A THESIS FOR THE DEGREE OF DOCTOR OF

PHILOSOPHY

Insights into Immunological Role of Teleost Complement System:

Genomic Structural Identification and Molecular

Characterization of Consecutive Nine Complement Components

of Rock Bream (*Oplegnathus fasciatus*)

W. D. NIROSHANA WICKRAMAARACHCHI

Department of Marine Life Sciences

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

REPUBLIC OF KOREA

FEBRUARY 2014

A THESIS FOR THE DEGREE OF DOCTOR OF

PHILOSOPHY

Insights into Immunological Role of Teleost Complement System:

Genomic Structural Identification and Molecular

Characterization of Consecutive Nine Complement Components

of Rock Bream (*Oplegnathus fasciatus*)

W. D. NIROSHANA WICKRAMAARACHCHI

Department of Marine Life Sciences

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

REPUBLIC OF KOREA

FEBRUARY 2014

Insights into Immunological Role of Teleost Complement System: Genomic Structural Identification and Molecular Characterization of Consecutive Nine Complement Components of Rock Bream (*Oplegnathus fasciatus*)
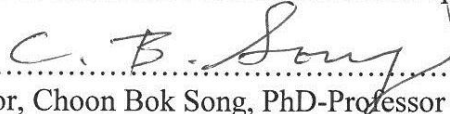
W. D. NIROSHANA WICKRAMAARACHCHI

(Supervised by Professor Jehee Lee)

A thesis submitted in partial fulfillment of the requirement for the degree of
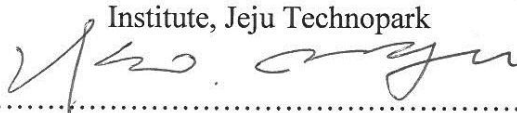
DOCTOR OF PHILOSOPHY

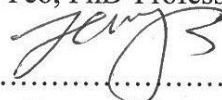### FEBRUARY, 2014

The thesis has been examined and approved by

....

Thesis Director, Choon Bok Song, PhD-Professor of Marine Life Sciences

....
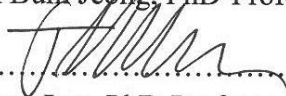
Mi-Hee Ko, PhD-Head of Division of Bioresources Research, Jeju Biodiversity Research Institute, Jeju Technopark

....

In-Kyu Yeo, PhD-Professor of Marine Life Sciences

....

Joon Bum Jeong, PhD-Professor of Marine Life Science

....

Jehee Lee, PhD-Professor of Marine Life Sciences

Date:........20.02.2014......

Department of Marine Life Sciences

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

Dedicated to beloved parents, my

ever loving wife and son

요약문

돌돔은 한국과 일본에서 중요한 양식어종으로 지난 10 년 동안 돌돔의 수요는 급격하게 증가하는데 반하여 그 생산량은 몇몇 이유로 인하여 정체되어 있는 실정이다. 돌돔양식산업은 질병에 매우 민감한데, 고밀도 사육과 밀식사육이 이루어지는 양식장에서는 매우 빈번하게 감염에 의한 질병이 발생한다. 어류의 면역체계에 대한 명확한 이해를 위해서는 어류의 면역 유전자원에 대한 확보가 매우 중요하고, 이러한 연구는 어류 질병의 효과적인 질병 제어 기술 개발에 기여하고 있다. 면역시스템은 몸에 이물질을 인식하고, 이를 제거하는 복잡한 네트워크이며, 척추동물에서 보체 시스템은 숙주에 침입하는 병원균의 파괴에 중요한 역할을 한다.

보체 구성요소들의 주요 분자의 기능에 관련된 분자적 특성과 발현에 대한연구는 면역체계에서 1 차적인 방어기작을 이해 할 수 있게 하므로 이러한 연구는 보체시스템에 대한 명확한 그림을 얻을 수 있고, 한국 돌돔 어류 양식의 감수성 증가와 질병을 방어하는 새로운 전략의 발전을 할 수 있을 것이다.

포유류의 보체 시스템은 보체의 직접적인 기능, 다른 신호전달경로의 활성화 등 많은 연구가 이루어져 있는데 반하여, 경골어류의 보체 시스템에 대해서는 연구가 부족한 실정이다. 따라서 돌돔의 전체 보체시스템의 연구는 경골어류의 보체 시스템을 위한 분자면역학에도 많은 도움을 줄 수 있을 것이다. 이 연구는 주요 보체 구성요소 (complement component 1 - 9) 그리고 sub unit 과 isoform 들 (전체 13 개 유전자)에 대한 genomic DNA 의 구조와 이들 유전자의 발현 수준을 확인하였다. 보체 구성성분 1q (C1Q), C2, C3, C4, C5, C6, C7, C8 (α, β, γ), C9 가 면역 자극 후 에 전사 (transcription) 수준에서 면역자극을 주지 않은 어류들과 어떠한 발현 차이를 나타내는지 비교 분석하였다. mRNA *in situ* hybridization 을 통하여 보체 C1q, C4, C5 가 조직내의 발현 패턴 및 분포를 확인하였다. 보체 C3, C4 그리고 C5 의 anaphylatoxin domain 부분을 cloning 하고 *E.coli* 발현 시스템을 이용하여 재조합 단백질을 overexpression 및 정체 하여 돌돔 head kidney primary cell 에서 respiratory burst activity 를 측정하였다. Web 기반의 prediction software 를 이용하여 보체 유전자의 promoter 에 존재하는 transcriptional factor binding site 를 확인하였다. 이 연구는 돌돔에 존재하는 보체 시스템에 대한 이해를 돕고, 이러한 보체 시스템에 대한 정보는 질병에 걸리지 않은 질 좋은 어류를 생산하고자 하는 돌돔 양식산업의 발전에 큰 도움을 줄 수 있을 것이다.

이 논문은 유전자의 기능에 따라 1 장에서는 initiative component 를, 2 장에서는 execution phase component, 마지막으로 3 장에서는 termination phase components 에 대한 내용으로 나누어 설명하고 있다. 1 장에서는 initiation component (Rb-C1q)의 3 가지 isoform 의 분자적 특성과 mRNA 발현 수준을 확인하였다. Glycoprotein 인 보체 C1q 는 보체시스템의 항체 의존적 classical pathway 를 개시한다. C1q A-like (RbC1qAL), C1q B-like (RbC1qBL), and C1q C-like (RbC1qCL)의 cDNA 는 각각 780, 720, 726 bp 의 염기서열로 이루어져 있으며, 260, 240, 242 개의 아미노산을 각각 암호화 하고 있다. 이들 3 개의 C1q 단백질의 아미노산 서열의 N-terminal 부위의 신호서열 (signal peptide)과 C-terminal 부위의 collagen-like region(s) (CLRs)를 가지고 있는 것을 확인 할 수 있었다. C1q 의 특징인 Gly-X-Y repeat 은 3 개 아미노산 서열 모두 가지고 있으며, 식세포 활성을 증가시키는 CLR-associated 서열은 RbC1qAL (^{49}GEKGEP^{54})과 와 RbC1qCL (^{70}GEKGEP^{75})의 두 개의 아미노산 서열에서만 확인 할 수 있었다. Rb1CqAL 과 RbC1qCL 은 6 개의 exon 으로 이루어진 데 반하여 RbC1qBL 은 단지 5 개의 exon 만으로 구성되어있다. 계통수는 3 개의 RbC1q cluster 는 어류 분류 군에 밀접하게 연관되어 있고, 3 개의 RbC1q 는 qPCR 결과 비장과 간에서 발현이 높았으며, 세균과 바이러스 등 면역자극을 주었을 때 간에서 초기에는 높은 발현이 나타났으며, 후기에는 lipopolysaccharide endotoxin 면역자극을 준 것이 높게 발현되었다. 이러한 결과를 바탕으로 통하여 RbC1q 가 세균 및 바이러스와 같은 감염에 대한 면역 반응의 시작에서 중요한 역할을 수행할 것으로 사료된다.

제 2 장에서는, 최종적으로 관여하는 보체와 핵심적인 기능을 수행하는 보체인 C3, 4, 5 의 분자적 특성 및 단백질의 기능 분석을 수행하였다. 보체 C3, C4, C5 의 cDNA 서열과 genomic DNA 구조 그리고 발현 차이를 비교하였다. C3 는 55,872 bp 의 genomic DNA (gDNA) 서열을 가지며, 총 43 개의 exon 이 42 개의 intron 에 의해 분리되어 있었다. ORF 서열은 1,663 개의 아미노산을 암호화하고, 아미노산 서열의 분자량은 187 kDa, 등전점은 5.8 을 가지고 있음을 확인하였다. 최종적으로 작용하는 보체인 Rb-C4 의 gDNA 16,362 bp 로 이루어져 있으며, 41 개의 exon 이 40 개의 intron 에 의하여 나뉘어져 있었다. ORF 부위는 1,710 개의 아미노산을 암호화 하고, 이는 191 kDa 의 분자량과 6.2 의 등전점을 나타내었다. Rb-C5 는 5061 bp 의 gDNA 로 이루어져 있으며 ORF 부위는 1,687 개의 아미노산을 암호화하며 187 kDa 의 분자량과 6.7 의 등전점을 가지고, 5'-UTR 은 29 bp, 3-UTR 은 249 bp 임을 확인 할 수 있었다. 이들 세 가지의 단백질 모두 신호서열을 포함하고 있었으며, 1,600 개 이상의 아미노산

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

서열과 8 개의 macroglobulin (MG) domain, linker domain (LNK), anaphylatoxin (ANA) domain, CUB domain, Thioester-containing domain (TED), 마지막으로 C-말단의 C345C domain 등 13 개 이상의 domain 을 가지고 있는 것을 확인 하였다. qPCR 을 통하여 Rb-C3, Rb-C4, Rb-C5 는 모든조직에서 발현이 관찰되었으며, 간조직에서 가장 많은 발현량을 나타내었다. 병원균인 *E. tarda, S. iniae* 와 박테리아 endotoxin 으로 잘 알려진 LPS, 돌돔 병원성 바이러스인 iridovirus 를 이용한 감염실험에서 head kidney 에서 Rb-C3 와 Rb-C5 의 mRNA 발현의 증가가 관찰되었고, Rb-C4 의 전사는 간조직과 비장에서 유의적으로 증가하는 결과를 나타내었다. 병원균 및 바이러스 감염 실험을 통하여 관찰된 유전자 발현 변화와 유전자 서열 유사도 및 계통분류학적 분석을 통하여 얻어진 결과들은 Rb-C3, C4, C5 가 기존에 밝혀진 포유류의 complement 들과 orthologous 임을 증명하고 있으며, 이는 돌돔의 보체시스템의 활성경로에 포함되는 면역적으로 중요한 유전자로서 역할을 할 것으로 예측할 수 있었다.

3 장에서는 보체시스템에서 lytic pathway 를 구성하는 C6, C7, C8 (α, β, γ), C9 유전자에 대한 gDNA 구조 및 분자적 특성을 서술하였다. C6, C7, C8, C9 는 single-chain glycoprotein 으로 세균 세포벽에 연속적으로 작용하여 세균을 용해시켜 방어하는 역할을 한다. C5 는 C5 convertase 에 의하여 C5a 와 C5b 로 나뉘어 지고 이때 C5b 는 다른 분자와 결합하여, 세균에 세포벽을 뚫을 수 있게 도움을 주는 하는 역할을 한다. cDNA library 를 구축하고 이를 이용하여 돌돔의 Rb-C6, C7, C8 (α, β, γ), C9 의 cDNA 서열을 확인하였고, 돌돔 gDNA 를 이용하여 bacterial artificial chromosome (BAC) library 를 구축하고 pools and super pools method 를 이용한 PCR screening 을 통해 genomic DNA 서열을 확인하였다. Lytic pathway 에 관여하는 components 중에서 Rb-C6 가 2,814 bp 로 가장 긴 ORF 서열을 가지고 있었으며, Rb-C8γ 가 가장 짧은 663 bp 의 ORF 를 가지고 있었다. 가장 긴 유전자 서열인 (gDNA sequence)는 18 개의 exon 과 17 개의 intron 으로 이루어진 Rb-C6 와 Rb-C7 이었고, Rb-C8γ 가 가장 짧았다. Rb-C8γ 만이 하나의 lipocalin domain 을 가지고 있었으며, 이를 제외한 Lytic pathway 에 관여하는 유전자 모두에서 thrombospondin type-1 (TSP-1) domain, low-density lipoprotein receptor domain class A (LDLR-A), membrane attack complex/perforin (MACPF) domain, epidermal growth factor like-1 (EGF-1) domain 이 확인되었다. Rb-C6 와 Rb-C7 은 2 개의 complement control protein (Sushi/CCP/SCR) domain 과 두 개의 factor I MAC module (FIMAC) domain 을 가지고 있었다. 모든 lytic pathway 관련 유전자의 promoter 부위에서 NF-kB, SP-1, C/EBP, AP-1, OCT-1 등과 같은 transcription factor 결합 부위를 포함하고 있음을 확인 할 수 있었다. 계통분류학적 분석과 아미노산 서열 비교분석을 통하여 Rb-C6, C7, C8a, C8B, Rb-

C9 은 Tetraodoniformes (puffer fish), Pleuronrctiformes (flounder), Salmoniformes (Rainbow trout), Cichlidae (Tilapia)와 진화적으로 가장 가까웠으며, 아미노산 서열 또한 가장 유사한 결과를 나타내었다. qPCR 을 통하여 모든 lytic pathway 에 관여하는 보체 유전자들이 모든 조직에서 constitutive 하게 발현되는 것을 확인 할 수 있었으며, Rb-C6 를 제외한 다른 유전자들은 간에서 가장 높은 발현량을 나타내었다. Rb-C6 는 다른 보체 component 들과 다르게 심장 세포에서 가장 높은 발현량을 보였으며 동시에 간에서도 상당히 높은 수준으로 발현되는 것을 확인 할 수 있었다.돌돔에 *E. tarda, S. iniae,* lipopolysaccharide endotoxin, rock bream iridovirus 등 을 주사하여 면역반응을 유도하였을 때 모든 lytic pathway 관련 보체 유전자들의 mRNA 는 head kidney 와 간조직에서 이들 면역 자극에 반응하여 유의 적인 발현변화를 나타내었다.

최종적으로 이 연구에서 수행한 돌돔 보체시스템에 관여하는 유전자들에 대한 종합적인 연구의 결과는 병원균의 침입에 노출되어 있는 양식 환경에서 어류의 생존을 위한 보체시스템의 역할을 이해하는데 매우 중요한 자료이며, 면역학적 관점에서 돌돔 뿐만이 아니라 경골어류들의 보체시스템의 대한 총체적인 이해를 위해서 중요한 자료로 사용 될 수 있을 것이다.

Summary

Rock bream (*Oplegnathus fasciatus*), is an important delicacy in Korea and Japan. The demand of the rock bream fish is tremendously increased during last decade, however production is stagnant in the same rate due to the few reasons. One of the bottlenecks in commercial aquaculture of rock bream is its susceptibility towards diseases (Zenke and Kim, 2008). The infectious disease outbreaks are very common in fish farms due to high stocking density and intensive culture techniques. The studies on immunological perspectives of fish and genetic background are important to gaining a detailed understanding of fish immune system and may immensely help in development of the effective disease control strategies. The immune system is a complex network of organs containing cells that recognize foreign substances in the body and eliminate them. In vertebrates, complement system plays a key role in destruction of the invading pathogens in the host. Understanding of molecular and expression-level functions of key molecules of the complement component is immensely help to evaluate the first line of defense mechanism. Thereby, researches will be able to obtained clear picture of whole complement system and will be able to develop novel strategies to prevent diseases and improve the sustainability of rock bream fish farming in Korea.

In mammalians, complement system is well studied in relation to the activations, direct functional characteristics, activation of other signaling pathways, drawbacks of the needless activations etc. However, it has been poorly studied in teleost fish or identified haphazardly in several fish species. Therefore, study of whole complement system in one of the teleost fish species is a challenging task for molecular immunologists. In this study, major complement components (complement component 1 to 9) and some of their sub units, isoforms (total of thirteen genes) have identified at genomic structural and transcriptional level. The complement component 1 (C1q), C2, C3, C4, C5, C6, C7, C8 (α, β and γ) and C9

were characterized at transcriptional levels upon immune challenges. The expression levels of healthy and immune-stimulated fish were compared. Furthermore, the mRNA *in situ* hybridization technique was used to confirm the expression pattern in complement C1q, C3 and C7. The protein level functions were tested in complement C3, C4 and C5 using their anaphylatoxin domains with the respiratory burst activity test on rock bream head kidney primary cells. The putative promoter regions of most of the genes were characterized using web based prediction software. Thereby, identified transcriptional factor binding sites were confirmed. This comprehensive study may lead to understanding complement system of rock bream fish and in broader sense, it could help in improving aquaculture and aid in the development of disease free, quality enriched fish.

This report is divided in to three chapters based on the features of the genes such as initiative components, execution phase components and termination phase components. In the first part of this study, three isoforms of complement initiation component (Rb-C1q) were characterized at genomic and transcriptional level. The complement C1q is the key glycoprotein that initiate antibody dependent classical pathway of the complement system. There isoforms were identified in rock breams which are resembled to human C1q. The full-length cDNAs of C1q A-like (RbC1qAL), C1q B-like (RbC1qBL), and C1q C-like (RbC1qCL) were consisted of 780, 720 and 726 bp of nucleotide sequence encoding polypeptides of 260, 240 and 242 amino acids, respectively. All three RbC1qs were possessed a leading signal peptide and collagen-like region(s) (CLRs) in the N-terminus, and a C1q domain at the C-terminus. The C1q characteristic Gly-X-Y repeats were presented in all three RbC1qs, while the CLR-associated sequence that enhances phagocytic activity was presented in RbC1qAL (^{49}GEKGEP^{54}) and RbC1qCL (^{70}GEKGEP^{75}). Moreover, the coding region was distributed across six exons in RbC1qAL and RbC1qCL, but only five exons were identified in RbC1qBL. Phylogenetic analysis revealed that the three RbC1qs were tightly

cluster with the fish clade. All three RbC1qs were most highly expressed in the spleen and liver, as indicated by quantitative real-time PCR (qPCR) tissue profiling. In addition, all three were transcriptionally responsive to immune challenges, with liver expression being significantly up-regulated in the early phase of infection with intact, live bacteria (*Edwardsiella tarda* and *Streptococcus iniae*) and virus (rock bream iridovirus) and in the late phase of exposure to purified endotoxin (lipopolysaccharide). These data collectively suggest that the RbC1qs may play defense roles as an innate immune response to protect the rock bream from bacterial and viral infections.

In the second chapter (Chapter II) of the study, molecular characteristics and functional properties of complement component 2, 3, 4, and 5 were described in-depth. The chapter II consists of two parts namely Part 1 and Part 2. In the part 1 of chapter II, genomic structural and transcriptional modulation of complement C2 was discussed. Furthermore, functional characterization of 5' flanking region of Complement C2 was accomplished by luciferase reporter assay. The complement C2 is one of the key molecules which make C3 convertase (C4b2b) in complement activation via classical pathway. In the present study, complement component 2 was identified and characterized at transcriptional level in rock bream (*Oplegnathus fasciatus*). The full-length cDNA of rock bream C2 (*Rb-C2*) was identified in the cDNA library and its genomic sequence was also obtained by screening and sequencing of rock bream BAC library. The complete genomic DNA of rock bream C2 is 7270 bp, consists of 19 exons interrupted by 18 introns. An open reading frame (ORF) is encoding a polypeptide of 779 amino acids. The predicted molecular weight of the *Rb-C2* polypeptide is 87 KDa and iso-electric point is 6.2. The *Rb-C2* possesses three of Sushi/CCP/SCR domains, a von Willebrand Factor domain (vWF), and a trypsin family domain, known to be important for the functions of C2 in vertebrates. In the promoter region, the luciferase reporter assay confirmed that the transcriptional factors such as SP-1, AP-1,

OCT and C/EBP are positively regulates the transcription of Rb-C2. The amino acid sequence of *Rb-C2* showed 69.7% and 59.2% identity to *Oreochromis niloticus* and *Oncorhynchus mykiss* C2, respectively. In tissue expression profile, *Rb-C2* transcripts were constitutively expressed in all the tissues whilst highest was observed in liver. In challenge experiments, *Rb-C2* transcripts were significantly up-regulated in liver and spleen tissues post challenge with *E. tarda*, *S. iniae*, lipopolysaccharide endotoxin and rock bream iridovirus.

In part 2 of the chapter II, the largest and key molecules belong to complement system were discussed in details and functional properties of their proteins were also characterized. Full-length cDNA sequences of complement component 3, 4 and 5 and their genomic structures, expression differences were compared. The complete genomic DNA of rock bream C3 is 55872 bp, consists of 43 exons interrupted by 42 introns. An open reading frame (ORF) is encoding a polypeptide of 1663 amino acids. The predicted molecular weight of the Rb-C3 peptide is 187 kDa and iso-electric point is 5.8. Interestingly, the largest complement protein found in the rock bream, Rb-C4 has a complete genomic of 16362 bp, consists of 41 exons interrupted by 40 introns. An ORF is encoding a polypeptide of 1710 amino acids. The predicted molecular weight of the Rb-C4 peptide is 191 kDa and iso-electric point is 6.2. Meanwhile, Rb-C5 has an ORF of 5061 bp encoding for a putative protein of 1687 amino acids of molecular mass 187 kDa and isoelectric point of 6.7 and 5′ UTR of 29 bp and 3′ UTR of 249 bp. Sequence analysis revealed that, all three peptide contain signal peptides, comprised of over 1600 amino acids arrange as 13 distinguished domains namely; 8 macroglobulin (MG) domains, linker domain (LNK), anaphylatoxin (ANA) domain, CUB domain, thioester-containing domain (TED) and carboxy-terminal C345C domain. The qPCR analysis confirmed that Rb-C3, Rb-C4 and Rb-C5 were constitutively expressed in all the examined tissues isolated from healthy rock bream, with highest expression occurring in liver. The pathogen challenge, including *E. tarda, S. iniae,* lipopolysaccharide endotoxin and rock

bream iridovirus led to up-regulation of *Rb-C3* and *Rb-C5* in head kidney. While, *Rb-C4* transcripts were significantly up-regulated in liver and spleen tissues post challenge with *E. tarda*, and lipopolysaccharide endotoxin. The observed response to bacterial and viral challenges and high degree of evolutionary relationship to respective orthologous, confirmed that *Rb-C3, Rb-C4* and *Rb-C5* are important immune genes, likely involved in activation pathway of the complement system of rock bream. Furthermore, all three complement components have the ability to release a short peptide chains upon activation to initiate inflammation in the acute phase, which occurs when pathogen breaches the first line of defense, and the immune system sends signaling molecules and defense cells to the site of infection to defend against foreign invasions. Collectively, findings of this chapter support to suggest that *Rb-C2, C3, C4* and *C5* may play significant role in host defense through the activation of effector mechanism of the complement system in rock bream.

Third chapter describes the genomic structural and molecular characterization of all lytic pathway components that have been identified in the rock bream. The complement component, 6, 7, 8 (α, β and γ) and 9 are single-chain glycoproteins that belong to the lytic pathway that sequentially arranged on the bacterial cell wall. This is initiated upon cleavage of C5 in to C5a and C5b by the C5 convertases and preceded as the other components bind to C5b. The full-length cDNA sequences were identified from a rock bream cDNA library, and their genomic sequences were obtained by screening and sequencing of rock bream bacterial artificial chromosome (BAC) genomic DNA library. The *Rb-C6* is the largest ORF found among the lytic pathway components (2814 bp) and the smallest is the *Rb-C8γ* (663bp). The largest genomic sequences (18 exons 17 introns) were identified in *Rb-C6* and *Rb-C7* orthologous whereas smallest was found in *Rb-C8γ* among the lytic pathway genes. When considering the domain architecture of the lytic pathway genes, all of them have common domains located in the polypeptide chains except *Rb-C8γ* which contains one lipocalin

domain. All others contain thrombospondin type-1 (TSP-1) domain, a low-density lipoprotein receptor domain class A (LDLR-A), membrane attack complex/perforin (MACPF) domain and epidermal growth factor-like-1 (EGF-1) domain. The domain structures of *Rb-C6* and *Rb-C7* contain two complement control protein (Sushi/CCP/SCR) domains and two factor I MAC module (FIMAC) domains. The promoter regions of all genes contain important putative transcription factor binding sites including those for NF-κB, SP-1, C/EBP, AP-1, and OCT-1. According to the phylogenic and pairwise comparison, *Rb-C6, Rb-C7, Rb-C8α, Rb-C8β* and *Rb-C9* have highest identity to fish from Order Tetraodoniformes (puffer fish), Pleuronrctiformes (flounder), Salmoniformes (rainbow trout) and Cichlidae (Tilapia). The qPCR analysis confirmed that all the lytic pathway genes were constitutively expressed in all examined tissues, isolated from healthy rock bream, and highest expression occurred in liver except *Rb-C6*. The *Rb-C6* was found to be highly expressed in heart cells, which is an exception to all other complements. Meantime, considerably high expression of *Rb-C6* was also observed in liver. In immune-challenged rock bream, all the lytic genes were responded to *E. tarda, S. iniae,* lipopolysaccharide endotoxin and rock bream iridovirus in head kidney and liver tissues. Taken together, the molecular characteristics, orthologous relationships and transcriptional response to pathogenic stimulants suggest that identified genes are belonging to lytic pathway genes of the complement system that involved in immune processes of rock bream.

In conclusion, the present approach in comprehensive study of complement system in rock bream fish *Oplegnathus fasciatus* will be important to understand its role in relation to survival of fish from challenging environment with abundant pathogenic threats and make significant insight into the immunological perspectives of the complement system in teleost.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 General introduction

The overall purpose of this study was to explore the functional existence of complement system in rock bream fish, *Oplegnathus fasciatus*. The complement system comprising of 35 distinct proteins is one of the main component of the innate immune system. The direct functional roles and activation, activation of downstream signaling pathways, drawbacks of the needless activations of this system have been extensively investigated in mammals. However, the prevalence, function and regulation of this system is relatively less demonstrated or identified haphazardly in several fish species. The discovery and functional characterization of the various components of the complement system within a single teleost fish species is a challenging task for molecular immunologists. In this study, with the advances made in transcriptomics and genomics, nine major complement components (Complement component 1-9) and some of their sub units, isoforms (total of thirteen genes) were identified and characterized at genomic/transcriptional level and functions of few have been elucidated. Understanding complement system as a part of the innate immune system in fish could help us in improving aquaculture and aid in the development of disease free, quality enriched fish. The primary objective of the present study was stated at the end of introduction part.

1.2 Perspectives in world aquaculture

The survival of each life form on the earth is competitive and challenging. The sources of food play a vital role in existence of every species on earth. Primarily, the plants are the only organisms that can produce their own food via the photosynthesis process by utilizing the solar energy, form the sun as the raw material. Humans, being the omnivorous and superior species in the world still find it difficult to satisfy their food needs. Hence, to fulfill the

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

unlimited needs of humans from the limited resources, humans have derived every means to find food sources to cater the demand of the population. In spite of the continuous efforts, one billion people have been reported to suffer from hunger (FAO report, 2012). The fisheries and aquaculture make crucial contributions to fulfill the food basket and in improvising the social status exclusively contributing to the prosperity of the society. Fish production contains two major arms, capture fishery and aquaculture. Aquaculture is expanding with development of new inventions, while capture fishery remain stable. Aquaculture industry can be considered as the most important and fast growing industry which will exceed that of other livestock industries and serves as the excellent sources of animal proteins within the next decade. In the last three decades (1980–2010), world fish production through aquaculture has expanded by almost 12 times, at an average annual rate of 8.8 percent (FAO report, 2012). The total production of the aquaculture has reached 60 million tons with an estimated total value of US$119 billion in 2010. However, the vital contributions from fisheries and aquaculture to global food security and economic growth remain constrained by an array of problems. One of the bottlenecks of the aquaculture is vulnerability to the adverse impacts of disease and environmental conditions. Disease outbreaks in the important aquatic industries such as Atlantic salmon in Chile, oysters in Europe, and marine shrimp farms in several countries in Asia in recent years, resulted in partial or sometimes total loss of production (FAO report, 2012). Like in any other farming, aquaculture is also overwhelmed with diseases. The protection of fish farming in particular from pathogenic microorganisms like bacteria, virus and parasites remains a challenging task (Kusuda and Salati, 1993; Toranzo et al., 2005).

1.3 Fish immunity

The vertebrate immune system consists of two arms namely specific adaptive and non-specific innate system. Evolutionarily, fish being the preliminary vertebrates possesses both the arms of immunity. Adaptive immunity first appeared about 450 million years ago in cartilaginous and bony fishes. It plays an important role in the protection against recurrent infections by generating 'memory' cells and specific soluble- and membrane-bound receptors, such as immunoglobulins (Ig) and T cell receptors (TCR), which allow for the fast and efficient elimination of the specific pathogens. The development of vaccines as one of the means of controlling infection relies on the principle of adaptive immunity, and vaccination technologies have revolutionized the fish farming industry, as they previously have domestic animal and human medicine. The major difference of the fish system with other vertebrates is that their metabolism and immune responses are temperature dependent. Therefore, development of vaccines for fish diseases was recognized as a primary requirement of disease control in aquaculture. Vaccine development in fish culture has reached a new era due to innovation of novel biotechnological tools and use of modern molecular biological methods such as gene silencing, RNA interference, micro RNA, and microarray. Identification and characterization of immune related genes is one of the fundamental requirements to understand the immune system and their role in the facilitation of functions of vaccine delivery.

1.3.1 Innate immunity

Innate immunity is an evolutionarily conserved, non-specific, first line of defense genetically inherited and not learned, adapted or permanently heightened because of exposure to microorganisms or vaccination. The innate immune system comprises of the cells and mechanisms that defend the host during microbe exposure and initial adaptive responses. The

3

innate immune system recognizes microbes through pattern recognition receptors (PRRs), which are receptors specific for molecular components of microorganisms that are not relevant for the host. Some innate immune responses are temporarily up regulated as a result of exposure to microbes, but the components of the innate immune system do not change permanently during an individual's lifetime. Unlike the adaptive immune system, it does not confer long-lasting or protective immunity to the host (Fearon and Locksley, 1996). Innate immune system is present in all classes of life ranging from microorganisms to well-developed mammals.

1.3.2 Adaptive immunity

The adaptive immune system is composed of highly specialized, systemic cells and processes that eliminate or prevent pathogenic growth. Thought to have arisen in the first jawed vertebrates, the adaptive or "specific" immune system is activated by the "non-specific" and evolutionarily older innate immune system (which is the major system of host defense against pathogens in nearly all other living organisms). The adaptive immune response provides the vertebrate immune system with the ability to recognize and remember specific pathogens (to generate immunity), and to mount stronger attacks each time when the pathogen is encountered. It is adaptive immunity because the body's immune system prepares itself for future challenges.

1.4 The complement system

The complement system was originally discovered in 1890s when it was found to aid or "complement" the killing of bacteria by heat-stable antibodies present in normal serum (Walport, 2001). The complement system is composed of more than 30 proteins, which are found as soluble proteins circulating in blood or as membrane-associated proteins. These

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

proteins play key roles in complement-mediated killing of pathogens through the lytic pathway (Gal et al., 2007; Korotaevskiy et al., 2009). The complement system exerts humoral effects in innate immunity and plays a crucial role in recognition, tagging and clearance of invading microorganisms (Nakao et al., 2011). It acts as a bridge in between innate and adaptive immune system. The activation of the complement system leads to a sequential cascade of enzymatic reactions resulting in the formation of the potent anaphylatoxins C3a and C5a that elicit a plethora of physiological responses that range from chemoattraction to apoptosis. Three distinct pathways are responsible for target recognition and formation of a protease complex for C3 activation: the classical, lectin and alternative pathways converge into a final common pathway, known as the cytolytic pathway. This cytolytic pathway is responsible for formation of the membrane attack complex (MAC) that ultimately causes lysis of the targeted foreign cell (Biesecker and Muller-Eberhard, 1980). The complement proteins exist as inactive zymogens under normal circumstances, which when sequentially cleaved by proteases becomes activated. On the other hand, some components of the system serve as the substrate of an earlier component of the cascade and then as an enzyme to activate a subsequent component. As a whole, there are four main functions of the complement system; lysis of microorganisms, promotion of phagocytosis, triggering inflammation and immune clearance.

The complement system itself has been highly conserved throughout evolution, and is present in both vertebrates and invertebrates. Its ancient origin is believed to trace back to Cnidaria, one of the most primitive metazoans (Dishaw et al., 2005). Furthermore, the complement system has been found in some protostomes, such as arthropods (Zhu et al., 2005) and mollusks (Castillo et al., 2009), and its anti-microbial function has been demonstrated in horseshoe crabs (Ariki et al., 2008), sea urchins (Smith et al., 1999) and ascidians (Nonaka and Azumi, 1999). In the Cyclostomatida primitive vertebrates, the

5

complement system is more similar to the invertebrates than that of mammals, based on the apparent lack of an antibody-recognizing activation cascade (the classical pathway-like) and a cytolytic pathway (Kimura et al., 2009). However, the complement system in fish appears to be fully equipped with all three of the C3-activation pathways and the cytolytic pathway, both of which exert many of the effector activities recognized in mammalian complement, such as targeted cell killing, opsonization, and anaphylatoxic leukocyte stimulation (Boshra et al., 2006).

Table 1. Functional protein classes in the complement system

Category/Pathway	Inactive form	Active form	Function
Classical pathway	C1q, C1r and C1s	C1 complex	Binding to the antigen-antibody complexes on the pathogen surfaces
	C2	C2b, C2a	Activating enzymes
Lectin pathway	Mannan-binding lectin (MBL)	MBL complex	Binds to the mannose on bacterial surface
	MASP-1 and MASP-2		
Alternative pathway	C3, Factor B, Factor D	Factor Bb, factor D,	Act as C3 convertase
Key activating molecules	C4 and C3	C4b, C3b	Membrane binding protein and opsonins
	C4, C3, and C5	C4a, C3a and C5a	Peptide mediators of inflammation
Lytic pathway	C5, C6, C7, C8 and C9	C5b, C6, C7, C8 (α, β and γ) and C9	Membrane attack- proteins
Regulators of complement system	CR1, CR2, CR3, CR4 and C1qR	CR1, CR2, CR3, CR4 and C1qR	Receptors for complement factors
	C1 inhibitor (C1INH) C4bp, CR1, MCP, DAF, Factor H, Factor I, Factor P, CD59	C1 inhibitor (C1INH) C4bp, CR1, MCP, DAF, Factor H, Factor I, Factor P, CD59	Complement-regulatory proteins
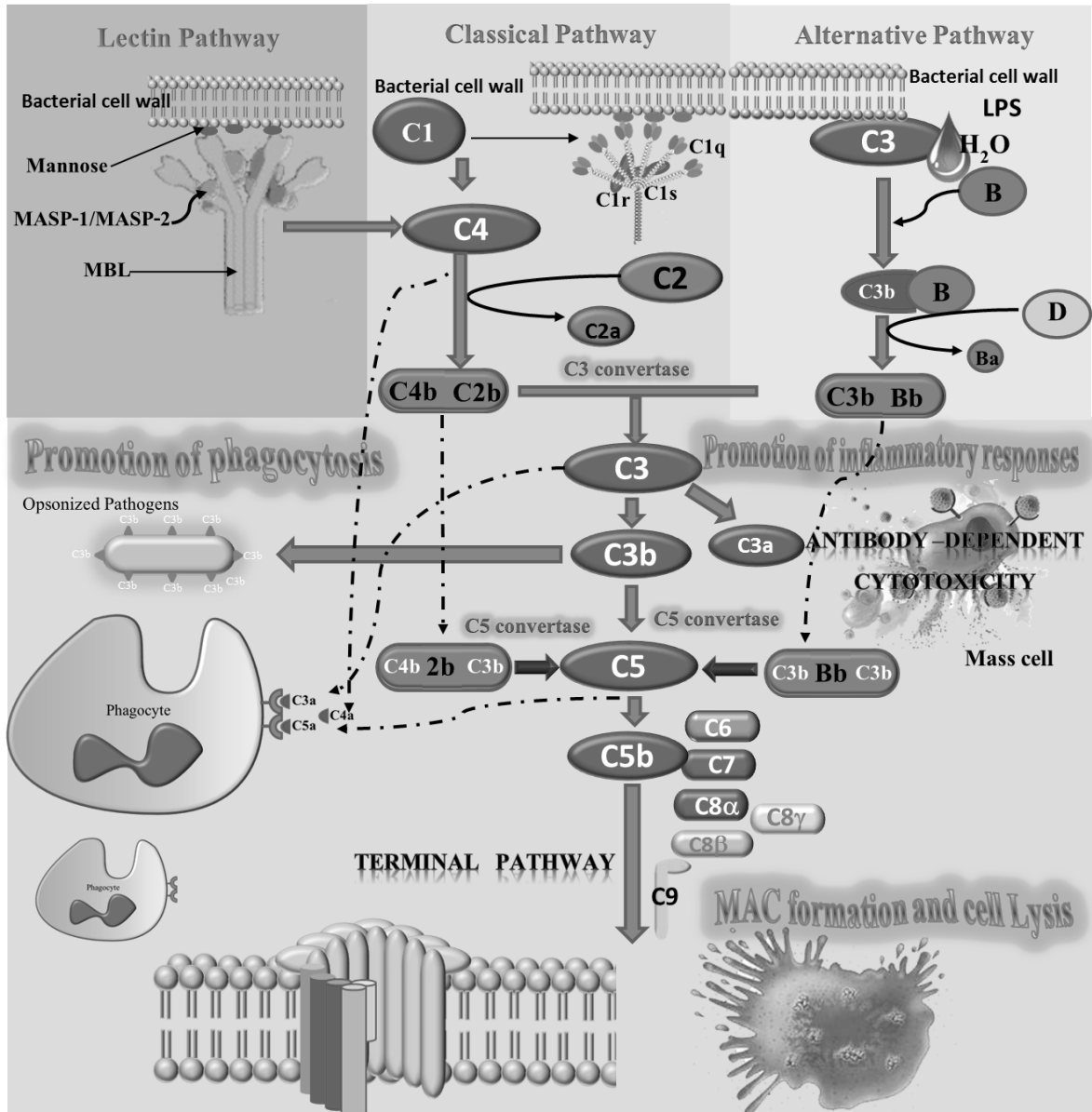
Figure 1. Schematic representation of complement system

1.4.1 Classical pathway

A wide variety of pathogenic microorganisms, including a number of bacteria and viruses, as well as many infected cells, efficiently activates the classical pathway after their recognition by antibodies. The primarily classical pathway is initiated by the C1 complex, which is circulating in the blood serum. The C1 complex consists of six molecules of C1q, two molecules of C1r and C1s. Activation is triggered upon interaction of the serum C1 complex

with antigen-antibody complexes or immune aggregates containing immunoglobulin G (IgG) or IgM (Wimmers et al., 2003). The C1 binding to immunoglobulins is mediated by its C1q subunit and involves the heterotrimeric C terminal globular 'heads' of the protein. Subsequently, C1s and C1r get activated and produce an enzyme that cleaves two complement components. First, it cleaves C4 into C4a and C4b, and the resulting C4b fragment binds covalently to -NH$_2$ or -OH groups on the protein in the immediate vicinity of the C1 molecule at the surface of the activator. Then bound C4b binds C2 in availability of magnesium. After that, C2 cleaved into C2a (small sub unit) and the glycoprotein C2b (large sub units) which binds noncovalently to a site on C4b to give the C4b2b complex, `the C3 convertase with a serine active site in C2b and exhibiting specific proteolytic activity towards C3. This complex protease is responsible for initiation of whole sets of biological activities of complement system (Arlaud and Colomb, 2001).

1.4.2 Lectin pathway

In addition to the well-characterized C1-mediated initiation of the classical pathway, a second activation route, the 'lectin pathway', was discovered in the early 1980s. It is triggered by a C1-like complex protease in which the recognition function is mediated by mannan-binding lectin (MBL), a member of the collectin family (Malhotra et al., 1994). Unlike other collectins, MBL is able to mediate the opsonization of microorganism through its multiple C-terminal C-type lectin modules. It has been reported that, MBL associated serine protease (MASP) harbors the proteolytic activity against C4 and C2 and activation of C3 also was reported at a low rate (Matsushita and Fujita, 1995). The MASP is a mixture of protease of MASP1 and MASP2, where MASP2 is involved in the cleavage of C4 (Rossi et al., 2001; Wong et al., 1999). MBL identifies ranges of oligosaccharides present in the surface of wide spectrum of microorganisms, including yeasts (*Candida albicans*), viruses (HIV, influenza

8

A), and a number of bacteria (*Salmonella montevideo*, *Listeria monocytogenes*, *Haemophilus influenzae* and *Neisseria meningitidis*) in human (Arlaud and Colomb, 2001).

1.4.3 Alternative pathway

The activation of the complement C3 involves C3, Factor B coupled with Factor D (adipsin), and Factor P (Properdin). The activation is initiated by the spontaneous hydrolysis of C3, which is abundant in the plasma activated by a H_2O molecule and reacting with factor B in the presence of Mg^{2+} ions to form the $C3(H_2O)B$ molecule. Then Factor D splits the bound Factor B into Bb and Ba, to form the $C3(H_2O)Bb$ enzymatic molecule. Afterward, properdin binds with the C3(H2O)Bb and stabilized the enzymatic molecular complex which subsequently break down other C3 molecule resulting more C3a and C3b molecules.

1.4.4 Execution mechanisms

The execution of the complement cascade is performed by the two key components namely C3 and C5. Complement protein C3 is a central mediator of the complement system whose activation is essential for all the important functions of this system. The complement C3 is acting like double-edged sword within the serum. It promotes inflammatory response against invading pathogens and on the other hand it facilitates or leads adoptive immune responses via proper antigenic activities (Sahu and Lambris, 2001).

Proteolytic cleavage of C4, C3 and C5 results in two subunits (a and b) from each component. All the "a" components (C4a, C3a and C5a) are called anaphylatoxins. They are small polypeptides released upon activation of the each molecule. These peptides bind to the receptors (CR1 and CR3) of granulocytes such as mast cells and basophils and trigger the release of vasoactives from intracellular granules and chemotaxis (Arlaud and Colomb, 2001). The resulting C3b molecule does opsonization of pathogen, which enhances the process of phagocytosis by the target phagocytes. Hence, it helps in elimination or clearance

9

of the immune complex from the blood stream and facilitates regular function without any needless damage (Nakao et al., 2011).

1.4.5 The lytic / cytolytic / termination pathway (Formation of Membrane Attack Complex)

One of the important effects of complement activation is the assembly of the terminal components of the complement to form a Membrane Attack Complex (MAC). The members of the MAC called lytic pathway genes belong to the membrane attack complex and perforin (MACPF) super family, which is one of the largest families of pore-forming molecules in the complement system. Lytic pathway includes the complement components C6, C7, C8 and C9. These are known as Terminal Complement Components (TCC) and are considered as key components of the cytolytic pathway. C6, C7, C8α, C8β, and C9 exhibit a high degree of sequence similarity, and all possess structural motifs for, Thrombospondin (TS), Low Density Lipoprotein Receptor (LDL-R), and Epidermal Growth Factor Precursor (EGFP) domains. Although the TCCs are structurally related, they differ in size and complexity. The MAC is assembled by sequential binding of C5b with C6, C7, and C8, along with incorporation of a variant number (6-16) of C9 molecules into the nascent complex. A study of carp serum determined that the molar ratios of isolated complement components C5b, C6, C7, C8, and C9 were 1:1:1:1:4, respectively.

Activation of the lytic pathway occurs when C5b associates with the target cell's membrane and undergoes conformational changes that expose its C6 binding site. The C5b-6 complex is then capable of binding to C7. Upon C7 binding to the C5b-6 dimer, the membrane binding sites are made available, which helps to stably incorporate the complex into the target membrane (Wurzner, 2000). Formation of the C5b-7 complex (known as the intermediate complex) can facilitate further interactions with other proximal cells, initiating

the process of reactive lysis (Thompson and Lachmann, 1970; Wurzner, 2000). Availability of C7 acts as the rate-limiting step of the membrane attack process, since unbound C5b-6 dissociates from the membrane without activating the cytolytic signal (Gonzalez et al., 2003). The C5b-7 complex binds to C8 and C9 to complete C5b-9 complex, which is MAC (Figure 2 (Corallini et al., 2009).
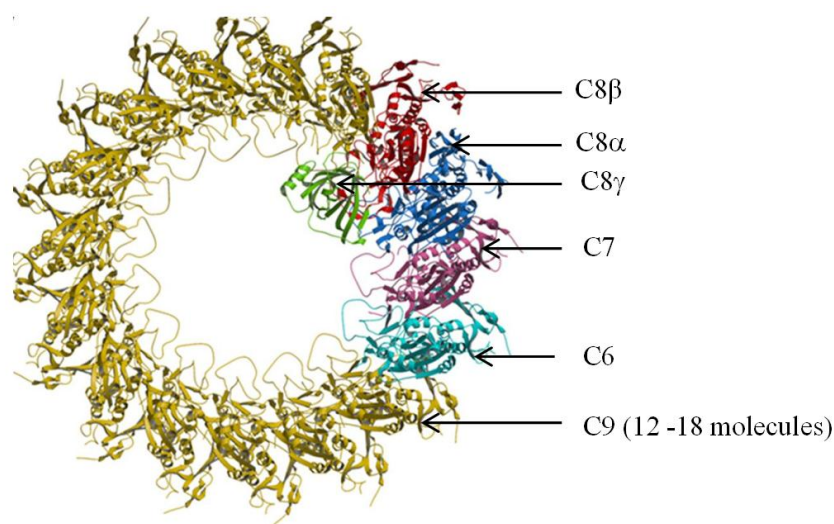
Figure 2. Crystallographic structural arrangement of complement lytic pathway genes in 'membrane pore' on the surface of bacterial cell wall

1.5 Complement control proteins and interrelated signaling

The complement system is one of the biochemical pathways, which needs high level of homeostasis. Deregulation of the complement system would be extremely damage to the host. All three pathways generate variants of C3 convertase, which cleaves C3 into C3a and C3b. These subcomponents trigger a series of further cleavage events that activates the complement cascade. This leads to immune defense responses such as degranulation of mast cells, increasing vascular permeability, smooth muscle contraction and initiation of the membrane attack pathway. However, activating molecules are tightly regulated by complement control proteins in each state. These regulatory proteins are found in much higher concentrations that complement proteins themselves prevent complement system

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

activation in 'self' tissues. In each level, there are proteins that pay regulatory role to avoid the host tissue damage due to needless or excessive activation. The excessive activation of these anaphylatoxin is controlled by the carboxypeptidase N which removes C-terminal Arg from each molecule and hence the degranulation and leucocyte chemo-attraction are terminated (Nakao et al., 2011). The complement system has been implicated in many autoimmune disorders, including systemic lupus erythematosus, multiple sclerosis and arthritis, and more recently has been suggested to have a pathophysiological role in Alzheimer's and other neurodegenerative disorders in mammals.

Table 2. Complement regulatory proteins

Name	Role in the regulation of complement activation
C1 inhibitor (C1INH)	Binds to activated C1r, C1s, removing it from C1q
C4-binding protein (C4BP)	Binds C4b displacing C2b:corfactor for C4b cleavage by I
Complement Receptor 1 (CR1)	Bind C4b, displacing C2b or C3b displacing Bb; cofactor for I
Factor H (H)	Binds C3b, displacing Bb; cofactor for I
Factor I (I)	Serine protease that cleaves C3b and C4b; aided by H, MCP, C4BP or CR1
Decay-accelerating factor (DAF)	Membrane protein that promotes C3b and inactivation by I
CD59 (protein)	Prevents formation of membrane-attack complex on autologous or allogeneic cells widely expressed on membranes

The complement system and innate immunity: http://www.ncbi.nlm.nih.gov/books/NBK27100/

1.6 Historical aspects and present status of the teleost complement system

The complement system was first identified as a heat-sensitive factor in fresh serum that 'complemented' or assisted the effects of specific antibody in the lysis of bacteria and red blood cells (Bordet and Gengou, 1901). First evidence suggested by Nuttal 1888; Buchner, 1889 and Ehrlich, 1899 was that the heat liable serum factor was required for bacterial activity of immune serum. The involvement of 11 unique plasma proteins in antibody mediated complement activation was discovered with the development of protein purification

techniques (Muller-Eberhard, 1975; Muller-Eberhard and Nilsson, 1960; Nelson et al., 1966) In 1970's detailed structural, functional and biosynthetic studies on the various components of the complement were identified in human (Paramaswara, 1998). In teleost, first attempt was to identify all the components as in mammals. The identification was accelerated by the novel molecular biological techniques such as homology cloning and use of genomic and transcriptomic data availability for several species (Nakao et al., 2011). According to the past studies, there is an evidence for the presence of the almost all mammalian components in the fish. Furthermore, these findings have led to the determination of the similarity among mammalian and teleost complement systems (Boshra et al., 2006; Nakao et al., 2006). The teleostan complement system was well characterized in common carp (Gonzalez et al., 2007a) rainbow trout and number of complement components were identified including C1r/C1s, C3, C4, C5, Bf/C2, Factor I and MASP. Some studies have shown that the CS components and functions of mammals are similar to those in rainbow trout (a bony fish) (Nonaka et al., 1981) and the nurse shark (a cartilaginous fish) (Jensen et al., 1981).

1.7 Selected species of the study

Rock bream or barred knifejaw or the striped beakperch *(Oplegnathus fasciatus)* belongs to the class Actinopterygii (Ray-fin fishes) and order Perciformes (perch-like). They are benthopelagic, inhibited in the coastal rocky reefs and associated waters with depth range from 1-10 meters. Rock bream is one of the most economically important fish species of Korean aquaculture. In recent years, significant losses in rock bream fish production have been recorded due to diseases outbreaks and researchers are directing their efforts towards understanding the underlying pathogenic mechanisms and immune responses to create a healthier and sustainable rock bream fish industry.

Figure 3. Rock bream fish *Oplegnathus fasciatus*

1.8 Objectives of the study

The primary objective of the present study was to explore molecular evidences for the existence of major complement components in one of the teleost counterparts, rock bream (*Oplegnathus fasciatus*). This broad research objective was accomplished via following areas.

- Molecular characterization of the genes involved in activation of complement system

- Genomic structural characterization of complement component C1 to C9

- Promoter level characterization of selected genes using luciferase reporter assay

- Transcriptional expression analysis of complement component C1 to C9 induced upon immune-stimulant challenges

- *In situ* hybridization of anti-sense RNA in selected complement components

- Biological activities of the proteins, providing insights into the function of the genes

2. Materials and Methods

2.1 Experimental animal

Healthy rock breams (∼50 g) were obtained from the Ocean and Fisheries Research Institute, Republic of Korea. They were acclimatized to laboratory conditions (salinity: 34±1‰, temperature: 24±1°C) in 3000 L tanks for one week prior to use them for any experimentation. Diet was comprised of a commercial feed with the daily ration equaling 3% (w/w) of body weight. To minimize stress factors, a maximum of 40 animals per 400 L were housed in each tank and the laboratory environmental conditions were uniformly maintained during the experiment.

2.2 Collection of rock bream tissue and peripheral blood cells

First, three healthy fish were taken and disinfected with 70% ethanol before isolation of tissues. Then, blood samples (0.75-1.0 mL/fish) were collected from the caudal vein (at 2.5 cm from the caudal fin) using a 1mL syringe containing heparin and immediately centrifuged at 3000 ×g (4°C) for 10 min to harvest the peripheral blood cells. After that, fish were dissected for collection of gill, liver, heart, spleen, intestine, head kidney, kidney, skin, muscle, and brain tissues. The isolated tissue samples were snap-frozen in liquid nitrogen and stored at -80°C until used for RNA extraction.

2.3 Isolation of the cDNA sequences of complement genes

A rock bream cDNA sequence database was established using the Roche 454 genome sequencer-FLX system (GS-FLX™), a next generation DNA sequencing (NGS) technology (DNA Link, Republic of Korea). Healthy rock bream fish with average weight of ~50 g, obtained from the Ocean and fisheries Research institute (Jeju, Republic of Korea) were

15

adapted to the laboratory conditions (salinity 34 ± 1‰, pH 8 ± 0.5 at 24 ± 1 °C) in 400 L tanks. Blood samples were harvested from the caudal vein of healthy, unchallenged fish using a 22-gauge needle and centrifuged immediately for 10 min at 3000 × g at 4 °C, to collect the hematic cells. Gill, liver, whole brain, kidney, head kidney, spleen, intestine, muscle, heart and skin tissues were harvested on ice from three healthy animals and immediately flash-frozen in liquid nitrogen and stored in -80 °C, until RNA extraction. Total RNA was isolated from individual tissues (pituitary gland, brain, gill, blood, liver, spleen, head kidney, and kidney) pooled from three healthy rock breams using the Tri Reagent™ (Sigma, USA) and following the manufacturer's instructions. The concentration and purity of RNA were evaluated using a UV-spectrophotometer (BioRad, USA) at 260 and 280 nm. Then, the polyadenylated mRNA was purified by means of an mRNA isolation kit (FastTrack® 2.0; Invitrogen, USA). First strand cDNA was synthesized from 1.5 μg of mRNA using Creator™ SMART™ cDNA library construction kit (Clontech, USA); amplification was performed with Advantage 2 Polymerase Mix (Clontech) under conditions of 95 °C for 7 s, 66 °C for 30 min and 72 °C for 6 min. Over-representation of the most commonly expressed transcripts was excluded by normalizing the synthesized cDNA using Trimmer-Direct cDNA normalization kit (Evrogen, Russia). Finally, a purified cDNA sample was sequenced using GS-FLX shotgun sequencing technology and library was created from the sequenced data (DNA Link, Republic of Korea). The candidate cDNA sequences with the highest homology to known complement components were identified by BLAST-X and BLAST-P search of the NCBI database (Altschul et al., 1990).

2.4 Construction of bacterial artificial chromosome library (BAC)

Rock bream obtained from the Jeju Special Self-Governing Province Ocean and Fisheries Research Institute (Jeju, Republic of Korea) were accustomed to the laboratory conditions.

Blood was harvested aseptically from the caudal vein using a sterile 1 mL syringe with 22 gauge needles. A Bacterial Artificial Chromosome (BAC) library was constructed (Lucigen ® Co. (Middleton, WI, USA) using randomly sheared genomic DNA isolated from peripheral blood cells of rock bream. The pSMART BAC vector was used to construct a library and nearly 100-120 kb size fragment resulted from randomly sheared genomic DNA was inserted into a vector. The corresponding vectors were then transformed to *E. coli* (92160 clones arrayed on 240 of 384-well microtiter plates) and library was prepared stored at -80 °C separately (Quiniou et al., 2003). For the screening purpose, the DNA was extracted from each clone separately and arranged by super pooling and pooling system, comprised of 20 super pools, 16 row pools and 24 column pools spanning the entire rock bream genome.

A two-step PCR-based screening of the BAC library was carried out for respective candidate genes, as per the manufacturer's instructions. Primers were designed based on the cDNA sequence identified from the cDNA database. Gene specific clones were isolated, cultured and purified with a Qiagen Plasmid Midi Kit (Hilden, Germany) and sequenced by Pyro-sequencing (GS-FLX 454, Macrogen, Republic of Korea). Once the genomic sequence was obtained by sequencing, the exon and intron structures were derived by comparing cDNA sequences of each gene. Genomic sequences were identified by cDNA alignment using the Spidey (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/) mRNA to genomic alignment program on NCBI or based on the prediction by the sequencing company predictions.

2.5 *In silico* characterization of CDS and protein sequences

Based on the highest identity to the known cDNA sequences of the other species in the National Center for Biotechnology Information (NCBI), complete cDNA sequences of complement genes were identified. Then, the amino acid sequences corresponding to the

coding sequences of each complement component were derived using DNAssist 2.2 (v3.0). The orthologous sequences of each gene were compared by BLAST-P of the NCBI and Ensembl databases (Altschul et al., 1990). Prediction of domains and motifs of protein was carried out using the PROSITE profile database (Bairoch et al., 1997) and SMART protein database (Letunic et al., 2009). Identity, similarity, and gap percentages were calculated by the EMBOSS pairwise alignment algorithms (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). Signal peptide sequence was identified using the SignalP v3.0 server (Bendtsen et al., 2004). Multiple sequence alignment was performed with ClustalW v2.0 using known corresponding amino acid sequences of other organisms (Thompson et al., 1994). The phylogenetic trees were constructed with the Molecular Evolutionary Genetic Analysis (MEGA) software package v5.0 using the Neighbor-Joining (NJ) method (Tamura et al., 2011) and bootstrapping values taken from 5000 replicates.

Sequencing results of the respective BAC clones were used to determine the exon and intron structures and the promoter regions of each gene. The exon and intron structures were derived by aligning the cDNA sequence with genomic sequence using Spidey program (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/). The putative transcriptional factor binding sites in 5' flanking regions were identified using TFSEARCH v1.3 according to the highest threshold score (90%) (Heinemeyer et al., 1998), Alibaba v2.1 (http://www.gene-regulation.com/pub/programs/alibaba2/index.htmL) and Neural Network Promoter Prediction in Berkeley Drosophila Genome Project (BDGP) (Hoskins et al., 2011). The genomic structure of the other species were obtained from NCBI and Ensembl genomic browser http://asia.ensembl.org/index.htmL. The structures were constructed using Gene Mapper v2.5 (http://genemapper.googlepages.com) in order to compare the structures.

The tertiary structure analysis of certain complement components were demonstrated by web based computer simulation models such as SWISS-MODEL (Arnold et al., 2006; Whang et al., 2011) and I-TASSER (Shen et al., 2011) and visualization of three Dimensional structures was done using PyMOL molecular graphic system 1.3 and POV-ray for windows v3.62 software.

2.6 Immune challenge experiment

The experiments were set-up in the Marine and Environmental Research Institute of Jeju National University. The rearing of fish, immune challenge experiments and tissue isolation were performed using the research facilities of Marine and Environmental Research Institute Jeju National University. To determine the immune responses of complement genes at transcriptional level, pathogenic bacteria (*Edwardsiella tarda* and *Streptococcus iniae),* purified LPS (055:B5 from *Escherichia coli*; Sigma), Polyinosinic-polycytidylic acid (Poly I:C; double stranded RNA) and live virus (Rock Bream Iridovirus (RBIV)) were used as immune-stimulants in time-course experiments.

2.6.1 Bacterial challenge

For the, *Edwardsiella tarda* and *Streptococcus iniae* challenge, bacteria were obtained from the Department of Aqualife Medicine at Chonnam National University (Republic of Korea). Bacteria were cultured on the Brain Heart Infusion (BHI) agar media plates supplemented with 1.5% NaCl plates at 30 ℃ for overnight. A seed culture of 5 mL using BHI broth was established from a single colony from the streaked plate. Bacterial cultures were incubated at 30 ℃ using 200 rpm shaking incubator until the OD_{600} reached the absorbance of 1.0-1.5. Then, the bacterial suspension was centrifuged at 3500 rpm for 20 min. Finally, bacterial pellet was re-suspended in PBS and cell numbers were adjusted to 5×10^6 CFU/mL by serial dilution and used as stock for injection. Then, each fish were intraperitoneally injected with *E.*

tarda (5×10^6 CFU/mL). For, *S. iniae*, fish were intraperitoneally injected with *S. iniae* (1×10^6 CFU/mL) suspended in $1 \times$ phosphate-buffered saline (PBS; 100 µL/animal).

2.6.2 LPS challenge

Sterile LPS stock was prepared by dissolving LPS at the rate of 1.25 µg/µL in PBS and filtered through a 0.2 µm filter. For the LPS challenge, the fish were injected with 100 µL LPS in PBS suspension (1.25 µg/µL).

2.6.3 Poly I:C challenge

In order to monitor the transcriptional changes of complement genes upon dsRNA injection *in vivo*, poly I:C was employed as an immune-stimulant. Sterile poly I:C stock was prepared by dissolving poly I:C at the rate of 1.5 mg/mL in PBS and filtered through a 0.2µm filter. A time course experiment was performed by intraperitoneally injecting the animals with 100 µL suspension of poly I:C stock.

2.6.4 Viral challenge

For the viral challenge, RBIV was prepared by homogenizing the kidney of a RBIV-infected rock bream with approximately 10 volumes of $1 \times$ PBS (1 mL PBS/100 mg tissue). The homogenate was centrifuged at $3000 \times g$ for 10 min, and the virus-containing supernatant was filtered through a 0.45 µm syringe filter and stored at -80°C until used. After thawing, the supernatant was diluted to obtain 10^3 TCID$_{50}$ of RBIV per 100 µL and intraperitoneally injected at a dosage of 100 µL (10^3 TCID$_{50}$ of RBIV per fish). Negative control groups were established by injecting PBS at an equal volume to the immune stimulants used in the experimental groups.

The control fish for all the challenge experiments were injected with equal volume of PBS. Rock bream head kidney, liver and spleen were excised from three randomly selected

20

fish from each of the un-challenged, immune-challenged and PBS-injected groups at 3, 6, 12, 24, and 48 h post-challenge. Tissues were immediately frozen using liquid nitrogen and stored at -80 ℃ until used for RNA extraction.

2.7 RNA extraction and cDNA synthesis

Total RNA was extracted using the TRIzol Reagent (Sigma-Aldrich) from rock bream blood, gill, liver, spleen, head kidney, kidney, skin, muscle, brain and intestine. The RNA concentration was determined at 260 nm using a UV-spectrometer (Bio-Red, USA) and diluted to 1 μg/μL.

Aliquots of 2.5 μg of RNA from each tissue were applied as template to the PrimeScript™ first-strand cDNA synthesis kit (TaKaRa, Japan). The RNA was incubated with 1 μL of 50 μM oligo(dT)$_{20}$ and 1 μL of 10 mM dNTP for 5 min at 65 °C and immediately cooled on ice. Then, 4 μL of 5× PrimeScript™ buffer, 0.5μL of RNase inhibitor (20 U), and 1μL of PrimeScript™ RTase (200 U) were added and the mixture was incubated for 1 h at 42 °C. The reaction was terminated by incubating at 70 °C for 15 min. The synthesized cDNA was diluted 40-fold and stored at -20 °C.

2.8 Quantitative real-time PCR (qPCR)

The synthesized cDNA was applied as template to the qPCR with gene-specific primers (designed by Primer 3.0) to amplify complement genes or the *β actin* (Accession No.FJ975145) reference gene. The qPCR was carried out in a 15 μL reaction volume contained 4 μL of cDNA (1:40 dilution), 7.5 μL of 2× SYBR Green Master Mix, 0.6 μL of each primer (10 pmol/μL), and 2.3 μL of PCR-grade water using the Thermal Cycler Dice™ real time system (TaKaRa) (Figure 4). The thermal cycling conditions were included one cycle of 95 °C for 3 min, followed by 45 cycles of 95 °C for 20 s, 58 °C for 20 s, and 72

°C for 30 s. To verify that each primer pair produced only a single product, a dissociation curve was generated for the product by heating from 60 °C to 95 °C at the end of the reaction.

Triplicate reactions were performed for each cDNA sample. The reaction carried out without cDNA sample was served as the negative control. Finally, the C_t values were converted into relative expression levels by the Livak $2^{-\Delta\Delta Ct}$ method. The expression levels for each and every tissues examined were normalized to that of the rock bream *β-actin* gene. For the expression profiles in normal healthy tissues, the relative mRNA levels were compared with that of expression in muscle to determine the relative tissue-specific profile. For the expression profiles in tissues from immune-challenged fish, the *β-actin* normalized mRNA levels were compared with the levels detected in the respective PBS-injected controls to calculate the fold-difference for each tissue.
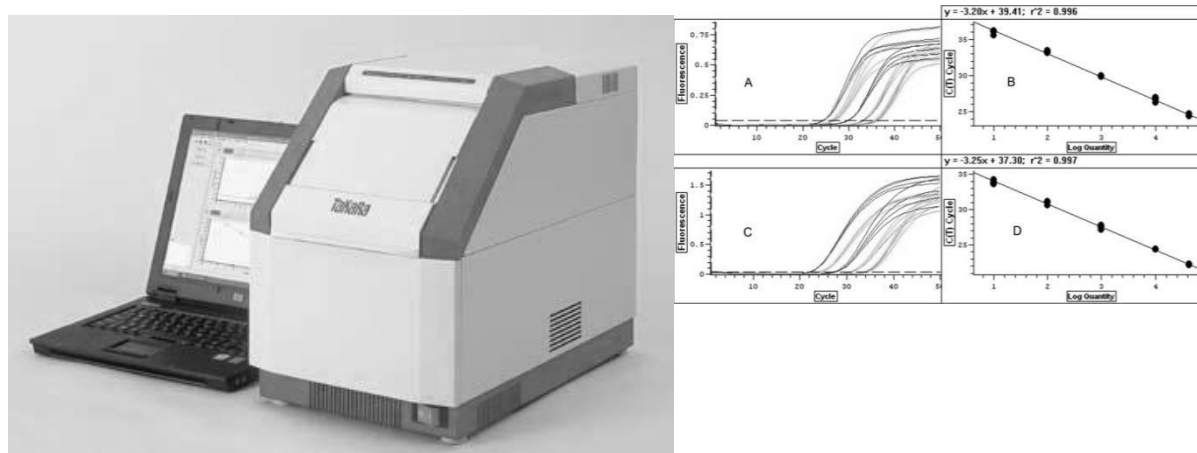
Figure 4. Quantitative real time PCR machine used to detect mRNA expression (TaKaRa, Japan)

2.9 Statistical analysis

The statistical significance was determined using the computer-based statistical software. All mRNA expression data were analyzed with Student's *t*-test in SPSS v16.0 for windows (SPSS Inc., USA) and significant differences of each time group vs. control (0 h) were

22

presented as mean ± SD in column charts. The differences at $P < 0.05$ was considered statistically significant.

2.10 Histological analysis of rock bream tissues

The spleen and liver tissues were dissected from healthy rock bream and fixed in 10% neutral buffered formalin. Fixed tissue samples were dehydrated through the ascending series of ethanol, cleared in xylene and embedded in melted paraffin. The fixed paraffin blocks containing tissues, were sectioned using microtome at 6 μm and mounted on glass slides. Sections were deparaffinized and rehydrated with xylene and series of ethanol. Tissue sections were stained with Hematoxylin and Eosin (H and E), and observed under light microscope.

2.11 RNA *in situ* hybridization (ISH)

Table 3. Information of selected genes for *in situ* hybridization

Gene	For Anti-sense probe	For sense probe	Amplicon size	Tissue /challenge
Rb-C1q-A	Spe I- A/CTAGT	Sac II- CCGC/GG	947 bp	Spleen tissue form Healthy fish
Rb-C1q-B	Spe I- A/CTAGT	Sac II- CCGC/GG	938 bp	Spleen tissue form Healthy fish
Rb-C1q-C	Sac II- CCGC/GG	Spe I- A/CTAGT	1067 bp	Spleen tissue form Healthy fish
Rb-C3	Sac I- GAGCT/C	Sac II- CCGC/GG	985 bp	Head kidney (*E. tarda* - 24h)
Rb-C7	Sal I – G/TCGAC	Sac II- CCGC/GG	1091 bp	Liver and Head kidney (healthy tissues and *E. tarda* - 12h)

The methodology was used only for molecular characterization of three isoforms of complement C1q, C3 and C7. All the reagents used in this experiment were RNAase free. To prepare the riboprobe templates for ISH, PCR products of each gene were cloned into pGEM®-T Easy vector (Promega, USA) using gene specific primers (Table 4, 8 and 14).

23

Digoxigenin (DIG)-labeled sense- and antisense-RNA probes of each gene were generated using DIG RNA Labeling Kit (SP6/T7) (Roche, Germany). To obtain the sense- and antisense- RNA probes, the plasmids were linearized with respective enzymes (Table 3) and transcribed with T7 and SP6 RNA polymerases.

Selected tissues were surgically excised from healthy rock bream and fixed in 4% paraformaldehyde in PBS (pH 7.4) at 4 °C for overnight. Tissue was mounted in 1.5% agarose and 2.5% sucrose solution followed by overnight immersion in 30% sucrose solution at 4 °C. The tissue was frozen in liquid nitrogen, sectioned (10~15 μm size) and air dried for 3 h. Each micro section was dehydrated in 100% methanol for 10 min at room temperature (RT) and rehydrated with series of methanol (75%, 50%, and 25%) in PBS. The slides were washed thrice with PBS for 3 min each and were incubated with 500 μL of prehybridization buffer (50% formamide, 5 x Saline sodium citrate buffer (SSC), 0.092M citric acid), in a humidified chamber for 2 h at RT. The pre-hybridization buffer was replaced with 150 μL of hybridization solution (prehybridization buffer with 50 μg/mL heparin, 500 μg/mL tRNA, 0.1% Tween-20) containing probe at 100 ng/μL DIG labeled RNA, which was heated at 80 °C for 5 min. For the post incubation, slides were gently submerged in 5x SSC for 10 min at 60 °C and transferred into 0.2x SSC buffer. The slides were washed for thrice with 0.2x SSC buffer at 60 °C and transferred to PBS at RT for 5 min. After blocking (with PBS plus 2% sheep serum, 2 mg/mL BSA) for 1 h, anti-DIG alkaline phosphatase conjugated antibody (Roche) was applied and incubated at 4 °C, overnight. The slides were washed several times with PBS and stained with BCIP/NBT in a humidified chamber at RT overnight. The sections were examined under a light microscope and images were captured at 400x magnification using Leica camera using an Leica camera adapter.

2.12 Construction of expression plasmid for recombinant protein expression

To produce recombinant proteins fused with Maltose Binding Protein (MBP), the coding sequences of anaphylatoxin domain (Rb-C3a, Rb-C4a and Rb-C5a) were identified and sequence specific primers were designed with restriction sites for *Eco*Rl and *Hind*III respectively (Table 8). The PCR amplification was carried out in a TaKaRa thermal cycler (TaKaRa, Japan) using 50 μL reaction mixture composed of 5 U of Ex*Taq* polymerase (TaKaRa), 5μL of 10x Ex*Taq* buffer, 8 μL of 2.5 mM dNTPs 80 ng of template 20 pmol of each primer. The thermal cycle program conditions were initial incubation at 94 ℃ for 3 min, followed by 35 cycles of 94 ℃ for 30 s, 55 ℃ for 30 s,72 ℃ for 30 s, and a final extension at 72 ℃ for 1 min. The respective amplified CDS were purified using the Accuprep® PCR purification kit (Bioneer Co., USA). The purified products and pMAL-c2X vector were digested with the respective endonucleases. The precise fragment was excised from electrophoretically resolved products and refined using Accuprep® gel purification kit (Bioneer Co., USA). Subsequently, it was ligated into linearized pMAL-c2X expression vector (New England Biolabs) using Mighty Mix (TaKaRa) and kept for overnight at 4 ℃. The recombinant plasmids were transformed into *E. coli* DH5α cells and the size of the recombinant plasmid and inserted sequences were confirmed by electrophoresis and sequencing, respectively (Macrogen Inc., Korea). Finally, sequence confirmed constructs were re-transformed into *E. coli* BL21 (DE3) cells and grown on LB/ampicillin agar plates at 37 ℃.

