clam (*Ruditapes philippinarum*) either by transcriptome, cDNA and/or BAC library screening and analysis. The full-length cDNA sequences from all three species were identified from respective transcriptome or cDNA libraries and additionally, genomic sequences from rock bream cystatin B (RbCyt B) and disk abalone cystatin B (AbCyt B) were identified in respective BAC libraries. Genomic DNA sequences from both rock bream and abalone were analyzed using Spiedy and TFSEARCH online tools in order to identify the exon-intron architecture and transcription factor (TF) binding sites in the promoter regions, respectively. Complete cDNA sequences were used to identify the coding regions and both amino acid and nucleotide sequences were compared with other known cystatin B sequences

from different taxonomic classes using ClustalW pairwise and multiple sequence analysis. Phylogenetic analysis was performed considering all reported cystatin families in order to establish the relationship between orthologous. In order to understand and compare the structural and functional relationship, 3D structural modeling was performed using I-TASSER server for protein structure and function prediction. The coding regions of three genes cloned individually into pMAL-c2X expression vector and the purified recombinant proteins were subjected to cysteine proteinase inhibitory assay under *in vitro* conditions. Finally, quantitative real-time PCR was performed in order to determine the tissue specific mRNA expressions in healthy animals and regulated transcriptional behavior after induction of animals with immune stimulants and live pathogens.

Genomic sequences of RbCyt B and AbCyt B were approximately 3.85Kb and 8.4 Kb in length including promoter regions, respectively. Genomic structure analysis indicated common structural architecture characterized by three exons and two introns in all compared cystatin B sequences from different taxonomic classes including RbCyt B and AbCyt B. In addition, some exons exhibited identical length in all species. Several transcriptional factor binding sites, which are activated by physiological and pathological stimuli, were found in both RbCyt B and AbCyt B promoter regions. The full-length cDNA sequences of RbCyt B (614 bp), AbCyt B (1967 bp) and Manila clam cystatin B (McCyt B; 509 bp) contain open reading frame (ORF) of 300 bp, 303 bp and 297 bp which encodes for a 100, 101 and 99 amino acid polypeptides, respectively. All three cystatin B molecules indicated the predicted molecular mass around 11 kDa and three polypeptides contained conserved cystatin-like (CY) domain characterized by cysteine protease inhibitor signature, conserved Gly, QxVxG motif and PW motif, confirming the resemblance of type 1 cystatins. In addition, characteristic features of the type 1 cystatins were found in all three polypeptides. The RbCyt B exhibited highest identity and similarity to that of European seabass (*Dicentrarchus labrax*) while

AbCyt B and McCyt B exhibited highest identity with the corresponding orthologous from human (*Homo sapiens*) and Pacific oyster (*Crassostrea gigas*), respectively. Phylogenetic tree analysis indicated clustering of three novel cystatin B members in the sub cluster of cystatin B, where RbCyt B positioned in teleost clade while AbCyt B and McCyt B positioned in mollusk clade. The 3D structural models of cystatin Bs, demonstrated the typical features of the stefin family proteinases. Importantly, the penta peptide conserved motif (QxVxG) in the first hairpin loop, PW motif, and the glycine residue located close to the N-terminal of the molecule manifest that the potential of forming a wedged-shaped edge which is highly compatible to the active site of the papain-like cysteine proteases, substantiating its structural and functional relationship with known cystatin B members.

In order to characterize the biochemical properties of cystatin B proteins, the coding sequences were cloned into the pMAL-c2X expression vector and expressed as a fusion protein with MBP. The recombinant proteins showed remarkable protease inhibitory activity against the proteolytic activity of papain on azo-casein hydrolysis. Over 80 % of the papain activity was inhibited by recombinant cystatin Bs at the concentration ratio of 1:1 with papain.

In the present study, we observed that all three cystatin Bs were universally expressed in different tissues examined where, significantly higher expressions were detected in liver, spleen and gills from rock bream as well as in hemocytes, gills, mantle and digestive tract from abalone and Manila clam, which are believed to be either function in immune regulation or frequently contact with microbes. However, after the immune stimulations with either live bacteria or pathogen associated molecules (PAMs) indicated significant induction of relative mRNA expression in immune tissues examined. Both head kidney and spleen from rock bream indicated significant up-regulation of relative mRNA expression after infection with live bacterial (*Edwardsiella tarda*) pathogen. Although, bacterial infections (*Vibrio parahaemolyticus* in abalone and *Vibrio tapetis* in Manila clam) in abalone and Manila clam

indicated up-regulation of relative mRNA expression in gill and hemocytes, induction with LPS indicated down-regulated transcriptional profile in both gill and hemocytes from Manila clam. In contrast, poly I:C-stimulated clams showed a complex behavior of its McCyt B transcriptional profile along with two early and late phase positively regulated responses in both gill and hemocytes.

In conclusion, the present approach in comparative analysis of cystatin B molecules from three species demonstrated common structural and functional features conserved in the evolutionary process regardless of species difference. Hence, this study will be valuable and make significant contribution to understand the properties of cystatin B gene in cellular biology of marine teleost and mollusk species in immunological perspectives.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

‰	Parts per thousand
3D	Three dimension
AbCyt B	Disk abalone cystatin B
AIDS	Acquired immunodeficiency syndrome
Am	Adductor muscle
ANOVA	Analysis of variance
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
bp	Base pair(s)
cDNA	Complementary deoxyribonucleic acid
CFU	Colony forming unit(s)
C-terminal	Carboxyl terminal, COOH-terminus
CY	Cystatin-like domain
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide-triphosphate
Dt	Digestive tract
ECM	Extra cellular matrix
EPMI	Progressive myoclonic epilepsy type 1
EST	Expressed sequence tag(s)
FAO	Food and Agriculture Organization of the United Nations
Ft	Foot
Gl	Gill
GS-FLX	Genome sequencer FLX
Hc	Hemocytes
HMWK	High molecular weight kininogens
Hp	Hepatopancreas
IFN	Interferon
IL	Interleukin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase(s)
kDa	Kilo Dalton
KGP	Lys-gingipain
LB	Luria-Bertani
LMWK	Low molecular weight kininogens
LPS	Lipopolysaccharides
MBP	Maltose binding protein
McCyt B	Manila clam cystatin B
MEGA	Molecular Evolutionary Genetics Analysis

MHC	Major Histocompatibility Complex
MMP	Matrix metalloproteases
mRNA	Messenger RNA
Ms	Muscle
Mt	Mantle
NCBI	National Center for Biotechnology Information
NC-IUBMB	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
nm	Nanometer(s)
NO	Nitric oxide
N-terminal	Amino terminal, NH_2-terminus
OD	Optical density
ORF	Open reading frame
p.i.	Post infection/ post induction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
Poly I:C	Polyinosinic:polycytidylic acid
PPB	Potassium phosphate buffer
qPCR	Quantitative real-time polymerase chain reaction
rAbCyt B	Recombinant disk abalone cystatin B fused with MBP
RbCyt B	Rock bream cystatin B
RCSB	Research Collaboratory for Structural Bioinformatics
RGP	Arg-gingipain
rMcCyt B	Recombinant Manila clam cystatin B fused with MBP
rProtein	Recombinant protein(s) fused with MBP
rRbCyt B	Recombinant rock bream cystatin B fused with MBP
SDS	Sodium dodecyl sulfate
Sp	Siphon
STAT	Signal transducer and activator of transcription
Taq	*Thermus aquaticus*
TF	Transcription factor
TNF	Tumor necrosis factor
TSS	Transcription start site
Tt	Testis
UTR	Untranslated region
α	Alpha
β	Beta
γ	Gamma
μL	Microliter(s)

1. INTRODUCTION

Importance of the Three Experimental Marine Organisms

Over the past decade, marine food sources have overtaken a large portion of the world food market, representing a promising supply of nutritional resources to compensate for the decrease in terrestrial sources that has resulted from overconsumption. This trend has promoted mariculture to the status of an economically important component of the global economy. Aquaculture industry has been developed with the advancing scientific technology and mariculture is one of the rapid developing industries in worldwide which provide essential fatty acids and essential amino acids along with enriched source of other nutrients. Many fish and shellfish species are used either as cultured or capture fisheries, where rock bream (*Oplegnathus fasciatus*) (Fig. 1A) is one of the most economically important and widely consuming fish species, mostly in eastern Asia. However, recently the decrease production was associated with disease outbreaks, especially the viral diseases (irido virus), bacterial diseases (*Edwardsiella tarda*) and unfavorable environmental conditions with the growing intensified industry. In contrast, shellfish culture also makes huge contribution to the industry and to the national economy especially in East Asian countries including Korea, Japan and China. Disk abalone (*Haliotis discus discus*) (Fig. 1B) is a valuable marine gastropod species (Phylum: mollusks), widely distributed and cultured in Asia (McBride, 1998). The flesh of abalones is considered to be a desirable and highly nutritious food, and is consumed either in raw or cooked form in a variety of different dishes. Manila clam (*Ruditapes philippinarum*) (Fig. 1C) also is an edible marine bivalve mollusk species, which is harvested in large scale by the mariculture industry, and naturally distributed in intertidal zones of Yellow sea bordering countries such as Korea, Japan, China and Philippines. Later, it was introduced as commercially valuable species to many parts of the world (Goulletquer, 1997) and according to the FAO fishery statistics current global aquaculture production of

Manila clam exceeds 3.5 million tons per year (FAO Fishery Statistics, http://www.fao.org/fishery/culturedspecies/Ruditapes_philippinarum/; accessed on 27/10/2012). However, significant mortality is present in marine mollusks including disk abalone and Manila clam aquaculture due to many reasons, including pathogens, stressful environments, pollutants, and disease outbreaks (Villena, 2003; Hooper et al., 2007). Recently, disease management is taken the most important and critical factor to minimize the production losses which target basically either by pathogen control with chemotherapeutics or by host protection with vaccines and immune-stimulants (Park, 2009). As such, there is significant interest in gaining a detailed understanding of the host-pathogen interactions and immune modulations in these mariculture creatures to prevent drastic losses of their harvest due to different pathogenic infections and the studies on immunogenetics will provide more precise approaches to develop new strategies of efficient disease control.

A. Rock Bream (*Oplegnathus fasciatus*) B. Disk Abalone (*Haliotis discus discus*) C. Manila Clam (*Ruditapes philippinarum*)

Fig. 1. Three marine species used in the research.

Cysteine Proteases and Their Multifunctional Role in Cellular Biology

Proteolysis is a universal mechanism mediated by proteases (also termed as peptidase and proteinase) where amide bonds (peptide bonds) hydrolyzed in protein catabolism. Proteases occur naturally in all organisms and are involved in multifaceted roles of physiological reactions from simple food protein digestion to complicated immune signaling

pathways (e.g., the blood-clotting cascade, the complement system, apoptosis pathways, and the invertebrate prophenoloxidase-activating cascade) (Vasiljeva et al., 2007; Brix et al., 2008). With the completion of human genome study, recently it was reported that about 2% of the all gene products are proteases in human genome (Turk and Turk, 2008).These proteases are currently categorized into six broad groups depending on the reactive groups at the active site involved in the catalysis by the International Union of Biochemistry and Molecular Biology, into serine proteases (EC 3.4.21), cysteine proteases (EC 3.4.22), aspartate (EC 3.4.23), metalloproteases (EC 3.4.24), threonine proteases (EC 3.4.25) and another unknown type of proteases (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)). Cysteine proteases, the most well studied protease is the thiol (-SH)-dependent group, which utilizes cysteine residues as nucleophiles to attack the amide bonds of a target protein (Chapman et al., 1997; Potempa et al., 2005), have been found in viruses, fungi, bacteria, protozoa, plants and mammals (Barrett, 1986; Mizuno et al., 1987; Sharma et al., 1989). The cysteine proteases classified into 21 families (C1–C21) according to distinctive sequences and tertiary structures; however, the majority of cysteine proteases belong to the C1 family which are largely known as papain-like proteases including the lysosomal cysteine proteases, cathepsins (Dubey et al., 2007). Traditionally, cysteine proteases had been known as the lysosomal mediators of terminal protein degradation; however recent findings in human biology have been illustrated more expanded role of cysteine proteinase including apoptosis, MHC class II immune responses, prohormone processing, and extracellular matrix remodeling which is important to the bone development (Chapman et al., 1997). Cysteine cathepsins are lysosomal proteases with housekeeping as well as highly specialized functions, for an example, matrix metalloproteases (MMP) are primarily responsible for the homeostasis of the extracellular matrix, whereas under disease conditions, cysteine proteases contribute to destruction of

extracellular matrix (Cudic and Fields, 2009; Bromme and Wilson, 2011). In the pathogenic micro-organisms including bacteria, fungi, viruses and parasites, cysteine proteases can act as virulence factors, which promote diseases in the host organisms (Takahashi et al., 1994; Mottram et al., 2004; Rudenskaya and Pupov, 2008). For an example, proteolytic enzymes produced by *Porphyromonas gingivalis* are important virulence factors, produced in large quantities to cause periodontal disease. It has been shown that these proteases can directly or indirectly degrade constituents of the periodontal tissues, destroy host defense elements, dysregulate coagulation and complement kallikerin-kinin cascades (Dubey et al., 2007). Furthermore, parasite derived cysteine proteases were reported to play key roles in immunoevasion, enzyme activation, virulence, tissue and cellular invasion as well as excystment, hatching and moulting (Sajid and McKerrow, 2002). Thus, many of imbalanced activity of cysteine proteinases may lead to wide range of human diseases including cancer (Joyce et al., 2004; Turk et al., 2004), rheumatoid arthritis, osteoarthritis, osteoporosis (Yasuda et al., 2005; Vasiljeva et al., 2007), tumorigenesis and multiple sclerosis (Berdowska and Siewinski, 2000) along with neurological disorders. Therefore, certainly the eukaryotic systems have evolved precise regulatory mechanisms for cysteine proteinases, modulating their expression, secretion, and maturation through specific degradation of mature enzymes and targeted inhibition largely through interaction with their endogenous inhibitors called cystatins (Rzychon et al., 2004).

Cysteine Protease Inhibitors

The activity of the cysteine proteinases are regulated by naturally occurring inhibitory proteins such as α2-macroglobulin and cystatins (Bobek and Levine, 1992). These inhibitors protect the cells from destructive endogenous or external proteolysis and/or could be involved in the control mechanism responsible for intracellular or extracellular protein

degradation by cysteine proteinases of host, bacterial, fungi, parasites and viral origin. The first cysteine proteinase inhibitor to be isolated and characterized was obtained from chicken egg white in 1963 and was shown to inhibit ficin, papain (Fossum and Whitaker, 1968; Sen and Whitaker, 1973) and cathepsin B and C (Keilova and Tomasek, 1974). Later, the name "cystatin" was applied to this protein (Barrett, 1981), and it was shown to be a very potent inhibitor of other cysteine proteinases of the papain superfamily, such as cathepsin H and L (Anastasi et al., 1983). Subsequently, cystatins have been isolated from many tissues and body fluids of humans and different animal species (Turk et al., 1985; Barrett et al., 1986) as well as from plants such as rice (Abe et al., 1988) and soybean (Brzin et al, 1990).

Cystatin Superfamily

The cystatins are competitive, reversible, tight binding cysteine proteinase inhibitors which display structural and functional similarities. Cystatins are considered as the largest and the best described group of natural exosite binding cysteine proteinase inhibitors, which obstruct the access of substrate without interacting with the enzyme catalytic center (Bode and Huber, 2000). Cystatins predominantly interact with cysteine proteases, such as the plant-derived papain and human lysosomal cathepsins B, H, and L (Barrett, 1987; Turk and Bode, 1991). The cystatins are now considered as cystatin superfamily, which is mainly divided into three families on the basis of their location, size and complexity of polypeptide chains as type 1 cystatins (Stefins) (Stefins A and B, also known as cystatins A and B), type 2 cystatins (Cystatins), and type 3 cystatins (Kininogens). Family 1 cystatins are comprised of cystatin A and B, which are unglycosylated inhibitors and characterized by low molecular weight (~11 KDa) and a single cystatin-like (CY) domain structure. This group of cytoplasmic proteins is composed of ~100 amino acid residues that lack of disulfide bond, signal peptide, and carbohydrate side chain (Barrett et al., 1986; Turk and Bode, 1991).

Family 2 cystatins characterized with molecular masses in the range of 13–14 kDa, and contain signal peptide and disulfide bonds at the carboxy terminus. However, some members may also be glycosylated or phosphorylated. Examples of these are cystatins E/M (glycosylated on N^{108}) and cystatins S and SN (consensus phosphorylation sites at S^{2} and S^{98}, respectively). The family 2 cystatins include cystatins C, D, E/M, F, G, S, SA, and SN (Cornwall and Hsia, 2003). In contrast to stefins, cystatins contain a signal sequence for secretion through the cell membrane to the extracellular space. Family 3 cystatins (kininogens) exhibit high structural complexity with more than one CY domain, which are large multifunctional glycoproteins with inhibitory activities (Kellermann et al., 1987). Kininogens are found in the plasma and secretions of mammalian species. Three types of kininogens are identified to date as the low-molecular-weight (LMWK), high-molecular-weight (HMWK), and T-kininogens which are single-chain glycoproteins with the molecular weight of ~50 to 120 kDa. They are comprised of three CY domains that resulted from gene duplication. They contain additional disulfide bonds and are also glycosylated. Kininogens have an additional polypeptide at the C terminus containing the bradykinin sequence, which can be cleaved by kallikrein (Bobek and Levine, 1992). Recently, another category of the cystatin superfamily, family 4, was reported. This new family includes cystatins with invertebrate origin and its members are mostly from parasitic nematodes (Khaznadji et al., 2005; Li et al., 2010).

Structure and Function of Cystatins

All cystatins contain a papain-binding site that binds to the catalytic site of papain-like proteinases, and inhibits them reversibly. This site is created by several conserved regions of the protein, including a glycine in the N-terminal region, a central glutamine-valine-glycine (QxVxG) motif in first hairpin loop of the protein, and a PW motif in the

second hairpin loop in the C terminal region (Björk and Ylinenjärvi, 1989; Margis et al., 1998; Rzychon et al., 2004). Each cystatin structure has a core of five-stranded anti-parallel β-sheets wrapped around a central five turn α-helix as shown by the ribbon model of the structure (Fig. 2).

