MOLECULAR CLONING, EXPRESSION, AND *IN SILICO* ANALYSIS OF A LONG TYPE PEPTIDOGLYCAN RECOGNITION PROTEIN FROM HALF-SMOOTH TONGUE SOLE, *CYNOGLOSSUS SEMILAEVIS* (ACTINOPTERYGII: PLEURONECTIFORMES: CYNOGLOSSIDAE)

Zhi T. QI [1, *], Qi H. ZHANG [1], Zi S. WANG [1], M. QIU [1], Y. HUANG [2], and Q. GAO [3, *]

[1] *Key Laboratory of Aquaculture and Ecology of Coastal Pool in Jiangsu Province, Yancheng Institute of Technology, Yancheng 224051, China*

[2] *Key Laboratory of Agricultural Environmental Microbiology, Ministry of Agriculture, College of Life Sciences, Nanjing Agricultural University, Nanjing, 210095, China*

[3] *State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*

Background. Half-smooth tongue sole, *Cynoglossus semilaevis* Günther, 1873, a marine teleost, is an important aquaculture species of great economic value. In recent years, its farm production increase coincided, however, with the number of reported cases of bacterial diseases. Further understanding of its immune response to bacterial pathogens can provide more information on pathogenesis and how to prevent disease using immune-related strategy. Peptidoglycan (PGN) recognition proteins (PGRPs) play important roles in the innate immunity against bacterial infection. In the presently reported study, a long type PGRP in half-smooth tongue sole (csPGRP-L) was cloned, and its sequence features, PGN binding ability, and mRNA expressions in different tissues after bacterial infection were also analyzed.

Materials and methods. The full length of csPGRP-L cDNA was obtained by RT-PCR and RACE-PCR method, and its sequence features were analyzed by multiple sequence alignment and phylogenetic tree. Meanwhile, its 3-D structure and PGN binding ability were analyzed by comparative modelling and molecular docking methods. Furthermore, the expressions of csPGRP-L in different tissues of healthy fish and fish infected with *Streptococcus dysgalactiae* were examined using quantitative real-time PCR method.

Results. The full length of csPGRP-L cDNA was 1509 bp (GenBank accession No. HQ909441), with a 1446 bp of open reading frame (ORF) encoding 481 amino acids (aa), which possessed several conserved PGRP family features, e.g., a typical PGRP domain at its C-terminal, 3-D structure. Molecular docking showed that the csPGRP-L also possessed the PGN-binding ability. csPGRP-L was constitutive expressed in all the selected tissues from healthy fish and following *S. dysgalactiae* infection its expression was up-regulated in a tissue-specific expression pattern.

Conclusion. The gene we cloned was exactly the homologue of vertebrates' long type PGRP in half-smooth tongue sole which was confirmed by several analyses and the up-regulation of csPGRP-L after bacterial infection suggest that csPGRP-L plays important role in antibacterial and anti-infective action.

Keywords: PGRP, gene clone, gene expression, bacterial stimulation, fish

INTRODUCTION

Peptidoglycan (PGN), belonging to the pathogen-associated molecular patterns (PAMPs), is the major component of bacterial cell wall. It is a polymer of β(1–4)-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) linked to a short peptide of D- or L- amino acids. The peptide at position 3 of PGN is different between Gram-negative bacteria and Gram-positive bacteria,

---

* Correspondence: Associate Prof. Zhitao Qi, 江苏省盐城市迎宾大道9号，盐城工学院, 224051, 中国 (China), phone: +86-88298612, e-mail: *qizhitao@ycit.edu.cn* or Associate Prof. Qian GAO, 湖北省武汉市东湖南路7号，中国科学院水生生物研究所, 430072, 中国 (China), e-mail: *gaoqian@ihb.ac.cn*.

with diaminopimelic acid (DAP) in Gram-negative bacteria and lysine in Gram-positive bacteria, respectively, which divide the PGN into two types: DAP-PGN and Lys-PGN (Dziarski and Gupta 2006, Li et al. 2007).

PGN-recognition proteins (PGRPs) were first identified as pattern-recognition receptors (PRRs) in insect which could specifically bind with PGN to initiate different antimicrobial pathways (Janeway and Medzhitov 2002). *Drosophila* PGRP-SA could bind with Lys-type PGN which activates the toll-like receptors (TLRs) pathway (Michel et al. 2001), PGRP-SD binds with DAP-type PGN and activates the TLRs pathway (Leone et al. 2008), while PGRP-LC and PGRP-LE bind with DAP-type PGN to activate the Immune Deficiency (Imd) pathway (Takehana et al. 2002, 2004, Lim et al. 2006).

In mammals only four types of PGRPs were identified, termed as PGLYRP1 for short type of PGRP, PGLYRP2 for long type PGRP and PGLYRP3 and 4 for intermediate type of PGRP to distinguish with insect PGRPs. Mammalian PGRPs also presented their selective binding specificity to PGN, e.g. human PGLYRP1 could bind with two types of PGN (Cho et al. 2005), while PGLYRP3 only bind with Lys-type PGN (Guan et al. 2004). Except for PRRs function, the mammalian PGRPs also possess immune-pathway regulation, amidase and bactericidal activities.

To date, PGRPs have been identified from several lower vertebrates. In fish, PGRP had been cloned in zebrafish, *Danio rerio* (Hamilton, 1822) (see Li et al. 2007); rockfish, *Sebastes schlegeli* Hilgendorf, 1880 (see Kim et al. 2010); grass carp, *Ctenopharyngodon idella* (Valenciennes, 1844) (see Li et al. 2013); and rainbow trout, *Oncorhynchus mykiss* (Walbaum, 1792) (see Jang et al. 2013). The research on fish PGRPs indicated that fish PGRPs might possess similar functions as mammalian PGRPs, e.g., PGN-lysis, the direct bactericidal action and the regulation of immune genes (Li et al. 2007, Chang and Nie 2008). However, the available information about fish PGRPs was less extensive, especially for representatives of Pleuronectiformes.

Half-smooth tongue sole, *Cynoglossus semilaevis* Günther, 1873, a marine pleuronectiform teleost, is an important aquaculture species in China and other Asian countries due to its delicious taste and economical value (Song et al. 2012). However, with transforming of its farming method from extensive to intensive type in recent years, the occurrences of the diseases, especially of bacterial diseases, increased greatly. Streptococcosis has been a major cause of mass mortality in marine aquaculture of Pleuronectiformes (see Park et al. 2013). Due to the resistances of pathogens to antibiotics and some effective fishery medicine having been found to be toxic to human and forbidden by law, the immune-related therapy might be one of good choice for the control of fish bacterial diseases. Thus, it is necessary to understand the immune responses of half-smooth tongue sole during bacterial infection. In the presently reported study, the gene of a long type peptidoglycan recognition protein (csPGRP-L)

was cloned from half-smooth tongue sole and its expression profiles under healthy condition and bacterial infection were detected by real-time quantitative PCR. Meanwhile, the 3-D structure of this protein was constructed using comparative modelling method and its PGN binding ability was analyzed by molecular docking method to reveal the interaction between fish PGRP and PGN.