2.13 Prokaryotic expression and purification of recombinant anaphylatoxin domains rock bream C3 (rRb-C3a) C4 (rRb-C4a) C5 (rRb-C5a)

A single colony from each plate was grown on 5 mL Luria-Bertani (LB) broth with 100 μg/mL ampicillin for overnight at 37 ℃. A starter culture (1mL from each) was inoculated

into rich medium containing 100 μg/mL of ampicillin and 0.2 % of D-glucose, and incubated at 37 °C with shaking at 200 rpm. When the OD_{600} of bacterial culture reached ~0.6, isopropyl-β-thiogalactopyranoside (IPTG) was added into each medium at a final concentration of 0.5 mM, and the incubation was continued for 4 h at 37 °C to induce the recombinant protein expression. The cells were harvested separately by centrifugation (3500 rpm for 30 min at 4 °C) and re-suspended in column buffer (20 mM Tris-HCl, 200 mM NaCl; pH 7.4) before storing at -20 °C for overnight. Following day, cells were thawed on an ice-water bath and subjected to cold sonication. The lysates of each sample were centrifuged (13000 rpm for 30 min at 4 °C) and recombinant protein was purified according to the manufacturer's instructions on ice. The crude rRb-C3a, rRb-C4a and rRb-C5a were mixed with pre-washed amylose resin and placed on ice for 2 h to facilitate the affinity binding. Then, extract-resin mixture was loaded onto a 1×5 cm column and washed with $12 \times$ volume of column buffer. Finally, by applying 4 mL of elution buffer (column buffer + 10 mM maltose) to the column, recombinant fusion-proteins were eluted in 0.5 mL aliquots respectively (Maina et al., 1988). The concentrations of the purified fusion-proteins were determined using Bradford method (Bradford, 1976). The purity and molecular masses of the each protein were evaluated on 12% Sodium Dodecyl Sulfate-polyacrylamide Gel Electrophoresis (SDS–PAGE) under reducing condition with low molecular weight marker (Enzynomics, Korea) using Bio-Red gel apparatus system. The resulted gels after 1h electrophoresis were stained with 0.05% Coomassie brilliant blue R-250 for 1 h followed by overnight de-staining and picture was taken through scanning.

2.14 Head kidney primary cell culture preparation

Head kidney tissues were excised under sterile conditions from freshly killed rock bream (*n* = 4) and minced into small pieces (approximately 1 mm^3 in size). After washed 3 times in

Leibovitz medium (L-15; Sigma-Aldrich) contained antibiotics (400 IU/mL penicillin and 400 μg/mL streptomycin). Then, the tissue was mechanically cut into small pieces and they were re-suspended in Leibovitz's L-15 medium supplemented with 20% fetal bovine serum (FBS) with 100 IU/mL penicillin, and 100 μg/mL streptomycin. The cells were then spread out in cell culture flask and culture was maintained at 24 °C for 5-7 days before the experiments. The macrophage cells that migrated out from the tissue were collected by prior wash with PBS followed by trypsin treatment. The trypsin was removed via centrifugation at 1000 rpm for 5 min and cell pellet were washed twice with PBS. Finally, cells were re-suspended in Leibovitz's L-15 medium supplemented with 20% Fetal Bovine Serum (FBS) with 100 IU/mL penicillin, and 100 μg/mL streptomycin. The cell confluence was observed over 80 % in 24 well cell culture plates before the activity test.

2.15 Respiratory burst activity assays

The NBT assay was used to measure the respiratory burst activity of the cells. The assay was executed as previously described with small modifications (Rotllant et al., 2004). This test was performed to the three recombinant proteins obtained from anaphylatoxin domains of Rb-C3, Rb-C4 and Rb-C5 genes (rRb-C3a, rRb-C4a and rRb-C5a). Adhered macrophage-like cells from head kidney primary culture were trypsinized and seeded in a 24-well tissue culture plate (SPL life sciences, Korea) with 90% confluence. The plate was incubated overnight at 24°C to facilitate cell adherence. The cells were then washed twice with PBS. The respiratory burst activity was then induced and measured with adherent cells. The cells were first treated with 100 μL of media containing NBT (1 mg/mL, Sigma-Aldrich, USA). Thereafter, the individual cells were treated separately with 20 μg of protein, 20 μg of protein + 20 μL of superoxide dismutase (SOD, 100 IU/mL) and 20 μL of 1% H_2O_2. Total reaction volume was adjusted to 260 μL by adding serum free L-15 medium and all treatments were

performed in triplicates. The 100 μL of media containing NBT alone was considered as negative control in the experiment. The plate was incubated at 24°C for 1 h period and then washed twice with PBS. The cells were fixed by adding 100% methanol kept for 3 min and washed with 70% methanol followed by air drying. The reduced form of NBT (formazan) was dissolved in 120 μL of KOH (0.112 g/mL) and 140 μL of Dimethyl Sulfoxide (DMSO). Finally, the absorbances of each treatment were measured at 620 nm using a spectrophotometer (Thermo Scientific, USA).

2.16 Promoter characterization of complement component 2

2.16.1 Construction of Rb-C2 promoter sequences

To obtain the 5' flanking region of the complement C2, BAC clone sequence was used. The promoter region was identified using web based prediction software and by comparing of the complete cDNA sequences with genome sequence as described in section 2.5. To clone the 5' flanking region Rb-C2 into pGL2 luciferase reporter vectors (promoter-less basic vector) (Promaga, USA), the sequence specific primers were design with *Kpn*I and *Xho*I restriction sites. The PCR amplification was carried out in a TaKaRa thermal cycler (TaKaRa, Japan) using 50 μL reaction mixture composed of 5 U of Ex*Taq* polymerase (TaKaRa), 5μL of 10 x Ex*Taq* buffer, 8 μL of 2.5 mM dNTPs, 80 ng of template, 20 pmol of each primer. The thermal cycle program conditions were initial incubation at 94 ℃ for 3 min, followed by 35 cycles of 94 ℃ for 30 s, 58 ℃ for 30 s, 72 ℃ for 30 s, and a final extension at 72 ℃ for 1 min. The respective amplified promoter constructs were purified using the Accuprep® PCR purification kit (Bioneer Co., USA). The purified products and pGL2 basic vector were digested with the respective endonucleases. The precise fragment was excised from electrophoretically resolved products and refined using Accuprep® gel purification kit (Bioneer Co., USA). Subsequently, it was ligated for overnight at 4 ℃ with linearized pGL2

28

expression vector (Promega, USA) using Mighty Mix (TaKaRa). The recombinant plasmids were transformed into *E. coli* DH5α cells. The size of the recombinant plasmid and inserted sequences were confirmed by electrophoresis and sequencing respectively (Macrogen Inc., Korea). Afterward, sequence confirmed constructs were re-transformed into *E. coli* DH5α cells and these cells were grown on LB/ampicillin agar plates at 37 ℃ in an oven. A single colony was up-scaled in LB/ampicillin broth for 16 h. Finally, plasmids were purified using midi prep isolation kit (Qiagen, Hilden, Germany) and plasmid concentrations were measured using Microplate Spectrophotometer (Thermo Scientific™ Multiskan™ GO, USA).

2.16.2 Transfection of vector constructs to the human hepatoblastoma cells (HepG3)

Human hepatoblastoma cells (HepG3, European Collection of Cell Cultures, and Cambridge, U.K.) were maintained in Eagle's minimum essential medium supplemented with 1% non-essential amino acids, 1 mM sodium pyruvate (BioWhittaker, Walkersville, MD, USA.), 2 m ML-glutamine, 10% (v/v) fetal calf serum (Harlan, Indianapolis, IN, U.S.A.), 100 units/mL penicillin, 100 μg/mL streptomycin and 2.5 μg/mL amphotericin (antibiotics from Roche,USA). The cells were grown overnight in a 35-mm size culture dishes with growth medium at 37 °C under 5 % CO_2. The cells were incubated until the monolayer reached 80% confluence and were washed with PBS. Cells were then pre-incubated with serum free medium for 20 min and transfected with transfection mixture (91μL medium, 8μL of Lipofectamine™ 2000 and 1μL of recombinant pGL2 vector containing 1.0 μg vector). The cells transfected only with the empty vector served as the transfection control. After 6 h of incubation, cells were incubated with normal growth medium for 24 h at 37 °C under 5% CO_2.

2.16.3 Luciferase assay

 The luciferase assay was performed to determine the promoter activity according to the E1500 standard protocol. The incubated cells were washed with PBS and 100 μL of cell lysis buffer was added to the each well and allowed 5-10 min for cell lysis. Then, all the cells were scraped and collected into a micro centrifuge tube. The cell lysate was kept for 40 min on ice for complete lysis of the cells. The luciferase activity was measured by mixing with luciferase substrate in an opaque 96-well microplate using luminometer (Promega GLOMAX$^{®}$ Multi) and reported as light units per mg of protein.

Chapter I

Molecular characterization of three complement-C1q isoforms as initiative components of the complement system

ABSTRACT

Complement component 1q (C1q) is a subcomponent of the C1 complex and the key protein that recognizes and binds to a broad range of immune and non-immune ligands to initiate the classical complement pathway. In the present study, we identified and characterized three novel C1q family members from rock bream, *Oplegnathus fasciatus*. The full-length cDNAs of C1q A-like (RbC1qAL), C1q B-like (RbC1qBL), and C1q C-like (RbC1qCL) consist of 780, 720 and 726 bp of nucleotide sequence encoding polypeptides of 260, 240 and 242 amino acids, respectively. All three RbC1qs possess a leading signal peptide and collagen-like region(s) (CLRs) in the N-terminus, and a C1q domain at the C-terminus. The C1q characteristic Gly-X-Y repeats are present in all three RbC1qs, while the CLR-associated sequence that enhances phagocytic activity is present in RbC1qAL (^{49}GEKGEP^{54}) and RbC1qCL (^{70}GEKGEP^{75}). Moreover, the coding region is distributed across six exons in RbC1qAL and RbC1qCL, but only five exons in RbC1qBL. Phylogenetic analysis revealed that the three RbC1qs tightly cluster with the fish clade. All three RbC1qs are most highly expressed in the spleen and liver, as indicated by qPCR tissue profiling. In addition, all three are transcriptionally responsive to immune challenge, with liver expression being significantly up-regulated in the early phase of infection with intact, live bacteria (*Edwardsiella tarda* and *Streptococcus iniae*) and virus (rock bream iridovirus) and in the late phase of exposure to purified endotoxin (lipopolysaccharide). These data collectively suggest that the RbC1qs may play defense roles as an innate immune response to protect the rock bream from bacterial and viral infections.

1. Introduction

The complement system can be activated by three major signaling cascades, known as the classical, lectin and alternative pathways, which lead to various but interactive biological processes, such as inflammation, phagocytosis, lysis, and the adaptive immune response (Sunyer and Lambris, 1998). The classical pathway itself is activated upon recognition and binding of antibody-antigen complexes by the complement component 1 (C1) complex, which is composed of the C1q recognition molecule and four serine protease proenzymes, including two units of C1r and two units of C1s (Arlaud et al., 2001).

C1q has been characterized as the primary link between innate immunity driven by classical pathway and acquired immunity mediated by the immunoglobulin (Ig) components of antibody-antigen complexes, IgG and IgM (Kishore and Reid, 2000). Moreover, its activity has been demonstrated as crucial to a broad spectrum of immunological processes, including phagocytic removal of bacteria, retrovirus neutralization, apoptotic cell clearance, immune cell adhesion, and growth modulation of dendritic cells, B cells, and fibroblasts (Kishore et al., 2004b). C1q mediates these effects via strong, targeted binding to a diverse set of host- and pathogen-derived ligands, such as the lipopolysaccharide (LPS) and porin bacterial endotoxins, the envelop proteins of some retroviruses, various phospholipids, β-amyloid fibrils, apoptotic cells, and pentraxins (Kishore et al., 2004a). The unique protein structure of C1q might likely mediate this diverse capacity to recognize and bind such a vast array of ligands.

The C1q subcomponent has been identified in nurse shark (Smith, 1998), channel catfish (Dodds and Petry, 1993), zebrafish (C1qA, C1qB and C1qC) (Hu et al., 2010), and mandarin fish (C1qL1 and C1qL2) (Lao et al., 2008). In addition, several studies of fish (Lao et al., 2008) and some other invertebrates (Gestal et al., 2010) have demonstrated that C1q

activity is important to immune defense mediated by the classical complement pathway against invading pathogens.

In this study, we attempted to characterize the full-length cDNA and genomic sequence of putative C1qs from rock bream. Our data on the temporal expression profile and transcriptional response in liver to infection by common pathogens, including *Edwardsiella tarda* and *Streptococcus iniae* bacteria and the rock bream iridovirus (RBIV), provide novel insights into the C1q-mediated defense mechanisms of rock bream.

Table 4. Description of primers used for amplification of genes in this chapter

Primer	Purpose	Amplicon size (bp)	Sequence, 5'-3'	Efficiency (%)	Accession No of target
RbC1qAL F	qPCR amplification	124	TGAGCAACAAGCTGGGATTCTGTG	103	JQ805140
RbC1qAL R	qPCR amplification		AAGGACTCGAGCCAAACCTTCTGT		
RbC1qBL F	qPCR amplification	188	AGAGGTCGACACTGCCATCAACTT	98.6	JQ805141
RbC1qBL R	qPCR amplification		TTATGTGACTGTCGCTGCCCTTCA		
RbC1qCL F	qPCR amplification	130	ATACCCAGACAAAGCCAGCGTCAT	100.8	JQ805145
RbC1qCL R	qPCR amplification		AGAGGCGTGGAACGCAAAGTAGTA		
Rb β-actin F	qPCR internal reference	108	TCATCACCATCGGCAATGAGAGGT	99.7	FJ975145
Rb β-actin R	qPCR internal reference		TGATGCTGTTGTAGGTGGTCTCGT		
C1qAL-ISH F	*In situ* hybridization	938	TGCTTCTGACAACCGGCCAAT	-	JQ805140
C1qAL-ISH R	*In situ* hybridization		AGACTCAATAATAAACTTTAATTTTGTTTTGGTTGCATTCACAA		
C1qCL-ISH F	*In situ* hybridization	1067	CTGAATCTTTGGTTTGCAGTCGGCT	-	JQ805145
C1qCL-ISH R	*In situ* hybridization		GGCCAGGTCTCTCTTGAAAGGAGATTTTAA		

2. Results

2.1 cDNA sequence characterization of rock bream C1qs

Three full-length C1qs, C1qA-like (RbC1qAL) and C1qC-like (RbC1qCL) from the rock bream cDNA library and C1qB-like (RbC1qBL) from the BAC library, were isolated and deposited in GenBank under the accession numbers JQ805140, JQ805145, and JQ805141, respectively. Sequence analysis indicated that the RbC1qAL cDNA is 1125 bp in

34

length. The ORF is 780 bp (excluding the termination codon) and encodes a polypeptide of 260 aa with an estimated molecular mass of 27.8 kDa. The 5′-untranslated region (UTR) is 96 bp in length and the 249 bp 3′-UTR contains a polyadenylation signal (^{1056}ATTAAA^{1061}) and an extended poly (A) tail (Figure 5a). The cDNA of RbC1qBL is 970 bp and contains an ORF of 720 bp (excluding the termination codon) encoding a polypeptide of 240 aa with an estimated molecular mass of 25.3 kDa. Alignment with the known sequence of C1q-like-1 protein from mandarin fish (Lao et al., 2008) revealed that the 5′-UTR of RbC1qBL is 71 bp in length and the 3′-UTR is 179 bp. The 3′-UTR contains two polyadenylation signals (^{863}ATTAAA^{868} and ^{926}AATAAA^{931}) (Figure 5b). The full-length cDNA of RbC1qCL is 1262 bp and consists of a 726 bp ORF (excluding the termination codon), a 5′-UTR of 146 bp, and a 3′-UTR of 390 bp with two polyadenylation signals (^{1053}ATTAAA^{1058} and ^{1154}ATTAAA^{1159}) and an extended poly A tail. In addition, the ORF of RbC1qCL encodes a polypeptide of 242 aa with an estimated molecular mass of 26.1 kDa (Figure: 5 c).
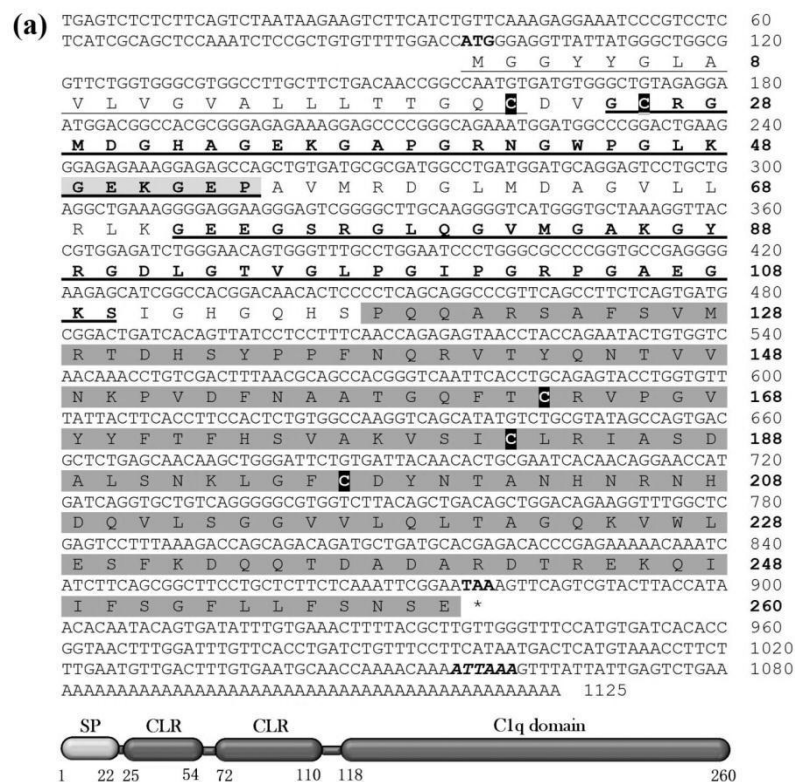
Figure 5. The complete nucleotide sequence with deduced amino acid sequence of (a) RbC1qAL, (b) RbC1qBL, and (c) RbC1qCL

The start and stop codons are in bold letters. The polyadenylation signals are in bold and italicized. The putative signal peptide (SP) is underlined. The collagen-like region (CLR), comprised of Gly-X-Y repeating triplets, is in bold and underlined. The C1q domain is highlighted in grey. Conserved cysteine residues are highlighted in black. The GEKGEP motif in CLR is highlighted in light grey. The two potential N-linked glycosylation sites in RbClqBL are underlined with waves. The schematic diagram of each C1q is illustrated below the sequence.
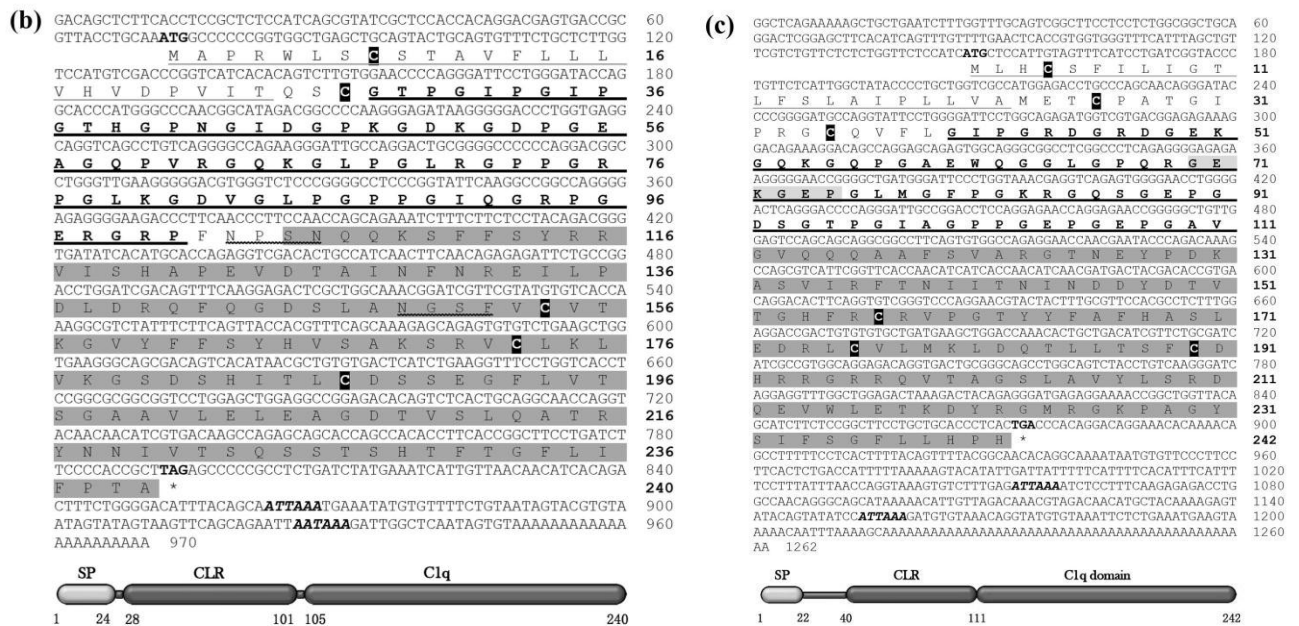
Figure 5. Continued.

The organization of C1q domain-containing (C1qDC) proteins has been characterized, and includes a leading signal peptide (SP), a CLR of variable size, and a globular C1q (gC1q) domain in the C-terminus (Mei and Gui, 2008). An SP was identified in all three rock bream C1qs (RbC1qs), and was 22 aa in RbC1qAL, 24 aa in RbC1qBL, and 22 aa in RbC1qCL. Motif analysis of the RbC1qs revealed that this specific sequence, responsible for stimulating phagocytic activity was located in the CLRs of both RbC1qAL ([49]GEKGEP[54]) and RbC1qCL ([70]GEKGEP[75]), but was absent from RbC1qBL. In addition, the characteristic Gly-X-Y (where X is often proline and Y is often hydroxylysine or hydroxyproline) repeats were clearly recognized in the CLRs of all three RbC1qs, and the number of repeats were 23 in

36

RbC1qAL, 24 in RbC1qBL, and 23 in RbC1qCL. The gC1q domain was present in the C-terminals of all three RbC1qs, and was 142 aa long in RbC1qAL, 135 aa long in RbC1qBL, and 131 aa long in RbC1qCL. Moreover, all three RbC1qs contained a total of five cysteine residues, three of which were located in the gC1q domains of each. Only RbC1qBL contained N-linked glycosylation sites ([103]NPSN[106] and [149]NGSF[152]), suggesting a unique glycoprotein function for this RbC1q.

```
S.chuatsi C1qL-1     1   MAPQWLSCSTA--VFLLLVHVGQVTTQSC-------GTPG------------IPCIPGTHCPNCIVCPKGKGDPEG-------GQP 60
O.fasciatus C1qBL    1   MAPRWLSCSTA--VFLLLVHVDPVITQSC-------GTPG------------IPCIPCTHCPNCIDCPKCDKCDPGEA-------GQP 60
D.labrax C1qB        1   MAAQWLSCSTA--VFLLLVHMDSVVTQSC-------GAPG------------IPCIPGTHGPNCKDCLKCEKCDPGEA-------VQP 60
D.rerio C1qB         1   MLFALMSAHVVPQLAIMLLLVTSSMSETCAGNK---GTPG------------TPCIPGVFGTCGKDGAKCEKGDPGEN-------EVQ 66
H.sapiens C1qC       1   MDVGPSSLPHLGLKLLLLLLLPLRGQANTGCY---GIPG------------MPCLPGAEGKDGYDGLPCPKGEPGIP-------AIP 66
H.sapiens C1qB       1   MMMKIPWGSIPVLMLLLLLGLIDISQAQLSCTGPP---AIPG------------IPCIPGTFGPNCQPCTPCIKGEKGLP-------GLA 68
D.labrax C1qC        1   MGGYYGLAVLVGVALLLTTGQCDVSCRGTDGRAGEDGIPGRDGWPGVKGEKGEPGIPGMPGRDGRDCGKCDPGTV-------GNV 81
O.niloticus C1qC     1   MLHHCLLIIGALLSLAISGLIAMETCPAT----GMPG------------MPCIPGLPGRDCRDCQKGEKCDPGAK-------LQG 62
O.fasciatus C1qCL    1   MLHCSFILIGTLFSLAIPLLVAMETCPAT----GIPRG------------CQVFLCIPGRICRDCEKCQKCQPGAE-------WQG 63
D.rerio C1qC         1   MFGGHLILVSLLSASLCLCLASADTCPAG----AMPG------------LPGIPGFFGFRGRQCMKGEKGDLGIP-------IKP 62
H.sapiens C1qA       1   MEGPRGWLVLCVLAISLASMVTEDLCR-------APDG------------KKGEAGRFGRFCRPCLKGEQGEPGAP-------GIR 60
O.fasciatus C1qAL    1   MGGYYGLAVLVGVALLLTTGQCDVGCR-------GMDG------------HACEKCAIGRNGWPGCLKGCEPAVMRDGLMDAGVL 67
S.chuatsi C1qL-2     1   MGGYYGLAVLVGVASLLTTGQCDVSCS-------CPAG------------HPCGAGAFGRNGWPGAKGEKGEPAARAEGPVDAGVL 67
D.rerio C1qA         1   MQPSAFFAFLWAGALFPFSFCQDEC-------VKHG------------RNGADGPNGRDGLPCPKGEKGEPALQVK-----LSS 60
                                                                                            <---©

S.chuatsi C1qL-1    61   --VRGQKGLEGLRGPPGRPGLKGDEGLEGPRGNPGRP-GEKGR-PFNPSN--QQKS--YFSYKRVMSYAPEV-DTDIFFNRQILPDLDQQ 141
O.fasciatus C1qBL   61   --VRGQKGLEGLRGPPGRPGLKGDVGLEGPGIQGRP-GERGR-PFNPSN--QQKS--FFSYRRVISHAPEV-DTAINFNREILPDLDRQ 141
D.labrax C1qB       61   --VKGQKGTGGPRGPPGRPGLKGDVGLPGPAGYAGRP-GEKGR-PFNPSN--QQKS--FFSYKRMISQAPEL-DTAMNFNRQILPDLDQQ 141
D.rerio C1qB        67   --MTGPKGDGFGKGGLPGRPGVKGPEGPQGPPGPPG-PKGQRGV-LSGKVAP-DQYF--VFSYKKSQKLEKILQDKLVVFDVPLITGIDGV 149
H.sapiens C1qC      67   G-IRGPKGQKGELGLPGHPGKNGPMGPPGMPGVPG-PMCIPGE-PGEEGRY-KQKFQSVFTVTR-QTHQPPAPNSLIRFNA-VLTNPQGD 150
H.sapiens C1qB      69   G-DHGEEGEKGDPGIPGNPGKVGPKGPMGPKGCGAP-GAPGP-KGESGDY-KATQKIAFSATRTINVPLRR-DQTIRFDH-VITNMNNN 152
D.labrax C1qC       82   G-LGFQKGDKGDGGLMGFSGKRGQSGDIGEPGVPGVA-GPLGE-PGEAGTI-GVRQQAAFSVARGTDQYPDK-ASIIRFTN-VITNINND 165
O.niloticus C1qC    63   G-LGPGKGEKGEIGGPKGPPGKRGQIGESGKPGTGA-GPPGE-PGEQGAV-VVQQQAAFSVARGTNEYPDK-ASVIKFTD-AITNINDD 146
O.fasciatus C1qCL   64   G-LGPQRGEKGELMGGFPGKRGQSGEEPGDSGTPGIA-GPPGE-PGEPGAV-GVQQQAAFSVARGTNEYPDK-ASVIRFTN-IITNINDD 147
D.rerio C1qC        63   G-DTVKKGERGAFGLKGFPGKRGPHGDIGIMGPPGPP-GEPGE-AGLVDVS-GSQLQSAFSVSRHTRIPPDA-NKVIRFSK-VITNPQGH 146
H.sapiens C1qA      61   TGIQGLKGDQGEPGPSGNPGKVGYPGPSGPLGARGIP-GIKGT-KGSPGNI-KDQPRPAFSAIR--RNPPMG-GNVVIFDT-VITNQEEP 143
O.fasciatus C1qAL   68   LRLKGEEGSRGLQGVMGAKGYRGDLGTVGLPGIPGRP-GAEGK-SIGHGQHSPQQARSAFSVMRTDHSYPPF-NQRVTYQN-TVVNKPVD 153
S.chuatsi C1qL-2    68   LMLKGEIGSRGLQGAMGPKGYRGDIGPAGRPGNPGRP-GADGK-SIGHGQHALQQAHSASSAIRTDVSYPRY-NQPVTYQN-TVVNKPGD 153
D.rerio C1qA        61   IALEELKGDMGVRGPPGEPGLEGLMGAIGPRGPPLG-PAGPRGSSVGADGAKASEKP--AFSVLRNEASQAQY-KQPVTFND-KLSDANDD 145
                              <---         CLR --->

S.chuatsi C1qL-1   142   FQGELLTNGSFTCVIKGVYFFSYHVS-AKSRVCLKLVKG----SDSHMTLCD-----SSEGFL-VTSGSAVLELEAGDTVSLQATK--YN 218
O.fasciatus C1qBL  142   FQGDSLANGSFVCVTKGVYFFSYHVS-AKSRVCLKLVKG----SDSHITLCD-----SSEGFL-VTSGAAVLELEAGDTVSLQATR--YN 218
D.labrax C1qB      142   LQGDALTNGTFTCLVKGIYFFESYHVS-ARSKVCLKIQIG----AEEKVKLCD-----SPKLIL-GTAASVVLKLNKGDKVSVQSTG--ES 222
D.rerio C1qB       150   LDGE----GYFDVTITGMYIYSYQIS-FQQSACLKIQIG----AEEKVKLCD-----SPKLIL-GTAASVVLKLNKGDKVSVQSTG--ES 222
H.sapiens C1qC     151   YDTS--T-GKFTCKVPGLYYFVYHAS-HTANLCVLLYRS----GVKVVTFCGH----TSKTNQ-VNSGGVLLRLQVGEEVWLAVND--YY 225
H.sapiens C1qB     153   YEPR---GHFRCRVPGLYYFTYHAS-SRGNLCVNLMRGRE-RAQKVVTFCDY----AYNTFQ-VTTGGMVLKLEQGENVFLQATD--KN 230
D.labrax C1qC      166   YQTET---GRFRCRVPGTYYFVFHAS-LEDKLCVLMKVD----QTLVTSFCDH----RRGKRQ-TTAGSVAVYLSRNQEVWLETKD--YR 240
O.niloticus C1qC   147   YNIN---TGRFRCRVPGTYYFVYHAS-LDEKLCVLLKRD----NMLLASFCDH----RQTGRQ-VTSGGLAVYVSKDQEVWLETKD--YR 221
O.fasciatus C1qCL  148   YDTVT---GHFRCRVPGTYYFAFHAS-LEDRLCVLMKLD----QTLLTSFCDH----RRGRRQ-VTAGSLAVYLSRDQEVWLETKD--YR 222
D.rerio C1qC       147   FSTDE---SKFVCKIPGTYYFVLHAS-SHDKKLCVILVHD----DKNLVSFCDH----NLSKSVSL-VVSGGAVLTLLKGDKVWNIEPFAG--LN 223
H.sapiens C1qA     144   YQNHS---GKYTCRVPGVYYFTFQVL-SQWEICLSIVSSSRGQVRRSLGFCDT----TNKGLFQVVSGGMVLQLQQGDQVWVEKDP--KK 223
O.fasciatus C1qAL  154   FNAAT---GQFTCRVPGVYYFTFHSV-AKVSICLRIASD---ALSNKLGFCDYNTANHNRNHDQVLSGGVKVLWLESFKDQQT 236
S.chuatsi C1qL-2   154   FNAAT---GHFTCRVPGVYYFTFHSV-AKVSMCLYIATD---APFNKLGFCDY-----NRNADQVLSGGVVLQLAARQKVWLESFRDQQT 231
D.rerio C1qA       146   FQIKT---GYFTCKVPGVYYFEVFHAS-SEGRLCLRLKSTS--APPVSLSFCDF----NSKSVSLVVSGGAVLTLLKGDKVWIEPFAG--D 223
                              ©                              ©                        ©

S.chuatsi C1qL-1   219   NIVTSQS-STSHTLAGFLIFPTT 240
O.fasciatus C1qBL  219   NIVTSQS-STSHTFTGFLIFPTA 240
D.labrax C1qB      219   NIVTSQS-STSHTFTGFLIFPTA 240
D.rerio C1qB       223   TVFSR---DTDCTFTGFMLFPIK 242
H.sapiens C1qC     226   DMVGIQ--GSDSVFSGFLLFPD 245
H.sapiens C1qB     231   SLLGME--GANSIFSGFLLFPDMEA 253
D.labrax C1qC      241   GMRGRP--AGYSIFSGFLLHPH 260
O.niloticus C1qC   222   AMRGTR--NGYSIFSGFLLHSH 241
O.fasciatus C1qCL  223   GMRGKP--AGYSIFSGFLLHPH 242
D.rerio C1qC       224   GMYATAD-RADSVFSGFLIHAH 244
H.sapiens C1qA     224   GHIYQGS-EADSVFSGFLIFPSA 245
O.fasciatus C1qAL  237   DADARDTREKQIIFSGFLLFSNSE 260
S.chuatsi C1qL-2   232   DPGVRDTREKQIIFSGFLLF 251
D.rerio C1qA       224   GGVGQMPKRLYAVFNGFLIYRNAE 247
```

Figure 6. Multiple alignment of deduced amino acid sequences of RbC1qs with putative orthologous

Completely conserved residues are highlighted in black, and highly conserved residues are highlighted in grey. Dashes in the amino acid sequences indicate gaps that were introduced to maximize alignment. Conserved cysteine residues are indicated by ©.

2.2 Amino acid sequence comparison and phylogenetic analysis of the RbC1qs

To investigate the evolutionary distance of RbC1qs from C1qs of other species, the pairwise identity and similarity were analyzed. RbC1qAL, RbC1qBL, and RbC1qCL displayed the highest identities with *Siniperca chuatsi* C1qL2 (78.5%), *Siniperca chuatsi* C1qL1 (87.1%), and *Oreochromis niloticus* C1qC (73.1%), respectively (Table 5). In contrast, the RbC1qs showed 30-45% identities with human C1qs. Multiple alignment analysis identified several conserved sequence regions among the C1qs of different species (Figure 6), including 19 Gly-X-Y repeats in the CLR and four cysteine residues.
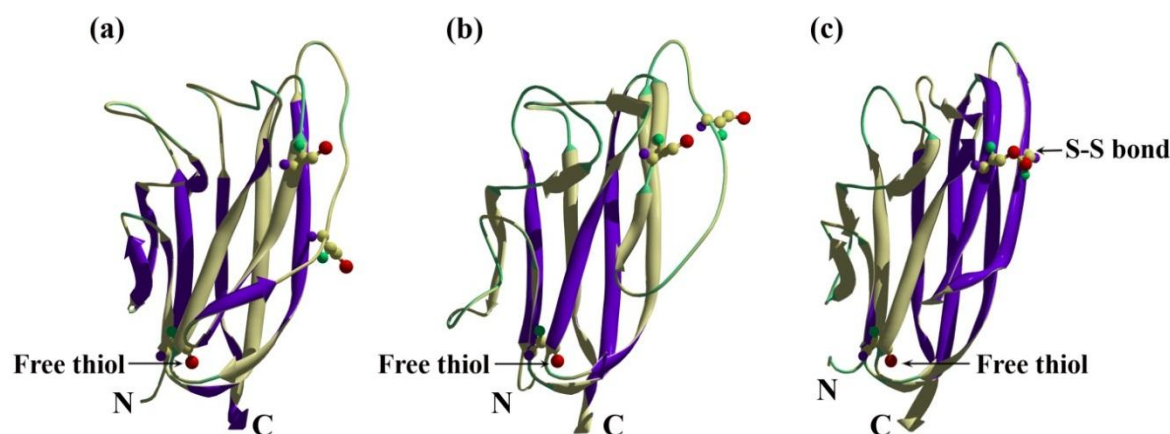
Figure 7. Modeled 3D structures of the globular domains of (a) RbC1qAL, (b) RbC1qBL, and (c) RbC1qCL

A ribbon diagram is displayed for each RbC1q, with the conserved cysteine residues and sulphur atom highlighted in red. N- and C-terminals are labeled as N and C, respectively.

Table 5. Percentage of interspecies amino acid sequence identity and similarity for RbC1qAL, RbC1qBL, and RbC1qCL

Identity (%) / **Similarity (%)**

	RbC1qAL	S. chuatsi C1qL2	H. sapiens C1qA	D. rerio C1qA	RbC1qBL	S. chuatsi C1qL1	D. labrax C1qB	O. niloticus C1qBL	S. salar C1qB	D. rerio C1qB	H. sapiens C1qB	RbC1qCL	O. niloticus C1qC	D. labrax C1qC	D. rerio C1qC	H. sapiens C1qC
RbC1qAL		**78.5**	39.8	36.4	31.2	32.7	30.5	29.7	33.1	27.5	33.7	33.8	37.0	44.0	34.0	33.5
S. chuatsi C1qL2	82.7		37.2	37.6	30.8	30.8	30.0	28.1	32.7	28.6	34.1	33.6	35.5	43.6	34.5	33.2
H. sapiens C1qA	54.2	51.4		35.9	32.7	32.2	33.3	32.4	37.3	30.8	40.3	34.0	34.8	37.4	37.3	37.6
D. rerio C1qA	51.2	54.2	51.0		32.9	32.6	32.9	30.7	34.0	29.7	32.7	32.5	35.7	33.7	34.0	34.5
RbC1qBL	46.5	45.0	51.0	47.0		**87.1**	82.5	57.6	48.2	35.5	38.4	34.0	32.7	31.2	32.5	36.9
S. chuatsi C1qL1	46.5	46.6	51.0	48.2	92.9		82.1	60.5	50.2	36.7	36.8	32.4	33.1	31.2	33.2	36.9
D. labrax C1qB	47.3	45.0	51.8	46.6	92.9	89.6		60.9	49.8	38.9	37.2	33.1	35.1	31.9	32.4	34.9
O. niloticus C1qBL	46.5	45.0	51.4	46.6	71.3	71.7	72.9		47.2	36.1	39.7	32.0	34.7	30.6	33.2	34.5
S. salar C1qB	47.7	48.2	53.8	49.4	62.2	61.4	62.2	61.0		41.8	36.8	34.4	36.7	31.6	36.2	34.3
D. rerio C1qB	44.6	43.4	48.6	47.0	55.4	54.5	57.0	56.2	61.4		35.0	30.3	32.1	30.3	31.2	33.5
H. sapiens C1qB	52.7	54.5	57.7	53.0	56.9	56.1	56.5	55.3	57.3	54.9		37.6	41.1	38.0	38.5	48.4
RbC1qCL	48.1	49.4	53.5	49.4	52.5	50.0	52.5	51.7	51.4	49.2	51.4		**73.1**	65.6	46.6	44.2
O. niloticus C1qC	51.5	54.2	53.1	51.8	53.1	53.1	52.7	53.5	53.0	48.3	53.8	83.1		59.6	50.4	46.0
D. labrax C1qC	58.5	55.4	54.2	48.1	47.7	46.5	46.5	48.1	50.4	46.2	51.9	76.5	71.2		45.1	43.6
D. rerio C1qC	50.0	49.8	55.9	49.8	48.8	50.0	49.6	48.0	52.2	50.0	54.5	63.1	68.0	58.8		46.6
H. sapiens C1qC	53.8	51.8	55.5	51.0	53.5	55.1	53.1	52.7	46.6	50.2	63.2	56.3	59.2	53.8	60.0	

Molecular modeling studies were performed to evaluate the conservation of folding structure in the rock bream gC1q domains by comparison with human gC1qs. Predicted structures of the rock bream gC1qs exhibited a jelly-roll topology with 10 β-strands (Figure 7), and also the position of cysteine residues in the gC1q regions is depicted in the each rock bream gC1q. Only RbC1qCL had a clearly demonstrated intra-chain disulfide bond between last two cysteine molecules, which were closest to one another. In addition, free thiol group was placed in similar position of the each RbC1qs.

A phylogenetic tree was constructed for the C1qs based on the three subcomponents from different species in order to ascertain the origin and formation of RbC1qs. As shown in Figure 8, the C1qs fell under three distinct clades. Each clade diverged into two vertebrate subclasses (fish and other vertebrates) and the RbC1qs clustered within the fish subclass.
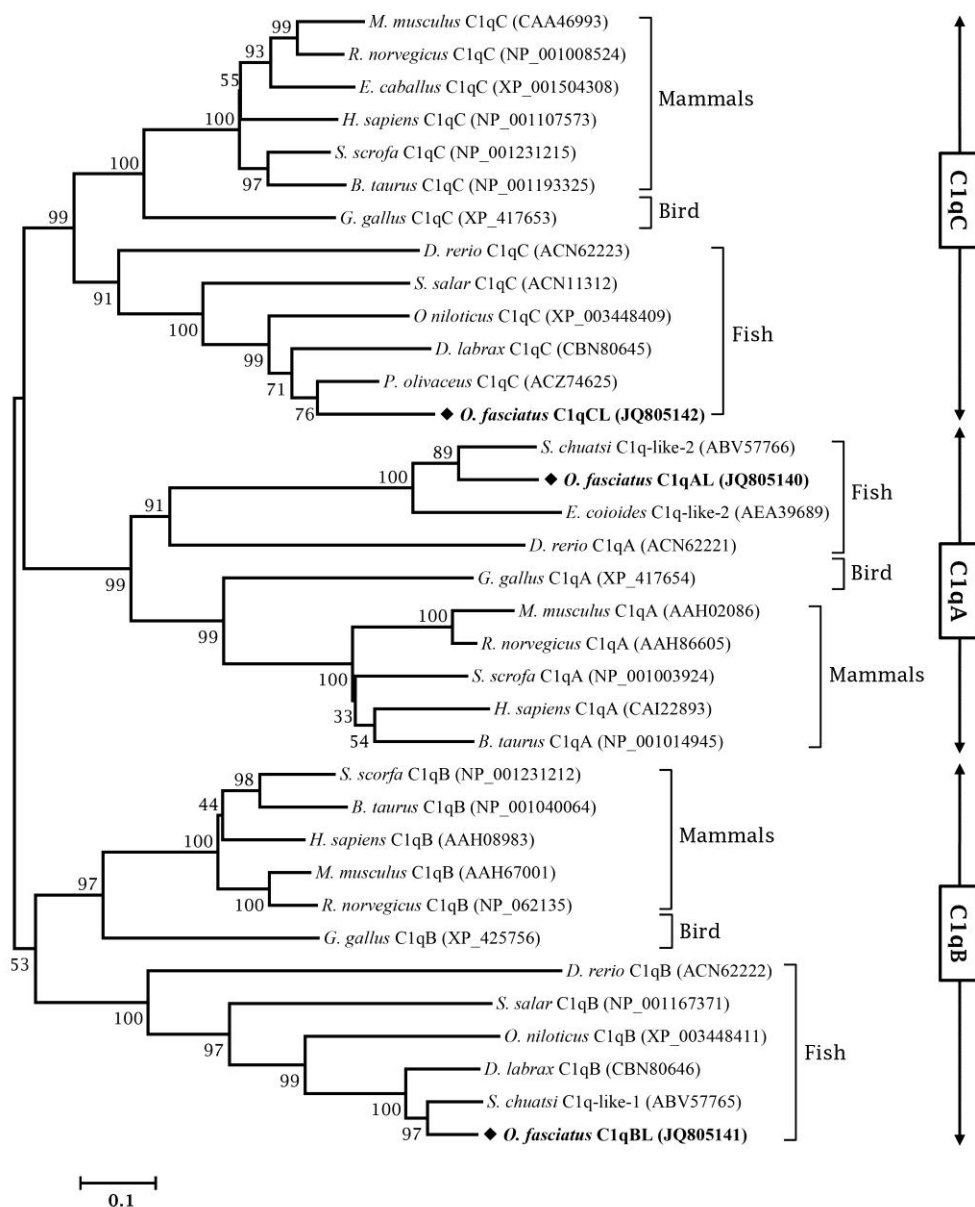
Figure 8. Phylogenetic tree of RbC1q family constructed by the neighbor-joining method
The numbers at each branch indicate the corresponding percentage of bootstrapping.

2.3 Genomic analysis of the RbC1qs

To determine the genomic structures of RbC1qs, BAC library were screened and sequenced by next generation sequencing strategy. The lengths of the genomic sequences of RbC1qAL, RbC1qBL and RbC1qCL were 5198 bp, 7560 bp and 5385 bp respectively and submitted to

the GenBank (Accession numbers JQ805142, JQ805143 and JQ805144). The ORFs of RbC1qAL, RbC1qBL, and RbC1qCL were distributed among six, five, and six exons, respectively, with the sixth exons being located in the 5′-UTRs (Figure 9). All the exon-intron boundaries of each were confirmed by the presence of splice acceptor (AG) and donor (GT) sites. Comparison of the RbC1qs genomic structures with those of other species demonstrated that two fish species, medaka (*Oryzias latipes*) and tetraodon (*Tetraodon nigroviridis*), exhibited a similar structural pattern even though tetraodon having one more extra exon in C1qA. The size of the first, second and fourth exons of RbC1qAL were comparable with those in the ORF of medaka C1qA; whereas, all of the exons of RbC1qBL were identical to those in the ORF of tetraodon C1qB. For RbC1qCL, the last two exons in the ORF were of similar sizes to those in both medaka and tetraodon ORFs. The ORFs of all mammalian C1qs are limited to two exons. Notably, zebrafish C1qs exhibit similar exon patterns to mammalian C1q counterparts.
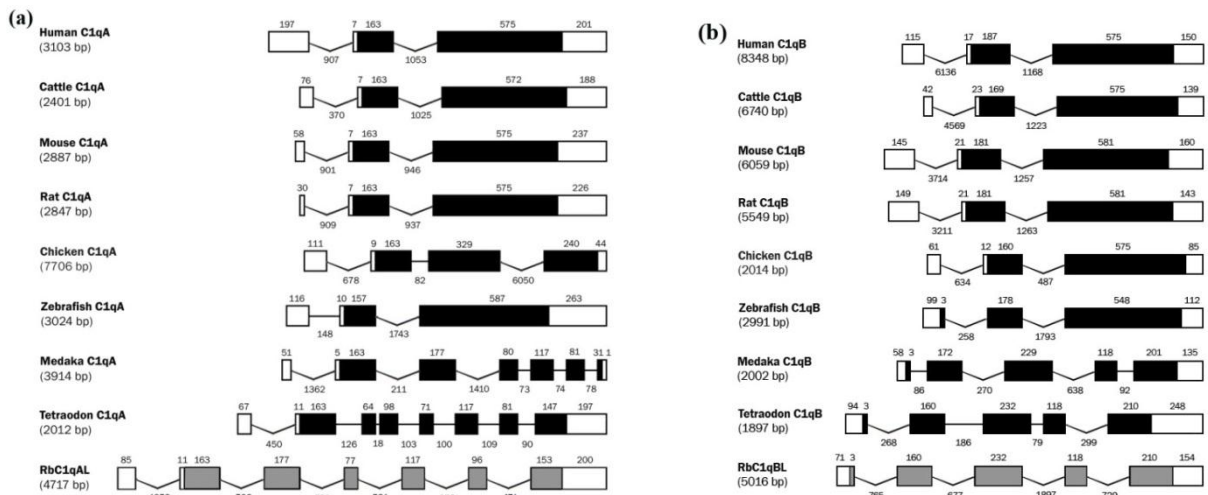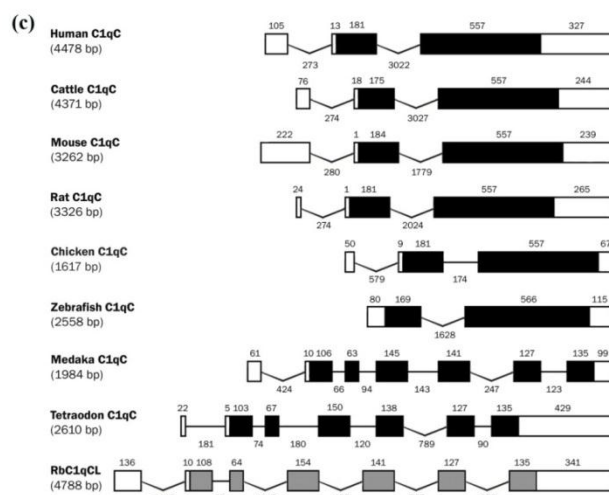
Figure 9. Comparison of genomic structure and organization of (a) RbC1qAL, (b) RbC1qBL, and (c) RbC1qCL among vertebrate species.
The genomic structures were obtained from the Ensembl database and included: C1qA of human (ENSG00000173372), cattle (ENSBTAG00000007153), mouse (ENSMUSG00000036887), rat (ENSRNOG00000012807), chicken (ENSGALG00000021569), zebrafish (ENSDARG00000044613), medaka (ENSORLG00000019405), and tetraodon (ENSTNIG00000006561); C1qB of human (ENSG00000173369), cattle (ENSBTAG00000011196), mouse (ENSMUSG00000036905), rat (ENSRNOG00000012749), chicken (ENSGALG00000004771), zebrafish (ENSDARG00000044612), medaka (ENSORLG00000017261), and tetraodon (ENSTNIG00000000825); C1qC of human (ENSG00000159189), cattle (ENSBTAG00000011193), mouse (ENSMUSG00000036896), rat (ENSRNOG00000012804), chicken (ENSGALG00000023605), zebrafish (ENSDARG00000095627), medaka (ENSORLG00000017257), and tetraodon (ENSTNIG00000006560).

2.4 Tissue distribution analysis of the RbC1qs

To determine the tissue-specific expression level of RbC1qs, qPCR was carried out using various tissues from healthy fish. Transcripts of all three RbC1qs were constitutively expressed in all tissues analyzed, and each C1q member displayed a similar expression pattern (Figure 10). The highest expression level was detected in spleen, followed by liver; the lowest expression levels were detected in muscle, skin, brain, intestine, and blood. *In situ* hybridization analysis also detected the strong expression of RbC1qs in spleen (Figure 12).
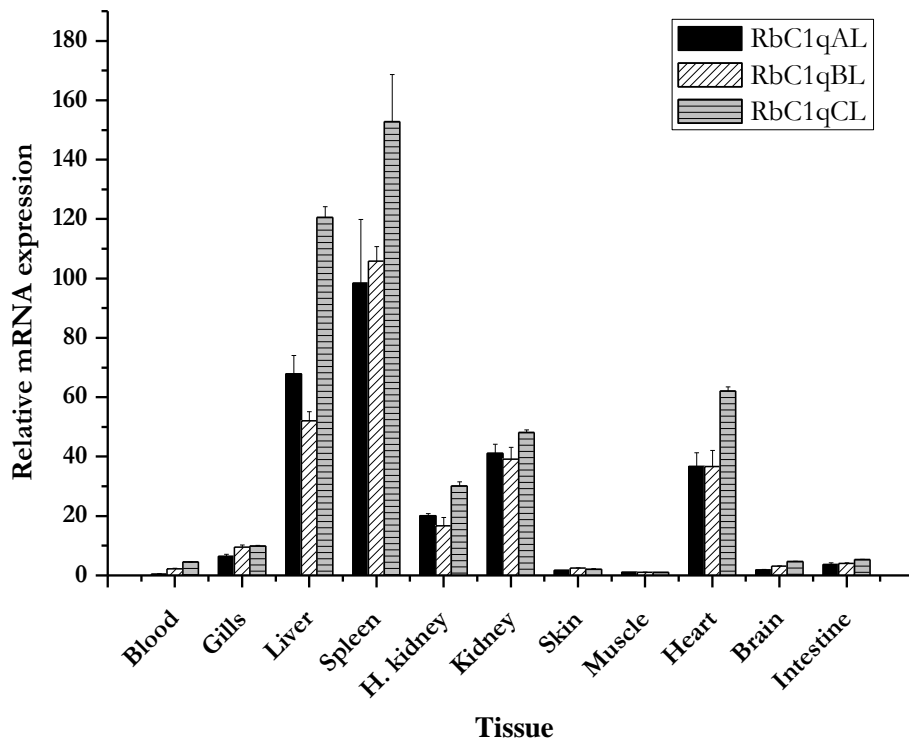
Figure 10. Tissue-specific transcriptional profile of RbC1q in healthy rock bream
The RbC1q mRNA detected in each tissue by qPCR was normalized to RbC1q mRNA detected in muscle. Data are presented as mean (n=3) ± SD.

2.5 Temporal expression analysis of RbC1qs after immune challenges

The transcriptional response of RbC1qs to bacterial and viral challenges were assessed in rock bream liver in order to determine the potential defense activity of these proteins against pathogen infection. Two live bacterial strains (*E. tarda* and *S. iniae*), a viral strain (RBIV), and a bacterial endotoxin (LPS) that are commonly encountered by rock bream in the natural environment were used as immune stimulants. All three RbC1qs showed similar expression patterns in response to each of the immune stimulants. At 3 h after the *E. tarda* challenge, levels of RbC1qAL, RbC1qBL, and RbC1qCL transcripts were significantly elevated

(*P*<0.05 vs. PBS control). Same time, the highest inductions were also observed respectively (8.8-, 11.2- and 14.9-fold, respectively), indicating an early phase response. After 24 h post infection (p.i.), the expression levels began to significantly decrease (Figure 11a). In contrast, the *S. iniae* challenge was led to up-regulation of RbC1qCL transcription at 12 h p.i., but of RbC1qAL and RbC1qBL transcription at 24 h p.i. (Figure 11b); compared to the *E. tarda* transcriptional response, the *S. iniae* response was mild. Collectively, these data showed that the RbC1qs actively respond to Gram-negative bacterium at the early phase of infection, as compared to Gram-positive species. The LPS, a major component of the Gram-negative bacterial cell wall, can induce the expression of number of immune-related genes in the host system. Induction of RbC1qs was examined after challenge with LPS. The highest expressions of RbC1qAL (1.3-fold), RbC1qBL (2-fold), and RbC1qCL (1.4-fold) were observed at 48 h after injection with LPS. Although significant down-regulation occurred between 6 h and 24 h p.i (Fig. 11c), the trends were similar among all three of the RbC1qs.
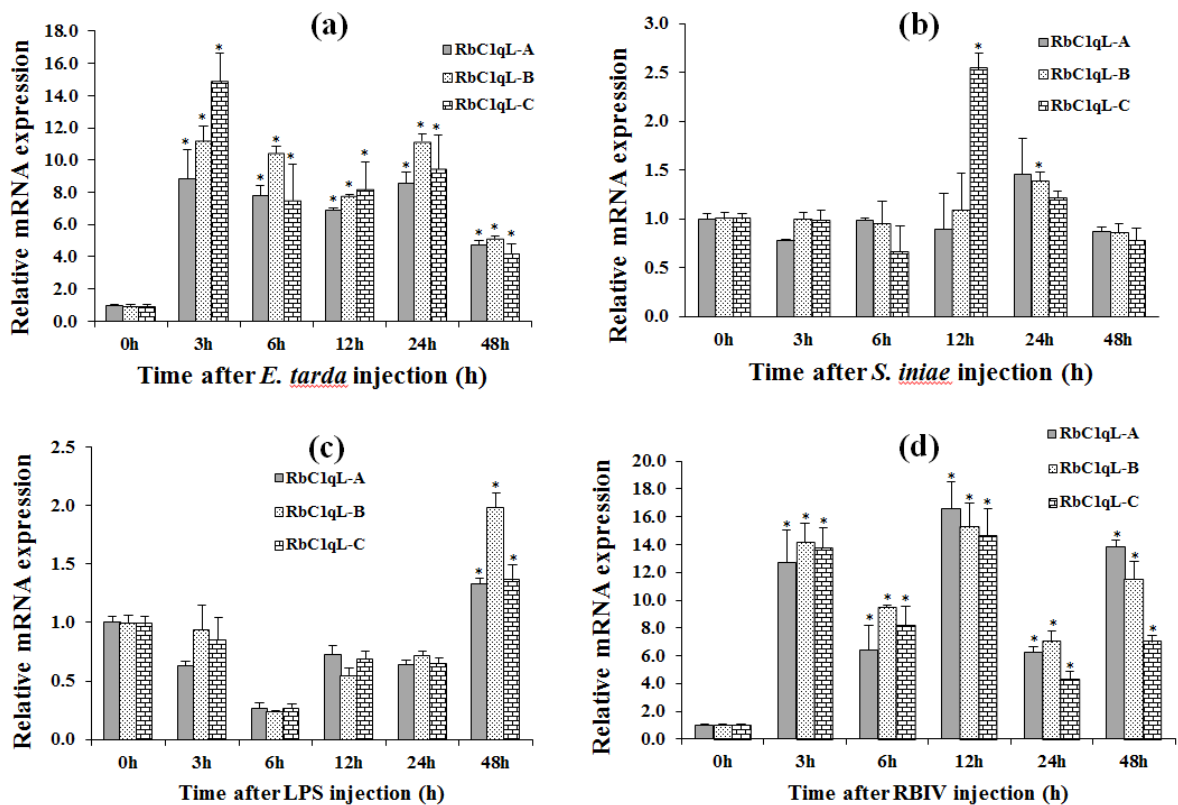
Figure 11. Temporal expression profile of RbC1q in liver after challenge with (a) *E. tarda*, (b) *S. iniae*, (c) LPS, and (d) rock bream iridovirus

The RbC1q mRNA detected by qPCR was normalized to β-actin expression. Data are expressed as mean fold-induction (n=3) relative to the PBS control ± SD. *t-test, $P<0.05$ vs. unchallenged control at 0 h.

When rock bream were infected with RBIV, the expression level of all three RbC1qs were significantly up-regulated in the liver within 3 h p.i. ($P<0.05$). However, the levels varied between 7- and 20-fold compared to PBS controls (Figure 11d).
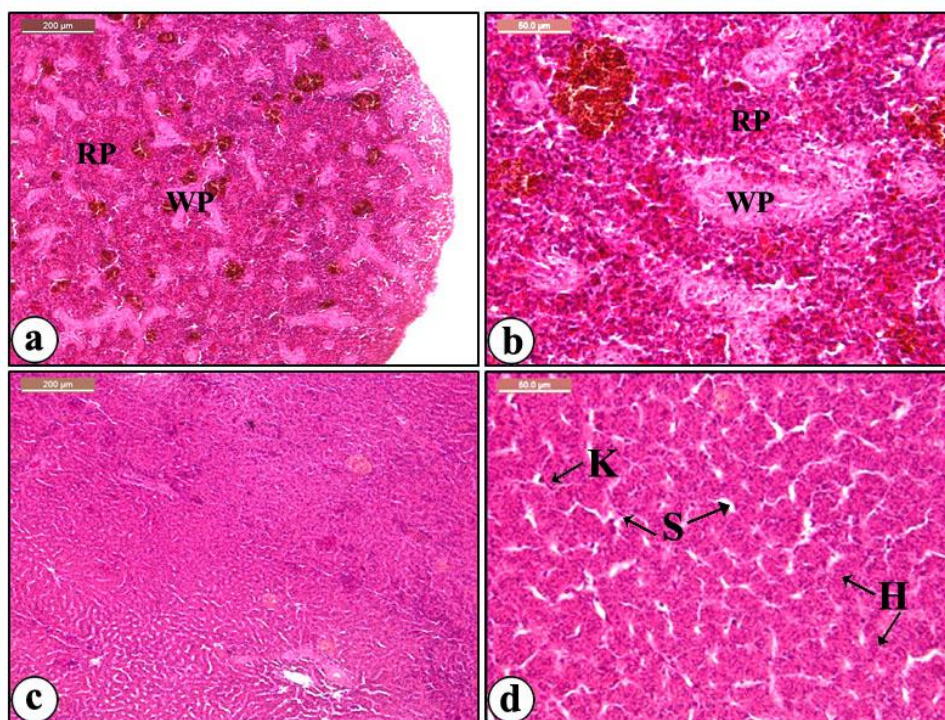
Figure 12. Histological examination of the spleen (a and b) and liver (c and d) tissues of rock bream.
RP: read pulp, WP: White pulp, H: hepatic cells, K: kupffer cell, S: sinusoid
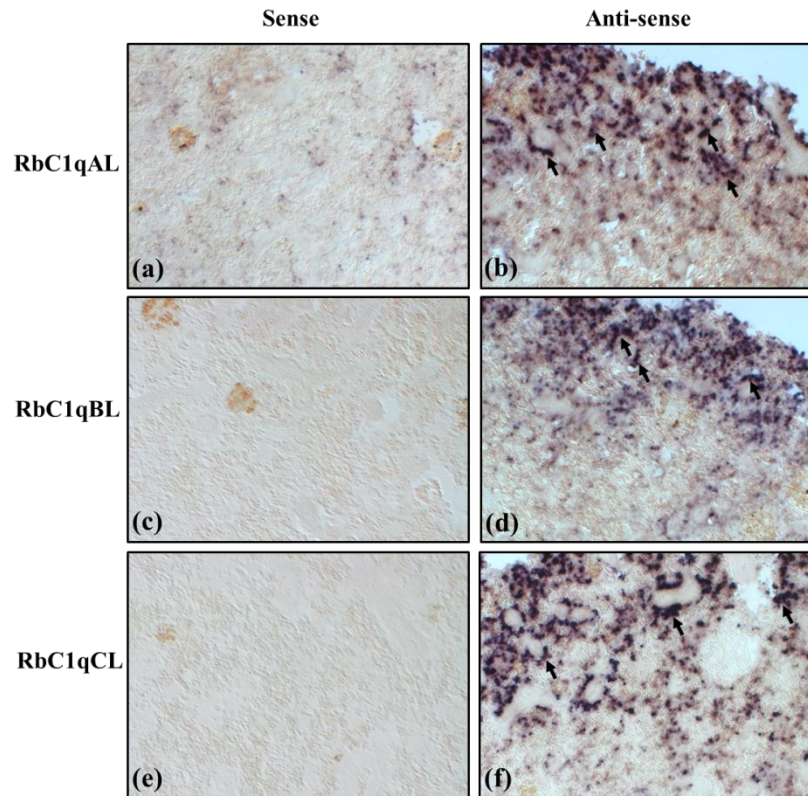
Figure 13. *In situ* hybridization detection of RbC1qs in spleen.
For each RbC1qs, DIG-labeled sense (a, c, e) and anti-sense (b, d, e) probes were used.
RbC1qs expressing cells were stained in dark purple and indicated by arrows.

3. Discussion

C1q, the first subcomponent of the C1 complex initiates the classical complement pathway in

mammalian complement system. The C1q is able to recognize various immune complexes

activated by pathogenic microorganisms such as viruses, bacteria and infected cells, which

subsequently resulted in proteolysis or phagocytosis.

Mammalian C1q is a hexamer, composed of 18 polypeptides that represent six C1qA,

C1qB, and C1qC chains. Each C1q polypeptide chain contains a collagen-like region (CLR)

and a C-terminal globular region known as the C1q domain (Sellar et al., 1991). Four

conserved cysteine residues are present in all C1q polypeptide chains identified to date. One

of these cysteines forms an inter-chain disulfide bond to generate A-B and C-C dimers, while

two others form an intra-chain disulfide bond. The fourth cysteine residue acts as a free thiol

46

group, which may interact with ligands. Several studies have been suggested that each C1q chain can bind independently to its preferred ligands (Kishore and Reid, 1999), thereby creating an exponential potential for different binding partners.

We have identified and characterized 3 C1q domain containing genes from rock bream. The characteristic features of SP, CLR and C1q domain were well identified in all 3C1q proteins. The CLR is composed of repeating triplets of Gly-X-Y (where X is often proline and Y is often hydroxylysine or hydroxyproline) (Bohlson et al., 2007), wherein the number of Gly-X-Y repeats can range from 14 to 100 (Innamorati et al., 2006). Gly-X-Y repeats were clearly recognized in the CLRs of all three RbC1qs, and the numbers of repeats were 23 in RbC1qAL, 24 in RbC1qBL, and 23 in RbC1qCL. This finding is similar to the number of Gly-X-Y repeats present in C1qA, C1qB, and C1qC of zebrafish and C1qL1 and C1qL2 of mandarin fish, which range from 20 to 25 (Hu et al., 2010; Lao et al., 2008). Experimental studies of human C1q have demonstrated that the CLR can interact with neutrophils and trigger the activation of NADPH, which will enhance the phagocytic activity via the production of microbicidal oxygen radicals (Ruiz et al., 1995). Later studies of the human mannan-binding lectin, which activates the CS via the lectin pathway, identified a specific sequence (GEKGEP) in the CLR that was responsible for stimulating phagocytic activity. In addition, several studies have reported that complement components are involved in phagocytosis of apoptotic cells (Gershov et al., 2000; Mevorach et al., 1998). Nauta et al.in 2002 (Nauta et al., 2002) demonstrated that the classical complement pathway could be initiated by binding of C1q directly to apoptotic cells through the globular head region. Therefore, we speculated that RbC1qAL and RbC1qCL may be involved in stimulating phagocytic activity in rock bream to eliminate pathogenic materials and/or apoptotic cells.

To study the evolutional mechanism of RbC1qs, we have compared the each amino acid sequences with other known vertebrate counterparts. The high identity was recognized

with fish species, i.e. RbC1qAL, RbC1qBL, and RbC1qCL presented the highest identities with *S. chuatsi* C1qL2, *S. chuatsi* C1qL1, and *O. niloticus* C1qC respectively. Whereas, the RbC1qs depicted very low identities with human C1qs. In the case of molecular modeling of each RbC1qs displayed similar pattern of topology as reported for the gC1qs of *H. sapiens*, *D. rerio*, *M. edulis*, and *B. cereus* (Carland and Gerwick, 2010; Hu et al., 2010). In addition, each RbC1qs depicted a free thiol group as shown in Figure 7. This has been previously demonstrated by *in vitro* study that the free thiol group in the gC1q region can interact with several immune complexes through IgG (Martin et al., 1990). However, teleosts do not possess IgG, and the presence of a free thiol group in all three of the RbC1qs may indicate the emergence of potential IgG-binding site in teleostean C1qs. Accordingly, we can speculate that the rock bream gC1qs have become evolutionarily altered to support an adaptive immune response. In addition, phenylalanine (F), glycine (G), and valine (V) were found to be completely conserved in the gC1q regions examined (Figure 6), suggesting that these residues may be essential to maintain the globular shape of the C1q.

According to the phylogenetic study, we recognized that fish group in each clade has clustered into three separate groups. Interestingly, RbC1qAL, RbC1qBL, and RbC1qCL clustered into the distinct fish groups of C1qA, C1qB, and C1qC, respectively. In addition, each zebrafish C1qs has evolved in separate sister clade in fish group. This information may indicate how the RbC1qs were conserved throughout evolution. Although the phylogenetic tree depicted a common ancestral origin for the three RbC1qs, RbC1qA and RbC1qC were categorized proximal to one another, while RbC1qB appeared to have evolved independently (Hu et al., 2010).

The genomic organization of C1qs revealed that fish C1qs have large number of exons (5-7) compare to other vertebrates (3-4) except in zebrafish. According to the genomic structural analysis of RbC1qs, it appears that the fish and other vertebrate C1qs may have

evolved via different evolutionary pathways, even though they originated from a common ancestor. The results from the phylogenetic analysis also support this premise.

The tissue specific expression of RbC1qs was investigated in different tissues and found that highest expression in spleen and liver. The C1 subcomponents are synthesized primarily in gut epithelium and a growing list of other cells (monocytes/macrophages, fibroblasts, endothelial cells, leukocytes, etc.) (Arlaud and Colomb, 2001). However, all the RbC1qs exhibited differential expression pattern to the general perception but similar to the other fish orthologous. In previous studies, C1q expression in zebrafish showed a very similar tissue distribution pattern (Hu et al., 2010), while mandarin fish exhibited the lowest expression level in liver (Lao et al., 2008). In humans, the main sources of C1q synthesis are myeloid cells, macrophages, monocytes, and dendritic cells (Rabs et al., 1986). Investigation of C1q expression in rat liver revealed that the immune-related Kupffer cells produced considerably higher amounts of C1q than other types of liver cells, such as the sinusoidal cells (Armbrust et al., 1997). Similarly, in mice, the interdigitating cells of the spleen, which are associated with high levels of monocyte-derived dendritic cells, exhibited high C1q expression (Lu et al., 2008). In this study, we have performed *in situ* hybridization technique to understand the localization of RbC1qs in spleen tissues and found that they highly expressed in white blood cells of white pulp region.

Temporal changes in RbC1qs expressions were analyzed using qPCR in bacterial and viral infected liver tissues, since other complement components were highly expressed in liver. Remarkably, all RbC1qs were exhibited similar expressional patterns. Strong early RbC1qs expressions were observed after Gram negative *E. tarda* infection, while recognizing mild expressions against Gram positive *S. iniae*. Previous studies of *Mytilus galloprovincialis* (Gestal et al., 2010), *Chlamys farreri* (Zhang et al., 2008), and *Argopecten irradians* (Kong et al., 2010) also demonstrated that the immune responsive expression of C1q occurred in the

49

early phase (1-6 h) of infection with Gram-negative and Gram-positive bacteria. Recently, a study of *S. iniae* infection in *D. rerio* revealed that C1q expression becomes significantly elevated at the early stage (within 24 h p.i.) of infection (Carland et al., 2012). The expressions of RbC1qs demonstrated strong early responses against RbIV infection compare to the PBS control. A previous study of C1q expression in mandarin fish showed no significant response to challenge with the infectious spleen and kidney necrosis virus in liver, but remarkable up-regulation in spleen (Lao et al., 2008), where endogenous expression is highest. Moreover, a study of C1q expression in monkey found up-regulation in the early asymptomatic stage of SIV (Simian Immunodeficiency Virus) infection (Depboylu et al., 2005). Although having contradicting observations, we can suggest that the RbC1qs possibly involve into the immunological responses beside bacterial and viral infections. In the case of LPS challenge, RbC1qs demonstrated a mild up regulation at 48h p.i. These results further indicate that the transcription of RbC1qs varies upon exposure to purified LPS and intact *E. tarda* bacteria, although the LPS is a component of the Gram-negative bacterial membrane. Moreover, amphioxus C1q has been reported as significantly induced by LPS at 48 h p.i. ($P<0.05$) (Yu et al., 2008). Experimental studies in mice have revealed that LPS can stimulate the maturation of dendritic cells, which will in turn reduce the expression of C1q (Armbrust et al., 1997; Castellano et al., 2004). In addition, C1q binding to purified LPS has been shown to be inefficient, while its binding to two other bacterial outer membrane proteins, presumably porins, was shown to be efficient and capable of initiating the classical complement pathway (Alberti et al., 1993). Such inefficient binding may explain the observed down-regulation of RbC1qs in the early phase of challenge with LPS. However, further studies need to be carried out to determine the C1q-related mechanisms of LPS recognition in fish complement system.

