**Two short peptidoglycan recognition proteins from *Crassostrea gigas*
with similar structure exhibited different PAMP binding activity**

Chuanyan Yang ^a, Lingling Wang ^{a, b}, Zhihao Jia ^c, Qilin Yi ^a, Qingsong Xu ^a,
Weilin Wang ^c, Changhao Gong ^a, Conghui Liu ^c, Linsheng Song ^{a, b, *}

Abstract

Peptidoglycan recognition protein (PGRP) is an essential molecule in innate immunity for both invertebrates and vertebrates, owing to its prominent ability in specifically recognizing bacterial peptidoglycan (PGN) and eliminating the invading bacteria. In the present study, the full length cDNA of two PGRP genes, *CgPGRPS2* and *CgPGRPS4*, were cloned from oyster *Crassostrea gigas*. Their amino acid sequences both contained one signal peptide, one typical PGRP/amidase domain with conserved catalytic residues responsible for amidase activity (55H, 90Y, 164H, 172C in *CgPGRPS2*, and 98H, 133Y, 207H, 215C in *CgPGRPS4*), and specific PGN recognition (84R, 85W, 104R, 109V in *CgPGRPS2*, and 127G, 128W, 147R, 152V in *CgPGRPS4*), and they shared 55.9% sequence similarity. The mRNA transcripts of *CgPGRPS2* and *CgPGRPS4* were constitutively expressed in all the examined tissues, including haemocytes, hepatopancreas, mantle, gonad, heart, adductor muscle and gill, with the highest expression level in adductor muscle and hepatopancreas, respectively. Both *CgPGRPS2* and *CgPGRPS4* proteins were mainly localized in the cytoplasm. The recombinant protein of *CgPGRPS2* (r*CgPGRPS2*) could bind lipopolysaccharide (LPS), PGN and mannan (Man), as well as various microorganisms including Gram-negative bacteria *Escherichia coli*, *Vibrio anguillarum*, Gram-positive bacteria *Staphylococcus aureus* and fungi *Yarrowia lipolytica*. The recombinant protein of *CgPGRPS4* (r*CgPGRPS4*) exhibited higher binding affinity to PGN, lower binding affinity to LPS, while no binding activity to Man and *Y. lipolytica*. The results indicated that *CgPGRPS2* and *CgPGRPS4* could function as pattern recognition

receptors (PRR) in the innate immune response of oyster, and they exhibited a certain degree of functional differentiation in recognition of Man.

Keywords: *Crassostrea gigas*; PGRP; Innate immunity; Pattern recognition receptor; PAMP binding activity

1 Introduction

Peptidoglycan recognition proteins (PGRPs) are conserved pattern recognition receptors that specifically recognize bacterial peptidoglycan (PGN) and play important roles in innate immune response in both invertebrates and vertebrates (Kang et al., 1998; Steiner, 2004; Dziarski, 2004). They were firstly identified from hemolymph of silkworm *Bombyx mori* with the activities to bind bacterial PGN and activate the downstream prophenoloxidase (proPO) system (Yoshida et al., 1996). By far, PGRP genes have been identified in many organisms, including mammals (Kang et al., 1998; Liu et al., 2001), fishes (Li et al., 2007; Sun and Sun, 2015), echinoderms (Coteur et al., 2007), arthropods (Werner et al., 2000; Udompetcharaporn et al., 2014; Dai et al., 2015) and molluscs (Ni et al., 2007; Su et al., 2007; Itoh and Takahashi, 2008; Itoh and Takahashi, 2009; Ye et al., 2014). Based on the predicted structures, PGRPs are divided into three classes, including the short extracellular PGRPs (PGRP-S) of 20-25 kDa, intermediate PGRPs (PGRP-I) of 40-45 kDa with two predicted transmembrane domains, and long PGRPs (PGRP-L) of up to 90 kDa intracellular or membrane- spanning proteins (Dziarski, 2004; Guan et al., 2005).