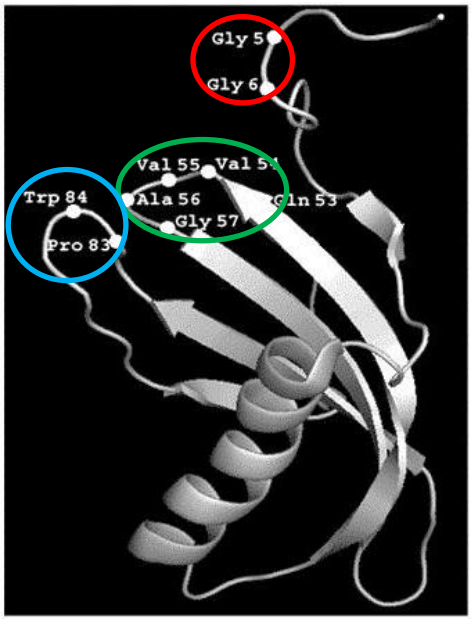

Fig. 2. Oryzacystatin three-dimensional structure. Ribbon representation of Oryzacystatin shows a five turn α-helix and five stranded antiparallel β-pleated sheet. The amino acids of the three functional domains of cystatins are encircled. (Source: Rodriguez-Mahillo et al., 2007).

Like cysteine proteinases, cystatins have been found in diverse organisms. Cystatins are involved in both biological and pathological processes which cysteine proteinases participate in, including protein homeostasis, inflammatory responses, antigen processing, metastasis, immune responses, and cathepsin dependent apoptosis (Synnes, 1998; Abrahamson et al., 2003; Kopitar-Jerala, 2006; Lefebvre et al., 2008; Shah and Bano, 2009). Additionally, there are many diseases observed with decreased cystatin levels, such as cancer, inflammatory diseases, osteoporosis, diabetes, neurodegenerative diseases and renal failure (Fi. 3) (Turk and Turk, 2008).

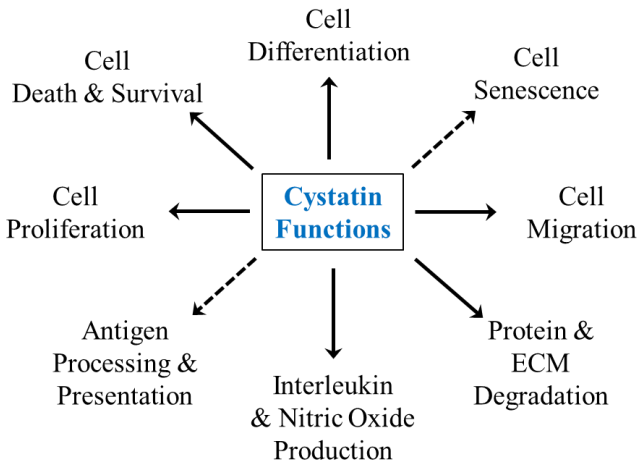

Fig. 3. Diverse biological functions of the cystatins. The cytoplasmic stefins and cell-secreted cystatins have been implicated in a multitude of biological processes, including cell differentiation, proliferation, survival, migration and interleukin production, leading under some instances to the increased production of NO. (Source: Keppler, 2006).

Cystatin in health and diseases

A large number of normal and pathological processes in the cellular environment are controlled by the balance between proteases and their inhibitors. The general physiological role of cystatins is believed to be the protection of cells from inappropriate endogenous or exogenous proteolysis by regulating the activity of cysteine proteinases both of host and of microbial origin. An increase level in local proteinases and proteinase inhibitors has been reported in the presence of inflammation (Keilova and Tomasek, 1977; Jarvinen, 1978), acquired immunodeficiency syndrome (AIDS) (Alavaikko et al, 1985), and malignancy (Okumichi et al., 1984; Sloane and Honn, 1984). It was also suggested that the cysteine proteinase, cathepsin B play a role in the cancer cell migration from their site of origin to secondary or metastatic sites by where they dissolve the extracellular matrix proteins to pave the way; however cystatins play a significant role at here to inhibit the cathepsin activity (Sloane and Honn, 1984). Cystatins have been shown to possess a wide range of effects in immune cells. For an example, cystatins involve in nitric oxide production by induction of tumor necrosis factor alpha (TNFα) and interleukin (IL) 10 synthesis through the interferon

gamma (IFNγ) activated murine macrophages. In turn, nitric oxide has inhibitory activity on cysteine proteases, especially those from parasitic protozoa (Vray et al., 2002). Furthermore, in the neurodegenerative disorder known as progressive myoclonus epilepsy of Unverricht-Lundborg type 1 (EPMI), a significant reduction in cystatin B activity and increased levels of cathepsin B, L, and S were observed in lymphoblastoid cells of EPMI (Rinne at al., 2002).

Cystatin B

Among cystatins, cystatin B is an intracellular cysteine proteinase inhibitor which is a tightly-binding reversible inhibitor of cathepsins by forming a dimer stabilized by non-covalent forces to inhibit C1 family cysteine proteases; papain and cathepsin B, H, and L (Anastasi et al., 1983; Cimerman et al., 2001; Lefebvre et al., 2008). Cystatin B (also known as stefin-B), was originally discovered as an inhibitor of cathepsin B (Lenney et al., 1979) and showed wide distribution among different cell types and tissues (Turk and Bode, 1991). Apart from its anti-protease activity, cystatin B has been shown to be involved in immune responses to bacterial and fungal infections, and in anti-apoptotic processes in the cerebellum (Takahashi et al., 1994; Di Giaimo et al., 2002). Cystatin B was also thought to play a role in protection against the proteinases leaking from lysosomes, and in biological defense system against invaders (Lefebvre et al., 2004; Li et al., 2010; Xiao et al., 2010). In contrast, defects in cystatin B gene cause progressive myoclonic epilepsy type 1 (EPM1), which is an autosomal recessive disorder characterized by severe, stimulus-sensitive myoclonus, and tonic–clonic seizures in human (Pennacchio et al., 1996; Riccio et al., 2005). Furthermore, the genes involve in proteolysis, apoptosis and glial activation including, cathepsin S, C1q B-chain of complement, β2-microglobulin, glial fibrillary acidic protein, apolipoprotein D, fibronectin 1 and metallothionein II were found to elevate the transcript level in neurological tissues from cystatin B knockout mice (Lieuallen et al., 2001). Hence, cystatin B is unique

among the members in cystatin superfamily which is involved in variety of pathological processes and having a unique structure containing a free cysteine in the N-terminal segment of the proteinase-binding region that facilitates the tight binding of target proteinases. Cystatin B reversibly inhibits the papain-like cysteine proteinases, and was found to play a protective role in organisms due to its omnipresent distribution in cells and tissues (Rahman et al., 1983; Abrahamson, 1994). It most likely prevents inappropriate proteolysis caused by the action of lysosomal cysteine proteinases, primarily the cathepsins B, H, K, L, and S (Turk et al., 1997). In addition, cystatin B has been implicated in innate immune responses to bacterial infections in vertebrate and invertebrate lineages including human, turbot fish (*Scophthalmus maximus*) and leech (*Theromyzon tessulatum*) (Suzuki et al., 2000; Lefebvre et al., 2004; Xiao et al., 2010), as well as in responses against viral infections (Luciano-Montalvo and Melendez, 2009). These observations may reflect the potential inhibitory action of cystatin B toward pathogen-derived cysteine proteinases which are secreted as virulent factors or as suppressors of host cysteine proteinases to protect the cells from inappropriate proteolysis.

Hence, this study has been focused on the molecular characterization of the cystatin B gene from three different marine species, including one teleost and two mollusk species; to investigate the role of cystatin B-like proteins in marine organisms from genomic to protein level. Furthermore, to understand the role in innate immune response, the temporal mRNA expression analysis was employed after different pathological infections.

2. MATERIALS AND METHODS

2.1. Experimental animals

2.1.1. Rock bream (*Oplegnathus fasciatus*)

Healthy rock breams with an average body weight of 45-50 g were provided by the National Fisheries Research and Development Institute (Republic of Korea), and they were maintained in 40 L tanks at 24 ˚C with sand-filtered, aerated seawater (salinity of 34 ± 1 ‰) at the Marine Molecular Genetics Laboratory of Jeju National University. Two weeks of acclimatization was carried out before any experimentation conducted. The fish were fed daily with commercial feed.

2.1.2. Disk abalone (*Haliotis discus discus*)

Healthy disk abalones were purchased from 'Youngsoo' commercial abalone farm in Jeju Island, Republic of Korea. They were maintained as 40 animals per tank in flat bottomed tanks (250 L) with sand-filtered aerated seawater at a salinity of 34 ‰ at 18 ±1˚C during the experimental period at Marine and Environmental Research Institute of Jeju National University. Abalones were acclimatized for 7 days before the experiment, and they were fed with fresh sea weed, *Undaria pinnatifida*, during the acclimatization period.

2.1.3. Manila clam (*Ruditapes philippinarum*)

Manila clams with average shell length of 35±5 mm were collected from the Eastern coastal region of the Jeju Island (Republic of Korea) and thereafter maintained in 250 L fiberglass tanks with controlled conditions same as in abalones at Marine and Environmental Research Institute of Jeju National University.

2.2. Identification of cystatin B sequences from three species

Rock bream cystatin B (RbCyt B) was originally identified by pyrosequencing of normalized cDNA using Roche GS-FLX system (DNA Link, Republic of Korea) as described in a previous report (Whang et al., 2011a). The RbCyt B cDNA sequence was identified by applying the basic local alignment search tool (BLAST; National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/BLAST). One of the ESTs was found to be homologous to known cystatin B. This sequence was cloned by amplification with gene-specific primers designed to the oligonucleotide sequences (RbCytB F2 and RbCytB R2; Table 1). Similarly, the full length cDNA sequence of abalone cystatin B (AbCyt B) was identified by analyzing the previously constructed abalone expressed sequence tag (EST) database sequences (Munasinghe *et al*., 2006). BLAST analysis indicated that one of the ESTs was homologous to previously identified cystatin Bs. This sequence was selected to design the BAC library screening primers and quantitative real-time polymerase chain reaction (qPCR) primers used in the study (Table 1). Finally the complete cDNA sequence encoding for Manila clam cystatin B (McCyt B) was identified using the same alignment tool at NCBI from our previously established Manila clam cDNA sequence database (Lee et al., 2011).

Table 1. Description of the primers used in the study

Name	Primer sequence (5'-3')	Orientation	Purpose
RbCytB F1	ATCTCTGCGCCACTTGATGCTGAT	Forward	qPCR and BAC screening
RbCytB R1	TCCCGGACACAATCTGTGTTGT	Reverse	qPCR and BAC screening
RbCytB F2	GAGAGAgaattcTCGATGATGTGTGGA GGAATCTCTGC	Forward	ORF amplification with EcoR I in pMAL c2X
RbCytB R2	GAGAGActgcagTTAGAAGTACGCAA TAGGGTCCTGGAG	Reverse	ORF amplification with Pst I in pMAL c2X
RbBeta F3	TCATCACCATCGGCAATGAGAGGT	Forward	β-actin for qPCR
RbBeta R3	TGATGCTGTTGTAGGTGGTCTCGT	Reverse	β-actin for qPCR
AbCytB F1	GTGGTGGTGCAACCGAAGTGAAAT	Forward	qPCR and BAC screening
AbCytB R1	TGCTACCACTTGTGAACGGAAGGA	Reverse	qPCR and BAC screening
AbCytB F2	GAGAGAgaattcATGTGTGGTGGTGCA ACCGAAG	Forward	ORF amplification with EcoR I site in pMAL c2X
AbCytB R2	GAGAGActgcagCTACTTGGCATCAA AATAATCCAGATCTGATGCA	Reverse	ORF amplification with Pst I in pMAL c2X
AbRib F3	TCACCAACAAGGACATCATTTGTC	Forward	Ribosomal protein for qPCR
AbRib R3	CAGGAGGAGTCCAGTGCAGTATG	Reverse	Ribosomal protein for qPCR
McCytB F1	GGAGGTGCAGGTGATGTTATG	Forward	McCyt B qPCR
McCytB R1	CCACTGAATCTTTTCCTGCTTTC	Reverse	McCyt B qPCR
McCytB F2	GAGAGAgaattcATGTGTGGAGGTGCA GGT	Forward	ORF amplification with EcoR I site in pMAL c2X
McCytB R2	GAGAGActgcagCTAGAAATATTCAA CAGCATCTTCTACTTTC	Reverse	ORF amplification with Pst I in pMAL c2X
McBeta F3	CTCCCTTGAGAAGAGCTACGA	Forward	β-actin for qPCR,
McBeta R3	GATACCAGCAGATTCCATACCC	Reverse	β-actin for qPCR,

2.3. Bacterial artificial chromosome (BAC) library construction and screening

Two separate BAC libraries of *O. fasciatus* and *H. discus discus* were constructed by Lucigen® Co. (Middleton, Wisconsin) using randomly sheared genomic DNA from rock bream whole blood cells and disk abalone gill tissues. Briefly, genomic DNA was randomly sheared and blunt end of large inserts (>100 kb) was ligated into pSMART® BAC vector to obtain an unbiased, full coverage library. Around 92,160 clones for each, possessing an average insert size of 120 Kb were arrayed in 240 of 384-well microtiter plates separately in each library. BAC clones inoculated into 384-deep well plates were grown individually. Aliquots of the grown cultures were pooled with other clones from the same plate, row or column pools for DNA preparation. Screening of the BAC-library was carried out with a PCR-based method (TaKaRa Bio, USA) following the manufacturer's instructions using gene-specific primers (Table 1). The identified clones were isolated from the corresponding

wells and confirmed by colony PCR with gene-specific primers. After confirmation, BAC DNA from positive clone was isolated and purified using QIAGEN Large-Construct Kit, following the manufacture's protocol and was subjected to sequencing by Roche (454) Genome Sequencer FLX (GS-FLX™) system (DNA Link, Inc., Korea).

2.4. In silico analysis of DNA and protein sequences

The amino acid sequences corresponding to RbCyt B, AbCyt B and McCyt B coding sequences were derived using DNAssist 2.2 (version 3.0). The orthologous sequences of cystatin B were compared by the BLAST search program at NCBI. Pairwise sequence alignment and multiple sequence alignment were performed by the EMBOSS Needle and ClustalW2 programs, respectively (http://www.ebi.ac.uk/Tools) (Thompson et al., 1994). Prediction of characteristic protein domains and conserved regions was carried out by the ExPASy-prosite database (http://prosite.expasy.org), MotifScan scanning algorithm (http://myhits.isb-sib.ch/cgi-bin/motif_scan) and Simple Modular Architecture Research Tool (http://smart.embl-heidelberg.de/). Some of the physicochemical properties of amino acid sequences were determined by the ExPASy ProtParam tool (http://web.expasy.org/protparam). Phylogenetic analysis was carried out by the Neighbor-Joining method with bootstrapping values taken from 1000 replicates using Molecular Evolutionary Genetics Analysis (MEGA) software, version 5 (Tamura et al., 2011).

The identified BAC clone sequences from rock bream and abalone BAC libraries were used to analyze the genomic structures and promoter regions. The potential transcription factor (TF) binding sites in the RbCyt B and AbCyt B promoter regions were predicted using the TFSEARCH version 1.3 (http://www.cbrc.jp/research/db/TFSEARCH.html) online database with threshold score at 90.0. The exon-intron structure was predicted by aligning the cDNA sequence with genomic sequence using the Spidey program

(http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey). Genome structures of cystatin B from other species were obtained from the Ensembl Genome Browser (http://www.ensembl.org/index.html) and visualized using Gene Maper version 2.5 (http://genemaper.googlepages.com) in order to compare the structures.

In order to reveal the tertiary structure of the three cystatin B molecules, computer simulation model was generated using I-TASSER (Roy et al., 2010; Zhang, 2008) online server and visualized the three dimensional (3D) structure using PyMOL Molecular Graphic System 1.3 and POV-Ray for windows v3.62 softwares.

2.5. Cloning and over-expression of recombinant cystatin B

To generate the recombinant proteins (rproteins) fused with maltose binding protein (MBP), the complete open reading frames (ORFs) were amplified from respective cDNAs by PCR with respective cloning primers (Table 1) containing EcoR I and Pst I restriction sites. The PCR amplification was carried out in a TaKaRa thermal cycler (TaKaRa Korea Biomedical Inc., Korea) using a 50 μL reaction mixture composed of 5 U of Ex*Taq* polymerase (TaKaRa), 5 μL of 10x Ex*Taq* buffer, 8 μL of 2.5 mM dNTPs, 80 ng of template, and 20 pmol of each primer. The thermal cycling conditions were: initial incubation at 94 ˚C for 3 min, followed by 35 cycles of 94 ˚C for 30 s, 55 ˚C for 30 s and 72 ˚C for 30 s, and a final extension at 72 ˚C for 2 min. The amplified products and pMAL-c2X vector were digested with the appropriate restriction enzymes, and confirmed by agarose gel electrophoresis. Appropriate sized fragments were excised and purified using the Accuprep Gel Purification Kit (Bioneer Co., Daejeon, Korea) and ligated into the linearized pMAL-c2X MBP-fused expression vectors by incubation with Mighty Mix (TaKaRa) for overnight at 4 ˚C. Cloned vectors were transformed into *Escherichia coli* DH5α cells. The size of the construct and sequences of the ORFs were confirmed by electrophoresis and sequencing,

respectively (Macrogen Inc., Korea). The recombinant fusion constructs were transformed in to *E. coli* BL21 (DE3) cells and grown on agar plates. A single colony was selected, inoculated in 5 mL of Luria-Bertani (LB) broth supplemented with 100 μg/mL of ampicillin, and grown overnight at 37 ˚C. Then, 1 mL of the inoculum was transferred into 500 mL of LB broth containing ampicillin, and grown at 37 ˚C in a shaking incubator (200 rpm). When the optical density (OD) reached 0.6 (at 600 nm), cultures were induced with isopropyl-β-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. After incubation at 20 ˚C for 8h, the cells were harvested by centrifugation and the pellet was re-suspended in column buffer (20 mM Tris-HCL, pH 7.4 and 200 mM NaCl). Recombinant cystatin Bs fused with MBP (rRbCyt B, rAbCyt B and rMcCyt B) were purified using the pMAL protein fusion and purification system (Maina et al., 1988). Samples collected at different steps in expression and purification process were analyzed on a 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under reduced conditions, with low molecular weight protein marker (Enzynomics, Korea). The resulted gel was stained with 0.05% Coomassie blue R-250, followed by a standard destaining procedure. The purified rprotein concentration was measured according to the Bradford method (Bradford, 1976) using bovine serum albumin (BSA) as the standard.