C1q is known to activate the classical pathway by binding with its ligands in antibody complexes, which represents a major link between the innate and adaptive arms of immunity (Kishore and Reid, 1999). Nevertheless, in some lower vertebrates, such as lampreys, and invertebrates, C1q has been shown to be dispensable for activation of the classical complement pathway. Due to the lack of antibodies, such as Ig, related to the adaptive immune response, it is possible that C1q may act as a lectin to stimulate the innate immune response (Matsushita et al., 2004). Previous studies have been reported that the classical pathway was originated in fish. Moreover, some studies have shown that fish are the oldest species, which have antibody-dependent (IgM) activation of the complement pathway (Boshra et al., 2006; Nonaka and Smith, 2000). Recent studies in zebrafish demonstrated involvement of C1q in the classical complement pathway (Hu et al., 2010); however, antibody-independent binding of C1q with a wide range of pathogens, including Gram-negative and -positive bacteria, viruses and parasites, has been reported in mammals. Human C1q has been shown to directly bind various outer membrane proteins and porins of Gram-negative bacteria, such as *Legionella pneumophila*, *Escherichia coli, Salmonella typhimurium, Salmonella Minnesota*, and *Klebsilla pneumonia* (Alberti et al., 1993; Mintz et al., 1995). Yet another study revealed that the C1q binding to OmpK36 porin of *Klebsiella pneumonia* was mediated by the globular region and showed that the binding site of OmpK36 was identical to that of IgG (Kojouharova et al., 2003). Several mammalian studies of virus infections have demonstrated C1q-mediated neutralization of the viral pathogens, including Human Immunodeficiency Virus (HIV), Human T Lymphotrophic Virus-1 (HTLV-1), Murine Leukemia Virus (MuLV), Herpes Simplex Virus (HSV), and Cytomegalovirus (CMV) (Thielens et al., 2002). Further studies of HIV, HTLV-1, and MuLV indicated that the C1q molecule directly interacts with the viruses to trigger the classical complement pathway (Mehlhop and Diamond, 2006). In the current study, expressional kinetic analysis of all three

RbC1qs was demonstrated a clear early phase up-regulation in response to experimental injection of *E. tarda* and RBIV. Our findings suggest that RbC1q likely interact with live Gram-negative bacteria and virus through an antibody-independent mechanism to mount early-phase immune responses. However, further studies are required to achieve a better understanding of the rock bream defensive mechanism against microbial pathogens and involvement of RbC1qs in the classical complement pathway.

In conclusion, we have isolated and characterized the full-length cDNA and genomic sequences of three complement component 1q genes, the subcomponents of the C1 complex in the classical complement pathway of rock bream. Structural and phylogenetic evidences suggest that these newly-identified molecules are three distinct members of the C1q family. Expression analyses revealed that while the RbC1qs are constitutively expressed in various tissues, they are most strongly expressed in myeloid-rich tissues. Furthermore, the responsive expression pattern of RbC1qs upon bacterial and viral challenge suggests that these proteins may play a major role in immune defense of rock bream.

Chapter II

Molecular and functional level characterization of complement C2, C3, C4 and C5 as execution phase components of the complement system

Part 1: Genomic structural, transcriptional and promoter level characterization of complement component 2 from rock bream *Oplegnathus fasciatus*

Abstract

With respect to the complement activation, factor B serves as the catalytic subunit of C3 convertase in alternative pathway whereas C2 plays the same role in classical pathway. Therefore, complement C2 is one of the key molecules, which make C3 convertase (C4b2b) in complement activation. In this study, homolog of complement component 2 was identified from rock bream (*Oplegnathus fasciatus*) and characterized at its transcriptional level. The full-length cDNA of rock bream C2 (*Rb-C2*) was identified from the cDNA library and its genomic sequence was obtained by screening and sequencing of rock bream BAC genomic DNA library. The complete genomic DNA of rock bream C2 was 7270 bp in length that was consisted of 19 exons interrupted by 18 introns. The open reading frame (ORF) of this gene was found to encode a polypeptide of 779 amino acids. The predicted molecular weight of the *Rb-C2* polypeptide was 87KDa and theoretical iso-electric point was 6.2. *Rb-C2* possessed three Sushi/CCP/SCR domains, a von Willebrand Factor A domain (vWFA), and trypsin family domain, which are known to be important for the functions of C2 in vertebrates. The amino acid sequence of *Rb-C2* showed 69.7% and 59.2% identity with C2 counterparts of *Oreochromis niloticus* and *Oncorhynchus mykiss*, respectively. In tissue expression profile, *Rb-C2* transcripts were observed to be constitutively expressed in all the tissues whilst highest was observed in liver. In challenge experiments, *Rb-C2* transcripts were significantly up-regulated in liver and spleen tissues post challenge with *Edwardsiella tarda*, *Streptococcus iniae*, an endotoxin lipopolysaccharide and rock bream iridovirus. Collectively, findings of this study support to suggest that *Rb-C2* may play a significant role in host antibacterial and anti-viral defense through the activation of complement system in rock bream.

1.1 Introduction

The complement system is activated by three different pathways. It is initiated by the C1 complex in the classical pathway and MBL along with MASP-1/MASP-2 in lectin pathway. In case of complement activation, factor B serves as the catalytic subunit of C3 convertase in alternative pathway. However, in the classical pathway, C4 is catalyzed by the C2 resulting C4bC2b (C3 convertase) and C2a as products. Furthermore, C2 acts as catalytic subunit of the C5 convertases (Shen et al., 2012b). In accordance with the literature, the complement C2 and Factor B in mammals were evolutionally arisen from common ancestor; hence, they share extensive homology at protein level. Moreover, they encode very much similar genomic structural organization, which are located on the same chromosome (Nonaka et al., 1998; Sunyer et al., 1998a). Additionally, both genes resembles the class III region of mammalian major histocompatibility complex (Carroll et al., 1984). The high identity and similarity between factor B and C2 genes in many teleost species have become critical for characterization without any differences in their biochemical activities. Further, functional evidence at the protein level will be required to determine whether these genes represent factor B (Bf) or C2; which would provide an inference to the time of divergence of Bf and C2 (Shen et al., 2012b). The complement C2/Bf is recognized as one of the mosaic protein composed of sequence repeats and a vWFA along with a serine protease domain (Zhong et al., 2012).

Up to date, several studies on C2 molecules have been conducted in fish including medaka, common carp, zebrafish, nurse shark, rainbow trout, catfish, large yellow croaker and grass carp (Kuroda et al., 1996; Nakao et al., 1998; Nakao et al., 2002; Seeger et al., 1996; Shen et al., 2012b; Shin et al., 2007; Sunyer et al., 1998a; Wei et al., 2009; Zhou et al., 2012). In this study, Rb-C2 gene was cloned and characterized at genomic and transcriptional levels.

Table 6. Description of primers used in this study

Name	Purpose	Primer Sequence (5'-3')
Rb-C2 -1F	qPCR amplification and BAC screening	CGAGGTGCAGTTGAAGTTCCATGT
Rb-C2-1R	qPCR amplification and BAC screening	TGACGTCGTCGATGCTACCACTTA
Rb- C2-3F	Promoter sequence 1	GAGAGAggtaccAGCTGTTTCTACATGAGAGAACGTTCAAGATG
Rb-C2 -4F	Promoter sequence 2	GAGAGAggtaccGCAGAGCATCCATCGACGCAATC
Rb- C2-5F	Promoter sequence 3	GAGAGAggtacc CTTGCATGCCCAAACCTGAACGA
Rb-C2 -6F	Promoter sequence 4	GAGAGAggtaccTGAGGTTTGGTTTGGGACTGGTCA
Rb- C2-7F	Promoter sequence 5	GAGAGAggtacc AAGTAAGAGCTCAAATCTGT GTTCTGTCCTA
Rb- C2-8R	Common reverse primer Promoter sequence	GAGAGActcgagTCTATATCGGTGTTGTTTGTGTCTCTGAGTGA
Rb-β-actin-3F	qPCR reference gene amplification	TCATCACCATCGGCAATGAGAGGT
Rb-β-actin-3R	qPCR reference gene - amplification	TGATGCTGTTGTAGGTGGTCTCGT

1.2. Results

1.2.1 *In silico* characterization of Rb-C2

Based on BLAST search analysis of EST data from our previously established rock bream normalized cDNA library, we identified a putative complement component 2/Bf and designated as *Rb-C2*. The complete sequence information were deposited in NCBI GenBank database under the accession number, KF744234. The complete cDNA sequence of *Rb-C2* was 2918 bp in length, including a 233 bp 5'-untranslated region (5′-UTR), a 2340 bp open reading frame (ORF) encoding 779 amino acids, and a 885 bp 3'-UTR with a polyadenylation signal. The predicted molecular mass of *Rb-C2* was 87 kDa and the theoretical isoelectric point was 6.2. According to SignalP prediction, the *Rb-C2* possesses a signal peptide of 22 residues, which is presumed to facilitate the extracellular localization. When consider about domain architecture of Rb-C2, it was consisted of three Sushi/CCP/SCR domains, a von

56

Willebrand Factor A domain (vWFA), and trypsin family domain, which were known to be important for the functions of C2 in vertebrates.

1.2.2 Characteristic features of 5' flanking region and genomic structure of Rb-C2

The genomic sequence of *Rb-C2* was identified by PCR-based screening of the BAC genomic library using *Rb-C2* gene-specific primers followed by sequencing of putative candidate clone using GS-FLX strategy. The putative genomic architecture of *Rb-C2* was identified through the comparison of the obtained genomic DNA sequence with the full-length cDNA sequence. Thereafter, the exon and intron structure was compared with the genomic structures of several other vertebrate species published in the Ensembl genome database. The *Rb-C2* genomic sequence is about 7230 bp in length and is composed of 19 exons interrupted by 18 introns. The 5'- and 3'-ends of all introns harbor canonical splicing motifs (according to the AG-GT rule). The numbers of exons are similar to that of human C2. However, all the other species considered in the comparison were consisted only 18 exons except zebrafish and tilapia. Nevertheless, the biggest genomic sequence was observed in zebrafish (21039bp) whereas tilapia C2 shows the smallest with in the considered species (Figure 14)
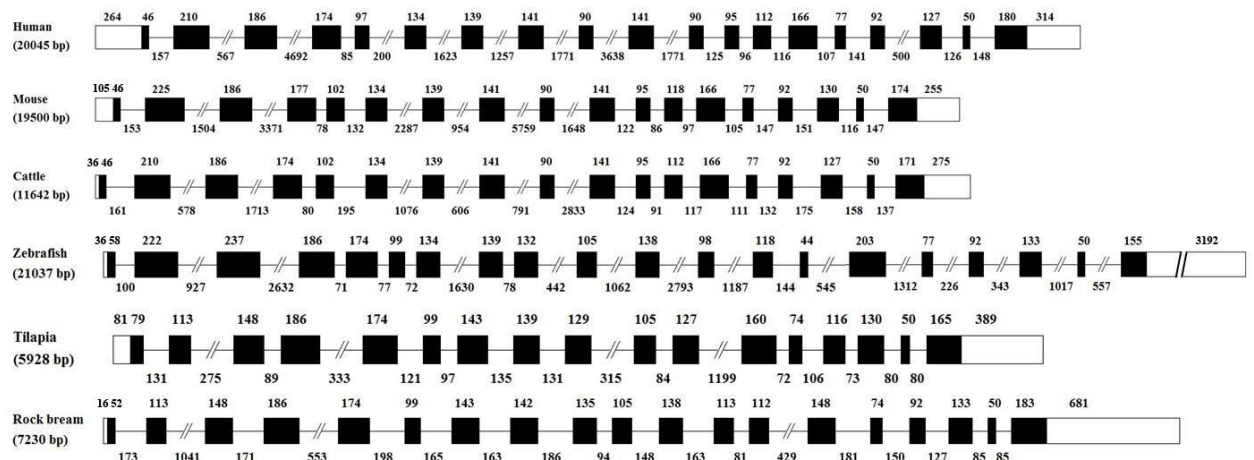
Figure 14. The genomic structural comparison of Rb-C2 with other orthologous

Exons are represented as black solid boxes. Un-translated regions are shown as light-colored boxes. Introns are represented as thin lines in-between the exons. Introns larger than 200 bp are indicated by interrupted lines. Numbers above the boxes and below the lines indicate the lengths of exons and introns respectively. Selected organisms are *Homo sapiens* (Gene ID: 717), *Bos taurus* (gene ID: 515440: XM_005223654.1), *Mus muscalus* (NM_013484.2), *Danio rerio* (XP_001921532.4), and *Oreochromis niloticus* (XP_003449339.1)

According to the Berkeley Drosophila Genome Project prediction analysis (BDGP), the putative transcription initiation (+1) site was identified at 233bp upstream of the translation start site (ATG). Putative TATA box was identified at 23 bp upstream of the +1 site. The promoter structure of *Rb-C2* is composed of multiple transcription factors binding sites including immune related and complement specific transcriptional regulatory factors. Among those identified sites, putative binding sites for transcription factors related to immune signaling were highlighted in Figure 15 and these include the CCAAT-enhancer binding protein (C/EBP), activator protein-1 (AP-1), octamer transcription factor (OCT-1), specificity protein-1 (SP-1), hepatocyte nuclear factor-3 (HNF-3), nuclear factor-1 (NF-1), and GATA-1.

```
TTATAAATATCTGAGGAGTCTGAGTACACTGCAACTTGGCTGAGGTCAGACATCTTTAGTCCGCTGTCAC   -947

TCTTTCAGGGCTCCTCTCCTCAAAGAGCCTGATGTTGTGATAAATACATCAGCGGTGACATTGCGAAGAG   -877
         SP-1                                              C/EBPα
                          -876
AGCTGTTTCTACATGAGAGAACGTTCAAGATGTTTTCCTGGGTGGTTTTCTTAACATTGTTCACTTTTTG   -807
                                                    C/EBPα    C/EBPα HNF-3b
CTCCAGCAGTAAACTTCCATTGAGGTGTAGTAAGCTTGCAGCCTTCACCCCTCCAGCACAGCAGAGCATC   -737
    -746         OCT-1
CATCGACGCAATCCAGATAAAAATGAAAAAAATAACACAATACACTTCACACCAAAGAATCTTCAAACTC   -667

CAGTTATCAAAATATCCATGTTTTCTGCTATTCTTCTGCCTTTTGCCTATTTGATCAGCAGCGACCTCTA   -597

TTTCTCTTGTTTTTGATGGTGTGTGAGAGGTTACAAAAGAGTTTTAGAAACTCCTTCACAGTGTTGAACA   -527
  C/EBPα
AAGCCGAGTCAGCTGAGAGGACTGAGCCCCAGGGAGGAGGAGGGAGGGCGGGGAAGGGGAGAAGAGGTGA   -457
      AP-1            YY-1         SP-1          SP-1
TGGTGGAGGTGGATTTGGGGGTCCCTGTTAAATCTCTATAATCCCTATAATGATCTTTTAAGCTCCTAAA   -387
           AP-2α          -366            OCT-1
ATGTTTTATTCACACGCACGCTTGCATGCCCAAACCTGAACGAGCAGGTGCACAAACACACAAATAGAAA   -317

TACACACCATTTTAAACCCCAAAACATTGGCACAGAGACAGTCTGAGTCATTTAGTAATAATTCTGGTTA   -247
   YY-1             C/EBPα                         C/EBPα
ATTAGACATAAAAACAACAACAGTGTTGGGGTGAGGTTTGGTTTGGGACTGGTCATTAATAGAGAGAAGT   -177
        GATA-1                    -215
GTTAAAGATATTGTGAGATTTGAAATTGCATTAGTTGCCCAAAGCTGGCTCATGACATTGCAGTTGAGTG   -107
               C/EBPα                    NF-1
TCCTAATAGACTTTTATTTAAAAGCTTTGGAGGAGCTTTTATGAGGCCACTCTTGGCTTTAAATGAACAA   -37
           OCT-1         SP-1                                    HNF-3
ATAACGTAAAAATGTCTCCAAAGAAACACTTGATTCACAGAATCCTAAAAAAAAAAAAGAAGAGAGAGAA   34
    TATA BOX   C/EBPα              +58
ACAGATCAAATGACCAAAAATGAAAGTAAGAGCTCAAATCTGTGTTCTGTCCTAAATTGCTCAATCCCTC   104
        C/EBPα                                              C/EBPα
CTTCCTTGTTTGGTGCTGCTGTGTTAATGGGTAATCAGTCATCAAATAAACTGTTGACTTCGGGTTGATA   174
               Common reverse primer    AP-1        HNF-3b
CAGACACTTGCTGTTTCTGTCACTCAGAGACACAAACAACACCGATATAGAGTCTGGTCATGTATATCAA   244
                                                        M  Y  I  K
```

Figure 15. 5' flanking region of Rb-C2

Putative transcription factor binding sites are shown in bold and underlined forms with the name of the corresponding factor indicated directly below the sequence. The bent arrow indicates the putative transcription initiation site. The selected promoter construct sites are represented by ash color boxes with bent arrow. Translated amino acids (uppercase) and start codons (italicized uppercase, in bold phase) are also shown in the figure.

According to the results of luciferase reporter assay, the highest activity was observed with respect to the -746 region and the least activity was observed regarding +58 region, which is 5' UTR region of the Rb-C2 promoter. The observed activity was similar corresponding to the both -215 and -366 regions. Therefore, we can intimate that YY-1 site and two C/EBPαsites are not significantly important in driving the basal transcription of *Rb-C2*. However, according to the detected activity, sites which were identified in between -366 to -746 region

59

(mainly including SP-1, OCT-1, and Ap-1) were found to be important in transcriptional

regulation of *RbC2*
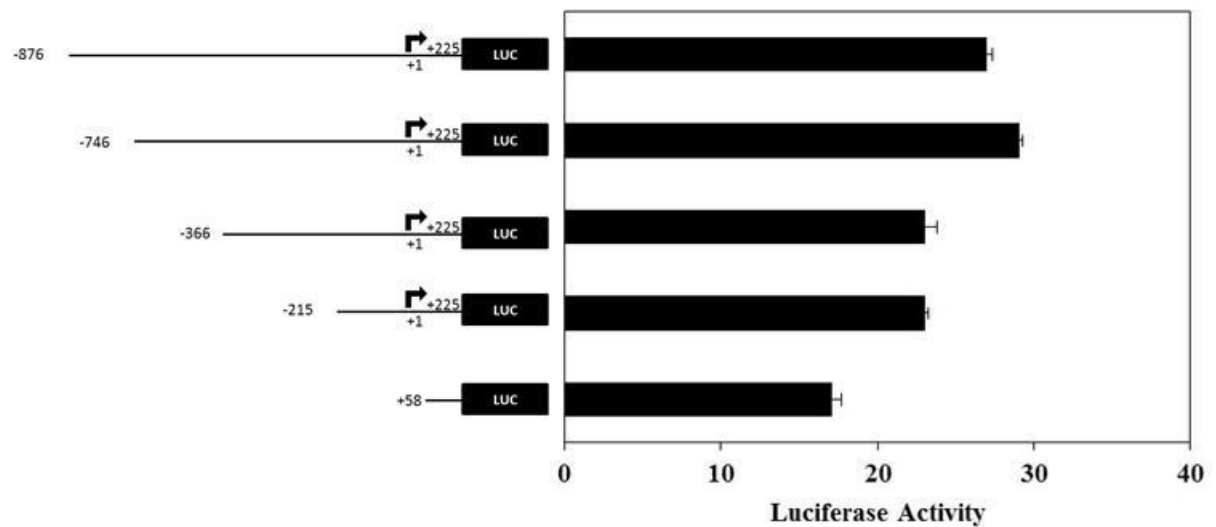
Figure 16. Promoter activity of the 5' region of the Rb-C2 gene in HepG3 cells
Transcriptional activity of promoter constructs was studied in the human hepatoma HepG3
cell Line. The cells were lysed and assayed for luciferase activity. The transcription
efficiency and background activity of pGL2 basic promoterless vector were also tested. The
relative activities are shown with the corresponding promoter constructs respectively.

1.2.3 Identity, similarity, and phylogenic relationship of Rb-C2

According to the pairwise sequence comparison, *Rb-C2* exhibited 69.7% amino acid identity

with its orthologous from Nile tilapia (*Oreochromis niloticus:* XP_003449339.1) and 66.5%

identity with its counterpart from Yellow croaker (*Larimichthys crocea*: GenBank ID;

ACS83542.1). However, compared to the other homologues of C2, Rb-C2 shares lesser

homology with its closer counterparts. For instance, it shows only over 30 % of identity with

mammals considered in the comparison (Table 7). In multiple sequence alignment, Rb-C2

showed higher degree of conservation among other vertebrates. More than 98% of cysteine

residues were found to be conserved across almost all of its orthologous, and certain

60

(presumably important) signature motifs were also found to be highly conserved among all of

the orthologous, considered herein (Figure 17).
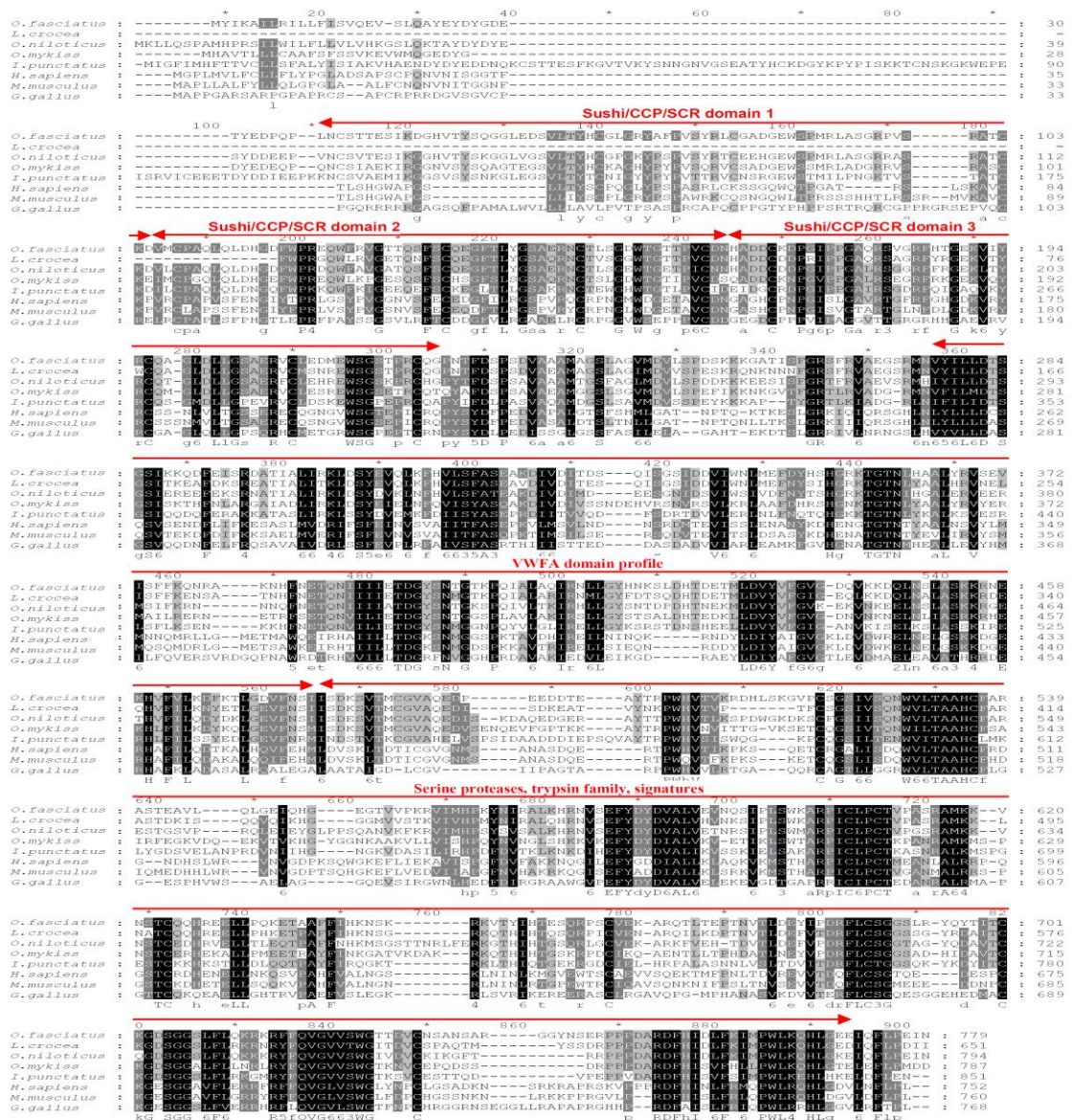
Figure 17. Multiple sequence alignment of Rb-C2
Identical residues among all orthologous' sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double-arrows. The GenBank accession number of each gene is presented under the C2 sub-cluster next to each species in Figure 18.

61

Table 7. Percentage of identity, similarity and gap of Rb-C2 with other C2 orthologous

Common Name	Species Name	Accession No.	AA	Identity %	Similarity %	Gap %
Nile tilapia	*Oreochromis niloticus*	XP_003449339.1	794	69.7	79.4	4.4
Yellow croaker	*Larimichthys crocea*	ACS83542.1	651	66.5	74.6	16.9
Rainbow trout	*Oncorhynchus mykiss*	BAB19788.1	787	59.2	73.7	5.7
Channel catfish	*Ictalurus punctatus*	NP_001244044.1	851	48.1	63.7	11.2
Common carp	*Cyprinus carpio*	BAA34707.1	833	47.9	63.0	13.2
Grass carp	*Ctenopharyngodon idella*	AFD18593.1	837	45.9	63.5	8.3
Nurse shark	*Ginglymostoma cirratum*	AAY56127.1	757	37.1	53.3	8.2
Human	*Homo sapiens*	NM_000063.4	752	36.7	54.0	9.1
Mouse	*Mus musculus*	NP_038512.2	760	36.3	53.6	8.1
Wild boar	*Sus scrofa*	NM_001101815.1	752	36.1	53.6	11.9
Cattle	*Bos taurus*	NP_001029664.1	750	35.2	52.3	12.4
Chicken	*Gallus gallus*	XP_417722.3	768	31.4	47.8	17.1

The evolutionary relationship was studied using the deduced amino acid sequences of Rb-C2 with sequences available in the NCBI database for the complement C2, C3 and C4. In phylogenetic tree, Rb-C2 was located within the fish clade and separated from other higher vertebrate C2 genes. Furthermore, C2 group was shown loosely associated with complement C3, C4 and C5 genes, which indicates the remote association with those genes (Figure 18).
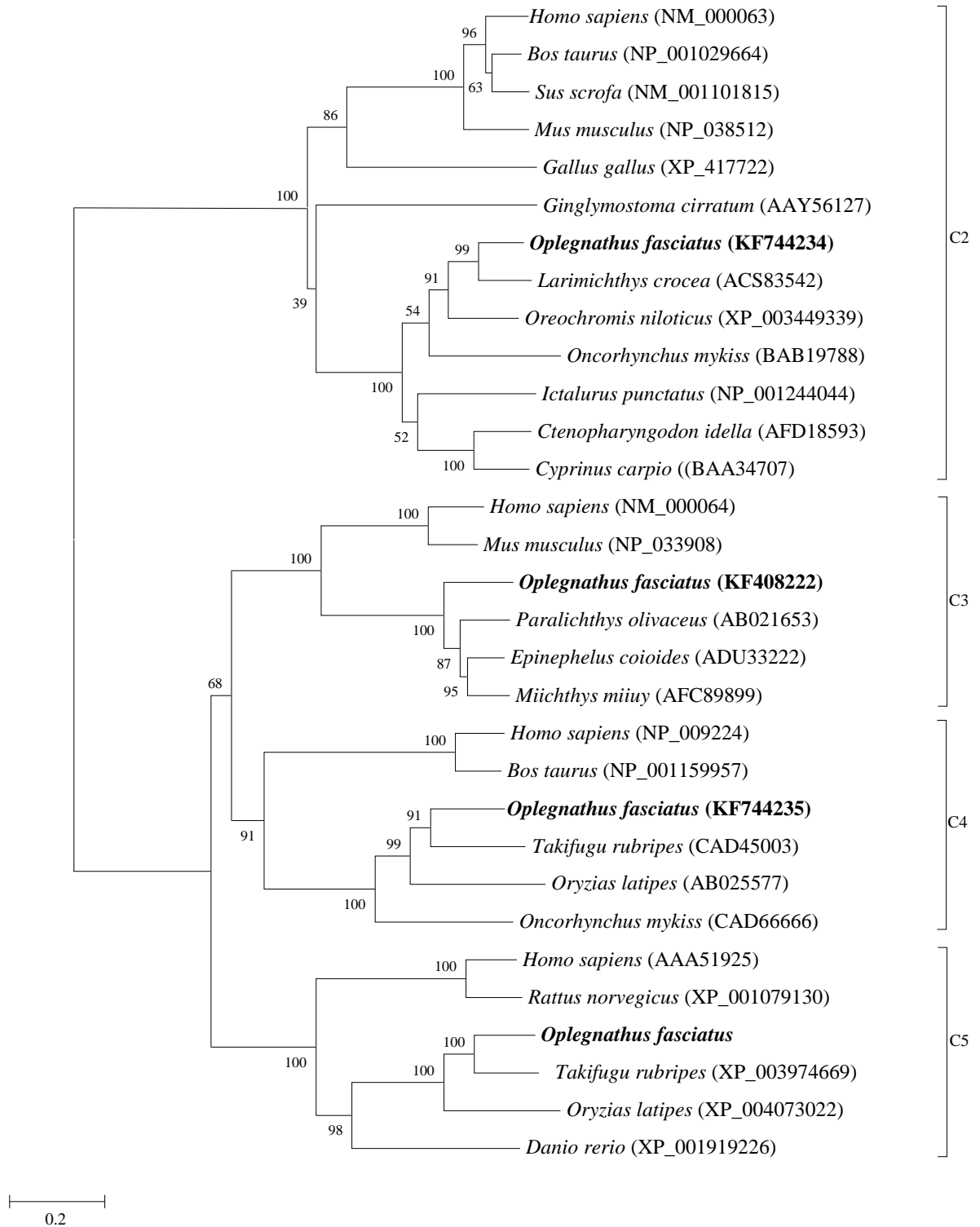
Figure 18. Phylogenetic analysis of Rb-C2, Rb-C3, Rb-C4 and Rb-C5 with their orthologous

The phylogenetic tree was constructed based on alignment of the full-length amino acid sequences. The numbers at the branches denote the bootstrap values from 5000 replicates. The GenBank accession numbers are shown within brackets next to each species

63

1.2.4 Tissue-specific mRNA expression of Rb-C3 in normal healthy fish

To examine the normal tissue distribution pattern of *Rb-C2* transcription, qPCR was carried out for blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine (Figure 19). *Rb-C2* was found to be constitutively expressed in all tissues examined, and showed highest in liver as in other complement components.
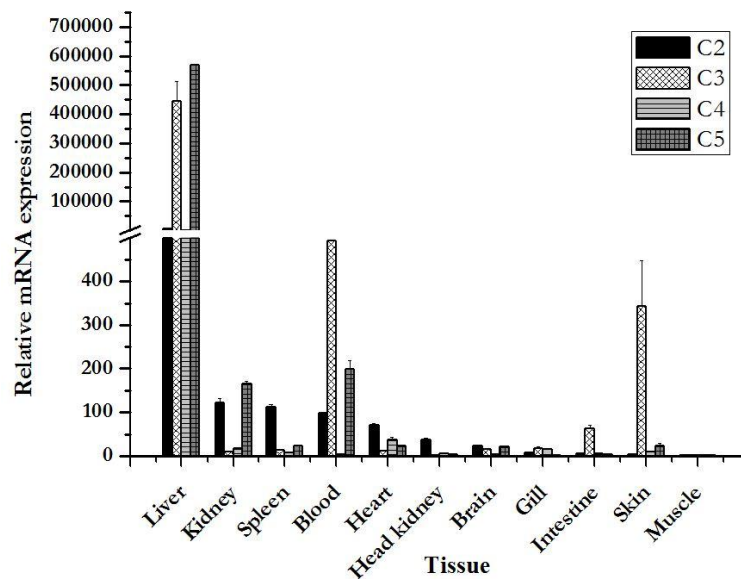
Figure 19. Comparison of tissue-specific expression of rock bream C2, C3, C4 and C5 transcripts by qPCR
The relative mRNA expression in peripheral blood cells, gills, liver, heart, spleen, intestine, head kidney, kidney, skin, muscle, and brain were calculated using β-actin as the reference gene. The bars represent the standard deviation (n = 3). Significant (fold) differences were calculated with respect to the relative expression in muscle.

1.2.5 Transcriptional modulation of the Rb-C2 in relation to immune stimulation

The mRNA levels of *Rb-C2* were measured in liver and spleen from rock breams challenged with *E. tarda, S. iniae*, RBIV, LPS and poly I:C by qPCR (Figure 20). The expression pattern upon Gram-negative bacteria *E. tarda* was displayed early response in liver. The amount of *Rb-C2* transcripts were significantly (*P*<0.05) increased from baseline (0 h) between 3 h and 6 h, and then declined to baseline at 12 h (p.i.), where the maximum level

was detected at 6 h p.i. (Figure 20). The biphasic patterns of induction were observed for the

S. iniae, iridovirus and poly I: C challenges. The transcriptional levels were increased 3 h to 6

h in each challenge treated and it was again reached to the baseline at 12 h. Then, it was

continued to increase until 48 h. When challenge with LPS, the Rb-C2 transcript level was

significantly unregulated at 24 h (p.i); in the meantime showed the same pattern as in other

challenges (Figure 20). In spleen tissue, transcriptional responses were observed in late phase

compared to the liver tissue. In all challenges, significant up-regulations were observed at 24

to 48 h p.i while the highest was reached at 24 h except in *S. iniae* challenge, which was

showed highest peak at 48 h.
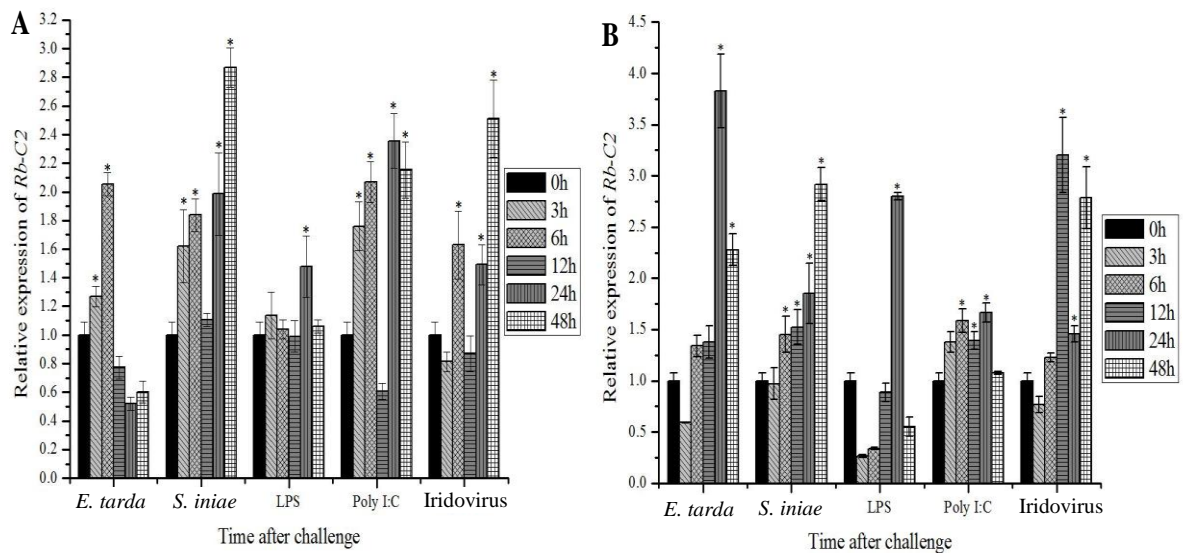
Figure 20. Relative expression of Rb-C2 mRNA in (A) liver (B) spleen after challenge of rock bream with *E. tarda*, *S. iniae*, LPS, and iridovirus, as detected by qPCR
The amount of mRNA relative to the internal control, β-actin, is expressed as mean ± SD (n = 3). Significant differences ($P < 0.05$ vs. untreated control at 0 h) are marked with asterisks.

1.3 Discussion

In this study of rock bream, we interested in identifying genomic structure of the complement

C2 and its unique biological features of active domains, transcriptional evidences for the

involvement in immune mechanism of rock bream in contrast to the mammalian counterpart.

Furthermore, as a representative gene of teleosts, we tested the promoter activity of *Rb-C2* using luciferase promoter assay.

The second component of complement system is a multi-domain serine protease that facilitates to form C3 convertases and C5 convertases. According to the past studies, the complement C2 is much similar to complement factor B in higher vertebrates. However, it seems to be different in number of amino acid in fish C2 and factor B. The identity and similarity may be very similar between fish isoform to that of human C2 or factor B. In rock bream, BAC screening was resulted a clone that resembles to the mammalian C2 than that of factor B.

Only one genomic structural study has been performed in the complement C2 gene in teleost. Therefore, this may be the second comprehensive study on genomic structure of complement C in teleost. The genome structure of the C2 genes quite varies among species considered in the comparison. Rb-C2 consists of 19-exon structure, which is similar to that of human. However, the size is one third of the size of Human. In other teleosts, C2 genomic architecture may contain variable numbers of exons.

The thioester site is highly reactive and readily hydrolyzed by water (Blandin and Levashina, 2004). In native C3, the thioester bond in α chain is protected within hydrophobic molecules that exposed only when C3 cleaved into C3b by C3 convertases (Sahu and Lambris, 2001). Upon activation of C3 by the proteolytic enzymes, thioester is transformed into free thiolate anion and form acyl-imidazole, which may be stabilized by Glutamic acid (Dodds et al., 1996; Gadjeva et al., 1998). This configurational change in the molecular structure is essential for binding ability of the active C3 into target organism (Law and Dodds, 1997). Rb-C3 also has the thioester site located between anaphylatoxin domain, MG8 and CUB regions. Tertiary structural arrangements of the Rb-C3 indicate that its thioester site may also be protected by the above domains (Figure. 5). Once C3a cleaved and

released, thioester site is exposed for hydrolysis. The β-α processing motif (RXXR) of human and some teleost are having four Arginine residues (RRRR) (Janssen et al., 2005; Zarkadis et al., 2001). However, Rb-C3 has three Arginine residues and one Lysine (RRKR), which is similar to spotted wolffish (Abelseth et al., 2003). However, in orange spotted grouper, C3 β-α processing motif possesses RKKR (ADU33222.1). The evidence above suggests that β-α processing motif in fish C3 gene may substitute Arginine with Lysine.

Past studies have shown that C2 is expressed in wide range of tissues and different life stages, such as juveniles and adults, in vertebrates (Shen et al., 2012b). The tissue-distribution study of *Rb-C2* showed that the liver was the main supplier of C2. Furthermore, liver-specificity is one of the unique features of complement genes in vertebrates (Morgan and Gasque, 1996). The earlier studies on rock bream complement lytic pathway genes including C7, C8α, C8β, and C9 also showed liver-specific expression, which provides further evidence to confirm the tissue specific expression patterns of complement genes (Wickramaarachchi et al., 2012; Wickramaarachchi et al., 2013a; Wickramaarachchi et al., 2013b). However, extra-hepatic cells in mammals, such as macrophages, monocytes, fibroblasts, B lymphocytes, polymorphonuclear leukocytes, type II pneumocytes, astrocytes and microglial cells, have been identified to produce complement C3 proteins (Lambris, 1988; Mishra et al., 2009; Rothman et al., 1989; Strunk et al., 1994; Strunk et al., 1985). In rock bream, kidney, spleen, and peripheral blood cells have been detected as the second large C2 producing tissue, which is also an evidence for extra-hepatic C2 production. Those extra-hepatic C2 is assumed to be important for modulation of immune responses in respective tissues.

Indeed, detail studies on the promoter regions of the complement component 2 in teleost fish are scanty. In our study, we tried to identify initiation sites and some of the important immune related transcription factor binding sites based on the predictions in the 5'

flanking region of *Rb-C2*. In the promoter region of *Rb-C2*, the luciferase reporter assay confirmed that the transcriptional factors such as SP-1, AP-1, OCT and C/EBP are positively regulates the transcription of Rb-C2.

In summary, we have identified the complete genomic sequence of complement component 2 from rock bream. Genome structure analysis of *Rb-C2* confirmed its substantial similarity with C2 of other teleost fish and higher vertebrates. Moreover, *Rb-C2* was found to be potentially induced by infection with *E. tarda*, *S. iniae,* LPS poly I: C and iridovirus, indicating its potential role in the rock bream immune defense system. According, to the results of luciferase assay, it was consisted of complement specific and liver specific active promoter region.

Part 2: Molecular characterization and functional analysis of complement component 3, 4, and 5 from rock bream *Oplegnathus fasciatus*

Abstract

The complement component 3(C3), complement component 4(C4), complement component 5(C5), are the key proteins of the complement system that activates effector mechanisms against invading pathogens. The full-length cDNA sequences of the C3, C4 and C5 sequences were identified from the rock bream cDNA database by BLAST-X search of cDNA sequences and then confirmed with genomic sequence. The *Rb-C3* gene contains 43 exons and 42 introns distributed in genomic sequence of 55870 bp and the cDNA of Rb-C3 is composed of a 4992 bp open reading frame (ORF). The *Rb-C4* gene contains 41 exons and 40 introns distributed in genomic sequence of 16437 bp and its cDNA is composed of a 5133 bp open reading frame (ORF). Meanwhile, *Rb-C5* is consisted of 5064 bp of ORF, which encodes polypeptide with molecular weight of 187 kDa. The all three genes had same domain architecture as eight macroglobulin (MG) domains, an anaphylatoxin domain, a thioester domain and a C345C domain in the C-terminus. The 5' flanking sequences of Rb-C3 and Rb-C4 were identified as a region abundant of immune related transcriptional factor binding sites, including those for NF-κB, SP-1, C/EBP, AP-1, c-Jun, ICSBP, HNF-1 and OCT-1. These transcription factor binding sites are common for most of the complement components. Quantitative real-time PCR analysis demonstrated that all three genes were constitutive expressed throughout healthy tissues from rock bream, with highest expressions were observed in liver. Similarly, *Rb-C3, Rb-C4 and Rb-C5* transcripts were up-regulated in head kidney in response to *E. tarda, S. iniae,* lipopolysaccharide endotoxin and rock bream iridovirus. Furthermore, *Rb-C3*-expressing cells were abundantly detected in head kidney after *E. tarda* infection by *in situ* hybridization assay. Furthermore, recombinant proteins of all anaphylatoxin domains were strongly stimulated the respiratory burst of head kidney cells

in vitro. Taken together, the molecular characteristics, orthologous relationships, transcriptional response to pathogenic stimulants and the ability to induce oxidative stress in immune cells suggest that *Rb-C3, Rb-C4* and *Rb-C5* are key functional components of the complement system and may be functionally involved in immune processes of rock bream.

2.1 Introduction

The complement component 3(C3), C4 and C5 are key acute phase proteins that circulates in the blood stream and get activated in protein level via classical, lectin or alternative pathways by photolytic cleavage (Janssen et al., 2005; Sahu and Lambris, 2001). Furthermore, their activation is essentially required for all the important functions performed by other downstream components of the complement system (Qi et al., 2011; Sahu and Lambris, 2001). Among the over 30 glycoproteins in the system, C3 is the most diverse, most abundant and third largest protein that interacts with more than 25 different proteins (Sahu and Lambris, 2001).The largest protein in the system is complement component 4 and the second largest is complement component 5. When considering structural and genomic level diversity, they have complicated domain architecture and the orientations of domains are located in a similar way in three proteins. All of the components consist of two main subunits, namely α chain and β chain that are linked together before activated by proteolytic enzymes. In human, C3, C4 and C5 proteins are comprised of over 1600 amino acids arrange as 13 distinguished domains namely; 8 macroglobulin (MG) domains, linker domain (LNK), anaphylatoxin (ANA) domain, CUB domain, thioester-containing domain (TED) and carboxy-terminal C345C domain. In multicellular organism, C3 and C4 have multiple functions, including acting as C3 convertases and as C5 convertases, which in turn activates the lytic pathway genes. The complement mediated inflammation and opsonization of invaded pathogen are performed by activated anaphylatoxin and C3b (Boshra et al., 2006). The functional role of

C3a (anaphylatoxin domain) has been well studied in mammalians and rainbow trout from teleost. The ability to induce respiratory burst responses of purified C3a have been demonstrated in rainbow trout (Rotllant et al., 2004). However, there is no any evidence for induction of the respiratory burst activity by recombinant proteins purified from prokaryotic expression systems.

To date, the complement system has been discovered in both deuterostomes and some lineages of protostomes such as arthropods, mollusks (Ariki et al., 2008; Castillo et al., 2009; Zhu et al., 2005). Moreover, the complement component 3 has been identified from vertebrates, some of arthropods and mollusks (Ariki et al., 2008; Castillo et al., 2009). In fish, complement C3 has been characterized at cDNA level in rainbow trout (Zarkadis et al., 2001), spotted grouper (Qi et al., 2011), Atlantic cod (Lange et al., 2005) and spotted wolfish (Abelseth et al., 2003). However, comprehensive immune-challenge study has not been carried out on complement C3 in fish. Furthermore, this is the first comprehensive study that reveals complete genomic structure, 3-dimentinal protein structural comparison, 5' flanking region analysis and protein level functional characterization of recombinant-expressed anaphylatoxin domain in teleost.

When considering the present status of complement C4, it has been identified from common carp (Mutsuro et al., 2005) rainbow trout (Boshra et al., 2004), Japanese flounder (Aoki and Hirono, 2006), nurse shark (Dodds et al., 1998) channel catfish (Dodds et al., 1998),medaka (Kuroda et al., 2000) and zebrafish (Samonte et al., 2002). However, there is no expression analysis or genomic structural identification and promoter identification have been performed in the past using new molecular biological tools. Therefore, this study is the first research work, which reveals the complete identification of the complement C4 gene in teleost fish.

The complment C5 have been identified in three fish species including rainbow trout (Franchini et al., 2001), common carp (Kato et al., 2003) and nurse shark (Graham et al., 2009). However, no detail studies have been performed to reveal its trasciptional responses upon pathogenic challenge.

In this study of rock bream, we aimed to identify three full-length cDNA sequences of C3 *(Rb-C3),* C4 *(Rb-C4)* and C5 (*Rb-C5*) their genomic structure from its BAC genomic DNA library. Modeling of 3-dimentinal protein structure along with comparison with human crystal structure and transcriptional factor binding site in 5' flanking region were determined. Apart from investigating their tissue-spcific expression patterns in healthy rock breams, we also evaluated the transcriptional modulations upon stimulated by live bacteria (Gram-positive and -negative), virus and immune stimulatory agents, lipopolysaccharide (LPS). Furthermore, respiratory burst activities of recombinant anaphylatoxin domains of each protein were also exmained using nitroblue tetrazolium (NBT) assay.

Table 8. Description of primers used for cloning and amplification of genes in this chapter

Name	Purpose	Primer Sequence (5'-3')
Rb-C3 -1F	qPCR amplification and BAC screening	TCCTCTGGGTAAACTGCGTCCAT
Rb-C3-1R	qPCR amplification and BAC screening	TACCTTATCATGGGCATGGGCAGA
Rb-C3 -2F	ORF amplification	TGCGGGTAGGAACAGCAGAATACA
Rb- C3-2R	ORF amplification	TTTACCGCAGCTGTATCCACTCCT
Rb-C3 -3F	ORF amplification	CTGAAAGGCAAGCATCAGCCAACA
Rb- C3-3R	ORF amplification	TCCTGAAAGTTGGCTGTGGACAGA
Rb-C3 -4F	ORF amplification	ATAACTGTGATAGGCGCAGCGAGT
Rb- C3-4R	ORF amplification	GCAAACACCTTGGCAACATAGGCT
Rb-C3 -5F	*In situ* hybridization –RNA probe synthesis	CTGAAAGGCAAGCATCAGCCAACA

Rb- C3-5R	*In situ* hybridization –RNA probe synthesis	TCTGTCCACAGCCAACTTTCAGGA
Rb-C3-6F	Anaphylatoxin domain amplification (*Eco*RI)	GAGAGAgaattcGCCAGTACAATACTGGACGTCAGAACC
Rb-C3-6R	Anaphylatoxin domain amplification (*Hind* III)	GAGAGAaagcttACTGCGTGCCAGTATGAGGC
Rb-C4 -7F	qPCR amplification and BAC screening	TGCTCAACAAGGGAAGACTCCACA
Rb-C4-7R	qPCR amplification and BAC screening	AAACCATACAGATGGCTCGGGTGT
Rb-C4 -8F	Anaphylatoxin domain amplification (*Eco*RI)	GAGAGAgaattcTCTGTGGACCTTCAACAGGAAATGATGA
Rb-C4-8R	Anaphylatoxin domain amplification (*Hind* III)	GAGAGAaagcttGCTTGCAGTCCTTCCAAGTCCTTTC
Rb-C5 -9F	qPCR amplification and BAC screening	AACTCCAGGGAAACACTGTGGTCA
Rb-C5-9R	qPCR amplification and BAC screening	TTTGTCCTGAGGCTCATAGACGCT
Rb-C5-10F	Anaphylatoxin domain amplification (*Eco*RI)	GAGAGAgaattcGCTCTGACTGATGAGGAGAAGACGA
Rb-C5-10R	Anaphylatoxin domain amplification (*Hind* III)	GAGAGAaagcttGTTGCGTCCCAGGATGAGGTT
Rb-β-actin-11F	qPCR reference gene amplification	TCATCACCATCGGCAATGAGAGGT
Rb-β-actin-11R	qPCR reference gene - amplification	TGATGCTGTTGTAGGTGGTCTCGT

2.2 Results

2.2.1 *In silico* characterization of Rb-C3

Based on BLAST search analysis of EST data from the rock bream normalized cDNA library, we identified a putative complement component 3 and designated it as *Rb-C3*. The sequence was deposited in NCBI under accession number KF408222. The complete cDNA sequence is 6155 bp, including a 17 bp 5'-untranslated region (5′-UTR), a 4989 bp open reading frame (ORF) encoding 1663 amino acids, and 1147 bp 3'-UTR with a polyadenylation signal (Figure 21). The predicted molecular mass of *Rb-C3* is 187 kDa and the theoretical isoelectric point is 5.8. According to SignalP prediction, the *Rb-C3* possesses a signal peptide of 29 residues, which is presumed to facilitate the extracellular localization (Figure 21). The mature protein of *Rb-C3* contains 13 domains similar to human C3, which are separated mainly into,

β and α chains by a linker [663]RRKR[666]. The molecular weights of deduced β and α chains are 73 kDa and 113 kDa, respectively. The domain structure of *Rb-C3* is composed of eight α_2 macroglobulin domains (MG 1-8) distributed in both chains, an anaphylatoxin domain which is conserved in complement C4 and C5 in vertebrates, thioester domain and common C-terminal C345C domain or netrin domain (NTR) (Fujito et al., 2010; Janssen et al., 2005). The anaphylatoxin domain of *Rb-C3* contains 78 amino acids and consists of 6 conserved cysteine residues where they form internal disulphide bond between designated two residues. Three connecting regions were identified such as a linker which forms a small hydrophobic core at C-terminus of β chain and the other regions either side of the TED domain (CUB regions). Moreover, the MG6 is separated into both β and αchains which ultimately get connected upon release of C3a anaphylatoxin domain (Figure 21). The NTR domain is a 130-residue domain containing six conserved cysteine residues, which are likely to form internal disulphide bonds. Moreover, the NTR is anchored by an anchor domain to the 8th MG domain in C3.
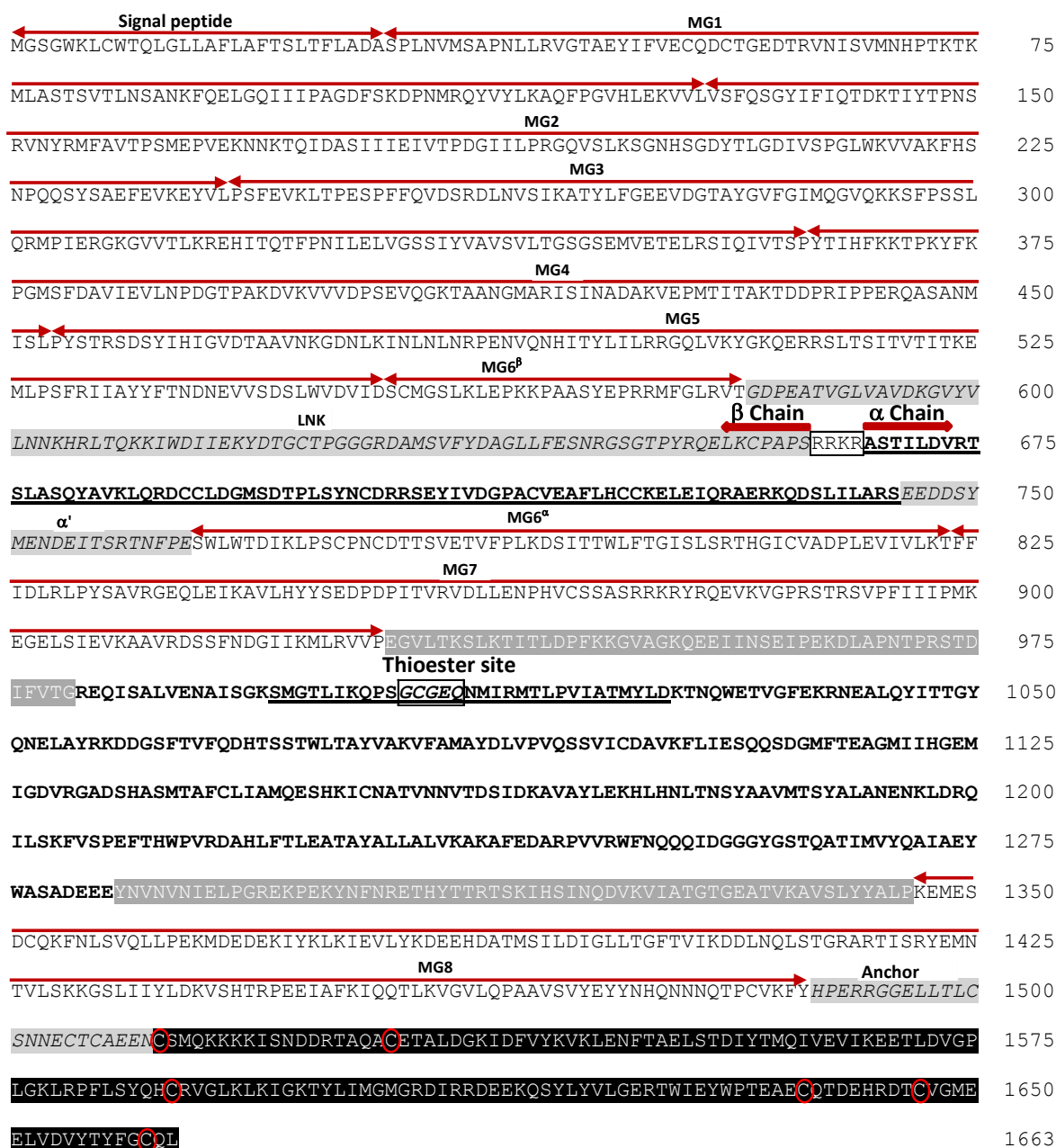
Figure 21. Protein sequence of Rb-C3

The α and β chain demarcation is indicated by arrow heads with corresponding labels. The potential cleavage processing site of α and β subunits is boxed. The domains identified; Signal peptide and macroglobulin (MG) 1-8 are indicated by double arrows with labeled above. An anaphylatoxin domain is bold and underlined. The five-residue thioester (TED) site (GCGEQ) is bold, italicized in box within the thioester domain (TED), CUB domain is indicated by white letters highlighted in gray either side of the TED domain and C345C / NTR domain is highlighted in black at the C-terminus. The linker (LNK), N terminus of C3α chain (α') and anchor are denoted by italicized black letters shaded in gray. The conserved cysteine residues of NTR are encircled.

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

2.2.2 Characterization of the Rb-C3 genomic structure and promoter region

The genomic sequence of *Rb-C3* was identified by PCR-based screening of the BAC genomic library using *Rb-C3* gene-specific primers and GS-FLX sequencing of putative candidate clone. The putative genomic sequence of *Rb-C3* was recognized by comparison with the full-length cDNA sequence. Thereafter, the exon and intron structure was compared with the genomic structures of several other vertebrate species published in the Ensembl genome database. The *Rb-C3* genomic sequence is about 55.87 kb and is composed of 43 exons interrupted by 42 introns. Interestingly, *Rb-C3* gene size is the largest among all the genomic sequences used in the comparison, even higher than that of human (42.81kb). The 5'- and 3'-ends of *Rb-C3* introns harbor canonical splicing motifs (according to the AG/GT rule) except 15th intron in which the GT is replaced by GC. The number of exons and introns were found to be varied between 40-44 in species considered in the comparison (Table 9 and Figure 22). Furthermore, the genomic sequence of *Rb-C3* is harbored few putative microsatellites in between exons namely (exon 8-9 (ACT)[5],exon 9-10 (GT)[19],exon 17-18 (GT)[28]…(GT)[6],27-28 (GT)[10],29-30(GT)[23],32-33(GT)[15],36-37(GT)[8] respectively.
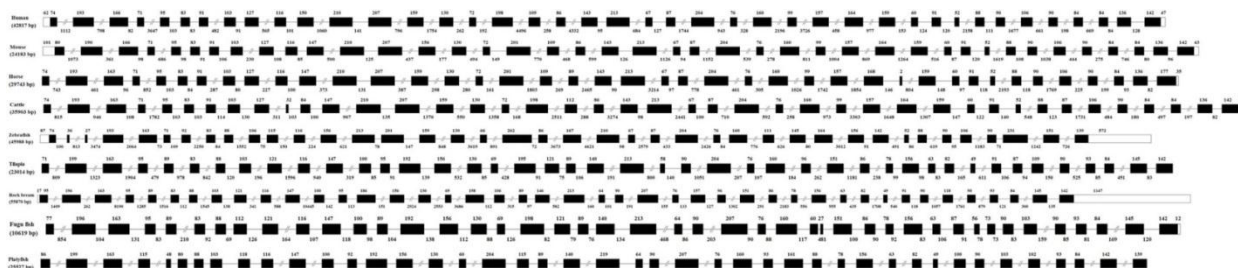
Figure 22. Genomic structural comparison of the Rb-C3

Exons are shown in black solid boxes. UTRs are shown in light-colored boxes. Introns are shown as thin lines between the exons, and introns larger than 200 bp are indicated by interrupted lines. Numbers above the boxes and below the lines indicate the lengths of exons and introns, respectively.

76

Table 9. Genomic sequence information

Species	Accession No	Genomic sequence size (bp)	Number of exons
Human	ENST00000245907	42817	41
Mouse	ENSMUST00000024988	24183	41
Horse	ENSECAT00000007684	29743	41
Cattle	ENSBTAT00000022979	35903	42
Zebrafish	ENSDART00000108737	45988	40
Tilapia	ENSONIT00000020689	23014	43
Rock bream	KF408222	55870	43
Fugu	ENSTRUT00000027125	10619	44
Platy fish	ENSXMAT00000012477	25527	43

According to the BDGP prediction analysis, the transcription initiation site was coinciding with the actual +1 site that was located 17 bp upstream of the ATG. The TATA box was not identified in the respective region. The promoter structure of *Rb-C3* was composed of co-promoter region and multiple binding sites for factors with functions related to the transcription regulation process. Among those identified sites, putative immune-related transcription factor binding sites were highlighted in Figure 23 and these include the CCAAT-enhancer binding protein (C/EBP), nuclear factor-kappa B (NF-κB), activator protein-1 (AP-1), octamer transcription factor (OCT-1), specificity protein-1 (SP-1), hepatocyte nuclear factor-1 (HNF-1), hepatocyte nuclear factor-3 (HNF-3), nuclear factor-1 (NF-1), interferon consensus sequence binding protein (ICSBP), activating transcription factors (ATF), GATA-1,CP-1 and C-Jun.

2.2.3 Identity, similarity, and phylogenic relationship of Rb-C3

Pairwise sequence analysis was revealed that *Rb-C3* exhibited 73.7% amino acid identity with the orthologous from orange-spotted grouper (*Epinephelus coioides:*ADU33222.1) and 72.7% with Mi-iuy croaker (*Miichthys miiuy*: AFC89899.1). In addition, it was showed over 45% identity with the orthologous from other fish species. When compared with other vertebrates, it was showed over 40% identity (Table 10). However, the identity and similarity values sheared with the sequences from invertebrate species such as sea cucumber and Arctic

lamprey were found to be relatively low. In multiple sequence alignment, Rb-C3 showed

higher degree of conservation among other vertebrates. All the cysteine residues were found

to be conserved across nearly all the orthologous, and certain (presumably important)

signature motifs also were highly conserved among all the orthologous (Figure 24).

```
GTGTTCAGTGAGGCAGTATGCTCTCTGAATGATTTTCTTGTTTG**TTTCTTTTTT**GTCTGGATGGATGTTC   -1254
                                              ICSBP
TCATGATTCAACCTTGTGTTCCATTGCTTGCTGATTT**TTGGGCCCCA**GTCCCTTGATTCCTTTAATATCG   -1184
                                         SP-1
**GCTATTTGCC**GATACCTAAACAGATCTAATACTTTTACAATTTTACTGAATGTGTATCTGACACATTGAA   -1114
  C/EBPα
ATAACAGGTGCAGACTTT**TAAATGTAGC**ACAC**CTCATTTT**TCACGAGATACTCAATCCAAATACCGAAGG   -1044
                    SP-1            ATF
TCCTGATTCAAAACCATAAAGA**TTATGAGTAT**AAATTATGAATATGATATGGATGAAAGTTTAACTGGCT   -974
                       OCT-1
GAATGTGACTCC**TAAAATTGAC**ATGCAAATTTTCACTCTCATACACAGTGAATGCTTGCAGTGTAGAGCC   -904
              C/EBPα
CAGCTATCATTAAGGAGAGCAGCATGCTGAAACACT**AGCGGGCTGC**GGTGGACAAAATATAGTATTTATT   -834
                                       SP-1
**TTAGTCAATA**CAGACCCATTAAAATATGCCCAGATTGCCTGCATTAACAATCCTTAGGATTGCCTTGGGG   -764
   HNF-1
CGTCCCTGTTTTATAGCATGCTGATATTTCT**ACAGCCAATT**GTGCAATATTCCTATTTCAAAATGATACA   -694
                                 CP-1
GTCCAG**TCCTCAGTTT**GCTTCCATATTA**TCATTTTTTA**ATGATGACAATGATCAATAAAGGTTGCTCTAA   -624
        ICSBP                     c-Jun
GTAGTAACCTGCCACATACTCTGTAACAATAACAAAACACTTCCAACGTCTAATAAACAGC**ATCAAAAAC**   -554
                                                                GATA-1
**A**GGTGCCAGTATGAGACCATTTAATGA**GAAATCTCTC**TCATTAGAC**ATTTCCTCTG**TCTCTTATCGTTAC   -484
                              NF-κB                   NF-κB
TAGATCTGACATTTTTATTTTTTACGTGGTTTTT**CTCATGGTTT**TTGTTGTGTTTTTGTTTGCCTTTTGT   -414
                                    c-Jun
GAAGCACTGTGTAATCACTGGGTTTGATATATATTTGGTAAGTGCTATATAAATAAAT**GTAATTATTT**TG   -344
                                                            C/EBPα
**ACAAAAAAAA**AATAAGAAAAAAGTCAAGGTCCCT**TTACTAACCATTAGTCAC**GGACTTCATCAGCAGTTG   -274
  HNF-3     OCT-1                        HNF-1       AP-1
GAGTTGAGCAAATCGTAGCACGAGAATTAGAGGAACCACTTTGCCTACCATCATATGTACTTATGTCAGA   -204

GATAGAAGAAATGTTTTTTGTACAACTG**TGTGCCAACT**CTGGGTTTAGGACTTCTAACAAGCAGA**TAAAT**   -134
                             NF-1
AA**TTT**ACCATACTAACAG**CCTAATGAATATGGCTGCA**CATCAATTG**CACATGCTGT**TACTTT**TGATATTG**   -64
C/EBPβ                  OCT-1     NF-1              c-Jun           ⮕ GATA-1
**AA**AACGCAGAGGTGCA**TGAGGAGGGG**GTAAGTGGATAAAAG**GAGGGTGTTG**CAGAGCGAGAGT**C**TACATT   7
  Sp-1                            Sp-1                        +1 site
GACAGCAGAT***ATG***GGATCAGGTTGGAAGCTGTGTTGGACCCAGCTGGGGCTATTGGCCTTTCTGGCCTTT   77
          M  G  S  G  W  K  L  C  W  T  Q  L  G  L  L  A  F  L  A  F
```

Figure 23. 5' flanking region of Rb-C3
Putative transcription factor binding sites are shown in bold phase and underlining while labeling the corresponding transcriptional factors directly below the sequence. The transcription initiation site is indicated by a bent arrow. Translated amino acids (upper case) and start codons (italicized uppercase, in bold phase) are shown.
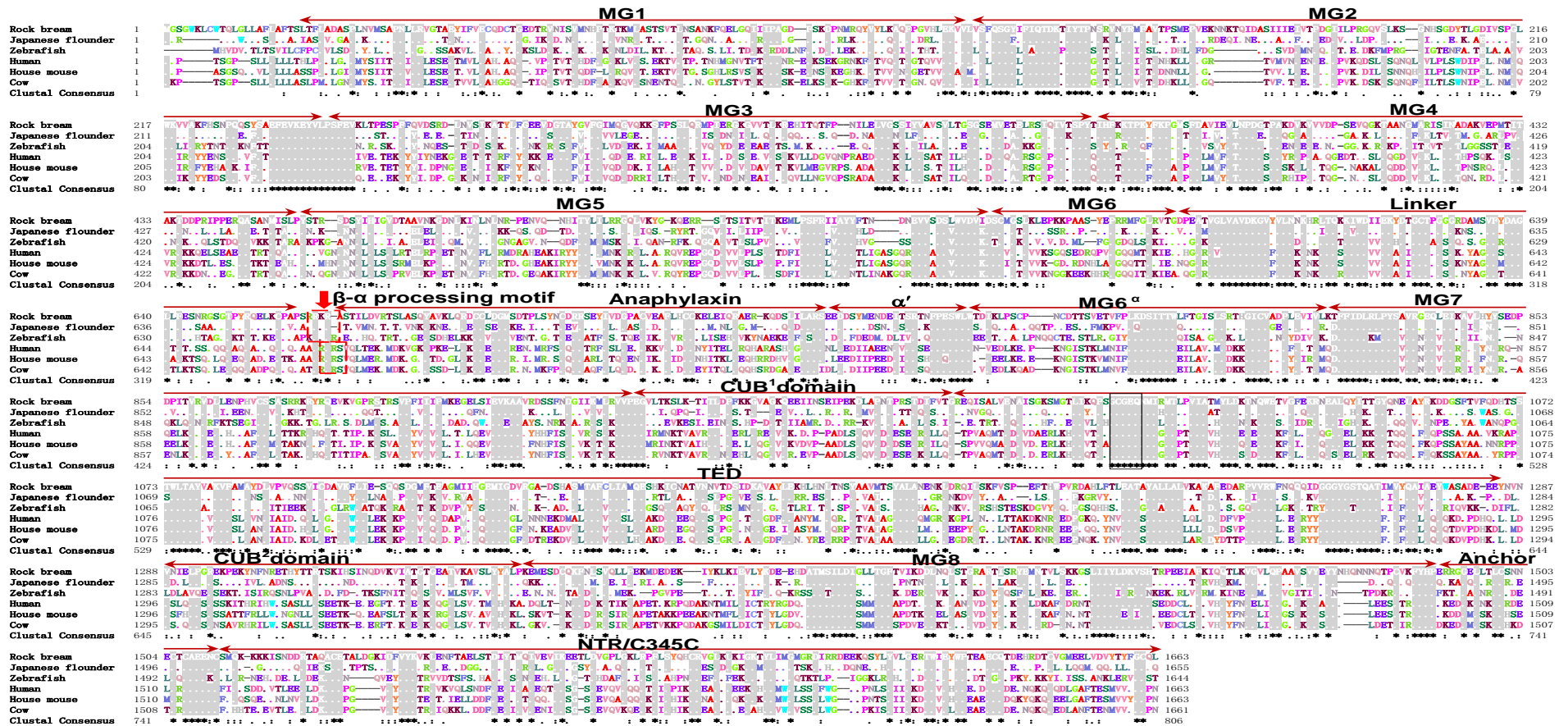
Figure 24. Multiple sequence alignment of Rb-C3 with its orthologous

Identical residues among all orthologous sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double-arrows. The cleavage sites of α and β subunits are boxed and thioester site is indicated by triangles, above the sequence alignment. Important domain regions are marked by double arrows. The corresponding GenBank accession numbers are presented under the C3 sub-cluster next to each species in phylogenetic tree except *Sparus aurata*: (HM543456.1), *Apostichopus japonicas*:(HQ214156.1).

79

Table 10. Percentage of identity, similarity, and gap of Rb-C3 with other C3 orthologous

Common Name	Species	Accession No.	AA	Identity %	Similarity %	Gap %
Orange-spotted grouper	*Epinephelus coioides*	ADU33222.1	1657	73.7	83.8	1.3
Mi-iuy croaker	*Miichthys miiuy*	AFC89899.1	1657	72.7	82.8	1.3
Japanese flounder	*Paralichthys olivaceus*	BAA88901.1	1655	70.1	81.9	1.3
spotted wolffish	*Anarhichas minor*	CAC29154.1	1662	68.8	80.8	2.2
Medaka	*Oryzias latipes*	AB025576.1	1657	67	79.3	1.2
Zebrafish	*Danio rerio*	XP_002660624.2	1644	47.9	65.2	3.5
Grass carp	*Ctenopharyngodon idella*	AAQ74974.1	1677	45.9	63.5	8.3
Cattle	*Bos taurus*	NP_001035559.2	1661	43.3	61	5.4
Mouse	*Mus musculus*	NP_033908.2	1663	42.7	60.6	4.8
Chicken	*Gallus gallus*	NP_990736.1	1652	42.6	60.2	5.8
Human	*Homo sapiens*	NM_000064.2	1663	42.3	60.5	5.3
Frog	*Xenopus (Silurana) tropicalis*	XP_002940096.1	1647	36.9	53.8	12.5
Arctic lamprey	*Lethenteron camtschaticum*	AB377282.1	1467	29.9	45.8	20.7
Sea cucumber	*Sparus aurata*	HM543456.1	1738	28	45.7	18

The evolutionary relationship of Rb-C3 with its homologues was studied using the deduced amino acid sequence of Rb-C3 with sequences available in the NCBI database for the complement C2, C3 and C4. In phylogenetic tree, Rb-C3 was located within the fish clade and separated from other higher vertebrate C3 genes. Furthermore, C3 group was shown closely associated with complement C4 gene group than C2 in phylogenetic tree, which indicates its close evolutionary relationship with C4 (Figure 18).