67 Compared to the high variety of the N-terminal sequence, the C-terminal amino acid
68 sequences of PGRP are highly conserved in almost all mammals and insects ([Werner](#)
69 [et al., 2000](#)), which contain one or more PGRP domains. The PGRP domains share
70 high homology with the amidase-2 of phage and bacteria, while their spatial folding
71 and active sites are more similar to lysozyme of bacteriophage T7 ([Kim et al., 2003](#)).
72 The diversity of PGRPs in both category and structure suggests that PGRPs perform
73 multifunction in innate immunity.

74 The increasing evidences from insects and mammals indicated that PGRPs could
75 function in innate immunity as pattern recognition receptor (PRR), signal regulation
76 molecules and effectors ([Royet et al., 2011](#)). PGRPs can recognize and bind bacterial
77 PGN ([Kang et al., 1998](#); [Yoshida et al., 1996](#)), lipopolysaccharide (LPS), lipoteichoic
78 acid (LTA) and fatty acids ([Sharma et al., 2012, 2013a and 2013b](#)), and then activate
79 the downstream immune responses, such as proPO system ([Yoshida et al., 1996](#)), Toll
80 pathway ([Michel et al., 2001](#); [Lemaitre et al., 2007](#)), the immune deficiency (IMD)
81 and c-Jun N-terminal kinase (JNK) pathway ([Steiner, 2004](#); [Choe et al., 2006](#); [Vincent](#)
82 [et al., 2006](#); [Maillet et al., 2008](#)). As signal regulation molecules, PGRPs could
83 negatively regulate the IMD pathway and protect the host from innocuous infection
84 ([Vincent et al., 2006](#); [Basbous et al., 2011](#)). Some PGRPs exhibit
85 N-acetylmuramoyl-L-alanine amidase activity ([Wang et al., 2003](#); [Mellroth and](#)
86 [Steiner, 2006](#)) and antimicrobial or bactericidal activity ([Lu et al., 2006](#); [Tydell et al.,](#)
87 [2006](#)), and some can function as opsonins to induce agglutination or phagocytosis
88 ([Ramet et al., 2002](#); [Coteur et al., 2007](#)). The multifunction of PGRPs in innate

immunity could help the host to effectively prevent the invasion of pathogenic microorganisms, and specifically control the level of symbiotic microorganisms.

Compared with the remarkable progress in mammals and other model organisms (Dziarski, 2004; Li et al., 2007; Royet et al., 2011), studies on PGRPs in marine invertebrates, especially in molluscs are still at the beginning. By far, several molluscan PGRPs have been identified from oysters (Itoh and Takahashi, 2008; Itoh and Takahashi, 2009; Zhang and Yu, 2013; Zhang et al., 2015), scallops (Ni et al., 2007; Su et al., 2007), razor clams (Wei et al., 2012), freshwater mussel (Ye et al., 2014) and squid (Michael et al., 2005; Joshua et al., 2009; Joshua et al., 2010). It has been demonstrated that most of the molluscan PGRPs could function in innate immunity as PRR. The mRNA expression of PGRPs in scallop (Ni et al., 2007; Su et al., 2007; Yang et al., 2010), razor clams (Wei et al., 2012) and freshwater mussel (Ye et al., 2014) could be significantly up-regulated by bacteria or PAMPs, and four PGRPs from Hawaiian bobtail squid could collaborate with each other to participate in the identification of symbiotic luminous bacteria (Michael et al., 2005; Joshua et al., 2009; Joshua et al., 2010). Moreover, some molluscan PGRPs could function as immune effectors. The recombinant CfPGRP-S1 (rCfPGRP-S1) could function as a bactericidal amidase to degrade PGN, agglutinate *Micrococcus luteus* and *Bacillus subtilis*, and inhibit the growth of *Escherichia coli* and *Staphylococcus aureus* (Yang et al., 2010). The recombinant CgPGRP-S1S (rCgPGRP-S1S) could bind and agglutinate *E. coli*, and induce secretion of granular contents by hemocyte degranulation (Iizuka et al., 2014). However, the structure and function of molluscan

PGRPs and their roles in innate immunity are still not well studied.