MATERIAL AND METHODS

Fish, bacterial stimulating, and tissues collecting. Healthy, 15~20-cm long, half-smooth tongue sole, *Cynoglossus semilaevis*, weighing 20~30 g, were purchased from a fish farm in Ganyu County, Lianyungang city, Jiangsu Province, China. The fish were acclimatized for at least one week before experiments.

Three fish were anesthetized with 50 mg · L^{-1} MS-222 (Shanghai Buxi Chemical Co., Ltd, China) and then dissected. The tissues including heart, gill, liver, spleen, stomach, muscle, kidney, and gut were collected for RNA extraction.

To investigate the effect of bacterial stimulation on the expression of csPGRP-L, 20 fish were randomly held in two 40-L tanks at 22°C with 10 fish in each tank. Samples of *Streptococcus dysgalactiae* were obtained from the State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, the Chinese Academy of Sciences, and cultured according to the method reported previously (Yang and Li 2009). The bacteria were suspended in sterile PBS at the concentration of 1.0×10^6 CFU · mL^{-1} as the challenge trial. 10 fish were injected intraperitoneally with 0.5 mL of 1.0×10^6 CFU · mL^{-1} bacteria and other 10 fish were injected with same amount of PBS as control. Tissues including heart, gill, liver, spleen, stomach, muscle, kidney, and gut from both groups were sampled at 12 h and 36 h at post injection for RNA extraction, respectively. The time for sampling was referred to previous reports in zebrafish, *Danio rerio*, and frog, *Xenopus tropicalis* (see Chang et al. 2007, Qi et al. 2011, respectively).

All of the experiments described comply with the country's regulations for the use of laboratory animals and have been approved by the Institutional Review Board of the Yancheng Institute of technology, China.

Total RNA extraction and first-strand cDNA synthesis. Total RNAs of sampled tissues were extracted using TRIzol Reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. The quantity and integrity of RNA was determined by measuring the absorbance at 260 and 280 nm and electrophoresis. After extraction, total RNAs were reverse-transcribed into cDNA using the SuperScript™ Reverse Transcript kit (Invitrogen, CA, USA). The cDNA were stored at –20 °C.

Cloning of csPGRP-L cDNA. To obtain the partial sequence of csPGRP-L, the PGRP-Ls of human (GenBank accession No. AAL05629) and zebrafish (GenBank accession No. ABF50497) were used to search the deposited EST sequences of half-smooth tongue sole at National Center for Biotechnology Information (NCBI) website[n] by Blastp program. An EST sequence (GenBank

accession No. GH231532) was predicted to be PGRP homolog. To confirm the correct of the sequence, a PCR was performed with 10 min at 94°C, 35 cycles of 1 min at 94°C, 30 s at 60°C, 45 s at 72°C, and 10 min at 72°C using a pair of specific primers (csPGRP-f1/csPGRP-r1) designed according to the obtained sequence. The PCR products were separated by agarose gel electrophoresis, extracted from gel using TaKaRa Agarose Gel Purification Kit (TaKaRa, Japan) and ligated into pMD18-T (TaKaRa, Japan) and sequenced. The result of DNA sequencing confirmed that the EST sequence was exactly correct.

Based on the confirmed sequence, 3′ and 5′ rapid amplification of cDNA ends (RACE) were conducted using gene-specific primers and adaptor primers (UPM) to obtain the full length of csPGRP-L cDNA. For 3′ RACE, the first-round PCR was performed using csPGRP-3out and UPM as primers as well as being subjected to an initial denaturation step of 94°C for 5 min, then 9 cycles of 94°C for 30 s, 64°C for 30 s, and 72°C for 90 s, followed by 29 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 90 s, with a final extension step of 72°C for 10 min. Products of the first-round PCR were then diluted (1 : 100) and re-amplified by a second-round PCR with csPGRP-3in/UPM as primers. For 5′ RACE, the first-round primers were csPGRP-5out/UPM, and the second-round primers were csPGRP-5in/UPM. The PCR products were analyzed as RT-PCR. All primers used in this analysis were listed in Table 1.

Sequence analysis. Homologous sequences were identified using the BLAST program at the NCBI website[*]. The deduced amino acid sequence was predicted using software at the ExPASy molecular biology server[**]. Signal peptides were identified using the SignalP program[***]. The protein family domains or motifs were predicted by Pfam software in the ExPASy server[**]. Multiple sequence alignments were conducted with CLUSTAL W1.8 and amino acid identities were determined by MegAlign from the DNAStar software suite (Clewley and Arnold 1997). Phylogenetic trees were generated using the neighbour-joining (N-J) method with 1000 bootstraps by MEGA 5.1.2 (Kumar et al. 2008).

Comparative modelling and molecular docking. For 3-D construction, the mature csPGRP-L was first used to search by BLASTP software against PDB, I-TASSER server[*****] (Zhang 2008) and Genesilico Metaserver[*****] (Kurowski and Bujnicki 2003) to find the suitable template for comparative modelling. All of these searching results suggested that the *Drosophila* PGRP-LB (PDB code: 1OHT) (Kim et al. 2003) was the best template for csPGRP-L, with the highest level of sequence identity and query coverage and less E-value, which also confirmed by FUGUE (Find Homologs of Uncharacterized Gene Products Using Environment-specific substitution tables) program[******] (Shi et al. 2001).

After the suitable template was confirmed, 3-D model of csPGRP-L was generated using the SWISS-Model server[*******] based on homology modelling method (Arnold et al. 2006). The obtained model was visualized using DeepView software (Johansson et al. 2012). The model quality was assessed using the PROCEHCK software (Laskowski et al. 1993).

The PGN binding ability of csPGRP-L was analyzed by molecular docking method. The manually docked csPGRP-L with PGN model, *Drosophila* PGRP-LE with tracheal cytotoxin (TCT) (a soluble type of DAP-PGN)

Table 1

Primers used in the study

Primer	Sequence (from 5′ to 3′)	Amplification
csPGRP-F1	TGGCATCCATCTGAGAAACA	Amplification of cDNA
csPGRP-R1	CTTGGGCAAGGACCACAC	
UPM-long	CTAATACGACTCACTATAGGGCAAG-CAGTGGTATCAACGCAGAGT	3′RACE
UPM-short	CTAATACGACTCACTATAGGGC	
csPGRP-3out	TGACCTGTCGCCGCTGGCTCTGCTC	
csPGRP-3in	AGTAACAGACAACGGTGAGGAACGAGGT	
csPGRP-5out	CCTCGTTCCTCACCGTTGTCTGTTACT	5′RACE
csPGRP-5in	GAGCAGAGCCAGCGGCGACAGGTC	
csPGRP-qF1	TGAAGACGCTGGCATTGGT	Real-time quantitative PCR
csPGRP-qR1	TGCGGACGGTTCGTAGGTG	
csactin-qF1	CAGCCATACTGTGCCCATCT	
csactin-qR1	TCCTTGATGTCACGCACGAT	

[*] http://www.ncbi.nlm.nih.gov.
[**] http://www.expasy.org.
[***] http://www.dtu.dk/services/SignalP.
[****] http://zhanglab.ccmb.med.umich.edu/I-TASSER.
[*****] https://genesilico.pl/meta2.
[******] http://tardis.nibio.go.jp/fugue.
[*******] http://swissmodel.expasy.org.