2.2.4 Tertiary structural model comparison of Rb-C3 protein

The protein homology modeling was carried out to compare the structural similarities and differences with crystal structure of complement C3 in human. The β and α chains of Rb-C3 were modeled as separate entities for easy comparison. In comparison to the human model, Rb-C3 was also showed the very similar structure characteristics at tertiary level. All the MG

80

domains (MG1-MG8) were highly resembled that in human C3 structure. In human, MG1-MG5 and half of MG6 are located in β chain, and half of MG6, MG7and MG8 belong to the α chain. The β chain end up with linker (LNK) domain and α chain begins with anaphylatoxin domain (Figure 25). The thioester domain is an important domain, which is buttressed by interactions with MG2, MG8 and CUB domains.
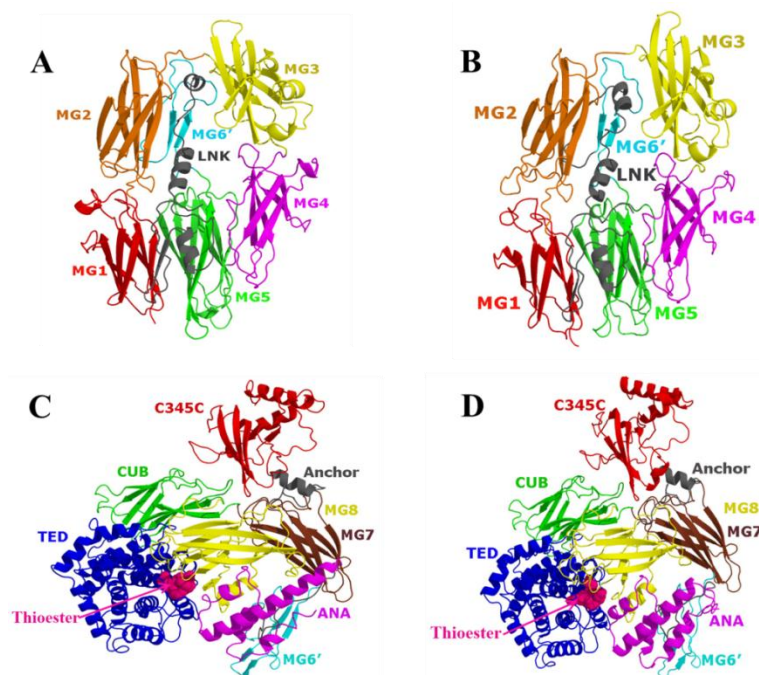
Figure 25. Comparison of 3D-molecular structures of β and α chains of human and rock bream C3

The A and B represent the β chain of human and rock bream respectively; C and D represent α chain of human and rock bream C3 s, respectively. Each domain was labeled with respective color near the domain. The thioester site is labeled in each α chain in pink arrow.

2.2.5 Tissue-specific mRNA expression of Rb-C3 in normal healthy fish

To examine the normal tissue distribution pattern of *Rb-C3* transcription, qPCR was carried out for blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine (Figure 19). *Rb-C3* was found to be constitutively expressed in all examined tissues, although the level of expression varied between each. As the all complement components, *Rb-C3* also indicated its liver-specific expression compared to the normalized expression level in muscle.

81

2.2.6 Modulation of Rb-C3 expression in response to immune stimulation

The complements are activated in protein level upon pathogen invasion. Out of all components, complement C3 is the most important protein and readily available in the blood stream of animals. However, it is still intriguing to explore the transcriptional regulation of *Rb-C3* under pathogenic challenge. Hence, modulation at RNA level of *Rb-C3* was measured in head kidney from rock breams challenged with *E. tarda, S. iniae*, RBIV, or LPS by qPCR (Figure 26). The expression pattern upon Gram-negative bacteria *E. tarda* displayed a biphasic expression response. The amount of *Rb-C3* transcripts were significantly ($P<0.05$) increased from baseline (0 h) between 3 h and 12 h, and declined to baseline at 24 h and again increased at 48 h post-injection (p.i.), giving the maximum level at 3 h p.i. (Figure 26). The same biphasic patterns of induction were observed for the LPS and iridovirus challenges as well. In *S. iniae* challenge, *Rb-C3* showed significant induction at 12 h p.i. The highest up-regulation was recorded for the LPS challenge at 24 h p.i (41-fold). In general, the overall extents of fold-inductions in *S. iniae* and iridovirus were lower than those in *E. tarda* and LPS challenges (Figure26). The *in situ* hybridization results were also confirmed that *Rb-C3* was strongly expressed in head kidney cells upon *E. tarda* challenge (Figure 27).
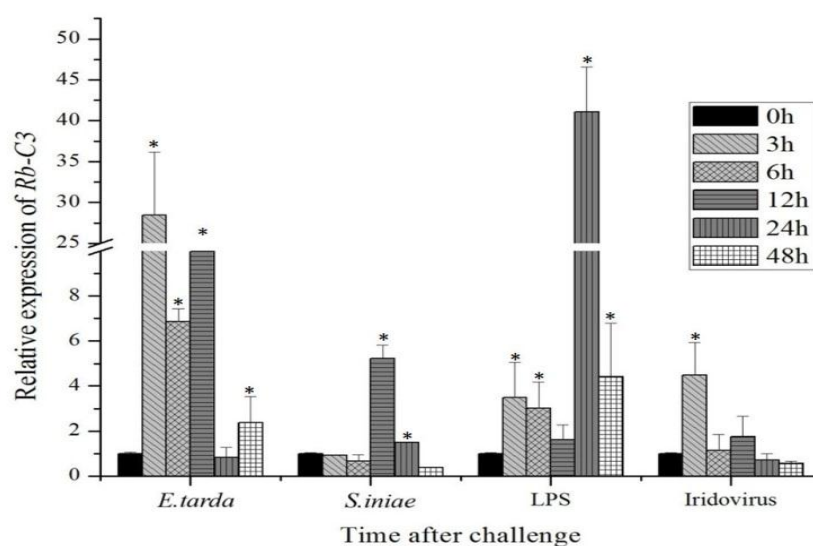
Figure 26. Relative expression of Rb-C3 mRNA in head kidney after challenge of rock bream with *E. tarda*, *S. iniae*, LPS, and iridovirus, as detected by qPCR
The amount of mRNA relative to the internal control, β-actin, is expressed as mean ± SD (n = 3). Significant differences ($P < 0.05$ vs. untreated control at 0 h) are marked with asterisks.
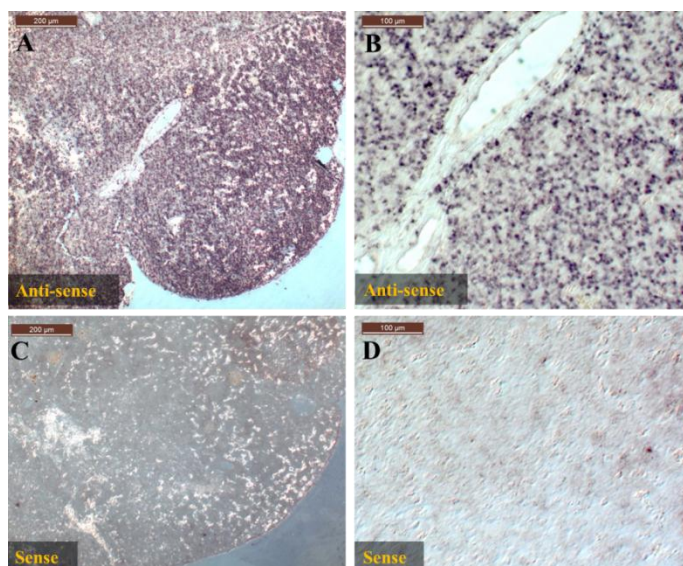
Figure 27. *In situ* hybridization detection of Rb-C3 RNA upon *E. tarda* challenge.
In situ hybridization in head kidney (A and B) DIG-labelled anti-sense probes, with low and high magnification, (C and D) DIG-labelled sense probes respectively.

2.2.7 Physiological characterization of Rb-C3a at protein level

2.2.7.1 Prokaryotic expression and purification of the soluble recombinant Rb-C3a

We introduced the anaphylatoxin domain (C3a) of Rb-C3 into pMAL-c2X expression vector and transformed the recombinant construct to express the Rb-C3a protein in *E. coli* BL21 (DE3) cells through IPTG-mediated induction. The purified protein was analyzed on SDS-PAGE (Figure 28 A). Purified recombinant Rb-C3a fusion protein appeared as a single band with a molecular size of ∼51.3 kDa, which is close to the size of calculated molecular mass of the fusion protein (Rb-C3a, 8.77 kDa + MBP, 42.5 kDa).

2.2.7.2 Respiratory burst activity of the Anaphylatoxin domain (C3a) of Rb-C3

The measurements of the respiratory activity over different treatments were plotted in Figure 28 B. Compared to the negative control (un-stimulated sample), absorbance of the samples in *Rb-C3a* treatment was significantly increased ($P < 0.05$). It was 1.4-fold increment of reactive oxygen species (ROS) production compared to negative control. Furthermore, in the *Rb-C3a*+SOD treatment, the active SOD was shown to be able to reduce ROS production level close to negative control (Figure 28B).
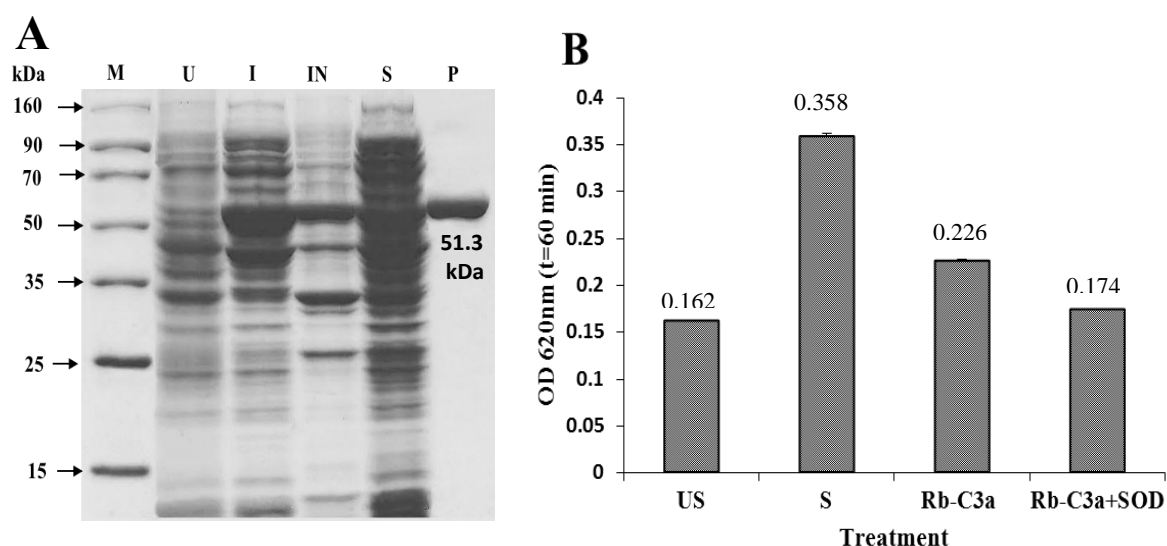
Figure 28. SDS-PAGE analysis of overexpressed and purified recombinant Rb-C3a protein (A) and Respiratory burst activity of the recombinant protein (B)
Lanes: M, molecular mass marker (sizes are indicated in kilodaltons); U, total cellular extract from *E. coli* BL21 (DE3) before IPTG induction; I, total cellular extract from *E. coli* BL21 (DE3) after IPTG induction; IN, insoluble fraction of proteins after IPTG-induction; S, soluble fraction of proteins after IPTG-induction; P, purified recombinant fusion protein Rb-C3a/MBP. **(B)** Comparison of respiratory burst activity of recombinant Rb-C3a. US, un-stimulated (-ve control); S, stimulated by H_2O_2(+ve control); Rb-C3a, 20 μg of Rb-C3a treated; Rb-C3a+SOD, 20 μg of Rb-C3a and 20 μL of 100 IU/mL of superoxide dismutase.

2.3 Results

2.3.1 *In silico* characterization of Rb-C4

Based on BLAST search analysis of EST data from the rock bream normalized cDNA library, we identified a putative complement component 4and designated it as *Rb-C4*. The sequence was deposited in NCBI under the accession number KF744235. The complete cDNA sequence was found to be 5444 bp in length, including a 17 bp 5'-untranslated region (5′-UTR), a 5133 bp open reading frame (ORF) encoding 1710 amino acids, and a 303 bp 3'-UTR with a polyadenylation signal. Interestingly *Rb-C4* was the largest ORF found in the rock bream complement system. The predicted molecular mass of *Rb-C4* was 191 kDa and the theoretical isoelectric point was 6.1. According to SignalP prediction, the *Rb-C4* possesses a signal peptide of 19 residues, which is presumed to facilitate the extracellular localization (Figure 29). The mature protein of *Rb-C4* contains 13 domains similar to human C4, which are separated mainly into, β and α chains by a linker ^{651}RQRR^{654}. The molecular weights of deduced β and α chains are 70 kDa and 118 kDa, respectively. The domain structure of *Rb-C4* is composed of eight α_2 macroglobulin domains (MG 1-8) distributed in both chains, an anaphylatoxin domain which is conserved in complement C4 and C5 in vertebrates, thioester domain (TED) with five-residue thioester site (^{991}GCGEQ^{995}) and common C-terminal C345C domain or netrin domain (NTR) (Fujito et al., 2010; Janssen et al., 2005). The anaphylatoxin domain of *Rb-C4* was contained 85 amino acids, which were consisted of 6 conserved cysteine residues where they form internal disulphide bond between designated two residues. Three connecting regions were identified such as linker, and two CUB regions either side of the TED domain. Moreover, the MG6 was separated into both β and α chains which ultimately get connected upon release of C4a anaphylatoxin domain (Figure 29). The NTR domain is a 130-residue domain containing six conserved cysteine

residues which were likely to formed internal disulphide bonds. Moreover, the C345C / NTR

was found to be anchored by an anchor domain to the 8th MG domain in C4.


```
                    Signal peptide                                      MG1
MRCYLFSVLFLILSTECSADRFFISAPGVFHVGVNEKVFVQMGGSHLNHPVTLYLEHETSSTIVSAKKTTVCTAE          75

GEIKTVELMIDSEIMSRVPQRHEHSYLVLVAESLSISGRKMSRVLVSKRRGNIFIQTDQPIYNPTQTVTYRIFTL         150
                                         MG2
DHTYRPKVEAFQISVFNAAGNRIMKSLKTAGGGILSGTFPIPDVSKTGTWKITAHYKDDEANAASREFKVQKFIL         225

PSFEVNIAMGQSYILLNSEQFDFTISAKYSHGEKVKGAYHCQFGVVEKDPTSGQKMKPVFIRGLELTGSVQDGTS         300
              MG3
GASLPIADLDMHLQKQKNQTLSKLQQSGAQLYLGVFVTNIQSGEIQEAEVHLPIISHRYTMDLSRTRSYFLPGYP         375
                          MG4
LDVVVVMRLPDGSPAAGVPVKIDVSPSTEEPWQGTTDQEGAVFPVFNIPTVAEITVQVSADGVQQRKVIRRASSP         450
                              MG5
SDSYLYLSFSNRMYSVGELFSVNYNTINSPKSGFIYYMVLSRGILTKQGSLGIGTSVKHNLPITPDMVPSFRLIG         525
                                MG6β
YYYDQRGNIIADSVWVDVRDECEIKVEVEEKGPFQPGKQSKLEFDLHGQRARVALLAVDKAFYAL HSDNKLTAKQ        600
            Linker                               β Chain          α Chain
VFSSMQSHDLGCSYGGGADPASVLTDAGLSFVSQSQSMWRRSFDCNSKSA RQRR SVDLQQEMMNLKANFSDEKLK        675
                                                                         Linker'
ECCTHGFFLIPMRLTCQQRAKRVSLVEANPVCADAFLKCCLEGERLRQKKIQEDALKGLGRTAS TAEIEEFFLDT        750
                              MG6α
ASQYIRRFFPP SFAFKEFDVNGKGSYSLALPDSITTWEIQVVTLSAATGFCVVKPSEVRAFKTAFVSLRLPYSVR        825
                          MG7
KYEQLSISPVIYNYGYNTLEVAVHMEQTEGLCSPGSATSTSFVNITVEQESSQLVSFSAVPMVTGSIPIKIRLYD        900
                          CUB1
MENNYGMDAVEKPLNVL TEGLQEKVEETQVLTLNGRSTKTIIIDGTLPDDTVPDSSSNIFVSMEG DGFGSSRVKN       975
            Thioester site
LLSPEKVASLIVLPT GCLEQ TMVRLAPTASALRYLDLSEQWFDLPAGSRDDALDKIESGYIRIHSTFKKADGSYG      1050
                          TED domain
PWYSVPSSNWVTALVVKVLSLVAQRQIAAFGVQGRQARVVPEEEIRQPVSFLLSIQKTDGSFGDQHPVLHRGVLK      1125

GKDQEASMTAFITLALHRSLQFLQKEGRNNAEASISRSTTYLLSNLEELQHPYAVAITAYCLAVCMPQGTDHSFV      1200

WTKLQTLATEGKNGCYLWTADANPRNRETGDAITIETTAYALLAAVELGHTKWADKAACWLTTQENYFGGYRSSQ      1275
                                                          CUB2
DTIMALEALAEYELKRSVSPEANLIAEFT VPGKADIVRLALENKKERVETDLKKLTGNNIVVQLTGIGNTKLKVV      1350

KAYHLLE PKDNCDQVSIRVTVEGKVKYTAPVIENYDYYDENYDIIEEKAARVPRSAIEWFDARTRNRRDLDNNLN      1425
                          MG8
SDEAVTYRICVSHSLSNNLTGMAIADITLLSGFEVDTQDLDMLKEPPEQYISHYEVSYGRVLIYFNELFESEECI      1500
                          Anchor
SFDAIQRVPIGLLQPAPAVFYDYYEPNRKCTVFY SAPKRSKLVSKLCSEDVCQCAERA CHKIQNTFQSVRGRRIT      1575
                  C345C/NTR domain
KDNRLQHA CFFPTVDYAYIVEVINVSMKSNFEMYKANVKEVLRSHGDILVGVDSVRVFAKRR CKGQLGLGKQYL      1650

IMGKDGSTTDSDSMMQYLLESNTWVERNPSEE CKKSAHRLA CAGFNEFTEEYKIDG CRQ                     1710
```

Figure 29. Protein sequence of Rb-C4

The α and β chain demarcation is indicated with arrowheads and labeled above. The potential cleavage processing site of α and β subunits is boxed. The domains identified; signal peptide and macroglobulin (MG) 1-8 are indicated by double arrows with labeled above. An anaphylaxin domain is bold and underlined. The five-residue thioester site (GCLEQ) is bold, italicized in box within the bold thioester domain (TED), CUB domain(CUN is indicated by white letters highlighted in gray either side of the TED domain and NTR domain is highlighted in black at the C-terminus. The linker (LNK), N terminus of C4α chain (α') and anchor are italicized in black letters and shaded in gray .The conserved cysteine residues are circled in NTR domain

2.3.2 Characterization of the Rb-C4 genomic structure and promoter region

The genomic sequence of *Rb-C4* was identified by PCR-based screening of the BAC genomic library of rock bream using gene-specific primers. Then GS-FLX sequencing was done to obtain its complete genomic sequence. The putative genomic sequence of *Rb-C4* was recognized by comparison with the full-length cDNA sequence. Thereafter, the exon and intron structures were compared with the genomic structures of several other vertebrate species published in the Ensembl genome database. The *Rb-C4* genomic sequence is about 16.43 kb and is composed of 41 exons interrupted by 40 introns. The 5'- and 3'-ends of *Rb-C4* introns were harbored canonical splicing motifs (according to the GT-AG rule). The number of exons and introns were found to be conserved in species considered in the comparison except medaka fish (Figure 30).

Table 11. Genomic sequence information

Species	Accession No	Genome size(bp)	Number of exons
Human	ENST00000435363	20657	41
Mouse	ENSMUST00000069507	15517	41
Zebrafish	ENSDART00000028349	29954	41
Rock bream	KF744235	16437	41
Fugu fish	ENSTRUT00000040860	9700	41
Medaka	NM_001104697.1	16362	42

Figure 30. Genomic structure of Rb-C4
Exons are shown in black solid boxes. UTRs are shown in light-colored boxes. Introns are shown as thin lines between the exons, and introns larger than 200 bp are indicated by interrupted lines. Numbers above the boxes and below the lines indicate the lengths of exons and introns, respectively.
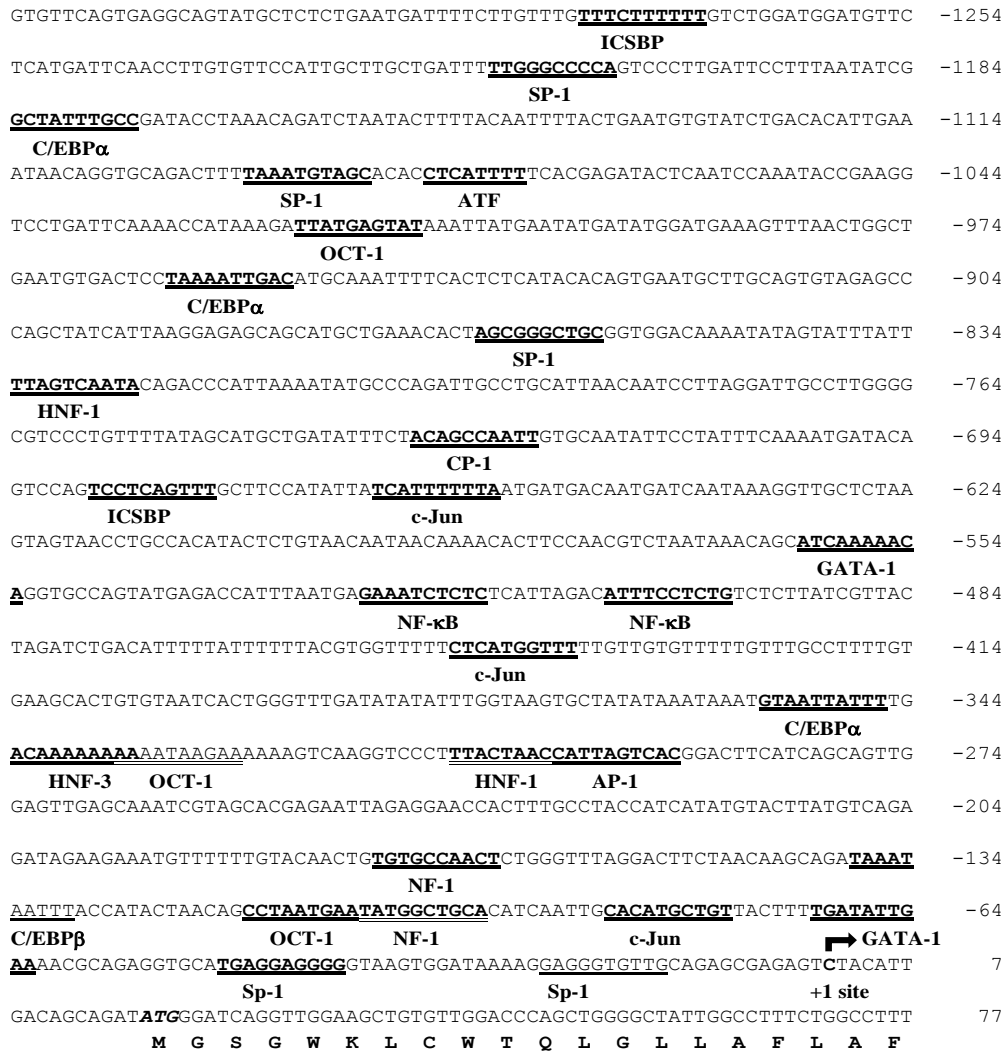
According to the BDGP prediction analysis, the transcription initiation site was coincided with the actual +1 site that was located 17 bp upstream of the ATG. The TATA box was not identified in the respective region. The promoter structure of *Rb-C4* was composed of co-promoter region and multiple binding sites for factors with functions related to the transcription regulation process. Among those identified sites, putative immune-related transcription factors were highlighted in Figure 31 and these include the CCAAT-enhancer binding protein (C/EBP), nuclear factor-kappa B (NF-κB), activator protein-1 (AP-1), octamer transcription factor (OCT-1), specificity protein-1 (SP-1), hepatocyte nuclear factor-1 (HNF-1), hepatocyte nuclear factor-3 (HNF-3), nuclear factor-1 (NF-1), interferon consensus sequence binding protein (ICSBP), GATA-1, and C-Jun.

```
CCTATGTATCTTATCTGAAACTCCTGGAGGATATTTTAAAAATGCAATTCAAAACTTTTTTTA      -1260
                                         OCT-1
ATGAATACGCTGAAATAAGTTGGAAATATGTACTGTTGACTGCGTATTGACATACTGTATCTACTATTAG  -1197
                     C/EBP
AGAAGGTCTCGGCAAGGTGGGAAATTGAGGAAAAATTTCAGAGACTCCAGGGATCAGTGCCTCAATCTGA  -1127
            SP-1
AGCACTTAGCTGGGATAACCATTTGGTGCTTAAATTCCTCCAAAAATCTTGACATGGATGTTTATTGGTG  -1057

TTTACATTTACACAAGAGCTATAGGTCTGAAGTAAAATATCAATAAAGATGTAAACATCTGATGCATTCA   -987

CTTGCAGTTTTGGTTTATTGTAAATGTAAACTTCAAAAAAGTGAATTCTCTGTCCAATCAATGTGTTTAA   -917
AP-1               OCT-1
TGTTGTATTGCTTAATGTATTCTGCTACACGCGATTATTTTTTGTAACTTTGTTTATTGCATCACCAAGC   -847
                                    C/EBPα                C/EBPα
TGTTCATGTAGCAAAAATAAAGTTTCTGACCATGTCTTTGTATATCTTTGTTATCTGCTGAGATCCCAAC   -777
            C/EBPα                                 GATA-1
CTGATAATAGTCTGTGTCTTTGTCATAAATCTTACAGCTGTTCCATTTTGCCAGATCAGAACACAGTTTG   -707
                                        YY1
GCAGCTTTTTAAATTCCATTCAGCATCAGTTCCCTGATGAGTACACACTTGATGGAGATTATAGTTACAA   -637
NF-1
GTAAACTACACAGGACATTTACATTGTGTTAAATAAGAAGCTTTAGCAACTACTGATAAGTTATATTAAA   -567
                        OCT-1                 GATA-1         C/EBPα
AAATCTGATTCACTAAACTGAAGGAATTGTTGTAATTAGGGAGTTTACAGATATACTTGAGCTCTACTCT   -497
                        C/EBPα        NF-κB         OCT-1
GCAAATCATTCCTGAACCATGTGTTGAAGAGTCAACAGCACCCACTACGTCCCAAATCCGAAGAGCACAA   -427
  C/EBPα                               C/EBPα

ACTTGATTTCTTGTAAATCACATGATCTGGGGGAAAACTGGCTGACTTGCAGCAACCAATGAGGGCACAT   -357
              Oct-1                   C-Jun
TGTGAAATAGAAACAAACACTTTCCGTGCAGCAAGCCAAGTAGTGATATAAAATGGACTTTTATAAAATC   -287
    ICSBP                            NF-1                              NF-κB
AAACATGCTGTAGTCTGAAATGCATCTGTCAGTCCACTATAATCTGATCTGTATTGTTAATCTCATGTTT   -217
                                                              C-Jun
TAGATACATCTGTTTAGATCTCATACTTTCATCTATGTAGAAATGTACACACTATTGAACAAACAAATGT   -147
                                                      HNF-3
GTCACACATGTATCTATACACATGAGTATACATGTTTATGTATTGCCAGTTAGCACATGCTATAGGAGAC    -77
                OCT-1                       NF-1
ATCACTCTGAGGAAAGATACTTCTTTTTAAATTATCTTAGAAGAAGAAAAAAACAGAGGGAGGGCACAGA    -7
    C/EBPα                       OCT-2                            SP-1
GAGATCCAACTATTGAGAAATACAGGGGTGGATGACTAACAAAACCTCAGTTGATCAAAAGCCTGGGTCA    64

GAGCAGGCAAAGAAACAGGAGCGGCACAATGAGGTGCTACCTCTTTTCGGTACTGTTCCTGATCTTGAGT   134
                            M   R   C   Y   L   F   S   V   L   F   I   L   S
```

Figure 31. 5' flanking region of Rb-C4

Putative transcription factor binding sites are shown in bold and underlined, with corresponding factor indicated directly below. The putative transcription initiation site is indicated by a bent arrow. Translated amino acids (uppercase) and start codons (italicized uppercase, in bold) are shown.

2.3.3 Identity, similarity, and phylogenic relationship of Rb-C4

Pairwise sequence analysis was revealed that *Rb-C4* was exhibited 66.5 % amino acid identity with its orthologous from Fugu fish (*Takifugu rubripes:* CAD45003.1) and 58.3% with Medaka (*Oryzias latipes*: AB025577.1). In addition, it was showed over 45% identity with the orthologous from other fish species. When compared with other vertebrates, it showed over 35.2% identity with human C4 (Table 12). In multiple sequence alignment, Rb-C4 showed higher degree of conservation among the species considered in the comparison. All the cysteine residues in the anaphylatoxin domain and NTR domain were found to be conserved across nearly all the orthologous, and certain (presumably important) signature motifs were also found to be highly conserved among all the orthologous (Figure 32).

Table 12. Percentage of identity, similarity, and gap of Rb-C4 with other C4 orthologous

Common Name	Species	Accession No.	AA	Identity %	Similarity %	Gap %
Fugu fish	*Takifugu rubripes*	CAD45003.1	1703	66.5	80.5	1.2
Medaka	*Oryzias latipes*	AB025577.1	1718	58.3	74.4	1.9
Rainbow trout	*Oncorhynchus mykiss*	CAD66666.1	1724	57.7	73.6	2.4
Common carp	*Cyprinus carpio*	AB037279.1	1716	48.6	66.4	3.5
Human	*Homo sapiens*	NP_009224.2	1744	35.2	52.7	9.5
Cattle	*Bos taurus*	NP_001159957.1	1741	34.9	51.9	9.8
Wild boar	*Sus scrofa*	NM_001123089.1	1741	34.5	51.6	9.5
Banded hound shark	*Triakis scyllium*	BAC82347.1	1693	34.3	53.2	10.7
Zebrafish	*Danio rerio*	ENSDARP00000031658	1779	33.1	53.4	8.4
Wild Turkey	*Meleagris gallopavo*	ACA64782.1	1699	29.7	46.8	12.9
Chicken	*Gallus gallus*	NP_001070701.1	1665	29	45.6	15.4

The evolutionary relationship of Rb-C4 was studied using the deduced amino acid sequences of Rb-C4 with sequences available in the NCBI database designated as complement C2, C3 and C4. In phylogenetic tree, Rb-C3 was located within the fish clade and separated from other higher vertebrate C3 genes. Furthermore, C3 group was shown closely associated with complement C4 gene group than C2 in phylogenetic tree, which indicates its close evolutionary relationship with C4 (Figure 18).

Figure 32. Multiple sequence alignment of the Rb-C4

Identical residues among all orthologous sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double-arrows. The cleavage site of α and β subunits is boxed and thioester sites are indicated in box. Important domain regions were marked by double-headed arrows.

2.3.4 Tertiary structural model comparison of Rb-C4 protein

According to the protein homology modeling, the Rb-C4 was found to resemble the human C4 crystal structure. The β and α chains of Rb-C4 were modeled as separate entities for easy comparison. All the MG domains (MG1-MG8) were highly resembled that of human structure. In human, MG1-MG5 and half of MG6 are located in β chain, and half of MG6, MG7and MG8 belong to the α chain. The β chain end up with linker (LNK) domain and α chain begins with anaphylatoxin domain (Figure 33). The thioester domain and its thioester site were buttressed by interactions with MG2, MG8 and CUB domains.
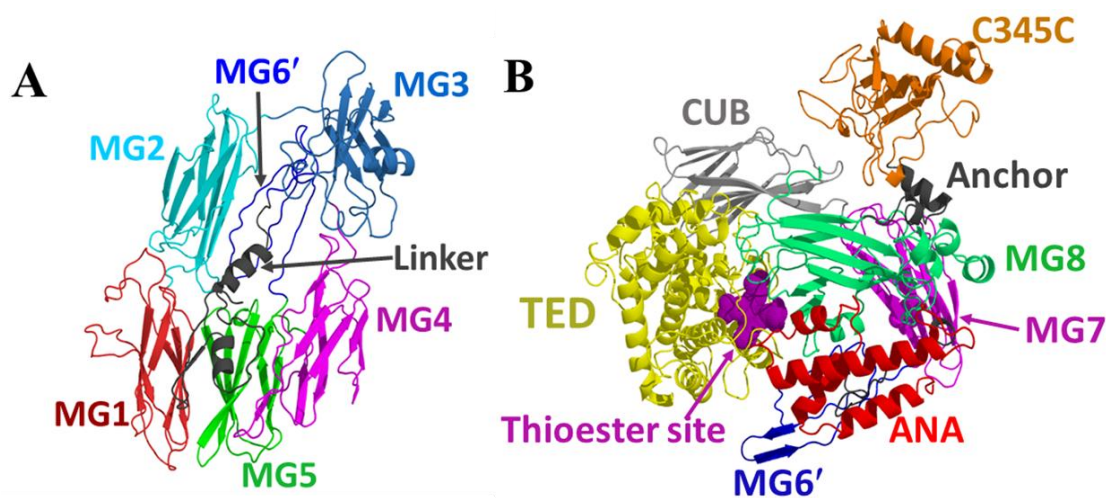
Figure 33. 3D-molecular structures of β and α chains of Rb-C4
The A and B represent the β chain and α chain of rock bream respectively. Each domain is labeled with respective color near the domain. The thioester site of α chain is labeled in purple indicated by purple arrow.

2.3.5 Tissue-specific mRNA expression of Rb-C4 in normal healthy fish

To examine the normal tissue distribution pattern of *Rb-C4* transcription, qPCR was carried out for blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine (Figure 18). *Rb-C4*was found to be constitutively expressed in all examined tissues, although

the level of expression varied between each. As the all complement components, *Rb-C4* also indicated its liver-specific expression compared to the normalized expression level in muscle.

2.3.6 Modulation of Rb-C4 expression in response to immune stimulation

The mRNA level of *Rb-C4* was measured in head kidney from rock breams challenged with *E. tarda, S. iniae*, RBIV, or LPS by qPCR (Figure 34). The expression pattern upon Gram-negative bacteria *E. tarda* was displayed an increasing trend until 24 h with the highest expression at 24 h post stimulation (P.S). The same patterns of induction were observed in response to *S. iniae,* LPS, poly I: C and iridovirus challenges as well. In *S. iniae* challenge, *Rb-C4* was showed significant induction at 6 h p.i. The highest up-regulation was recorded for the LPS challenge at 24 h p.i. According to the overall challenge experiment, *E. tarda* could eminently induce the Rb-C4 transcript level at 24 h p.s. compared to the other stimuli (Figure 34).
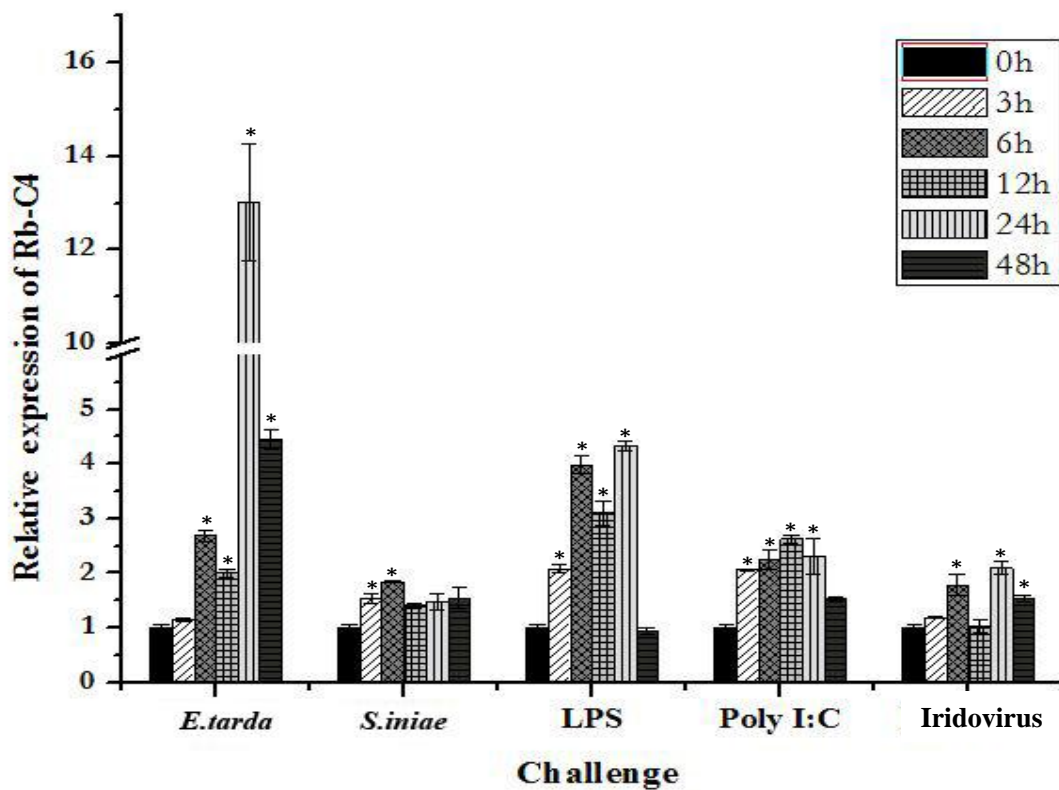
Figure 34. Relative expression of Rb-C4 mRNA in head kidney after challenge of rock bream with *E. tarda*, *S. iniae*, LPS, Poly I: C and iridovirus, as detected by qPCR
The amount of mRNA relative to the internal control, β-actin, is expressed as mean ± SD (n = 3). Significant differences ($P < 0.05$ vs. untreated control at 0 h) are marked with asterisks.

2.3.7 Recombinant protein level activity of Rb-C4 anaphylatoxin domain

2.3.7.1 Prokaryotic expression and purification of the soluble recombinant Rb-C4a

We introduced the anaphylatoxin domain (C4a) of Rb-C4 into pMAL-c2X expression vector and transformed the recombinant construct to express the Rb-C3a protein in *E. coli* BL21 (DE3) cells through IPTG-mediated induction. The purified protein was analyzed on SDS-PAGE (Figure 35 A). Purified recombinant Rb-C4a fusion protein appeared as a single band with a molecular size of ~52.13 kDa, which was close to the size of calculated molecular mass of the fusion protein (Rb-C4a, 9.63 kDa + MBP, 42.5 kDa).

2.3.7.2 Respiratory burst activity of the anaphylatoxin domain (C4a) of Rb-C4

The measurements of the respiratory burst activity over different treatments were plotted in Figure 35 B. Compared to the negative control (un-stimulated sample), absorbance of the samples in *Rb-C4a* treatment was significantly increased ($P < 0.05$). It could prominently increase the production of reactive oxygen species (ROS) compared to negative control. Furthermore, in the case of *Rb-C4a*+SOD treatment, the active SOD was shown to be able to reduce ROS production level close to negative control (Figure 35B).
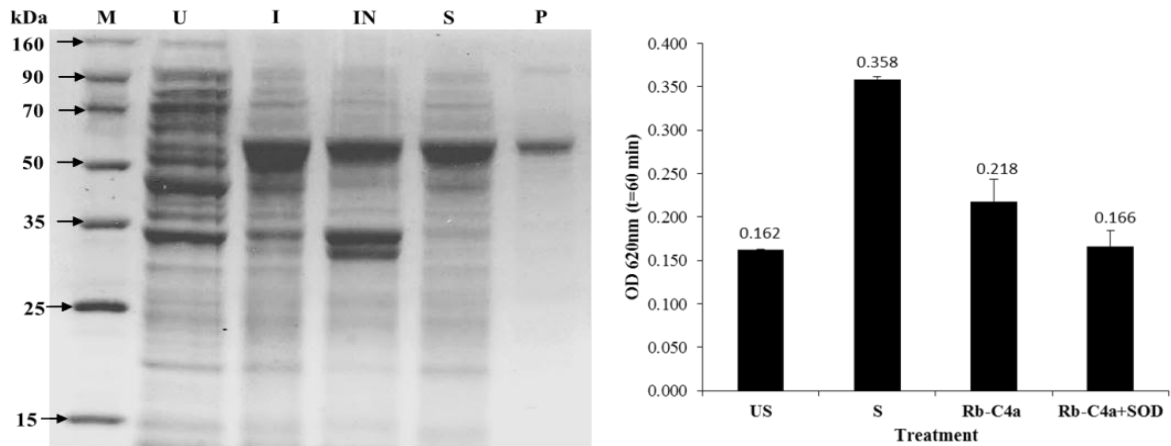
Figure 35. SDS-PAGE analysis of overexpressed and purified recombinant Rb-C4a protein (A) and respiratory burst activity of the recombinant protein (B)
Lanes: M, molecular mass marker (sizes are indicated in kilodaltons); U, total cellular extract from *E. coli* BL21 (DE3) before IPTG induction; I, total cellular extract from *E .coli* BL21 (DE3) after IPTG induction;   IN, insoluble fraction of proteins after IPTG-induction; S, soluble fraction of proteins after IPTG-induction; P, purified recombinant fusion protein Rb-C3a/MBP. **(B)** Comparison of respiratory burst activity of recombinant Rb-C4a. US, un-stimulated (-ve control); S, stimulated by H_2O_2 (+ve control); Rb-C4a, 20 µg of Rb-C4a treated; Rb-C4a+SOD, 20 µg of Rb-C4a and 20 µL of 100 IU/mL of superoxide dismutase.

2.4. Results

2.4.1 *In silico* characterization of Rb-C5

Based on BLAST search analysis of EST data from the rock bream normalized cDNA library, we identified a putative complement component 5 and designated it as *Rb-C5*. The sequence was deposited in NCBI GenBank database under the accession number of KF408222. The complete cDNA sequence was 5342 bp in length, including a 29 bp 5'-untranslated region (5′-UTR), a 5064 bp open reading frame (ORF) encoding 1687 amino acids, and a249 bp 3'-UTR with a polyadenylation signal. The predicted molecular mass of *Rb-C5* was187 kDa and the theoretical isoelectric point was 6.7. According to SignalP prediction, the *Rb-C5* possessed a signal peptide of 19 residues, which was presumed to facilitate the extracellular localization (Figure 36). The mature protein of *Rb-C5* was contained 13 domains similar to human C5,

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

which are separated mainly into β and α chains by a linker ^{676}RPKR^{679}. The molecular weights of deduced β and α chains were 73 kDa and 111 kDa, respectively. The domain structure of *Rb-C5* was composed of eight α_2 macroglobulin domains (MG 1-8) distributed in both chains, an anaphylatoxin domain which is also conserved in complement C4 and C3 in vertebrates, thioester domain (TED) and common C-terminal C345C domain or netrin domain (NTR) (Fujito et al., 2010; Janssen et al., 2005). The anaphylatoxin domain of *Rb-C5* was contained 76 amino acids and consisted 6 conserved cysteine residues where they formed internal disulphide bond between designated two residues. Three connecting regions were identified such as linkers and the other regions at either side of the TED domain (CUB regions). Moreover, the MG6 is separated into both β and α chains which ultimately get connected upon release of C3a anaphylatoxin domain (Figure 36). The NTR domain which composed of 130-residues contains six conserved cysteine residues which are likely to form internal disulphide bonds. Moreover, the NTR is anchored by an anchor domain to the 8th MG domain in C5.

```
                Signal peptide                                                  MG1
◄──────────────────────────────►  ◄──────────────────────────────────────────────────────────────►
MKVCVLLMCVCCLCWRTEADPRSYLITAPLSLRLDAVETVLLQLFGFTEEVRAYVYLKTSMAPDHVVLAQEVVTL        75

─────────────────────────────────────────────────────────►◄
NTQNNHQAAAQVRLHPGRLGKSVSHVILHVQSPEINQHLSIAVSRSNGFLFIQTDKPLYTPHQSVKVRAFSLNQE       150
                                          MG2
─────────────────────────────────────────────────────────────────────────►◄
LRPANRSVFLTFKDPDHTTVDIVEMFDVNNGIPSMQNPFKIPIKPKLGIWSIEASYSDDFTTSAKTDFEVKEYVL       225
                                 MG3
─────────────────────────────────────────────────────────────────►
PSFSILVEPEVNFISYGSFDHFNFKISARYLRGAPVADGEVFLRYGYISGKSPPVLIPSSVSRERLSSSGEVDVT       300

─────────────────────────────────────────────────────────────────►◄
INMTKVLSKHDGPKNLNSLIGKYLYITVLVQEDTGGITQEAEFAAVKFVKSPYSLSLVSTPPFIKPGLPYNIQVL       375
                                MG4
─────────────────────────────────────────────────────────────────►
VKDHLDKPVNQVRVRLVERQLIRQGRENEEIPCPDSANSQSDGLAVFICNTPSEGIKAVLKFETADPNLPAASQA       450

───────────────────────────────────────────────────►◄
SLSLEAVAYHSPNQRYLYIDPPLPSRSLEVGRYSNIKVYSATPAYVPVRALSYLVLSKGKVVDFGSRKFVSTSDN       525
                          MG5                                          MG6ᵝ
─────────────────────────────────────────────────────────►◄
KQTLSFEVTTSMVPSIRLLVYYILYGEGTSELVADSVWLDVKDKCVNGLQTDLSYRRRDYKPKENLQLDIQTNKE       600
                                 Linker                                   β Chain
                  ◄───────────────────────────────────────────────────────────────►
GLVALSAV*DTALFTLRPNYRDPVTMVLRQIEHSDQGCGGGGGRDSADVFRLAGLTFITNANAHPSTSDEACTAVV*       675
        α Chain
◄──────────►
RPKR**ALTDEEKTKKAESYGRQKTCCELGMRYIPKGVTCHHFAKQVYRNPKHHLCREAFKACCEFIQQHLGQDHNL**  750
            Linker'                              MG6ᵅ
**ILGRN***ELGSDFDQAPSLVRSYFPESWMWE*VQRVSPGKLSLTRPLPDTLTTWKVKAVGMFKDGMCVAEPVQVSVSL  825
                                  MG7
PLSVDIPLPYQVVRGEQLELKGSVYNQQLDNIKYCVTLTVGPALCLLQSQPAAGGLHSTACSWTSLSARGVGRVT       900
                                    CUB¹
FTVLGLEPGEHTLTFTLKTRHGQKDILEK*KLRVVPEGVRKEMFSGGTLDPQGLYGSQKTTVEVKNYLPTNMVPNA*     975
**AVERMLTIN**GEVVGDILSVMHSPEGLRQLVNLPAGSAEAELGGLLPLVQVHQYLETTRSWDVLGADIHKNSADLR 1050

**QKIKEGLVSISSFRRGDSSYSMWSKRESNTWLTALVVKALSVVDPVVDSVADSDRNHQKLSESVAWLIRSAQQPD** 1125
                                   TED domain
**GSFAEESSFRPNKVVAAGTDAVGQSVYLTSFVLIALHRATSIKDQILQLRFHEDSKRSAANYISQHALAVKSVYV** 1200

**RAVATYALTLHDPNSMAASELLSSLEREGRQKGHPAVLRYWQESSVTGDWLKPDQSSGLTVETTAYVLLTVLLK**G 1275
                                                               CUB²
**RIADANPILSWLTQDQRYGEGFYSIQDIVLTLEAVTEYHRVVPRAV**LSQEINIRYKTKGALGRVQLSQSRPVATP 1350
                           ◄──────────────────────────────────────────────────────
IQVTNNEDIIASTGFGRGVSNVKMKTVYYETTSS TQTCNFDLTVEVARSPSLRSPHLVACAKYNPPPNEMFTESS     1425
                                  MG8
LTVMKIQLPTGVEAYLEDLKQFRDAAEPLISHYELQGNTVVIQADAVPSEVFLCVSFRIRTGFRVSGVTESLFSV      1500
◄──────────────────►           Anchor
YEPQDKGSMCTKQF*TYQDQKLQSLCVGDQCQCMTAA*Ⓒ AYRGKIDVTLTAAQRTDETⒸ KPNIKYAYKVTVKSSAA     1575
                                    C345C/NTR domain
EGDFMTYTASVVEVLKNTDKAFEAVSSGTEVDLLKKITⒸ NAVDIQNNKQYLVMGASGSEVTASHGFKYRLPLDSE     1650

ALVELWPTDⒸSSPEⒸLDYISQLEDFALDLQLASⒸPDS                                            1687
```

Figure 36. Protein sequence of Rb-C5

The α and β chain demarcation is indicated by arrowheads while labeling, respectively. The potential cleavage processing site of α and β subunits is boxed. The domains identified; signal peptide and macroglobulin (MG) 1-8 are indicated by double arrows with labeled above. An anaphylatoxin domain is bold and underlined. The thioester domain (TED) is in bold phase and CUB (CUB[1] and CUB[2]) domain is indicated by white letters highlighted in gray at either side of the TED domain and NTR domain is highlighted in black at the C-terminus. The linker (LNK), N terminus of C3α chain (α') and anchor are denoted by italicized black letters and shaded in gray. The conserved cysteine residues of C345C / NTR are encircled.

Signal peptide MG1

Oplegnathus fasciatus	:	MKVCVLLMCVCCLCWRTEADPRS...	: 109
Takifugu rubripes	:	----------------MAPPLR...	: 92
Oreochromis niloticus	:	------------MWERVRTPGGP...	: 51
Danio rerio	:	-MARLLFCIFNLFVCFTVQENV...	: 108
Homo sapiens	:	-MGLLGILCFLIFLGKTWGGQEQT...	: 110
Bos taurus	:	-MGLWGILCFLIFLGKSWGGQEQT...	: 110

MG2

Oplegnathus fasciatus	:	INQHLS...	: 221
Takifugu rubripes	:	INQHLS...	: 204
Oreochromis niloticus	:	-LRYTITL...	: 150
Danio rerio	:	FTDNVK...	: 219
Homo sapiens	:	FSKSKR...	: 220
Bos taurus	:	FSKAKK...	: 220

MG3

Oplegnathus fasciatus	:	EYVLP...	: 330
Takifugu rubripes	:	EYVLP...	: 313
Oreochromis niloticus	:	EYVLP...	: 259
Danio rerio	:	EYVLP...	: 328
Homo sapiens	:	EYVLH...	: 329
Bos taurus	:	EYVLH...	: 329

MG4

Oplegnathus fasciatus	:	QBDTGG...	: 438
Takifugu rubripes	:	QBDTGG...	: 421
Oreochromis niloticus	:	EDSTGG...	: 367
Danio rerio	:	QES TGG...	: 440
Homo sapiens	:	IES TGG...	: 436
Bos taurus	:	IESTGG...	: 436

MG5

Oplegnathus fasciatus	:	AAPNLF...	: 550
Takifugu rubripes	:	AAPVLF...	: 526
Oreochromis niloticus	:	AACASS...	: 479
Danio rerio	:	TLD---ASW...	: 547
Homo sapiens	:	TLDAPDL...	: 547
Bos taurus	:	PEDLLD...	: 547

MG6β Linker

Oplegnathus fasciatus	:	GBGTG...	: 658
Takifugu rubripes	:	GBGIG...	: 634
Oreochromis niloticus	:	GBGTG...	: 587
Danio rerio	:	GEEKA...	: 659
Homo sapiens	:	GEQTA...	: 655
Bos taurus	:	GEQTA...	: 655

β Chain   α Chain   Anaphylaxin domain

Oplegnathus fasciatus	:	NANHP...	: 758
Takifugu rubripes	:	NANQA...	: 732
Oreochromis niloticus	:	NANTAP...	: 693
Danio rerio	:	NANAKA...	: 764
Homo sapiens	:	NANDD...	: 755
Bos taurus	:	NANDD...	: 755

Linker   MG6α

Oplegnathus fasciatus	:	SDFDQAPSL...	: 867
Takifugu rubripes	:	ADFDQAPSL...	: 844
Oreochromis niloticus	:	ADFDLAPSL...	: 802
Danio rerio	:	FLMDAPLSQ...	: 872
Homo sapiens	:	TLLPVSKPE...	: 863
Bos taurus	:	SLLPVTKPE...	: 863

Figure 37. Multiple sequence alignment of Rb-C5 with its orthologous
Identical residues among all orthologous' sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double headed arrows. The cleavage site of α and β subunits is boxed. Important domain regions are marked by double headed arrows while labeling respectively.
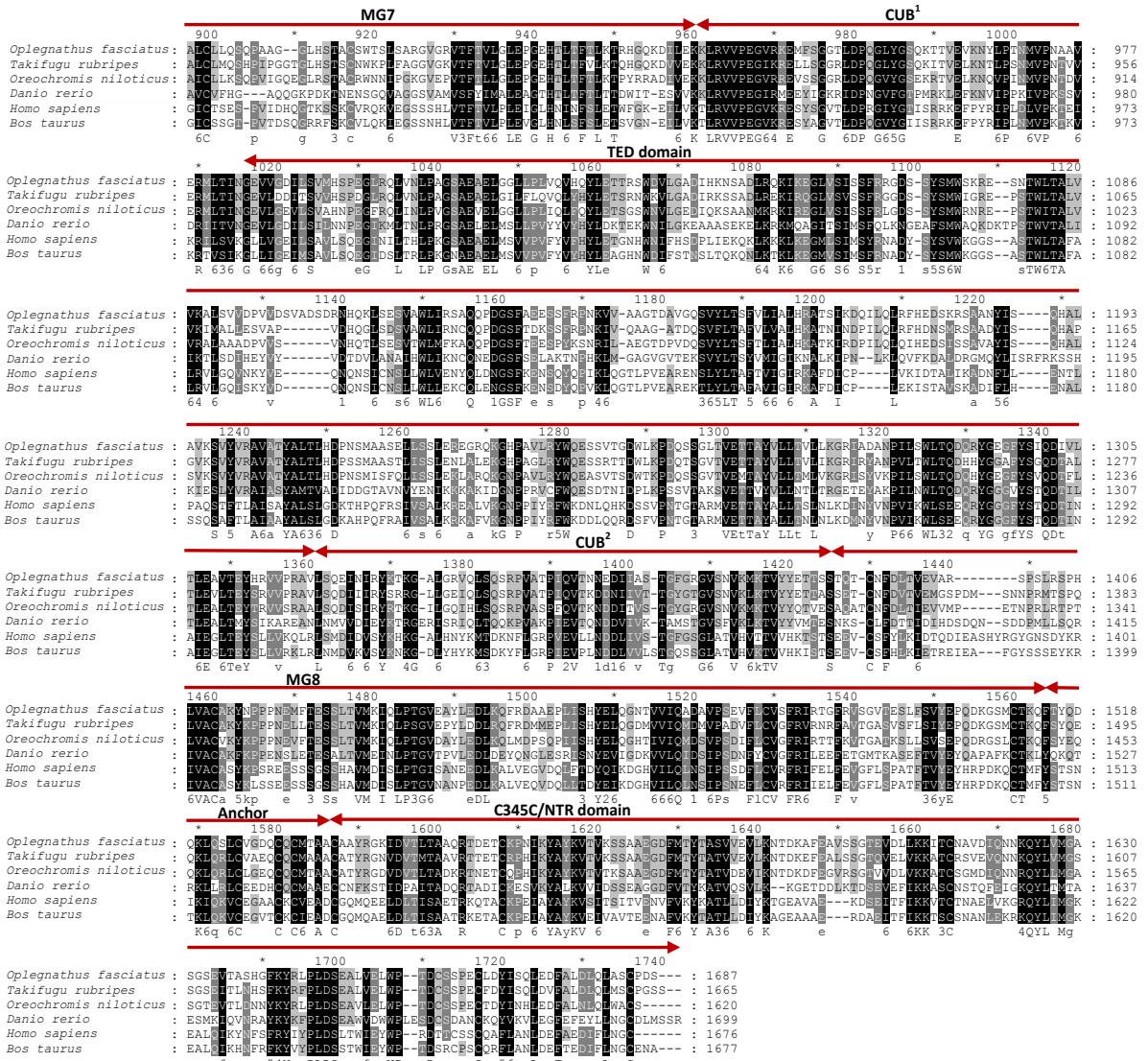
Figure 37. Continued

2.4.2 Identity, similarity, and phylogenic relationship of Rb-C5

Pairwise sequence analysis was revealed that *Rb-C5* was exhibited 74.6% amino acid identity with its orthologous from Puffer fish (*Takifugu rubripes:* XP_003974669.1) and 70.4% with Ice Blue Zebra (*Maylandia zebra*: XP_004571725.1). Also, it was showed over 45% identity with the orthologous from other fish species. When compared with other vertebrates, it was showed over 35 % identity (Table 13). In multiple sequence alignment, Rb-C5 was showed higher degree of conservation among other vertebrates. Most of the cysteine residues were

99

found to be conserved among almost all the orthologous, and certain (presumably important) signature motifs were also found to be highly conserved among all the orthologous (Figure 37).

Table 13. Percentage of identity, similarity, and gap of Rb-C5 with other C5 orthologous

Common Name	Species	Accession No	AA	Identity%	Similarity%	Gap%
Puffer fish	*Takifugu rubripes*	XP_003974669.1	1665	74.6	85.8	2.7
Ice Blue Zebra	*Maylandia zebra*	XP_004571725.1	1666	70.4	83.7	2.9
Nile Tilapia	*Oreochromis niloticus*	XP_003455551.1	1620	66.8	79.3	7.3
Medaka/Ricefish	*Oryzias latipes*	XP_004073022.1	1682	63.7	79.7	1.5
Common carp	*Cyprinus carpio*	BAC23058.1	1684	43.8	62.7	4.9
Zebrafish	*Danio rerio*	XP_001919226.3	1699	43.2	62.1	6.1
Cattle	*Bos taurus*	NP_001160088.1	1677	38.3	57.2	5.8
Human	*Homo sapiens*	AAA51925.1	1676	37.7	57.9	5.6
Norway rat	*Rattus norvegicus*	XP_001079130.2	1688	37.7	57.2	5.7
Pig	*Sus scrofa*	NP_001001646.1	1677	37.3	58.2	4

The evolutionary relationship of Rb-C5 was studied using the deduced amino acid sequences of Rb-C5 with sequences available in the NCBI database for the complement C2, C3 and C4. In phylogenetic tree, Rb-C5 was located within the fish clade and was separated from other higher vertebrate C5 genes. Furthermore, C5 group was shown closely associated with complement C3 and C4 gene groups than C2 in phylogenetic tree, which was indicated its close evolutionary relationship with C3 and C4 (Figure18).

2.4.3 Tertiary structural model comparison of Rb-C5 protein

The protein homology modeling was carried out to identify the structural similarity of Rb-C5 with the crystal structure of complement C5 in human. The β and α chains of Rb-C5 were modeled as separate entities for easy comparison. In comparison to the human model, Rb-C5 also showed the very similar structure characteristics at tertiary level. All the MG domains

(MG1-MG8) were highly resembled to that of human structure. In human, MG1-MG5 and half of MG6 are located in β chain, and half of MG6, MG7and MG8 belong to the α chain. The β chain end up with linker (LNK) domain and α chain begins with anaphylatoxin domain (Figure 38). The Thioester domain is an important domain, which is buttressed by interactions with MG2, MG8 and CUB domains. However, thioester site was not identified with in the thioester domain of the Rb-C5.
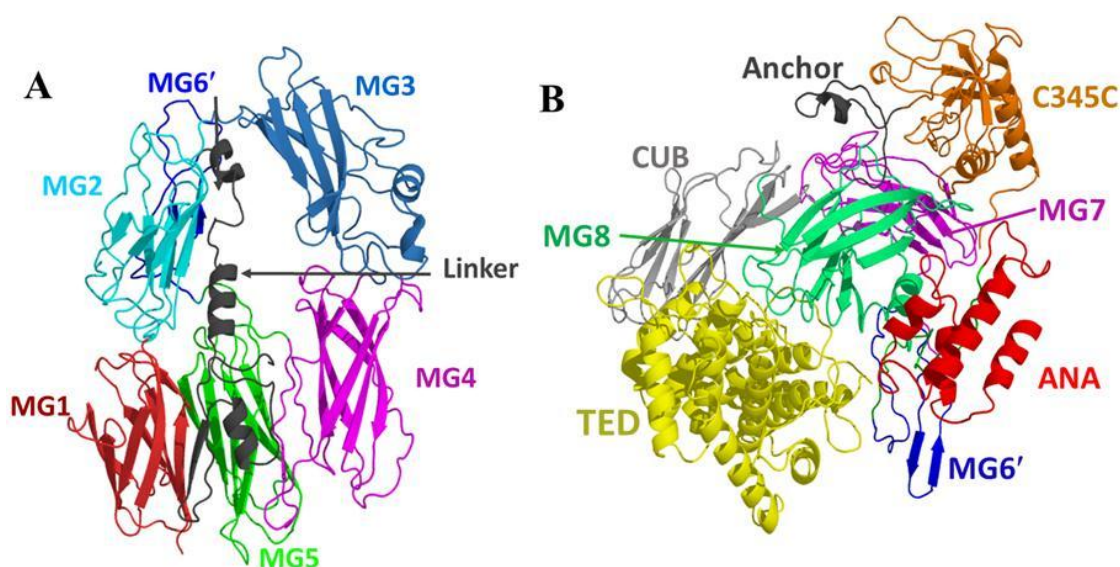
Figure 38. 3D-molecular structures of β and α chains of Rb-C5
The A and B represent the β chain and α chain Rb-C5, respectively. Each domain is labeled with respective color near the domain.

2.4.4 Tissue-specific mRNA expression of Rb-C5 in normal healthy fish

To examine the normal tissue distribution pattern of *Rb-C5* transcription, qPCR was carried out using blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine of healthy rock breams (Figure 19). *Rb-C5* was found to be constitutively expressed in all examined tissues, although the level of expression varied between each. As the all complement components, *Rb-C5* also indicated its liver-specific expression compared to the normalized expression level in muscle.

2.4.5 Modulation of Rb-C5 expression in response to immune stimulation

Out of all components, complement C5 is the most important protein and readily available in the blood stream of animals. In order to test its mRNA expression, immune challenge experiments were employed and expression levels were compared in head kidney of rock breams challenged with *E. tarda, S. iniae*, RBIV, or LPS by qPCR (Figure 39). The expression pattern upon Gram-negative bacteria *E. tarda* displayed a biphasic expressional response. The amount of *Rb-C5* transcripts were significantly ($P<0.05$) increased from baseline (0 h) between 3 h and 6 h, declined to baseline at 12 h and again increased at 24 h post-injection (p.i.), with giving the maximum level at 48 h p.i. (Figure 39). The same biphasic patterns of induction were observed for the LPS poly I: C and iridovirus challenges as well. In *S. iniae* challenge, *Rb-C5* was showed significant induction at 3 h to 48h p.i., where the highest up-regulation was reported at 6h p.i. Out of all challenges, transcriptional induction upon Poly I: C stimulation at 3h p.i. was found to show the highest fold-difference.
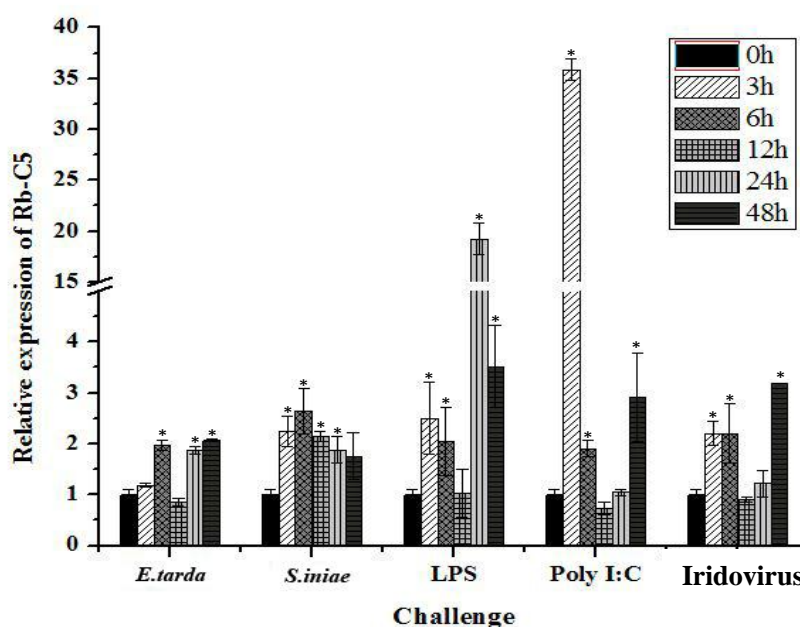
Figure 39. Relative expression of Rb-C5 mRNA in head kidney after challenge of rock bream with *E. tarda*, *S. iniae*, LPS, Poly I: C and iridovirus, as detected by qPCR.
The amount of mRNA relative to the internal control, β-actin, is expressed as mean ± SD (n = 3). Significant differences ($P < 0.05$ vs. untreated control at 0 h) are marked with asterisks.

2.4.6 Physiological characterization of Rb-C5a at protein level

2.4.6.1 Prokaryotic expression and purification of the soluble recombinant Rb-C5a

We introduced the anaphylatoxin domain (C5a) of Rb-C5 into pMAL-c2X expression vector and transformed the recombinant construct to express the Rb-C5a protein in *E. coli* BL21 (DE3) cells through IPTG-mediated induction. The purified protein was analyzed on SDS-PAGE (Figure 40A). Purified recombinant Rb-C5a fusion protein was appeared as a single band with a molecular size of ~51.36 kDa, which is close to the size of calculated molecular mass of the fusion protein (Rb-C5a, 8.86 kDa + MBP, 42.5 kDa).

2.4.6.2 Respiratory burst activity of the anaphylatoxin domain (C5a) of Rb-C5

The measurements of the respiratory activity over different treatments were plotted in Figure 40B. Compared to the negative control (un-stimulated sample), absorbance of the samples in *Rb-C5a* treatment was significantly increased ($P < 0.05$). It was showed 1.4-fold increment of reactive oxygen species (ROS) production compared to negative control. Furthermore, in the *Rb-C3a*+SOD treatment, the active SOD was shown to be able to reduce ROS production level below to negative control (Figure 40B).
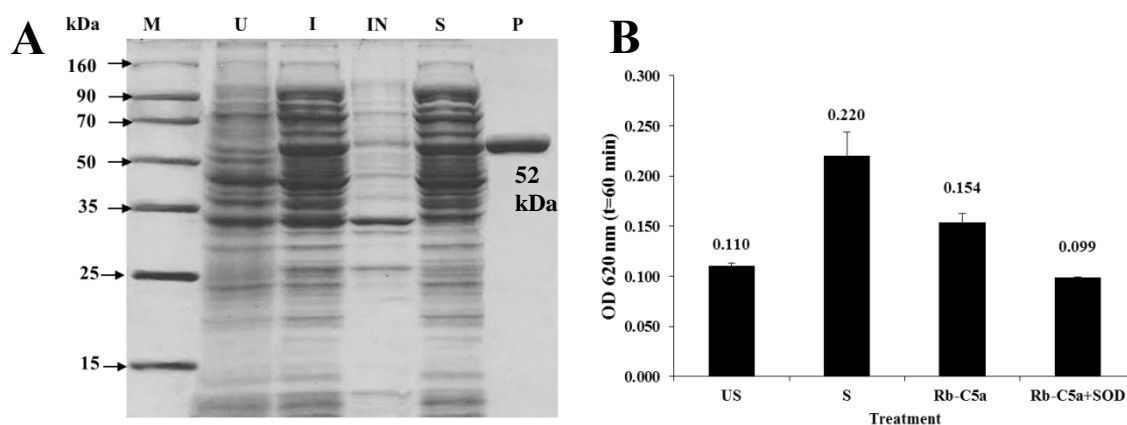
Figure 40. SDS-PAGE analysis of overexpressed and purified recombinant Rb-C5a protein (A) and respiratory burst activity of the recombinant protein (B)

103

Lanes: M, molecular mass marker (sizes are indicated in kilodaltons); U, total cellular extract from *E.coli* BL21 (DE3) before IPTG induction; I, total cellular extract from *E.coli* BL21 (DE3) after IPTG induction; IN, insoluble fraction of proteins after IPTG-induction; S, soluble fraction of proteins after IPTG-induction; P, purified recombinant fusion protein Rb-C5a/MBP. **(B)** Comparison of respiratory burst activity of recombinant Rb-C5a. US, un-stimulated (-ve control); S, stimulated by H_2O_2 (+ve control); Rb-C5a, 20 μg of Rb-C5a treated; Rb-C5a+SOD, 20 μg of Rb-C5a and 20 μL of 100 IU/mL of superoxide dismutase.

2.5 Discussion

In this study of rock bream, we interested in identifying genomic structure of the complement C3, C4 and C5 and their unique biological features of active domains, transcriptional evidences for the involvement in immune mechanism of rock bream in contrast to the mammalian counterpart.