In the oyster genome, nine PGRP genes were predicated (Zhang et al., 2015), and four of them, CgPGRPS2 (AB425337.1), CgPGRPS3 (CGI_10027186), CgPGRPS4 (CGI_10027187) and CgPGRPS5 (CGI_10001545) were only consisted of a signal peptide and a PGRP domain (Fig.S1A), which shared significant sequence identity (Fig.S1B). The highest sequence similarity was observed between CgPGRPS3 and CgPGRPS4 (78.5%), while the lowest similarity was observed between CgPGRPS2 and CgPGRPS5 (36%) (Table S1). In our previous study, the highest expression level of CgPGRPS2 mRNA was observed in adductor muscle, while CgPGRPS3, CgPGRPS4 and CgPGRPS5 were all highly expressed in hepatopancreas (Fig.S2). Moreover, the mRNA expression of CgPGRPS2 could be regulated by LPS, PGN and GLU, while CgPGRPS3, CgPGRPS4 and CgPGRPS5 were not (Fig.S3). On the contrary, the mRNA expression levels of CgPGRPS4 from the early D-shaped larvae to juvenile were higher according to the transcriptional analysis by Zhang et al. (Zhang et al., 2012), while that of CgPGRPS2 could hardly be detected during the whole embryonic development stages (Fig.S4). Therefore, we selected CgPGRPS2 and CgPGRPS4 as representatives for further study because of (1) their moderate sequence similarity, (2) their similar but not identical phylogenesis, (3) their different tissue distribution, and (4) their different mRNA expression profile after PAMPs challenge and during the larval development. The objective of the present study was to seek the possible functional assignment of CgPGRPS2 and CgPGRPS4 in the innate immune response of oysters by comparing their expression profile, cellular

133 localization, as well as PAMP and microbial binding activities.

135 **2 Materials and Methods**

136 **2.1. Oysters and tissue collection**

137 Adult oysters, *C. gigas* about 2 years old (average shell length of 13.0 cm) were
138 collected from a local farm in Qingdao, Shandong Province, China, and maintained in
139 aerated fresh water at $15 \pm 2^{\circ}\text{C}$ for one week before processing.

140 Six tissues including hepatopancreas, mantle, gonad, heart, adductor muscle and
141 gill from six oysters were collected as parallel samples. Hemolymph from these six
142 oysters was also collected and immediately centrifuged at $800\times g$, 4°C for 10 min to
143 harvest the haemocytes. All these samples were stored at -80°C after addition of 1 mL
144 TRIzol reagent (TaKaRa) for RNA extraction.

146 **2.2. RNA isolation and cDNA synthesis**

147 Total RNA was isolated from haemocytes and tissues of oyster using Trizol reagent
148 (TaKaRa) following its protocol. The first strand cDNA synthesis was carried out
149 based on Promega M-MLV RT Usage information using the DNase I
150 (Promega)-treated total RNA as template and oligo(dT)-adaptor as primer ([Table 1](#)).
151 The reaction was performed at 42°C for 1 h, terminated by heating at 95°C for 5 min.
152 The cDNA mix was diluted to 1:100 and stored at -80°C for subsequent gene cloning
153 and SYBR Green fluorescent quantitative real-time PCR (qRT-PCR).

2.3. Cloning and sequence analysis of the full-length cDNA of CgPGRPS2 and CgPGRPS4

Sequence information of CgPGRPS2 (AB425337.1) and CgPGRPS4 (EKC26200) was retrieved from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.gov>). Two pairs of gene specific primers were designed to clone the full cDNA sequence of CgPGRPS2 (CgPGRPS2-F and -R) and CgPGRPS4 (CgPGRPS4-F and -R), respectively ([Table 1](#)). The PCR product was gel-purified and cloned into the pMD19-T simple vector (TaKaRa). The resulting sequences were verified and subjected to cluster analysis.

The protein sequence information for homologous and phylogenetic analysis was retrieved from NCBI and shown in [Table S2](#). The protein domains were predicted with the simple modular architecture research tool (SMART) version 7.0 (<http://smart.embl-heidelberg.de/>). Multiple sequence alignment was performed with the ClustalW multiple alignment program (<http://www.ebi.ac.uk/clustalw/>). An un-rooted neighbor-joining (NJ) tree was constructed with MEGA 6.06 software package. To derive confidence value for the phylogeny analysis, bootstrap trials were replicated 1000 times.