2.6. Cysteine protease inhibition assay

To analyze the activity of purified recombinant proteins (rRbCyt B, rAbCyt B and rMcCyt B), *in vitro* papain inhibitory assay was performed as reported previously, with some modifications (Xiao et al., 2010). Briefly, papain (P3375; Sigma-Aldrich, USA) was dissolved in 0.1 M potassium phosphate buffer (PPB), pH 7.4, to obtain 1 mg/mL solution. Subsequently, different volumes of the rprotein and papain were mixed together to generate the following rProtein/papain concentration ratios in 125 μL of final volumes: 0/1 (control),

1/32, 1/16, 1/8, 1/4, 1/2, 1/1, 1.5/1. The mixtures were then incubated at 25 ˚C for 10 min, after which 125 µL of azo-casein (0.5% in PPB) was added to the mixture and was allowed to react by incubating at 37 ˚C for 30 min. The reaction was terminated by adding an equal volume of 10% trichloroacetic acid. The same procedure was followed simultaneously for the MBP as the control and both positive (without rprotein) and negative (without papain) controls were implemented. Finally, the mixture was centrifuged and the supernatant was subjected to measure the OD_{440}. The relative activity was calculated as: 100 × [1-(OD_{440} of test sample /OD_{440} of control)] (Xiao et al., 2010). All the samples were tested in three replicates at each point.

2.7. Isolation of tissues from experimental animals

Healthy animals described in section 2.1 were selected to examine the tissue specific mRNA expression profile of cystatin B gene. Animals were carefully dissected on ice and tissues (blood, gills, liver, spleen, brain, intestine, kidney, head kidney, and muscle from rock bream; gills, mantle, muscle, hemolymph, hepatopancreas, digestive tract and testis from abalone and hemolymph, adductor muscle, mantel, siphon, gills and foot from clam) were isolated carefully from three replicate animals. The blood from rock breams was collected using a heparin contained syringe fitted with a 22-gauge needle from the caudal vein. Collected blood was immediately transformed into micro tubes and centrifuged at 3000× *g* for 10 min at 4 ˚C to collect the blood cells. The abalone hemolymph was withdrawn from the cephalic arterial sinus, accessed anteriorly at the angle between the foot and the head using a syringe fitted with a 22-gauge needle. Collected hemolymph was immediately transferred into micro tubes and was centrifuged as previous. The supernatant was removed and hemocytes were collected. Hemolymph (1-2 mL/clam) from clam was collected from the posterior adductor muscle using a 26 G syringe as same to the abalone to collect the

hemocytes. All the tissue samples were snap-frozen in liquid nitrogen, and stored at -80 ˚C until use for RNA extraction.

2.8. Immune challenge experiment for temporal mRNA expression analysis

Rock breams were immune-challenged with live *Edwardsiella tarda* bacterium. *O. fasciatus* were intra-peritoneally (i.p) injected either with 100 μL live *E. tarda* (5×10^3 CFU/μL) in phosphate buffered saline (PBS) or equal volume of PBS per fish as described previously (Whang et al., 2011b). Tissue samples from head kidney and spleen were collected at 0h (un-injected control), 3h, 6h, 12h, 24h and 48h post-infection (p.i.). Three replicates of fish were obtained from each group to isolate tissues at each time point and, pooled tissue samples were subjected to total RNA extraction. Tissue collection was conducted as described in section 2.7.

To determine the immune responses of the AbCyt B, immune stimulation/challenge experiment was devised and conducted using pathogenic Gram-negative bacterial species, *Vibrio parahaemolyticus*. Animals were injected with 100 μL of 1×10^4 CFU/mL of bacteria in saline (0.9% NaCl) and the same volume of saline injected group was used as a control. Gill and hemolymph tissue samples were collected at different time points (0, 3, 6, 12, 24 and 48 h) after the infection as described in section 2.7 and tissues were snap-frozen and stored at -80 ˚C. Four replicates were used to collect the tissues at each time point.

Clams were challenged with *Vibrio tapetis* (a Gram negative bacterial pathogen), LPS (an endotoxin of Gram negative bacterial cell walls), or poly I:C. As described previously, 100 μL of *V. tapetis* (1.9×10^6 cells/mL in 0.9% saline), LPS (100 μg/clam in 0.9% saline), or poly I:C (100 μg/clam in 0.9% saline) was injected directly into adductor muscle (Umasuthan et al., 2012). A group of un-injected clams served as negative controls. Gills and hemocytes

tissues were collected at 0 (un-injected), 3, 6, 12, 24 and 48 h of p.i. from control and challenged groups, including five clams from each group. Total RNA was extracted from pooled tissue samples using at least five clams and used for cDNA synthesis.

2.9. Total RNA extraction and cDNA synthesis

Total RNA was isolated from pooled tissues at each time point from three different species separately, using the TRIzol reagent (Sigma-Aldrich) according to the manufacturer's instructions. Concentration of purified RNA was determined by measuring at A_{260} on a spectrophotometer (BioRad, USA). Purified RNA samples were diluted to 1 μg/μL and cDNAs were synthesized from 2.5 μg of RNA by using the PrimeScript™ First Strand cDNA Synthesis kit (TaKaRa Bio Inc., Japan), following the manufacturer's protocol. The resultant cDNA was then diluted 40-fold (total 800 μL) before storage at -20 ˚C.

2.10. Quantitative real-time PCR (qPCR) analysis

Synthesized cDNA was used to quantify the mRNA expression level in different tissues at different time points in both the immune-challenged and control groups by the qPCR technique using gene-specific primers (Table 1). Each reaction was carried out in a 15 μL of reaction mixture containing 4 μL of 1:40 diluted template cDNA, 7.5 μL of 2× SYBR Green Mix, 0.6 μL of each primer (10 pmol/μL), and 2.3 μL of PCR-grade water, and by using the Thermal Cycler Dice Real Time System (Model TP800; TaKaRa Bio Inc.). The real time (RT)-PCR cycle program consisted of one cycle at 95 ˚C for 10 s, followed by 45 cycles of 95 ˚C for 5 s, 58 ˚C for 20 s, and 72 ˚C for 20 s, and finally, 1 cycle of 95 ˚C for 15 s, 60 ˚C for 30 s and 95 ˚C for 15 s. The baseline was automatically set to maintain consistency. The mRNA expression was determined by the $2^{-\Delta\Delta CT}$ (Livak) method (Livak and Schmittgen,

2001) and was tested in triplicates. Beta-actin gene expression was detected as the internal control (Accession No. FJ975145) for the rock bream and all the RbCyt B expression values were normalized to the beta-actin values. The tissue specific expression values were compared with that of muscle and immune challenged values were normalized to the PBS control at each time point and compared with non-injected (0h) control. For abalones, the relative mRNA expression was calculated by first normalizing to expression values of the gene codes for abalone ribosomal protein L5 (Wan et al., 2011) as the internal control (Accession No. EF103443) and then was normalizing to the expression values of the saline injected control group at each time point and compared with un-injected (0h) control . Tissue specific values were compared with expression in muscle. In Manila clam, beta-actin gene expression was used as the internal control (Zhang et al., 2011). The expression level of adductor muscle was considered as the basal level, by which expression in all other tissues was compared. To determine the relative expressions, the observed expression after challenge was first normalized to the beta-actin and saline-injected controls and then compared with expression in the unchallenged (0h) group.

2.11. Statistical analysis

To determine the statistical significance between the experimental and control groups, all the mRNA expression analysis data were subjected to either Student's t-test or one-way analysis of variance (ANOVA) in SPSS 16.0 for Windows (SPSS, Chicago, IL, USA). Differences were considered statistically significant at $P < 0.05$ and all the data were represented as means ± standard deviation.

3. RESULTS

3.1. Identification and characterization of cystatin B sequences

The cDNA sequence of RbCyt B was identified in rock bream multi-tissue normalized cDNA GS-FLX database by BLAST analysis, and this sequence was used to isolate the respective colony from a BAC library. The complete genomic sequence was determined and deposited in the GenBank database under accession number JQ287496. The full-length cDNA was found to contain a 5'-untranslated region (UTR) of 87 bp and a 3'-UTR of 224 bp with a polyadenylation signal sequence (AATAAA) that was found 11 bp upstream of the poly (A) tail (Fig. 4A). Its ORF was found to be 300 bp in length, encoding a polypeptide of 100 amino acids. Analysis of the amino acid sequence revealed the presence of a CY domain and a cysteine protease inhibitor signature, which suggests that the protein may belong to the cystatin superfamily. Analysis with the SignalP program indicated that there was no signal peptide, and the ExPASy Prosite database indicated that no carbohydrate side chains or disulfide bonds associated with the protein. The estimated molecular mass of RbCyt B is 11 kDa and its isoelectric point is 6.6. Collectively, these data suggest that the newly-identified RbCyt B is a member of the family 1 cystatin.

```
  1 CCTGACTCAACCTGAACGCCTCACTGAAACAGACTCACTCGGTGTCTTTCGTCTTTTCGC
 61 TCATTTTAAACCCTAAAAACTGTCAAAATGTCGATGATGTGTGGAGGAATCTCTGCGCCA
  1                            M  S  M  M  C  G  G  I  S  A  P
121 CTTGATGCTGATGAAGACATCCAGAAAATGTGTGATAACGTAAAACCTCATGCAGAGGAG
 12  L  D  A  D  E  D  I  Q  K  M  C  D  N  V  K  P  H  A  E  E
181 AAAGCAGGGAAGAAATATGATGTTTTCACAGCCAAGACATACACAACACAGATTGTGTCC
 32  K  A  G  K  K  Y  D  V  F  T  A  K  T  Y  T  T  Q  I  V  S
241 GGGACCAACTACTTCATAAAGATCCATGTGGGAGGAGACGATCATGTTCACCTTCGTGTT
 52  G  T  N  Y  F  I  K  I  H  V  G  G  D  D  H  V  H  L  R  V
301 TACAAAAAACTCCCCTGTCATGGAGGAGGCCTTGAGCTGAGTGGCATGCAGCACTCCAAG
 72  Y  K  K  L  P  C  H  G  G  G  L  E  L  S  G  M  Q  H  S  K
361 AGCCTCCAGGACCCTATTGCGTACTTCTAATGGGATCCCAAACCAAACAAAGCACCAGTG
 92  S  L  Q  D  P  I  A  Y  F  *
421 CCTACAGGCTGCCATCTCATTATAACTACTTATTATCAGATTATCCTGACTCATTACATA
481 AAACAATATTCTCATTCCTAACATGTCATAACATACTGTATACGTACTGTAGATCATGAT
541 GTTAGTATGTCAAAATGGAACTATTTTGCACAAGATGAATTAATAAAGATCAAAAAAAAA
601 AAAAAAAAAAAAAA
```

Fig. 4A. Nucleotide and deduced amino acid sequences of RbCyt B. The numbers of nucleotides and deduced amino acid residues are shown in the left margin. The start codon (ATG) and the stop codon (TAA) are underlined. The cystatin-like domain is shaded, and the cysteine protease inhibitor signature is indicated in boldface letters. The ^{48}QIVSG^{52} motif is boxed in the middle of domain and conserved N-terminal glycine residue (G^{6}) is marked in bolted italics, and the P^{76}C^{77} motif is indicated in underlined boldface. The poly (A) signal is boxed, and the poly (A) tail is underlined at the end of the nucleotide sequence.

The full length AbCyt B cDNA sequence of 1967 bp contains a 303 bp open reading frame (ORF) encodes for 101 amino acids, a 115 bp 5'-UTR and a 1546 bp 3'-UTR with a poly-adenylation signal (AATAAA). AbCyt B is comprised of a single CY domain with a conserved glycine residue near the N-terminal, a glutamine-valine-valine-alanine-glycine (QVVAG) motif at the middle of domain and a variant of proline-tryptophan (PW) motif with tyrosine (PY) substitution at C-terminal region. However, no signal peptide or disulfide bonds were identified. The predicted molecular mass of the derived protein was shown 11 kDa and the isoelectric point was predicted as 5.49. The complete cDNA and deduced amino acid sequences of AbCyt B are shown in Fig. 4B. The complete genomic sequence was deposited in GenBank under accession number JQ653304.

```
   1  GAAGCAAACAAGTCA CATGAGGCGTTCCCC GCTGATTACGGAAAA CACACAGATATTTAG
  61  GAGCTGCGGTTGTTG CTGTTGTTTGCTGTG AAAGTTTGATTAAAC GAAGATGTCG  ATG
   1                                                              M
 119  TGTGGTGGTGCAACC GAAGTGAAATCTGCA ACAGAAGAAGTGCAA AAACTCTGTAATGAG
   2   C  G  G  A  T   E  V  K  S  A   T  E  E  V  Q   K  L  C  N  E
 179  GTTCGAGAGGCCTTG CAAACACAAGCAGGG AGAACGTTTGGAGCG TACAAAGCCATATCC
  22   V  R  E  A  L   Q  T  Q  A  G   R  T  F  G  A   Y  K  A  I  S
 239  TTCCGTTCACAAGTG GTAGCAGGAACCAAC TACTTTGTTAAGGTC CAGGTGGATGAAAAC
  42   F  R  S  Q  V   V  A  G  T  N   Y  F  V  K  V   Q  V  D  E  N
 299  GATGAACACTTTCAC CTGAGGATATTCGCC CCCCTCCCCTACACC AACTCCCCACCCTCC
  62   D  E  H  F  H   L  R  I  F  A   P  L  P  Y  T   N  S  P  P  S
 359  CTTGCTGGCTATCAG ACTGGACACACCGCT GCATCAGATCTGGAT TATTTTGATGCCAAG
  82   L  A  G  Y  Q   T  G  H  T  A   A  S  D  L  D   Y  F  D  A  K
 419  TAGTTGCCGTCCAAG TTTACACCATGTGCT GGACAAGCTCTTGAT AGCTTATGTTCATGG
       *
 479  GATCTTTACATAATC TCTCAGTATAACATA TCAGTAGCTAAAGCT CACGCTTCATGTGTT
 539  ATCTCCCTTGAATAT CTTCAGGAAATGATC TGTTCATAAATCATT TTGAAATAAAAAAAC
 599  AAAACAGTTATCATG ATTAATCTGGTACAT AAAAACAACATCCAT TTAGTGTATTTTAAA
 659  CAAATAAATGGATAT TTTTTACACTTTTTG TCTCGGTGGTGAGCG TCCCCTGCAGCCTGT
 719  ATCAGTTGTATTGTG TACACTCAAGAAGAA CTGCCATTAACACCA TGGAACAACATCATG
 779  GTGGTGAGATGCTTG TAAACACCAAAACAC AAGATTATGATGCAA GTATGCTCATTTCCT
 839  TCCTCAGATCAGCAT CACGTCCGCCGCACA AGGATATGTCAAATT TGTGTTTGTATTGTT
 899  TTGCTTCCCTTTGGG ACTACAGAATGTAAA AAAATGAATTTGGTT GACTATTGTCCCTAC
 959  TACAACAGGTTGTTT TCAGAAATAGTATAT TTGCTGCTGTTTAGC GTTTCATAGTTCTCA
1019  TTTTGTAACCTACCA TTTTCTATTATACTT GTTTTGTCTATCAGT ATGTAGGCTGCTGCA
1079  GGAATGCCCATATCA TGTTATGTCACTCTG TTGTGACAATGCTTC CTAGAGAGCTGCCTG
1139  TATGATCTGCTCAGA GGGGTTACAGGGATA AATATATACAGCCAG TATCTTGCTTATGTG
1199  TATTCCCCGTGCTTG TGGGTGTCACCAGTG GTACCAGAAAACGTC CGAACCTGCAACACT
1259  TTGACTTTTCCGTCC ATATTAAGCTGACAC ACCATGATATAACTG GAATACTGTTCAAAG
1319  TGTTGTTAAACCCAA CTCGCTCACTCATAT CTATACAATATAAAA ATGGTTAAAGTTGCT
1379  GAAATTACAAATTGA TGTGTGCTGTTTGAA AAGCTGAAAACTTTC TTACATTTTCATGTG
1439  AAAATGCTATAAGGC TGCATTTGCAGATTC TGACATTCTTGATTG TCCATGAAATTCTTG
1499  AATGTTCATGTAAAA TGTCAGATTTCTTTG GAAACGAGGACCAAA AGTTACAGTGCAGGT
1559  GAATTTTGATAATTT ATTATTGTCCATGAT CTCCTGATCCTTTGC AAACGAACATTATAG
1619  CAGTTTCTGCACAGT AGATACATATTTAAT ATGAAAGTTACTGAT TCAGAGGAATCTGCT
1679  CTTGGGAGTGTTTGT AGTCGGTTCCACCTG ACAGACTTATCTTAG CTGCCGTCTCCGGTT
1739  GTTTCATCTCACTGC CACTGAACCCTTCAT TTCACGTGTTGGGAT CTGTGGACTTAATTA
1799  ATCAGCAGTTCTGCA TGAAAATGCAAGACA GGAATTGTGAAACGA CAAGACCTTTTATCA
1859  GTGTCTATTTGTTGA GCCATAGATGAAATT AAGGTAACAGTTGTG ATGTTGACTTTTTTA
1919  TCTGCTGTTGAAATT TTCACATAAACAAC A ATAAA CCATTCTTTT TACA
```

Fig. 4B. Nucleotide and deduced amino acid sequences of AbCyt B. The start (ATG) and stop (TAG) codons are underlined. The cystatin-like domain is shaded and cysteine protease inhibitor signature sequence is in boldface. The conserved G^{3}, $^{45}QVVAG^{49}$ motif and $P^{74}Y^{75}$ motif are showed in bold italic, boxed and underlined boldface, respectively. The poly (A) signal (AATAAA) is in italic boldface at the end of nucleotide sequence.