Complement C3 was predicted to be emerged over 700 million years ago and it harbors characteristics features of the α2-macroglobulin (α2M) family (Janssen et al., 2005; Sunyer et al., 1998b). The C3 is most abundant complement components in serum basically circulate as an inactive protein (Sahu and Lambris, 2001). Furthermore, it is the central component which is activated by three signaling pathways (Nakao et al., 2003). Therefore, it plays a vital role in activating complement cascade and other immune pathways. The complement C3 ,C4 and C5 share common sequence characteristics and domain architecture. They are the highest molecular weight proteins and contain two chains with disulphide bonds. All three proteins consisted of similar domain architecture as 8 domains of $α_2$ macroglobulins, β-α processing motif (RXXR), anaphylatoxin domain, thioester domain and C345C / nectin domain (NTR). The anaphylatoxin domain is consisted of six cysteine residues with intra-disulphide bonds and plays a critical role in protecting thioester site in the native C3. Thus, it prevents spontaneous activation of the C3 and helps to avoid needless tissue damage in the host. However, there is no direct structural contact between anaphylatoxin domain and TED (Janssen et al., 2005). The netrin domain is an about 130-residue domain consisted of 6

104

conserved cysteine residues found in complement C3, C4 and C5. Furthermore, it has been recorded in secreted frizzled-related proteins, and type I procollagen, C-proteinase enhancer proteins (PCOLCEs), as well as in the N-terminal parts of tissue inhibitors of metalloproteinases (TIMPs) (Banyai and Patthy, 1999; Trexler et al., 2001). It has been associated with many other domains such as CUB domain in the complement C3, WAP, Kazal, Kunitz, Ig-like, laminin N-terminal, laminin-type EGF or frizzled (Trexler et al., 2001). The NTR domain of Rb-C3 Rb-C4 and Rb-C5 are also consisted of six conserved cysteine residues among vertebrates and commented to the CUB domain in tertiary structure may imply that it is similar to the identified features of other NTR domains in vertebrates. Therefore, considering the sequence features and domain architecture, we can therefore hypothesize that they are belong to the execution phase of the complement pathway and act as in other vertebrates.

As described by the previous reports, anaphylatoxin domains enhanced respiratory burst activity of leukocytes in mammals (Ehrengruber et al., 1994; Elsner et al., 1994a; Elsner et al., 1994b). The functional role of anaphylatoxin domains in teleost complement system is poorly understood. In our study, we used recombinantly expressed anaphylatoxin domain of Rb-C3, Rb-C4 and Rb-C5 to measure their potent activities using NBT reduction assay. In the assay, recombinant proteins of Rb-C3, Rb-C4 and Rb-C5 were able to significantly induced respiratory burst activity in head kidney cells. The results were also compared with positive control (H_2O_2 treatment) and SOD treatment. The ROS generating ability of H_2O_2 in human neutrophils was demonstrated by Winn et.al (Winn et al., 1991). SOD is known to be a potent convertor of O_2^- into O_2 and H_2O_2. Therefore, the reduction of ROS production in the SOD treated sample confirmed all three proteins are able to induced O_2^- production. To sum up, we can confirm that Rb-C3, Rb-C4 and Rb-C5 are biologically active and play a vital role in complement-mediated elimination of invaded pathogen in rock beam.

The thioester site is highly reactive and readily hydrolyzed by water (Blandin and Levashina, 2004). In native C3 and C4, the thioester sites in α chain are protected within hydrophobic molecules that exposed only when C3 cleaved into b component by C3 convertases (Sahu and Lambris, 2001). Upon activation of C3 by the proteolytic enzymes, thioester is transformed into free thiolate anion and form acyl-imidazole, which may be stabilized by Glutamic acid (Dodds et al., 1996; Gadjeva et al., 1998). This configurational change in the molecular structure is essential for binding ability of the active C3 into target organism (Law and Dodds, 1997). Rb-C3 also has the thioester site located between anaphylatoxin domain, MG8 and CUB regions. Tertiary structural arrangements of the Rb-C3 indicate that its thioester site also may be protected among the above domains (Figure.5). Once C3a cleaved and released, thioester site is exposed for hydrolysis. The β-α processing motif (RXXR) of human and some teleost are having four Arginine residues (RRRR) (Janssen et al., 2005; Zarkadis et al., 2001). However, Rb-C3 has three Arginine resides and one Lysine (RRKR), which is similar to Spotted wolfish (Abelseth et al., 2003). However, in orange spotted grouper, C3 β-α processing motif possesses RKKR (ADU33222.1). The above evidence suggests that β-α processing motif in fish C3 gene may contain alternative Lysine residues besides Arginine. The Rb-C4 harbors RQRR as the β-α processing motif whereas Rb-C5 contains RPKR. Therefore, it can be concluded that RXXR may exists as β-α processing motif in fish complement C3,C4 and C5. Complement C5 is homologous to that of C3, C4 and other α_2-macroglobulin superfamily proteins. However, it does not contain a Cys-Gln thioester bond (Fredslund et al., 2008).

Past studies have shown that C3 is expressed in wide range of tissues and different life stages, such as juveniles and adults, in vertebrates (Lange et al., 2004; Lovoll et al., 2006; Mishra et al., 2009). The tissue-distribution study of *Rb-C3, Rb-C4 and Rb-C5* showed that the liver was the main producer of these genes. Furthermore, the *Rb-C3* transcripts were

constitutively expressed in many tissues, which are consistent with the previous reports on teleost fish, such as common carp and Atlantic salmon (Lovoll et al., 2007; Mishra et al., 2009). Furthermore, liver-specificity is one of the unique features of complement genes in vertebrates (Morgan and Gasque, 1996). The Earlier studies on rock bream complement lytic pathway genes including C7, C8α, C8β, and C9 also showed liver-specific expression, which provides further evidence to confirm the tissue specific expression patterns of complement genes (Wickramaarachchi et al., 2012; Wickramaarachchi et al., 2013a; Wickramaarachchi et al., 2013b). However, extra-hepatic cells in mammals, such as macrophages, monocytes, fibroblasts, B lymphocytes, polymorphonuclear leukocytes, type II pneumocytes, astrocytes and microglial cells, have been identified to produce complement C3 proteins (Lambris, 1988; Mishra et al., 2009; Rothman et al., 1989; Strunk et al., 1994; Strunk et al., 1985). In rock bream, peripheral blood cells have been detected as the second large C3 and C5 producing tissue, which is also an evidence for extra-hepatic production. Those extra-hepatic expression of complements are assumed to be important for modulation of immune responses in respective tissues.

Several complement C3 genes identified from teleost and invertebrates have showed strong transcriptional response to LPS challenge and live bacterial challenge (Giacomelli et al., 2012). The similar treascriptional responses were observed in *Rb-C3, Rb-C4* and *Rb-C5.* In our study, highest expression of *Rb-C3* was observed in early time-pints (3 h) after pathogenic challenges, which may provide evidence that C3 is immediately activating gene upon identifying target molecules in the host. However, the highest expressions of complement lytic pathway genes (C7, C8 and C9) in rock bream were previously observed in late time-pints (12 h or beyond) after pathogenic challenges (Wickramaarachchi et al., 2012; Wickramaarachchi et al., 2013a; Wickramaarachchi et al., 2013b). Therefore, there is a chronological relationship of *Rb-C3* and lytic pathway gene expression.

107

Indeed, no detail study has been conducted on the promoter regions of the complement component 3,4 or 5 in teleost fish. In our study, we tried to identify initiation sites and some of the important immune related transcription factor binding sites based on the predictions in the 5' flanking regions of *Rb-C3 and Rb-C4*. The canonical TATA or CCAAT box was not found in the putative core-promoter region in both genes. Therefore, they may be categorized to the TATA-less complements which include human C4 (Yu, 1991),C2 (Horiuchi et al., 1990),C1 INH (Carter et al., 1991), CR1(Vik and Wong, 1993). According to our analysis, several important putative transcriptional factor-binding sites were identified in the 5′-flanking region of *Rb-C3* (Section 3.2). Several studies have been carried out in promoter regions of the complement genes of the human, such as some specific transcription factors or region of C3 (Juan et al., 1993; Wilson et al., 1990), C7 (Gonzalez et al., 2003), complement factor H (Juan et al., 1993), and complement factor I (Minta et al., 1998). Furthermore, the promoter regions of lytic pathway genes in fish were studied based on prediction, such as grass carp C7 (Shen et al., 2012a), grass carp C9 (Li et al., 2007) and rock bream C7, C8 and C9 (Wickramaarachchi et al., 2012; Wickramaarachchi et al., 2013a; Wickramaarachchi et al., 2013b). The *Rb-C3 and Rb-C4* contain C/EBP, AP-1 and HNF-3 transcription factor binding sites in common with most of the complement components, which indicates these factors may play important role in transcription of complement genes. However, further studies based on empirical data are required to delineate the functional elements of *Rb-C3* promoter.

The comparative genomic structural study has not been performed yet in the complement C3 and C4 genes in teleost. Therefore, this is the first comprehensive study on genomic structure of complement C3 in teleost. In our analysis, some of exons are highly conserved among the fish and other selected vertebrates. However, both genes were found to be having biggest genomic structures compared to other complement components.

In summary, we have identified the complete genomic sequence of complement component 3, 4 and 5 from rock bream. This is first reveal of the C3 and C4 genomic structures and the protein structure homology modeling in fish. Genome structure analysis of all three genes confirmed their substantial similarity with C3 of other teleost fish and higher vertebrates. Moreover, *Rb-C3 Rb-C4* and *Rb-C5* were found to be potentially induced by infection with *E. tarda*, *S. iniae,* LPS and iridovirus, indicating their potential role in the rock bream immune defense system. *In vitro* assay with recombinantly expressed their anaphylatoxin domains may prove their significance in generating oxidative stress to immune cells. This study is the first study, which demonstrate respiratory burst activity *in vitro* condition by the recombinantly expressed anaphylatoxin domain. We anticipated that the findings of this study will not only increase our understanding of the rock bream complement genes, but will also promote the broader understanding of the genomic diversity, biological significance of key regulatory molecules of the teleost complement system.

Chapter III

Molecular characterization, expression analysis of complement C6, C7, C8 (α, β, γ) and C9 as lytic components of the complement system

Abstract

Termination of activated-complement system began with the activation of lytic pathway genes results in formation of membrane attack complexes (MACs) that disrupt lipid bilayer of targeted bacteria. The complement component, 6, 7, 8 (α, β and γ) and 9 are single-chain glycoproteins belong to the lytic pathway which are sequentially arranged on the bacterial cell wall. This process is initiated upon cleavage of C5 in to C5a and C5b by the C5 convertases and preceded as the other components bind to C5b. This chapter describes genomic structural and molecular characterization of the all lytic pathway components that have been identified from rock bream. The full-length cDNA sequences were identified from rock bream transcriptome database, and their genomic sequences were obtained by screening and sequencing of rock bream BAC library. The largest and smallest ORFs were found in *Rb-C6* (2814bp) and *Rb-C8γ* (663bp), respectively. According to the domain architecture of the genes in lytic pathway, the common domains; thrombospondin type-1 (TSP-1) domain, a low-density lipoprotein receptor domain class A (LDLR-A), membrane attack complex/perforin (MACPF) domain and epidermal growth factor-like-1 (EGF-1) domain were identified except in Rb-C8γ which bared lipocalin domain. In addition, Rb-C6 and Rb-C7 contained two complement control protein (Sushi/CCP/SCR) domains and two factor I MAC module (FIMAC) domains. The promoter regions of all genes contained important putative transcription factor binding sites including those for NF-κB, SP-1, C/EBP, AP-1, and OCT-1. According to the phylogenic and pairwise comparison, Rb-C6, Rb-C7, Rb-C8α, Rb-C8β and Rb-C9 have higher identity to fish from Order; Tetraodoniformes (puffer fish), Pleuronrctiformes (flounder), Salmoniformes (rainbow trout) and Cichlidae (Tilapia). Quantitative real-time PCR analysis confirmed that all the lytic pathway genes except *Rb-C6* were constitutively expressed in all examined tissues, isolated from healthy rock bream, and with highest expression in liver. The *Rb-C6* was found to be highly expressed in heart cells

111

that was an exception to all other complements. However, considerably high expression of *Rb-C6* was also observed in liver. In immune-challenged rock bream, all the lytic genes were responded to *E. tarda, S. iniae,* lipopolysaccharide endotoxin and rock bream iridovirus in head kidney and liver tissues. Taken together, the molecular characteristics, orthologous relationships and transcriptional response to pathogenic stimulants suggest that these identified genes are belonging to the lytic pathway of the complement system that are involving in immune processes of rock bream.

1. Introduction to lytic pathway genes

Complement activation via the classical, lectin or alternative pathways leads to formation of the membrane attack complex (MAC) on the surface of complement-opsonized cells (Nakao et al., 2011), ultimately causing lysis of the targeted foreign cell. MAC itself is a glycoprotein aggregate, composed of the terminal complement component (TCC) plasma proteins C5b, C6, C7, C8, and C9 that regulate the lytic pathway of the complement system. The MAC is assembled by sequential binding of C5b with C6, C7, and C8 and incorporation of an alternative number (12-18) of the C9 molecules into a transmembrane pore (Muller-Eberhard, 1986). A study on carp serum complement was revealed that different TCC molecules were interacted with molar ration of 1:1:1:1:4 in the MAC formation respectively (Nakao et al., 1996).

These TCC proteins belong to the same gene family as the perforins, which are the lytic proteins of natural killer cells and cytotoxic lymphocytes. The genes of the C6–C9 TCCs and perforins share the unique membrane attack complex/perforin-like (MACPF) domain. The MACPF is one of the common domains identified in many different functional proteins. They exhibit limited sequential similarity but contain a signature MACPF motif (Y/W-G-T/S-H-F/Y-X_6-G-G) (Ponting, 1999). Furthermore, terminal-complement

112

components (TCC) C6, C7, C8α, C8β, and C9 exhibit a high degree of sequence similarity, and all possessed structural motifs for, thrombospondin (TS), low density lipoprotein receptor (LDL-R), and epidermal growth factor precursor (EGFP) domains (DiScipio et al., 1988; Hobart et al., 1995). Interestingly, C6 and C7 were contained two complement control protein (Sushi/CCP/SCR) domains, two factor I MAC module (FIMAC) domains and extra TSP-1 also was presented in the C6. Therefore, the highest domain complexity was observed in the C6 gene in the lytic pathway. On the other hand, the C8 has most complex subunit arrangement comprising of three genetically distinct proteins (C8α, C8β, and C8γ). They are arranged as a disulfide-linked C8α-γ heterodimer that is noncovalently associated with C8β (Lebioda L, 2005). The C8γ has dis-similar structure to that of other lytic pathway genes where it harbors only lipocalin domain which is able to bind small hydrophobic molecules and large soluble macromolecules (Papanastasiou and Zarkadis, 2006). The C9 is the one which mostly contributes to cover the pore wall of the MAC on the cell wall at least 12 molecules may contribute to form a single pore. Therefore, C9 plays a key role in formation of membrane attack complex.

Activation of the lytic pathway occurs when C5b associates with the target cell membrane and undergoes conformational changes that expose its C6 binding site. The C5b-6 complex is then capable of binding to C7. Upon C7 binding to the C5b-6 dimer, the membrane binding sites are made available, which helps to stably incorporate the complex into the target membrane (Wurzner, 2000). Formation of the C5b-7 complex (known as the intermediate complex) can facilitate further interactions with other proximal cells, initiating the process of reactive lysis (Thompson and Lachmann, 1970) (Wurzner, 2000). Availability of C7 acts as the rate-limiting step of the membrane attack process, since unbound C5b-6 dissociates from the membrane without activating the cytolytic signal (Gonzalez et al., 2003). When the C5b-7 complex binds to C8 and C9 to form the complete C5b-9 complex,

phosphorylation of the Bcl-2-associated death promoter (BAD) protein is induced which leads to blockade of the FLICE-inhibitory protein (FLIP) and allows for subsequent activation of caspase-8 (Corallini et al., 2009).

To date, several lytic pathway genes have been characterized at the molecular level, in mammalian species and animals that serve as mammalian's food sources, such as fish. In the teleost fishes, C7, C8β and C9 have been defined in trout (Kazantzi et al., 2003; Papanastasiou and Zarkadis, 2005; Tomlinson et al., 1993; Zarkadis et al., 2005), C6 (Shen et al., 2011) and C7 (Shen et al., 2012a) in grass carp, C8β and C9 in Japanese flounder (Katagiri et al., 1999), C8β in carp (Nakao et al., 1996), and C9 in puffer fish (Yeo et al., 1997). In addition, C7 and C9, along with their promoter regions, have been predicted in grass carp via bioinformatics analysis (Li et al., 2007; Shen et al., 2012a). However, no comprehensive immune-challenge studies had been conducted on any of these complement factors for any of these fishes.

This chapter describes mainly on the lytic pathway genes of the rock bream and their unique features in relation to the formation of the mambrane attack complex. Indeed, the genomic structure of C6 (*Rb-C6*),C7 *(Rb-C7),* C8α (*Rb-C8α),* C8β (*Rb-C8β),* C8γ (*Rb-C8γ)* and C9 (*Rb-C9*) were identified using bacterial artificial chromosome (BAC) library sequencing and their evolutionary relationships were determined by phylogenetic tree and pairwise alignments. Finally, tissue-specific expression analysis was carried out to determine the normal expression profile and immune responsiveness of each gene upon bacteria (Gram-positive and -negative), lipopolysaccharide (LPS) endotoxin, and viral challenges.

Table 14. Description of primers used for the amplification of genes included in this chapter

Name	Target	Primer Sequence(5'-3')
Rb-C6-1F	RT-PCR amplification and BAC screening	TTGTGGTCGATTGCAAGGTGCTTC
Rb-C6-1R	RT-PCR amplification and BAC screening	TGCAGGTATCAAACTCCTCCTGGT
Rb-C7-1F	RT-PCR amplification and BAC screening	TCAGCTTGGAGCTCTGCGATGTAT
Rb-C7-1R	RT-PCR amplification and BAC screening	AGATTGTCCCTGGTTTGCAGTCCT
Rb-C7-2F	*In situ* hybridization- RNA probe synthesis	TGAAGAAGCTCCACCTGAAAAGAGC
Rb-C7-2R	*In situ* hybridization- RNA probe synthesis	AGGGCATTCTGACGCATTCTGA
Rb-C8α -1F	qPCR amplification and BAC screening	TGTCCAACAGGAACCACAGAGTGT
Rb-C8α-1R	qPCR amplification and BAC screening	ATTACAGCGAAGGGACTGGCTGAT
Rb-C8β -1F	qPCR amplification and BAC screening	TTCCATTGGCTTTGGCATTCCAGG
Rb-C8β-1R	qPCR amplification and BAC screening	TCTTGCTAGAAGCACGGCGAATCT
Rb-C8γ -1F	qPCR amplification and BAC screening	AGCTCGTCCTGAGCAGAACATTGA
Rb-C8γ-1R	qPCR amplification and BAC screening	ACCTGCTCAAGAGTTTCAGGGTGA
Rb-C8γ -2F	ORF verification	TTGTTCTGTGTGTTTGGGGCTAAGTACA
Rb-C8γ-2R	ORF verification	ATCAAAATAAGAGCCTTCCAGCCTTGTCA
Rb-C9-1F	qPCR amplification and BAC screening	TGCTTGCTCTGACACAGAGTTCCA
Rb-C9-1R	qPCR amplification and BAC screening	TTACGCACAGGTTCACAGTCCTCA
Rb-β-actin-2F	qPCR reference gene	TCATCACCATCGGCAATGAGAGGT
Rb-β-actin-2R	qPCR reference gene	TGATGCTGTTGTAGGTGGTCTCGT

2. Results and discussion of rock bream complement C6

2.1 Results

2.1.1 *In silico* characterization of Rb-C6

Based on identity and similarity with other orthologous in BLAST search analysis of EST

data from the normalized cDNA library, we identified a putative complement component 6

from rock bream, and designated it as *Rb-C6*. The sequence was deposited in NCBI under

accession number KF682442. The complete cDNA sequence was 4553 bp, including 234 bp 5'-untranslated region (5′-UTR), 2817 bp of open reading frame (ORF) encoding 938 amino acids, and 1502 bp of 3'-UTR with a polyadenylation signal (Figure 41). The predicted molecular mass of *Rb-C6* was 106 kDa and the theoretical isoelectric point was 6.3. According to SignalP prediction, the *Rb-C6* was possessed a signal peptide ranging from the 1st amino acid to the 22nd residue, which was presumed to facilitate the extracellular localization (Figure 41). The mature protein contains three thrombospondin type-1 (TSP-1) repeat domains, a low-density lipoprotein receptor domain class A (LDLR-A) domain, an epidermal growth factor-like-1 (EGF-1) domain, two complement control protein (Sushi/CCP/SCR) domains, two factor I MAC module (FIMAC) domains, and a membrane attack complex/perforin (MACPF) domain.
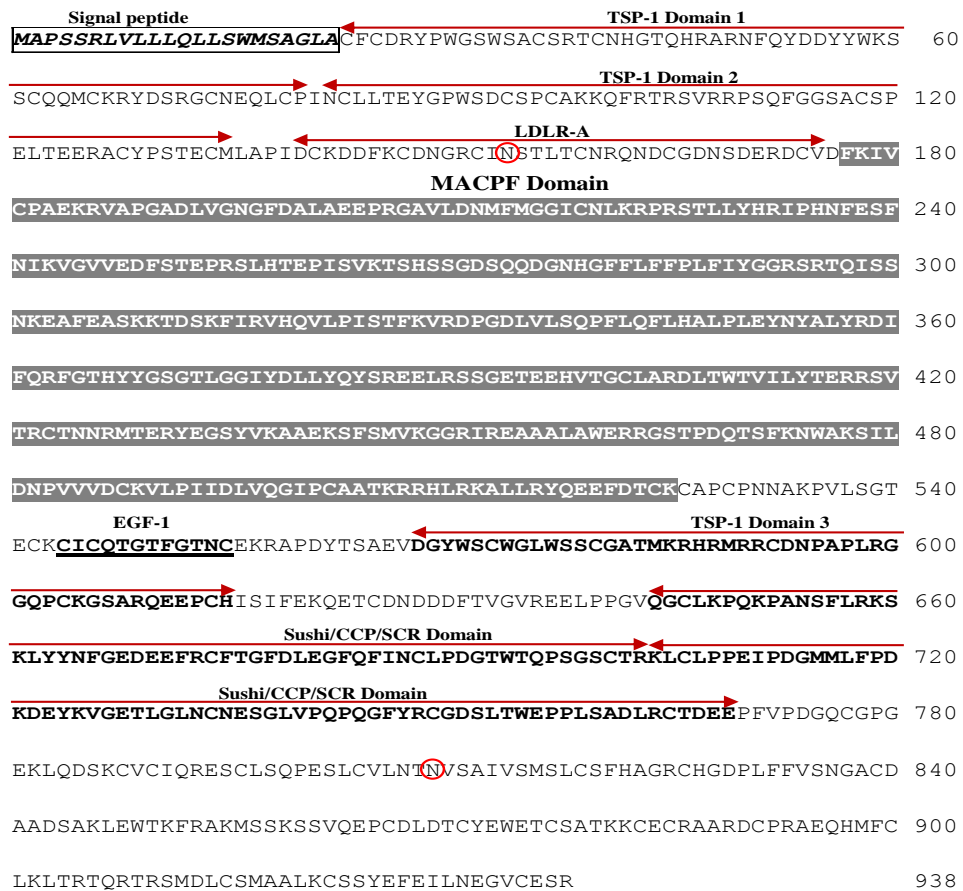
Figure 41. Amino acid sequence of Rb-C6

Predicted signal peptide is boxed with bold italic letters. Conserved domains are indicated as double headed arrows. Characteristic MACPF domain is highlighted in grey color

2.1.2 Characterization of the Rb-C6 genomic structure and promoter region

PCR-based screening of the BAC genomic library using *Rb-C6* gene-specific primers led to the identification of only one positive BAC clone that contained the sequence of interest (Table 14). The putative genomic sequence of *Rb-C6* was recognized by comparison with the full-length cDNA sequence and then compared by genome alignment with known C6 genomic DNA structures of several other vertebrate species published in the GenBank. The *Rb-C6* genomic sequence was about 16.2 kb and was composed of 18 exons interrupted by 17 introns. The 5'- and 3'-ends of all introns harbor canonical splicing motifs (according to the GT-AG rule). The number of exons and introns were found to be well conserved among all the C6 orthologous. Furthermore, *Rb-C6* was consisted of un-translated region as a separate exon in the 5'-UTR (Figure 42).
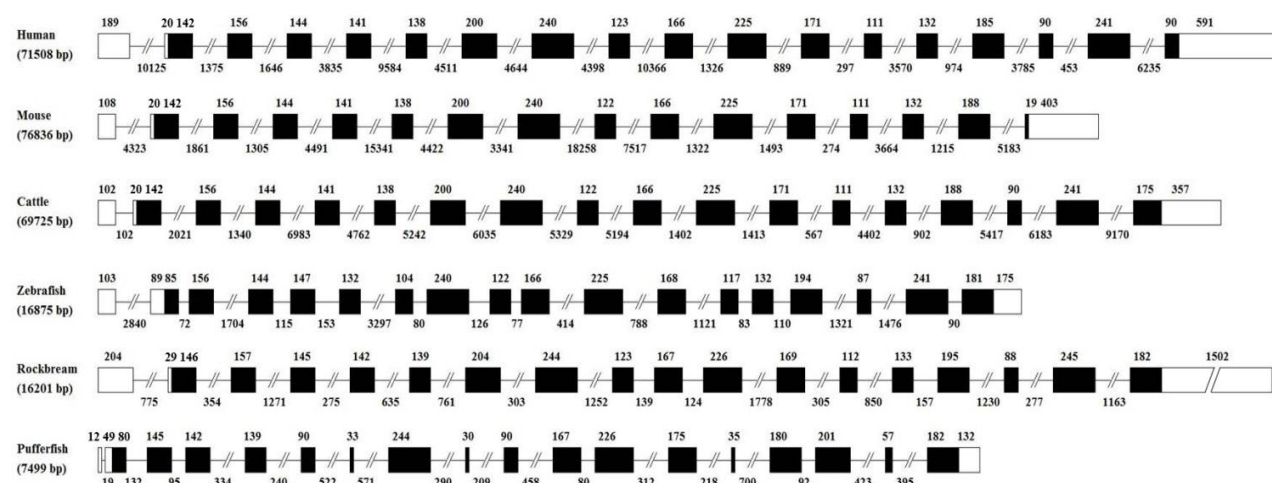
Figure 42. Genomic structure comparison of Rb-C6
Exons are shown in black solid boxes. UTRs are shown in light-colored boxes. Introns are shown as thin lines between the exons, and introns larger than 200 bp are indicated by interrupted lines. Numbers above the boxes and below the lines indicate the lengths of exons and introns, respectively. Accession numbers are Puffer fish: XP_003965156.1, Zebrafish: BC164035.1, Cattle: NP_001039444.1, Mouse: NM_016704.2, and Human: BC035723.1.

117

According to the BDGP prediction analysis, the promoter of *Rb-C6* has complex structure with a co-promoter region and contains multiple binding sites for proteins with functions related to the transcription initiation process. A transcription initiation site (+1 site) was found at 234 bp upstream of the start codon (ATG). Binding sites for immune-related transcription factors were also identified, and these included the CCAAT-enhancer binding protein (C/EBP), nuclear factor-kappa B (NF-κB), activator protein-1 (AP-1),octamer transcription factor (OCT-1), specificity protein-1 (SP-1), hepatocyte nuclear factor-1, (HNF-1) hepatocyte nuclear factor-3 (HNF-3), and nuclear factor-1 (NF-1) (Figure 43).

```
TCTCTGCACAGGCACAATAATGTAACCAAACCTGGCAATCTCAATGTAACGGCAAGACCTCAGGTAAGCT  -1040
                                                            GATA-1
CTTAACACAACGCCCAAGAAGTAGTTCCAGCACAAAACTCCCTCTGAAACTACAAATTGTTGTTTTATGA   -970
                              NF-κB
ATGTGACTTATGAATTAAACAAATGAGATACAGAATGTTAATTAGTGAGCTTTAGAGGGGTTGGCATATG   -900
             HNF-3B                                      SP-1
TATAATTTTGAACTTTGGACAGAGCCAGGCTACTGTTTCCCCTGCCTCCAGTCCTTATGCTAAGCTAGGC   -830
                                   NF-κB
TAACCAAGTCCTGACTTCGGCTCCGCACTTAACACACAGAAAGAAAGCGAATAAGAGTATTACCCAAAAT   -760
                                                          HNF-1
GTTGAACTATTCCTTAACGTAATCGTACAAGAGCAGTAAAAAAAGACATTTTTTGCTTGGGAATTAAATT   -690

CATATCTGAAATTTTTGAAAATCTTGGCGCTAAATTTTGTTGTATTTGCATTTCAGTGATTTCAACGTTT   -620
                         YY-1               OCT-6
GCTTCCAGTTGTGCATCCTCTCAGCATTCACATCTCTTTTTCCATGAAGAAGATTTTTTGAACTTGGCAC   -550
                                                             NF-1
TGTGTCTGTGTATCTTTTGCAACACTTTACATACTCTTTTCATCTTCTTACCCCCAAAGTCTTGGAAAAA   -480

TTGAACAATGACATACCTAACCAGATAACAGTATTTTATTCAACAATGTATGTCAAAGTGTTCCAGTCAT   -410
                  GATA-1
TCACTGGGCAACAACCTTTACATTTCTCTGTATACATATCACTTCTCCCAGTTGCAATTTGTGTTGCAGT   -340
                            OCT-1                    C/EBPα
TTAACAATTTAGGCAGAATTCACTTATGTGCCAATTACAACATGTACCAATTGCGAGTTTGTTTAGTGCC   -270
                           NF-1
AGATTGTGCAAATACTTGGATAACTAAAGAAAGTGTGGAAGGAGAATAAGTTTTCTCTCCCTCAGTATCT   -200
    C/EBPα                              SP-1
CCAACTGTCAAGCCTCAAACACAAAACATCTGGTTTACGTCTCTGCAGCCGGGAAACCCCAGAGCTCTTT   -130
                  HNF-3                    SP-1
ACTTCCCGTAACAGTGAGGGACAAGAGGGTTATGCAAGAACTTCATGTGCTCCTATGTTTTATGATCATT    -60
               SP-1            OCT-1               C/EBPα
GTTCTTATTAAGTAGTTTTTTATTTTGACATGTTTTTTTGAACACAGATTGGGTAATACAAGTCAAGAAAC     11
           C/EBPα                                        +1 site
AACTGGATTCCAATGAAAAATAACAGAGGGAGGTTTTAAATTGCTGCTGGCAACTCAGTCAGATTCATCT     81

CTGCCTGCAGTGCAGTCGACTTAGTAAGAAGACTCTTACAGTAACTGCCCAAAAATACAGTCAAACACAG    151

GAGCTACTAAATGACATGCTTTAATATCTGCTGAGGAAGACGTCAGTATATGTGAGGAGGTGAAACCACT    221

TTGAACTTCCCTGATGGCTCCTTCCAGCCGGTTGGTCCTGCTGCTACAGCTCCTCAGCTGGATGTCAGCT    291
               M  A  P  S  S  R  L  V  L  L  L  Q  L  L  S  W  M  S  A
```

Figure 43. 5' Flanking region of Rb-C6

Putative transcription factor binding sites are shown in bold and underlined, with corresponding factor indicated directly below. The putative transcription initiation site is indicated by a bent arrow. Translated amino acids (uppercase) and start codons (italicized uppercase, in bold) are shown.

2.1.3 Identity, similarity, and phylogenic relationship of Rb-C6

Pairwise sequence analysis was revealed that *Rb-C6* has 75.6% amino acid identity with the orthologous from Nile tilapia (*Oreochromis niloticus*: XP_003440120.1) and 73.7% with the orthologous from rainbow trout (*Takifugu rubripes:* XP_003965156.1). It also showed >50% identity with the orthologous from other fish species, including grass carp (*Ctenopharyngodon idella*: ADV18976.1), rainbow trout (*Oncorhynchus mykiss*: NP_001118093.1), and zebrafish (*Danio rerio*: NP_956932.1). Furthermore, it showed > 40% identity/similarity to the other higher vertebrates such as chicken, cattle and human. Multiple sequence alignment of *Rb-C6* indicated that each motif or domain has higher similarity with the corresponding domain of other considered species. All the cysteine residues were found to be conserved across nearly all the orthologous, and certain (presumably important) signature motifs were highly conserved among all the orthologous. However, in the FIMAC domain was found to be lowest identity/similarity with the same of other orthologous (Figure 44).
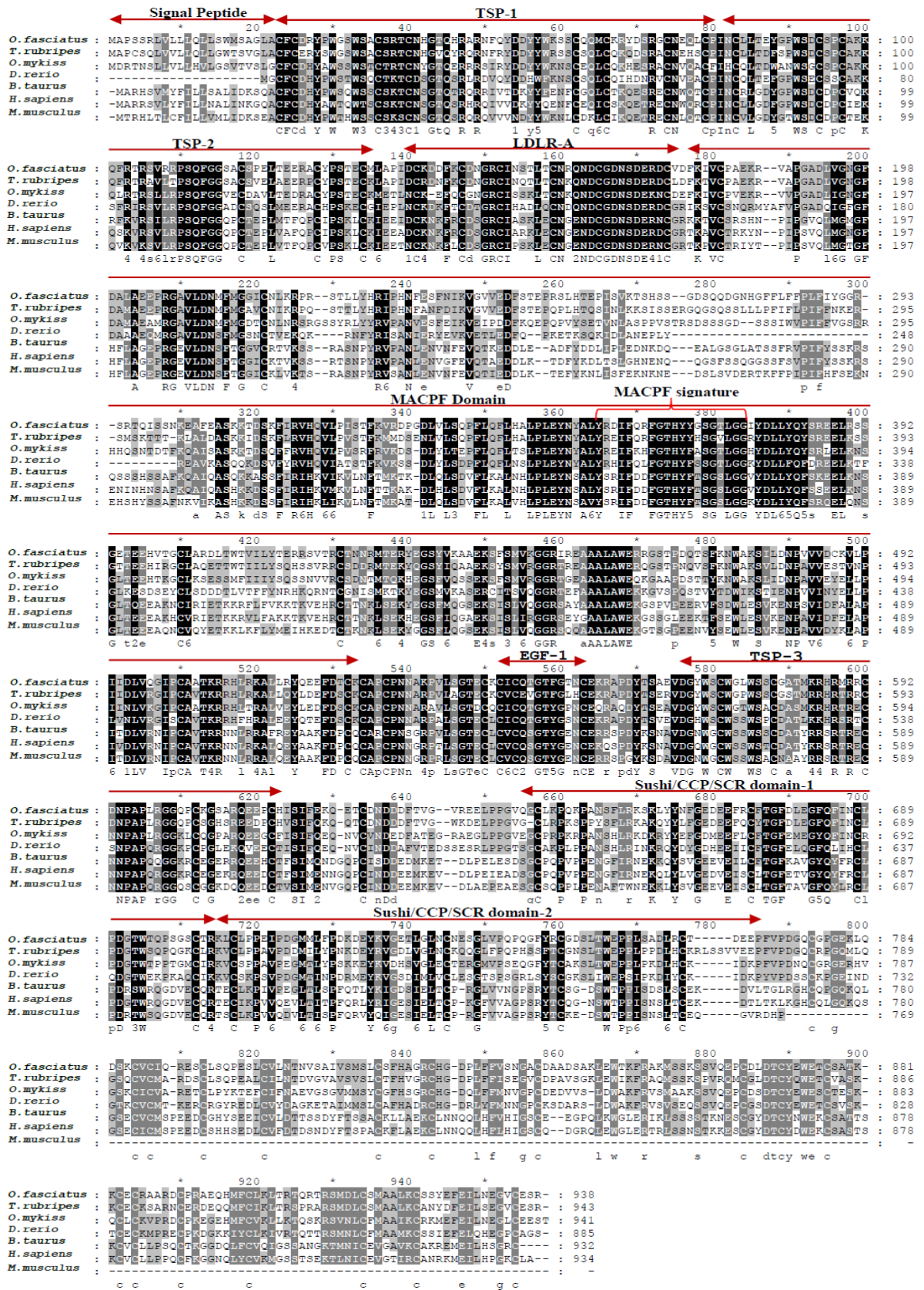
Figure 44. Multiple sequence alignment of Rb-C6 and C6 orthologous from other vertebrates

Identical residues among all orthologous' sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double-arrows.

The evolutionary relationship was studied using the deduced amino acid sequences of Rb-C6 and available sequences for the complement components published in the NCBI database. The results were depicted as a phylogenetic tree as in Figure 45. The phylogenetic tree represents each of the lytic pathway genes (C6, C7, C8α, C8β, C8γ and C9) as separate clusters. It clustered within the highest identify orthologous from fish.
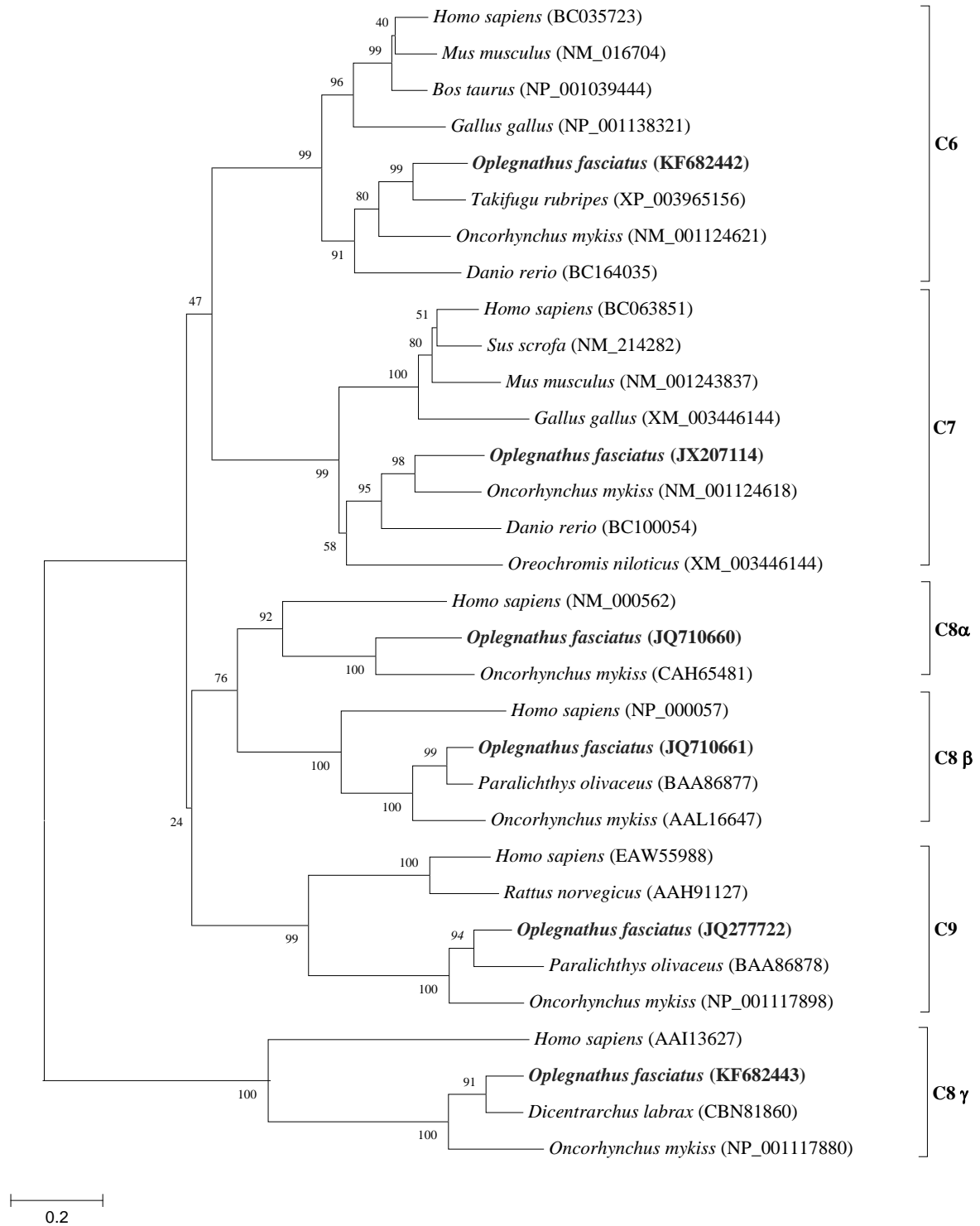
Figure 45. Phylogenic tree of lytic pathway of genes of rock bream
The tree is based on alignment of the full-length amino acid sequences. The numbers at the branches denoted the bootstrap values from 5000 replicates. The GenBank accession numbers are shown within brackets next to each species.

123

2.1.4 Tissue-specific mRNA expression of Rb-C6 in normal healthy conditions

To examine the normal tissue distribution pattern of *Rb-C6* transcription, qPCR was carried out for blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine (Figure 46). *Rb-C6* was found to be constitutively expressed in all examined tissues, although the level of expression varied between each. The most abundant expression was detected in heart and liver, compare to the normalized expression level in blood (lowest expression), and marginal expression was detected in muscle, skin and gills.
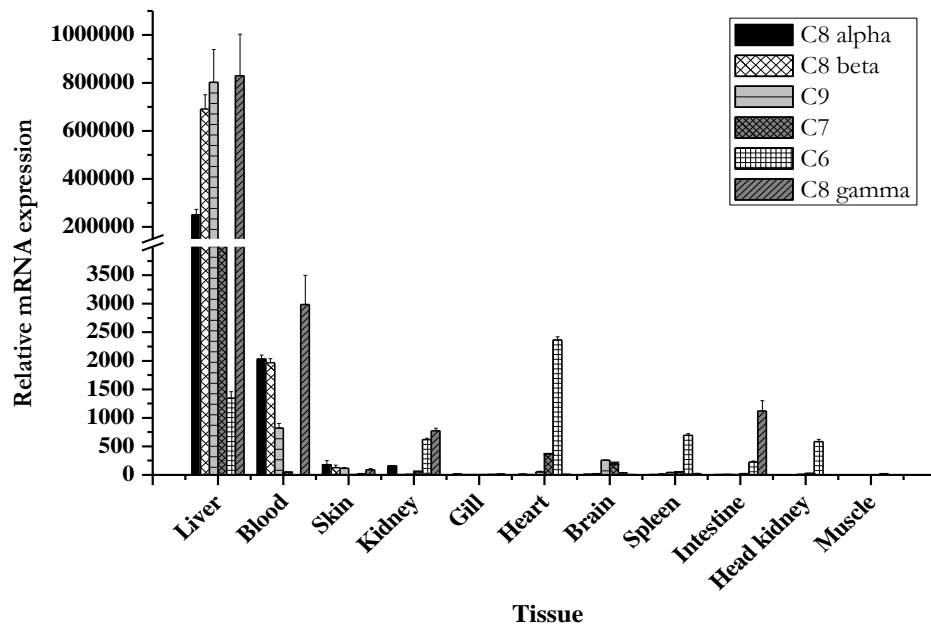
Figure 46. Analysis of the tissue-specific expression of all lytic pathway genes of rock bream
The relative mRNA expression in peripheral blood cells, gills, liver, heart, spleen, intestine, head kidney, kidney, skin, muscle, and brain were calculated using β-actin as the reference gene. The bars represent the standard deviation (n = 3). Significant (fold) differences were calculated with respect to the relative expression in muscle.

2.1.5 Modulation of Rb-C6 expression in response to immune stimulation

To explore the functions of *Rb-C6* under pathogenic conditions, immune challenge was performed and the respective mRNA expression was analyzed by qPCR in spleen tissue from rock breams. The expression pattern of *Rb-C6* in spleen of rock breams challenged with the Gram-negative bacteria *E. tarda* is shown in Figure 47. The up-regulated transcription levels were observed in 6h - 48h while showing highest at 48 h (6.6-fold increase relative to the un-injected control at 0 h, $P<0.05$). The highest induction was observed in the spleen after challenge with *E. tarda* and iridovirus. Even though *S. iniae*, LPS and poly I:C challenges were led to up-regulate the level of transcription of *Rb-C6,* the fold inductions in each challenge were lower than that of *E .tarda* and iridovirus.
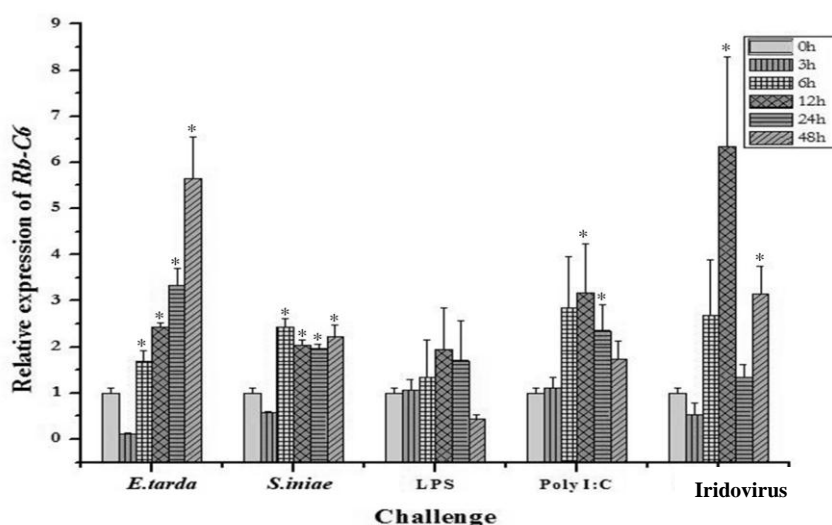
Figure 47. Relative mRNA expression of *Rb-C6* in spleen after challenge of rock bream with *E. tarda*, *S. iniae*, LPS, Poly I:C and iridovirus, as detected by qPCR
The amount of mRNA relative to the internal control, β-actin, is expressed as mean ± SD (n = 3). Significant differences ($P < 0.05$ vs. untreated control at 0 h) are marked with asterisks.

2.2 Discussion

Evidence has been suggested that the lytic pathway genes trace back to the deuterostome lineage and that they evolved as functional mediators of direct pathogen killing through their formation of the MAC on target organisms (Zarkadis et al., 2005). In human genome, only the genes encoding for the MAC components (C6, C7, C8α, C8β, and C9) and perforin contains the MACPF domain, suggesting its importance in direct membrane-related killing (Zarkadis et al., 2005). Not surprisingly, studies of teleost species have been identified all the complement components of the lytic pathway that are known in humans. Thus, it can be suggested that the teleost fish consists of a fully functional complement system. The Rb-C6 possesses all the motifs found in other complement lytic genes. In additions to others, it contains two CCP motifs and two FIMAC domains, which are important for the interaction with C5. In the past, the complement C6 has been studied in rainbow trout (Chondrou et al., 2006) and grass carp (Shen et al., 2013). However, transcriptional responses of C6 upon broader panel of various pathogen-related stimuli are scanty. In the present study, a MAC component (*Rb-C6)* gene encoding an MACPF super family protein of the complement lytic pathway was isolated, sequenced, and characterized at the genomic structural level from the teleost rock bream. Furthermore, the gene expression profile of this newly discovered C6 was defined for the healthy organism and in response to *E. tarda*, *S. iniae,* LPS, Poly I:C and RBIV exposures.

Indeed, in the rock bream genome, C6 lies 1.6 kb downstream of the newly identified *Rb-C7*, which suggests that the panel of rock bream MAC components may occur in the same or adjacent chromosomal levels. However, according to our comprehensive screening of the rock bream BAC library, only one C6 isoform exists in the rock bream genome. Assembly of the MAC on the surface of bacteria and other pathogenic organisms involves the sequential interaction of the C5b, C6, C7, C8 and C9 complement proteins (Muller-Eberhard, 1986).

The *Rb-C6* is composed of various characteristic domains including TSP-1, LDLR-A, MACPF, EGF-1, CCP and FIMAC as other teleostean TCC genes. (Chondrou et al., 2006; Papanastasiou and Zarkadis, 2005). The structural arrangement of the domains is also resembled to other TCC genes. To date, several hundreds of MACPF-containing orthologous bearing MACPF signature ((Y/W)G(T/S)H(F/Y)X$_6$GG), have been identified. (Ponting, 1999). Similarly, the MACPF domain of *Rb-C6* contains this signature domain, which further supports it's potential role as a complement lytic pathway gene. However, few proteins of the MACPF super family have been functionally characterized (Lovelace et al., 2011). The presence of a second FIMAC domain, as found in *Rb-C6*, may increase the affinity and specificity of the C6 protein's interactions with the C5b. Presence of highly conserved cysteine residues in FIMAC domains increases the structural integrity of secondary and tertiary structures of the C6 protein, which facilitates the firm binding of the C5b-6 complex.

Many of the C6 orthologous, including *Rb-C6*, were showed a similar exon-intron structural pattern containing 18 exons interrupted by 17 introns. However, the size of the introns was varies among these orthologous, which explained the differences in total length (ranging from 7.5 kb of the puffer fish to 76.8 kb for the mouse protein). Interestingly, all the C6 genes in the comparison including Rb-C6 were separated exonic region in 5' UTR, which does not, included in the cording region, which was common in selected species in the comparison.

Although teleost research has recently focused on identification of complement genes, to the best of our knowledge detailed deletion study has not been carried out to understand the TCC gene promoter region and only one 5' sequence region (in grass carp C6 and C7) has been analyzed using a computer-based approach (Shen et al., 2012a, Shen et al., 2013). Deletion studies of the human C7 promoter were identified a single transcription start site at 115 bp upstream from the initiation codon (Gonzalez et al., 2003), while the *in silico* study of

127

the grass carp promoter predicted it at 39 bp upstream of the start codon. In the *Rb-C6* promoter region, a putative transcriptional initiation site was identified at 234 bp upstream of the start codon. However further studies are necessary to confirm the actual site of transcription initiation.

According to our analysis, several important putative transcriptional binding sites were located in the 5′-flanking region of *Rb-C6*. In comparison, the human C7 (Gonzalez et al., 2003), grass carp C7 (Shen et al., 2012a), grass carp C6 (Shen et al., 2013) and *Rb-C6* contain C/EBP and HNF-3 transcription factor binding sites. In addition, the *Rb-C6* promoter region contains putative transcription factor binding sites for AP-1, IRF-1, OCT-1, SP-1, and NF-1, which are present in other TCC genes' promoters (Li et al., 2007; Wickramaarachchi et al., 2012). This promoter profile may imply that the mechanisms regulating the *Rb-C6* gene transcription may be similar to those of other fish TCC genes and mammalian C6. The OCT-1 is involved in regulation of some T cell-specific cytokines' expression, such as interleukin (IL)-2, IL-4, and IL-5. Thus, the *Rb-C6* promoter contains several important putative transcription factor binding sites that are similar to those previously identified in the complement lytic pathway genes. Moreover, most of these sites are associated with factors related to immunity, which supports the hypothesis that *Rb-C6* may have an immune function, similar to that described for the C6s in human and grass carp. However, further studies are needed to confirm these binding sites as functional elements of the *Rb-C6* promoter.

In mammals, the liver is the primary production site for the majority of the complement components (Alper et al., 1980; Witzel-Schlomp et al., 2001). The expression pattern of complement C6 is slightly deviated from that, hence it has highest expression in spleen followed by liver and heart in grass carp. However, the highest Rb-C6 was detected in heart followed by liver and spleen. Taken together, these results may indicate that *Rb-C6* behaves in a similar manner as C6 genes described previously.

According to the immune-responsive expression analyses of *Rb-C6*, Gram-negative bacteria can effectively trigger the mRNA expression in head kidney between 6 h to 48 h p.i. The induced transcription level was highest at 24 h p.i., which also represented the highest fold-induction among all the immune stimulants examined in this study. A previous study conducted in grass carp also found that C6 expression was induced in head kidney by the Gram-negative bacteria *Aeromonas hydrophila* between 4 h and 168 h (Shen et al., 2012a). To the best of our knowledge, only two studies of C6 orthologous have included analyses of pathogen-induced expression conditions, and both of those used Gram-negative bacteria (Gonzalez et al., 2007b; Shen et al., 2012a). Other TCC genes, however, have been studied in response to Gram-positive bacteria, LPS, and dsDNA virus, and all showed remarkable inducement of expression upon exposure to these stimulants (Wickramaarachchi et al., 2012). In addition, recent microarray studies have provided evidences of up-regulation of immune genes in response to pathogenic challenges. Zebrafish C9 transcripts were shown to be increased in both fins and organs (liver, spleen and head kidney) at 48 h post-challenge with viral hemorrhagic septicemia virus (VHSV) (Encinas et al., 2010). Similarly, *Rb-C6* was also up-regulated significantly in both head kidney and liver at 48 h p.i. of RBIV. Therefore, we hypothesized that *Rb-C6* may also be involved in the virus-mediated immune response of rock bream.

TCCs may be activated in response to general microbial infections (Li et al., 2007), but the regulatory mechanism underlying such a general activation of TCC genes' transcription has yet to be fully elucidated. In the current study, the up-regulated transcriptional profile of *Rb-C6* upon exposure to Gram-negative bacteria and RBIV can be substantiated through the presence of immune signaling transcription factor binding sites in the promoter region, such as NF-κB, AP-1, and IRF-1 (virus-specific) which are themselves regulated by toll-like receptor-dependent immune signaling pathways (Hawlisch and Kohl,

2006; Martin and Wesche, 2002). Furthermore, the observed robust expression of *Rb-C6* transcript levels in the spleen of healthy fish and its significant enhancement accompanying immune challenge likely reflect the obligatory requirement of C6 molecules for MAC formation.

In summary, we have identified the complete genome sequence of complement component 6 from rock bream. Genome structure analysis of *Rb-C6* was confirmed its significant similarity with C6s of other teleosts and classified it as a TCC protein family member. Moreover, *Rb-C6* up-regulation was accompanied via experimental challenge with *E. tarda*, *S. iniae,* LPS, poly I:C and RBIV, indicating a potential role for this gene and its encoded protein in the immune defense system of rock bream.

3. Results and discussion of rock bream complement C7

3.1 Results

3.1.1 *In silico* characterization of Rb-C7

Based on identity and similarity with other orthologous in BLAST search analysis of EST data from the normalized cDNA library, we identified a putative complement component 7 gene from rock bream, and designated it as *Rb-C7*. The sequence was deposited in NCBI under accession number JX207114. The complete cDNA sequence is 2823 bp, including a 32 bp 5'-untranslated region (5′-UTR), a 2490 bp open reading frame (ORF) encoding 830 amino acids, and a 301 bp 3'-UTR with a polyadenylation signal (Figure 48 and 50). The predicted molecular mass of *Rb-C7* was 91 kDa and the theoretical isoelectric point was 6.8. According to SignalP prediction, the *Rb-C7* was possessed a signal peptide ranging from the 1st amino acid to the 24th residue, which was presumed to facilitate the extracellular localization (Figure 48). The mature protein contains two thrombospondin type-1 (TSP-1) repeat domains, a low-density lipoprotein receptor domain class A (LDLR-A) domain, an

130

epidermal growth factor-like-1 (EGF-1) domain, two complement control protein (Sushi/CCP/SCR) domains, two factor I MAC module (FIMAC) domains, and a membrane attack complex/perforin (MACPF) domain. The MACPF domain and signature were identified using the ScanProsite program in the PROSITE database (de Castro et al., 2006).

```
ATG AAGTTGGATTTG GCTGCAGCTCCGTCT TCCTTTGCACTGTTG CTCATCTTTCTGTCT    92
  M   K   L   D   L   A   A   A   P   S   F   A   L   L   L   I   F   L   S   20
CCAGCTTGCTGTCAG CAGCCTGTGAACTGC AGATGTGGGCCTTTT GGAGCGTGGTCTGAG   152
  P   A   C   C   Q   Q   P   V   N   C   R   W   G   P   F   G   A   W   S   E   40
TGTGATGGCTGCTCC AGAACAAAGGTCAG ACTCGCCATGTGGAG GTGTACGCCCAGTTT   212
  C   D   G   C   S   R   T   K   V   Q   T   R   H   V   E   V   Y   A   Q   F   60
GGGGGCACGCCGTGT TCAGGAGAAGCCACA CAAACACAACCATGT GTCCCACAGAAACGA   272
  G   G   T   P   C   S   G   E   A   T   Q   T   Q   P   C   V   P   Q   K   R   80
TGTCCCCTGGAAACA GGCTGTGGAGACAGG TTCCGCTGCACCTCT GGTCAGTGTATCAGC   332
  C   P   L   E   T   G   C   G   D   R   F   R   C   T   S   G   Q   C   I   S   100
CAGTCTCTGGTGTGT AACGGAGACCAGGAC TGTGAGGATGGTCTG GATGAGAGAAGCTGT   392
  Q   S   L   V   C   N   G   D   Q   D   C   E   D   G   L   D   E   R   S   C   120
GGTCGAGACAGCAGC CACTACCCATGTGAC CTCGACAAAAACGCCT CCCAACTCTGACGTC   452
  G   R   D   S   S   H   Y   P   C   D   L   D   K   T   P   P   N   S   D   V   140
ACAGGCAGAGGGTAC AACGTGCTGACAGGC AAACTGAGAGCAGGT GTGATCAACACTCTG   512
  T   G   R   G   Y   N   V   L   T   G   K   L   R   A   G   V   I   N   T   L   160
AGCTTTGGAGGGCAG TGCAGGAAGGTGTTC AGTGTGGACCATAAA GTCTATTACAGACTG   572
  S   F   G   G   Q   C   R   K   V   F   S   G   D   H   K   V   Y   Y   R   L   180
CCACAGAACATCCTC AGGTACAACTTTGAG GTGAAAGTAGACAAT GACGAGAGCGATGAA   632
  P   Q   N   I   L   R   Y   N   F   E   V   K   V   D   N   D   E   S   D   E   200
TCCTACGAGAGCTCC TGGTCCTACACGCAG CACATCCAGGCTAAT GCCTTGGTGGGACAC   692
  S   Y   E   S   S   W   S   Y   T   Q   H   I   Q   A   N   A   L   V   G   H   220
GACCGTCGCACTTTC CACAAAGAGCTCACT GACAACAAGGCCCAC AAACTTCTAATCCTG   752
  D   R   R   T   F   H   K   E   L   T   D   N   K   A   H   K   L   L   I   L   240
AAGAATAAGGTGGAG CTGGCCCAGTTCCAA AACTCAGCCCCCCAA TATCTCACACTGGCC   812
  K   N   K   V   E   L   A   Q   F   Q   N   S   A   P   Q   Y   L   T   L   A   260
GAAGGCTTCTGGAAG GCCTTGTCTTCGCTG CCGTTAACGTACGAC TACTCTGCCTACCGC   872
  E   G   F   W   K   A   L   S   S   L   P   L   T   Y   D   Y   S   A   Y   R   280
CAGCTATTTCAGACA TACGGCACACATTAC CTCTCTGAGGGCTCA CTGGGAGGTGAATAC   932
  Q   L   F   Q   T   Y   G   T   H   Y   L   S   E   G   S   L   G   G   E   Y   300
CAAGGCTTGTTGGAA TTAGACCGCCAGGCG CTCGAATGGACCAGT ACAACAAATATAGAT   992
  Q   G   L   L   E   L   D   R   Q   A   L   E   W   T   S   T   T   N   I   D   320
TACCAGAGGTGTTGG AGAAAAGTGAAACGC CGTCTATTCAGGAAG AAAGTCAAGACCACC   1052
  Y   Q   R   C   W   R   K   V   K   R   R   L   F   R   K   K   V   K   T   T   340
TGTGAAAAACTGACA AAAGCTTTATCATCC AGAGACGGACACAAC GTAAACAAAATATCC   1112
  C   E   K   L   T   K   A   L   S   S   R   D   G   H   N   V   N   K   I   S   360
ATTAAAGTCAACGTG ATCGGAGGAGATCCG GCCTTAACATATACT CTGACTTCTCTGGAT   1172
  I   K   V   N   V   I   G   G   D   P   A   L   T   Y   T   L   T   S   L   D   380
GTGGAAAATCCTGAA GCCAACGGACAGAGC TATGACAATTGGGCC TCGTCTGTCAAAGAC   1232
  V   E   N   P   E   A   N   G   Q   S   Y   D   N   W   A   S   S   V   K   D   400
TTCCCTGAAGTCACA GACCAGAAGCTGCGT CCTCTGTATGAGCTG GTGAAGGAAGTGCAG   1292
  F   P   E   V   T   D   Q   K   L   R   P   L   Y   E   L   V   K   E   V   Q   420
TGTGCAGGTTTGAAG AAGCTCCACCTGAAA AGAGCCACAGAGGAG TACCTGGCAGAGGAG   1352
  C   A   G   L   K   K   L   H   L   K   R   A   T   E   E   Y   L   A   E   E   440
CATCCCTGTCACTGC CGGCCCTGCCAGAAC AATGGCCAGCCGCTG CTGACGGGCTCCGAA   1412
  H   P   C   H   C   R   P   C   Q   N   N   G   Q   P   L   L   T   G   S   E   460
TGTCGCTGCGTCTGC CGGCCGGGAACCTCA GGACAAGCTGCGAG AGGGGAGCTGTGACT   1472
  C   R   C   V   C   R   P   G   T   S   G   Q   A   C   E   R   G   A   V   T   480
GGAGAACAACCAGGG GTGATCCATGGCAGC TGGAGCTGCTGGTCC TCCTGGGGATCCTGC   1532
  G   E   Q   P   G   V   I   H   G   S   W   S   C   W   S   S   W   G   S   C   500
TCTGGAGGTCAAAGG TCAAGGACCCGAAGC TGCAACAACCCCGCC CCCAGCAGAGGTGGC   1592
  S   G   G   Q   R   S   R   T   R   S   C   N   N   P   A   P   S   R   G   G   520
CACCACTGCACTGGA CTGCAGGTGGAGCAG AAGCCTTGTGAGGAC GCAGACATCCAGTAC   1652
  H   H   C   T   G   L   Q   V   E   Q   K   P   C   E   D   A   D   I   Q   Y   540
TTACAGATGATGGAG CCTCAGTGCTTCAGT CTCTCTGTGACTCCA CCAAAGATGTGTGGG   1712
  L   Q   M   M   E   P   Q   C   F   S   L   S   V   T   P   P   K   M   C   G   560
CCGCCTCCAAACCTC AGGAATGGATTTATT CAGAATCCCAGAGAT TTTTACCTGGTTGGA   1772
  P   P   P   N   L   R   N   G   F   I   Q   N   P   R   D   F   Y   L   V   G   580
AACACGGCGGATAC TCCTGCATAGATGGT CATTACCTCAGTGGA GACGCTGGCCTCAA   1832
  N   T   A   E   Y   S   C   I   D   G   H   Y   L   S   G   D   A   V   A   Q   600
TGCATGGAGAACCAG AGGTGGAGGACAGGA GCGATGGTTTGTAAA AGTTCCTCTTGTGTG   1892
  C   M   E   N   Q   R   W   R   T   G   A   M   V   C   K   S   S   S   C   V   620
ATTCCTCTGCTCAAC AGTGAAGTTGTAGCA ACGCCCGCCAAATCA GCCTATCAGATTGGA   1952
  I   P   L   L   N   S   E   V   V   A   T   P   A   K   S   A   Y   Q   I   G   640
GACAGAGTACCTTG TCCTGTCCTGCGGGG TCAGTGCTGGACGCT GAGGTGTCAGAAATC   2012
  D   R   V   S   L   S   C   P   A   G   S   V   L   D   A   E   V   S   E   I   660
ATGTGCAGTCCCAGT CTGCAGTGGTCTCCT TCTCCAGCCAGCGCG CATTGTGAAGCTGCG   2072
  M   C   S   P   S   L   Q   W   S   P   S   P   A   S   A   H   C   E   A   A   680
CCCACAACTCCCTCT CCTTCATCTGGCCTG AAGTGTAAGCAGTGA GAGAATGTAGGAAAA   2132
  P   T   T   P   S   P   S   S   G   L   K   C   K   Q   W   E   N   V   G   K   700
ACGGACTGTGTGT AAATGCCCAAGTCAG TGCTTGACTTCGCTG CAGTTGTGCCAGA   2192
  T   D   C   V   C   K   M   P   S   Q   C   L   T   S   L   Q   L   C   A   R   720
TTCGGCTCAAGCCAA ACCCGTTACTTGGA GTTTGTCAGCTTGGA GCTCTGCGATGTATG   2252
  F   G   S   S   Q   T   R   L   L   G   V   C   Q   L   G   A   L   R   C   M   740
GGGCGAACCTTCACA TTGGCCAGGGACGGT GAGTGTGTGTGGCCA GAGGAGACATCCACA   2312
  G   R   T   F   T   L   A   R   D   G   E   C   V   W   P   E   E   T   S   T   760
TCCTGCGAGGACTGC AAACCAGGGACAATC TGTCAGGAACCGGCG AGAAAATGCGTTTGT   2372
  S   C   E   D   C   K   P   G   T   I   C   Q   E   P   A   R   K   C   V   C   780
CAGAATGCGTCAGAA TGCCCTAAAGACTCC GCCCCCCTGTGCGTC AGCTCTGGTGCTGGC   2432
  Q   N   A   S   E   C   P   K   D   S   A   P   L   C   V   S   S   G   A   G   800
GTTGCCACAACGATG ACCGAGTGCGAGGTT GGGGCCAGAGGCGT GCTGGGGAGCAGGTC   2492
  V   A   T   T   M   T   E   C   E   V   G   A   R   R   C   A   G   E   Q   V   820
AATGTGATCAGTATT GACGCGCATGTCCAGAA TAA AGATGGCTGTTG TTTTCCTGAGCAGTC   2552
  N   V   I   S   I   D   A   C   P   E                                             830
TCACGATGTATATAA AGGAGGAACTTGTTT TTTTCTCTCAATGCC TGAATTGTAAATAGA   2612
GTAGTATTTTTTCTG GAACGAACACTTGAA TGATGTAAATACAGT ATGTTCCCCTCTGT   2672
CCTGTGGTGTCTTTT TATATTTGTCATGGA GGACTGATGATAGAC AACATTTGCATAAAG   2732
TAAACATAAGGTTCT GTTCTAAAGGTTTGA TAAAATCTTA AATAA A TGCATTTtATAAAC   2792
TGTTTTTTTTTtAGA TGACGTTAATAATAA A                                    2823
```

Figure 48. Nucleotide and deduced amino acid sequence of Rb-C7 (JX207114)

The nucleotide sequence (upper) and the deduced amino acid sequence (lower) are numbered.
The start codon (ATG), the stop codon (TAA), and the polyadenylation signal sequence
(AATAAA) are boxed. The predicted signal peptide is underlined, and the thrombospondin

type-1 (TSP-1) repeat profiles are double-underlined. The low-density lipoprotein receptor domain class A signature (LDLR-A) is underlined and italicized. The membrane attack complex/perforin (MACPF) domain profile is in bold, and the MACPF MAC/perforin domain is in bold and shaded in gray. The MACPF domain signature is in bold, underlined, and shaded in gray. The epidermal growth factor-like-1 (EGF-1) domain signature is in white bolded letters, and shaded in gray. The Sushi/CCP/SCR domain profiles are in bold and wave-underlined. The FIMAC domains are dot-dash underlined with gray shading.

3.1.2 Characterization of the Rb-C7 genomic structure and promoter region

PCR-based screening of the BAC genomic library using *Rb-C7* gene-specific primers led to the identification of only one positive BAC clone that contained the sequence of interest. The putative genomic sequence of *Rb-C7* was recognized by comparison with the full-length cDNA sequence and then compared by genome alignment with known C7 genomic DNA structures of several other vertebrate species published in the Ensembl sequence database. The *Rb-C7* genomic sequence (JX207114) was about 10.4 kb and was composed of 18 exons interrupted by 17 introns. The 5'- and 3'-ends of all introns harbor canonical splicing motifs (according to the GT-AG rule). The number of exons and introns were found to be well conserved among all the C7 orthologous (Figure 49).
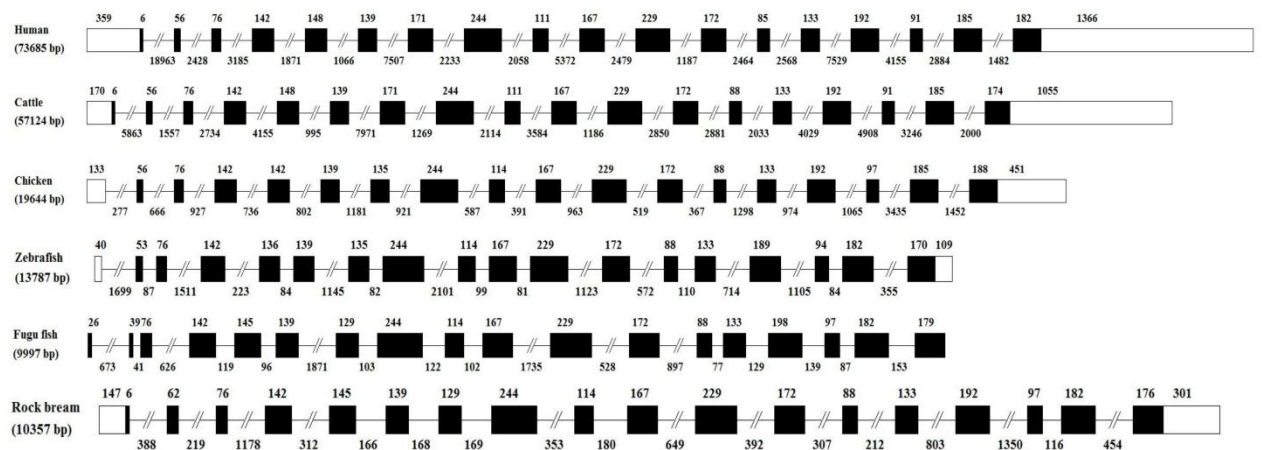
Figure 49. Genomic structure of Rb-C7 compared to genome structures of other C7 genes

Exons are shown in black solid boxes. UTRs are shown in light-colored boxes. Introns are shown as thin lines between the exons, and introns larger than 200 bp are indicated by interrupted lines. Numbers above the boxes and below the lines indicate the lengths of exons and introns, respectively. The C7 genes shown represent *Homo sapiens* (ENST00000313164), *Bos taurus* (ENSBTAT00000056663), *Gallus gallus* (ENSGALT00000023947) *Danio rerio* (ENSDART00000132741), and *Takifugu rubripes* (ENSTRUT00000033558).