2.4. Detection of CgPGRPS2 and CgPGRPS4 mRNA transcripts in different tissues

The mRNA transcripts of CgPGRPS2 and CgPGRPS4 in different tissues were analyzed by qRT-PCR. All reactions were performed as previously described ([Yang et](#)

al., 2010). The information of all primers used in this assay was shown in Table 1. The expressions of CgPGRPS2 and CgPGRPS4 were normalized to that of elongation factor 1 α (EF1- α) gene for each sample. The $2^{-\Delta\Delta CT}$ method was used to analyze the expression level of CgPGRPS2 and CgPGRPS4 (Livak and Schmittgen, 2001). All data were given in terms of relative mRNA expressed as mean $\pm$ SE ($N \geq 4$). The data were subjected to one-way analysis of variance (one-way ANOVA) followed by a multiple comparison (S-NK). Differences were considered significant at $p < 0.05$.

2.5. Preparation of recombinant proteins and their polyclonal antibodies of CgPGRPS2 and CgPGRPS4

The cDNA fragments encoding the mature peptide of CgPGRPS2 and CgPGRPS4 were amplified with gene-specific primers (Table 1). A *Bam*HI site and an *Xho*I site were added to the 5' end of sense primers and antisense primers with the stop codon deletion, respectively. The PCR fragments were cloned into pMD19-T simple vector (TaKaRa), digested completely by restriction enzymes *Not*I and *Kpn*I (NEB), and then cloned into the *Bam*HI/*Xho*I sites of expression vector pET-30a (Novagen).

The incubation of the strains *E. coli* Transetta (DE3) with recombinant plasmid (pET-30a-CgPGRPS2 and pET-30a-CgPGRPS4), the induced expression of the recombinant proteins of CgPGRPS2 and CgPGRPS4 (designated rCgPGRPS2 and rCgPGRPS4), and the purification of rCgPGRPS2 and rCgPGRPS4 were performed as previously described (Yang et al., 2010). The resultant proteins were detected by SDS-PAGE and their concentration was quantified by BCA kit (Beyotime).

For the preparation of polyclonal antibodies, rCgPGRPS2 and rCgPGRPS4 were immunized to 6 weeks old mice to acquire polyclonal antibody as previous description (Zhou et al., 2013).

2.6. Immunohistochemistry assay of CgPGRPS2 and CgPGRPS4 in haemocytes

Hemolymph were obtained from oysters after one week culture in filtered aerated seawater at 18°C and immediately centrifuged at 800×g, 4°C for 10 min. Modified-Leibovitz L-15 mediums (Gibco) (Cao et al., 2003) were used to suspend the haemocytes. The haemocyte suspension was added into confocal dishes pre-coated with gelatin solution (Gelatin 5 g L⁻¹, CrK (SO₄)₂·12H₂O 0.5 g L⁻¹), and was allowed to adhere to the wall for 3 h. The following experiments including the fixation of haemocytes, the block of dishes, and the incubation of antibody were performed as previously described (Jia et al., 2015), except that the primary antibody incubation was performed by using 200 µL anti-CgPGRPS2 and anti-CgPGRPS4 (diluted 1:800 in blocking buffer). After three times of washing with PBST, DAPI (diluted 1:10000 in PBS) was added into the dishes to stain the nucleus. After the last three times of washing, the dishes were mounted in buffered glycerin for observation by Laser Scan Confocal Microscope (ZEISS).

2.7. PAMP binding assay

The binding activities of rCgPGRPS2 and rCgPGRPS4 to PAMPs were examined by Enzyme-linked immunosorbent assay (ELISA) based on PAMP binding assay

according to previous description (Jia et al., 2015). Briefly, 20 μg of lipopolysaccharide (LPS) from *E. coli* 0111:B4, Lys-type peptidoglycan (PGN) from *Staphylococcus aureus*, and mannan (Man) from *Saccharomyces cerevisiae* (Sigma) in 100 μL of carbonate-bicarbonate buffer (50 mmol L^{-1} , pH 9.6) were coated to the 96-well microliter plates (Costar) by incubation at 4°C overnight, respectively. After washed three times with PBST for 5 min per time, the wells were blocked with 200 μL of 5% BSA in PBS at 37°C for 1 h, and then incubated with 100 μL of rCgPGRPS2 and rCgPGRPS4 at different concentration with the presence of 0.1 mg mL^{-1} BSA and 1 mmol L^{-1} ZnCl_2 at 37°C for 1 h. The following antigen-antibody reaction and the calculation of apparent dissociation constant (K_d) were carried out according to previous description (Jia et al., 2015).