The complete cDNA sequence of McCyt B was determined to consist of 509 nucleotides, comprising a 297 bp open reading frame (ORF) that encodes 99 amino acids, an 85 bp 5′-UTR, and a 127 bp 3′-UTR. The 3′-UTR contains a polyadenylation signal

(^{404}AGTAAA^{409}) and a single RNA instability motif (^{368}ATTTA^{372}) (Fig. 4C). The predicted molecular mass was 11 kDa and the theoretical isoelectric point was 5.9. The complete cDNA sequence and amino acid sequence was deposited in the NCBI GenBank database under the accession number JQ972711.

```
-85    GGTGAAAGTGTAAAGGTTTCGTCTTGCTTTGAACAGTTTGTGATAATTTCCGTTTCTA
-27  CTGTAGATTCATCTTAATATGGCTCAAATGTGTGGAGGTGCAGGTGATGTTATGCCAGCT
 1                                M  C  G  G  A  G  D  V  M  P  A
 34  GATGAGGAAGTGAAAGGTTATTGCAATGAGGTGAAAGCTGACATACTGAAGAAAGCAGGA
 12   D  E  E  V  K  G  Y  C  N  E  V  K  A  D  I  L  K  K  A  G
 94  AAAGATTCAGTGGAAATATTTGAACCTGTCCACTACAGAAGCCAGCTTGTAGCTGGAGTC
 32   K  D  S  V  E  I  F  E  P  V  H  Y  R  S  Q  L  V  A  G  V
154  AATTACTTTGTAAAGATCAGAATTGGCTCTGGTGGAGAATGTTTACATGCTAGAATTTTC
 52   N  Y  F  V  K  I  R  I  G  S  G  G  E  C  L  H  A  R  I  F
214  AAGGGTCTTCCTCATACAGGAGGGAACTTAGAAGTGAGCAGCGTACAAACAAATAAGAAA
 72   K  G  L  P  H  T  G  G  N  L  E  V  S  S  V  Q  T  N  K  K
274  GTAGAAGATGCTGTTGAATATTTCTAGTTGTTGAAAGGCAGAGTTAGCAGGGACACTTGA
 92   V  E  D  A  V  E  Y  F  *
334  TTGCACAACAGTGGTCAATAATTAGCCAAGCTTGATTTAAGGGCTTTATTTTACCTTTAA
394  ATATTATAACAGTAAATACTCAACTTAAAAA
```

Fig. 4C. Nucleotide and deduced amino acid sequences of McCyt B. The start (ATG) and stop (TAG) codons are underlined. The cystatin-like domain is shaded and Cystatin protease inhibitor signature sequence is in boldface. The conserved G^{3}, ^{46}QVVAG^{50} motif and $P^{75}Y^{76}$ motif are showed in italic, boxed and underlined boldface respectively. The poly (A) signal (AGTAAA) is in italic boldface at the end of nucleotide sequence.

3.2. In silico analysis and comparison of cystatin B

3.2.1. Promoter region analysis

The RbCyt B genomic DNA sequence is ~3.85 Kb in length including a promoter region (Fig. 6). The putative promoter region contains a TATA box at -25 bp upstream from the transcription start site (TSS). Additionally, several TF binding sites were identified throughout the promoter region, including those for myeloid zinc finger protein-1 (Mzf 1), sex-determining region Y (SRY), specificity protein 1 (SP-1), CdxA, activating protein 1 (AP-1), Nkx2, AML-1A, GATA-1, and SOX 5 (Fig. 5A). The AbCyt B genomic sequence is

approximately 8.4 Kb in length including promoter region. The putative promoter region and TF binding sites of the AbCyt B were shown in Fig. 5B. Two enhancer box (E-box) TF binding sites and an initiator (Inr) motif sequence (CCATTTT) were identified -36 bp and -253 bp upstream to the transcription initiation site, respectively. Apart from Inr motif, several TF binding sites including CdxA, AML-1A, heat shock factor (HSF), NIT 2, Sox 5, erythroid TF (GATA-1), CAAT box, ecotropic viral integration site-1 (Evi-1), and TBP related factor (TRF 1) were identified in the anticipated region.

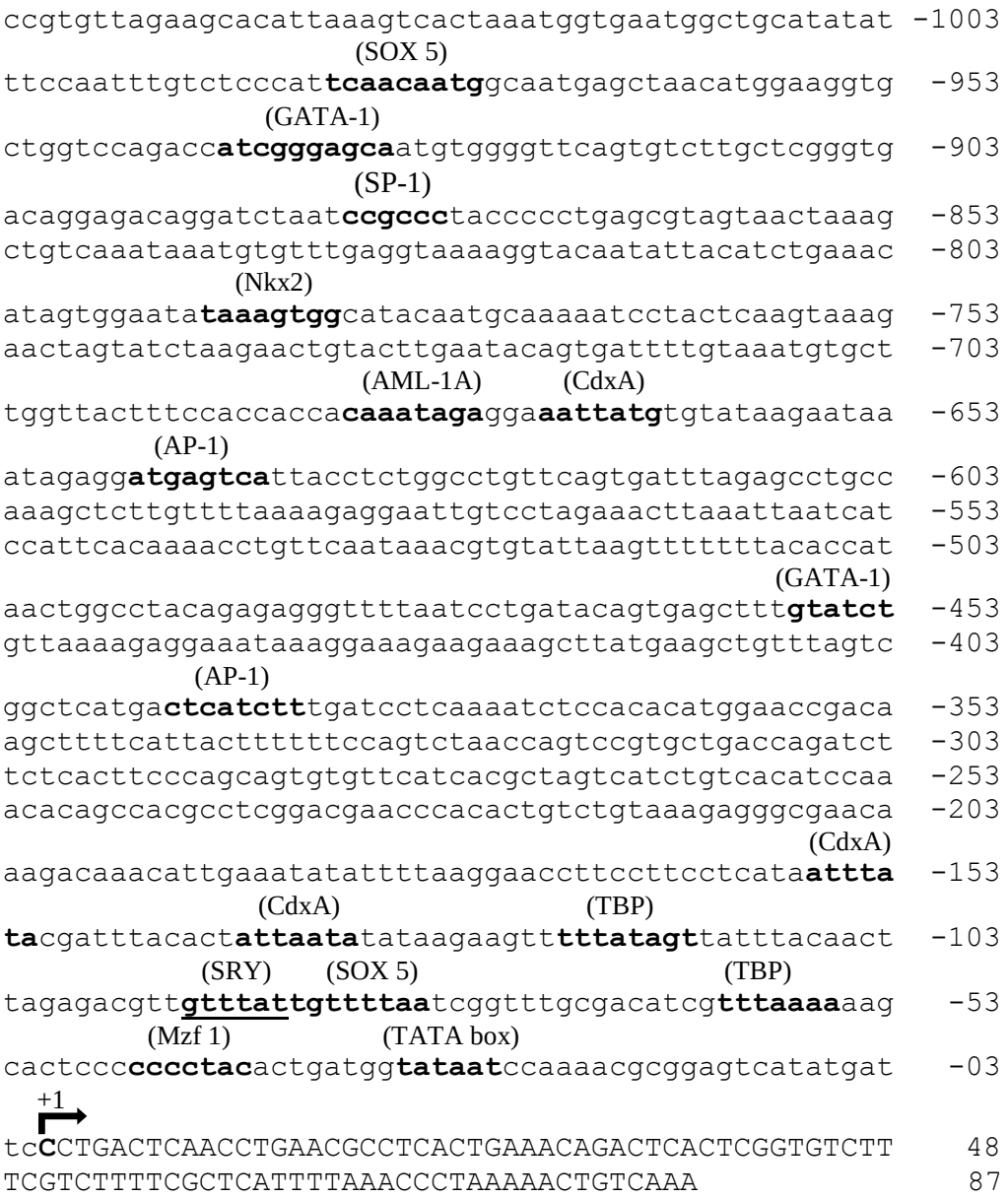

Fig. 5A. Putative promoter region and 5'-UTR region of the RbCyt B gene. The regions are encompassed by sequences -1052 to +87, relative to the transcription start site (+1; bent arrow). Putative transcription factors binding sites are in boldface. The 5'-UTR region is shown in block letters.

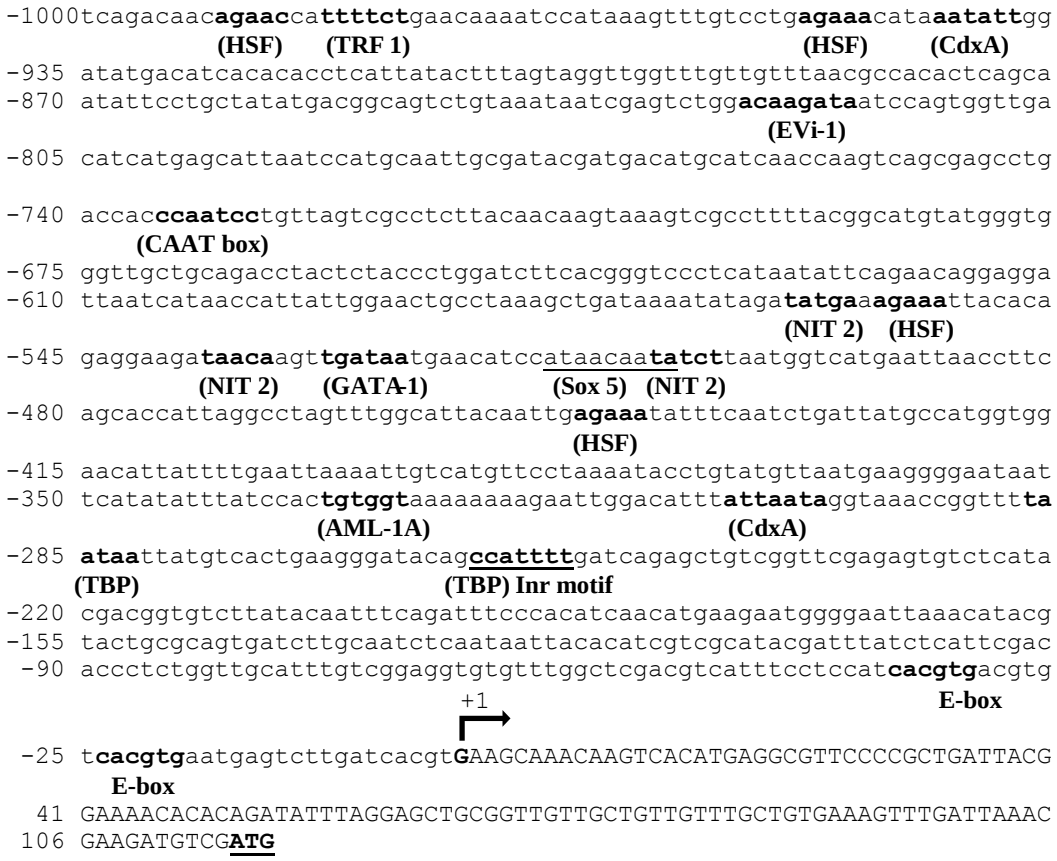

Fig. 5B. Putative promoter region and 5'-UTR region sequence of AbCyt B from -1000 to +118 bp form transcription initiation site (+1; bent arrow) with TF binding sequences predicted by TFSEARCH online database. The 5'-UTR region is denoted by block letters and translation start cordon is underlined.

3.2.2. Genomic structure comparison

According to the complete cDNA sequence alignment with the respective genomic sequence by the Spidey program, three exons interrupted by two introns were identified in both RbCyt B and AbCyt B genes and they were compared with those of human, cattle, rat, chicken and zebrafish cystatin B genes (Fig. 6). The intron-exon junction structures were typical, with donor and acceptor (GT and AG) dinucleotide sequences indicating the presence of possible splice sites in both RbCyt B and AbCyt B. The second exon showed identical length to other known cystatin B sequences, while the first and third exons extensively varied in length. In the third exon, the coding region exhibited identical length in all compared sequences except in AbCyt B, whereas the 3'-UTRs showed remarkable

variations in length. The second intron was much longer in both experimental species than that in any of the other cystatin B genes.

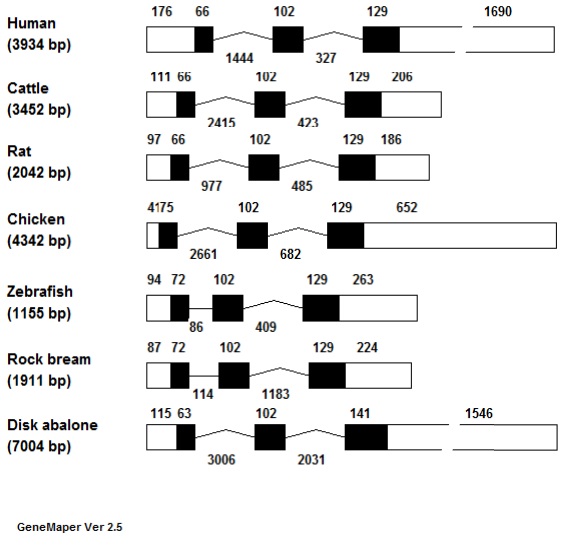

Fig. 6. Schematic representation and comparison of rock bream and disk abalone cystatin B genomic structures with that of other species. UTRs, coding regions and introns represent by white boxes, black boxes and black lines respectively.

3.2.3. Pairwise and Multiple sequence alignment analysis

Pairwise sequence alignment with known cystatin B orthologous revealed that RbCyt B shared up to 72.5% amino acid identity and 88.2% amino acid similarity with other known cystatin B sequences. However, AbCyt B and McCyt B indicated relatively low percentage identity and similarity (Table 2A) with its orthologs. AbCyt B protein was 32.7 – 44.7% identical to the cystatin B proteins from mammals, chicken, fish, mollusks, and other invertebrates. In contrast, AbCyt B exhibited relatively lower level of amino acid identity to the other two experimental species with lowest identity to the RbCyt B (Table 2B). McCyt B exhibited the highest identity and similarity with cystatin B from Pacific oyster (*Crassostrea gigas*), and the lowest identity and similarity were with that from clam worm (*Perinereis aibuhitensis*) (Table 2C).

In multiple homologue alignment analysis identified three conserved regions in cystatin B genes, including an N-terminal glycine, a glutamine-valine-glycine (QxVxG) motif, and a variant of proline-tryptophan (PW) motif at C-terminal region, all of which are important for the biological activity of the molecule.

Table 2A. Amino acid identity and similarity of RbCyt B to other known cystatin B proteins

Organism	Species	Accession No.	Identity (%)	Similarity (%)
Human	*Homo sapiens*	NP000091	46.1	60.8
Norway rat	*Rattus norvegicus*	NP036970	45.1	59.8
Chicken	*Gallus gallus*	NP001185577	50.0	64.7
European seabass	*Dicentrarchus labrax*	CBN81974	72.5	88.2
Japanese flounder	*Paralichthys olivaceus*	ACC86114	68.6	78.4
Disk abalone	*Haliotis discus discus*	JQ653304	32.7	55.8
Pacific oyster	*Crassostrea gigas*	ADI33157	45.1	63.7
Manila clam	*Ruditapes philippinarum*	JQ972711	40.2	62.7
Duck leech	*Theromyzon tessulatum*	AAN28679	41.0	60.0

Table 2B. Amino acid identity and similarity of AbCyt B to other known cystatin B proteins

Organism	Species	Accession No.	Identity (%)	Similarity (%)
Human	*Homo sapience*	NP000091	44.7	61.2
Norway rat	*Rattus norvegicus*	NP036970	44.1	62.7
Chicken	*Gallus gallus*	NP001185577	44.1	65.7
Zebrafish	*Danio rerio*	NP001096599	40.4	59.6
Rock Bream	*Oplegnathus fasciatus*	JQ287496	32.7	55.8
Pacific Oyster	*Crassostrea gigas*	ADI33157	44.1	65.7
Manila clam	*Ruditapes philippinarum*	JQ972711	38.2	57.8
Clamworm	*Perinereis cultrifera*	AAN28679	44.1	53.9
Duck leech	*Theromyzon tessulatum*	AAN28679	42.7	65.0

Table 2C. Amino acid identity and similarity of McCyt B to other known cystatin B proteins

Organism	Species	Accession No.	Identity (%)	Similarity (%)
Human	*Homo sapiens*	NP000091	36.0	58.0
Chicken	*Gallus gallus*	NP001185577	37.0	59.0
Norway rat	*Rattus norvegicus*	NP036970	37.0	62.0
Sable fish	*Anoplopoma fimbria*	ACQ58093	44.7	62.1
Rock Bream	*Oplegnathus fasciatus*	JQ287496	40.2	62.7
Disk abalone	*Haliotis discus discus*	JQ653304	38.2	57.8
Pacific oyster	*Crassostrea gigas*	ADI33157	47.0	63.0
Duck leech	*Theromyzon tessulatum*	AAN28679	37.5	58.0
Clamworm	*Perinereis aibuhitensis*	ACL12062	34.0	54.4

In RbCyt B, the QxVxG motif was identified as a glutamine-isoleucine-valine-serine-glycine (QIVSG), while in AbCyt B and McCyt B it was QVVAG and QLVAG sequences, respectively (Fig. 7). In this study, we observed that there were two conserved glycine residues adjacent to each other in the N-terminal region of all cystatin B sequences except human and Norway rat cystatin B sequences and a highly conserved cysteine residue at the N-terminal of all compared sequences. However the PW motif was not identified as highly conserved motif, which showed the tryptophan (W) substitution with either Y in abalone, C in rock bream and European seabass or H in all other species compared in the study.