(Lim et al. 2006) was subjected to 200 cycles of energy minimization by Discovery Studio/Minimization program. **Real-time quantitative PCR.** Real-time quantitative PCR (qPCR) was done to analyze the basal and bacterial stimulated expression level of csPGRP-L using CFX96 Touch[TM] Real-time PCR detection system (Bio-Rad). RNA extracted from the selected tissues was digested with RNase-free DNase I (TaKaRa, Japan) and transcribed into cDNA using a PrimeScript[TM] RT reagent kit (Invitrogen, USA) according to the manufacture's instruction. The concentration of obtained cDNA was measured using BioPhotometer plus (Eppendorf, Germany), and then they were diluted into the same concentration using TE buffer for qPCR analysis. The gene specific primers were designed according to the csPGRP-L and β-actin (control gene) cDNA sequence (Table 1). The 20 μL reaction mixture of qPCR contained 10 μL SYBR® *Premix Ex Taq*[TM] (TaKaRa, Japan), 0.5 μL of each gene specific primer, 1.0 μL cDNA template and 8 μL double-distilled H_2O. The PCR was performed using a 3-step method, an initial denaturation step at 95°C for 3 min, followed by 40 cycles of 95°C for 5 s, 58°C for 30 s, and finally 95°C for 10 s and then the melting curve were analyzed from 65°C to 95°C to confirm the specificity of amplification. All samples were performed in triplicate. Basal expression level for target gene in each normal tissue was measured relative to the expression of β-actin gene, respectively. Fold changes of csPGRP-L in different tissues after bacterial stimulation were analyzed using $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). All the qPCRs data were expressed as the mean ± SD and analyzed by Microcal Origin 6.0 software. Comparison between challenged and control group was performed using Student's *t*-test method, with *P*-value < 0.05 set to be statistically significant.

RESULTS
Cloning and sequence analysis of csPGRP-L. The full cDNA length of csPGRP-L was obtained using RT-PCR and RACE-PCR. The csPGRP-L cDNA was totally 1509 bp in length (GenBank accession No. HQ909441), with a 25 bp of 5′-untranslated region (UTR), 119 bp of 3′-UTR, and 1446 bp of open reading frame (ORF) encoding 481 amino acids (aa). There was a polyadenylation signal (aataaa) at 80 bp upstream of the polyA tail. A 22-aa signal peptide was predicted for csPGRP-L, giving mature csPGRP-L of 459 aa with a calculated molecular weight of 51.3 kDa and isoelectric point (pI) of 6.25. The C-terminal region of csPGRP-L contained a typical PGRP domain with 138 aa in length predicted by Pfam software (Fig. 1).

Multiple sequence alignment of csPGRP-L with other known PGRP-L molecules from human, mouse, frog, and zebrafish showed that the C-terminals of PGRP-L, containing PGRP domain, were highly conserved. Some conservative amino acid residues necessary for the biological functions of PGRP-L were also found in csPGRP-L, e.g., H344, Y381, H456, and C465 for amidase activity, and H370, W376, and T463 for Zn^{2+} binding (Fig. 2).

Phylogenetic tree analysis showed that the vertebrate PGRPs were divided into three clusters with one containing the mammalian PGLYRP3 and PGLYRP4, another one containing PGRP-S and the other one containing all the PGRP-Ls and some PGRP-Ss from Atlantic salmon, *Salmo salar* Linnaeus, 1758; zebrafish, *Danio rerio*; and grass carp, *Ctenopharyngodon idella*. csPGRP-L was well clustered with the PGRP-Ls of Atlantic croaker, *Micropogonias undulatus* (Linnaeus, 1766); and Japanese flounder, *Paralichthys olivaceus* (Temminck et Schlegel, 1846) with a bootstrap value of 79% (Fig. 3).
Tissue distribution of csPGRP-L transcript. The expressions of csPGRP-L in the tissues from healthy fish were detected using qRT-PCR by normalizing to the expression level of β-actin. csPGRP-L was constitutively expressed in all the selected tissues, highly in gill, moderately in heart, intestine, gut, kidney, and spleen and lowly in liver (Fig. 4a).
Transcript change of csPGRP-L after bacterial stimulation. The *Streptococcus dysgalactiae* infection could significantly increase the expression level of csPGRP-L in several selected tissues. Here, we found that the expression levels of csPGRP-L, in the tissues including heart, intestine, gill, kidney, muscle, and stomach, were significantly up-regulated at 12 h post injection (*P* < 0.05) compared with control group. The fold changes in the heart and gills remained highly induced (*P* < 0.05) at 36 h post injection, e.g., 5-fold increasing in the gills (Fig. 4b).
3-D Structure and PGN binding ability of csPGRP-L. We first generated a theoretical 3-D model of csPGRP-L, with *Drosophila melanogaster* PGRP-LB (PDB:1OHT) as template according to the sequence identity (37.5%), by using the SWISS-MODEL package. csPGRP-L contained a mixed alpha helices and beta sheets structure (β1-β2-α1-β3-β4-β5-α2-β6-α3), similar to the structure of template and human PGLYRP2 (Fig. 5a). Superimposition between the csPGRP-L and the structure of the template showed high similarities (Fig. 5b). The quality assessment of the model by using PROCHECK program confirmed the reliability of predicted model.

Then, the PGN docking with csPGRP-L was simulated by molecular superimposition using *Drosophila* PGRP-LE in complex with TCT (Lim et al. 2006) to discriminate the PGN-binding ability of csPGRP-L. It was clear that there was a cleft that was formed by helix α1 and four loops (β3-α1, α1-β4, β5-β6, β6-α2). The TCT (DAP-PGN) (in yellow colour) was held in an extended conformation in this cleft. The docking results showed that the Arg373 of csPGRP-L superimposed fully on the Arg254 of *Drosophila* PGRP-LE (Fig. 6a). Hydrogen bonds are formed between TCT and Thr324, Arg352, Gly355, Arg373, Arg380, His381, Asn386 and His434 of csPGRP-L (Fig. 6b), four of which are retained in the PGRP-LE-PGN structure (Arg233, Arg254, Asn267 and His314).