2.8. Microbial binding assay

Microbial binding activity was measured following as previous description with some modification (Lee and Söderhäll, 2001). The overnight cultured Gram-negative bacteria *E. coli* and *Vibrio anguillarum*, Gram-positive bacteria *S. aureus*, fungi *Yarrowia lipolytica* were used to detect the microbial-binding ability of rCgPGRPS2 and rCgPGRPS4. The microbes were suspended in TBS ($\text{OD}=2.0$), and incubated with 100 μL of 1 mg mL^{-1} rCgPGRPS2 and rCgPGRPS4 in the presence of 1 mmol L^{-1} ZnCl_2 under rotation at 4°C overnight. After 5 times washes, the bound proteins were dissociated from the microorganisms by loading buffer and analyzed by 12% SDS-PAGE. Proteins were transferred on to a 0.45 μm pore nitrocellulose membrane

and the membrane was blocked in 5% skim milk powder at room temperature for 1 h after three times of wash in TBST. The membrane was then incubated with the polyclonal antibody (diluted 1:1000 in blocking buffer) as primary antibody at 4°C overnight and HRP-labeled goat-anti-mouse IgG (AB clonal, diluted 1:4000 in blocking buffer) as second antibody at room temperature for 1 h. After four times of wash in TBST, the membrane was incubated briefly in Western lighting ECL substrate system (Thermo Scientific, USA) before scanned by Typhoon FLA 9500 rapid laser imager (GE, USA).

3 Results

3.1. The molecular characters and phylogeny of CgPGRPS2 and CgPGRPS4

The amplified nucleotide sequences of CgPGRPS2 (AB425337.1) and CgPGRPS4 (EKC26200) were consistent with their previous submission in NCBI. The open reading frame (ORF) of CgPGRPS2 encodes a polypeptide of 200 amino acids with a molecular mass of 22.4 kDa and a theoretical isoelectric point of 8.73, while the open reading frame (ORF) of CgPGRPS4 encodes a polypeptide of 235 amino acids with a molecular mass of 25.6 kDa and a theoretical isoelectric point of 6.85. SMART analysis revealed that both of the deduced proteins contained one signal peptide and one typical PGRP/amidase domain ([Fig. 1A](#)).

Multiple alignments of amino acid sequences revealed significant sequence identity of PGRP domains (from 38.7 to 76.4%) within the CgPGRP protein family ([Table. S2](#)). It was noteworthy that four catalytic residues responsible for amidase activity

were well conserved in both CgPGRPS2 and CgPGRPS4 (55H, 90Y, 164H, 172C in CgPGRPS2, and 98H, 133Y, 207H, 215C in CgPGRPS4), while four residues involved in specific PGN recognition showed a certain degree of variation (84R, 85W, 104R, 109V in CgPGRPS2, and 127G, 128W, 147R, 152V in CgPGRPS4) (Fig. 2). Moreover, the 13 possible substrate binding sites found in *Drosophila melanogaster* PGRP seems conserved at different degrees in CgPGRPs (Fig. 2).

Based on the multiple sequences alignment from *Homo sapiens* (4 HsPGRPs), *Danio rerio* (5 DrPGRPs), *D. melanogaster* (17 DmPGRPs) and *C. gigas* (9 CgPGRPs), an unrooted phylogenetic tree was constructed by using Neighbor Joining method (NJ) (Fig. 1B). In the phylogenetic tree, most PGRPs from the same species first gathered together, and formed a few major branches. CgPGRPS2 and CgPGRPS4 together with another 4 short type CgPGRPs (CgPGRPS1, CgPGRPS3, CgPGRPS6 and CgPGRPS7) were first clustered into the major branch of CgPGRPs, and then met the short type CgPGRPS5 after bunching with one HSPGRP (HsPGLYRP2) and three DrPGRPs (DrPGRP6, DrPGRPL and DrPGRP2). Finally, the 7 short type CgPGRPs were clustered with one long type CgPGRP (CgPGRPL1), and then clustered with three HSPGRPs (HsPGLYRP1, 3 and 4) and 14 DmPGRPs.