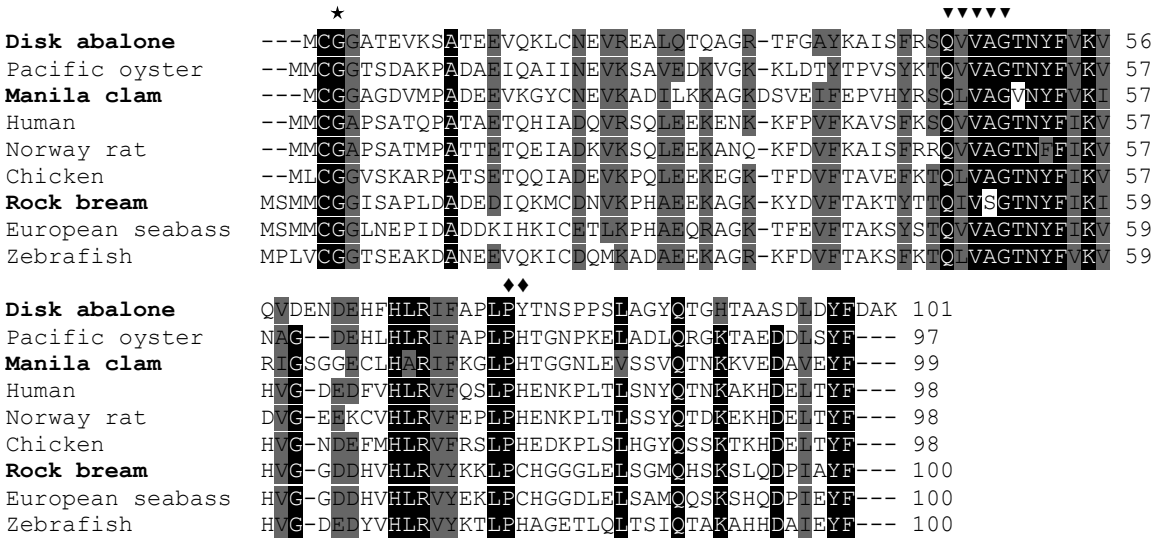

Fig. 7. Alignment of the RbCyt B, AbCyt B and McCyt B amino acid sequences with selected known cystatin B sequences. The conserved G, QxVxG motif and variant of PW motif are marked with (★), (▼) and (♦) respectively. Residues that are identical among the sequences are given in a black background, and those that are similar among the sequences are given in a gray background.

3.2.4. Computer based molecular modeling

Tertiary structures of RbCyt B, AbCyt B and McCyt B were determined using the computer based simulation modeling strategy of I-TASSER online server. The generated models exhibited 89 ±7% accuracy based on TM-score value and top ten threading templates from the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank used by

the server belong to the stefin family proteinases. Moreover, the normalized Z-score values of the threading alignments exceeded 1, substantiating the higher degree of confidence of the predicted structures. The main body of the molecules comprised of five stranded anti parallel β sheet wrapped around a central five turned alpha helix but with an extension of carboxy terminus running along the back of the sheet. Furthermore, two hairpin loops, one of which contains a highly conserved amino acid sequence (QxVxG), can be prominently identified in the structures along with a conserved amino acid residue at the N terminal and variant of PW motif at second hairpin loop in all three molecules in similar manner (Fig. 8). Altogether these features are important in forming a hydrophobic wedged shaped edge in the active site of the molecule.

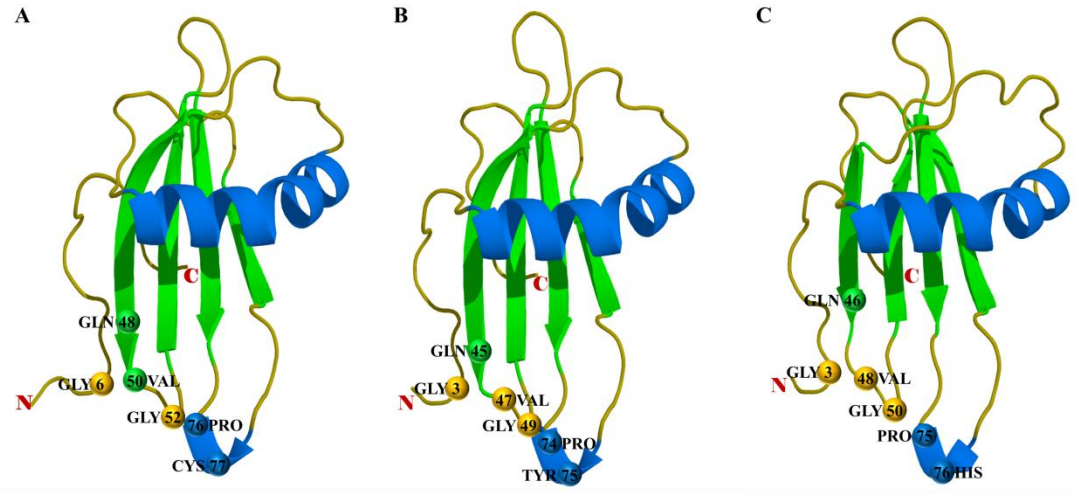

Fig. 8. Predicted three-dimensional structure of A) RbCyt B B) AbCyt B and C) McCyt B based on human stefin B. The molecule consists mainly of a straight five turn α-helix (blue) and antiparallel β-pleated sheets (green). The C-terminal (C), N-terminal (N), conserved G, QxVxG motif, and PW motifs are indicated.

3.3. Phylogenetic analysis of cystatin B

The evolutionary tree generated by phylogenetic analysis considering all four cystatin family sequences from mammals, avian, teleost, mollusks, annelids and nematode species. Different types of cystatins were clearly clustered into 4 different phylogenetically related

families, where our interested members were clustered within type 1 cystatins (Fig. 9). RbCyt B was clustered in the clade with teleostean homologues for cystatin B with the closest evolutionary proximity to the cystatin B from European seabass (*Dicentrarchus labrax*), while AbCyt B exhibited closest affinity to McCyt B, which was placed in the sub-cluster within cystatin B from mollusks.

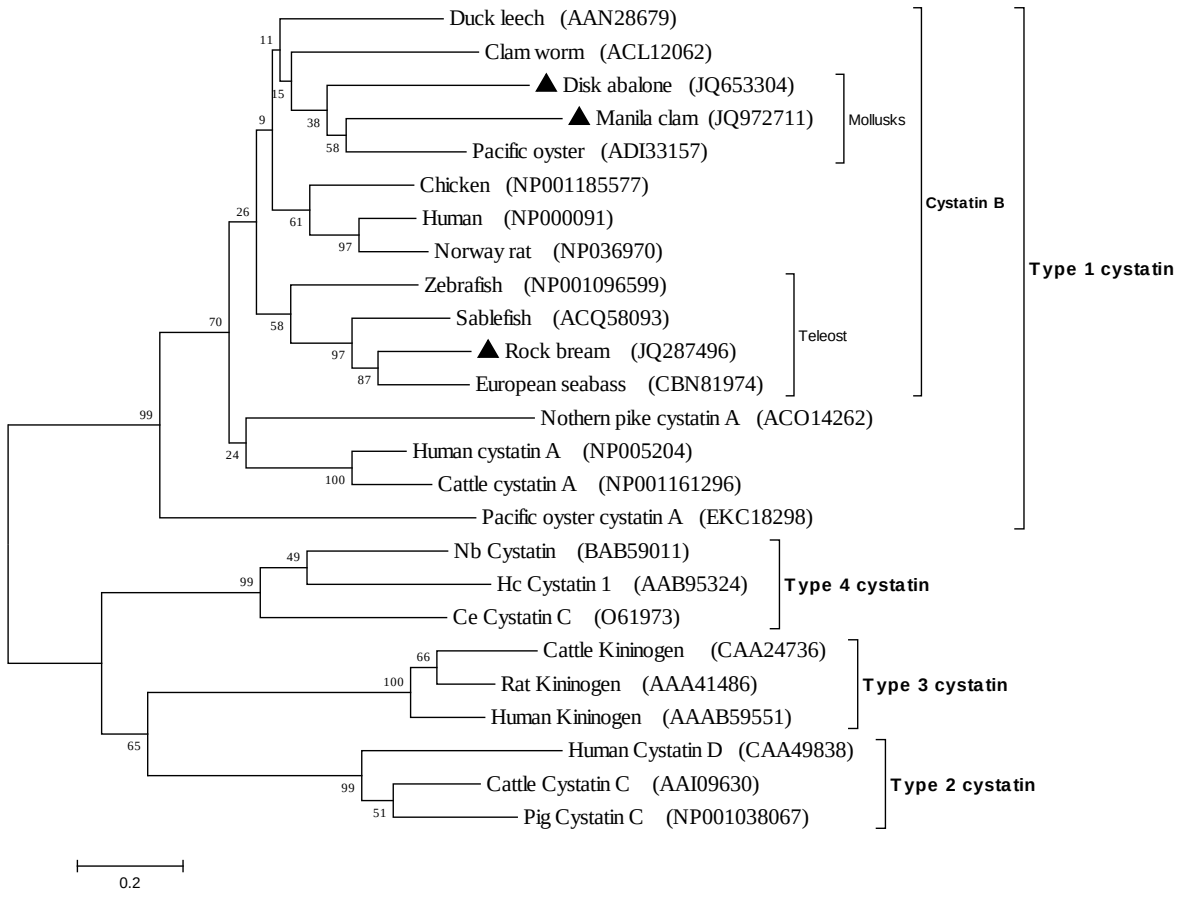

Fig. 9. Phylogenetic tree of cystatin superfamily members from different species. The tree was constructed using the Neighbor-Joining method based on 1000 bootstrap replications. Nb, *Nippostrongylus brasiliensis*; Hc, *Haemonchus contortus*; Ce, *Caenorhabditis elegans*.

3.4. Over-expression and recombinant protein purification

In order to characterize the biochemical properties of recombinant cystatin Bs, the coding sequences (ORF) were cloned into the pMAL-c2X expression vector and expressed as a fusion protein with MBP. The purified proteins were analyzed on a 12% SDS-PAGE gel

system and the SDS-PAGE resolved a single purified protein band for each, confirming successful purification of the fusion protein with an apparent molecular mass of approximately 53.5 KDa, which was compatible with the predicted molecular mass (MBP 42.5 + cystatin B 11 KDa) (Fig. 10).

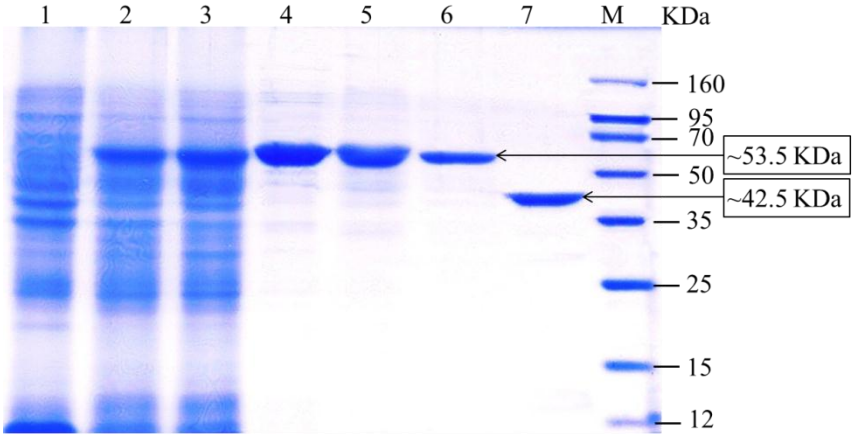

Fig. 10. SDS-PAGE analysis of over-expressed and purified recombinant fusion proteins from three species. Lanes: 1, total cellular extract from *E. coli* BL21 (DE3) prior to IPTG induction; 2, crude extract of cell lysate after IPTG induction; 3, supernatant of IPTG-induced cell lysate; 4, purified recombinant fusion protein of RbCyt B; 5, purified recombinant fusion protein of AbCyt B; 6, purified recombinant fusion protein of McCyt B; 7, purified recombinant MBP; M, low molecular weight protein marker (Enzynomics, Korea).

3.5. Cysteine protease inhibition assay

Cysteine protease inhibitory activity of purified recombinant proteins was examined against the activity of papain enzyme on azo-casein hydrolysis under *in vitro* conditions. The percent inhibitory activity was plotted against the rprotein: papain concentration ratio and the inhibitory activity of rproteins was detected to be concentration-dependent (Fig. 11). The 1:1 concentration ratio exhibited the highest inhibitory activity against papain in all rproteins. Further increases in the rprotein concentrations did not significantly increase the inhibitory activity (P<0.05), as evidenced by a plateau shape on the graph and suggesting that the 1:1 ratio might provide the optimal conditions for the reaction elevation of its activity at provided

conditions. As a control, the papain inhibition experiment was conducted with the purified MBP alone, in place of the recombinant fusion protein product. The results demonstrated a negligible papain inhibitory activity, compared to the results obtained with the recombinant fusion proteins. Therefore, the MBP protein was considered as an inert component of the fusion protein.

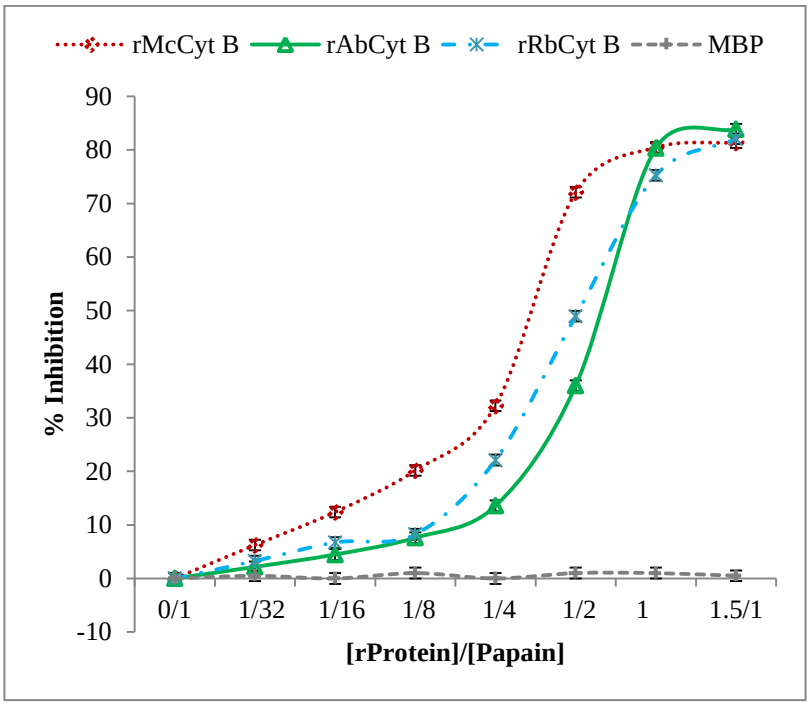

Fig. 11. Papain inhibitory activity profile of recombinant fusion proteins and maltose binding protein (MBP). Data are represented as means ± standard deviation (n=3).

3.6. Tissue specific transcriptional profile of cystatin B

Cystatin B mRNA expression in different tissues from healthy animals was analyzed by qPCR. The results showed that cystatin B was constitutively expressed in all tissues examined in three different species. RbCyt B showed lowest mRNA expression in muscle and highest expression in liver (Fig. 12A). However, both AbCyt B (Fig. 12B) and McCyt B (Fig. 12C) were prominently expressed in hemocytes while their least expression was detected in muscle and adductor muscle tissues, respectively.

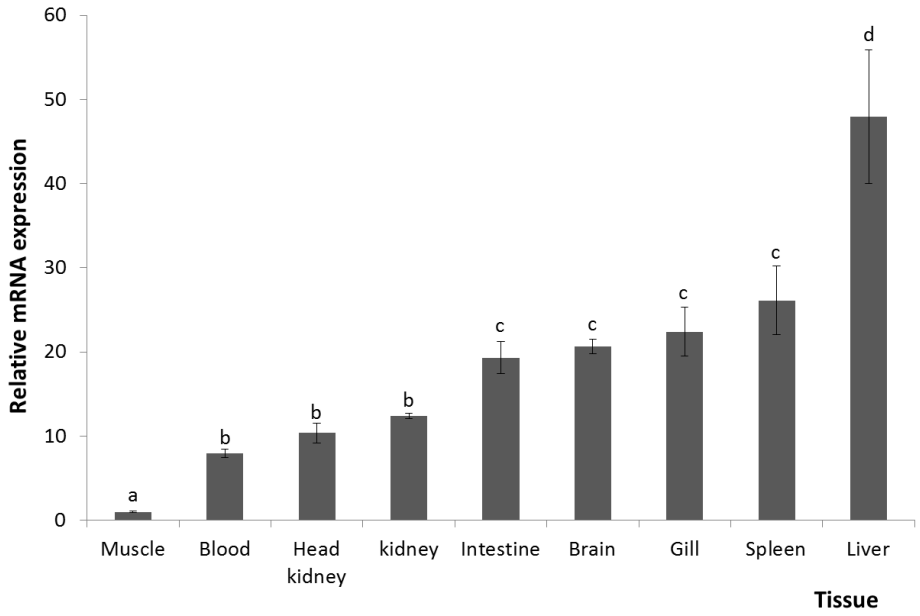

Fig. 12A. Tissue-specific mRNA expression of RbCyt B in healthy rock breams. The relative expression values were calculated by comparing to the detected expression in muscle tissue. Data are represented as means ± standard deviation (n=3). Statistical analysis was performed by one-way ANOVA followed by Duncan's Multiple Range test using the SPSS 16.0 program. Data with different letters are significantly different ($P < 0.05$) among different tissues.