DISCUSSION
In the presently reported study, a full cDNA length of a long type PGRP was cloned from half-smooth tongue sole, *Cynoglossus semilaevis*, a marine teleost, using

RT-PCR and RACE-PCR method. The obtained sequence was exactly the homologue of the vertebrates' long type PGRP which was confirmed by several analyses:

• The csPGRP-L was 481 aa in length, possessing similar aa length with the reported PGRP-L in zebrafish (433 aa) (Chang et al. 2007) and frog (497 aa) (Qi et al. 2011) (Fig. 1);

• csPGRP-L possessed a typical PGRP domain at its C-terminal region (from 330 aa to 468 aa) (Fig. 1). The PGRP domain, about 200 aa, was well conserved in PGRP during evolution. But the number of PGRP domain in different type of PGRP was different, e.g., vertebrates' short and long type PGRP only have 1 PGRP domain, whist mammalian PGLYRP3 and PGRLYRP4 have two PGRP domains;

• Phylogenetic tree showed that csPGRP-L was well clustered with the reported PGRP-L from the fish belongs to Pleuronectiformes (Fig. 3);

```
   1  agaagcagcctctgaaacatgcaacATGACTTTGTCTGTTTTTATGACTTTCACTACAGC
   1                            M  T  L  S  V  F  M  T  F  T  T  A
  61  TCTGCTCTGCTTTTTCTCTGTCTGCAGCAGGCCAATTGGCATCCATCTGAGAAACATGGA
  13   L  L  C  F  F  S  V  C  S  R  P  I  G  I  H  L  R  N  M  D
 121  CAGTTTTATTCGTGCTGTGCAACAGCTTGAAGACTCAAATCCTGACCTGTCGCCGCTGGC
  33   S  F  I  R  A  V  Q  Q  L  E  D  S  N  P  D  L  S  P  L  A
 181  TCTGCTCAGAGCTCTGCGGAGGACTGCTGGCCACGATGATACAATGACTATCCATTTCCT
  53   L  L  R  A  L  R  R  T  A  G  H  D  D  T  M  T  I  H  F  L
 241  GGGTTCTTCAAATAACCTCAGTGACACTGAGGGCCTAGAGGCAGCCATACTTAATGCCTC
  73   G  S  S  N  N  L  S  D  T  E  G  L  E  A  A  I  L  N  A  S
 301  ATCTTTCAGTTTCTTTGACAAAGCCATCCACCACATAGTAACAGACAACGGTGAGGAACG
  93   S  F  S  F  F  D  K  A  I  H  H  I  V  T  D  N  G  E  E  R
 361  AGGTGTGGTCCTTGCCCAAGATGGCACCACAGTCGCCCTTGGGCCCTTACTGCTAGGGAT
 113   G  V  V  L  A  Q  D  G  T  T  V  A  L  G  P  L  L  L  G  I
 421  TGAAGCAGGACTGAAAGCCAGGGTGGAGGGAAAGCCCTCTGATGGGATCTTCTCTCTAAC
 133   E  A  G  L  K  A  R  V  E  G  K  P  S  D  G  I  F  S  L  T
 481  CTTGGGCAGGACACTTGGTCTGTCCTTCCTCAGCCTGCAGGACCTCCCACCATCTAATCG
 153   L  G  R  T  L  G  L  S  F  L  S  L  Q  D  L  P  P  S  N  R
 541  CCTGGGGCCAAATGGCTGCTGGGACAGTGTGCATCGCCCTAAAATGTTCAAGCTGTCTCA
 173   L  G  P  N  G  C  W  D  S  V  H  R  P  K  M  F  K  L  S  Q
 601  GCCCGCCACTCTGGCCACTGATGCTATGATTAATGGTGGTATGGACGGAGCCATACTTGG
 193   P  A  T  L  A  T  D  A  M  I  N  G  G  M  D  G  A  I  L  G
 661  TGTAGACATCAGTGACCTCAATAGTTCCGAAGAGCCAAAGAGCCTAAGTAAAATTTTGAA
 213   V  D  I  S  D  L  N  S  S  E  E  P  K  S  L  S  K  I  L  K
 721  AGACTACTACAGTTTTATCTTGCAGGAAGAGCATGATCTGGACACTCTGACCAGGCATGT
 233   D  Y  Y  S  F  I  L  Q  E  E  H  D  L  D  T  L  T  R  H  V
 781  AAGCCAAAAGCGAAGAGAGAGCTCTAGAGCACTGCTAGAGTCACGTGATCTTCATGGCGA
 253   S  Q  K  R  R  E  S  S  R  A  L  L  E  S  R  D  L  H  G  E
 841  GGTGATGAAGACGCTGGCATTGGTGTGGAAGCTGGAGATGACTGAATGGATCGCTTTCGA
 273   V  M  K  T  L  A  L  V  W  K  L  E  M  T  E  W  I  A  F  D
 901  CAATGAAATAGGGAAATCAGTGGAGGATGGACTGCATGAATTTGTGCACAACTACTGGGA
 293   N  E  I  G  K  S  V  E  D  G  L  H  E  F  V  H  N  Y  W  D
 961  CTGCCCTCCCATCATCCCTCGCTGTCAGTGGGGTGCAAAAGCTCACGAGAGTACGCCCAC
 313   C  P  P  I  I  P  R  C  Q  W  G  A  K  A  H  E  S [T  P  T
1021  CCCGCTGTCTCTGCCACTGCAGTTTCTGTATGTCCACCACACCTACGAACCGTCCGCACC
 333   P  L  S  L  P  L  Q  F  L  Y  V  H  H  T  Y  E  P  S  A  P
1081  CTGTTTATCCTTCACCCAGTGCTCTCGTGACATGAAGGCCATGCAGCGTTTCCACCAGGA
 353   C  L  S  F  T  Q  C  S  R  D  M  K  A  M  Q  R  F  H  Q  E
1141  GGACCGTGGCTGGGGAGACATAGGATACAGTTTTGTGGTGGGCTCCGATGGTTACGTCTA
 373   D  R  G  W  G  D  I  G  Y  S  F  V  V  G  S  D  G  Y  V  Y
1201  TGAGGGCAGGGGATGGAAACACCTCGGCAGACACACCAGGGGGCACAATCACCTCGGCTA
 393   E  G  R  G  W  K  H  L  G  R  H  T  R  G  H  N  H  L  G  Y
1261  TGGGGTGTCGATCATCGGTAACTACACTACAAGGCTTCCGTCACGTCATGCAGTGGACTT
 413   G  V  S  I  I  G  N  Y  T  T  R  L  P  S  R  H  A  V  D  L
1321  GTTGCGACATCGTCTGGTCAAATGTGCCATTGACGGAGGACGACTTAGGACAAACTTCAC
 433   L  R  H  R  L  V  K  C  A  I  D  G  G  R  L  R  T  N  F  T
1381  CATCCATGGCCACAGGCAGGTGGTGAACACCCTCTGCCCTGGAGATGCCTTCTTTTCAGA
 453   I  H  G  H  R  Q  V  V  N  T  L  C  P  G  D  A] F  F  S  E
1441  AATACGAAGCTGGGAACACTTTAAGGATTGAccgttaatgaaataaaataaataacgtac
 473   I  R  S  W  E  H  F  K  D  *
1501  atttgctgtaaacctgatacgaatgcacattgctacgttcaaactatagtaaaaatgttt
1561  ttttttaacaaaaaaaaaaaaaaaaaaaaa
```

Fig. 1. Nucleotide and deduced amino acid sequence of PGRP-L from half-smooth tongue sole, *Cynoglossus semilaevis*; The signal peptide is underlined; Initial codon (ATG) and stop codon (TGA) is bold and underlined, and putative poly-adenylation signal (aataaa) is underlined with a grey line; The PGRP domain at the C-terminal region of csPGRP-L is in a frame

- csPGRP-L possessed the conserved 3-D structure, similar to the reported long type PGRP, e.g., human PGLYRP2 (Fig. 5);
- csPGRP-L possessed potential PGN binding ability. The PGN binding ability was an ancient feature of PGRPs and important for its functional exertion (Dziarski 2004). Our molecular docking analysis showed that csPGRP-L contained a conserved PGN binding cleft formed by helix α1 and four loops (β3-α1, α1-β4, β5-β6, β6-α2), similar to *Drosophila* PGRP-LE (Fig. 6). Arg373 of csPGRP-L was fully superimposed with Arg254 of *Drosophila* PGRP-LE. This residue was also conserved in PGRP molecules, e.g., PGRP-LCx (Mellroth et al. 2005), PGRP-LB (Kim et al. 2003), and human PGRP-S (Kumar et al. 2005), that are known to interact with PGN.