3.2. The tissue distribution of CgPGRPS2 and CgPGRPS4 mRNA

The qRT-PCR technique was employed to investigate the distribution of CgPGRPS2 and CgPGRPS4 mRNA transcripts in different tissues of oysters with EF1- α gene as internal control. For each gene, there was only one peak at the

corresponding melting temperature in the dissociation curve analysis, indicating that the PCR products were specifically amplified (data not shown). The CgPGRPS2 mRNA transcripts were constitutively expressed in all the detected tissues, and the highest expression level was found in adductor muscle, which was 25.44-fold ($p < 0.01$) of that in hepatopancreas (Fig. 3A). The relative expression levels of CgPGRPS2 in haemocytes, gonad, heart, gill and mantle were 7.39-fold ($p < 0.01$), 3.00-fold, 1.92-fold, 1.45-fold and 1.22-fold of that in hepatopancreas, respectively (Fig. 3A). The CgPGRPS4 mRNA transcripts were also detected in all the tested tissues with the highest expression level in hepatopancreas, which was 147.63-fold ($p < 0.01$) of that in mantle (Fig. 3B). The following high expression was observed in haemocytes, which was 34.72-fold ($p < 0.01$) of that in mantle, respectively (Fig. 3B). The relative expression levels of CgPGRPS4 in adductor muscle, gill, heart and gonad were 8.31-fold ($p < 0.01$), 2.66-fold, 2.35-fold and 1.76-fold of that in mantle, respectively (Fig. 3B). Additionally, the basal mRNA expression levels of CgPGRPS2 in haemocytes and adductor muscle were 42.42-fold ($p < 0.01$) and 348.02-fold ($p < 0.01$) of CgPGRPS4, respectively, while that of CgPGRPS4 in hepatopancreas was 0.99-fold ($p > 0.05$) of CgPGRPS2.

3.3. The recombinant proteins of CgPGRPS2 and CgPGRPS4, and their antibodies

After IPTG induction for 4 h, the whole cell lysate of *E. coli* Transetta (DE3) with pET-30a-CgPGRPS2 and pET-30a-CgPGRPS4 were analyzed by SDS-PAGE. Two

bands of ~25 kDa (rCgPGRPS2 with His-tag) and ~29 kDa (rCgPGRPS4 with His-tag) were visualized by Coomassie brilliant blue staining, which were consistent with the predicted molecular mass. The rCgPGRPS2 and rCgPGRPS4 were purified by a Ni²⁺ chelating Sepharose column (Fig. 4A, lane 3 and Fig. 4B, lane 3) and used to prepare antibodies. Two clear reaction bands of rCgPGRPS2 and rCgPGRPS4 with high specificity were observed in Western blotting (Fig. 4A, lane 4 and Fig. 4B, lane 4).

3.4. Cellular localization of CgPGRPS2 and CgPGRPS4 in oyster haemocytes

Fluorescence confocal microscopy was employed to detect the subcellular localization of endogenous CgPGRPS2 and CgPGRPS4 in oyster haemocytes. The nucleus stained by DAPI was observed in blue and the CgPGRPS2 and CgPGRPS4 immunoreactive areas were stained in green. The positive signals of both CgPGRPS2 and CgPGRPS4 appeared mainly in the cytoplasm (Fig. 5).