According to the BDGP prediction analysis, the promoter of *Rb-C7* has complex structure with a co-promoter region and contains multiple binding sites for proteins with functions related to the transcription initiation process. A transcription initiation site (+1 site) was found at 32 bp upstream of the start codon (ATG), and a TATA (TATATTA) box was also been identified at -25 to -31 upstream from the putative +1 site. Binding sites for immune-related transcription factors were also identified, and these included the CCAAT-enhancer binding protein (C/EBP), nuclear factor-kappa B (NF-κB), activator protein-1 (AP-1), interferon regulatory factor-1 (IRF-1), octamer transcription factor (OCT-1), specificity protein-1 (SP-1), hepatocyte nuclear factor-3 (HNF-3), and nuclear factor-1 (NF-1) (Figure 50).

```
                                                      TACTCCAAAAAAAAGTTC -1433

ATATAAGTTTATATATGGAAAATAAATGAGGATAGGAATGGAAAGTAAAAAAAGGAAAAAAATAAGAATG -1363
                                 IRF-1
ATGCATCGTAGTAATAATGGTAGGTGGAGACGAGCTGGGTCTTGTGGGATTTGCAAAACAAGCCCCTGGC -1293
                                             OCT-1
CAACATTCCCTACCACAGCCCAGCCTAGAAGAGAACCCTGCCGCCGGAGGACAAGGAGTTTATGAAAGAC -1223
                                        SP-1
CAGAAAGGAAAGAAGCAGTCTCCACACACTGTCTGCTTCATCTCTAATGGCAACAGGTAGACATACTGTG -1153

GAGCTGCAGTGTGACATCTGTGTCTTGGTGCTAAACACTAAAGAAGAAAATAAATCACCACTAGATGGCA -1083

ACCACAACTAACTATGTAAGTAGGTCGATGACGAATCACTGCACACTGGCTACGATCCAGTCAGAGCTCT -1013

CTGTTTATCCGGCGGTGGCTTTCAAGTTGGATAAGATCCAGATGTGGTGAGGGTGCTTTTTGTTTGGTTG  -943

TTCCTGTTCTCTGGCACAAACTGCGGGATGTGTCCCCTTTGTAAAGCAAACTTACAGAGATTAAGTCAAC  -873
                                 NF-κ
ATGATTAAGTCACACACAATAAATAACAACTCATTGTGGTTGTAAGTGTGCCCAAGACAAAAATAAAACA  -803
                                                             HNF-3
TTTCAAGTCCAGGAGAATTTTGTACAGCATATTTTGAGATATTTGTTGCAGGGTGTATGTAGGGTTTGAT  -733

GACTGTGGTGTGCAGATGGAGTAAAGGAACATTTTTGCTTTATCTATGGAATTGAGTTTATATTTACAAA  -663
                                    OCT-1
TATAAAACAAACAGGTAAAAACAAAACTAAAAACTAAGAAATAGCACTTAATAAAAAACTGTTTTGAATA  -593
OCT-1             HNF-3
AGATATTGCGTTTCTGGCTAAAATTTGAAATGTGGATTGAATTAATACTAGCATGCTGGCATAAAAAGAA  -523

AATTGGCATTACAAATGGAGAAGCGGATGTTATTTCCACTTGCATTCTGTGTTTACAGAGCAACAGAGTC  -453

TTTCAAATATGGGCCAAATTGGGATTTTTTGGACCTTTGTTCAAATCGTGCAGACTCAGCTTCTGACAGC  -383
        NF-1
AGGATAACCACTTGTGACGAAACTACTCCTGTCGCAATCCGGTGAAACAGTTCTGACAAAAAAGATTTCA  -313
                          C/EBPα
CCGCACTCCGTCTTGTACCCCTTCAGCACTCAGAGCAAACAGGATTATCAAAATGTTTACTGGAAATAAA  -243
                                                           C/EBP
CTCACAGTGTTGTTGGAAACAATCCAGCGCAGATTAAAGAGCAAACAAAGCCTGAGTCCAGGCTGCACT  -173
                                        C/EBPα        AP-1
GGAGCTGTGGACAACAACACTGGGTCTCTGTTCATGACTCACTGAGTGAAACGTGAGATGGATGACAGTT  -103
     SP-1
TCCTCACTGTGGGAATTCCAGAACTTGGGGATCTGTTGAGTAATCCCCCCCCACCCCACCCCCCCTTTGA   -33
         NF-κB                    C/EBPα           SP-1
CTATATTACCCAGAGTGTCACAGTGCTGGACCAGACGACAGGGCTGAACTCTGGGGAGCACATCATGAAG    38
 TATA box                    +1 site                                M  K
```

Figure 50. 5'-flanking region of the Rb-C7 (JX207114)
Putative transcription factor binding sites are shown in bold and underlined with the corresponding factor identified below. The putative TATA box is boxed and identified below. The transcription initiation site (+1 site) is indicated by a bent arrow. The start codon of the ORF (italicized uppercase, in bold) and the translated amino acids (uppercase, in bold) are shown.

135

3.1.3 Identity, similarity, and phylogenic relationship of Rb-C7

Pairwise multiple sequence analysis was revealed that *Rb-C7* has 76.5% amino acid identity with the orthologous from Japanese flounder (*Paralichthys olivaceus*: AB020964) and 64.2% with the orthologous from rainbow trout (*Oncorhynchus mykiss:* NM_001124618). It also showed >50% identity with the orthologous from other fish species, including grass carp (*Ctenopharyngodon idella*: JN655169), silver carp (*Hypophthalmichthys molitrix*: JN806258), and zebrafish (*Danio rerio*: BC100054). The identity dropped below 50%, but remained high (at 43%) for orthologous from the higher vertebrate species, including pig (*Sus scrofa*: NM_214282), mouse (*Mus musculus*: NM_001243837), and human (*Homo sapiens*: BC063851) (Table 15). Multiple sequence alignment of *Rb-C7* was indicated that each motif domain has higher similarity with the corresponding domain of other considered species. All the cysteine residues were found to be conserved across nearly all the orthologous, and certain (presumably important) signature motifs were also highly conserved among all the orthologous (Figure 51).

Table 15. Percentages of identity, similarity, and gaps of Rb-C7 with C7 vertebrate orthologous

Common Name	Species	Accession No.	AA	Identity, %	Similarity, %	Gap, %
Japanese flounder	*Paralichthys olivaceus*	AB020964	805	76.5	84.0	3.0
Rainbow trout	*Oncorhynchus mykiss*	NM_001124618	808	64.2	76.4	3.4
Grass carp	*Ctenopharyngodon idella*	JN655169	821	55.2	69.0	2.3
Sliver carp	*Hypophthalmichthys molitrix*	JN806258	827	55.1	69.0	2.5
Zebrafish	*Danio rerio*	BC100054	829	53.7	67.9	3.0
Atlantic salmon	*Salmo salar*	NM_001139773	845	48.3	62.8	5.2
Chicken	*Gallus gallus*	XM_424774	834	44.8	62.6	2.8
Pig	*Sus scrofa*	NM_214282	843	44.2	58.8	5.5
African frog	*Xenopus laevis*	NM_001091647	830	43.7	60.1	2.4
Human	*Homo sapiens*	BC063851	843	43.6	59.3	6.1
Mouse	*Mus musculus*	NM_001243837	845	43.4	59.2	4.1

AA, amino acid. GenBank accession numbers are shown.

The evolutionary relationship was studied using the deduced amino acid sequences of *Rb-C7* and all available sequences for the complement components published in the NCBI database. The results were depicted as a phylogenetic tree as in Figure 45. The phylogenetic tree represents each of the lytic pathway genes (C6, C7, C8α, C8β, and C9) as separate clusters, with C3 from sea cucumber (*Apostichopus japonicas*) forming a distinctive outer group. *Rb-C7* clustered together with its orthologous and was closely associated with the Japanese flounder orthologous.
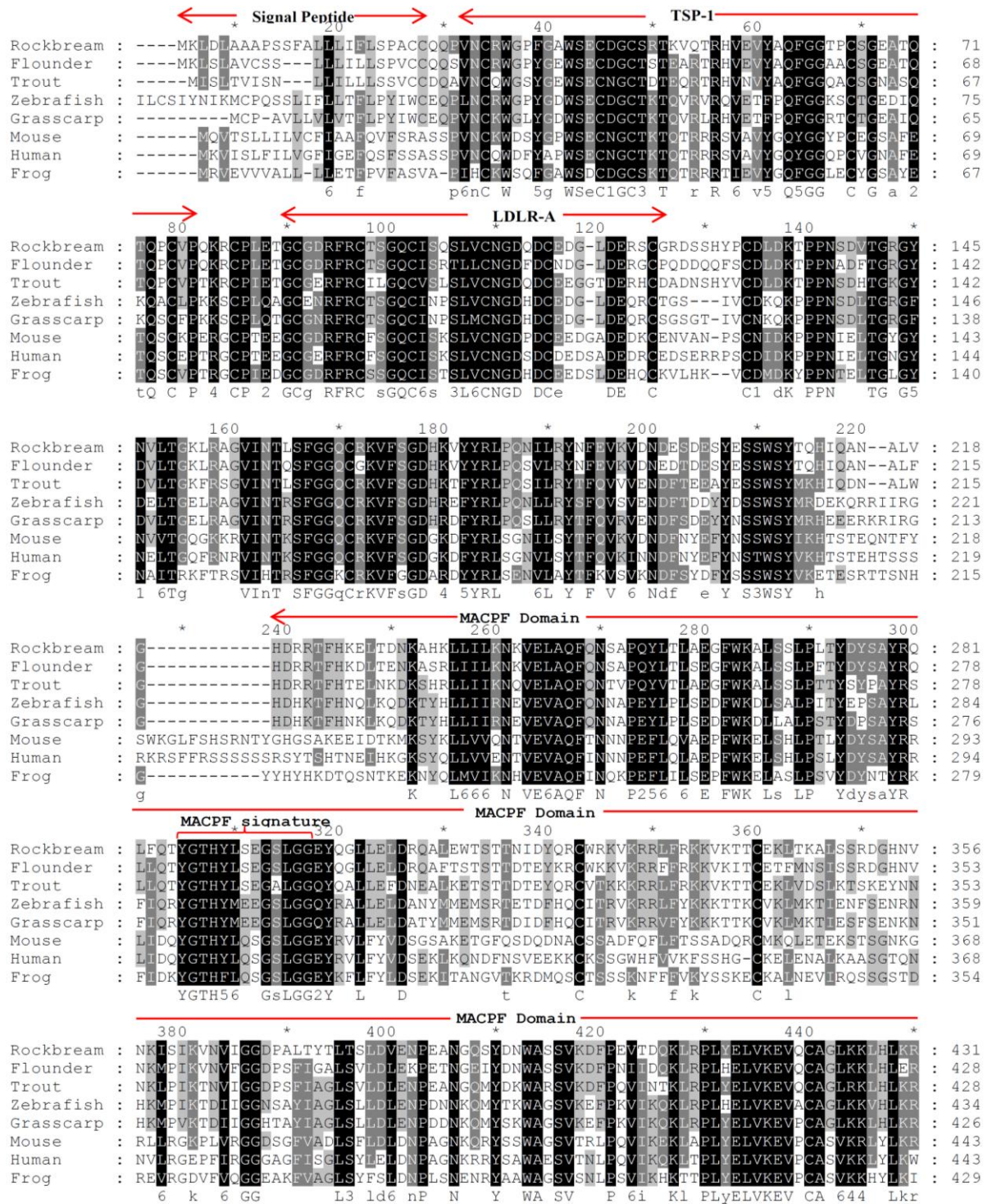
Figure 51. Multiple sequence alignment of Rb-C7 and C7 orthologous from other vertebrates

Identical residues among all orthologous' sequences are denoted by uppercase letters. Similar residues (with single discrepancies among the orthologous) are denoted by lowercase letters or numbers. Gradations of similarity frequency are denoted by dark-to-pale shading (100%, 80% and 60% respectively). Missing amino acids are denoted by dashes, and the conserved domain is indicated by double-arrows. The GenBank accession number of each gene is presented under the C7 sub-cluster next to each species.

EGF-1   TSP-1
                460   *   480   500   520
Rockbream : ATEEYLAEEHPCHCRPCQNNGQPLLTGSECRCVCRPGTSGQACERGAVTGEQPGVIHGSWSCWSSWGSCSGG--Q : 504
Flounder : ATEEYLSEEHPCHCRPCQNNGQPLLSGSVCRCVCRPGTSGPACQTGAVIGEQPGLIHGSWSCWSSWGSCSGG--R : 501
Trout   : ATEEYLAEEHPCHCRPCHNNGQPLLTGTECKCVCRPGTSGPACESGTVIGEQPGVIHGSWSCWSSWGSCSQS--K : 501
Zebrafish : ALEAYLEEQSPCHCRPCQNNGMAVLSEGVCTCVCRPGTSGNACQNGHVLGEQPGVIEGCWSCWSAWSSCSHG--Q : 507
Grasscarp : ALEAYLEEQSPCHCSPCQNNGMAVLTEGVCTCVCRPGTSGNACQNGHVLGEQPGVIEGSWSCWSAWSSCSRG--Q : 499
Mouse   : ATEEYFDEFDPCHCRPCQNGGLAIVVETQCQCLCKPYTFGSACEQGVLVGDQAGGVDGCWNCSWSWSPCVQG--K : 516
Human   : ALEEYLDEFDPCHCRPCQNGGLATVEGTHCLCHCKPYTFGAACEQGVLVGNQAGGVDGCWNCSWSPCVQG--K : 516
Frog   : ATEQYINEESPCRCKPCKNKGVPGVEGTQCVCYCKPYTYGPACQYGILAEGDPAVVDGSWSCWSSWSPCLKGTRR : 504
            A E Y E PChC PCqN G   6   C C C4P T G AC2 G 6 g qpg 6 G WsCWSsW C g

                        TSP-1
              *   540   *   560   *   580   *   600
Rockbream : RSRTRSCNNPAPSRGGHHCTGLQVEQKPCEDADIQYLQMMEPQCFSLSVTPPKMCGPPPNLRNGFIQNPRDFYLV : 579
Flounder : MSRTRSCNNPAPSRGGQHCTGLRAEQKPCEDPQLQYLQTMEPQCFSLSVTPPKTCGPPPNLRNGFIQNPRDFYVV : 576
Trout   : MSRSRTCTNPAPRRGGLHCVGLPTEHTSCEESDLQHLKMMEPQCFGLSISTPKACRAPPALRNGFVLSPRDVYLV : 576
Zebrafish : KSRTRSCNNPTPRNGGKNCIGETIERRSCEEPDFEHLKILEPHCFDPTLTPVKTCKTPPLANGFVLDPKDIYTV : 582
Grasscarp : KSRTRSCNNPAPRNGGRNCIGETIERKNCEDPDFEILKMMEPQCFDPTLTPVKTCKTPPPLTNGFVLDPKDIYPV : 574
Mouse   : RTRSRECNNPPPRDDGKSCLGETTESKQCEDQDLEKLRLLEPHCFHSSLAPKEFCLSPPALKDGFVQGEGTMVPV : 591
Human   : KTRSRECNNPPPSGGGRSCVGETTESTQCEDEELEHLRLLEPHCFPLSLVPTEFCPSPPALKDGFVQDEGTMFPV : 591
Frog   : RVRNRVCNNPRPSGGGLNCIGDSIASELCEDSELEHLRTVEPHCEDISIVPTEFCPPAELLENGYVQDANPPYPA : 579
            R R CnNP P gG C G   e   CEd   2 L 6EP CF   36 p   C pP L 1G56   v

             Sushi/CCP/SCR domain-1   Sushi/CCP/SCR domain-2
              *   620   *   640   *   660
Rockbream : GNTAEYSCIDGHYLSGDAVAQCMENQRWRTGAMVCKSSSCVIPLLNSEVVATPAKSAYQIGDRVSLSCPAGSVLD : 654
Flounder : GNTVEFSCTEGHYLSGDAVTRCTENQTWTPRTTVCKSTTCGIPQLGAEVMATPTKAAYQIGDGVSLSCPAGSLLD : 651
Trout   : GSKIEYSCIEGHYITGETVAECTDNNSWTRGPLECKSATCDVPSLENDVTGTPWKVIYQIGDSVYLSCPVG-MMR : 650
Zebrafish : GKKIEYICTDGYDRIGNPFAECTESLTWRISPMECKKSECDAPAVLRNVIVVPLKQSYRIGDSVTLSCPSG-MQK : 656
Grasscarp : GRKIEYTCADGHDHIGNPVAECTENLTWKMSPMECKKTACDPPAVLSNVIATPSKQTYLIGESVTLSCPSG-MQK : 648
Mouse   : GQSVVYACDEGYSLIGDPVARCGEDLQWLVGEMHCQKLACVLPGEMNGMQSHPQKPFYMVGEKVTLSCVG-MSL : 665
Human   : GKNVVYTCNEGYSLIGNPVARCGEDLRWLVGEMHCQKIACVLPVEMDGIQOSHPQKPFYTVGEKVTVSCSGG-MSL : 665
Frog   : GRSIAYTCKLGYTLVGDSMARCGEDLKWQAGQMKCQQIMCSPPVLPGILLVAPEKESYEIGDKITLSCARG-LDL : 653
            G   5 C G   G   aC e   W   C   C P   6   PK Y 6G 6 6SC G 6

              680   *   700   720   *   740   *
Rockbream : AEVSEIMCSPSLQWSPSPASAHCEAAPT-TPSPSSGLKCKQWENVGKTDCVCKMPSQCLTSLQLCARFGS-SQTR : 727
Flounder : GEVSEVICSPSLQWSPSPAGARCKAAPT-APPPPSGLKCKPWETAGKNKCVCKMPFQCSASLLLCARLG--SAPK : 723
Trout   : ECMAEIVCSSSLQWSPSPDSIRCREVPK-GPTPPLGLNCKLWETPGKSQCVCKLPYQCQPSLQLCATLHT-GRTS : 723
Zebrafish : DCEGEIRCRLGLSWYPDPKSVRCNPVEA-VPTQPS-LLCKPWEKPGTGQCVCKMPFECKTSLHVCATVRP-GRVN : 728
Grasscarp : ECEAEIMCRLGLSWYPAPESVRCNPVEA-VPTQPS-LFCKPWEKQAKGQCVCKMPYECKASLHVCASVRP-GRVN : 720
Mouse   : ECPSTLLCGSSLKWSPELKDAKCVQKATSTPPPAAP-ACQPWEKLQNSKCICKMPYECGSSLDVCARDESRNKIL : 739
Human   : ECPSAFLCGSSLKWSPEMKNARCVQKEN-PLTQAVP-KCQRWEKLQNSRCVCKMPYECGPSLDVCAVQDERSKRIL : 738
Frog   : ECPNTISCTSSLTWLPSIKNAECLAKATPTTAQSLGSQCQPWEKFEDGKCSCRMPYECGASLDVCAVDGRNNKNV : 728
            g   C sL W P   C   p   C WE   C C46P 2C SL 6CA

              FIMAC
              *   760   *   780   *   800   *   820
Rockbream : LLGVCQLGALRCMGRTFTLARDGECVWPEETSTSCEDCKPGTICQEPARKCVCQNASECPKDSAPLCVSSGA-GV : 801
Flounder : LLSTCQLGALQCLGRSFSLTSDADGQWPEG-----------AFCQESVR--------ECPTDSAPLCVSS---GV : 776
Trout   : KLGMCQLGALQCLGRSFKLVKDSDCDWSDQG---------LTSCQE-----------ECFDTSVQLCVELPDSAV : 778
Zebrafish : RMSVCQLGALQCLGQTFTLLQDSACSWPETKFTSCQDCHQWETCDGSKCDCK--DPEDCPDDSAHVCVSLMG-GA : 800
Grasscarp : GMSICQLGAFQCLGQTFSLLQDSACSWPETKFTSCQDCHQWETCDGFKCECI--DPDDCSDDSAHLCVSIKG-GA : 792
Mouse   : SLTICKMHVVHCHKRNYTLVGKESCTLPSPAEKVCGACPIWSKCDAQNGKCICREASECQTAGYRICVEVNG--K : 812
Human   : PLTVCKMHVLHCQGRNYTLTGRDSCTLPASAEKACGACPLWGKCDAESSKCVCREASECEEEGFSICVEVNG--K : 811
Frog   : ALTVCKMHALQCLGRKYTLTNDAKCAFPVATERSCDSCHPWEQCKVQTNKCVCREAKTCGNEGISICVNVDG--K : 801
            6   C C6a C gr 5 L d C p   c c   C   c   C   C   6CV

              FIMAC
              *   840   *
Rockbream : ATTMTECEVGARRCAGEQVNVISIDACPE---- : 830
Flounder : VTTMTECEVGARRCGGELMNVIGIDACPE---- : 805
Trout   : AVTMSECEVGARRCQGEQFDVVSIKACSAL--- : 808
Zebrafish : PEMMSECEAGAWRCRGKEMNVLRIGACQS---- : 829
Grasscarp : PEMMSECEAGAWRCQGKEITVLSIGDCQA---- : 821
Mouse   : EETMSECEAGILRCRGQSISISITAIKPCAEEAAQ : 845
Human   : EQTMSECEAGALRCRGQSISVTSIRPCAAETQ- : 843
Frog   : LQTMSECECEAGIRKCQGENAIVVSLSPCDV---- : 830
            tM3ECE Ga 4C G   6 6   C

Figure 51. Continued.

3.1.4 Tissue-specific mRNA expression of Rb-C7 in normal healthy conditions

To examine the normal tissue distribution pattern of *Rb-C7* transcription, qPCR was carried out for blood, gill, liver, spleen, head kidney, kidney, skin, muscle, heart, brain, and intestine (Figure 46). *Rb-C7* was found to be constitutively expressed in all tissues examined, although the level of expression varied between each. Compared to the normalized expression level in muscle (lowest expression), the most abundant expression was detected in liver and marginal expression was detected in gills and skin.

3.1.5 Modulation of Rb-C7 expression in response to immune stimulation

After challenges with the *E. tarda, S. iniae*, LPS and RBIV, the amount of Rb-C7 transcripts in head kidney were significantly ($P<0.05$) up-regulated from baseline (0 h) within 12 h and through 48 h post-injection (p.i.) (Figure 52). The highest fold-induction was detected at 24 h p.i. of *E. tarda* (224-fold increase relative to the un-injected control at 0 h). This result indicates that Gram-negative bacteria can effectively trigger Rb-C7 expression in head kidney. Similarly, a previous study conducted in grass carp also found that C7 expression was significantly induced in head kidney by the Gram-negative bacteria *Aeromonas hydrophila* (Shen et al., 2012a). In comparison with head kidney, Rb-C7 expression in liver tissue showed similar patterns with the same challenges; however, the overall fold-inductions were much lower than those in head kidney tissue (Figure 52) Moreover, in contrast to the acute response in head kidney, *Rb-C7* expression in RBIV-challenged liver was showed a latent response, with up-regulation only at 48 h p.i. (Figure 52). Furthermore, *in situ* hybridization results were confirmed the qPCR data (Figure 52). *Rb-C7* mRNA-expressing cells were observed in both control and *E. tarda* challenged head kidney and liver tissues. Compared to control, significantly higher number of hybridized *Rb-C7* mRNA-expressing cells and

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

increased intracellular expression level were shown in both tissues at 12 h post challenge. TCCs have been demonstrated to be activated in response to general microbial infections (Li et al., 2007; Wickramaarachchi et al., 2012). However, the regulatory mechanism underlying such a wide spectrum transcriptional activation has not yet been fully elucidated. The observed robust expression of *Rb-C7* transcript levels in the complement-related organ (liver) of healthy fish and its significant enhancement accompanying immune challenge suggest the obligatory requirement of C7 molecules for MAC formation.
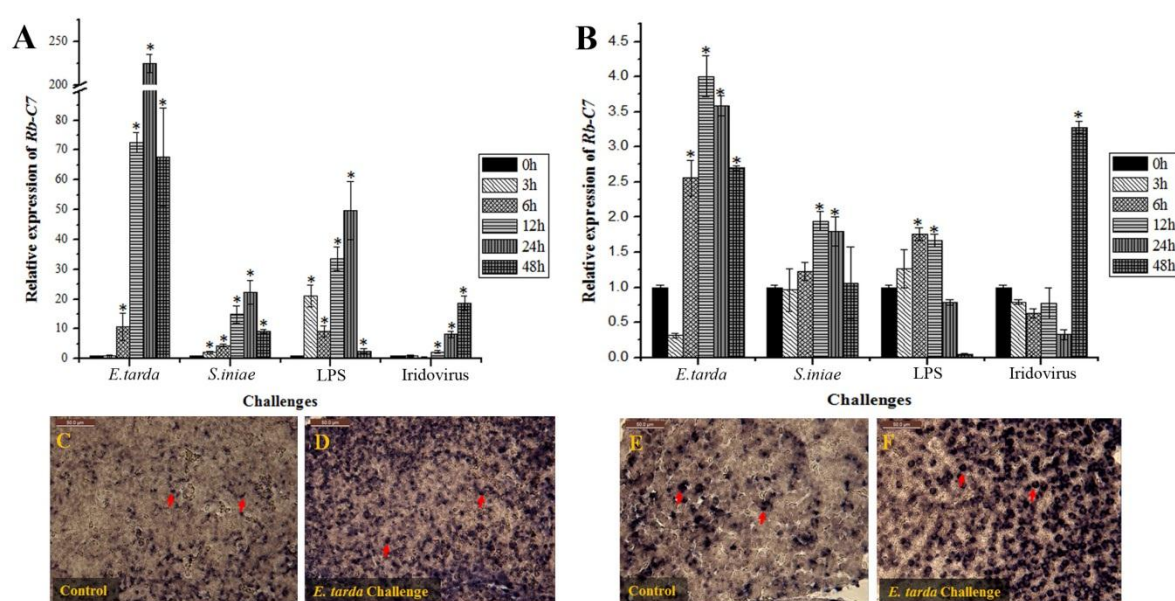
Figure 52. Expression analysis of Rb-C7 gene in immune-challenged rock bream tissues
Relative expression of *Rb-C7* after challenge with LPS and intact pathogens in head kidney (A) and in liver (B) were quantified by qPCR assay. *In situ* hybridization assay was conducted to investigate *Rb-C7* hybridized mRNA distribution in head kidney of healthy fish (C), head kidney after challenge (12 h) with *E. tarda* (D), liver of healthy fish (E) and liver after challenge (12 h) with *E. tarda* (F). In qPCR data, the amount of Rb-C7 mRNA relative to β-actin is expressed as mean ± SD (n=3). * $P<0.05$ vs. un-injected control at 0 h. mRNA *in situ* hybridization data, hybridized cells are shown in dark purple in each tissue section (representatively indicated by arrowheads).

3.2 Discussion

Not surprisingly, studies of teleost species have identified all the complement components of the lytic pathway that are known in humans; in particular, the rate-limiting C7 factors of the teleost species have shown similar molecular characteristics, domain architecture, tissue expression, and transcriptional response to bacterial stimuli (Papanastasiou and Zarkadis, 2005; Shen et al., 2012a; Zarkadis et al., 2005). However, no study to date has investigated a teleost C7 transcriptional response to a broader panel of various pathogen-related stimuli. In the present study, a MAC component (*Rb-C7*) gene encoding an MACPF super family protein of the complement lytic pathway was isolated, sequenced, and characterized at the genomic structural level from the teleost rock bream. Furthermore, the gene expression profile of this newly discovered C7 was defined for the healthy organism and in response to *E. tarda*, *S. iniae, LPS,* and RBIV exposures.

Previous studies have been indicated that the evolutionary diversity of the C7 orthologous resulted from a chromosomal duplication event that was believed to have been occurred in a common ancestor of the teleost lineage (Ravi and Venkatesh, 2008). Indeed, in the rock bream genome, C6 lies 1.6 kb downstream of the newly identified *Rb-C7*, which suggests that the panel of rock bream MAC components may occur in the same or adjacent chromosomal levels. According to our comprehensive screening of the rock bream BAC library, two C7 exists in the rock bream genome, similar to rainbow trout which carries two C7 genes (Papanastasiou and Zarkadis, 2005).

The *Rb-C7* gene also contains this identical domains signature. Up to date, several hundred MACPF-containing orthologous have been identified, and they all contain the MACPF signature motif ((Y/W)G(T/S)H(F/Y)X_6GG) within the MACPF domain (Ponting, 1999). Similarly, the MACPF domain of *Rb-C7* contains this signature domain, which further supports its' potential role as a complement lytic pathway gene. Studies of fish TCCs have

suggested that the presence of at least one C-terminal TSP-1 may be one of the unique features of these orthologous (Yeo et al., 1997); indeed, *Rb-C7* featured two TSP-1 domains in its C-terminus. The FIMAC domain has been shown to mediate the protein's stability and to enhance the binding ability of the C5b-6 complex, which is a characteristic of the C-terminal (Thai and Ogata, 2004). The presence of a second FIMAC domain, as found in *Rb-C7*, may increases the affinity and specificity of the C7 protein's interactions with the C5b-6 complex (DiScipio, 1992). Presence of highly conserved cysteine residues in FIMAC domains increases the structural integrity of secondary and tertiary structures of the C7 protein, which facilitates the firm binding of the C5b-6 complex. Furthermore, the cysteine residues of the *Rb-C7* protein were showed remarkably high conservation with those from all other species evaluated in our study.

Many of the C7 orthologous, including *Rb-C7*, show a similar exon-intron structure consisting of 18 exons interrupted by 17 introns. However, the size of the introns varies between these orthologous, which explains the differences in total length (ranging from 9 kb of the *Takifugu rubripes* (fugu) protein to 73 kb for the human protein). However, there was a trend noticed among the fish C7 orthologous, in which a smaller number of the introns were lengthy. For an example, while the zebrafish C7 contains six introns that are longer than 1 kb, the fugu and rock bream C7s contain only two such lengthy introns. Detailed analysis of the C7 exons was revealed that exons 3, 4, 6, 8, 10, 11, 12, and 14 were identical in length for all species examined; regardless, the genomic structure of *Rb-C7* was the most similar to that of the fugu C7.

Although teleost research has recently focused on identification of complement genes, to the best of our knowledge no detailed deletion study has been carried out to understand the C7 promoter region and only one 5' sequence region (in grass carp C7) has been analyzed using a computer-based approach (Shen et al., 2012a). Deletion studies of the human C7

promoter identified a single transcription start site at 115 bp upstream from the initiation codon (Gonzalez et al., 2003), while the *in silico* study of the grass carp promoter predicted it at 39 bp upstream of the start codon. In the *Rb-C7* promoter region, a putative transcriptional initiation site and TATA box were found at 32 bp and 57 bp upstream of the start codon, respectively. This locale was close to the grass carp predicted transcription start site, but future deletion studies are necessary to identify the actual site of transcription initiation.

According to our analysis, several important putative transcriptional binding sites were located in the 5′-flanking region of *Rb-C7*. In comparison, the human C7 (Gonzalez et al., 2003), grass carp C7 (Shen et al., 2012a), and *Rb-C7* contain C/EBP and HNF-3 transcription factor binding sites. In addition, the *Rb-C7* promoter region contains putative transcription factor binding sites for NF-κB, AP-1, IRF-1, OCT-1, SP-1, and NF-1, which are present in other TCC genes' promoters (Li et al., 2007; Wickramaarachchi et al., 2012). This promoter profile may imply that the mechanisms regulating the *Rb-C7* gene transcription may be similar to those of other fish TCC genes and mammalian C7. Among the putative *Rb-C7* transcription factors, NF-κB is an important eukaryotic transcription factor that has been implicated in the regulation of a large number of genes that control basic cellular and tissue-specific processes, including immune and inflammatory responses, development, cellular growth, and apoptosis (Long et al., 2004; Pikarsky et al., 2004). IRF-1 is another putative binding factor of *Rb-C7*, belongs to the IRF family of proteins that play important roles in the interferon-mediated immune response (Barnes et al., 2002). Likewise, Oct-1 is involved in regulation of some T cell-specific cytokines' expression, such as interleukin (IL)-2, IL-4, and IL-5. Thus, the *Rb-C7* promoter contains several important putative transcription factor binding sites that are similar to those previously identified in the C7s from human and grass carp. Moreover, most of these sites are associated with factors related to immunity which supports the hypothesis that *Rb-C7* may have an immune function, similar to that described

144

for the C7s in human and grass carp. However, further studies are needed to confirm these binding sites as functional elements of the *Rb-C7* promoter.

In mammals, the liver is the primary production site for the majority of the complement components (Alper et al., 1980; Witzel-Schlomp et al., 2001). The expression patterns of C7s in teleost fishes reflect this phenomenon; for an example, the highest expression level of C7 in grass carp was detected in liver, followed by head kidney, trunk kidney, and intestine, respectively. However, the C7 expression in rainbow trout was most robust in liver, brain, heart and intestine, and completely absent in spleen and kidney. In our rock bream study, we also observed that expression of *Rb-C7* was extremely high in liver, as compared to other tissues; moreover, the responsive of *Rb-C7* expression in liver to immune stimulation, suggests that *Rb-C7* may act similarly to the other complement components that belong to the acute-phase proteins that are generally of hepatic origin (Witzel-Schlomp et al., 2001). This hypothesis is further supported by the presence of liver-specific transcription factor binding sites, such as HNF-3 and C/EBP, in the promoter region (Dogra and May, 1997). However, *Rb-C7* expression in gill, another highly reactive immune tissue, was very low. Grass carp and rainbow trout C7 expression was previously found to be reduced in gill, and those results were comparable to another study of murine C9-responsiveness to LPS challenge in which no expression of C9 was detected in lungs of healthy mice (Bolger et al., 2007). Taken together, these results may indicate that *Rb-C7* behaves in a similar manner to the other previously described TCC genes.

According to the immune-responsive expression analyses of *Rb-C7*, Gram-negative bacteria can effectively trigger expression in head kidney between 6 h to 48 h p.i. The induced transcription level was highest at 24 h p.i., which also represented the highest fold-induction among all the immune stimulants examined in this study. A previous study conducted in grass carp also found that C7 expression was induced in head kidney by the

145

Gram-negative bacteria *Aeromonas hydrophila* between 4 h and 168 h (Shen et al., 2012a). To the best of our knowledge, only two studies of C7 orthologous have included analyses of pathogen-induced expression conditions, and both of those used Gram-negative bacteria (Gonzalez et al., 2007b; Shen et al., 2012a). Other TCC genes, however, have been studied in response to Gram-positive bacteria, LPS, and dsDNA virus, and all showed remarkable inducement of expression upon exposure to these stimulants (Wickramaarachchi et al., 2012). Recent microarray studies have also provided evidence of up-regulation of immune genes in response to pathogenic challenges. Zebrafish C9 transcripts were shown to be increased in both fins and organs (liver, spleen and head kidney) at 48 h post-challenge with viral hemorrhagic septicemia virus (VHSV) (Encinas et al., 2010). Similarly, *Rb-C7* was also up-regulated in both head kidney and liver at 48 h p.i. of RBIV. Therefore, we hypothesize that *Rb-C7* may also be involved in the virus-mediated immune response of rock bream.

TCCs may be activated in response to general microbial infections (Li et al., 2007), but the regulatory mechanism underlying such a general activation of TCC genes' transcription has yet to be fully elucidated. In the current study, the up-regulated transcriptional profile of *Rb-C7* upon exposure to Gram-negative bacteria and RBIV can be substantiated through the presence of immune signaling transcription factor binding sites in the promoter region, such as NF-κB, AP-1, and IRF-1 (virus-specific) which are themselves regulated by toll-like receptor-dependent immune signaling pathways (Hawlisch and Kohl, 2006; Martin and Wesche, 2002). Furthermore, the observed robust expression of *Rb-C7* transcript levels in the complement-related organ (liver) of healthy fish and its significant enhancement accompanying immune challenge likely reflect the obligatory requirement of C7 molecules for MAC formation.

In summary, we have identified the complete genome sequence of complement component 7 from rock bream. Genome structure analysis of *Rb-C7* was confirmed its significant

similarity with C7s of other teleost fishes and classified it as a TCC protein family member. Moreover, *Rb-C7* up-regulation was accompanied experimental challenge with *E. tarda*, *S. iniae,* LPS, and RBIV, indicating a potential role for this gene and its encoded protein in the immune defense system of rock bream.

4. Results and discussion of rock bream complement C8

4.1 Results

4.1.1 *In silico* analysis of Rb-C8α, Rb-C8β and Rb-C8γ

Based on BLAST-X analysis of ESTs data from the normalized cDNA library, we identified putative, complete cDNA homologues of *Rb-C8α*, *Rb-C8β* and *Rb-C8γ* and compared with exons of genomic sequences obtained from BAC library. The alignment results showed that ORF of cDNA completely matched with the derived ORF of genomic DNA of all genes. Hence, the genomic sequences were deposited in NCBI under accession numbers of JQ710660 , JQ710661 and KF682443 respectively.

The complete cDNA sequence of *Rb-C8α* was composed of 2070 bp, including 64 bp of 5'-UTR, 1794 bp of ORF encoding 598 amino acids, and a 212 bp of 3'-UTR, with polyadenylation (^{2047}AATAAA ^{2052}) signal (Figure 53 A). The predicted molecular mass of *Rb-C8α* was 66 kDa and a theoretical isoelectric point was 5.7. According to the SignalP server prediction, the *Rb-C8α* was contained a signal peptide representing a cleavage site at 26-27 in the N-terminus. The mature protein contained a low-density lipoprotein receptor domain class A signature (LDLR-A), a epidermal growth factor-like domain (EGF-1), a thrombospondin type-1(TSP-1) repeat profiles, four potential *N*-glycosylation sites (Asn-X-Ser/Thr) and MACPF domain profile. Within the MACPF domain profile, MACPF domain and ^{348}FGTHYVTEGTMGG^{360} of MACPF signature was identified. Furthermore, sequence

^{512}CVCKSGYHGDAC^{523} of EGF-1 was also identified. *Rb-C8α* was contained only one

TSP-1 domain in the downstream of the EGF-1 signature.

A

```
ATG GAGACATTTATA AATGTATCCCTGACT TTATGCACACTGTAC CTGTTTATCAGTTTT   124
 M   E   T   F   I   N   V   S   L   T   L   C   T   L   Y   L   F   I   S   F    20
AGCCCAACAGTGAAC GCATCCAGGTGGCCA TGGACAACTGCTGAT AACAGCGGGGCTGGG   184
 S   P   T   V   N   A   S   R   W   P   W   T   T   A   D   N   S   G   A   G    40
ACAAGGAGGACCAGA GCTGTAAACAGGCCG ATTCCATCGACTGT AAGGTGGGGGGCTGG   244
 T   R   R   T   R   A   V   N   R   P   I   P   I   D   C   K   V   G   G   W    60
TCATCTTGGACTCCT TGCAACTCCTGCACA GATAAAAGTTCCGT TTTCACTACTTGGAG   304
 S   S   W   T   P   C   N   S   C   T   D   K   K   F   R   F   H   Y   L   E    80
AAGCCCTCACAGTTT GGTGGCACAGAGTGC CTTGAGACACTGTGG GAGAATCTGGCATGT   364
 K   P   S   Q   F   G   G   T   E   C   L   E   T   L   W   E   N   L   A   C   100
CCAACAGGAACCACA GAGTGTCTGGTGTCA GATTACTGTGGAGAG AGCTTCACCTGCAAT   424
 P   T   G   T   T   E   C   L   V   S   D   Y   C   G   E   S   F   T   C  (N)  120
GAAACAGGGCGCTGC ATCAGCCAGTCCCTT CGCTGTAATGGAGAG TCAGACTGTGATGAT   484
 E   T   G   R   C   I   S   Q   S   L   R   C   N   G   E   S   D   C   D   D   140
TCTTCTGATGAAGAC GGCTGCGTAGACTTC AACCTGAGAGAAGAC AAGTGTTCCACCCTA   544
 S   S   D   E   D   G   C   V   D   F   N   L   R   E   D   K   C   S   T   L   160
TTGTCAATCCCTGGC ACTGAACGTGGCACT CAGGGCTACAATGTG TTGACTAGCGAATTT   604
 L   S   I   P   G   T   E   R   G   T   Q   G   Y   N   V   L   T   S   E   F   180
GTGGATCATGTCCTG GACCCGAGGTACTTT GGAGGAAAGTGCGAA TATGTCTATAATGGT   664
 V   D   H   V   L   D   P   R   Y   F   G   G   K   C   E   Y   V   Y   N   G   200
GAATGGAGGAAATTC AACTATGACGCATTC TGTGAGAACCTGCAC TACAATGAAGATGAG   724
 E   W   R   K   F   N   Y   D   A   F   C   E   N   L   H   Y   N   E   D   E   220
AAGAACTACAGGAAA CCTTACAACTACCAT GCCTACCGTTTTGTG GCTGAAGCTACGTCA   784
 K   N   Y   R   K   P   Y   N   Y   H   A   Y   R   F   V   A   E   A   T   S   240
GAGGGCTCTCAGGAA TACTTTGACGATATA GTGAGCTGCTGAGA GCGAGGAAAGGTATG   844
 E   G   S   Q   E   Y   F   D   D   I   V   S   L   L   R   A   R   K   G   M   260
AGGTCTTCCAATGGC GGGGTCACATTTGGG ATCTATCATGTGGAA GTGGGACTGAGTGGA   904
 R   S   S   N   G   G   V   T   F   G   I   Y   H   V   E   V   G   L   S   G   280
AGTCAGGAATCTGAC TTTGTCCACAACATA ACACGGCATAAAAGC GAGGACCTAGGGTTC   964
 S   Q   E   S   D   F   V   H  (N)  I   T   R   H   K   S   E   D   L   G   F   300
GTTAGGCTTTCCTCC AAAGTGCAAACTGCC CACTTCAAAATGAGA AGTAATAAGTTGATG  1024
 V   R   L   S   S   K   V   Q   T   A   H   F   K   M   R   S   N   K   L   M   320
CTTCATGAAGATTTC TACATTTCACTCATG GAGCTCCCTGAGCAG TATGACTTTGGGATG  1084
 L   H   E   D   F   Y   I   S   L   M   E   L   P   E   Q   Y   D   F   G   M   340
TACTCTCGCTTCTTC AACACATTTGGCACC CACTATGTCACAGAG GGAACTATGGGAGGA  1144
 Y   S   R   F   F   N   T   F   G   T   H   Y   V   T   E   G   T   M   G   G   360
ACCCTTGAATATGTT GTGGTTGTCAACAAA ACATCTATGGCAGAA TCAAAGCTTGAGGGT  1204
 T   L   E   Y   V   V   V   V  (N)  K   T   S   M   A   E   S   K   L   E   G   380
GAACAAGCTGGTCGA TGTTTAGGTGCCTCT CTTGGTCTAAGCATT CCCATCGGCTTTGTA  1264
 E   Q   A   G   R   C   L   G   A   S   L   G   L   S   I   P   I   G   F   V   400
ACAGTAAATCCCAAA GTGGGGTTGGGTCCG TTTTGCGGTAAAGAA GGAGCGTTTGATGGA  1324
 T   V   N   P   K   V   G   L   G   P   F   C   G   K   E   G   A   F   D   G   420
GCCACGCACTCTAGT TCACGTGTGATTGAG GACATTATTACTCTG GTGAAGGGGGGAATC  1384
 A   T   H   S   S   R   V   I   E   D   I   I   T   L   V   K   G   G   I   440
ACACACAGCAGCAGT GGCCTGTTGGCTATC AGGAACCCTGACACC TACAGGAACTGGGGA  1444
 T   H   S   S   G   L   L   A   I   R   N   P   D   T   Y   R   N   W   G   460
GCGACTCTCAAATAC AACCCAACACTCATT GAATATGAGACCATG CCAATTTATGAGCTA  1504
 A   T   L   K   Y  (N)  P   T   L   I   E   Y   E   T   M   P   I   Y   E   L   480
GTGCGCCTCAGCACA CTGACCACATA GGTGCGAGACTTGCC AACTTGCAGAGGGGC  1564
 V   R   L   S   T   A   A   D   H   I   G   A   R   L   A   N   L   Q   R   G   500
TGGGATGAGTATCTG CAGCAGTTCGACTCC TGTCGCTGCGCCCCC TGCAGGCACAACGGG  1624
 W   D   E   Y   L   Q   Q   F   D   S   C   R   C   A   P   C   R   H   N   G   520
ATCCCTGCCCTCACG GGCACTTCCTGTTCA TGCGTCTGTAAGTCA GGCTACCATGGAGAT  1684
 I   P   A   L   T   G   T   S   C   S   C   V   C   K   S   G   Y   H   G   D   540
GCCTGTGAAAAGACT CTCAGAGGAGATACC AAGACAGACGGGGCG TGGTCGTGCTGGGGG  1744
 A   C   E   K   T   L   R   G   D   T   K   T   D   G   A   W   S   C   W   G   560
GCCTGGTCACCCTGT GCATCAGGGAGGAAG ACTCGGACAAGAGCC TGTAATAATCCTGCC  1804
 A   W   S   P   C   A   S   G   R   K   T   R   T   R   A   C   N   N   P   A   580
CCTGATGGGGGCGGA GTAACCTGCCTGGGC TCCTCCTCTCAGACT CAGCGCTGT TAG AGT  1864
 P   D   G   G   V   T   C   L   G   S   S   S   Q   T   Q   R   C               598
GACGTTTCCCTCAGA CTACGATGCTCTTGA TGTTTATTCAGAATA ATCTGTTTGTGTTTT  1924
TTTCAATTAATAAAA ACAGCTCTTTCACAA TCTTTTTGTTCATGT CAGAGCTCTCAAAAA  1984
TAAAATGAAATCTGT GTCCCTTATCAGGAC ACATTGTATGACATG CTGTATTTCAAAGTA  2044
TG AATAAA TGTGTGT TGTGCAGACTT                                            2070
```

Figure 53. Nucleotide and deduced amino acid sequences of Rb-C8α (A) and Rb-C8β (B)
Rb-C8γ (C)

The nucleotide sequence (upper) and the deduced amino acid sequence (lower) are numbered. Similar domains and motifs are found in *Rb-C8α* and *Rb-C8β* as follows; start codons, stop codons and polyadenylation signal sequences (AATAAA) are indicated in boxes. Predicted signal peptides are underlined, thrombospondin type-1 (TSP-1) repeat profiles are double underlined. Low-density lipoprotein receptor domain class A signatures are underlined and italicized. EGF-1 like domain signatures are dashed underlined. MACPF is bold and *N*-glycosylation sites are circled. MACPF domains are bold and shaded in gray. MACPF domain signatures are bold, underlined and shaded in gray. In the *Rb-C8γ*, lipocalin domain is highlighted in gray.

149

Figure 53. Continued.

C

```
                                                         TCTATTTGTT  -130
CTGTGTGTTTGGGGC TAAGTACAGCAGATT GGGAATAGTAGACTG CAGACAGAGGCAGAG  -120
GGTACAGAAAACCCA AAACAAAAGAAGAAA AGAGAAAAAAGACAA GTAAGAACACAGGAA   -61
CCTGGGCAATCCTAT CCGAGAAGAAGACTG ATTCTTCGCTGTATT ATTGTCAGTGCTAGG   -1
ATGGCTGGTGGATGG CGTTGCATGTTGGCA GTGGTGGCGCTGATG TGTGTGTGTCTGTCG   60
 M  A  G  G  W   R  C  M  L  A   V  V  A  L  M   C  V  C  L  S   20
GGGTCCACTGACGCT GTGGGGGGGGCTAGG AGTCGGCCGCGACCC CAGAGACGACCTAAG  120
 G  S  T  D  A   V  G  G  A  R   S  R  P  R  P   Q  R  R  P  K   40
AAGCCAAAGGTTGAA CCTGTAGACGTGACC CCACCCGCACAGAAC ATAGACATACAGCGG  180
 K  P  K  V  E   P  V  D  V  T   P  P  A  Q  N   I  D  I  Q  R   60
ATGACAGGAATATGG TACCTGCTGAACACT GCCTCCAAATGCTCT TACCTGATAAATCAT  240
 M  T  G  I  W   Y  L  L  N  T   A  S  K  C  S   Y  L  I  N  H   80
GGAACCCAAGTGGAG CCCACAGTCATGAAC CTCACACGTTCCTCT GCCTCTGACCAAACA  300
 G  T  Q  V  E   P  T  V  M  N   L  T  R  S  S   A  S  D  Q  T  100
CTGTCTGTTAGCACT AAAACTCGACATAAT CACCAGTGTTGGGAG ATATTACAGGTCTAC  360
 L  S  V  S  T   K  T  R  H  N   H  Q  C  W  E   I  L  Q  V  Y  120
CATCTAACACCAACC CCAGGCCATCTAACA CTTAAAGGAGCTCGT CCTGAGCAGAACATT  420
 H  L  T  P  T   P  G  H  L  T   L  K  G  A  R   P  E  Q  N  I  140
GACATAGTGATCGGG GATACGGACTACGAC TCCTATGCGATCATG TACTACCAGAAACGT  480
 D  I  V  I  G   D  T  D  Y  D   S  Y  A  I  M   Y  Y  Q  K  R  160
GGCAAGATCACCCTG AAACTCTTGAGCAGG TCTGTAGACGATCTG TCAGAGCCAATGTTG  540
 G  K  I  T  L   K  L  L  S  R   S  V  D  D  L   S  E  P  M  L  180
ACCAAGTTTGAGCAG GTTGCTGAAAAACAA AATTTGGGACTGGCA TACCTCTTCCCCTTC  600
 T  K  F  E  Q   V  A  E  K  Q   N  L  G  L  A   Y  L  F  P  F  200
CCCACCTACAGTCAC TGTGGTGCTGTGGAC CAGGACCATGTAATC AACTGTGTCCCCACA  660
 P  T  Y   S  H   C  G  A  V  D   Q  D  H  V  I   N  C  V  P  T  220
TGCTGAAGAGGCAGA GCTGCGCAACAGACA GATCCAGCCTCGCAG TGTTGGGCAGTTCTC  720
 C                                                              221
ACACGAATATCTGAA TATTGTATCATTTTT ATATGTGGATATTAGC AAGAAAGTTGTTGA  780
CTTCAAATTGTAACA CTGTCAACTATAAAT GTTACCAGTTACGAAT AAATAAAACTAAAT  840
GGAGC                                                           844
```

Figure 53. Continued.

The complete cDNA sequence of *Rb-C8β* was composed of 1952 bp, including 27 bp of 5'- UTR, 1761 bp of ORF encoding 587 amino acids and a 164 bp 3'-UTR, including polyadenylation ([1920]AATAAA [1925]) signal (Figure 53 B). The predicted molecular mass of *Rb-C8β* was 66 kDa and theoretical isoelectric point was 7.2. According to the SignalP server prediction, the *Rb-C8β* was contained a signal peptide representing a cleavage site at 34-35 in the N-terminus. The mature protein contained two TSP-1 repeat profiles, LDLR-A domain, EGF-1 domain and MACPF domain profile. Within the MACPF domain profile, MACPF domain and [335]YGTHYITEAGLGG[347] of MACPF signature were identified. Further,

150

sequence [518]CVCPVGYTGQGC[529] of EGF-1 was also identified. Two TSP-1 profiles were located either side of the MACPF domain profile.

The complete CDS of *Rb-C8γ* was composed of 666 bp and partial 5' flanking region was consisted of 130 bp and complete 3' UTR was contained 186 bp including polyadenylation ([827]AATAAA[832]) signal (Figure 53 C). The predicted molecular mass of *Rb-C8γ* was 25 kDa and a theoretical isoelectric point was 8.6. According to the SignalP server prediction, the *Rb-C8γ* was contained a signal peptide representing a cleavage site at 25-26 in the N-terminus. The mature protein contained a single domain called lipocalin.

4.1.2 Genomic DNA organizations and promoter regions of Rb-C8α, Rb-C8β and Rb-C8γ

The *Rb-C8α* genomic sequence (JQ710660) was about 8 kb, consisted of 11 exons interrupted by 10 introns. The 5' and 3' ends of all introns were carried canonical splicing motifs (GT-AG rule). The number of exons and introns were well-conserved in most of species considered for comparison. Further, the length of exons and introns were variable in length, in species used for comparison. Moreover, within the fish the variability was minimal and lengths of middle exons (4-10) were conserved. Interestingly, medaka fish contains extra small exon (5 bp) which was slightly deviated from other species (Figure 54 A).
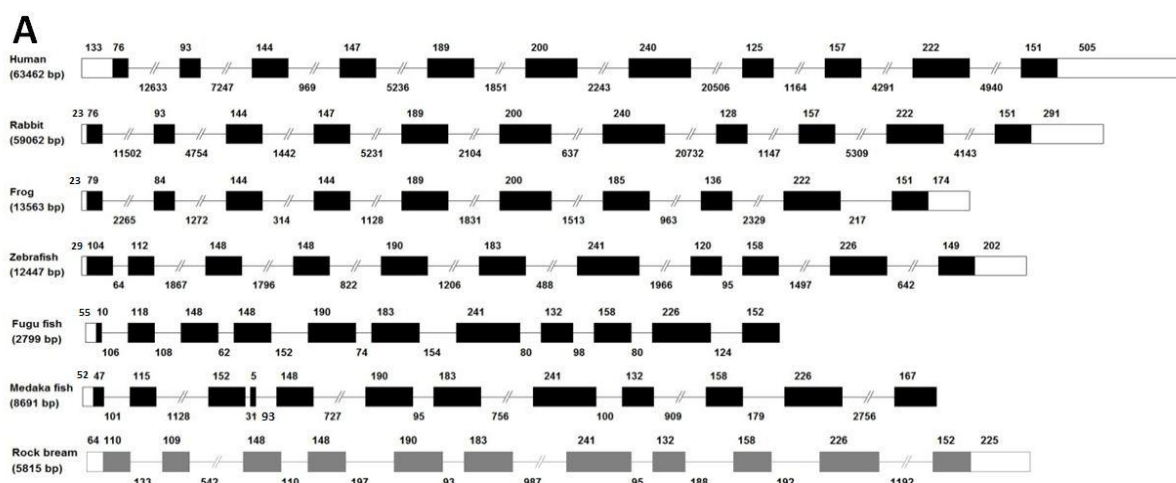
Figure 54. Genomic structure of Rb-C8α with other selected 8α genome structures (A) and Rb-C8β with other selected 8β genome structures (B) Rb-C8γ with other selected 8γ genome structures (C)

Exon represented in back and gray solid boxes, un-translated regions are shown in light-colored boxes. Intron are represented as thin line i-between exons. Introns lager than 250 bp are indicated with interrupted lines. Numbers above the boxes and below the lines indicated the lengths of exons and introns respectively. Selected organisms for **A**: *Homo sapiens* (ENST00000361249), *Oryctolagus cuniculus* (NM_001082255), *Danio rerio* (ENSDART00000147283),*Takifugu rubripes* (ENSTRUT00000013219) *Oryzias latipes* (ENSORLT00000012303) and *Xenopus (Silurana) tropicalis* (NP_001005445) and for **B**: *Homo sapiens* (ENST00000371237), *Mus musculus* (ENSMUST00000031663), *Bos taurus* (ENSBTAT00000033407), *Danio rerio* (ENSDART00000057736), *Takifugu rubripes* (ENSTRUT00000015084), *Gasterosteus aculeatus* (ENSGACT00000021601) and for **C**: *Homo sapiens* (AAI13627.1), *Mus musculus* (NP_001258706.1), *Bos taurus* (NP_001103546.1), *Danio rerio* (NP_957157.1), *Takifugu rubripes* (XP_003967902.1), *Oreochromis niloticus* (XP_003443552.1)
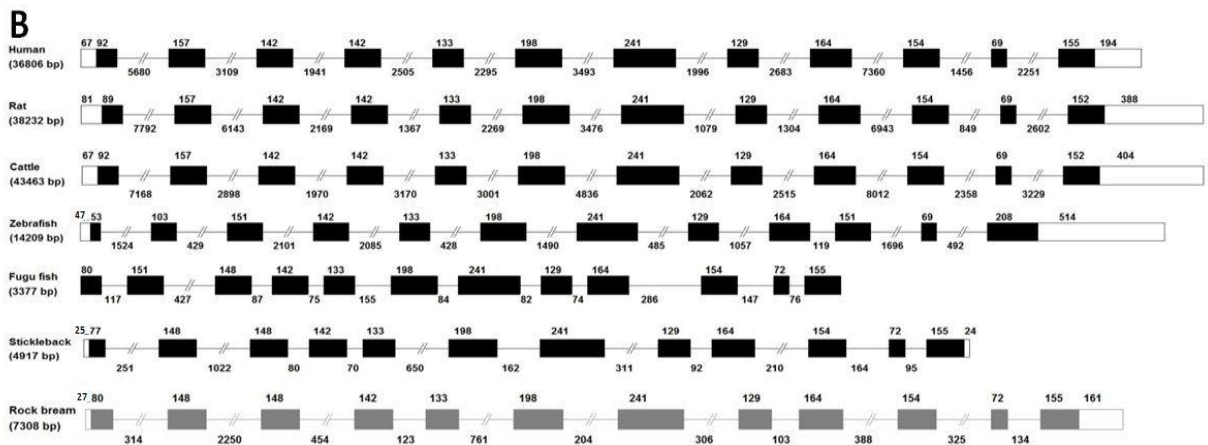
Figure 54. Continued.
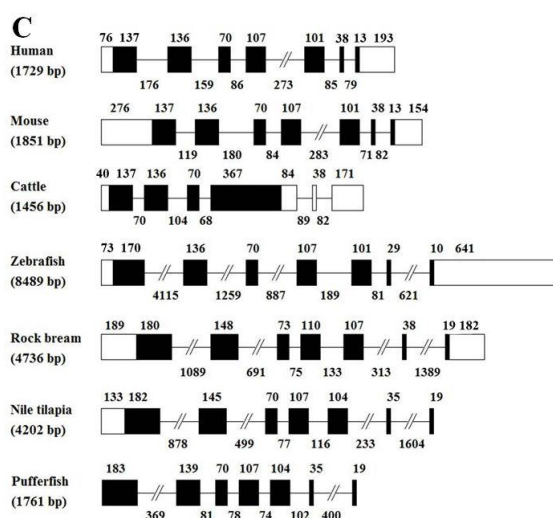
Figure 54. Continued.

The *Rb-C8β* genomic sequence (JQ710661) was about 10 kb, consisted of 12 exons interrupted by 11 introns. The 5' and 3' ends of all introns were showed canonical splicing motifs (GT-AG rule). The number of exons and introns were well conserved in all of the species considered for comparison, with introns of varying lengths. Moreover, lengths of middle exons (4 -9) were conserved in all considered species (Figure 54 B).

The *Rb-C8γ* genomic sequence (KF682443) was about 4736 bp, consisted of 7 exons interrupted by 6 introns. The 5' and 3' ends of all introns were showed canonical splicing motifs (GT-AG rule). The number of exons and introns were well conserved in all of the species considered for comparison, with introns of varying lengths (Figure 54 C) In the BAC screening, 5' region of Rb-C8γ was unidentified as the sequence end up with BAC vector sequence.

Typically, promoter region has a complex structure consisting of multiple functional binding sites for proteins involved in the transcription initiation process. *Rb-C8α* has 64 bp of 5′-UTR in accordance with complete cDNA. Interestingly, we could identify transcription initiation site at 64 bp upstream of the start cordon (ATG) and no TATA box was found in the upstream of the 5' flanking region of genomic sequence. *Rb-C8β* has 17 bp 5'-UTR in accordance with complete cDNA. Prediction analysis of *Rb-C8β* promoter region was revealed the presence of a transcription initiation site (+1 site) at 27 bp upstream of the start cordon (ATG). TATA (CATAAAGA) box was also been identified at -23 to -30 upstream from the initiation site (+1). *Rb-C8α* and *Rb-C8β* contain almost identical immune related sites such as CCAAT-enhancer binding protein (C/EBP), nuclear factor kappa-B (NF-κB), nuclear factor-1 (NF-1), activator protein-1 (AP-1), interferon consensus sequence binding protein (ICSBP), octamer transcription factor (OCT-1), specificity protein-1 (SP-1), YY-1, GATA-1,and HNF (Figure 55 A and B).

A

```
                                                          ATTATTATATTACTGGTTTCTTT  -1137
AGTCCTCTATGACAGTAAACTGAATATCTTTGGGTTGTGGACAAAACAAGACATTTGAGGATGTCACCTTGGGCT  -1452
                            CP-2                         NF-κB
TTGGGAAACAGCGATCAACATTTTTCACCATTTTATGACATTTTATGGACCAAACAACTAATCGATTAATCACGA  -1377
                        YY-1
AAATGATCAGCAGATTAATCAATAATGAAAATAATCGTTAGTTGCAGCCCTAATTACTAGAAGTGTTTCTTACTT  -1302
                        OCT-1                         OCT-1          ICSBP
TCACAGACAATTTACTTAAGACTGTTAATTTAGGGGGCCAAAACCCAATCCTGACAAAATTACAAAAACAATATT  -1227
            C/EBPγ                                                  SOX-5
TCCTGTCTCATTTATCTGGTTCCAACTTCTCAAATGTGAATATTTGCTGGTTTCCTCAGTCCTCCATGATTGTAA  -1152
            GATA-1
ACTGAATATGTTTGGACTGTTGGGTTTGACGAAACAGGACATCTGAAGGCATTCTTCAGTTTTCTGACATTTTAT  -1077
AAACCAAACGATTAACTGATTAACCAAGAAAATATCAAACAGATTAATCAATAATGAAAATAAAATTGTTTATCC  -1002
ACAGAATACGTGGTGAACTGTTGATGTTGAGGTCATAACAGGTACACTGAAAAGACGATGCTGCTGAATTTTTTA   -927
ATATAATTTCTCAATGCTGAGAGCACCACAAAAGAAATTCCATTTGCAGCCATCGTGTTGAGGTGGAAGAGGATA   -852
                            NF-kappa     NF-1                    GATA-2
TCTCAAGAGATGGATATCTGAAATCATGAGCAAATAAAACCACAATGTGAATGAGAGGATGTCTTGTAATTTGGG   -777
            NF-κB                                                       YY-1
TGGATCAACCCTTTAATATACTATAGTAATTAGAAATGTGATAACTTAGCTAAGCCATGAACTTGTATGGTATAA   -702
        OCT-1                 C/EBPα              AP-1
ACTAGTGAAGCAAACTGCATTAAAACGCAGTGTATTAAACATTGGGTACCAGCTAGCACAAATGTAGCTATCTAA   -627
        C/EBPα                                                            YY-1
ATGTAGCTTAGCTTTCTTCTAACGGTCCTCAAGCTAAAGAACAAATGTTTAAAAGTTACCTCTTCCATTGTAGCT   -552
CGTTTTGTTCATCACTGTTCAGGTTTGGACATTTGGAAAGCGGTCTAGACTGAAATGAAGTGGACGCTGTCGGCT   -477
                        SRF                             Sp-1
GAGTTGACAGATTTAGCGAGACGAATCCAACAAACTGTTTATGCTAAAACAAACAATTTGAAAGCGCCGCCGTGA   -402
                                            HNF-3              Sp-1
GACGCACCTGCGCCCTGCGCGTGACCTGATAACGTGCACGTTCATGCTGTTTGATTAGCAAATCAGAAACGTCGG   -327
            Sp-1                                        NF-κB
AAACCACTTTATATGTGGATAGGTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA   -252
ATAATCACTGGAATACATATATTGTAAGAAGAAAAAGGTTTACCTTGTAGTTTGAGGTGtTTGAAGTACTTTTAC   -177
                TBP         GATA-1
AGTTATAGTGAAAGTTTTTGTGTATATTGGGACAAAGTGCTCTGGCTGGACTGGAGTGTTTGCGGGGGTAAAGAC   -102
                                                        Sp-1
GAGAAAGATCATTGAAAAACTTTCCCAACTGTTGCAAGAGCATGTGTTTACGCTTGATCAATCATTCACTGTGGT    -27
ICSBP                       SRF      C/EBPα
                                        ⌐→
TGGTCAAGACTGCAGTGGCAAAAGTCACAACACAACAGCTGTTTTGTGATGTGGCATTTTAACCTTAAGGATTAG     49
            C/EBPα         +1 site
CTGCTCGTTTTCAGTATGGAGACATTTATAAATGTATCCCTGACTTTATGCACACTGTACCTGTTTATCAGTTTT    124
            M   E   T   F   I   N   V   S   L   T   L   C   T   L   Y   L   F   I   S   F     20
```

Figure 55. 5' flanking region of Rb-C8α (A) and Rb-C8β (B)
Putative transcription factor binding sites are shown in bold and underlined with corresponding factor identified below. The putative TATA box is boxed and identified below. The transcription initiation site is indicated by a bent arrow. Translated amino acids (uppercase, in bold) and start codons (italicized uppercase, in bold) are shown.

B

```
ACACTAATACAAATCCTCTATGCAGTCGATCAATGTCATGTGACTATGTGATATATTAGTATAATCACCACTTGT   -1414
                                   AP-1
TCTTTTTCATGCTGATTATAGTATTTAATCACATGAAGAGTAGCACAGTGTCATTCAAAAGTAAAAGTACAGAAG   -1339
        Oct-1
TATTCAAAGCAAAATGCACTTAAAGTATCAAAAGTAGAAGAAGTCATTATGCAAAGTGGAGCTAGATTTAACCAC   -1264
   Oct-1
TTTTTAGATACTTGTTGTTTATTTCAGTGGTTCCCAACCTACTGTAGGAGTTGGGCCCCCTTCCAAAGGGTCACA   -1189
                                                    SP-1
AGATAACTCTGGGGGGTCATGAGATGATTAATGGGGCAGAAAAGGGGaAAAAGGTTCTGCTCCACATATTTGTAT   -1114
CCATTTTTTGGACTTTTCTACAGTCTTTGCTTTTGTTGTGAAATAATGGAAAATTTGACCCCTTTGAGCCTCAAA   -1039
CGGTTATTTTAATGAAACCATCTGAGAAGTTTAAAGGGGAAATCACTTTTGGTGGAACTGCTAAAACTCAGCCTT   -964
                                       NF-κB
TTGTAAAGAGCCACAGGCCAGAAAGGTTGAGAACCACTGGTTTCAGTTTTAGCAATGCTTTATATCTTATAAGCT   -889
C/EBPα          NF-1                           ICSBP
TATCATATGGTTATGTTTATATAAAATCTAAAAGTAAGAAAAGTAACTCCAGCTGGCAAATAAAGATAGTGGAGT   -814
GATA-1
AAAATGTACATGTACGAGTAAAATACAAATATTCATGTCGAGTACAAGTACCTCAAATAAATACAGCACATGTAT   -739
                 Oct-1
TTAATTGTAATTGTACTTAAATACAGTACTGTAGTAATTGTACTTAGTTACTTTGCACCACTGCTGTCCTTACAA   -664
GCATGGATCTACCAATCGCATCCCATTGCATATACAGCAAGGACAACATATCTTAATATAGTGTATTACATTATT   -589
                      C/EBPβ
CGTAATATCAGTAATAGCAGCATTAGTACCAGTTATTAGATATTAGTATAAGTATTATGTGGCTTTTCACTGTGT   -514
                              HNF-1
TAGTAAACTACAGCCGAGTCATGTGATTGTGTCACTCCGTAAAATCCAAAATGAGTGAAAGTAAAGTTTAAAGGT   -439
                         C/EBPβ
CGTAAGTACACAAACATGACTCAATGTTCGACCACAGAAGTCTGTAACAGGAAGTCACTTTATTACAAAGGTTTA   -364
                 Ap-1
ATTTTCCACCGCTGTTGTATGTTATCTAAAGTAGGTTTTCTAGCCATTTGAGATTCATCACATCACTAAAGATTA   -289
                               NF-1
TACATAATAAAGTTTATGATAACTGAAAAGACATCCACACCTCCTGGGCTGCCTTTTAAACGTGCCTGCAGATTC   -214
                        SP-1
CAAACAGCTTCATTATAAACCTGAGGTCTATTTCATACCTGTGTTAAATACACACTGACACTAAGACCCAATACC   -139
ACAACACCTAGTTTTGTGGCTCTGCATTGTTGTGTATTCAGCTCCACATCTGACTCTAATGGCAGCAGTGAGTTC   -64
                                               C/EBPα          NF-1
```

```
ATATCCAGGCCAGTATTGATCAGTGTGTGTTTACATAAAGAGCTGAGCAGTGTCGTCTTACTTCTTCACAGTCAG    12
    Sp-1   C/EBPα              TATA box                          +1 site
TCATAACACACAACCATGCTTTCTGTGGTGATACGCATGAGTGCGCTCAGTTTGCAGTGCTGTCTTCTCCATGTT    87
            M  L  S  V  V  I  R  M  S  A  L  S  L  Q  C  C  L  L  H  V            20
```

Figure 55. Continued.

4.1.3 Phylogenic studies of Rb-C8α , Rb-C8β, and Rb-C8γ

According to the pair-wise alignment results, *Rb-C8α* has 62% amino acid identity with rainbow trout (*Oncorhynchus mykiss*) and 55% with ayu fish (*Plecoglossus altivelis*). However, it shares 33.7 % identity with its homologue *Rb-C8β* and more than 40 % identity

155

is shared with other higher vertebrates (Table 16). *Rb-C8β* has highest amino acid identity (83 %) to Japanese flounder (*Paralichthys olivaceus*) and 67 % with rainbow trout (*Oncorhynchus mykiss*). *Rb-C8γ* has highest amino acid identity (77.5 %) to both European sea bass (*Dicentrarchus labrax:* CBN81860.1) and Fugu (*Takifugu rubripes* XP_003967902.1). Multiple sequence alignments of *Rb-C8α*, *Rb-C8β* and *Rb-C8γ* were indicated that recognized domain in each gene has higher similarity with their respective orthologous. The cysteine residues and important signature motif were highly conserved among all the examined organisms (Figure 56 A and B).