Interestingly, we found that csPGRP-L only possessed signal peptide but no transmembrane domain, which sug-

gest it might be a secreted protein. This feature was different with the reported PGRP-Ls. The mammalian PGLYRP2 possessed both signal peptide and trans-membrane domain, whist PGRP-L in the lower vertebrates, e.g., frog and zebrafish, only had trans-membrane domain (Chang et al. 2007, Qi et al. 2011). Thus, the long type of PGRP might play its roles at different cellular sites.

Further analysis of the phylogenetic tree gave light on the evolutionary story of vertebrates' PGRPs. There were at least four gene duplications occurred during PGRPs' evolution. The first round of gene duplication produced the ancestral proto-PGRP nearly happened in the stem lineage leading to all jawed vertebrates. The second round occurred just after the first round produced proto-PGRP-S and then the proto-PGRP-S diverged into pro-PGRP-L and pro-PGRP-I in the third round gene duplication. The fish PGRP-L were produced by the forth gene duplication from proto-PGRP-L (Montaño et al. 2011).
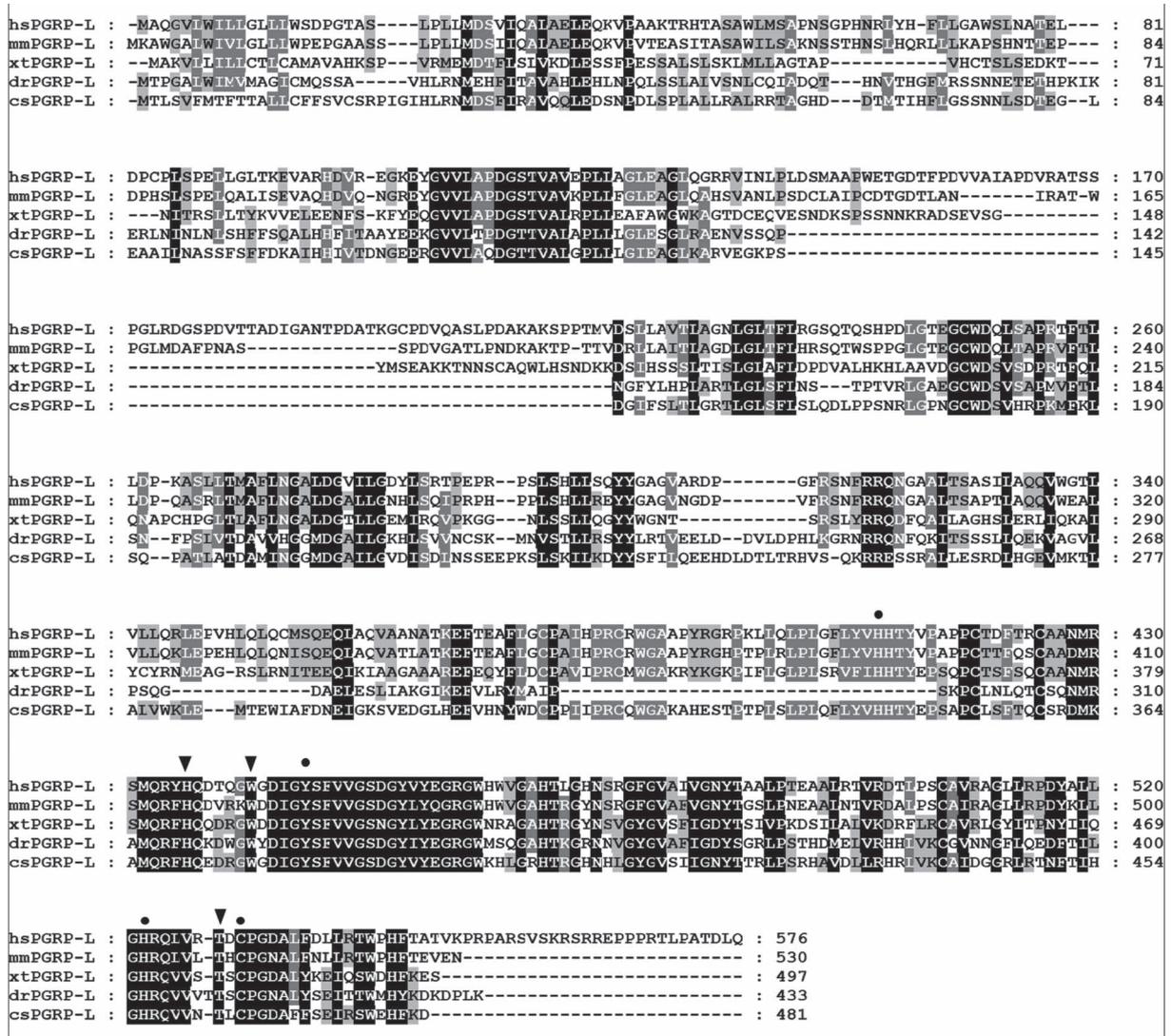

Fig. 2. Alignment of amino acid sequences of PGRP-L from half-smooth tongue sole, *Cynoglossus semilaevis* (csPGRP-L), human, *Homo sapiens* (hsPGRP-L), mouse, *Mus musculus* (mmPGRP-L), frog, *Xenopus tropicalis* (xtPGRP-L), and zebrafish, Danio rerio (drPGRP-L); The amino acids needed for amidase activity was marked by solid black cycle and for Zn^{2+} binding by solid triangle above the sequences, respectively