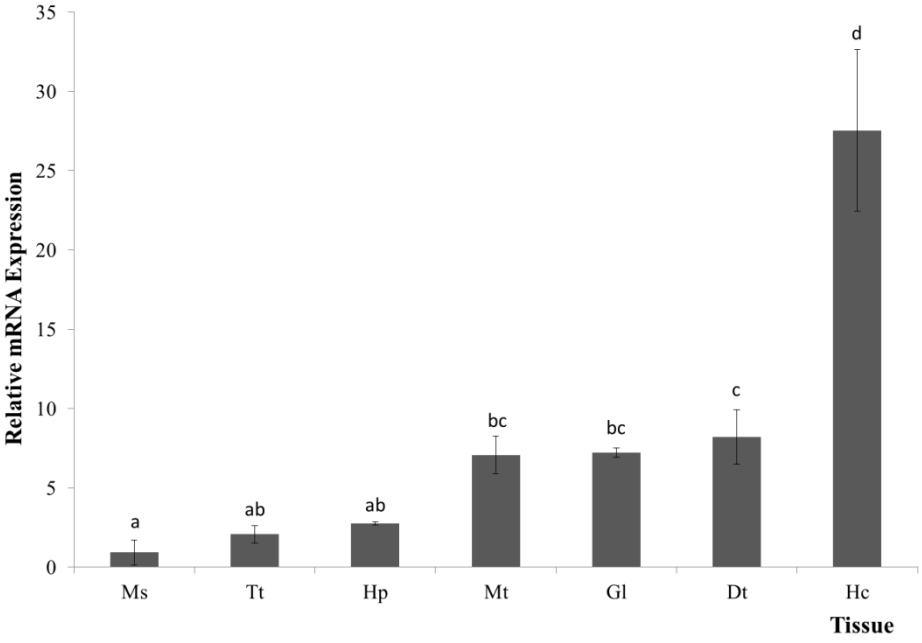

Fig. 12B. Tissue specific mRNA expression of AbCyt B in healthy disk abalones. The relative expression values were calculated by comparing to the detected expression in muscle tissue. Abbreviations: Ms, muscle; Tt, testis; Hp, hepatopancreas; Mt, mantle; Gl, gill; Dt, digestive tract and Hc, hemocytes. Data are represented as means ± standard deviation (n=3). Statistical analysis was performed by one-way ANOVA followed by Duncan's Multiple Range test using the SPSS 16.0 program. Data with different letters are significantly different ($P < 0.05$) among different tissues.

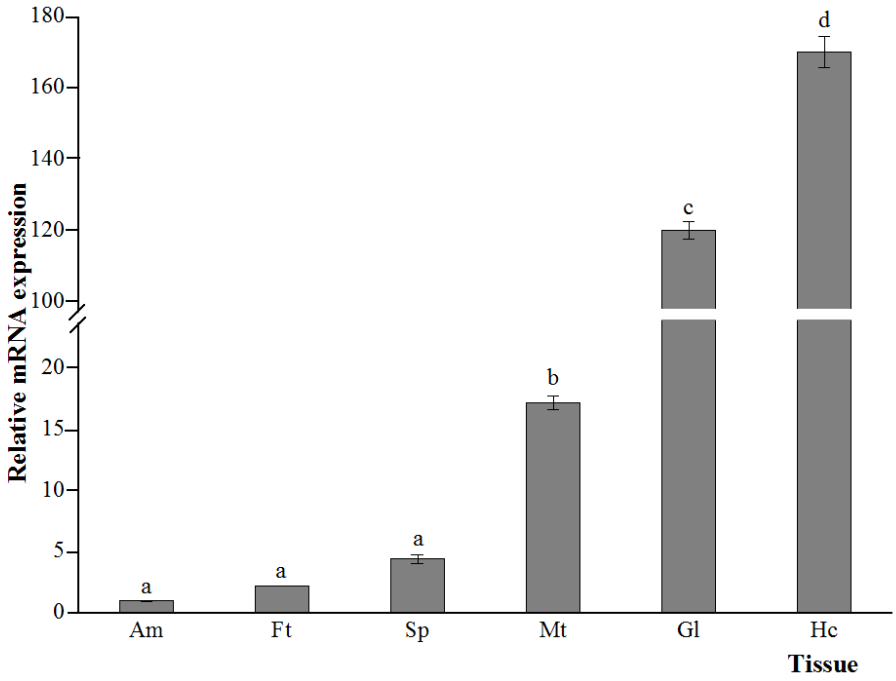

Fig. 12C. Tissue specific mRNA expression of McCyt B in healthy Manila clams. The relative expression values were calculated by comparing to the detected expression in adductor muscle. Abbreviations: Am, adductor muscle; Ft, foot; Sp, siphon; Mt, mantle; Gl, gill; Hc, hemocytes. Data are represented as means ± standard deviation (n=3). Statistical analysis was performed by one-way ANOVA followed by Duncan's Multiple Range test using the SPSS 16.0 program. Data with different letters are significantly different ($P < 0.05$) among different tissues.

3.7. Immune-regulated transcriptional profile of cystatin B

The transcriptional response analyzed after the immune challenge in rock breams with live *E. tarda* caused significant up-regulation of RbCyt B mRNA expression in head kidney at 12 h and 24 h p.i. ($P < 0.05$, vs. PBS control; Fig. 13A). Similarly, in spleen also it was detected significant up-regulation at 24 h and 48 h p.i. ($P < 0.05$, vs. PBS control; Fig. 13B). However, a significant down-regulation of RbCyt B expression was detected at 6h p.i. only in head kidney and the level of expression was very high in head kidney compared to the spleen (Fig. 13).

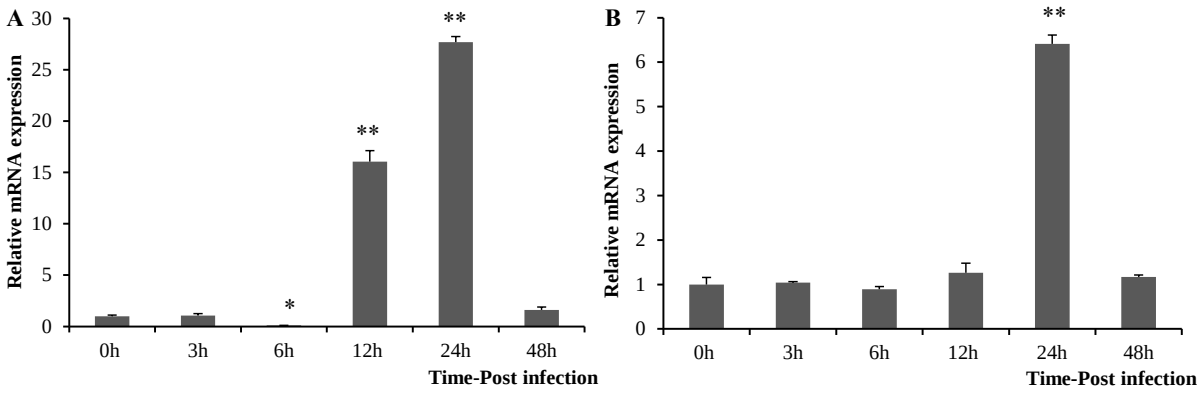

Fig. 13. RbCyt B mRNA expression in *E. tarda*-infected head kidney (A) and spleen (B) tissues. The expression levels detected by qPCR were normalized to beta-actin expression and calculated by comparing to that at the 0h (un-injected control) time point in each challenge. The data are represented as means ± standard deviation (n=3). Significant expression increase ($P < 0.05$ or $P < 0.01$) compared with control is indicated with '*' or '**' respectively.

In disk abalone, relative mRNA expression showed significant up-regulation ($P < 0.05$) in gills at 24h of p.i. with bacteria, and dropped down to the basal level at 48h of p.i. (Fig. 14A). However, in hemocytes, it was shown as relatively early response compared with the profile in gills (Fig. 14B). The significant ($P < 0.05$) induction was detected as early as 6 h of p.i., but afterward the expression was significantly down-regulated compared with the un-injected control (0h). In contrast, significant transcriptional modulation of McCyt B was observed in gill tissue and hemocytes, both of which are involved in immunity. Infection with the live bacterial pathogen, *V. tapetis*, exhibited a late phase increase in mRNA transcription of McCyt B. Exceptionally, in hemocytes significant elevations of McCyt B was detected at 3 h of p.i. in addition to the 48 h p.i. ($P < 0.05$), whereas in gill, relative mRNA level was induced only at 48 h p.i. (4-fold higher relative mRNA expression level than the basal expression; Fig. 15A). However, systemic challenge with a well-characterized bacterial cell wall component, LPS, remarkably influenced McCyt B relative mRNA expression in hemocytes, with significant down-regulation being detected at every time point after the endotoxin injection ($P < 0.05$; Fig. 15B). Moreover, except at 3 h p.i., gill tissue also

exhibited a generally down-regulated transcriptional profile upon LPS exposure. In contrast, McCyt B expression was significantly down-regulated from 6 h to 24 h p.i. in gill, but only under-expressed at 12 h p.i. ($P < 0.05$) in hemocytes. In contrast, poly I:C-stimulated clams showed a complex transcriptional profile of McCyt B, along with random positive and negative regulations (Fig. 15C).

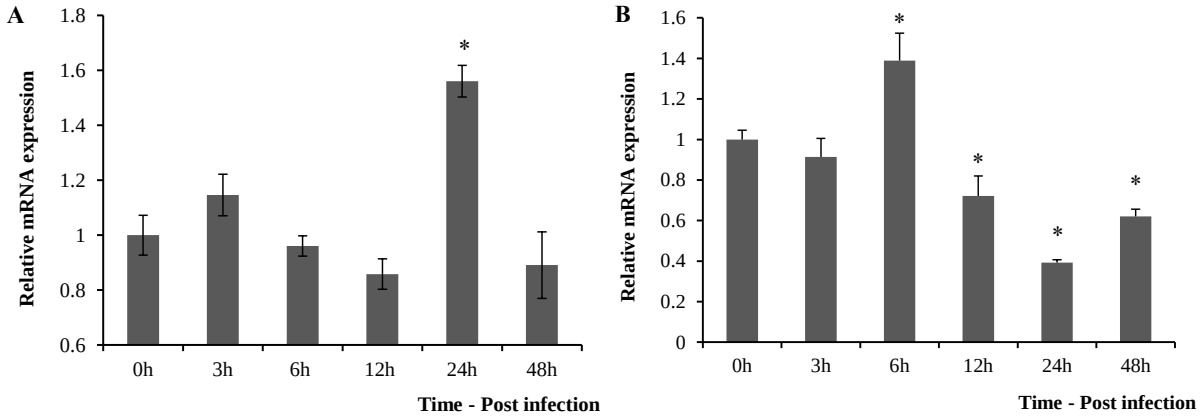

Fig. 14. AbCyt B mRNA expression after bacterial challenge in gill (A) and hemocytes (B). The relative mRNA expression was calculated by the $2^{-\Delta\Delta CT}$ method using abalone ribosomal protein L5 as the internal control and normalizing to the saline injected control at each time point. The data are represented as means ± standard deviation (n=3). Statistical analysis was performed by student's t-test. Asterisk indicates significant differences ($P < 0.05$) to the 0h (non-injected control).

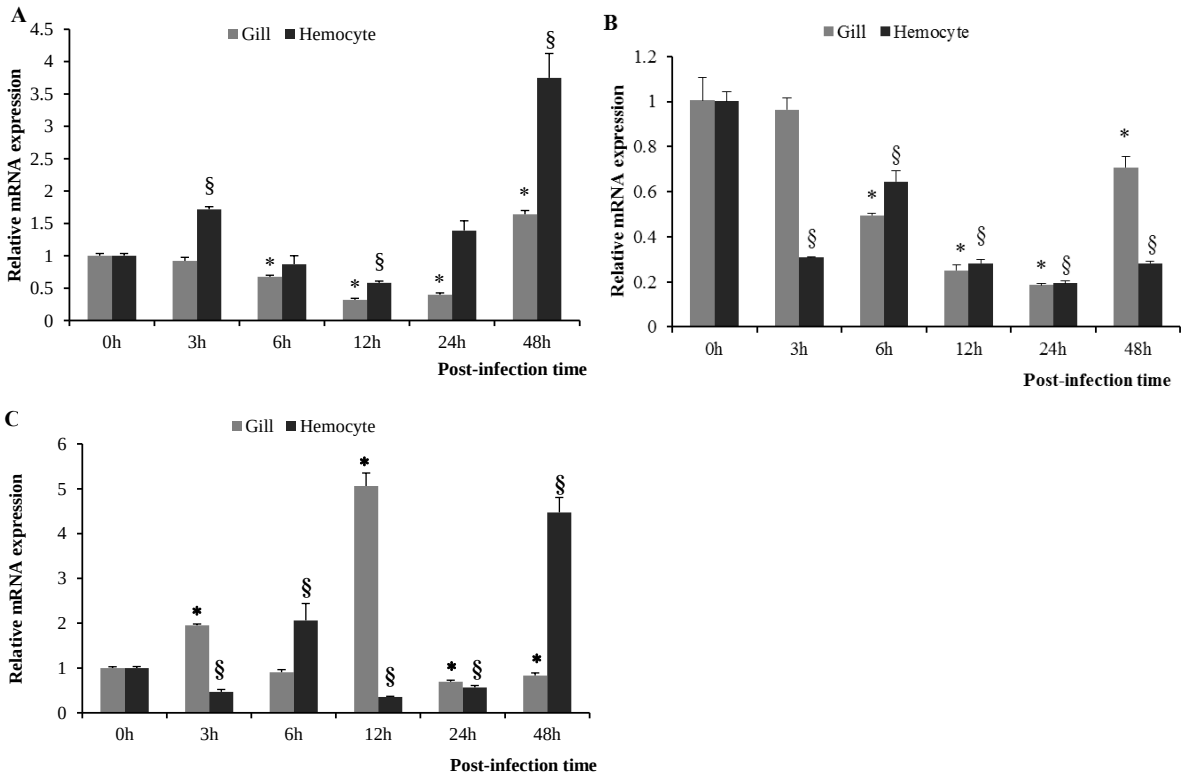

Fig. 15. Expression profile of McCyt B in gill and hemocytes upon (A) *V. tapetis* induction, (B) LPS induction, and (C) poly I:C induction. The relative mRNA expression was calculated by the $2^{-\Delta\Delta CT}$ method using beta-actin as the internal control and normalizing to the saline injected control at each time point. The data are represented as means ± standard deviation (n=3). Statistical analysis was performed by student's t-test. Symbols indicate significant differences ($P < 0.05$) to the 0h (non-injected control).

4. DISCUSSION

Cystatins function as reversible inhibitors of papain-like cysteine proteinases which constitute the largest and best described group of natural cysteine protease inhibitors (Bode et al., 1990). Interestingly, recent findings reported that cystatin B was involved in immune responses against invading bacteria in teleost, crustaceans and annelids (Lefebvre *et al*., 2004; Li et al., 2010; Xiao et al., 2010). In the present study, we isolated and characterized a gene related to cystatin superfamily from three different marine species, which possesses *in vitro* cysteine protease inhibitory activity and *in vivo* immune responses against invading pathogens.

According to the genomic structure comparison, all reported cystatin B sequences exhibited a common genomic structural architecture with three exons interrupted by two introns which confirm the unique genomic structure shared between different taxonomic classes. Furthermore, the second exon showed identical length to other known cystatin B sequences from all the classes, while the first exon coding region was observed to vary in length depending on the taxonomic class. Moreover, third exon coding region exhibited identical length in all compared sequences except in disk abalone which was little longer than all others. The 5'-UTRs and 3'-UTRs were showed remarkable variations in length within compared members.

In most of the protein coding genes, the TATA box facilitates transcription factor (TF) II D (TBP) binding to the promoter region, which is generally located at 25-30 bp upstream of the transcription start site (TSS) within the core promoter region, where core promoter can be extended ~35 bp upstream/ downstream to the transcription initiation site (Smale and Kadonaga, 2003). In RbCyt B, this TATA box sequence was identified within the typical region; however in AbCyt B it would not be identified. Most of the human core promoters

were reported to contain the consensus Inr motif (YYAN(T/A)YY) and considerable number of promoters were Inr motif containing TATA-less genes (Yang et al., 2007). Similar to the TATA box, the Inr motif facilitates the binding of TF II D (TBP) (Xi et al., 2007). In our study we also identified the Inr motif sequence in the promoter of AbCyt B. Additionally two E-boxes were identified within first -36 bp region, which are able to enhance the transcription of AbCyt B (Chaudhary and Skinner, 1999). Another TBP binding sequence and CAAT box sequence were identified, which function as regulatory elements for transcription initiation (Xi et al., 2007). Moreover, RbCyt B promoter indicated several potential TF binding sites. Mzf 1 is a TF belonging to the Krüppel family of zinc finger proteins, and has been reported to act as a tumor growth suppressor in the hemopoietic compartment (Gaboli et al., 2001). AP-1 mediates gene regulation in response to various physiological and pathological stimuli, such as cytokines, growth factors, stress signals, bacterial and viral infections, and oncogenic stimuli (Hess et al., 2004). In addition, SP-1 also can either activate or repress the transcription of the following gene in response physiological and pathological stimuli (Marin et al., 1997). Hence, the presence of these TF binding sites in the putative promoter region hints that the RbCyt B gene may be activated in response to some physiological and pathological stimuli. Similarly in AbCyt B, presence of HSF, NIT 2, GATA-1 and Evi-1 TF binding sites in promoter region suggest that AbCyt B may involve in physiological processes (Sorger, 1991; Bard-Chapeau et al., 2012).