A

```
O. fasciatus   -------------METFINVSLTLCTLYLFISFSPTVNASRWPWTTADNSG--AGTRRTR  45
O. mykiss      MSSFSSRSSFISNMNRLICVLLLGSYILLLVLNKSPTVDATEYSWNMAETRTGQSAIRRVR 60
P. altivelis   -------MIIIFNMANKVICGLLGSYILLSLNRSFFVESARNNWNSTKSRG---SARVFR  50
H. sapiens     ------------MFAVVFFILSLMTCQPGVTAQEKVN----------------QRVR     29
R. norvegicus  ------------MFTVAFFGLSLMAWHSGVTAQEKVN----------------QRVT     29
                           *         *        :. :     *:               :

O. fasciatus   AVNRPIPIDCKVGGWSSWTPCNSCTDKKFRFHYLEKPSQFGGTECLETLWENLACPTGTT 105
O. mykiss      SVNKPAPINCKMKMWSSWSPCDSCTDKKFRFRYMEKASQFGGRQCLDSQWEELACPTAQA 120
P. altivelis   DVSRPTPIDCVMQKWSSWSSCNSCTVKKHRFRHLERPSQFGGTNCLEAQWETLNCPETKE 110
H. sapiens     RAATPAAVTCQLSNWSEWTDCFPCQDKKYRHRSLLQPNKFGGTICSGDIWDQASCSS-ST  88
R. norvegicus  RAVTPGAVACQLSDWSEWTDCFPCQNEKYRYRSLLQPSKFGGTICSGDIWDKANCDS-PT  88
                .   *  :: *  : **.*: *  .*     :*.*.:  : :.:*** *  *:      *

                   ← ═══════════════ LDLR-A
O. fasciatus   ECLVSDYCGESFTCNETGRCISQSLRCNGESDCDDSSDEDGCVDFNLREDKCSTLLSIPG 165
O. mykiss      VCWEPDICGERFTCNATGRCVSQALRCNGEVDCDDESDETDCEQVDDRQDKCSTLLPIPG 180
P. altivelis   KCLVADFCEKGFGCSESDRCISKDLLCNGEDDCDDWSDEADCGEVKIREDKCSTLIPIPG 170
H. sapiens     TCVRQAQCGQDFQCKETGRCLKRHLVCNGDQDCLDGSDEDDCEDVRAIDEDCSQYEPIPG 148
R. norvegicus  PCLRQAQCGQDFQCKRETGRCLKRHLVCNGDQDCLDGSDENCDDARVIEDDCRQYEPIPG 148
                *          *  :  * *  :.**:.:  * ***:  ** * ***  .*  :     ::.*    .***

O. fasciatus   TERGTQGYNVLTSEFVDHVLDPRYFGGKCEYVVYNGEWRKFNYDAFCENLHYNEDEKNYRK 225
O. mykiss      AGRGTQGFNILTGEFVDHVLDPQYYGGQCEYVVYNGEWRKLIYDPFCENLHYNEDEKNYRK 240
P. altivelis   AEHGSQGYNILRGDFVEKVLDPGYFGGLCEYVVYNGEWRKFKYESFCENLSYNEDEKNYRK 230
H. sapiens     SQKAALGYNILTQEDAQSVYDASYYGGQCETVYYNGEWRELRYDSTCERLYYGDDEKYFRK 208
R. norvegicus  SERAALGYNILTQEEGQSVYDAKYYGGQCETVYYNGDWRRLQYDPTCERLYYGEDEKYFRK 208
                :  :.:  *:*:*   :   : * * . *:** ** ****:**.: *:. **.* *.:*** :*

                                                              ═════════════ ←
O. fasciatus   PYNYHAYRFVAEATSEGSQEYFDDIVSLLRARKGMRSSNGGVTFGIY------HVEVGLS 279
O. mykiss      PYNFHTYRFMAQATSEGSSEYYEDMATLLKARKTEDSFNLGVTVGIR------YVEFGVS 294
P. altivelis   PYNFHSYRFLAQATSEGSNEYYEDMAQLIKAMKTDDSGHVGVTVGVQ------YVEVGLS 284
H. sapiens     PYNFLKYHFEALADTGISSEFYDNANDLLSKVKKDKSDSFGVTICIGPAGSPLLVGVGVS 268
R. norvegicus  PYNFLKYHFEALADTLISSEFYDDANDLFSKIQRDKSQSNSVTFGISPAKSPITLDASVS 268
                ***:  *:* *  *     *.*:::  *:    * .   * .**.* :        .:*

                ─────────────────── MACPF Domain ───────────────────
O. fasciatus   GSQESDFVHNITRHKSEDLGFVRLSSKVQTAHFKMRSNKLMLHEDFYISLMELP-EQYDF 338
O. mykiss      GNVESALLTNLTKYTNQELCGFIRLQSKVQTAQFKMRSEGLMLHEDMYLSLMELPPEEKYDF 354
P. altivelis   ASLGVKMLKNITKYSSKDVGFVRLFSRVQTARFKIRSSGLMLSTRMLSALLDLP-TVYDF 343
H. sapiens     HSQDTSFLNELNKYNEKKFIFTRIFTKVQTAHFKMRKDDIMLDEGMLQSLMELP-DQYNY 327
R. norvegicus  WSDESSFMKELSKYNEKKYSFMRVSTKVQTAHFKMRRHNIVLDEGMMESLMELP-EQFNY 327
                   .     :: ::.:.:.:.  * *: ::*****:**:*   ::*    :  :*::** :::

                ─────────────────── MACPF Domain ───────────────────
                       ⌐──── MACPF signature ────⌐
O. fasciatus   GMYSRFFNTFGTHYVTEGTMGGTLEYVVVVNKTSMAESKLEGEQAGRCLGASLGLSIPIG 398
O. mykiss      GLYSRFLNTYGTHYVTQGIMGGTLEYVAVVNTTAMKTSKIDAEQLKGCLGGSIGISSPIG 414
P. altivelis   GQYSQFLNTYGTHYVTQGTMGGTLEYVAVIDKVAMEKIEKNGFHIAGCLGASFGLAGPLG 403
H. sapiens     GMYAKFINDYGTHYITSGSMGGIYEYILVIDKAKMESLGITSRDITTCFGGSLGIQYEDK 387
R. norvegicus  GMYSKFINDYGTHYITSGTMGGIYEYVLVLDKEKMRIHGITVEDVKKCIGGGVGLQFGEK 387
                * *:*:* :*****:*.* *** **: *:.  *     *      *:*.:..*:

                ─────────────────── MACPF Domain ───────────────────
O. fasciatus   ---FVTVNPKVGLG-PFCGKEGAFDGATHSSSRVIEDIITLVKGGITHSSSGLLAIRNPD 454
O. mykiss      KTKQVEVGGKLEVK-GCKG-TGSYEKEMYGSSSLIKDIVTLVKGGSTGGSGGLLAIKDPD 472
P. altivelis   ----QNVEGELKVRGKVCPSGGTFNSGRKTESSFIEDIFTVVKGGTIETTGSLLSIQDPE 459
H. sapiens     ------INVGGGLSGDHCKKFGGGKTERARKAMAVEDIISRVRGGSSGWSGGLAQNRSTI 441
R. norvegicus  ------IIGEGDLSGESCVMTGDGNQDKRKKDLAVEDIISRVQGGSSRWSTGLAHNSSAI 441
                       :      :        *         ::**.: *:** :   : .*    ..

                ─────────────────── MACPF Domain ──────────────── →
O. fasciatus   TYRNWGATLKYNPTLIEYETMPIYELVRLSTAADHIGARLANLQRGWDEYLQQFDSCRCA 514
O. mykiss      TYRKWGLSLKYNPNLIEFETLPIFELVRLSTAGDHVGARPAHLRRAWEEYLLQFNSCRCA 532
P. altivelis   TYRRWGETLKYNPALIDFEMLPIHELVRLSTAADQAGVRIAHLRSALDAYLLEFNSCRCA 519
H. sapiens     TYRSWGRSLKYNPVVIDFEMQPIHEVLRHTSLG-PLEAKRQNLRRALDQYLMEFNACRCG 500
R. norvegicus  TYRSWGRSLKYNPVVIDFEMQPIYQLLQHTNLG-PLETKRRNLRRALDQYLMEFNACRCG 500
                *** **:***** :*::* **.:::: :. .   .:  :*: . : ** :*::***.

                     EGF-1
                ─────                                 ←─────────────────
O. fasciatus   PCRHNGIPALTGTSCSCVCKSGYHGDACEKT---LRGDTKTDCAWSCWGAWSPCASCRKT 571
O. mykiss      PCRHDGIPVLSQTSCHCICKQGFRGEACEET---LRKDSTTDGAWSCWGAWSSCQSGSKT 589
P. altivelis   PCKYNGIPVFIQSKCSCLCKQGYRGDACEET---LRKNSVTDGSWSCWSSWSSCQSGKMS 576
H. sapiens     PCFNNGVPILEGTSCRCQCRLGSLGAACEQT---QTEGAKADGSWSCWSSWSVCRACIQE 557
R. norvegicus  PCFNNGVPILEGTSCSCQCSMGRQGPACEQT---SMVLEAMEGAKADGRWSCWSSWSECRGGSRE 560
                **  :*:*  :  :.* *  *   *  ***.  .: :** ****.:** *.*

                ─────── TSP-1
                ─────
O. fasciatus   RTRACNNPAPDGGVTCLGSSSQTQRC 598
O. mykiss      RRRSCDNPQPDGG--AACLGSSSQNQRC 615
P. altivelis   RRRTCDNPPPDKG--AACLGSSSQSRNC 602
H. sapiens     RRRECDNPAPQNGGASCPGRKVQTQAC 584
R. norvegicus  RRRQCNNPAPQNGGAPCLGKSLQTQAC 587
                * *  *:** *: * ..* *  . *.: *
```

Figure 56. Multiple sequence alignment of Rb-C8α (A) and Rb-C8β (B) in comparison with C8 sequences of other vertebrates using ClustalW (v2.1) program

The identical residues are indicated black shaded with asterisk (*). Conserved and semi-conserved residues are indicated by gray shading, semicolon (:) and dots (.) respectively. Missing amino acid is marked by dashes; the conserved domain is indicated by double arrows. The GenBank accession numbers of these genes are shown in Table 16.

157

B

```
O. fasciatus    MLSVVIRMSALSLQCCLLHVTLS-LLLSEVA------VTTASNEEPSVREARSVGDQVVV  53
P. olivaceus    MFRVAIPRSALNLHSCLLHVTLSLVLISKAA------ITTAGNEDSDVREARSVSDQQVV  54
O. mykiss       --MNHKLKPTVGLGYCLLCAALCLLLLRDVA------IAGSGEEPSGVREARSVGTQVAV  52
H. sapiens      MKNSRTWAWRAPVELFLLCAALGCLSLPGSRGERPHSFGSNAVNKSFAKSRQMRSVDVTL  60
R. norvegicus   MKTG-AQVWRALAKSCLLCAALGCLHLPGARGEKPDFFETNAVNGSLVRSRPVRSVDVTP  59
                                **  .:*   :   :    . . . ..: . :. . :
                        <------------------- TSP-1 ------------------->
O. fasciatus    HPVDCVLSEWTAWSRCDTCQKKRYRYAKLDQPSQFGGEPCNFHGREEEACIVSSRYTC--  111
P. olivaceus    HPVDCVISDWSAWSRCDTCQKKRYRYAKLDQPSQFGGEPCHFHDMEDEACDVPDRYTC--  112
O. mykiss       QPVDCVQSEWSSWTRCDVCRKKRYRYAKLVQPSQFGGEPCHVQGKEVEPCSPPSRYDCTH  112
H. sapiens      MPIDCELSSWSSWTTCDPCQKKRYRYAYLLQPSQFHGEPCNFSDKEVEDCVTN--RPCG-  117
R. norvegicus   APTDCQLSTWSSWTACDPCQKKRYRHTYLLRPSQFYGELCDFSDKEVEDCVTN--RACR-  116
                  *  **  *  *:::*: ** *:*****::  *  :**** **  *.. . * * *   *
                    <------------------- LDLR-A ------------------>
O. fasciatus    DNVPLCEGFLCTQTGRCIHRTLQCNGEDDCGDMSDEVGCKTASKPCRETVEEYWGIENLA  171
P. olivaceus    DSIPLCEGFLCTQTGRCIHRTLQCNGEDDCGDMSDEVGCKKVPKPCRQEAEEYWGIENLA  172
O. mykiss       DETPLCEGFLCTYTGRCVPIDIRCNGDDDCGDWSAEKGSPKVPKACKQEAQEYHGIENLA  172
H. sapiens      -SQVRCEGFVCAQTGRCVNRRLLCNGDNDCGDQSDEANCRRIYKKCQHEMDQYWGIGSLA  176
R. norvegicus   -SQVRCEGFVCAQTGRCVNRRLLCNGDNDCGDQSDEANCRRIYKKCSQDMEQYWAIGNLA  175
                 .    ****:*: ****:     * ***::**** * * .. * * . ::* .* .* **
O. fasciatus    KGININLNSNLEGVVLDNRYYAGSCLPHYIQNVRFRKPYNLQQYTLQTKGSYDFTAQSFET  231
P. olivaceus    KGININLNSNLEGLVLDNRYYAGSCLPQYIQDVRFRKPHNLQQYTLETKGSYDFNVQSFES  232
O. mykiss       KGINILHSHLEGSVIDNRYYAGSCLPHYIQDVRFRKPYNLQQYTLETKGTYDFKLQSFES  232
H. sapiens      SGINLFTNSFEGPVLDHRYYAGGCSPHYILNTRFRKPYNVESYTPQTQGKYEFILKEYES  236
R. norvegicus   SGINLFTNTFEGPVLDHRYYAGACSPHYILNTNFRKPYNVESYTPQTQGKYEFALTEYES  235
                .***:: . :** *:*:***** .* *:** ::.*****:*::.** :*:*.*:* *::
                                         <-------------- MACPF Domain -------------
O. fasciatus    YSDYMDHTAKERATQTVVSIGFGIPGLFEFGFNFNNAKNTRSVKKIRRASSKTNSFVRAK  291
P. olivaceus    YSDYMDYSMRERMTQTIVSIGFAIPGIAEFGFNYNNAKVTRSIQKIRRASSKINSFVSAK  292
O. mykiss       YSEFVHYTMTERSSKTTVSIGFALPGVAEFGFNYADSKYSKEKKIRRASRKENSFVQAK  292
H. sapiens      YSDFERNVTEKMASKSGFSFGFKIPGIFELGISSQSDRGKHYIRRTKRFSHTKSVFLHAR  296
R. norvegicus   YFDFEHNVTEKATSKSSFKFGEKLDGLVEFGVRKESNEGRHYISRTKRFSHTKSKFLHAR  295
                * ::         :   ::: ..:** :*: **.      :. :      .* *: *:
                              <--------------- MACPF Domain ---------------
                                                  <-- MACPF signature -->
O. fasciatus    AELELAQYMLKSDDLMLHPEFLQRLRSLPQSYVYGEYRQIYRDYGTHYITEAGLGGQYEH  351
P. olivaceus    AELELAQYMLRSDDLMLHPEFLQRLRSLPQAYVYGEYRQIYRDYGTHYITEAALGGEYEH  352
O. mykiss       AELQLARYILKSEDLMLHPEFFIRLRALPQSYNYGEYRQIYRDYGTHYITEATLGGDYEY  352
H. sapiens      SDLEVAHYKLKPRSLMLHYEFLQRVKRLPLEYSYGEYRDLFRDFGTHYITEAVLGGIYEY  356
R. norvegicus   SVLEVAHYKLKSRQLMLHYEFLQRVKSLPLEYSYGEYRDLRDFGTHFITEAVLGGIYEY  355
                : *::*:* *:. ****.*** **: *: .** *. * ***** ** *.*:*:*** *** **
                         <-------------- MACPF Domain --------------
O. fasciatus    TIILNKERLARTDYTLDDYKRCTQAGLKVQANIQGVYLSLGAHGGSCNGLLNEMGEDTER  411
P. olivaceus    TIIILDKEKLAKTDYSLEDYKSCTQAGLKIGANIYGVYVSAGIEGGSCNGLLNEMGEDTAI  412
O. mykiss       TVILDKEKLEKTGYSLEAYKNCEQIVLKVGANIKGVYVTVGLEGGGCDGLLNEMGEDTVK  412
H. sapiens      TLVMNKEAMERGDYTLNNVHACAKNDFKIGGAIEEVYVSLGVSVGKCRGILNEIKDRNKR  416
R. norvegicus   TLIMNKDAMERGDYTLDHVSACAGGGFQIGGNVYKVYLKLGVSEKKCSDILNEIKDRNKR  415
                *::::*:  :  . *.*:        *        :::*. :  **:.  *    .:***: .:
                                  <------------- MACPF Domain -------------
O. fasciatus    GSMVEDFVAVVKGGSSESITALVSKKLPTPQLMGLWGEGVRFNPDFIRSTTRPLYELVTS  471
P. olivaceus    GSSVEDFVAVVRGGSSESITGLVSKKLPTPQLMRLWGEGVRFNPDFIRKTTRPLYELVTS  472
O. mykiss       GSMVEDYVVSGGDSESITWLAAKNLPTPPLMRLWGEAVHYNLDFIRSVTRPLYELVTA  472
H. sapiens      DTMVEDLVVLVRGGASEHITTLAYQELPTADLMQEWGDAVQYNPAIIKVKVEPLYELVTA  476
R. norvegicus   RTMVEDLVVLVRGGTSEYITSLAYKDLPTAELMKEWGDAVQYNPAIIKLKAEPLYELVTA  475
                 :  *** *.:* ** ** ** **  *. .:.***. **   **:.*::*   :*:  ..:********:
                                                        <--- EGF-1 --->
O. fasciatus    RDFSQDATLKRNLKRAISEYLAEASSCRCAPCHNNGVAVLKGTRCDCVCPVGYTGQCGEI  531
P. olivaceus    KDFSHDATLKRNLKRAISEYLAESSSCRCAPCHNNGVAVLRGTRCDCVCPTGYTGRGCEI  532
O. mykiss       RDFSSANSLKKNLRRAIAEYLEESSSCRCAPCRNNGLAVLKGSRCDCECVCPVGYSGLGCEI  532
H. sapiens      TDFAYSSTVRQNMKQALEEFQKEVSSCHCAPCQGNGVPVLKGSRCDCICPVGSQGLACEV  536
R. norvegicus   TDFAYSSTVKQNMKKALEEFQMEVSSCRCAPCRNNGVPILKESRCDCICPAGFQGVACEV  535
                 **:     ::::::*::*:** *:  * ***:****:.**:..::*: :**::*** .* .**:
                        <------------------- TSP-1 ------------------->
O. fasciatus    TQRRKDLAIEGSWSCWGAWSSCSGRTMTRSRQCNNPAPSNGGLACRGLQQESSECF  587
P. olivaceus    TQRKKQIATDGSWSCWGAWSSCSGRKMSRSRQCNNPVPSDGGLACRGLQQESTDCF  588
O. mykiss       TQR-PDIGIDGSWSCWGSWSPCRGRSKTRSRQCNNPAPSSGGIACRGLQMETTDCF  587
H. sapiens      SYR-KNTPIDGKWNCWSSWSSCSGRRKTRQRQCNNPPPQNGGSPCSGPASETLDCS  591
R. norvegicus   TNR-KDIPIDGKWSCWSDWSFCSGGRKTRQRQCNNPAPQRGGSPCSGPASETLDC-  589
                : *   :   :*.*.*.**. **.* * .  :*.***** *. ** .* *   *: :*
```

Figure 56. Continued.

158

Table 16. Percentages of identity, similarity, and gaps of Rb-C8α, Rb-C8β and Rb-C8γ with C8 orthologous from vertebrates

Gene	Species	Accession No.	AA	Identity%	Similarity %	Gap%
Rb-C8α	*Oncorhynchus mykiss* (Rainbow trout)	CAH65481	615	62.1	75.6	3.7
	Plecoglossus altivelis (Ayu)	FR714485	602	55	71	1.3
	Xenopus tropicalis (Frog)	NP_001005445	584	42.5	61	4.6
	Rattus norvegicus (Rat)	NP_001100140	587	42.2	57.7	6.4
	Homo sapiens (Human)	NM_000562	584	42	58.3	5.9
	Oryctolagus cuniculus (European rabbit)	P98136	585	41.2	55.4	8.3
	Bos taurus (Cattle)	NP_001039750	589	40.4	57.5	5.1
	Oplegnathus fasciatus (Rock bream) – *C8β*	JQ710661	587	33.7	49.5	10.1
Rb-C8β	*Paralichthys olivaceus* (Japanese flounder)	BAA86877	588	83	91.2	0.2
	Oncorhynchus mykiss (Rainbow trout)	AAL16647	587	67.4	80	2.7
	Homo sapiens (Human)	NP_000057	591	47.5	65.3	2.3
	Rattus norvegicus(Rat)	NP_001178688	589	47	63.4	3.7
	Macaca mulatta (Monkey)	XP_001114456	591	47	65.4	3
	Oryctolagus cuniculus (European Rabbit)	P98137	590	46	63.3	7.4
Rb-C8γ	*Oreochromis niloticus* (Nile Tilapia)	XP_003443552.1	222	77	88.7	0.5
	Takifugu rubripes (Puffer fish)	XP_003967902.1	220	77.5	85.1	1.4
	Dicentrarchus labrax (European seabass)	CBN81860.1	206	77.5	83.3	7.7
	Oryzias latipes (Medaka)	XP_004081042.1	221	70.1	79.6	0
	Oncorhynchus mykiss (Rainbow trout)	NP_001117880.1	221	66.1	79	2.7
	Salmo salar (Salmon)	NP_001139902.1	221	64.3	76.3	2.7
	Danio rerio(Zebrafish)	NP_957157.1	209	37.8	57.3	8.9
	Homo sapiens (Human)	AAI13627.1	202	28.3	49.3	10.3
	Bos taurus (Cattle)	NP_001103546.1	237	21.3	34.9	31.6

In phylogenetic tree, examined fish species formed two discrete clusters. *Rb-C8α* was sub-clustered with rainbow trout and *Rb-C8β* sub-clustered with Japanese flounder which showed highest identity respectively. Furthermore, they are clearly separated from the other TCC genes (Figure 45).

4.1.4 Tissue expression profiles of Rb-C8α, Rb-C8β and Rb-C8γ

The tissue specific mRNA expression profile was determined by qPCR using gene-specific primes designed from each gene. Rock bream β-actin gene was used to determine the relative mRNA values of each transcript. Furthermore, results were compared with the expression level of muscle to determine fold differences with respect to each tissue. In our study, eleven tissues were selected for mRNA expression namely from gills, liver, heart, spleen, intestine, head kidney, kidney, skin, muscle, brain and peripheral blood cells. According to the results, constitutive expressions of *Rb-C8α, Rb-C8β* and *Rb-C8γ* were detected in all tissues examined, abundant expressions were observed in liver, and marginal expression showed in gill, heart, brain, spleen, intestine, and head kidney, compared to the liver. The lowest expression was observed in muscle ($P< 0.05$) (Figure 46)

4.1.5 Transcriptional modulation of Rb-C8α, Rb-C8β and Rb-C8γ during immune stimulation

The mRNA expression levels were observed in liver and head kidney tissues, following the immune challenge experiments. The data was obtained in different time points, and relative expression was observed by qPCR using the gene specific primers. The relative transcription levels were calculated using rock bream *β actin* gene as a reference. The results were further compared with the respective PBS values (control) and fold difference were calculated for each tissue with respect to the *un-injected* control ($P<0.05$).

The expression pattern of Rb-C8α in rock bream liver in response to bacterial and viral challenges is shown in Figure 57. Transcription up-regulation of *Rb-C8α* upon Gram-negative bacteria (*E. tarda*) was showed highest induction at 12 h and the amount of transcripts were significantly increased from 0 h to 12 h post injection (p.i.), and reached the

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

highest level at 12 h p.i. (30-fold increase compared to the un-injected control (0 h), ($P<0.05$). Subsequently the expression levels were gradually decreased at 24 h, and again tend to increase from later time points. Following Gram-positive bacterial (*S. iniae*) challenge, the total *Rb-C8α* mRNA expression was significantly ($P<0.05$) enhanced in liver, as compared to the controls. Post-challenge examination, from 3 h to 48 h, revealed alternating increases and decreases in the expression level, with the maximum level (1.75 fold) was reached at 48 h p.i. Liver cells of the iridovirus-challenged animals exhibited a significant ($P<0.05$) and persistent up-regulation from 3 h to 48 h p.i., reaching peak expression (15-fold) at 48 h.

Rb-C8β transcripts were also followed the same comparable pattern of immune stimulation upon bacterial and viral challenges (Figure 58). However, relative fold differences were varied in each challenge. With respect to Gram-negative bacteria (*E. tarda*) the maximum fold difference was 20 in 12 h compared to un-injected control and, *Rb-C8β* expression for *S. iniae* was slightly fluctuated up to 24 h post-injection, followed with an increase (1.5-fold) at 48 h. Over the experiment time course, transcriptional response to iridovirus was gradually increased until 48 h p.i., and highest was reached at 48 h, which was indicated the late-phase response to the viral-induced immune challenge.

Similarly, transcriptional responses of each transcript were examined in head kidney for *E. tarda* challenge. In addition, the maximum fold increase was observed at 24 h p.i. In time course experiment, *Rb-C8α* transcript showed 250 fold differences at 24 h p.i. whereas *Rb-C8β* was up-regulated by 57 fold at 24 h (Figure 59).

In *Rb-C8γ*, the significant up-regulations were observed after rock breams challenge with Gram positive (*S. iniae*) and negative bacteria (*E. tarda*) in liver tissue. The significant up-regulation of *Rb-C8γ* transcripts were observed at 3 h, 6 h upon *E. tarda* challenge in liver.

161

On the other hand, it was up-regulated at 3h, 12 h to 48 h upon *S. iniae* p.i. challenge in liver

which showing a bipartite induction corresponding to the time interval (Figure 60).
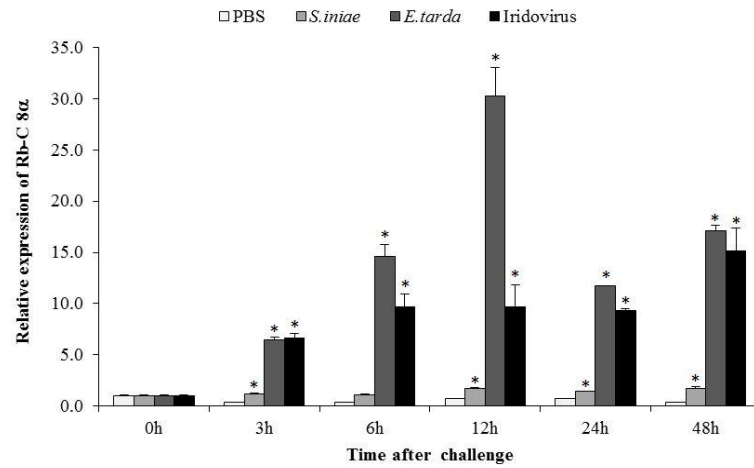
Figure 57. The temporal expression of liver Rb-C8α mRNA after rock bream challenge with different pathogens with respect to PBS (control) detected by qPCR
The expression of Rb-C8α mRNA relative to the internal control β-actin is expressed as mean±SD (n=3). Significant differences (*P<0.05* vs. un-injected control at 0 h) are marked with an asterisks.
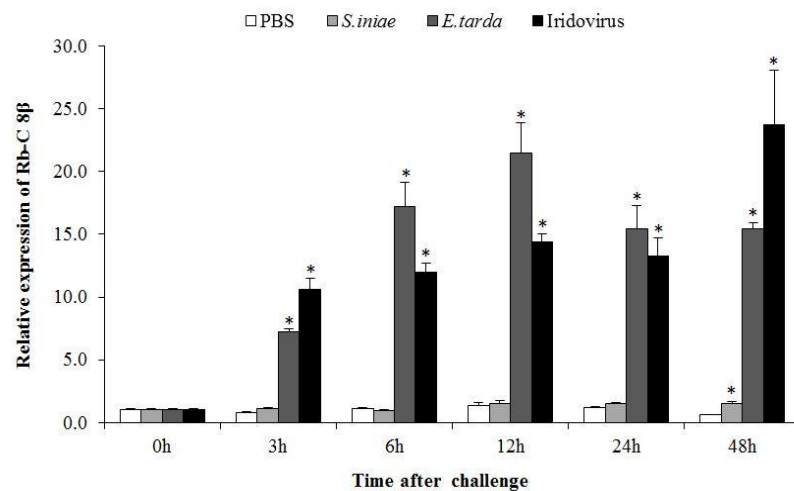
Figure 58. The temporal expression of liver Rb-C8β mRNA after rock bream challenge with different pathogens with respect to PBS (control) by qPCR
 The expression of Rb-C8α mRNA relative to the internal control β-actin is expressed as mean±SD (n=3). Significant difference (*P<0.05* vs. un-injected control at 0 h) are marked with asterisks.

162
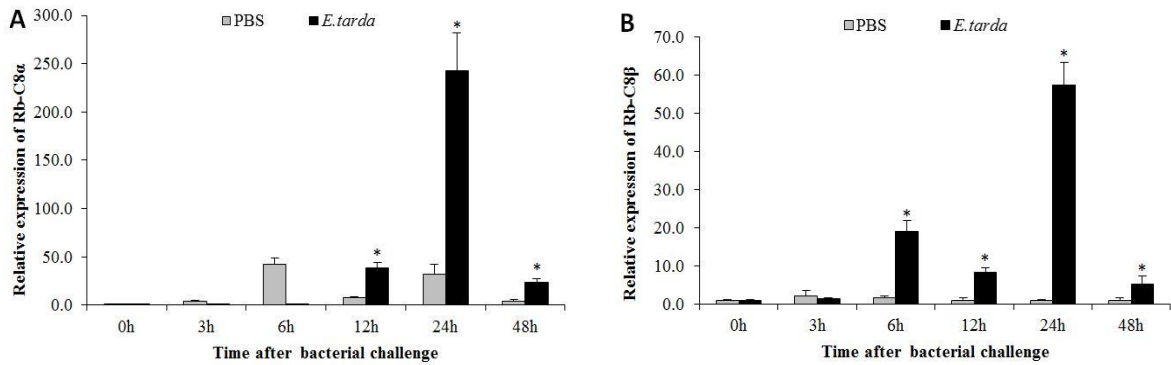
Figure 59. Temporal expression of head kidney Rb-C8α mRNA (A) head kidney Rb-C8β mRNA (B) after rock bream challenge with Gram-negative bacteria (*E. tarda*) and PBS (control) detected by qPCR

The amount of each mRNA relative to the internal control β-actin is expressed as mean±SD (n=3). Significant difference (*P<0.05* vs. un-injected control at 0 h) are marked with asterisks.
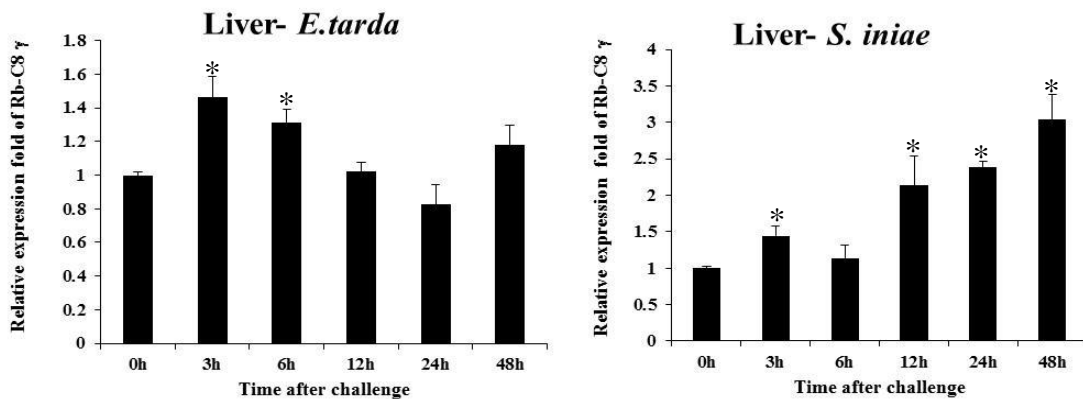
Figure 60. The relative expression of Rb-C8γ mRNA in liver after rock bream challenge with different pathogens

The expression of Rb-C8γ mRNA relative to the internal control β-actin is expressed as mean±SD (n=3). Significant difference (*P<0.05* vs. un-injected control at 0 h) are marked with asterisks.

4. 2 Discussion

In the present study, two genes encoding for MACPF super family protein and one lipocalin domain containing protein of complement lytic pathway were isolated, sequenced, and characterized from rock bream, *O. fasciatus.* Further, their mRNA distribution in various tissues and transcriptional response to pathogenic challenges were defined.

The MACPF super family includes more than 100 structurally different proteins and they are involved in different biological functions in the multi-cellular organism, of which a few have been defined at molecular level (Chondrou et al., 2006). Moreover, MACPF super family proteins exhibit limited sequence similarity, whilst all contain the MACPF signature motif (Ponting, 1999). Furthermore, the MACPF domain is conserved in bacteria, fungi, mammals and plants (Law SKA, 1988). Therefore, identification of the other domain features in classifying MACPF super family proteins is important. *Rb-C8α* and *Rb-C8β* share a common signature motif ((Y/W)G(T/S)H(F/Y)X$_6$GG) identified in the middle of the domain which is highly conserved among invertebrate and vertebrate species of MACPF super family. MACPF domains of *Rb-C8α* and *Rb-C8β* include 227 and 226 amino acids, respectively, and both have estimated molecular mass of 40 kDa. In MACPF signature motif of *Rb-C8α*, tyrosine (Y) is replaced with phenylalanine (F), deviating from the common signature motif of TCC proteins. Both genes exhibit the characteristic domains of the TCC such as TSP-1, LDLR-A, MACPF, and EGF (Chondrou et al., 2006; Li et al., 2007; Ponting, 1999). Furthermore, two TSP-1 domains have been identified in fish C8 orthologous. However, *Rb-C8α* contains only one TSP-1 domain in the downstream of the MACPF domain which is a disagreement with TCC proteins of other fish species (Katagiri et al., 1999; Stanley and Herz, 1987; Yeo et al., 1997). Moreover, each TSP-1 contains two *C*-mannosylation motifs (WXXWXXW). This is one of distinctive features present in fish. *Rb-C8α* has four potential *N*-glycosylation sites (Asn-X-Ser/Thr) which are similar to *O. mykiss* (Chondrou et al., 2006) and is important in post transcriptional modification such as folding and assembly. Therefore, presence of *N*-glycosylation sites and *C*-mannosylation motifs in *Rb-C8α* demonstrates its role in MAC formation and hence it needs complex structural arrangements. In agreement with studies done in trout (Kazantzi et al., 2003) and Japanese flounder (Katagiri et al., 1999), no *N*-glycosylation sites were found in *Rb-C8β*. In contrast, mammalians have two sites

(Chondrou et al., 2006), and in general, complement system proteins may contain 1-8 *N*-glycosylation sites depending on their functions. Cysteine residues were remarkably conserved among all evaluated species in our study. However, the domain boundaries and sequence similarities in the terminal complement component family between species have been shown to be ambiguous (Katagiri et al., 1999). The phylogenetic tree led to clear separation between two homologs as main clusters and within the each, they were sub clustered with highest identity fish C8 orthologous. The number of amino acids in *Rb-C8α* was more close to *P. altivelis* and interestingly *Rb-C8β* and C8β of *O. mykiss* have same number of amino acids. Moreover, *Rb-C8α* and *Rb-C8β* exhibit highest identity to rainbow trout, *O. mykiss* and to Japanese flounder *P. olivaceus*, respectively. The above unique and conserved characteristics, together with the high degree of similarity to other known orthologous, suggests that the *Rb-C8α* and *Rb-C8β* sequences belongs to the MACPF super family and the TCC of lytic pathway of the complement system.

In genomic structures, *Rb-C8α* has 11 exon and 10 introns structure which compatible with other orthologous except frog (exons: 10) and medaka fish (exons: 12). However, *Rb-C8β* has 12 exon and 11 intron structure which is conserved among all the species considered. Length of introns in higher vertebrates are much larger than those of fish, and within the fish, intron size vary between species. *Rb-C8α* and *Rb-C8β* were showed the highest structural similarity to fugu fish. In *Rb-C8α*, sizes of the 4th, 5th, 6th, 7th, 9th, and 10th exons are found to be identical in all the fish species used in the comparison. Moreover, in *Rb-C8β,* 4th, 5th, 6th, 7th and 9th exons were conserved among the examined species.

Although teleost research has recently focused on identification of complement lytic pathway genes, according to our knowledge promoter analysis has been performed only in grass carp, which was based on the prediction method for TCC genes (Li et al., 2007). In the *Rb-C8α* promoter region, a putative transcriptional initiation site was found at 64 bp upstream

of the start cordon (ATG) and no putative TATA box was found. However, in promoter region of *Rb-C8β*, a putative transcriptional initiation site and TATA box were found at 27 bp and 50 bp upstream of the start codon (ATG), respectively. According to our analysis, several important putative transcriptional binding sites were located in the 5'-flanking region of *Rb-C8α* and *Rb-C8β* (section 3.2.). In agreement with previous studies, C/EBP, and AP-1 were identified as common transcription factor binding sites for TCC genes (Li et al., 2007). In addition, Sp-1, OCT-1, YY-1, ICSBP, GATA-1, NF-κB, and NF-1 were found to be unique to several TCC genes. Presence of the above transcription factor binding sites in *Rb-C8α* and *Rb-C8β* implies they may mediate transcription as in other TCC genes. In contrast, human C9 (TCC family) gene contains transcription factor binding site for activator protein-2 (AP-2), yinyang-1 (YY-1), brain-2 (BRN-2), granulocyte macrophage colony stimulating factor (GM-CSF), and interleukin 6 (IL-6) (Witzel-Schlomp et al., 2001). Therefore, we hypothesize that, presence of different transcription factor binding sites indicate transcription process of human is marginally different from the transcription of TCC genes of teleost. Furthermore, a binding site for NF-κB, an important eukaryotic transcription factor, was found in *Rb-C8α* and *Rb-C8β*. NF-κB is involved in the control of a large number of normal cellular and tissue-specific processes, such as immune and inflammatory responses, development, cellular growth, and apoptosis (Long et al., 2004; Pikarsky et al., 2004). ICSBP, another putative binding factor of *Rb-C8α* and *Rb-C8β*, belongs to the interferon regulatory factor (IRF) family of proteins that plays an important role in interferon-mediated immune response (Barnes et al., 2002). Likewise, Oct-1 is involved in regulation of some T cell-specific cytokines expression, such as IL-2, IL-4 and IL-5. In conclusion, the *Rb-C8α* and *Rb-C8β* promoters contain several important putative transcriptional factor binding sites similar to those previously identified in other TCC genes from human and teleost fish. Moreover, most of these sites are associated with factors related to immunity, suggesting that *Rb-C8α*

and *Rb-C8β* may have an immune function, similar to that described for the C8s in other fish. However, further studies based on empirical data are needed to delineate the functional elements of *Rb-C8α* and *Rb-C8β* promoters.

The tissue expression profiles of *Rb-C8α* and *Rb-C8β* genes were examined by qPCR analysis. Both transcripts were detected in all examined tissues, including liver, peripheral blood cells, brain, skin, heart, spleen, intestine, head kidney, kidney, gills, and muscle, indicating a constitutive transcription of *Rb-C8α* and *Rb-C8β* in healthy rock bream. The expressions of *Rb-C8α* and *Rb-C8β* in liver were extremely high compared to the other tissues. In mammals, the liver is the primary production site for the majority of the complement components (Witzel-Schlomp et al., 2001). Moreover, rainbow trout C8β transcript was expressed in liver, intestine, and kidney but absent in spleen, brain (Kazantzi et al., 2003). Therefore, our results agree with the general insight that, in healthy fish, *C8* homologues are mainly expressed in liver and constitutively expressed in other tissues. As a whole, it can be suggested that *Rb-C8α* and *Rb-C8β* are similar to most complement components, which belong to the type of acute-phase proteins and are generally consisted of hepatic origin (Witzel-Schlomp et al., 2001).

TCC genes are direct immune related genes which are known to be induced by the different type of immune stimulants. A previous study conducted in grass carp, found that C9 expression was induced in liver by Gram-negative bacteria (*Flavobacterium columnare*) at 12 h p.i. (Li et al., 2007). According to the immune-responsive expression results, Gram-negative bacteria can effectively trigger expression of *Rb-C8α* in liver and transcription level was highest at 12 h p.i. compared to the *un-injected* controls. The *E. tarda* time course challenge experiment was yielded a clear pattern of induction and produced the highest induction among all the immune stimulants. Approximately 20-fold high induction was detected in Gram-negative bacterial challenge in both transcripts. Furthermore, they showed

very high up-regulation in head kidney which demonstrated that it expressed in other immune tissues. Therefore, we suggested that *Rb-C8α* and *Rb-C8β* genes sensitive to the Gram-negative bacterial challenge.

In virus challenge, continuous transcriptional induction was occurred throughout the 48 h post-iridovirus challenge and it showed increasing trends with the time. We observed that, *Rb-C8β* transcripts were also followed the similar pattern of induction in response to bacterial and viral stimulants. To the best of our knowledge, no studies of C8α and C8β orthologous have been performed under virus-induced expression conditions. Therefore, our expression pattern could not be comparatively explained. However, few induced-expression studies on TCC genes were found in other species that can be used to interpret our data. In microarray (GSE19049) and qPCR array (GSE19503) studies in virus-challenged zebrafish showed that increased expression of C9 transcripts in fins and organs like liver, spleen and head kidney, by 15.3-fold and 2.7-fold, respectively, at 48 h post challenge with viral hemorrhagic septicemia virus (VHSV) (Encinas et al., 2010).

The complement C8γ gene is one the smallest protein identified in the rock bream complement system. It was consisted of only one domain named as lipocalin domain which is responsible for making hydrophobic conditions. The C8γ is covalently bind to the C8α chain and both dimer is associated noncovalently with the C8β chain (Lovelace et al., 2011). The C8γ gene has been identified in human, mouse, rabbit and rainbow trout from teleost. However, it is unique to the complement system having only one extraordinary domain and the functions of C8γ yet to be elucidated.

In summary, we have identified the full-length genome sequences of complement component 8α, 8β and 8γ from rock bream. Genome structures analysis of *Rb-C8α, Rb-C8β* and *Rb-C8γ* were confirmed their substantial similarity with orthologous of other teleost fish and showed that they belong to the TCC protein family. Moreover, *Rb-C8α Rb-C8β* and *Rb-*

C8γ were found to be potentially induced by infection with *E. tarda*, *S. iniae,* and iridovirus, indicating their potential role in the rock bream immune defense system. However, the mechanisms of *Rb-C8α, Rb-C8β* and *Rb-C8γ* that mediates immune responses to acute-phase infection require further studies.

5. Results and discussion of rock bream complement C9

5.1 Results

5.1.1 Molecular characterization of Rb-C9

Based on BLAST-X analysis of EST data from the normalized cDNA library, we identified a putative complement component 9 from rock bream and designated as *Rb-C9*. The sequence was deposited in NCBI under accession number JQ277722. The complete cDNA sequence was 2112 bp, including a 252 bp 5'-untranslated region (5′-UTR), 1782 bp of open reading frame (ORF) encoding 593 amino acids, and a 75 bp 3'-untranslated region with a polyadenylation signal (Figure 61). The predicted molecular weight of *Rb-C9* was 66 kDa and the theoretical isoelectric point was 6.1. According to the SignalP server prediction, the *Rb-C9* contains a signal peptide representing a cleavage site at 26-27 amino acid residues in the N-terminus. The mature protein contains two thrombospondin type-1 (TSP-1) repeat profiles, a low-density lipoprotein receptor domain class A signature, an epidermal growth factor (EGF)-like domain, four potential N-glycosylation sites (Asn-X-Ser/Thr), and an MACPF domain profile. Within the MACPF domain profile, the MACPF domain and signature were identified using the ScanProsite programs in the PROSITE database (de Castro et al., 2006).

169

```
                         ←───────────────────────
Rock bream        -----------MRIEVALQLGFCGLCLTLALLGEG--------MGVDLPDPPAVNCVWSR 41
Puffer fish       -----------MRTEAALQLGFCALCVMLALLGEG--------MGRELPDPPAVNCVWSR 41
Japanese flounder ------------------------------------------MAVELPDPPGVNCRWGL 17
Common mummichog  ----------------------TGPHLAFFDAG--------LGAVLPEPAAVNCQWSR 28
Rainbow trout     ----------MAEVAAALLLGFCAVAVSVSVSGDENRTGKRDTREVKVAPPDPVDCVWSR 50
Sweet fish        ----------MRTFSTVLVGFSVLCLTGAVLGEDR---AIPKRMTRDVEPARIDCQWSR 46
Grass carp        ----------MKALAALFVTICIFYQANGKSITDHMQ-GPRREVRQINDPASIDCKMSA 48
Cattle            ------------MSAGQRFAFAICILEISLLRAGPTPSYDP--AERQGTPLPIDCRMSS 45
Human             ------------MSACWSFAVAICILEISILTAQYTTSYDPELTESSGSASHIDCRMSP 47
Rat               MLLRTPGLPRRSGMASGVTITLAIAIFALEINAQAPEPTPRE--EPSADALLPIDCRMST 58
                                                                    ::* .

                    ─────────────── TSP-1 ────────────────→
Rock bream        WSEWSSCDPCTRTRRRSRSVEVFGQFDGESCQG-SLGDRESCIPRGACDLPPRPACSDTE 100
Puffer fish       WAPWSSCDPCTNTRRRSRGVEVFGQFAGIACQG-SVGDREYCITNAKCNLPPPRECSDSE 100
Japanese flounder WSQWSLCDPCTNTRRRSRTVEVFGQFGGESCQG-SLGDREHCTADTECARPPPPQCLDSE 76
Common mummichog  WSEWSSCDSCTKSRRRSRGIEAFGQFDGESCMD-SIGDRQACQPSESCLTPTLLPCLESE 87
Rainbow trout     WSEWTPCNSCTKIRHRSRSVEVFGQFGGKPCQGQPIGEQQACTSDAVCEQALPSECSSTE 110
Sweet fish        WSEWTPCDPCSKSQRRSRGVEVFGQFQGMACEG-SIGEDRECNPAGDCNRPPLHLCSDSE 105
Grass carp        WSEWTSCDPCTNTTHRSRGIEVFGQFKGSRCIY-PIGERRPCKPSTKCQMDPPPVCKSSQ 107
Cattle            WSEWSKCDPCLKQMFRSRSIEIFGQFNGRKCVD-AVGDRQQCVPTEACE-DPEEGCGN-D 102
Human             WSEWSQCDPCLRQMFRSRSIEVFGQFNGKRCTD-AVGDRRQCVPTEPCE-DAEDDCGN-D 104
Rat               WSQWSQCDPCLKQRFRSRSIEVFGQFQGKSCAD-ALGDRQHCEPTQECE-EVQENCGN-D 115
                    *: *  *:.*   *** :* **** *  .  .:*: . *  . *    * . :

                      ──────────── LDLR-A ─────────→  ←──── MACPF-2 ──────
Rock bream        FQC--ESGYCIKRRLMCNGDYDCE--DGSDE-DCE-PVRKPCGALVLDTNEQGRTAGYGI 154
Puffer fish       FQC--ESGSCIKLRLMCNGDYDCE--DGSDE-DCE-PLRKTCPPTVLDTNEQGRTAGYGI 154
Japanese flounder FQC--ESGSCIKKRLMCNGDYDCE--DGSDE-DCD-PTNKPCKSAVLENNEQGRTAGYGI 130
Common mummichog  FRCTTELQTCIKKRLECNGDYDCE--DGSDE-DCE-PTRTPCGASVLENNEQGRTAGYGI 143
Rainbow trout     FTC--ESGACIKLRLSCNGDYDCE--DGSDE-DCE-PVRKPCGTKLYDTNEQGRTAGYGI 164
Sweet fish        FQC--DSGFCVKKALMCNGDLDCD--DMSDE-TCDNPVRLPCGPVSLENNEQGRTAGYGI 160
Grass carp        WQC--ASGICISKHLKCNGDHDCGESDTSDEDDC-DDIRSPCGKTAVFESDIATQAGYGI 164
Cattle            FQC--GTGRCIKNRLLCNEDNDCG--DYSDEDNCEQDPRPPCRNRVVEESELARTAGFGI 158
Human             FQC--ESTGRCIKRKLLCNGDNDCG--DFSDEDCES DPRPPCRDRVVEESELGRTAGYGI 160
Rat               FQC--ETGRCIKRKLLCNGDNDCG--DFSDESLCES DPRLPCRDRVVEESELGRTAGYGI 171
                    : *      *:.   *  ** * **    * ***    . .*  . .:.  **:**

                    ──────────────── MACPF-2 ────────────────
Rock bream        NVLGADPRMNPFNNDYFNGRCDNVRNPNTGQRDRLPWNIGVLNYETLVEETVSREIYEDT 214
Puffer fish       NILGADPRMNPFNNDFFNGRCDKVRNPNTLQLDRLPWNIGVLNYQTLVEETASREIYEDS 214
Japanese flounder NILGSDPRMNPFNNDYFNGRCQIVNNPNSATRDRLPWNVGVLNYETLAEETVSREIYEDT 190
Common mummichog  NILGADPRRNPFNNDYFNGMCNKVKNINNNEYNRLPWNVGLLNYETIAEETVSKEIYEDT 203
Rainbow trout     NILGMEPRINPFNNDYFNGMCNKVKNINNNEYNRLPWNVGLLNYETIAEETVSKEIYEDT 224
Sweet fish        NILGADPRMNPFNNDFFNGKCIKVRNPTSRNLDRLPWNVGSLNYETRAEDLTSREIYEDT 220
Grass carp        NVLGSGPRMNPFNNRLYNGQCNRIRDPTTLEYNRLPWNIGVLNYETKVEESSSKQMYEDI 224
Cattle            NILGMDPLSTPFDNQYYNGLCDRVWDGNTLTYYRRPWNVASLTYDTKADKNFRTENHEES 218
Human             NILGMDPLSTPFDNEFYNGLCNRDRDGNTLTYYRRPWNVASLIYETKGEKNFRTEHYEEQ 220
Rat               NILGMDPLGTPFDNEFYNGLCDRVRDGNTLTYYRKPWNVAFLAYETKADKNFRTENYEEQ 231
                    *:** *  .**:* :** *   :     . * ***:. *   *:*  :. : :*:

                    ─────────────── MACPF-2 ───────────
Rock bream        HSLLREMLKEMSFKIDVGLSFKFSPSEQSL(N)V--------------------------- 247
Puffer fish       YSLLREMLKEMSIKVDAGLSFKFKSTEPSMSNN--------------------------- 247
Japanese flounder HSLLREILKEMVKVDAGLSFKFKLSEESMSKS--------------------------- 223
Common mummichog  YSIVKEMLKETTFKVDAGFSFKFTATEPSLANT--------------------------- 236
Rainbow trout     YTLLRELMTETKLTVSAGLNLKFTPTEKSMAKSNTTV---------------------- 261
Sweet fish        HSLINEITKETTFSIGVGGSLKFSPSEKSKKKSNTTV---------------------- 257
Grass carp        HSLIKEITTESSKTFEAGLSVKFSPTEPSSADEGGAAGSAGNAGSAGDVASPGGGGEAAS 284
Cattle            IQILRTIIEEKKLNFNAGLSVKYTPVEAIEKNKCVDLEHSDKGSTSSPSK---------- 268
Human             IEAFKSIIQEKTSNFNAAISLKFTPTETNKAEQCCEETAS---SIS------------- 263
Rat               FEMFKTIVRDRTTSFNANLALKFTITEAPIKKVGVDEVSPEKNSSKP------------- 278
                    ..  :  :  .  ... .   .*:.  * .

                    ──────────────── MACPF-2 ────────────
Rock bream        ----------------------------SLKLGVDHQYEKKTMIRDVTETSNVF(N)KSF 277
Puffer fish       ----------------------------SLKLDASLEYEKKTMIKDVSELTNIKNKSF 277
Japanese flounder ----------------------------SAKLDLNVEYEKKTMIKEVTEYSTIKNKSF 253
Common mummichog  ----------------------------TIGAEAGAEFSKKNVLKEVTEYSVIKNKSF 266
Rainbow trout     ----------------------------SGGVGLDAEYDRTQMIKEVSEYTTIKNKSF 291
Sweet fish        ----------------------------TGGVGADVDYSRRNLIKEISEYTKVKKGSF 287
Grass carp        PSKNGGTEAEGPGVAGEAKGSDGGQNDMVISGSVEFDTKTEKNNIIKQLSEITTTKRKTF 344
Cattle            ----------------------------LAAEEAKFRFTYSKDDIYRLLSSYSAKQEKMF 299
Human             ----------------------------LHGKGSFRFSYSKNETYQLFLSYSSKKEKMF 294
Rat               ----------------------------KDSSVDFQFSYFKKENFQRLSSYLSQTKKMF 309
                                                :    :  .       .    . *
```

Figure 61. Multiple sequence alignment of Rb-C9 and C9 of other vertebrates. The amino acid sequences were compared by the ClustalW (v2.1) multiple sequence alignment program

Identical residues are indicated by black shading asterisks (*). Conserved and semic-conserved residues are indicated by gray shading, semicolon (:) and dots (.), respectively. Missing amino acids are denoted by dashes, and the conserved domain is indicated by double arrows. The N-glycosylation sites are circled. The GenBank accession numbers of each gene is presented in phylogenic tree next to each species.

```
                                ┌──────────────────MACPF-2──────────────────
                                                                  ┌─MACPF signature─┐
Rock bream         MRVKCRVQLSTYRLRTRDLQVANEFLEHVKSLPLQYEKGIYFAFLEDYGTHYTRNGKSGG 337
Puffer fish        MRVKCRLQLSTYRMRSHQLQVADEFVAHVKSLPLEYEKGIYYAFLEDYGTHYTKNGKSGG 337
Japanese flounder  MRVKCRVQLSTYRMRSRDLKVADEFLAHINTLPLEYEKGIYFAFLEDYGTHYTKNGKSGG 313
Common mummichog    MRVKCRVQLSSYRMRSRELNVAEEFLIHVRSLPVVYEKGIYFAFLEDYGTHYTKNGKSGG 326
Rainbow trout      MRVNGRLQLSTYRMRSRDLQVAGEFLEHVKSLPLEYEKGQYFSFLEDYGTHYTRNGKSGG 351
Sweet fish         IRVKGKVEISKYRLISHDLQVTDRFLEDVMRLPLEYEKGMYFAFVEDYGTHYTKNGVSGG 347
Grass carp         MRVKCRVEIATYRMRQRGLEVSSTFLDDVDALPLTYEKGQYFAFLEDYGTHFTKNGKSGG 404
Cattle             LHVKCKVHLGRFVMRSRDVMLQTTFLDSINTLPTTYEKGEYFAFLETYGTHYSSSGSLGG 359
Human              LHVKGEIHLGRFVMRNRDVVLTTTFVDDIKALPTTYEKGEYFAFLETYGTHYSSSGSLGG 354
Rat                LHVRGMIQLGRFVMRNRGVMLTTTFLDDVKALPVSYEKGEYFGFLETYGTHYSSSGSLGG 369
                   ::*.* ::.*. : : : : : : *: : ** **** *:.*:* ****:: .*  **

                                ┌──────────────────MACPF-2──────────────────
Rock bream         EYELIYVLDQDAIRQFN̲LTERKIQECVKLGITAELGGIP-----FTEAKGKLGLD-SCEP 391
Puffer fish        EYELVYVLNQDTIKAKNLTERKIQECLKIGIEAEFATTS-----VQDGKAHAKLN-KCDD 391
Japanese flounder  EYELIYVLNQDTIKAKNLTERKIQECIKIGITADFD---------ELAKGNVGKK-DCDD 363
Common mummichog    EYELIYILNQDAIKAKNMTQRGVQTCVKAGISADFGFPT-----GGSAKAHVRPE-YCDN 380
Rainbow trout      EYQLVYVLNQDTIKDKKLTERKLQDCIKVGISANFDTNI-----GIGGDAHIRPG-HCKD 405
Sweet fish         EYELVYVLNQDAIKEKKITEREVQDCVKVGITANFDTNI-----GIGGEAHVKPD-YCNK 401
Grass carp         EIEFIYVMNEDILKQKKVTESTIKDCVEVGIKGSIQFSK-----IAEAERHVKPVGNCDT 459
Cattle             LYELIYVLDKKSMEQKDIELRDVQRCLGFDLDLSLKVG-------VEVTGNFDSK-LCSK 411
Human              LYELIYVLDKASMKRKGVELKDIKRCLGYHLDVSLAFSE------ISVGAEFNKD-DCVK 407
Rat                LYELIYVLDKASMKEKGVELSDVKRCLGFNLDVSLYTPLQTTLEGPSLTANVNHS-DCLK 428
                    *:*:*::::: :. * :   ::  *:  :  ::            .      *

                                ┌──────────────────MACPF-2──────────────────
Rock bream         VTTKKKDETHGKAVVDKVMTSVKGGTLQTAVAMRAKLNKEG-VMDITTYQNWARSITDAP 450
Puffer fish        VTTKSQGDVEGKAVVDNVMTSVKGGSLESAVTMRAKLNKEG-VMDIATYQNWARTIASAP 450
Japanese flounder  VTTKEEGSIHGKAVVDKVMTSVKGGTLESAVTMRAKLNRDG-VMDLPTYQNWARTITNAP 422
Common mummichog    PTNTQTATAEGKALVDKVMTSVKGGTLETAVTMRAQLNKDG-VMNIDAYRNWARSIADVP 439
Rainbow trout      TVNKNTAEKEGKALVDKVITLVRGGTLEAAVAMRTQITKEG-LMDVTTYQNWARTVGDAP 464
Sweet fish         VTPKNTDSKEGSAAVDRVLTLVRGGTTDTASAMRTKLDKDA-VMDLATIREWARSLADNP 460
Grass carp         ITDKPTDDRDTRGVIDKVLIAVRGGSPATAVAMKSQLTKDG-ILDYTQYVEWAKTVSALP 518
Cattle             KGMGQTETNPEADLFDDVITFIRGGTRKYATELKEKLLRGARMINVTDFVNWAASLNHAP 471
Human              RGEGRAVNITSENLIDDVVSLIRGGTRKYAFELKEKLLRGT-VIDVTDFVNWASSINDAP 466
Rat                TGDGKVVNISRDHIIDDVISFIRGGTRKQAVLLKEKLLRGAKTIDVNDFINWASSLDDAP 488
                       .* *:  :::**:  *  :: :: : ::       :**  ::  *

                                ┌──────────────────MACPF-2──────────────────▶
 Rock bream        ALINSQPEPIYTLIPLDMPDANTRISNLKQATADYVAEYSVCKCKPCQNGGTLALLDGKC 510
 Puffer fish       ALINSEPEPIYMLIPTDIPGANSRIANLKQATADYVAEYNVCKCRPCHNGGTLALLDGKC 510
 Japanese flounder ALLYSEPEPIYMLEMPGANSRIANLKRATADYVAEYNICKCKPCHNGGTIAMLDGKC 482
 Common mummichog   ALINSEAEPIYNLIPLDTPDANNRISNLKKAIEEYVAEYNICKCKPCQNGGTIALIDGKC 499
 Rainbow trout     ALLSSEPEPIQTLIPLSMPDANTRRLNMQRATQEYEAEYSVCKCKPCHNGGSLALLDGKC 524
 Sweet fish        ALLKSEPEPIYTIIPLSFPNVNAFRTNIKRAIEEYVAEYNVCKCQPCQNGGSVILIDGSC 520
 Grass Carp        ALIHSDPEPIYNAIPLDFFPDAQPRRDNLRRAIDDYVAEYSVCKCQPCQNGGTVIQIDGEC 578
 Cattle            VLISQKLVPIYDLIPVKMKDAHLKKQNLERAIEDYINEFSVRKCQPCQNGGTVVLLDGEC 531
 Human             VLISQKLSPIYNLVPVKMKNAHLKKQNLERAIEDYINEFSVRKCHTCQNGGTVILMDGKC 526
 Rat               ALISQKLSPIYNLIPLTMKDAYAKKQNMEKAIEDYVNEFSARKCYPCQNGGTAILLDGQC 548
                   .*: .. **  :*   ..  *:::*  :* *:. **  .*:***:   :**.*

                                ┌──────────────────TSP-1──────────────
Rock bream         ICLCPHLFEGLACQNFKRDKVRTGGKSTRPTVNQEⓄNSCWSSWSNCRAEKRTRTRSCNT 570
Puffer fish        ICMCSNLFEGLGCQNFKGDKARVP--AARPAVTQEGNWSCWSSWSNCQGQKRSRTRYCNT 568
Japanese flounder  MCLCSFLFEGMGCQNFKDKAQHAG--TRPTVNHEGNWSCWSAWSSCSGGKRTRARNCNT 540
Common mummichog    ICMCTGQFEGLACQNYRADKAANQA-GQRPPVLQEGNWSCWGPWSPCSGGKHSRTRRCRT 558
Rainbow trout      LCLCLPQFEGLACQDAKADNNKNTK-TPVESVPQEGNWSCWAAWSGCSGGKRIRTRSCNT 583
Sweet fish         SCLCTPEFEGLACQTLKSDKRRPTS------VTEKGNWACWSAWSSCSAGRRQRSRLCKT 574
Grass Carp         KCMCLPGSEGVACQIVDMEMVKGKS------FEQLGNWGCWTPWSPCSGRRRKRTRTCNT 632
Cattle             VCSCPKEFKGVACEIKK--------------------------------------------- 548
Human              LCACPFKFEGIACEISKQKISEGEDNLFDVIKDKNMHFGVI------------------- 567
Rat                MCSCTIKFKGIACEISKQR------------------------------------------ 567
                    * *   :*:.*:

                                ┌──────────────────────────▶
Rock bream         DGLSGGAVCRGDTSSEEHCIAEK 593
Puffer fish        EGVLG-AECRGEIRSEEYC---- 586
Japanese flounder  DGLPD-ANCRGDTNSDEFC---- 558
Common mummichog    DGLSETVKCRGEATSEEYC---- 577
Rainbow trout      QGLSD-ATCRGDIVTEDYC---- 601
Sweet fish         GQLSG-ATCLGNPSSDDYC---- 592
Grass carp         RGVPN-GICKGDTTSEDYC---- 650
Cattle             ----------------------
Human              ----------------------
Rat                ----------------------
```

Figure 61. Continued.

5.1.2 Genomic structure and promoter region of Rb-C9

Genomic library screening for *Rb-C9* led to the identification of only one positive BAC clone that contained the sequence of interest. The putative genomic sequence of *Rb-C9* was recognized by comparison with the full-length cDNA sequence. The *Rb-C9* genomic sequence (JQ277722) was about 6 kb, and was composed of 11 exons interrupted by 10 introns. The 5'- and 3'-ends of all introns were carried canonical splicing motifs (by the GT-AG rule). The number of exons and introns were well-conserved in all the C9 orthologous were considered in this analysis. Furthermore, while the exons and introns of C9 orthologous vary in length those within C9 from fish species were shown less variability (Figure 62).
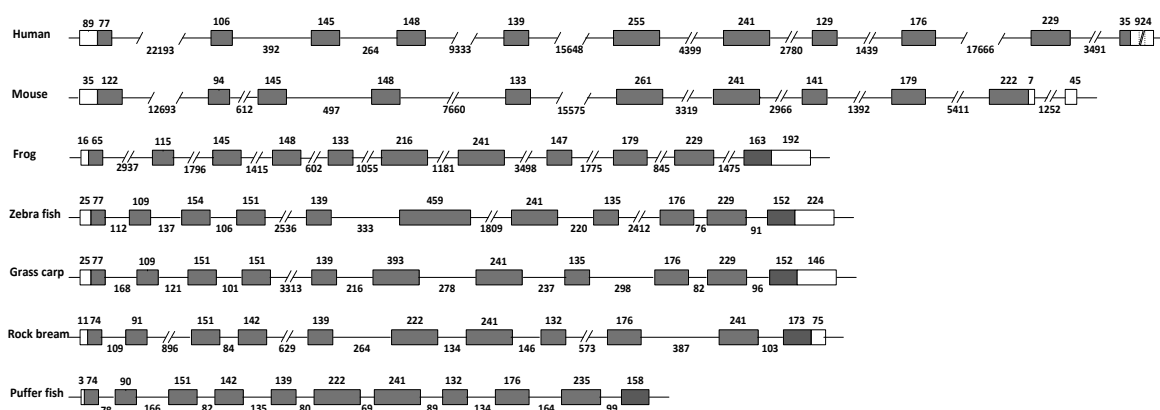
Figure 62. Genomic structure of Rb-C9 compared to other selected C9 genome structures

Exons are represented in gray solid boxes. Un-translated regions are shown as light-colored boxes. Introns are represented in thin lines in-between the exons. Introns larger than 600 bp are indicated as interrupted lines. Numbers above the boxes and below the lines are indicated the lengths of exons and introns respectively. Selected organisms are *Homo sapiens* (CAA69849), *Mus musculus* (AAH11137), *Danio rerio* (ENSDART00000147132), *Ctenopharyngodon idella* (ABN49522), *Takifugu rubripes* (AAC60288) and *Xenopus (Silurana) tropicalis* (NW_003163744).

Typically, the promoter region of any gene has a complex structure consisting of multiple functional binding sites for proteins involved in the transcription initiation process. According to the BDGP prediction analysis of the *Rb-C9* promoter region, a transcription initiation site (+1 site) was recognized at 252 bp upstream of the start codon (ATG). A TATA

172

(TATATAA) box was also been identified at -26 to -32 upstream from the initiation site. Putative immune-related transcription factor binding sites were identified, including the CCAAT-enhancer binding protein (C/EBP), nuclear factor kappa B (NF-κB), activator protein-1 (AP-1), interferon consensus sequence binding protein (ICSBP), octamer transcription factor (OCT-1), specificity protein-1(SP-1), HNF-1, HNF-3, CdxA, YY-1, GATA-1, GATA-3, Nkx-2, SRY, and XFD-3 (Figure 63).

```
                                           TCCTGACTGACATTAGAGGCATCAGGGAATAGTTGGACTGTGTTT   -1534
                                                           Sp-1
ATTTAAGTCAGCAGAAAATACTTGTCAGCATGCTGAATACTCTAGAATTTAACATCAGTTGTACCCTCTGAGATG   -1489
  C/EBP-alpha                    OCT-1
GATAAAAGCTGATGATTTGTCAGCCTCTCACCGCTGCCAAACCTCTGTGTTCATGAATAATGATTCACCTGATCA   -1414
      AP-4
TGAATGATTAGCTAATTTGATTATAATCCTAATAAGTCCTGTCTACTTTTAAATTCACTCCTGCTTTCATGTCTC   -1339
                        C/EBP-beta
TACACTGAGAAGTGTTGTTTCCCATTTTCTGTAGTTTATACACACTGCACTGCACTCCCTCCTTTTGAAGAACA   -1264
                NF-Kappa         GATA-1               Sp-1
AAGTACTGCAGACTCAGCAAAAGCCTCTCAGATCACATGACAGATCATTAGCTGCAACCGTTTTAAAACTTCCTC   -1189
                                                                    NF-Kappa
TGCATTGTGCAAGGCTCGTTTTAATATTTTTCATAAAAGAAAACAGTGTTTGACTTGTCAGAGGAAGAAATTAGT   -1114
      C/EBP-gama
ACAAGTGTATGCATATAAACAGTTTTTCTCTCTATATTGGTGTGTTCCCTCAGCCTATACAAAAAAAACAGTGCG   -1039
            GATA-3
TGAATGTAAAGACATGTTGTCTGTTCGATGTTGCCATCTTGTGGACAAAGAGTTGCACTTCAGCGCACGCTCAGC   -964
                                  YY-1
AAACCTGTGCAGCATGAACCCAATCGCAGCGCATGCCACTTCCAAACGAGTGTGTGGGTGGTTGAATCTTGATAG   -889
                     GATA-3
TCTTTCCATTCAGTTTCCCAATCTCAGACGATGACCCAAATGTACCATGGACGTTTCTCGTCAATGTGGTACAAC   -814
          ICSBP              Ap-1
AGTCAAAACCTGCCTCTTTAAATATTTGCTCTGCTTTAGTGAAAACCTGCCAGAATGATTGCATAATCCTGGAGA   -739
HNF-1              Oct-1
AAAAAAAAACAAATTAACAAACTGTCTCCAAATAACAGAAGAATTCTTAAAATaAAAACCAAAACACAAGAAATG   -664
        HNF-1
ACTTTTGGCCTTGTGCATAAAAAGTGTGTATGCACTAAAGGCCACACTTGACGCCGCAGTATTAAGAGTTTGTCA   -589
GCTTTTCTGTGCTGGCCCAAGTTCCCCTCCGCCTTTCACCCAGTGCATGCTGGGATACTTTCCAGCTTCCTATGA   -514
  GATA-1                    Sp-1
CTCTGACTGGATACAGAACAAATGACTGAATATATCATTTCTATAAAAACATCTTAACATAACATGTACTGCCAA   -439
                HNF-3                GATA-1              C/EBP-alpha
ATATTGCACTGACCACTGTTTCCTATTTGTGTTGTACATTAAAGTCGTAGGATGGATTGTGTCACATGGCTCTTG   -364
GCCCCTGAATCCCTAGTTCACATCATTTAAACTGCGCATTTTTATTGATTTTTATTTTCATGATTCATACCTTCCAA   -289
                                      SRY          ICSBP       Ap-1
ATTAGATCATTAAATAAATCTAACTTTTTCTGCAGCAAGTTCCGATAATTGGGAAGTATAGTATTAAAAGTAGAT   -214
              SRY                       Nkx-2 C/EBP-alpha         GATA-1
TATTTTaAAaAAAAACTTGCATAAACATTAAAATATATTTTTAAGTATAAGTATCTGTTATAGAGTCATGTGCTC   -139
            OCT-1    Nkx-2                                        AP-1
ATCAATCTGAGTGTTTGTTTTATTGGCTTTAGTGAATCAAGGTGTAGTTTCAACCAGTGGTTAAGCACAGAACT   -64
          HNF-3     OCT-1       HNF-1           Sp-1
AAAATAGCGGTTCACAGATTAATATAACAGTTATATAACAGTTTTGCCGAGATGATGGGAAGATGTGAATCATCA    12
                     CdxA       TATA box                    +1 site
CTTTTTACTTGTTTTTGGAGAGGCTTTGCTCACACTGTGAAGAACGAAATATACTTGCAGGATGAGTATGGTGTTT    87
  Nkx-2
TAACAAACACTGTAAACTCTGGTTCGGGAGTTTCCAAGCCCTGAATCAATATCGCTGTACTATTGACTGTGACAA   162
  XFD-3                       NF-kappaB                                XFD-3
CAACAACATGGTCCTTTTGCAGTTGGATAATCATTGACAGAGAGGTAGCTTAAAAGACTAATGCACCATTTGTCT   237
CTTCAAAGGATCATCATGAGGATTGAGGTCGCTCTCCAGCTGGGCTTCTGTGGCCTGTGCCTTACTCTGGCGCTT   312
             M  R  I  E  V  A  L  Q  L  G  F  C  G  L  C  L  T  L  A  L
```

Figure 63. Sequence of the Rb-C9 5'-flanking region (JQ277722)
Putative transcription factor binding sites are shown in bold and underlined with corresponding factor identified below. The putative TATA box is boxed and identified below. The transcription initiation site is indicated by a bent arrow. Translated amino acids (uppercase, in bold) and start codon (italicized uppercase, in bold) are shown.

5.1.3 Similarity and phylogenic relationship of Rb-C9

According to the NCBI-BLAST results, *Rb-C9* has 74% amino acid identity with puffer fish (*Takifugu rubripes*: AAC60288) and 70.2% with Japanese flounder (*Paralichthys olivaceus:* BAA86878). It also shares 63.2% identity with common mummichog (*Fundulus heteroclitus* : AAR87007) and more than 45% identity with rainbow trout (*Oncorhynchus mykiss*:NP_001117898), sweet fish (*Plecoglossus altivelis*: CBX31962) and grass carp (*Ctenopharyngodon idella*: ABN49522). Multiple sequence alignment of *Rb-C9* was indicated that each domain has higher similarity with the same entities of other considered species. All the cysteine residues were conserved in almost all species and important signature motifs were highly conserved among all the examined organisms (Figure 61).

In order to prove the evolutionary relationship, a phylogenetic tree (Figure 45) was constructed based on the deduced amino acid sequences of *Rb-C9* and known C9 sequences. According to the results, examined fish species were formed a discrete sub-cluster. *Rb-C9* was clustered with puffer fish, which also had the highest identity. All C9 genes originated from a common base, were indicated origin from a common ancestor.