csPGRP-L was constituted expressed in all the detected tissues and the same case was also observed in other lower vertebrates' PGRP-L. But the transcript levels of PGRP-L were selectively expressed in different tissues. Rainbow trout PGRP-L were most abundant in the intestine, and at a low level present, in increasing order, in the gill, kidney, head kidney, liver, eye, skin, and spleen (Jang et al. 2013). Frog PGRP-L was highly expressed in liver, lung and intestine, and moderately in heart and stomach and lowly in spleen and kidney (Qi et al. 2011). Whist, zebrafish PGRP-L was detected in head kidney, skin, gills, eye, trunk kidney, liver and spleen, with strong
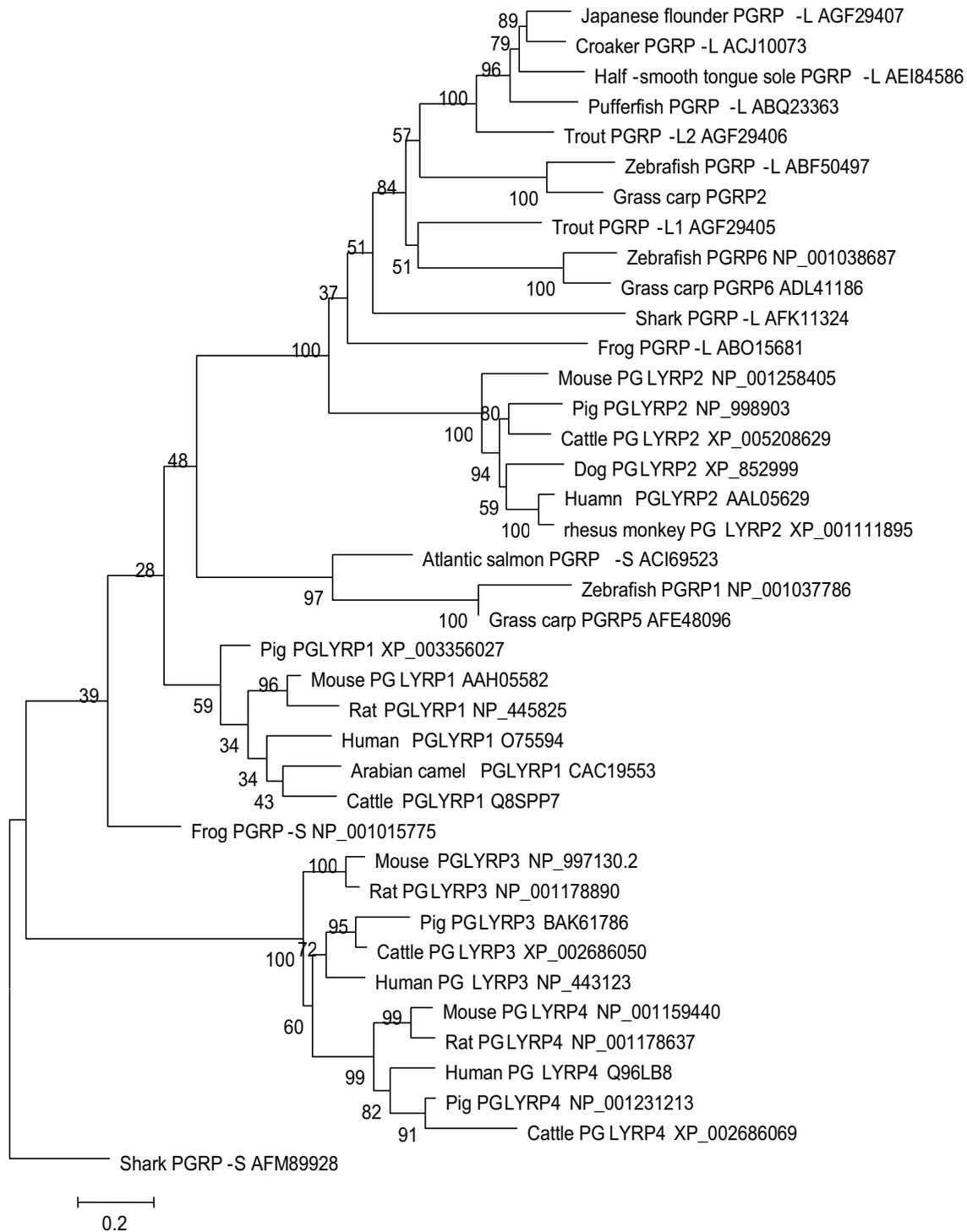
Fig. 3. Phylogenetic tree analysis of relation of PGRP from half-smooth tongue sole, *Cynoglossus semilaevis*, with that in other vertebrates; Phylogenetic tree was constructed based on the deduced amino acid sequences of PGRP using N-J method in MEGA 5.1. The GenBank accession numbers of the sequences used for this analysis was shown behind each name

detection in intestine (Li et al. 2007). The constitutive expression pattern of PGRP-L indicated that it possessed a wide secreted basis and might exert function in a wide range of tissues in teleost.

Several bacterial could induce the expression of PGRP-L in a tissue-specific pattern. The zebrafish PGRP-L was significantly up-regulated in spleen and liver, significantly down-regulated in kidney and intestine after *Flavobacterium columnare* infection (Chang et al. 2007). The PGRP-L of red drum, *Sciaenops ocellatus* (Linnaeus, 1766), was up-regulated in kidney, spleen, and liver after *Streptococcus* infection (Li et al. 2012). Here, our results also found that csPGRP-L was up-regulated in several tissues after *Streptococcus* infection but the expression pattern was different with previous reported. csPGRP-L was significantly up-regulated in the tissues including heart, intestine, gill, kidney, muscle and stomach at 12 h post injection ($P < 0.05$) and remained highly induced in gill and heart at 36 h post injection ($P < 0.05$). The highly induced expression of csPGRP-L in gill indicated that csPGRP-L might contribute to the defence against pathogens in those tissues that contact directly with pathogens in the surrounding water. Previous study found that zebrafish PGRP-Ls could be detected in the lumen of blood vessels using immunohisto-