The amino acid sequences identified from three experimental species contain around 100 amino acid residues and they were identified to be consisted of only a single CY domain including a cysteine protease inhibitor signature sequence. However there were no signal peptide, disulfide bonds and carbohydrate side chains identified, which indicate these three molecules are sharing characteristic features of the type 1 cystatin members (Turk and Bode, 1991; Abrahamson, 1994; Pemberton, 2006). Furthermore, all three cystatin molecules

indicated the predicted molecular mass around 11 kDa, which further confirm that they are belonging to the type 1 cystatins. Moreover, three characteristic conserved regions, N-terminal Gly, QxVxG motif and PW motif of cystatin superfamily members were found in the amino acid sequences, which are important for the cysteine protease inhibitory activity. The N-terminal Gly residue is located at 3rd position in both AbCyt B and McCyt B, however in RbCyt B it was at 6th position. The 3rd position for the G residue was homologues with human, rat, chicken and Pacific oyster sequences, whereas the 6th position was homologues to the fish cystatin B sequences considered in the multiple sequence alignment. The conserved QxVxG motif sequence is mostly presented as QVVAG sequence in number of animal species including human, rat, European seabass and Pacific oyster, which is exactly matched with the corresponding motif in AbCyt B sequence. However, this motif sequence in RbCyt B and McCyt B had two and one residue substitutions as glutamine-isoleucine-valine-serine-glycine (QIVSG) and glutamine-leucine-valine-alanine-glycine (QLVAG), respectively. Although, PW motif at second hairpin loop has a highly conserved P residue in all three molecules, the W residue was substituted by cysteine (C) in rock bream, tyrosine (Y) in abalone and histidine (H) in Manila clam. Ochieng and Chaudhuri also reported that the human stefin B had an H substitution instead W in its PW motif (Ochieng and Chaudhuri, 2010). Similarly, in our study it was observed as PH in many cystatin B sequences from different species used in the study. These conserved regions together are known to form a hydrophobic wedge-shaped edge which is highly complementary to the active site of the papain-like cysteine proteinases in order to inhibit their protease activity (Turk and Bode, 1991). In addition to these three regions, Pol and Björk (2001) reported a free Cys 3 in N-terminal segment of the proteinase-binding region in human and bovine cystatin B that is important for tight binding of target proteinases. In the present study, cysteine at 2nd position of the AbCyt B and McCyt B amino acid sequences and 5th position of the RbCyt B amino

acid sequence was found to be homologous to that of Cys 3 in human and bovine cystatin B. Moreover, this C residue was identified as a highly conserved residue among all cystatin B sequences used in the study.

The cystatin superfamily members was initially divided into three families (Barrett et al., 1986; Turk and Bode, 1991); however, a fourth family, consisting exclusively of cystatins from invertebrates, has now been described recently (Khaznadji et al., 2005; Li et al., 2010). The hierarchical clustering analysis performed in the present study considered all four families and in addition to the well described family 1-3, we also demonstrated a fourth family of cystatin, which was reported in recent studies and mostly consisted of nematode parasites (Khaznadji et al., 2005; Li et al., 2010). However, the family 1 and 3 consists of members from both vertebrate and invertebrate species and our interested members also clustered within type 1 cystatins and they were clearly clumped within cystatin B, where RbCyt B separated into teleost clad and AbCyt B and McCyt B felt into mollusk clad, further affirming their orthologous nature towards cystatin B variants. Moreover, this close evolutionary relationship provides evidence to support the notion that these three cystatins originated from a common ancestor.

According to our approach in generating a 3D structural model of cystatin B, in order to understand the structural-functional relationship, all three cystatin B molecules exhibited the typical features of the stefin family proteinases as described in the result section. The anticipated 3D models encompass the characteristic features of stefin family protease inhibitors, especially related to the human stefin/ cystatin B (Stubbs et al., 1990). Importantly, the penta peptide conserved motif (QxVxG) in the first hairpin loop, PW motif, and the glycine residue located close to the N-terminal of the molecule manifest that the potential of forming a wedged-shaped edge which is highly compatible to the active site of the papain-

like cysteine proteases, substantiating its structural and functional relationship with the known cystatin B members.

In order to characterize the biochemical properties of cystatin B proteins, the coding sequences were cloned into the pMAL-c2X expression vector and expressed as a fusion protein with MBP. The rproteins showed remarkable protease inhibitory activity against the proteolytic activity of papain to the azo-casein hydrolysis, suggesting these proteins are biologically active. The papain activity was >80% inhibited by recombinant cystatin Bs at the ratio of 1:1 concentration with papain. Further increases in the rprotein concentration did not significantly increase the inhibitory activity ($P < 0.05$), as evidenced by a plateau shape on the graph and suggesting that the 1:1 ratio might provide the optimal conditions for the reaction elevation of its activity at provided conditions. These results agree with several previously reported characterizations of cystatins from invertebrates. For example, when the concentration ratio of cystatin from jellyfish *Cyanea capillata* to papain reached 1:1, the activity of papain was completely inhibited (Yang et al., 2003). rEsCystatin in Chinese mitten crab (*Eriocheir sinensis*) and Limulus cystatin in Horseshoe crab (*Tachypleus tridentatus* L-Cystatin) were also reported 89% and 90% inhibition of papain activity at 300 μg mL^{-1} rEsCystatin and at 1:1 ratio of papain to Limulus cystatin, respectively (Agarwala et al., 1996; Li et al., 2010). Binding of cystatins to cysteine proteases may or may not result in conformational changes of either protein. For example, cystatin binding to papain does not induce any conformational change in either protein, whereas binding to cathepsin B involves the initial displacement of a loop that subsequently occludes the active site and facilitates tight binding (Pemberton, 2006). It has also been reported that cystatins do not form a covalent bond with cysteine proteases, but cover the active site cleft and effectively block the access to the active site (Calkins and Sloane, 1995). Previous studies reported that cathepsin B which is a key cysteine proteinase expressed extra-cellularly on the surface of tumor cells,

plays a key role in tumor cell invasion, and anti-protease activity of cystatin possesses a critical role at tumor cell invasion by inhibiting the activity of cathepsins (Ochieng and Chaudhuri, 2010). Moreover, in microbes, especially bacteria, cysteine proteases have elastin degradation properties, which may lead to a pathogenesis. For example, Arg-gingipain (RGP) and Lys-gingipain (KGP) produced by bacteria play a major role in pathogenesis and act as a virulence factor (Otto and Schirmeister, 1997). Therefore, it is clear that inhibitory activity of cystatin B plays an essential role in host immune defense system. However, further studies are needed for detailed clarification of the exact role of cystatin B in immunity.

Cystatins have been extensively studied in a wide variety of organisms. Functional roles in the progression of pathogenic processes, ranging from cancers to infectious diseases, have been recognized. Studies on mammalian cystatin B showed direct correlations with EPM1, which is a degenerative disease of the central nervous system caused by a mutation of cystatin B gene (Pennacchio et al., 1996). In addition to regulation of cysteine protease activity, Laitala-Leinonen et al. showed that cystatin B inhibits bone resorption by down-regulating intracellular cathepsin K activity despite increased osteoclast survival (Laitala-Leinonen et al., 2006). Bacterial-derived cysteine proteinases can act as virulence factors (Otto and Schirmeister, 1997) and host-derived cystatins can counteract with these molecules, aiding in immune defense by inhibiting their activity. Hence, cystatins are known to be involved in a broad spectrum of physiological and pathological processes, and are accordingly distributed throughout different types of cells and tissues in various species (Turk and Bode, 1991; Synnes, 1998; Abrahamson et al., 2003; Pemberton, 2006; Lefebvre et al., 2008). In the present study, we observed that all three cystatin Bs were universally expressed in different tissues examined. However, significantly higher expressions were observed in liver, spleen and gills from rock bream as well as in hemocytes, gills, mantle and digestive tract of mollusks, which are believed to be either function in immune regulation or

frequently contact with microbes. Similar high expression of cystatin B in hemolymph was reported in Horseshoe crab and Chinese mitten crab EsCystatin (Agarwala et al., 1996; Li et al., 2010). Hemolymph, gill and mantle are important tissues involved in innate immune functions in mollusks. Cellular responses in first line innate immunity of mollusks are carried out by circulating hemocytes which can kill microbes by phagocytosis and cytotoxic reactions (Pruzzo et al., 2005). Therefore, we suggested that AbCyt B and McCyt B might play a critical role associated with hemocytes. The gill tissue is a key respiratory organ whose surface is exposed to the external environment and frequently in contact with pathogenic microbes, and it was believed to be involved with the mollusks immune system (Chakraborty et al., 2010). In the present study, the lowest cystatin B expression from three species was detected in muscle in rock bream and abalone or adductor muscle in Manila clam; however, a contradictory result for the mRNA expression in muscle was reported in turbot cystatin B (*Sm*CytB) and EsCystatin in Chinese mitten crab where it was the highest expression (Li et al., 2010; Xiao et al., 2010). Although the turbot *Sm*CytB expression was strongest in muscle under normal (healthy) conditions, it was up-regulated in immune-related tissues, such as kidney, spleen, liver and brain, upon bacterial challenge.

Furthermore, studies in leech (*T. tessulatum*), Chinese mitten crab (*E. sinensis*), and turbot (*S. maximus*) demonstrated that cystatins were involved in immune defense activity against invading bacteria and fungi (Lefebvre et al., 2004; Li et al., 2010; Xiao et al., 2010). In our study, immune challenge of rock bream with *E. tarda* caused significant up-regulation of RbCyt B mRNA expression in head kidney and spleen. These results were highly comparable with a previous study reporting the responsiveness of turbot cystatin B (*Sm*CytB) against *E. tarda* infection (Xiao et al., 2010). Thus, it could be concluded that RbCyt B may play a major role in immune defense against invading bacteria in rock bream. Similarly, the relative mRNA expression level of AbCyt B was significantly up-regulated ($P < 0.05$) in gill

and hemocytes as a response to the pathogenic bacteria in disk abalone. It was also observed some down-regulated expressions in hemocytes in latter stage. Similar results were reported in EsCystatin against bacteria as down-regulation at 3 h p.i. and up-regulation of mRNA expression at 24 h p.i. in hemolymph (Li et al., 2010). However, significant transcriptional modulation of McCyt B was observed in gill tissue and hemocytes, both of which are involved in immunity. Systemic challenge with a well-characterized bacterial cell wall component; LPS remarkably influenced McCyt B mRNA expression in hemocytes, with significant down-regulation being detected at every time point after the endotoxin injection. Moreover, except at 3 h p.i., gill tissue also exhibited a generally under-expressed transcriptional profile upon LPS exposure. Upon infection with the live bacterial pathogen, *V. tapetis*, transcriptional response of McCyt B was observed to be initially as down-regulation, however in late phase the expression was significantly increased in both tissues examined. The overall transcriptional modulation of cystatin B that was observed in Manila clam upon *V. tapetis* stimulation suggests a late phase active mechanism for McCyt B in host pathology. These observations strongly indicate a potential active involvement of cathepsin-like proteinases in Manila clam tissues in response to Gram negative bacterial infections, whereby cystatins would be made less abundant in the affected tissues. Correspondingly, the expression of invertebrate linage cystatins from duck leech and Chinese mitten crab are reported to be augmented at the transcriptional level upon bacterial induction (Lefebvre et al., 2004; Li et al., 2010). In duck leech, the transcript level of its cystatin B homolog was elevated in their large coelomocytes at 2 h p.i. upon *E. coli* stimulation, at 3 h and 24 h p.i. upon *Micrococcus luteus* stimulation, and at 2 h p.i. upon induction with a mixture of the two bacteria. Likewise, in Chinese mitten crab, the mRNA level of a family 1 cystatin member was enhanced in hemolymph upon *Listonella anguillarum* infection at 24 h p.i., while *Pichia pastoris* infection enhanced transcription at 3 h (peak), 6 h, 12 h, and 48 h p.i. However,

when the Chinese mitten crab was challenged with *L. anguillarum*, cystatin B expression was down-regulated at 3 h p.i, which agreed with down-regulated transcriptional profile observed for McCyt B in response to *V. tapetis* (Li et al., 2010). Cathepsins, well-known cysteine proteases, are reported to be involved in immune responses through the toll like receptor signaling pathway (Creasy and McCoy, 2011). A previous study in *Drosophila melanogaster* showed that bacterial infections up-regulated the expression of cathepsins (De Gregorio et al., 2001). As lysosomal cysteine proteases play major role in MHC class II-mediated antigen presentation by processing and degradation of antigens to peptides are known to be protease activity-dependent, a considerable down-regulation of cystatin gene expression thereby can be expected during the bacterial infection (Turk et al., 2000). However, due to scanty information on cysteine proteases and cystatin B-like genes from mollusks, it is unclear about the functional relationship between cysteine proteases and their inhibitors in immune responses. Further studies might elucidate the link between cysteine protease inhibitor and their role in bacterial infections. We report these results for the first time in mollusks species, and these results suggest that AbCyt B and McCyt B are immune responsive and inducible by bacterial infection.

In our study, transcriptional response of an invertebrate cystatin upon challenge with the viral dsRNA mimic immune stimulant, poly I:C, was investigated in Manila clam. Poly I:C-stimulated clams showed a complex behavior of transcriptional profile of McCyt B, along with two early and late positively regulated responses in both gill and hemocytes. The early response elicited in gill, compared to that observed in hemocytes, may be attributed to the frequent exposure of gill tissues to various pathogens, through its continuous interaction with the outer environment (Verdot et al., 1999; Vray and Hartmann, 2002). As cystatins, including cystatin B, are known to play an inhibitory role on cysteine proteases, such as cathepsin B, H and L (Turk and Bode, 1991), a considerable down-regulation of cystatin

gene expression is expected to occur during an immune response to an invasive pathogen. Furthermore, in human placental macrophages, down-regulated expression of cystatin B was reported to facilitate the inhibition of human immunodeficiency virus replication via a STAT-1 interaction (Luciano-Montalvo and Melendez, 2009). On the other hand, cystatin expression can be elevated under some other pathological conditions, reflecting the stimulatory role of these molecules to induce production of different cytokines, such as IL-10 and TNFα, in mammalian macrophages (Verdot et al., 1999; Vray and Hartmann, 2002).

CONCLUSIONS

The complete genomic sequences encoding cystatin B from rock bream and disk abalones, as well as the complete cDNA sequence from Manila clam were identified. The genomic structure of cystatin B share common features between taxonomic classes. The amino acid sequences from all three species exhibit high similarity to cystatin B homologues from different taxonomic classes, especially the rock bream cystatin B to the teleosts and disk abalone and Manila clam cystatin Bs to the mollusks. Furthermore, all these three molecules harbor characteristic structural features and motifs of family 1 cystatin, implying all three proteins are cystatin B homologs. Moreover, the recombinant cystatin B proteins possessed an *in vitro* papain inhibitory activity, affirming its cysteine protease suppressive function. Finally, according to the overall transcriptional behavior of these three cystatin B genes upon *in vivo* immune challenges, we can suggest that they are responsive to different pathological conditions specific to the each species, acting directly on pathogen-derived or host immune-related factors, possibly leading to inhibition of cysteine proteinases. Collectively, these observations speculate the putative involvement of cystatin Bs in the immune defense against invading pathogens in tested species.

REFERENCES

Abe K, Emori Y, Kondo H, Arai S, Suzuki K. 1988. The NH2-Terminal 21 Amino Acid Residues Are Not Essential for the Papain-Inhibitory Activity of Oryzacystatin, a Member of the Cystatin Superfamily. J. Biol. Chem. **263**:7655-59.

Abrahamson M, Alvarez-Fernandez M, Nathanson CM, 2003. Cystatins. Biochem. Soc. Symp. **70**:179-99.

Abrahamson M. 1994. Cystatins. Meth. Enzymol. **244**:685-700.

Agarwala KL, Kawabata S, Hirata M, Miyagi M, Tsunasawa S, Iwanaga S. 1996. A cysteine protease inhibitor stored in the large granules of horseshoe crab hemocytes: purification, characterization, cDNA cloning and tissue localization. J. Biochem. **119**: 85-94.

Alavaikko M, Aine R, Rinne A, Jarvinen M, Blanco G, Apaja-Sarkkinen M, Hopsu-Havu. VK. 1985. Behaviour of Dendritic Reticulum Cells Possessing Immunoreactive Acid Cysteine-Proteinase Inhibitor in Human Lymphoid Secondary Follicles and in Follicular-Centre Cell Lymphomas. Int. J. Cancer. **35**:319-25.

Anastasi A, Brown MA, Kembhavi AA, Nicklin MJH, Sayers CA, Sunter DC, Barrett AJ. 1983. Cystatin, a Protein Inhibitor of Cysteine Proteinases. Biochem. J. **211**:129-38.

Bard-Chapeau EA, Jeyakani J, Kok CH, Muller J, Chua BQ, Gunaratne J, et al., 2012. Ecotopic viral integration site 1 (EVI1) regulates multiple cellular processes important for cancer and is a synergistic partner for FOS protein in invasive tumors. PNAS doi: 10.1073/pnas.1119229109.

Barrett AJ, 1987. The cystatins: a new class of peptidase inhibitors. Trends Biochem. Sci. **12**: 193-96.

Barrett AJ, Fritz H, Grubb A, Isemura S, Jarvinen M, Katunuma N, Machleidt W, Muller-Esterl W, Sasaki M, Turk V. 1986. Nomenclature and Classification of the Proteins Homologous with the Cysteine-Proteinase Inhibitor Chicken Cystatin. Biochem. J. **236**:312.

Barrett AJ. 1981. Cystatin, the Egg White Inhibitor of Cysteine Proteinases. Meth. EnzymoL. **80**:771-78.

Barrett AJ. 1986. An introduction to the proteinases. In Proteinase Inhibitors; Barrett, AJ., Salvesen, G., Eds.; Elsevier: Amsterdam. pp 3-22.

Berdowska I, Siewinski M. 2000. The role of cysteine cathepsins and their inhibitors in physiological and neoplastic processes. Postepy biochemii. **46**:73-84.

Björk I, Ylinenjärvi K. 1989. Interaction of chicken cystatin with inactivated papains. Biochem. J. **260:**61-8.

Bobek LA, Levine MJ. 1992. Cystatins – Inhibitors of cysteine proteinases. Crit. Rev. Oral Biol. Med. **3**(4):307-32.

Bode W, Engh R, Musil D, Laber B, Stubbs M, Huber R, et al., 1990. Mechanism of interaction of cysteine proteinases and their protein inhibitors as compared to the serine proteinase-inhibitor interaction. Biol. Chem. Hoppe Seyler. **371**:111-118.

Bode W, Huber R. 2000. Structural basis of the endoproteinase-protein inhibitor interaction. Biochem. Biophys. Acta. **1477**:241-52.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

Brix K, Dunkhorst A, Mayer K, Jordans S. 2008. Cysteine cathepsins: cellular roadmap to different functions. Biochimie. **90**:194-207.

Bromme D, Wilson S. 2011. Role of Cysteine Cathepsins in Extracellular Proteolysis. In W.C. Parks and R.P. Mecham (eds.), Extracellular Matrix Degradation, Biology of Extracellular Matrix 2, Verlag Berlin Heidelberg. pp 23-51.