5.1.3 Tissue distribution analysis of Rb-C9 mRNA

The tissue-specific mRNA expression profile was determined by qPCR using gene specific primers designed from the *Rb-C9* sequence. The rock bream β-actin gene was used to determine the relative *Rb-C9* mRNA values. Furthermore, results were compared with the expression level of muscle to determine the tissue-specific expression with respect to each tissue. Analysis of 11 tissues was revealed that the most abundant expression was in liver, and marginal expression was in brain, gills, skin, heart and spleen, as compared to muscle. The lowest expression was observed in muscle ($P < 0.05$) (Figure 46).

5.1.4 Transcriptional modulation of Rb-C9 expression during immune stimulation

Following immune challenge, the expression responsiveness of *Rb-C9* was investigated by measuring mRNA expression levels in liver and peripheral blood cells at various time points post-challenge, by qPCR. The relative transcription levels were calculated using rock bream β-actin gene as a reference. The results were further compared with the respective PBS-injected controls to calculate the fold-difference for each tissue.

The expression pattern of *Rb-C9* in rock bream liver in response to challenge with the Gram-negative bacteria *E. tarda* is shown in Figure 63A. The amount of transcripts was significantly increased from 0 h to12 h post injection (p.i.), and the highest level was recorded at 12 h p.i. (58-fold in increase compared to the un-injected control (0 h), $P < 0.05$). Subsequently, the expression levels were decreased gradually, until 48 h p.i. when the expression was 16-fold higher than the un-injected control and significantly up-regulated compared to the PBS control. Following challenge with the Gram-positive bacteria, *S. iniae*, the total *Rb-C9* mRNA expression was significantly enhanced in liver, as compared to the controls ($P < 0.05$). Post-challenge examination, from 6 h to 48 h, was revealed alternating increases and decreases in the expression level and the maximum level was reached at 24 h p.i. (Figure 63B). Following LPS challenge, the total *Rb-C9* expression was significantly enhanced at 12 h p.i. and gradually decreased at 24 h p.i. as compared with the un-injected control (Figure 64A). Finally, iridovirus challenge led to significantly up-regulated *Rb-C9* expression in liver at 3 h p.i. ($P < 0.05$), and the highest level was recorded at 6 h p.i. The pattern of transcriptional modulation in response to iridovirus challenge was similar to that induced by the Gram-negative bacteria challenge (Figure 64 B). A significant expression response for *Rb-C9* mRNA in peripheral blood cells were not recorded by any of the immune stimulants (data not shown).
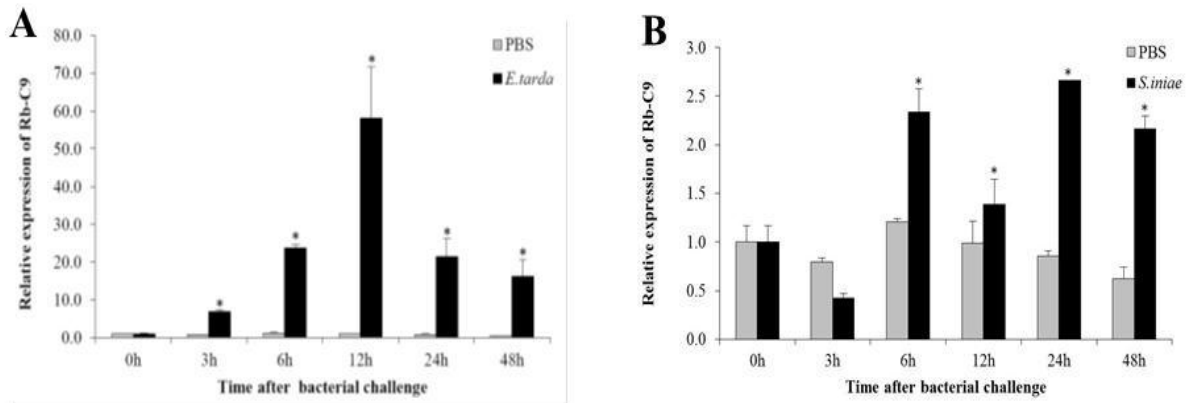
Figure 64. The temporal expression of liver Rb-C9 mRNA after rock bream challenge with (A) Gram-negative bacteria (*E. tarda*) (B) Gram-positive bacteria (*S. iniae*) and PBS (control) detected by qPCR

The amount of *Rb-C9* mRNA relative to the internal control β-actin is expressed as mean ± SD (n=3). Significant differences (*P* < 0.05 vs. un-injected control at 0 h) are marked with asterisks.
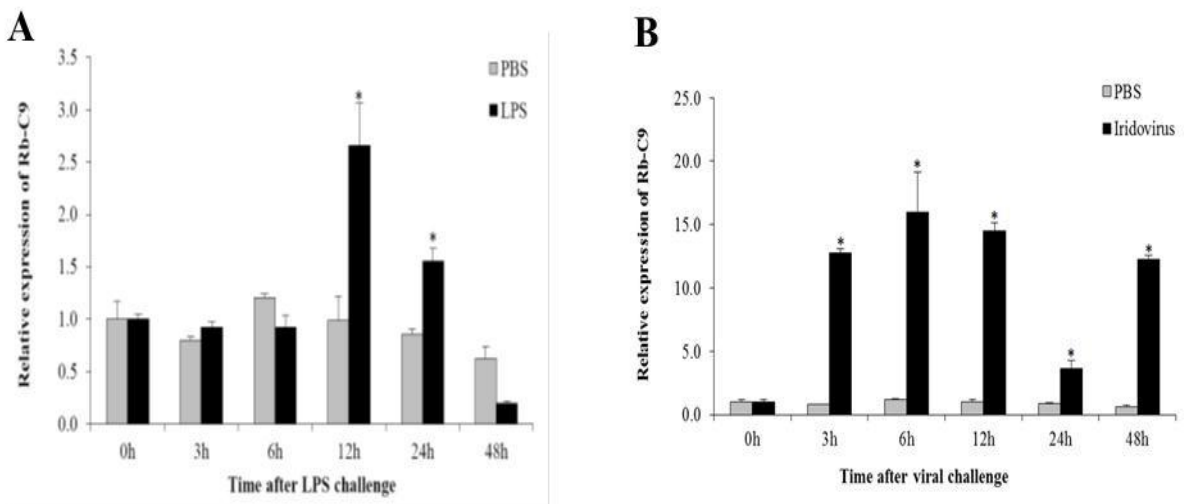
Figure 65. The temporal expression of liver Rb-C9 mRNA after rock bream challenge with (A) lipopolysaccharide (LPS) (B) iridovirus and PBS (control) detected by qPCR

The amount of Rb-C9 mRNA relative to the internal control β-actin is expressed as mean ± SD (n=3). Significant difference (*P* < 0.05 vs. un-injected control at 0 h) is marked with an asterisk.

176

5.2 Discussion

Although complement proteins have been intensively studied in multicellular organisms, less attention has been given to molecular characterization of genes involved in the lytic pathway of the complement system in teleost fish. However, the lytic pathway is an important effector mechanism of the complement system for destruction of invading pathogens. Assembly of the membrane attack complex on the surface of bacteria and other pathogenic organisms involves the sequential interaction of complement proteins C5b, C6, C7, C8, and C9 (Muller-Eberhard, 1986). Although MAC formation starts with assembly of the membrane-bound tetrameric complex, using the first four complement components, sequential binding of 12-18 C9 molecules is critical to the formation of a cylindrical transmembrane pore. Therefore, C9 is an important molecule involved in MAC that is responsible for weakening the target cell's membrane integrity and causing its lysis (Lovelace et al., 2011).

The characteristic domains of the TCC (TSP-1, LDL-A, MACPF, and EGF) (Chondrou et al., 2006; Li et al., 2007; Papanastasiou and Zarkadis, 2005) are present in rock bream C9. To date, several hundred MACPF-containing proteins have been identified in a variety of organisms; however, functions are known for only a few (Lovelace et al., 2011). Even though the MACPF proteins exhibit limited sequence similarity, they all contain the MACPF signature motif $((Y/W)G(T/S)H(F/Y)X_6GG)$ (Ponting, 1999). The MACPF signature motif was identified in *Rb-C9*. To the best of our knowledge, only a few fish C9 genes have been characterized so far, and all of them feature a second TSP-1 domain in the C-terminal region (Katagiri et al., 1999; Stanley and Herz, 1987; Yeo et al., 1997). Moreover, the TSP-1 domains all contain two C-mannosylation motifs (WXXWXXW). This feature appears to be unique to fish. The *Rb-C9* gene also contains all of these characteristic domains, and their relative positions are very similar to the other TCCs. However, the domain boundaries of sequence similarity between species in the TCC family have been shown to be

ambiguous (Katagiri et al., 1999). *Rb-C9* has four potential N-glycosylation sites (Asn-X-Ser/Thr), similar to that in rainbow trout (Chondrou et al., 2006). In contrast, the mammalian C9s have two N-glycosylation sites. For the cysteine residues, however, there was remarkably high conservation among all species evaluated in our study. Based on the collective results from our domain and motif analyses, *Rb-C9* was determined to be identical to the other fish orthologous.

The exon-intron structure of C9 has been conserved in all identified species (Li et al., 2007). While all C9s are composed of 11 exons and 10 introns, the introns vary in size between animal classes and genus. Human and mouse C9s are much larger than those in fish, and there is remarkably high variability in intron lengths between fish C9s. However, fish C9 shows a trend towards a smaller number of lengthy introns. For example, zebrafish C9 contains three large introns (1.8 kb, 2.4 kb and 2.5 kb) and grass carp has only one large intron (3.3 kb). All of the other introns were less than 1 kb in length. While *Rb-C9* showed the highest structural similarity to the C9 from puffer fish, the sizes of the 5th, 7th, and 9th exons were identical to those from all the fish species used in the comparison.

Although teleost research has recently focused on identification of complement genes, according to our knowledge, only one promoter analysis has been performed, namely in grass carp C9, which was based on the prediction method for TCC genes (Li et al., 2007). In the *Rb-C9* promoter region, a putative transcriptional initiation site and TATA box were found at 252 bp and 279 bp upstream of the start codon (ATG), respectively. This is in stark contrast to the human C9, which lacks the TATA box, but harbors an AT rich region at 53 bp upstream of the translational start codon (ATG) (Witzel-Schlomp et al., 2001). In grass carp, the TATA box was identified at 327 bp upstream of the start codon but no transcriptional initiation site (+1) was identified (Li et al., 2007). According to our analysis, several

important putative transcriptional binding sites were located in the 5′-flanking region of *Rb-C9*.

In humans, the C9 gene contains transcription factor binding sites for activator protein-2 (AP-2), yinyang-1 (YY-1), brain-2 (BRN-2), granulocyte macrophage colony stimulating factor (GM-CSF), interleukin-6 (IL-6) (Witzel-Schlomp et al., 2001). YY-1 is a *Gli-Kruppel*-type zinc finger protein that controls the transcription of a large number of viral and host cellular genes (Kim et al., 2006) and is evolutionarily conserves throughout all vertebrate lineages. However, YY-1 function differs as an activator, repressor or initiator of transcription (Witzel-Schlomp et al., 2001) depending on the sequence context of YY1-binding sites with respect to other regulator elements (Thomas and Seto, 1999). YY-1 is also known to interact with several key proteins, including TBP, TAFs, TFIIB and SP-1. Similarly, *Rb-C9* also contains putative binding sites for AP-2, YY-1, and SP-1, all of which may mediate transcription as in humans. In contrast, both rock bream and grass carp C9s contain putative binding sites for C/EBP, AP-1, OCT-1, and GATA-1. This may imply that regulation of fish C9 is marginally different from mammalian C9. Furthermore, a binding site for NF-κB, an important eukaryotic transcription factor, was found in *Rb-C9*. NF-κB is involved in the control of a large number of normal cellular and tissue-specific processes, such as immune and inflammatory responses, development, cellular growth, and apoptosis (Long et al., 2004; Pikarsky et al., 2004). Some nucleotide deletion studies have been demonstrated that SRY, along with CdxA and GATA, may act as major negative regulators of gene transcription (Dubin and Ostrer, 1994; Rath et al., 2008). The GATA transcription factor family has been found to participate directly in several signal-transduction pathways (Zhou and Ouyang, 2003). ICSBP, another putative binding factor of *Rb-C9*, belongs to the interferon regulatory factor (IRF) family of proteins that plays an important role in interferon-mediated immune response (Barnes et al., 2002). Likewise, Oct-1 is involved in regulation of

some T cell-specific cytokines expression such as IL-2, IL-4 and IL-5. In conclusion, the *Rb-C9* promoter contains several important putative transcriptional factor binding sites similar to those previously identified in C9s from human and grass carp. Moreover, most of these sites are associated with factors related to immunity, suggesting that *Rb-C9* may have an immune function, similar to that described for the C9s in human and grass carp. However, further studies based on empirical data are needed to delineate the functional elements of *Rb-C9* promoter.

In mammals, the liver is the primary production site for the majority of the complement components (Alper et al., 1980; Law SKA, 1988; Witzel-Schlomp et al., 2001). Moreover, the C9 expression patterns in teleost fish, such as grass carp (Li et al., 2007) and rainbow trout (Chondrou et al., 2006), are similar to those observed in rock bream. In grass carp, C9 was mainly detected in liver, and subsequently expressed in head kidney, spleen, skin, brain, and kidney respectively. In rainbow trout, C9 expression was more pronounced in liver, kidney, spleen, brain and intestine, but absent in heart. Therefore, our results agree with the notion that, in healthy fish, C9 transcripts are mainly expressed in liver and constitutively expressed in most other tissues. In our study, we also observed that the liver expression of *Rb-C9* was extremely high compared to the other tissues and responsive to immune stimulation, suggesting that *Rb-C9* is similar to most complement components, which belongs to the type of acute-phase proteins and are generally of hepatic origin (Witzel-Schlomp et al., 2001). This is further confirmed by the presence the liver specific transcription factor binding sites such as HNF-1, HNF-3, and C/EBP in the promoter region (Dogra and May, 1997). However, *Rb-C9* expression in gill, another highly immune reactive tissue, was very low. Intriguingly, C9 expression was found to be reduced in gill of both grass carp and rainbow trout. Since a previous study of murine C9-responsiveness to LPS

challenge not revealed any expression of C9 transcripts in the lungs of healthy mice (Bolger et al., 2007), and it confirmed that *Rb-C9* also behaves in a similar manner to other C9 genes.

According to the immune-responsive expression results, Gram-negative bacteria can effectively trigger expression of *Rb-C9* in liver at 3 h to 12 h p.i. Transcription level was highest at 12 h p.i. compared to the *un-injected* controls and produced the highest fold-induction among all the immune stimulants. A previous study conducted in grass carp, also found that C9 expression was induced in liver by Gram-negative bacteria (*Flavobacterium columnare*) at 12 h p.i. (Li et al., 2007). Although the *Rb-C9* was highly induced by the live bacterial challenge, it showed only a very low response in liver to the purified LPS immune-modulator (2.6-fold). To the best of our knowledge, only a few studies of C9 orthologous have been investigated under pathogen-induced expression conditions (Law SKA, 1988). Therefore, lack of other related data on induced expression of C9 make extra effort to explain our data on rock bream. In peripheral blood cells, *Rb-C9* expression was not affected by bacterial or viral challenges. Similar results were reported for C5 mRNA expression in common carp infected with *Ichthyophthirius multifiliis* (Gonzalez et al., 2003). Following viral challenge, *Rb-C9* mRNA was highly up-regulated at 3 h p.i. and thereafter showed a cyclic expression pattern. In contrast, microarray (GSE19049) and qPCR array (GSE19503) studies in virus-challenged zebrafish showed that, expression of C9 transcripts were increased in both fins and organs (liver, spleen and head kidney) by 15.33- and 2.71-fold, respectively, at 48 h post challenge with viral hemorrhagic septicemia virus (VHSV) (Encinas et al., 2010). Therefore, we hypothesized that, *Rb-C9* may be involved in the virus-mediated immune response of rock bream. As a whole, TCC may be activated in protein level as a response to the microbial infections (Li et al., 2007). However, the regulatory mechanism underlying the transcriptional activation of TCC genes are still unclear. In our study, the up-regulated transcriptional profile of *Rb-C9* upon Gram-negative bacteria and iridovirus challenge can be

substantiated through the presence of immune signaling transcription factor binding sites in the promoter region such as NF-κB, AP-1, and ICSBP(virus specific) which are expressed through toll-like receptors dependent immune signaling pathways (Hawlisch and Kohl, 2006; Martin and Wesche, 2002) Furthermore, intense expression of C9 transcript level in liver and above speculated, activation of multiple defense pathways can be correlated with the excessive requirement to C9 molecules to form MAC (12-16 molecules per one MAC) (Nakao et al., 1996)

In summary, we have identified the full-length genome sequence of complement component 9 from rock bream. Genome structure analysis of *Rb-C9* was confirmed its substantial similarity with C9 of other teleost fish and showed that it belongs to the TCC protein family. Moreover, *Rb-C9* was found to be potentially induced by infection with *E. tarda*, *S. iniae,* LPS and iridovirus, indicating its potential role in the rock bream immune defense system. However, the mechanisms that which *Rb-C9* mediates immune response to acute-phase infection require further study.

6. General discussion

The complement system is consisted of more than 35 proteins, which belong to both acute phase glycoproteins and cell surface bound proteins. When considering the major 9 components, most of them present in execution phase and lytic pathway, share common domain architecture. The structurally dissimilar complements within the 9major components are complement C1, C2, C3, C6 and C8γ. Furthermore, many of the complement factors may have at least two isoforms. In our study, we identified, three isoforms of complement C1q, four isoforms of complement C3, two isoforms of C7. The two isoforms of complement component 3 were found to be occurred sequentially in rock bream genome. Similarly,

complement C6 and complement C7 were also found in another chromosome as consecutive genes.
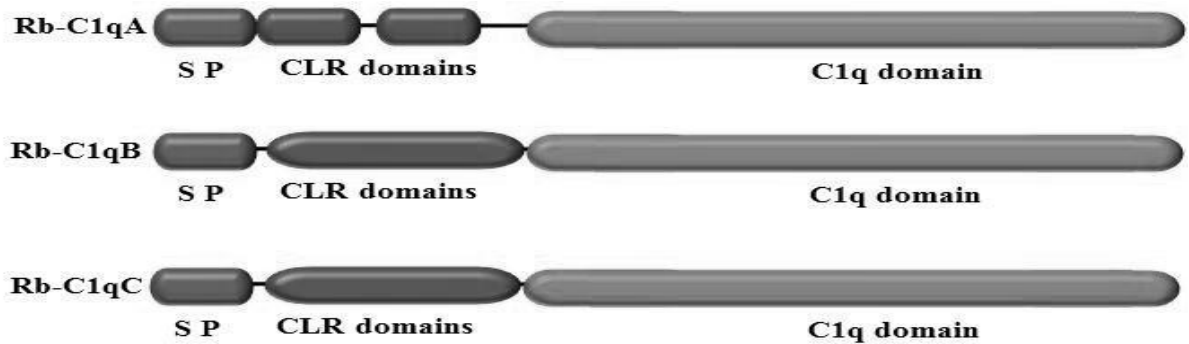
Figure 66. Schematic representations of three isoforms of complement C1q genes

In higher vertebrates, anaphylatoxins, C3a, C4a and C5a are highly cationic peptides characterized by cysteine residues that form three intra-chain disulfide bonds imparting significant chemical and physical stability to the peptides. Hence, recombinant Rb-C3a, Rb-C4a and Rb-C5a were able to induce respiratory burst activity in immune cells like macrophages and other immune cells.
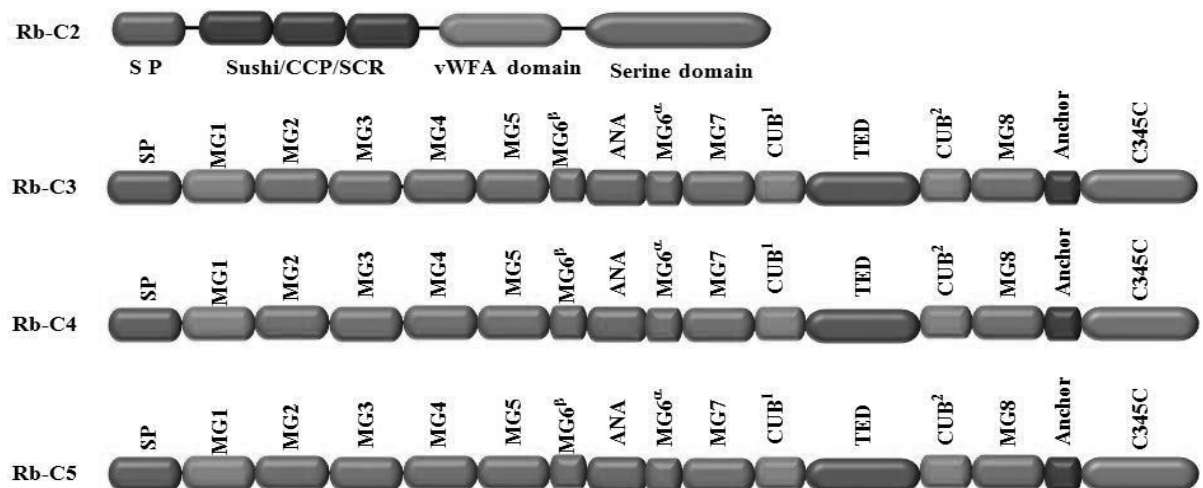
Figure 67. Schematic representation of complement C2, C3, C4 and C5 genes

The MACPF domains of the rock bream genes also harbors low sequential similarity compared to similarity between orthologous counterparts (Figure 69). However, all of them contain characteristic MACPF signature ((Y/W)G(T/S)H(F/Y)X$_6$GG). The molecular mass and number of amino acid are varied between 37 - 41 kDa of molecular weight and 320 – 362 of amino acid respectively. In MACPF signature motif of *Rb-C8α*, tyrosine (Y) is replaced with phenylalanine (F), deviating from the common signature motif of TCC proteins. When consider domains of the lytic pathway genes, MACPF, TSP-1, LDLR-A, EGF-1 are common for all except C8γ. However, numbers of domains in each gene were varied among them. To the best of our knowledge, only a few fish C9 genes have been characterized so far, and all of them have a second TSP-1 domain in the C-terminal region [21,23,38]. Moreover, the all TSP-1 domains contain two C-mannosylation motifs (WXXWXXW). This feature appears to be unique to fish.
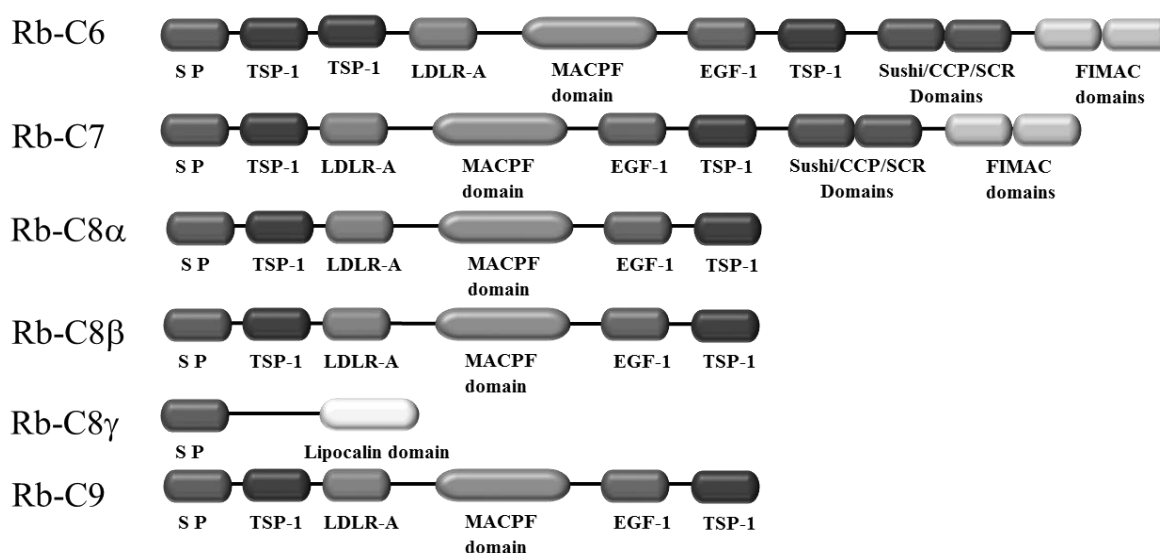
Figure 68. Domain organization of the lytic pathway genes of the complement system of the rock bream
Abbreviations: SP; signal peptide, TSP-1; thrombospondin type I domain, LDLR-A; low-density lipoprotein receptor domain class A, MACPF; The membrane attack complex/perforin EGF-1 like domain, Sushi/CCP/SCR domain; complement control protein/ short consensus repeats

```
                     *         20          *         40          *         60          *
Rb-C8α : -REDKCSTLLSIPGT-ERGTQGYNVLTSEFVDHVLDPRYFGGKCEYVYNGEWRKFNYDAFCENLHYNEDEK :  69
Rb-C8β : ASKPCRETVEEYWGI-ENLAKGINILNSNLEGVVLDNRYYAGSCLPHYIQNVR------------------ :  52
Rb-C7  : -SHYPCDLDKTPPNS-DVTGRGYNVLTGKLRAGVINTLSFGGQCRKVFSGDHKVY--------------- :  53
Rb-C9  : VRKPCGALVLDTNEQGRTAGYGINVLGADPRMNPFNNDYENGRCDNVRNPNTGQR--------------- :  55
Rb-C6  : -FKIVCPAEKRVAPGADLVGNGFDALAEEPRGAVLDNMFMGGICNLKRPRSTLLY--------------- :  54
                 G  1 L          v 1       G C

                    80          *        100          *        120          *        140
Rb-C8α : NYRKPYNYHAYRFVAEATSEGSQEYFDDIVSLLRARKGMRSSNGGVTFG-------------IYHVEVGLS : 127
Rb-C8β : -FRKPYNLQQYTLQTKGSYDFTAQSFETYSDYMDHTAKERATQTVVSIGFGIPG--------LFEFGFNFN : 114
Rb-C7  : -YRLPQNILRYNFEVKVDNDESDESYESSWSYTQHIQAN--------------------------ALV :  94
Rb-C9  : -DRLPWNIGVLNYETLVEETVSREIYEDTHSLLREMLKEMSFKIDVGLSFKFSPSEQSLSNVSLKLGVDHQ : 125
Rb-C6  : -HRIPHNFESENIKVGVVEDFSTEPRSLHTEPISVKTSHSSGDSQQDGNHGFFLFFP------LFIYGGRS : 118
                 R P N                3 2

                     *        160          *        180          *        200          *
Rb-C8α : GSQESDFVHNITRHKSEDLGFVRLSSKVQTAHFKMRS-NKLMLHEDFYISLMELPEQYDFGMYSRFFNTFG : 197
Rb-C8β : NAKNIRSVKKIRRASSKTNSFVRAKAELELAQYMLKS-DDLMLHPEFLQRLRSLPQSYVYGEYRQIYRDYG : 184
Rb-C7  : GHDRRTFHKELTDN--KAHKLLILKNKVELAQFQNSAPQYLTLAEGFWKALSSLPLTYDYSAYRQLFQTYG : 163
Rb-C9  : YEKKTMIRDVTETSNVKNKSFMRVKGRVQLSTYRLRT-RDLQVANEELEHVKSLPLQYEKGIYFAFLEDYG : 195
Rb-C6  : RTQISSNKEAFEASKKTDSKFIRVHQVLPISTFKVRDPGDLVLSQPFLQFLHALPLEYNYALYRDIFQRFG : 189
                 f6r         6         5         L 6     F    6   LP Y    Y        5G

                    220          *        240          *        260          *        280
Rb-C8α : THYVTEGTMGGTLEYVVVVNKTSMAESKLEGEQAGRCLGASLGLSIPIGFVTVNPKVGLGPFCGKEGAFDG : 268
Rb-C8β : THYITEAGLGGQYEHTIILNKERLARTDYTLDDYKRCTQAGLKVGANIQGVYLSLGAHGGSCNGLLNEMGE : 255
Rb-C7  : THYLSEGSLGGEYQGLLELDRQALEWTSTTNIDYQRCWRK-----VKRRLFRKKVKTTCEKLTKALSSRDG : 229
Rb-C9  : THYTRNCKSGGEYELIYVLDQDAIRQKNLTERKIQECVKLGITAELGGIPFTEAKGKLGLDSCEPVTTKKK : 266
Rb-C6  : THYYGSGTLGGIYDLLYQYSREELRSSGETEEHVTGCLARDLT-----WTVILYTERRSVTRCTNNRMTER : 255
                 THY    g  GG y           6      t        C

                     *        300          *        320          *        340          *
Rb-C8α : ATHSSSRVIEDIITLVKGGITHSSSGLLAIRNP----DTYRN--WGAILKYNPTLIEYETMPIYELVRLST : 333
Rb-C8β : DTERGS-MVEDFVAVVKGGSSESITALVSKKLP----TPQLMGLWGEGVRFNPDFIRSTTRPLYELVTSRD : 321
Rb-C7  : HNVNKISIKVNVIGGDPALTYTLTSLDVENPEA----NGQSYDNWASSVKDFPEVTDQKLRPLYELVKEVQ : 296
Rb-C9  : DETHGKAVVDKVMTSVKGGTLQTAVAMRAKLNKEGVMDITTYQNWARSITDAPALINSQPEPIYTLIPLDM : 337
Rb-C6  : YEGSYVKAAEKSFSMVKGGRIREAAALAWERRG-STPDQTSFKNWAKSILDNPVVVDCKVLPIIDLVQGIP : 325
                 vkgg                          W   6   P          P6y L6

                    360          *        380
Rb-C8α : AADHIGARLANLQRGWDEYLQQFDSCR : 360
Rb-C8β : FSQDATLKRN-LKRALSEYLAEASSCR : 347
Rb-C7  : CAG---LKKLHLKRATEEYLAEEHPCH : 320
Rb-C9  : PDAN--TRISNLKQATADYVAEYSVCK : 362
Rb-C6  : CAAT---KRRHLRKALLRYQEEFDTCK : 349
                 4   L  a   Y 2  C
```

Figure 69. Multiple sequence alignment of MACPF domains of lytic pathway genes.

7. Conclusion

In the present study, we demonstrated the existence and evolutionary preservation of nine major consecutive complement components, two sub components and two isoforms (total of 13 genes) at cDNA and genomic sequence levels within one teleost species; rock bream, *Oplegnathus fasciatus*. The study portrays a clear picture of diversity of complement components within the teleost complement system. According to our knowledge, this is the first report on the molecular characterization and expression analysis of major complement components in a non-mammalian counterpart, teleost fish. Furthermore, employed experimental methods provide complete and sound understanding of the immunological role of complement system as it summarized the *in vivo* mRNA expression of each gene upon diverse immune stimulants/ pathogen challenges and the recombinant proteins were determined to generate reactive oxygen species *in vitro*. In addition, the biological activities of the recombinant proteins of key components demonstrate their ability to generate inflammatory reaction in live cells *in vitro*, providing insights into the functional role of the genes in rock bream.

Finally, the present approach in comprehensive study of complement system in rock bream fish, *Oplegnathus fasciatus* will be important to understand its role in relation to survival of fish from challenging environment with abundant pathogenic threats and make significant progress into the immunological perspectives of the complement system in teleost.

8. Bibliography

Abelseth, T.K., Stensvag, K., Espelid, S., Nygaard, R., Ellingsen, T., Bogwald, J., Dalmo, R.A., 2003. The spotted wolffish (*Anarhichas minor* Olafsen) complement component C3: isolation, characterisation and tissue distribution. Fish & Shellfish Immunology 15, 13-27.

Alberti, S., Marques, G., Camprubi, S., Merino, S., Tomas, J.M., Vivanco, F., Benedi, V.J., 1993. C1q binding and activation of the complement classical pathway by *Klebsiella pneumoniae* outer membrane proteins. Infection and Immunity 61, 852-860.

Alper, C.A., Raum, D., Awdeh, Z.L., Petersen, B.H., Taylor, P.D., Starzl, T.E., 1980. Studies of hepatic synthesis *in vivo* of plasma proteins, including orosomucoid, transferrin, alpha 1-antitrypsin, C8, and factor B. Clinical Immunology and Immunopathology 16, 84-89.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403-410.

Aoki, T., Hirono, I., 2006. Immune relevant genes of Japanese flounder, *Paralichthys olivaceus*. Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics 1, 115-121.

Ariki, S., Takahara, S., Shibata, T., Fukuoka, T., Ozaki, A., Endo, Y., Fujita, T., Koshiba, T., Kawabata, S., 2008. Factor C acts as a lipopolysaccharide-responsive C3 convertase in horseshoe crab complement activation. Journal of Immunology 181, 7994-8001.

Arlaud, G.J., Colomb, M.G., 2001. Complement: Classical Pathway, eLS. John Wiley & Sons, Ltd.

Arlaud, G.J., Gaboriaud, C., Thielens, N.M., Rossi, V., Bersch, B., Hernandez, J.F., Fontecilla-Camps, J.C., 2001. Structural biology of C1: dissection of a complex molecular machinery. Immunological Reviews 180, 136-145.

Armbrust, T., Nordmann, B., Kreissig, M., Ramadori, G., 1997. C1Q synthesis by tissue mononuclear phagocytes from normal and from damaged rat liver: up-regulation by dexamethasone, down-regulation by interferon gamma, and lipopolysaccharide. Hepatology 26, 98-106.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

Bairoch, A., Bucher, P., Hofmann, K., 1997. The PROSITE database, its status in 1997. Nucleic Acids Research 25, 217-221.

Banyai, L., Patthy, L., 1999. The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. Protein Science : a publication of the Protein Society 8, 1636-1642.

Barnes, B., Lubyova, B., Pitha, P.M., 2002. On the role of IRF in host defense. Journal of Interferon & Cytokine research : The Official Journal of the International Society for Interferon and Cytokine Research 22, 59-71.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340, 783-795.

Biesecker, G., Muller-Eberhard, H.J., 1980. The ninth component of human complement: purification and physicochemical characterization. Journal of Immunology 124, 1291-1296.

Blandin, S., Levashina, E.A., 2004. Thioester-containing proteins and insect immunity. Molecular Immunology 40, 903-908.

Bohlson, S.S., Fraser, D.A., Tenner, A.J., 2007. Complement proteins C1q and MBL are pattern recognition molecules that signal immediate and long-term protective immune functions. Molecular Immunology 44, 33-43.

Bolger, M.S., Ross, D.S., Jiang, H., Frank, M.M., Ghio, A.J., Schwartz, D.A., Wright, J.R., 2007. Complement levels and activity in the normal and LPS-injured lung. American Journal of Physiology. Lung Cellular and Molecular Physiology 292, L748-759.

Bordet, J., Gengou, O., 1901. Sur l'existence de substances sensibilisatrices dans la plupart des serum antimicrobiens. Ann. Inst. Pasteur 15, 289−302.

Boshra, H., Gelman, A.E., Sunyer, J.O., 2004. Structural and functional characterization of complement C4 and C1s-like molecules in teleost fish: insights into the evolution of classical and alternative pathways. Journal of Immunology 173, 349-359.

Boshra, H., Li, J., Sunyer, J.O., 2006. Recent advances on the complement system of teleost fish. Fish & Shellfish Immunology 20, 239-262.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72, 248-254.

Carland, T.M., Gerwick, L., 2010. The C1q domain containing proteins: Where do they come from and what do they do? Developmental and Comparative Immunology 34, 785-790.

Carland, T.M., Locke, J.B., Nizet, V., Gerwick, L., 2012. Differential expression and intrachromosomal evolution of the sghC1q genes in zebrafish (*Danio rerio*). Developmental and Comparative Immunology 36, 31-38.

Carroll, M.C., Campbell, R.D., Bentley, D.R., Porter, R.R., 1984. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. Nature 307, 237-241.

Carter, P.E., Duponchel, C., Tosi, M., Fothergill, J.E., 1991. Complete nucleotide sequence of the gene for human C1 inhibitor with an unusually high density of Alu elements. European Journal of Biochemistry / FEBS 197, 301-308.

Castellano, G., Woltman, A.M., Nauta, A.J., Roos, A., Trouw, L.A., Seelen, M.A., Schena, F.P., Daha, M.R., van Kooten, C., 2004. Maturation of dendritic cells abrogates C1q production *in vivo* and in vitro. Blood 103, 3813-3820.

Castillo, M.G., Goodson, M.S., McFall-Ngai, M., 2009. Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. Developmental and Comparative Immunology 33, 69-76.

Chondrou, M.P., Mastellos, D., Zarkadis, I.K., 2006. cDNA cloning and phylogenetic analysis of the sixth complement component in rainbow trout. Molecular Immunology 43, 1080-1087.

Corallini, F., Bossi, F., Gonelli, A., Tripodo, C., Castellino, G., Mollnes, T.E., Tedesco, F., Rizzi, L., Trotta, F., Zauli, G., Secchiero, P., 2009. The soluble terminal complement complex (SC5b-9) up-regulates osteoprotegerin expression and release by endothelial cells: implications in rheumatoid arthritis. Rheumatology 48, 293-298.

de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research 34, W362-365.

Depboylu, C., Schafer, M.K., Schwaeble, W.J., Reinhart, T.A., Maeda, H., Mitsuya, H., Damadzic, R., Rausch, D.M., Eiden, L.E., Weihe, E., 2005. Increase of C1q biosynthesis in brain microglia and macrophages during lentivirus infection in the rhesus macaque is sensitive to antiretroviral treatment with 6-chloro-2',3'-dideoxyguanosine. Neurobiology of Disease 20, 12-26.

DiScipio, R.G., 1992. Formation and structure of the C5b-7 complex of the lytic pathway of complement. The Journal of Biological Chemistry 267, 17087-17094.

DiScipio, R.G., Chakravarti, D.N., Muller-Eberhard, H.J., Fey, G.H., 1988. The structure of human complement component C7 and the C5b-7 complex. The Journal of Biological Chemistry 263, 549-560.

Dishaw, L.J., Smith, S.L., Bigger, C.H., 2005. Characterization of a C3-like cDNA in a coral: phylogenetic implications. Immunogenetics 57, 535-548.

Dodds, A.W., Petry, F., 1993. The phylogeny and evolution of the first component of complement, C1. Behring Institute Mitteilungen, 87-102.

Dodds, A.W., Ren, X.D., Willis, A.C., Law, S.K., 1996. The reaction mechanism of the internal thioester in the human complement component C4. Nature 379, 177-179.

Dodds, A.W., Smith, S.L., Levine, R.P., Willis, A.C., 1998. Isolation and initial characterisation of complement components C3 and C4 of the nurse shark and the channel catfish. Developmental and Comparative Immunology 22, 207-216.

Dogra, S.C., May, B.K., 1997. Liver-enriched transcription factors, HNF-1, HNF-3, and C/EBP, are major contributors to the strong activity of the chicken CYP2H1 promoter in chick embryo hepatocytes. DNA and Cell Biology 16, 1407-1418.

Dubin, R.A., Ostrer, H., 1994. Sry is a transcriptional activator. Mol Endocrinol 8, 1182-1192.

Ehrengruber, M.U., Geiser, T., Deranleau, D.A., 1994. Activation of human neutrophils by C3a and C5A. Comparison of the effects on shape changes, chemotaxis, secretion, and respiratory burst. FEBS Letters 346, 181-184.

Elsner, J., Oppermann, M., Czech, W., Dobos, G., Schopf, E., Norgauer, J., Kapp, A., 1994a. C3a activates reactive oxygen radical species production and intracellular calcium transients in human eosinophils. European Journal of Immunology 24, 518-522.

Elsner, J., Oppermann, M., Czech, W., Kapp, A., 1994b. C3a activates the respiratory burst in human polymorphonuclear neutrophilic leukocytes via pertussis toxin-sensitive G-proteins. Blood 83, 3324-3331.

Encinas, P., Rodriguez-Milla, M.A., Novoa, B., Estepa, A., Figueras, A., Coll, J., 2010. Zebrafish fin immune responses during high mortality infections with viral haemorrhagic septicemia rhabdovirus. A Proteomic and Transcriptomic Approach. BMC Genomics 11, 518.

Fearon, D.T., Locksley, R.M., 1996. The instructive role of innate immunity in the acquired immune response. Science 272, 50-53.

Franchini, S., Zarkadis, I.K., Sfyroera, G., Sahu, A., Moore, W.T., Mastellos, D., LaPatra, S.E., Lambris, J.D., 2001. Cloning and purification of the rainbow trout fifth component of complement (C5). Developmental and Comparative Immunology 25, 419-430.

Fredslund, F., Laursen, N.S., Roversi, P., Jenner, L., Oliveira, C.L., Pedersen, J.S., Nunn, M.A., Lea, S.M., Discipio, R., Sottrup-Jensen, L., Andersen, G.R., 2008. Structure of and influence of a tick complement inhibitor on human complement component 5. Nature Immunology 9, 753-760.

Fujito, N.T., Sugimoto, S., Nonaka, M., 2010. Evolution of thioester-containing proteins revealed by cloning and characterization of their genes from a cnidarian sea anemone, *Haliplanella lineate*. Developmental and Comparative Immunology 34, 775-784.

Gadjeva, M., Dodds, A.W., Taniguchi-Sidle, A., Willis, A.C., Isenman, D.E., Law, S.K., 1998. The covalent binding reaction of complement component C3. Journal of Immunology 161, 985-990.

Gal, P., Barna, L., Kocsis, A., Zavodszky, P., 2007. Serine proteases of the classical and lectin pathways: similarities and differences. Immunobiology 212, 267-277.

Gershov, D., Kim, S., Brot, N., Elkon, K.B., 2000. C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. The Journal of Experimental Medicine 192, 1353-1364.

Gestal, C., Pallavicini, A., Venier, P., Novoa, B., Figueras, A., 2010. MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. Developmental and Comparative Immunology 34, 926-934.

Giacomelli, S., Melillo, D., Lambris, J.D., Pinto, M.R., 2012. Immune competence of the *Ciona intestinalis* pharynx: complement system-mediated activity. Fish & Shellfish Immunology 33, 946-952.

Gonzalez, S., Martinez-Borra, J., Lopez-Larrea, C., 2003. Cloning and characterization of human complement component C7 promoter. Genes and Immunity 4, 54-59.

Gonzalez, S.F., Buchmann, K., Nielsen, M.E., 2007a. Complement expression in common carp (*Cyprinus carpio* L.) during infection with *Ichthyophthirius multifiliis*. Developmental and Comparative Immunology 31, 576-586.

Gonzalez, S.F., Chatziandreou, N., Nielsen, M.E., Li, W., Rogers, J., Taylor, R., Santos, Y., Cossins, A., 2007b. Cutaneous immune responses in the common carp detected using transcript analysis. Molecular Immunology 44, 1664-1679.

Graham, M., Shin, D.H., Smith, S.L., 2009. Molecular and expression analysis of complement component C5 in the nurse shark (*Ginglymostoma cirratum*) and its predicted functional role. Fish & Shellfish Immunology 27, 40-49.

Hawlisch, H., Kohl, J., 2006. Complement and Toll-like receptors: key regulators of adaptive immune responses. Molecular Immunology 43, 13-21.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L., Kolchanov, N.A., 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Research 26, 362-367.

Hobart, M.J., Fernie, B.A., DiScipio, R.G., 1995. Structure of the human C7 gene and comparison with the C6, C8A, C8B, and C9 genes. Journal of Immunology 154, 5188-5194.

Horiuchi, T., Macon, K.J., Kidd, V.J., Volanakis, J.E., 1990. Translational regulation of complement protein C2 expression by differential utilization of the 5'-untranslated region of mRNA. The Journal of Biological Chemistry 265, 6521-6524.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H., Yang, L., Boley, N., Andrews, J., Kaufman, T.C., Graveley, B.R., Bickel, P.J., Carninci, P., Carlson, J.W., Celniker, S.E., 2011. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. Genome Research 21, 182-192.

Hu, Y.L., Pan, X.M., Xiang, L.X., Shao, J.Z., 2010. Characterization of C1q in teleosts: insight into the molecular and functional evolution of C1q family and classical pathway. The Journal of Biological Chemistry 285, 28777-28786.

Innamorati, G., Bianchi, E., Whang, M.I., 2006. An intracellular role for the C1q-globular domain. Cellular Signalling 18, 761-770.

Janssen, B.J., Huizinga, E.G., Raaijmakers, H.C., Roos, A., Daha, M.R., Nilsson-Ekdahl, K., Nilsson, B., Gros, P., 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437, 505-511.

Jensen, J.A., Festa, E., Smith, D.S., Cayer, M., 1981. The complement system of the nurse shark: hemolytic and comparative characteristics. Science 214, 566-569.

Juan, T.S., Wilson, D.R., Wilde, M.D., Darlington, G.J., 1993. Participation of the transcription factor C/EBP delta in the acute-phase regulation of the human gene for complement component C3. Proceedings of the National Academy of Sciences of the United States of America 90, 2584-2588.

189

Katagiri, T., Hirono, I., Aoki, T., 1999. Molecular analysis of complement component C8beta and C9 cDNAs of Japanese flounder, Paralichthys olivaceus. Immunogenetics 50, 43-48.

Kato, Y., Nakao, M., Mutsuro, J., Zarkadis, I.K., Yano, T., 2003. The complement component C5 of the common carp (*Cyprinus carpio*): cDNA cloning of two distinct isotypes that differ in a functional site. Immunogenetics 54, 807-815.

Kazantzi, A., Sfyroera, G., Holland, M.C., Lambris, J.D., Zarkadis, I.K., 2003. Molecular cloning of the beta subunit of complement component eight of rainbow trout. Developmental and Comparative Immunology 27, 167-174.

Kim, J.D., Hinz, A.K., Bergmann, A., Huang, J.M., Ovcharenko, I., Stubbs, L., Kim, J., 2006. Identification of clustered YY1 binding sites in imprinting control regions. Genome Research 16, 901-911.

Kimura, A., Ikeo, K., Nonaka, M., 2009. Evolutionary origin of the vertebrate blood complement and coagulation systems inferred from liver EST analysis of lamprey. Developmental and Comparative Immunology 33, 77-87.

Kishore, U., Gaboriaud, C., Waters, P., Shrive, A.K., Greenhough, T.J., Reid, K.B., Sim, R.B., Arlaud, G.J., 2004a. C1q and tumor necrosis factor superfamily: modularity and versatility. Trends in Immunology 25, 551-561.

Kishore, U., Ghai, R., Greenhough, T.J., Shrive, A.K., Bonifati, D.M., Gadjeva, M.G., Waters, P., Kojouharova, M.S., Chakraborty, T., Agrawal, A., 2004b. Structural and functional anatomy of the globular domain of complement protein C1q. Immunology Letters 95, 113-128.

Kishore, U., Reid, K.B., 1999. Modular organization of proteins containing C1q-like globular domain. Immunopharmacology 42, 15-21.

Kishore, U., Reid, K.B., 2000. C1q: structure, function, and receptors. Immunopharmacology 49, 159-170.

Kojouharova, M.S., Tsacheva, I.G., Tchorbadjieva, M.I., Reid, K.B., Kishore, U., 2003. Localization of ligand-binding sites on human C1q globular head region using recombinant globular head fragments and single-chain antibodies. Biochimica et biophysica acta 1652, 64-74.

Kong, P., Zhang, H., Wang, L., Zhou, Z., Yang, J., Zhang, Y., Qiu, L., Wang, L., Song, L., 2010. AiC1qDC-1, a novel gC1q-domain-containing protein from bay scallop Argopecten irradians with fungi agglutinating activity. Developmental and comparative immunology 34, 837-846.

Korotaevskiy, A.A., Hanin, L.G., Khanin, M.A., 2009. Non-linear dynamics of the complement system activation. Mathematical biosciences 222, 127-143.

Kuroda, N., Naruse, K., Shima, A., Nonaka, M., Sasaki, M., 2000. Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish. Immunogenetics 51, 117-128.

Kuroda, N., Wada, H., Naruse, K., Simada, A., Shima, A., Sasaki, M., Nonaka, M., 1996. Molecular cloning and linkage analysis of the Japanese medaka fish complement Bf/C2 gene. Immunogenetics 44, 459-467.

Kusuda, R., Salati, F., 1993. Major bacterial diseases affecting mariculture in Japan. Annual Review of Fish Diseases 3, 69-85.

Lambris, J.D., 1988. The multifunctional role of C3, the third component of complement. Immunology Today 9, 387-393.

Lange, S., Bambir, S., Dodds, A.W., Magnadottir, B., 2004. An immunohistochemical study on complement component C3 in juvenile Atlantic halibut (*Hippoglossus hippoglossus* L.). Developmental and Comparative Immunology 28, 593-601.

Lange, S., Dodds, A.W., Gudmundsdottir, S., Bambir, S.H., Magnadottir, B., 2005. The ontogenic transcription of complement component C3 and Apolipoprotein A-I tRNA in Atlantic cod (*Gadus morhua* L.)--a role in development and homeostasis? Developmental and Comparative Immunology 29, 1065-1077.

Lao, H.H., Sun, Y.N., Yin, Z.X., Wang, J., Chen, C., Weng, S.P., He, W., Guo, C.J., Huang, X.D., Yu, X.Q., He, J.G., 2008. Molecular cloning of two C1q-like cDNAs in mandarin fish *Siniperca chuatsi*. Veterinary Immunology and Immunopathology 125, 37-46.

Law, S.K., Dodds, A.W., 1997. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. Protein Science : a publication of the Protein Society 6, 263-274.

Law SKA, R.K., 1988. Complement, in: D, M. (Ed.), In Focus Series IRL Press, Oxford, p. 82.

Lebioda L, S.J., 2005. Complement protein C8, in: Lambris D, M.J. (Ed.), Structural Biology of the Complement System. CRC Press, p. 233.

Letunic, I., Doerks, T., Bork, P., 2009. SMART 6: recent updates and new developments. Nucleic Acids Research 37, D229-232.

Li, L., Chang, M.X., Nie, P., 2007. Molecular cloning, promoter analysis and induced expression of the complement component C9 gene in the grass carp *Ctenopharyngodon idella*. Veterinary Immunology and Immunopathology 118, 270-282.

Long, F., Liu, H., Hahn, C., Sumazin, P., Zhang, M.Q., Zilberstein, A., 2004. Genome-wide prediction and analysis of function-specific transcription factor binding sites. In Silico Biology 4, 395-410.

Lovelace, L.L., Cooper, C.L., Sodetz, J.M., Lebioda, L., 2011. Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. The Journal of Biological Chemistry 286, 17585-17592.

Lovoll, M., Johnsen, H., Boshra, H., Bogwald, J., Sunyer, J.O., Dalmo, R.A., 2007. The ontogeny and extrahepatic expression of complement factor C3 in Atlantic salmon (*Salmo salar*). Fish & Shellfish Immunology 23, 542-552.

Lovoll, M., Kilvik, T., Boshra, H., Bogwald, J., Sunyer, J.O., Dalmo, R.A., 2006. Maternal transfer of complement components C3-1, C3-3, C3-4, C4, C5, C7, Bf, and Df to offspring in rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 58, 168-179.

Lu, J.H., Teh, B.K., Wang, L., Wang, Y.N., Tan, Y.S., Lai, M.C., Reid, K.B., 2008. The classical and regulatory functions of C1q in immunity and autoimmunity. Cellular & Molecular Immunology 5, 9-21.

Maina, C.V., Riggs, P.D., Grandea, A.G., 3rd, Slatko, B.E., Moran, L.S., Tagliamonte, J.A., McReynolds, L.A., Guan, C.D., 1988. An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene 74, 365-373.

Malhotra, R., Lu, J., Holmskov, U., Sim, R.B., 1994. Collectins, collectin receptors and the lectin pathway of complement activation. Clinical and Experimental Immunology 97 Suppl 2, 4-9.

Martin, H., Kaul, M., Loos, M., 1990. Disulfide bridge formation between C1q and IgG in vitro. European Journal of Immunology 20, 1641-1645.

Martin, M.U., Wesche, H., 2002. Summary and comparison of the signaling mechanisms of the Toll/interleukin-1 receptor family. Biochimica et Biophysica Acta 1592, 265-280.

Matsushita, M., Fujita, T., 1995. Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. Immunobiology 194, 443-448.

Matsushita, M., Matsushita, A., Endo, Y., Nakata, M., Kojima, N., Mizuochi, T., Fujita, T., 2004. Origin of the classical complement pathway: Lamprey orthologue of mammalian C1q acts as a lectin. Proceedings of the National Academy of Sciences of the United States of America 101, 10127-10131.

Mehlhop, E., Diamond, M.S., 2006. Protective immune responses against West Nile virus are primed by distinct complement activation pathways. The Journal of Experimental Medicine 203, 1371-1381.

Mei, J., Gui, J., 2008. Bioinformatic identification of genes encoding C1q-domain-containing proteins in zebrafish. Journal of Genetics and Genomics = Yi chuan xue bao 35, 17-24.

Mevorach, D., Mascarenhas, J.O., Gershov, D., Elkon, K.B., 1998. Complement-dependent clearance of apoptotic cells by human macrophages. The Journal of Experimental Medicine 188, 2313-2320.

Minta, J.O., Fung, M., Turner, S., Eren, R., Zemach, L., Rits, M., Goldberger, G., 1998. Cloning and characterization of the promoter for the human complement factor I (C3b/C4b inactivator) gene. Gene 208, 17-24.

Mintz, C.S., Arnold, P.I., Johnson, W., Schultz, D.R., 1995. Antibody-independent binding of complement component C1q by *Legionella pneumophila*. Infection and Immunity 63, 4939-4943.

Mishra, J., Sahoo, P.K., Mohanty, B.R., Das, A., 2009. Sequence information, ontogeny and tissue-specific expression of complement component C3 in Indian major carp, *Labeo rohita* (Hamilton). Indian Journal of Experimental Biology 47, 672-678.

Morgan, B.P., Gasque, P., 1996. Expression of complement in the brain: role in health and disease. Immunology Today 17, 461-466.

Muller-Eberhard, H.J., 1975. Complement. Annu. Rev. Biochem. 44, 697.

Muller-Eberhard, H.J., 1986. The membrane attack complex of complement. Annual Review of Immunology 4, 503-528.

Muller-Eberhard, H.J., Nilsson, U., 1960. Relation of a b glycoprotein of human senun of the complement systern. Journal of Experimetal Medicine 111, 217.

Mutsuro, J., Tanaka, N., Kato, Y., Dodds, A.W., Yano, T., Nakao, M., 2005. Two divergent isotypes of the fourth complement component from a bony fish, the common carp (*Cyprinus carpio*). Journal of Immunology 175, 4508-4517.

Nakao, M., Fushitani, Y., Fujiki, K., Nonaka, M., Yano, T., 1998. Two diverged complement factor B/C2-like cDNA sequences from a teleost, the common carp (*Cyprinus carpio*). Journal of Immunology 161, 4811-4818.

Nakao, M., Kato-Unoki, Y., Nakahara, M., Mutsuro, J., Somamoto, T., 2006. Diversified components of the bony fish complement system: more genes for robuster innate defense? Advances in Experimental Medicine and Biology 586, 121-138.

Nakao, M., Matsumoto, M., Nakazawa, M., Fujiki, K., Yano, T., 2002. Diversity of complement factor B/C2 in the common carp (*Cyprinus carpio*): three isotypes of B/C2-A expressed in different tissues. Developmental and Comparative Immunology 26, 533-541.

Nakao, M., Mutsuro, J., Nakahara, M., Kato, Y., Yano, T., 2003. Expansion of genes encoding complement components in bony fish: biological implications of the complement diversity. Developmental and Comparative Immunology 27, 749-762.

Nakao, M., Tsujikura, M., Ichiki, S., Vo, T.K., Somamoto, T., 2011. The complement system in teleost fish: progress of post-homolog-hunting researches. Developmental and Comparative Immunology 35, 1296-1308.

Nakao, M., Uemura, T., Yano, T., 1996. Terminal components of carp complement constituting a membrane attack complex. Molecular Immunology 33, 933-937.

Nauta, A.J., Trouw, L.A., Daha, M.R., Tijsma, O., Nieuwland, R., Schwaeble, W.J., Gingras, A.R., Mantovani, A., Hack, E.C., Roos, A., 2002. Direct binding of C1q to apoptotic cells and cell blebs induces complement activation. European Journal of Immunology 32, 1726-1736.

Nelson, R.A., Jensen, J., Gigli., Tamuna, N., 1966. Methods for the separation, purification and measurement of nine components of hemolytic complement in guinea-pig serum. Immunochemistry 3, 111.

Nonaka, M., Azumi, K., 1999. Opsonic complement system of the solitary ascidian, *Halocynthia roretzi*. Developmental and Comparative Immunology 23, 421-427.

Nonaka, M., Kuroda, N., Naruse, K., Shima, A., 1998. Molecular genetics of the complement C3 convertases in lower vertebrates. Immunological Reviews 166, 59-65.

Nonaka, M., Smith, S.L., 2000. Complement system of bony and cartilaginous fish. Fish & Shellfish Immunology 10, 215-228.

Nonaka, M., Yamaguchi, N., Natsuume-Sakai, S., Takahashi, M., 1981. The complement system of rainbow trout (*Salmo gairdneri*). I. Identification of the serum lytic system homologous to mammalian complement. Journal of Immunology 126, 1489-1494.

Papanastasiou, A.D., Zarkadis, I.K., 2005. Gene duplication of the seventh component of complement in rainbow trout. Immunogenetics 57, 703-708.

Papanastasiou, A.D., Zarkadis, I.K., 2006. The gamma subunit of the eighth complement component (C8) in rainbow trout. Developmental and Comparative Immunology 30, 485-491.

Paramaswara, B., 1998. Characterisation of the Promoter for the Human Factor I Graduate Department of Cellular and Molecular Pathology. University of Toronto.

Pikarsky, E., Porat, R.M., Stein, I., Abramovitch, R., Amit, S., Kasem, S., Gutkovich-Pyest, E., Urieli-Shoval, S., Galun, E., Ben-Neriah, Y., 2004. NF-kappaB functions as a tumour promoter in inflammation-associated cancer. Nature 431, 461-466.

Ponting, C.P., 1999. Chlamydial homologues of the MACPF (MAC/perforin) domain. Current Biology : CB 9, R911-913.

Qi, Z.H., Liu, Y.F., Wang, W.N., Wu, X., Xin, Y., Lu, Y.F., Wang, A.L., 2011. Molecular characterization and functional analysis of a complement C3 molecule in the orange-spotted grouper (*Epinephelus coioides*). Fish & Shellfish Immunology 31, 1284-1290.

Quiniou, S.M., Katagiri, T., Miller, N.W., Wilson, M., Wolters, W.R., Waldbieser, G.C., 2003. Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. Genetics, Selection, Evolution : GSE 35, 673-683.

Rabs, U., Martin, H., Hitschold, T., Golan, M.D., Heinz, H.P., Loos, M., 1986. Isolation and characterization of macrophage-derived C1q and its similarities to serum C1q. European Journal of Immunology 16, 1183-1186.

Rath, B., Pandey, R.S., Debata, P.R., Maruyama, N., Supakar, P.C., 2008. Molecular characterization of senescence marker protein-30 gene promoter: identification of repressor elements and functional nuclear factor binding sites. BMC Molecular Biology 9, 43.

Ravi, V., Venkatesh, B., 2008. Rapidly evolving fish genomes and teleost diversity. Current Opinion in Genetics & Development 18, 544-550.

Rossi, V., Cseh, S., Bally, I., Thielens, N.M., Jensenius, J.C., Arlaud, G.J., 2001. Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. The Journal of Biological Chemistry 276, 40880-40887.

Rothman, B.L., Merrow, M., Bamba, M., Kennedy, T., Kreutzer, D.L., 1989. Biosynthesis of the third and fifth complement components by isolated human lung cells. The American Review of Respiratory Disease 139, 212-220.

Rotllant, J., Parra, D., Peters, R., Boshra, H., Sunyer, J.O., 2004. Generation, purification and functional characterization of three C3a anaphylatoxins in rainbow trout: role in leukocyte chemotaxis and respiratory burst. Developmental and Comparative Immunology 28, 815-828.

Ruiz, S., Henschen-Edman, A.H., Tenner, A.J., 1995. Localization of the site on the complement component C1q required for the stimulation of neutrophil superoxide production. The Journal of Biological Chemistry 270, 30627-30634.

Sahu, A., Lambris, J.D., 2001. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. Immunological Reviews 180, 35-48.

Samonte, I.E., Sato, A., Mayer, W.E., Shintani, S., Klein, J., 2002. Linkage relationships of genes coding for alpha2-macroglobulin, C3 and C4 in the zebrafish: implications for the evolution of the complement and Mhc systems. Scandinavian Journal of Immunology 56, 344-352.

Seeger, A., Mayer, W.E., Klein, J., 1996. A complement factor B-like cDNA clone from the zebrafish (*Brachydanio rerio*). Molecular Immunology 33, 511-520.

Sellar, G.C., Blake, D.J., Reid, K.B., 1991. Characterization and organization of the genes encoding the A-, B- and C-chains of human complement subcomponent C1q. The complete derived amino acid sequence of human C1q. The Biochemical Journal 274 (Pt 2), 481-490.

Shen, Y., Zhang, J., Xu, X., Fu, J., Li, J., 2012a. Expression of complement component C7 and involvement in innate immune responses to bacteria in grass carp. Fish & Shellfish Immunology.

Shen, Y., Zhang, J., Xu, X., Fu, J., Li, J., 2013. A new haplotype variability in complement C6 is marginally associated with resistance to Aeromonas hydrophila in grass carp. Fish & Shellfish Immunology 34, 1360-1365.

Shen, Y., Zhang, J., Xu, X., Fu, J., Liu, F., Li, J., 2012b. Molecular cloning, characterization and expression of the complement component Bf/C2 gene in grass carp. Fish & Shellfish Immunology 32, 789-795.

Shen, Y.B., Zhang, J.B., Xu, X.Y., Li, J.L., 2011. Molecular cloning, characterization and expression analysis of the complement component C6 gene in grass carp. Veterinary Immunology and Immunopathology 141, 139-143.

Shin, D.H., Webb, B., Nakao, M., Smith, S.L., 2007. Molecular cloning, structural analysis and expression of complement component Bf/C2 genes in the nurse shark, *Ginglymostoma cirratum*. Developmental and Comparative Immunology 31, 1168-1182.

Smith, L.C., Azumi, K., Nonaka, M., 1999. Complement systems in invertebrates. The ancient alternative and lectin pathways. Immunopharmacology 42, 107-120.

Smith, S.L., 1998. Shark complement: an assessment. Immunological Reviews 166, 67-78.

Stanley, K.K., Herz, J., 1987. Topological mapping of complement component C9 by recombinant DNA techniques suggests a novel mechanism for its insertion into target membranes. The EMBO Journal 6, 1951-1957.

Strunk, R.C., Fleischer, J.A., Katz, Y., Cole, F.S., 1994. Developmentally regulated effects of lipopolysaccharide on biosynthesis of the third component of complement and factor B in human fibroblasts and monocytes. Immunology 82, 314-320.

Strunk, R.C., Whitehead, A.S., Cole, F.S., 1985. Pretranslational regulation of the synthesis of the third component of complement in human mononuclear phagocytes by the lipid A portion of lipopolysaccharide. The Journal of Clinical Investigation 76, 985-990.

Sunyer, J.O., Lambris, J.D., 1998. Evolution and diversity of the complement system of poikilothermic vertebrates. Immunological Reviews 166, 39-57.

Sunyer, J.O., Zarkadis, I., Sarrias, M.R., Hansen, J.D., Lambris, J.D., 1998a. Cloning, structure, and function of two rainbow trout Bf molecules. Journal of Immunology 161, 4106-4114.

Sunyer, J.O., Zarkadis, I.K., Lambris, J.D., 1998b. Complement diversity: a mechanism for generating immune diversity? Immunology Today 19, 519-523.

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. Molecular Biology and Evolution 28, 2731-2739.

Thai, C.T., Ogata, R.T., 2004. Complement components C5 and C7: recombinant factor I modules of C7 bind to the C345C domain of C5. Journal of Immunology 173, 4547-4552.

Thielens, N.M., Tacnet-Delorme, P., Arlaud, G.J., 2002. Interaction of C1q and mannan-binding lectin with viruses. Immunobiology 205, 563-574.

Thomas, M.J., Seto, E., 1999. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene 236, 197-208.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673-4680.

Thompson, R.A., Lachmann, P.J., 1970. Reactive lysis: the complement-mediated lysis of unsensitized cells. I. The characterization of the indicator factor and its identification as C7. The Journal of Experimental Medicine 131, 629-641.

Tomlinson, S., Stanley, K.K., Esser, A.F., 1993. Domain structure, functional activity, and polymerization of trout complement protein C9. Developmental and Comparative Immunology 17, 67-76.

Toranzo, A.E., Magariños, B., Romalde, J.L., 2005. A review of the main bacterial fish diseases in mariculture systems. Aquaculture 246, 37-61.

Trexler, M., Banyai, L., Patthy, L., 2001. A human protein containing multiple types of protease-inhibitory modules. Proceedings of the National Academy of Sciences of the United States of America 98, 3705-3709.

Vik, D.P., Wong, W.W., 1993. Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele. Journal of Immunology 151, 6214-6224.

Wei, W., Wu, H., Xu, H., Xu, T., Zhang, X., Chang, K., Zhang, Y., 2009. Cloning and molecular characterization of two complement Bf/C2 genes in large yellow croaker (*Pseudosciaena crocea*). Fish & Shellfish Immunology 27, 285-295.

Whang, I., Lee, Y., Kim, H., Jung, S.J., Oh, M.J., Choi, C.Y., Lee, W.S., Kim, S.J., Lee, J., 2011. Characterization and expression analysis of the myeloid differentiation factor 88 (MyD88) in rock bream *Oplegnathus fasciatus*. Molecular Biology Reports 38, 3911-3920.

Wickramaarachchi, W.D., Wan, Q., Lee, Y., Lim, B.S., De Zoysa, M., Oh, M.J., Jung, S.J., Kim, H.C., Whang, I., Lee, J., 2012. Genomic characterization and expression analysis of complement component 9 in rock bream (*Oplegnathus fasciatus*). Fish & Shellfish Immunology 33, 707-717.

Wickramaarachchi, W.D., Whang, I., Kim, E., Lim, B.S., Jeong, H.B., De Zoysa, M., Oh, M.J., Jung, S.J., Yeo, S.Y., Kim, S.Y., Park, H.C., Lee, J., 2013a. Genomic characterization and transcriptional evidence for the involvement of complement component 7 in immune response of rock bream (*Oplegnathus fasciatus*). Developmental and Comparative Immunology 41, 44-49.

Wickramaarachchi, W.D., Whang, I., Wan, Q., Bathige, S.D., De Zoysa, M., Lim, B.S., Yeo, S.Y., Park, M.A., Lee, J., 2013b. Genomic characterization and expression analysis of complement component 8α and 8β in rock bream (*Oplegnathus fasciatus*). Developmental and Comparative Immunology 39, 279-292.

Wilson, D.R., Juan, T.S., Wilde, M.D., Fey, G.H., Darlington, G.J., 1990. A 58-base-pair region of the human C3 gene confers synergistic inducibility by interleukin-1 and interleukin-6. Molecular and Cellular Biology 10, 6181-6191.

Wimmers, K., Mekchay, S., Schellander, K., Ponsuksili, S., 2003. Molecular characterization of the pig C3 gene and its association with complement activity. Immunogenetics 54, 714-724.

Winn, J.S., Guille, J., Gebicki, J.M., Day, R.O., 1991. Hydrogen peroxide modulation of the respiratory burst of human neutrophils. Biochemical Pharmacology 41, 31-36.

Witzel-Schlomp, K., Rittner, C., Schneider, P.M., 2001. The human complement C9 gene: structural analysis of the 5' gene region and genetic polymorphism studies. European Journal of Immunogenetics : Official Journal of the British Society for Histocompatibility and Immunogenetics 28, 515-522.

Wong, N.K., Kojima, M., Dobo, J., Ambrus, G., Sim, R.B., 1999. Activities of the MBL-associated serine proteases (MASPs) and their regulation by natural inhibitors. Molecular Immunology 36, 853-861.

Wurzner, R., 2000. Modulation of complement membrane attack by local C7 synthesis. Clinical and Experimental Immunology 121, 8-10.

Yeo, G.S., Elgar, G., Sandford, R., Brenner, S., 1997. Cloning and sequencing of complement component C9 and its linkage to DOC-2 in the pufferfish *Fugu rubripes*. Gene 200, 203-211.

Yu, C.Y., 1991. The complete exon-intron structure of a human complement component C4A gene. DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene. Journal of Immunology 146, 1057-1066.

Yu, Y., Huang, H., Wang, Y., Yu, Y., Yuan, S., Huang, S., Pan, M., Feng, K., Xu, A., 2008. A novel C1q family member of amphioxus was revealed to have a partial function of vertebrate C1q molecule. Journal of Immunology 181, 7024-7032.

Zarkadis, I.K., Duraj, S., Chondrou, M., 2005. Molecular cloning of the seventh component of complement in rainbow trout. Developmental and Comparative Immunology 29, 95-102.

Zarkadis, I.K., Sarrias, M.R., Sfyroera, G., Sunyer, J.O., Lambris, J.D., 2001. Cloning and structure of three rainbow trout C3 molecules: a plausible explanation for their functional diversity. Developmental and Comparative Immunology 25, 11-24.

Zenke, K., Kim, K.H., 2008. Functional characterization of the RNase III gene of rock bream iridovirus. Archives of Virology 153, 1651-1656.

Zhang, H., Song, L., Li, C., Zhao, J., Wang, H., Qiu, L., Ni, D., Zhang, Y., 2008. A novel C1q-domain-containing protein from Zhikong scallop *Chlamys farreri* with lipopolysaccharide binding activity. Fish & Shellfish Immunology 25, 281-289.

Zhong, L., Zhang, F., Chang, Y., 2012. Gene cloning and function analysis of complement B factor-2 of *Apostichopus japonicus*. Fish & Shellfish Immunology 33, 504-513.

Zhou, M., Ouyang, W., 2003. The function role of GATA-3 in Th1 and Th2 differentiation. Immunologic Research 28, 25-37.

Zhou, Z., Liu, H., Liu, S., Sun, F., Peatman, E., Kucuktas, H., Kaltenboeck, L., Feng, T., Zhang, H., Niu, D., Lu, J., Waldbieser, G., Liu, Z., 2012. Alternative complement pathway of channel catfish (*Ictalurus punctatus*): molecular characterization, mapping and expression analysis of factors Bf/C2 and Df. Fish & Shellfish Immunology 32, 186-195.

Zhu, Y., Thangamani, S., Ho, B., Ding, J.L., 2005. The ancient origin of the complement system. The EMBO Journal 24, 382-394.