chemistry methods, indicating fish PGRP-L might play some role in the non-immune tissues (Li et al. 2007). Here, we found that the transcript level of csPGRP-L in heart indicate that there was secrete basis for the csPGRP-L in the non-immune tissues. After *Streptococcus* infection the highly induced expression of csPGRP-L also confirmed this and suggested that during bacterial infection the host immune system need protect the important tissues to sustain its living. The expression of csPGRP-L in other tissues including intestine, kidney, muscle, and stomach was back to normal expression level suggest that csPGRP-L might play different roles in some tissues at different stage of bacterial infection.
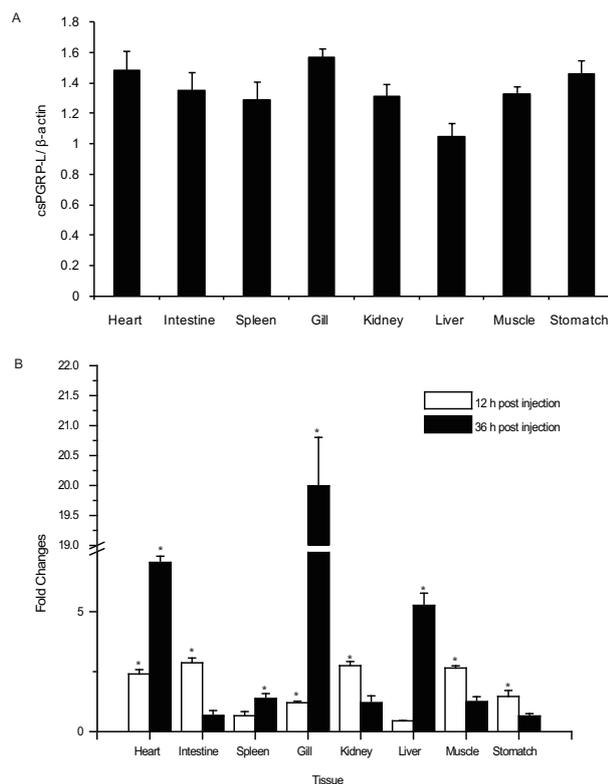
Fig. 4. Gene expression of PGRP-L in the tissues from healthy half-smooth tongue sole, *Cynoglossus semilaevis*, (**A**) and the fish infected by *Streptococcus dysgalactiae* for 12 and 36 h (**B**) by qRT-PCR; β-actin was used for internal control for producing relative fold value compared with that in liver for basal expression analysis and $2^{-\Delta\Delta CT}$ method was used for bacterial stimulated analysis. Data were expressed as mean ± SE; *$P < 0.05$
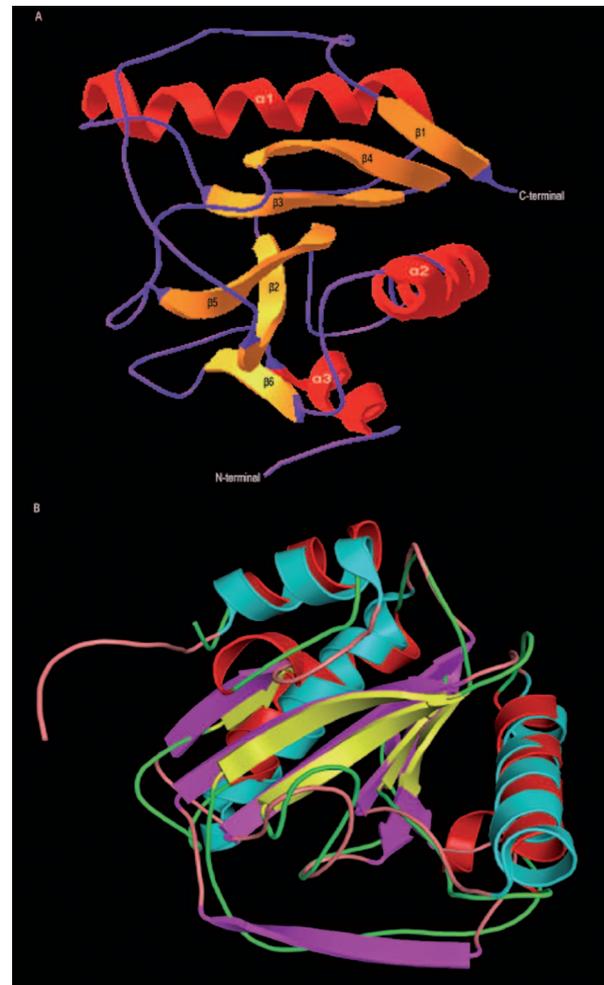

Fig. 5. (**A**) 3-D structure of PGRP-L from half-smooth tongue sole, *Cynoglossus semilaevis*; The α helices are red, loops blue, and β sheets—yellow; (**B**) Superimposition of csPGRP-L model with template 1OHT; The α helices, β sheets and loops of csPGRP-L are red, yellow, and green, and those of 1OHT are cyan, carmine, and brown, respectively
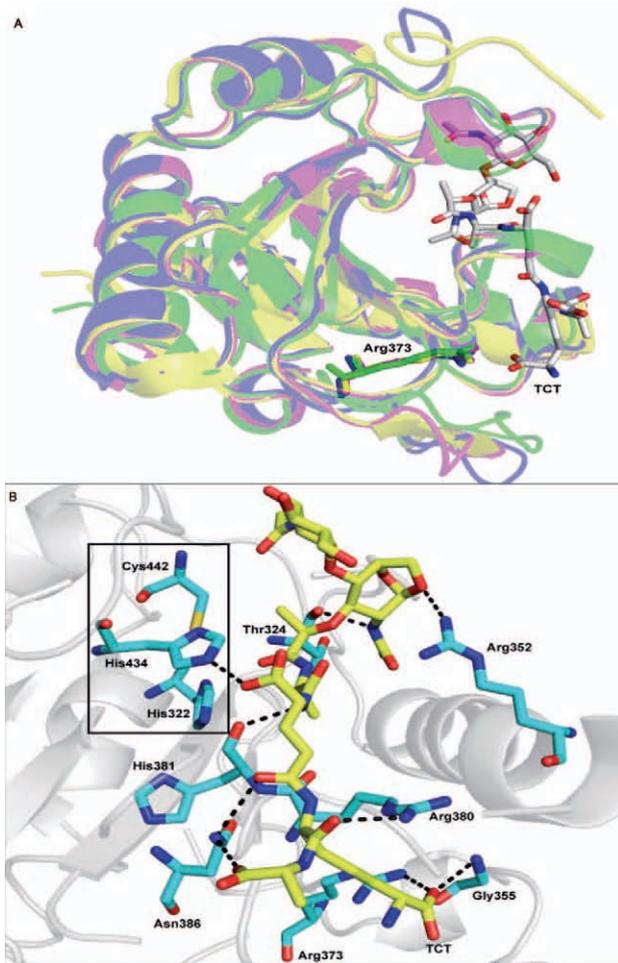
Fig. 6. Determining the PGN binding ability and intermolecular contacts in the PGRP-L from half-smooth tongue sole, *Cynoglossus semilaevis*, with TCT complex; (**A**) Superimposition of csPGRP-L (magenta), PGRP-SD (green), PGRP-LB (yellow), and PGRP-LE (blue) in complex with TCT (grey rod); The side chain of Arg373 in csPGRP-L fully superimposed on that of Arg90, Arg92, and Arg254 in PGRP-SD, PGRP-LB, and PGRP-LE, respectively; (**B**) Intermolecular contacts between csPGRP-L and TCT at the PGN-binding site; TCT was shown as yellow rod, csPGRP-L in grey, and the contacting residues in cyan; Hydrogen bonds were shown as dashed lines; Residues of active site conserved in csPGRP-L were highlighted in the inner frame

REFERENCES

Arnold K., Bordoli L., Kopp J., Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22** (2): 195–201. DOI: 10.1093/bioinformatics/bti770

Chang M.X., Nie P. 2008. RNAi suppression of zebrafish peptidoglycan recognition protein 6 (zfPGRP6) mediated differ-

entially expressed genes involved in Toll-like receptor signalling pathway and caused increased susceptibility to *Flavobacterium columnare*. Veterinary Immunology and Immunopathology **124** (3–4): 295–301. DOI: 10.1016/j.vetimm.2008.04.003

Chang M.X., Nie P., Wei L.L. 2007. Short and long peptidoglycan recognition proteins (PGRPs) in zebrafish, with finding of multiple PGRP homologs in teleost fish. Molecular Immunology **44** (11): 3005–3023. DOI: 10.1016/j.molimm.2006.12.029

Cho J.H., Fraser I.P., Fukase K., Kusumoto S., Fujimoto Y., Stahl G.L., Ezekowitz R.A. 2005. Human peptidoglycan recognition protein S is an effector of neutrophil-mediated innate immunity. Blood **106** (7): 2551–2558. DOI: 10.1182/blood-2005-02-0530

Clewley J.P., Arnold C. 1997. MEGALIAN. The multiple alignment module of LASERGENE. Pp. 119–120. *In*: Swindell S.R. (ed.) Sequence data analysis guidebook. Book series: Methods in Molecular Biology Vol. 70. Humana Press, Totowa, NJ, USA. DOI: 10.1385/0-89603-358-9:119