Brzin J, Ritonja A, Popovic T, Turk V. 1990. Low Molecular Mass Protein Inhibitor of Cysteine Proteinases from Soybean. Biol. Chem. Hoppe-Seyler **371** (Suppl.):167-170.

Calkins CC, Sloane BF. 1995. Mammalian cysteine protease inhibitors: biochemical properties and possible roles in tumor progression. Biol. Chem. Hoppe Seyler **376**: 71-80.

Chakraborty S, Ray M, Ray S. 2010. Toxicity of sodium arsenite in the gill of an economically important mollusc of India. Fish Shellfish Immunol. **29**: 136-48.

Chapman HA, Riese RJ, Shi GP. 1997. Emerging roles for cysteine proteases in human biology. Ann. Rev. Physiol. **59**:63-88.

Chaudhary J, Skinner MK, 1999. Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. Mol. Endocrinol. **13**:774-86.

Cimerman N, Mesko BP, Krasovec M, Suskovic S, Kos J. 2001. Serum concentration and circadian profiles of cathepsins B, H and L, and their inhibitors, stefins A and B, in asthma. Clin. Chim. Acta **310**:113-122.

Cornwall GA, Hsia N. 2003. A new subgroup of the family 2 cystatins. Mol. Cell. Endocrinol. **200**:1-8.

Creasy BM, McCoy KL. 2011. Cytokines regulate cysteine cathepsins during TLR responses. Cell. Immunol. **267**:56-66.

Cudic M, Fields GB. 2009. Extracellular proteases as targets for drug development. Curr. Protein Pept. Sci. **10**(4):297-307.

De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. 2001. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. Proc. Natl. Acad. Sci. USA **98**: 12590-95.

Di Giaimo R, Riccio M, Santi S, Galeotti C, Ambrosetti DC, Melli M. 2002. New insights into the molecular basis of progressive myoclonus epilepsy: a multiprotein complex with cystatin B. Hum. Mol. Genet. **11**:2941-50.

Dubey VK, Pande M, Singh BK, Jagannadham MV. 2007. Papain-like proteases: Applications of their inhibitors. African Journal of Biotechnology. **6**(9): 1077-86.

Fossum K, Whitaker JR. 1968. Ficin and Papain Inhibitor from Chicken Egg White. Arch. Biochem. Biophys. **125**:367-75.

Gaboli M, Kotsi PA, Gurrieri C, Cattoretti G, Ronchetti S, Cordon-Cardo C, Broxmeyer HE, Hromas R, Pandolfi PP. 2001. Mzf1 controls cell proliferation and tumorigenesis. Genes Dev. **15**:1625-30.

Goulletquer P. 1997. A bibliography of the manila clam *Tapes philippinarium*. IFREMER, RIDRV-97.02/RA/La Tremblade, 122 p.

Hess J, Angel P, Schorpp-Kistner M. 2004. AP-1 subunits: quarrel and harmony among siblings. J. Cell Sci. **117**:5965-73.

Hooper C, Day R, Slocombe R, Handlinger J, Benkendorff K. 2007. Stress and immune responses in abalone: Limitations in current knowledge and investigative methods based on other models. Fish Shellfish Immunol. **22**:363-79.

Jarvinen M. 1978. Purification and Some Characterization of the Human Epidermal SH-Protease Inhibitor. J. Invest. Dermatol. **71**:114-18.

Joyce JA, Baruch A, Chehade K, Meyer-Morse N, Giraudo E, Tsai FY, Greenbaum DC, Hager JH, Bogyo M, Hanahan D. 2004. Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. Cancer Cell. **5**:443-53.

Keilova H, Tomasek V. 1974. Effect of Papain Inhibitor from Chicken Egg White on Cathepsin B. Biochem. Biophys. Ada **334**:179-86.

Keilova H, Tomasek V. 1977. Naturally Occurring Inhibitors of Intracellular Proteinases. Acta Biol. Med. Ger. **36**:1873-81.

Kellermann J, Thelen C, Lottspeich F, Henschen A, Vogel R, Muller-Esterl W, 1987. Arrangement of the disulphide bridges in human low-Mr kininogen. Biochem. J. **247**: 15-21.

Keppler D. 2006. Towards novel anti-cancer strategies based on cystatin function. Cancer Letters. **235**,(2):159-76.

Khaznadji E, Collins P, Dalton JP, Bigot Y, Moire N. 2005. A new multi-domain member of the cystatin superfamily expressed by Fasciola hepatica. Int. J Parasitol. **35**:1115-25.

Kopitar-Jerala N. 2006. The role of cystatins in cells of the immune system. FEBS Lett. **580**: 6295-301.

Laitala-Leinonen T, Rinne R, Saukko P, Väänänen HK, Rinne A. 2006. Cystatin B as an intracellular modulator of bone resorption. Matrix Biol. **25**:149-57.

Lee Y, Whang I, Umasuthan N, De Zoysa M, Oh C, Kang DH, et al. 2011. Characterization of a novel molluscan MyD88 family protein from manila clam, *Ruditapes philippinarum*. Fish Shellfish Immunol. **31**:887-93.

Lefebvre C, Cocquerelle C, Vandenbulcke F, Hot D, Huot L, Lemoine Y, et al. 2004. Transcriptomic analysis in the leech *Theromyzon tessulatum*: involvement of cystatin B in innate immunity. Biochem J. **380**:617-25.

Lefebvre C, Vandenbulcke F, Bocquet B, Tasiemski A, Desmons A, Verstraete M, et al., 2008. Cathepsin L and cystatin B gene expression discriminates immune coelomic cells in the leech *Theromyzon tessulatum*. Dev. Comp. Immunol. **32**:795-807.

Lenney JF, Tolan JR, Sugai WJ, Lee AG. 1979. Thermostable endogenous inhibitors of cathepsins B and H. Eur. J. Biochem. **101**:153-61.

Li F, Gai X, Wang L, Song L, Zhang H, Qiu L, Wang M, Siva VS. 2010. Identification and characterization of a Cystatin gene from Chinese mitten crab *Eriocheir sinensis*. Fish Shellfish Immunol. **29**:521-29.

Lieuallen K, Penancchio LA, Park M, Myers RM, Lennon GG. 2001. Cystatin B-deficient mice have increased expression of apoptosis and glial activation genes. Hum. Mol. Genet. **10**: 1867-71.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. **25**:402-8.

Luciano-Montalvo C, Melendez LM. 2009. Cystatin B associates with signal transducer and activator of transcription 1 in monocyte-derived and placental macrophages. Placenta. **30**:464-7.

Maina CV, Riggs PD, Grandea AG, Slatko BE, Moran LS, Tagliamonte JA, et al., 1988. An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene **74**:365-73.

Margis R, Reis EM, Villeret V. 1998. Structural and phylogenetic relationships among plant and animal cystatins. Arch. Biochem. Biophys. **359**:24-30.

Marin M, Karis A, Visser P, Grosveld F, Philipsen S. 1997. Transcription Factor Sp1 Is Essential for Early Embryonic Development but Dispensable for Cell Growth and Differentiation. Cell. **89**(4):619-28.

McBride SC. 1998. Current status of abalone aquaculture in the Californias. J. Shellfish Res. **17**(3):593-600.

Mizuno K, Nakamura T.; Takada, K, Sakakibara S, Matsuo H. 1987. A membrane-bound, calcium-dependent protease in yeast alpha-cell cleaving on the carboxyl. Biochem. Biophys. Res. Commun. **144**:807-14.

Mottram JC, Coombs GH, Alexander J. 2004. Cysteine peptidases as virulence factors of *Leishmania*. Curr. Opin. Microbiol. **7**:375-81.

Munasinghe H, Kang HS, Lee J. 2006. Analysis of Digestive Gland Expressed Sequence Tag Library from the Disk Abalone, *Haliotis discus discus*. J. World Aquac. Soc. **37**:96-106.

Ochieng J, Chaudhuri G. 2010. Cystatin superfamily. J. Health Care Poor Underserved. **21**: 51-70.

Okumichi T, Nishiki M, Takasugi S, Yamane M, Ezaki H. 1984. Purification of Thiol Protease Inhibitor from Human Lung Cancer Tissue. Hiroshima J. Med. Sci. **33**:801-8.

Otto HH, Schirmeister T. 1997. Cysteine Proteases and Their Inhibitors. Chem. Rev. **97**:133-72.

Park SL. 2009. Disease Control in Korean Aquaculture. Fish Pathol. **44**(1):19-23.

Pemberton PA. 2006. Cystatins, In: Geoffrey, J.L., Steven, D.S. (Eds.), Encyclopedia of respiratory medicine, Elsevier. pp. 511-17.

Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, D'Amato E, Ramirez L, Faham M, Koskiniemi M, Warrington JA, Norio R, De la Chapelle A, Cox DR, Myers RM. 1996. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). Science **271**:1731-4.

Pol E, Björk I. 2001. Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases. Protein Sci. **10**:1729-38.

Potempa J, Golonka E, Filipek R, Shaw LN. 2005. Fighting an enemy within: cytoplasmic inhibitors of bacterial cysteine proteases. Mol. Microbiol. **57**:605-10.

Pruzzo C, Gallo G, Canesi L. 2005. Persistence of vibrios in marine bivalves: the role of interactions with haemolymph components. Environ. Microbiol. **7**:761-72.

Rahman KM, Khan HM, Haq JA. 1983. Incidence of cutaneous diphtheria in Bangladesh. Bangladesh Medical Research Council bulletin. **9**:49-53.

Riccio M, Santi S, Dembic M, Di Giaimo R, Cipollini E, Costantino-Ceccarini E, et al., 2005. Cell-specific expression of the epm1 (cystatin B) gene in developing rat cerebellum. Neurobiol. Dis. **20**:104-14.

Rinne R, Saukko P, Järvinen M, Lehesjoki AE. 2002. Reduced cystatin B activity correlates with enhanced cathepsin activity in progressive myoclonus epilepsy. Ann. Med. **34**(5):380-85.

Rodriguez-Mahillo AI, Gonzalez-Munoz M, Gomez-Aguado F, Rodriguez-Perez R, Corcuera MT, Caballero ML, Moneo I. 2007. Cloning and characterisation of the *Anisakis simplex* allergen Ani s 4 as a cysteine-protease inhibitor. Int. J. Parasitol. **37**:907–17.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. **5**:725-38.

Rudenskaya GN, Pupov DV. 2008. Cysteine proteinases of microorganisms and viruses. Biochemistry (Mosc). **73**:1-13.

Rzychon M, Chmiel D, Stec-Niemczyk J. 2004. Modes of inhibition of cysteine proteases. Acta biochimica Polonica. **51**:861-73.

Sajid M, McKerrow JH. 2002. Cysteine proteases of parasitic organisms. Molecular & Biochem. Parasitol. **120**:1-21.

Sen LC, Whitaker JR. 1973. Some Properties of a Ficin-Papain Inhibitor from Avian Egg White. Arch. Biochem. Biophys. **158**:623-32.

Shah A, Bano B. 2009. Cystatins in Health and Diseases. Int. J. Pept. Res. Ther. **15**:43-8.

Sharma A, Padwal-Desai S, Ninjoor V. 1989. Intracellular hydrolases of *Aspergillus parasiticus* and *Aspergillus flavus*. Biochem. Biophys., Res. Commun., **159**:464-71.

Sloane B, Honn K. 1984. Cysteine Proteinases and Metastasis. Cancer Metastasis Rev. **3**:249-63.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. Annu. Rev. Biochem. **72**:449-79.

Sorger PK. 1991. Heat shock factor and the heat shock response. Cell **65**:363-6.

Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, et al., 1990. The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J. **9**:1939-47.

Suzuki T, Hashimoto S, Toyoda N, Nagai S, Yamazaki N, Dong HY, et al. 2000. Comprehensive gene expression profile of LPS-stimulated human monocytes by SAGE. Blood. **96**:2584-91.

Synnes M. 1998. Purification and characterization of two cysteine proteinase inhibitors from the skin of Atlantic salmon (*Salmo salar* L.). Comp. Biochem. Physiol. B Biochem. Mol. Biol. **121**:257-64.

Takahashi M, Tezuka T, Katunuma N. 1994. Inhibition of growth and cysteine proteinase activity of *Staphylococcus aureus* V8 by phosphorylated cystatin alpha in skin cornified envelope. FEBS Lett. **355**:275-8.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731-9.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids res. **22**:4673-80.

Turk B, Turk D, Turk V, 2000. Lysosomal cysteine proteases: more than scavengers. Biochim. Biophys. Acta. **1477**: 98-111.

Turk B, Turk V, Turk D. 1997. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. Biol. Chem. **378**:141-50.

Turk V, Bode W. 1991. The cystatins: protein inhibitors of cysteine proteinases. FEBS lett. **285**:213-19.

Turk V, Brzin J, Lenarcic B, Locnikar P, Popovic T, Ritonja A, Babnik J, Bode W, Machleidt W.1985. Structure and Function of Lysosomal Cysteine Proteinases and Their Protein Inhibitors. In: E. A. Khairallah, J. S. Bond, and J. W. C. Bird, Eds. Intracellular Protein Catabolism, Alan R. Liss, New York. pp. 91-103.

Turk V, Kos J, Turk B. 2004. Cysteine cathepsins (proteases)--on the main stage of cancer? Cancer Cell. **5**:409–10.

Turk V, Turk B. 2008. Lysosomal Cysteine Proteases and Their Protein Inhibitors: Recent Developments. Acta Chim. Slov. **55**:727–38.

Umasuthan N, Revathy KS, Lee Y, Whang I, Choi CY, Lee J. 2012. A novel molluscan sigma-like glutathione S-transferase from Manila clam, *Ruditapes philippinarum*: cloning, characterization and transcriptional profiling. Com. Biochem. Physiol. C Toxicol. Pharmacol: CBP. **155**:539-50.

Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V, Turk B 2007. Emerging roles of cysteine cathepsins in disease and their potential as drug targets. Curr. Pharm. Des. **13**:387–403.

Verdot L, Lalmanach G, Vercruysse V, Hoebeke J, Gauthier F, Vray B. 1999. Chicken cystatin stimulates nitric oxide release from interferon-gamma-activated mouse peritoneal macrophages via cytokine synthesis. Eur. J. Biochem. FEBS. **266**:1111-7.

Villena AJ. 2003. Applications and needs of fish and shellfish cell culture for disease control in aquaculture. Rev. Fish. Biol. Fish **13**:111-40.

Vray B, Hartmann S, Hoebeke J. 2002. Immunomodulatory properties of cystatins. CMLS, Cell. Mol. Life Sci. **59**:1503-12.

Wan Q, Whang I, Choi CY, Lee JS, Lee J. 2011. Validation of housekeeping genes as internal controls for studying biomarkers of endocrine-disrupting chemicals in disk abalone by real-time PCR. Comp. Biochem. Physiol. C Toxicol. Pharmacol. **153**:259-68.

Whang I, De Zoysa M, Nikapitiya C, Lee Y, Kim Y, Lee S, Oh C, Jung SJ, Oh MJ, Choi CY, Yeo SY, Kim BS, Kim SJ, Lee J. 2011a. Molecular characterization and expression analysis of Cathepsin B and L cysteine proteases from rock bream (*Oplegnathus fasciatus*). Fish Shellfish Immunol. **30**:763-72.

Whang I, Lee Y, Lee S, Oh MJ, Jung SJ, Choi CY, Lee WS, Kim HS, Kim SJ, Lee J. 2011b. Characterization and expression analysis of a goose-type lysozyme from the rock bream *Oplegnathus fasciatus*, and antimicrobial activity of its recombinant protein. Fish Shellfish Immunol. **30**:532-42.

Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z. 2007. Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. Genome Res. **17**:798-806.

Xiao PP, Hu YH, Sun L. 2010. *Scophthalmus maximus* cystatin B enhances head kidney macrophage-mediated bacterial killing. Dev. Comp. Immunol. **34**:1237-41.

Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene **389**:52-65.

Yang Y, Cun S, Peng L, Xie X, Wei J, Yang W, et al. 2003. cDNA cloning, identification and characterization of a novel cystatin from the tentacle of *Cyanea capillata*. Biochimie. **85**:1033-9.

Yasuda Y, Kaleta J, Brömme D. 2005. The role of cathepsins in osteoporosis and arthritis: rationale for the design of new therapeutics. Adv. Drug Deliv. Rev. **57**:973-93.

Zhang L, Zhao J, Li C, Su X, Chen A, Li T, Qin S. 2011. Cloning and characterization of allograft inflammatory factor-1 (AIF-1) from Manila clam *Venerupis philippinarum*. Fish Shellfish Immunol. **30**(1):148-53.

Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. **9**:40.

ACKNOWLEDGEMENT

First and foremost I want to thank my supervisor, Prof. Jehee Lee for his support, understanding and patience during the past two years over this project. Without his guidance as a great mentor, this work would not have been possible, and I certainly wouldn't be here. It has been an honor to be a student of his research laboratory and his valuable assessments and insights have given me an opportunity to grow academically and professionally throughout the research process. I would also like to thank Dr. Ilson Whang and Dr. Qiang Wan for taking their valuable time and providing great input into this work.

Secondly, I lend my sincere thanks to Dr. Mahanama De Zoysa for introducing and recommending me to Prof. Lee, and helping in my research works by spending his valuable time. I would also like to acknowledge Dr. B.C. Jayawardena, who offered this opportunity to me through Dr. De Zoysa.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Choon-Bok Song and Prof. Seungheon Lee, for their encouragement, insightful comments, and hard questions.

I am also greatly indebted to companionship of a small circle of close friends, including Dr. Janaka, Kalpa, Niroshana, Anushka, Sanjaya, Uma and all research students in my lab namely, Dr. Youngdeuk, Saranya, Yucheol, Sukkyoung, Hyowon, Minyoung and Sung-do who have continuously supported me throughout these years. Without the love and support of my family, this would have been a very hard journey. I thank my father for teaching me good values such as hard work and appreciation for the gift of life and my mother without whom I would not be the person that I am today. My love for them is eternal. Finally, I am dedicating this work to my beautiful fiancé, Heshani, for giving me an unwavering love and support. I dare not even imagine how it would have been without you! I strive every day to make you proud!