Dziarski R. 2004. Peptidoglycan recognition proteins (PGRPs). Molecular Immunology **40** (12): 877–886. DOI: 10.1016/j.molimm.2003.10.011

Dziarski R., Gupta D. 2006. Mammalian PGRPs: novel antibacterial proteins. Cellular Microbiology **8** (7): 1059–1069. DOI: 10.1111/j.1462-5822.2006.00726.x

Guan R., Malchiodi E.L., Wang Q., Schuck P., Mariuzza R.A. 2004. Crystal structure of the C-terminal peptidoglycan-binding domain of human peptidoglycan recognition protein Iα. Journal of Biological Chemistry **279** (30): 31873–31882. DOI: 10.1074/jbc.M404920200

Janeway C.A.jr., Medzhitov R. 2002. Innate immune recognition. Annual Review of Immunology **20**: 197–216. DOI: 10.1146/annurev.immunol.20.083001.084359

Jang J.H., Kim H., Cho J.H. 2013. Rainbow trout peptidoglycan recognition protein has an anti-inflammatory function in liver cells. Fish and Shellfish Immunology **35** (6): 1838–1847. DOI: 10.1016/j.fsi.2013.09.015

Johansson M.U., Zoete V., Michielin O., Guex N. 2012. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics **13**: 173. DOI: 10.1186/1471-2105-13-173

Kim M.-S., Byun M., Oh B.-H. 2003. Crystal structure of peptidoglycan recognition protein LB from *Drosophila melanogaster*. Nature Immunology **4** (8): 787–793. DOI: 10.1038/ni952

Kim M.Y., Jang J.H., Lee J.-W., Cho J.H. 2010. Molecular cloning and characterization of peptidoglycan recognition proteins from the rockfish, *Sebastes schlegeli*. Fish and Shellfish Immunology **28** (4): 632–639. DOI: 10.1016/j.fsi.2009.12.023

Kumar S., Nei M., Dudley J., Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics **9** (4): 299–306. DOI: 10.1093/bib/bbn017

Kumar S., Roychowdhury A., Ember B., Wang Q., Guan R., Mariuzza R.A., Boons G.-J. 2005. Selective recognition of synthetic lysine and *meso*-diaminopimelic acid-type pepti-

doglycan fragments by human peptidoglycan recognition proteins Iα and S. Journal of Biological Chemistry **280** (44): 37005–37012. DOI: 10.1074/jbc.M506385200

Kurowski M.A., Bujnicki J.M. 2003. GeneSilico protein structure prediction meta-server. Nucleic Acids Research **31** (13): 3305–3307. DOI: 10.1093/nar/gkg557

Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography **26** (2): 283–291. DOI: 10.1107/S0021889892009944

Leone P., Bischoff V., Kellenberger C., Hetru C., Royet J., Roussel A. 2008. Crystal structure of *Drosophila* PGRP-SD suggests binding to DAP-type but not lysine-type peptidoglycan. Molecular Immunology **45** (9): 2521–2530. DOI: 10.1016/j.molimm.2008.01.015

Li J.H., Chang M.X., Xue N.N., Nie P. 2013. Functional characterization of a short peptidoglycan recognition protein, PGRP5 in grass carp *Ctenopharyngodon idella*. Fish and Shellfish Immunology **35** (2): 221–230. DOI: 10.1016/j.fsi.2013.04.025

Li M.-F., Zhang M., Wang C.-L., Sun L. 2012. A peptidoglycan recognition protein from *Sciaenops ocellatus* is a zinc amidase and a bactericide with a substrate range limited to Gram-positive bacteria. Fish and Shellfish Immunology **32** (2): 322–330. DOI: 10.1016/j.fsi.2011.11.024

Li X., Wang S., Qi J., Echtenkamp S.F., Chaterjee R., Wang M., Boons G.-J., Dziarski D., Gupta D. 2007. Zebrafish peptidoglycan recognition proteins are bactericidal amidases essential for defense against bacterial infections. Immunity **27** (3): 518–529. DOI: 10.1016/j.immuni.2007.07.020

Lim J.-H., Kim M.-S., Kim H.-E., Yano T., Oshima Y., Aggarwal K., Goldman W.E., Silverman N., Kurata S., Oh B.-H. 2006. Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. Journal of Biological Chemistry **281** (12): 8286–8295. DOI: 10.1074/jbc.M513030200

Livak K.J., Schmittgen T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25** (4): 402–408. DOI: 10.1006/meth.2001.1262

Mellroth P., Karlsson J., Håkansson J., Schultz N., Goldman W.E., Steiner H. 2005. Ligand-induced dimerization of *Drosophila* peptidoglycan recognition proteins in vitro. Proceeding of the National Academy of Sciences of the United States of America **102** (18): 6455–6460. DOI: 10.1073/pnas.0407559102

Michel T., Reichhart J.-M., Hoffmann J.A., Royet J. 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature **414** (6865): 756–759. DOI: 10.1038/414756a

Montaño A.M., Tsujino F., Takahata N., Satta Y. 2011. Evolutionary origin of peptidoglycan recognition proteins in vertebrate innate immune system. BMC Evolutionary Biology **11**: 79. DOI: 10.1186/1471-2148-11-79

Park M.A., Kwon M.G., Hwang J.Y., Jung S.H., Kim D.-W., Park J.-Y., Kim J.-S., Na Y.-J., Kim M.-Y., Kim D.-S., Chae S.-H., Seo J.S. 2013. Genome sequence of *Streptococcus parauberis* strain KCTC11980, isolated from diseased *Paralichthys olivaceus*. Genome Announcements **1** (5): e00780-13. DOI: 10.1128/genomeA.00780-13

Qi Z.-T., Zhang Q.-H., Wang Z.-S., Wang A.-M., Huang B., Chang M.-X., Nie P. 2011. Cloning and expression analysis of a long type peptidoglycan recognition protein (PGRP-L) from *Xenopus tropicalis*. Zoological Research **32** (4): 371–378.

Shi J., Blundell T.L., Mizuguchi K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. Journal of Molecular Biology **310** (1): 243–257. DOI: 10.1006/jmbi.2001.4762

Song W., Li Y., Zhao Y., Liu Y., Niu Y., Pang R., Miao G., Liao X., Shao C., Gao F., Chen S. 2012. Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). PLoS One **7** (12): e52097. DOI: 10.1371/journal.pone.0052097

Takehana A., Katsuyama T., Yano T., Oshima Y., Takada H., Aigaki T., Kurata S. 2002. Overexpression of a pattern-recognition receptor, peptidoglycan recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. Proceeding of the National Academy of Sciences of the United States of America **99** (21): 13705–13710. DOI: 10.1073/pnas.212301199

Takehana A., Yano T., Mita S., Kotani A., Oshima Y., Kurata S. 2004. Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity. The EMBO Journal **23** (23): 4690–4700. DOI: 10.1038/sj.emboj.7600466

Yang W.M., Li A.H. 2009. Isolation and characterization of *Streptococcus dysgalactiae* from diseased *Acipenser schrenckii*. Aquaculture **294** (1–2): 14–17. DOI: 10.1016/j.aquaculture.2009.05.018

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9**: 40. DOI: 10.1186/1471-2105-9-40