3.5. PAMPs binding activities of rCgPGRPS2 and rCgPGRPS4

PAMPs binding assay of rCgPGRPS2 and rCgPGRPS4 was performed based on the absorbance at 450 nm. The apparent dissociation constant (K_d) values were calculated using Prism 5.00 software (GraphPad software) with a one-site binding model and nonlinear regression analysis. Compared to rCgPGRPS4, rCgPGRPS2 showed a higher binding affinity to LPS in a dose-dependent manner, and the K_d of rCgPGRPS2 and rCgPGRPS4 to LPS was 3.26×10^{-6} M and 8.76×10^{-6} M, respectively (Fig. 6A and B). Both rCgPGRPS2 and rCgPGRPS4 showed a high

binding affinity to PGN in a dose-dependent manner, and their K_d was 3.84×10^{-6} M and 4.78×10^{-6} M, respectively (Fig. 6C and D). In addition, rCgPGRPS2 showed a high binding affinity to Man in a dose-dependent manner ($K_d = 3.31 \times 10^{-6}$) (Fig. 6E), while rCgPGRPS4 did not bind Man (data not shown). In general, rCgPGRPS2 exhibited broader and stronger PAMP binding activities than rCgPGRPS4.

3.6. Binding activities of rCgPGRPS2 and rCgPGRPS4 to microbes

To further investigate *in vitro* microbial binding activity, rCgPGRPS2 and rCgPGRPS4 were incubated with formaldehyde-killed Gram-negative bacteria (*E. coli* and *V. anguillarum*), Gram positive bacteria (*S. aureus*) and yeast (*Y. lipolytica*). After stringent washes, the proteins bound to the microbes were eluted and subjected to Western Blot analysis. Purified rCgPGRPS2 and rCgPGRPS4 were set as positive control which shows a clear band (Fig. 7A and B, lane 5), while no significant signal was detected in the negative control (data not shown). Both rCgPGRPS2 and rCgPGRPS4 were detected in *E. coli*, *V. anguillarum* and *S. aureus* groups, and rCgPGRPS2 showed higher binding affinity towards *E. coli* and *S. aureus* than rCgPGRPS4 (Fig. 7A and B, lane 1, 2 and 3). In the *Y. lipolytica* group, there was significant band for rCgPGRPS2, while no band was detected for rCgPGRPS4 (Fig. 7A and B, lane 4).

4 Discussion

Since the first PGRP was identified in 1996, a larger number of PGRPs have been

identified in many organisms which share high conservation and contain at least one PGRP domain in the C-terminal amino acid sequences (Yoshida et al., 1996; Werner et al., 2000). In the present study, CgPGRPS2 and CgPGRPS4 both contained one signal peptide and one typical PGRP/amidase domain (Fig. 1A), and they shared similar structural characteristics to the short extracellular PGRPs previously identified (Werner et al., 2000; Ni et al., 2007; Su et al., 2007; Itoh et al., 2009; Dai et al., 2015; Sun and Sun, 2015). Multiple sequence alignments revealed significant sequence identity of PGRP domains (from 38.7 to 76.4%) within the PGRP protein family in oysters. It was noteworthy that four catalytic residues responsible for amidase activity were well conserved in both CgPGRPS2 and CgPGRPS4 (55H, 90Y, 164H, 172C in CgPGRPS2, and 98H, 133Y, 207H, 215C in CgPGRPS4) (Fig. 2), while four residues involved in specific PGN recognition showed a certain degree of variation (84R, 85W, 104R, 109V in CgPGRPS2, and 127G, 128W, 147R, 152V in CgPGRPS4) (Fig. 2), and this was consistent with the previous study by Zhang and Yu (Zhang and Yu, 2013). Moreover, 13 possible substrate binding sites in DmPGRPs showed a varied degree of sequence identity in CgPGRPs (Fig. 2). These structural features suggested that CgPGRPS2 and CgPGRPS4 could function as PRRs to mediate protein-ligand and protein-protein interactions and as effectors to eliminate pathogen similar to other PGRPs.

Although CgPGRPS2 and CgPGRPS4 shared high homology and their structural features were highly similar to each other, their mRNA expression profile was obviously different. Their mRNA transcripts were constitutively expressed in

haemocytes, hepatopancreas, mantle, gonad, heart, adductor muscle and gill, indicating that they could be involved in some essential physiological processes of oysters (Fig.3). The highest mRNA expression level of CgPGRPS2 was found in adductor muscle (Fig.3A). Though similar results were observed in *Branchiostoma japonicum* (Yao et al., 2012) and *Cynoglossus semilaevis* (Sun and Sun, 2015), the highest expression of CgPGRPS2 in the non-immune organ of adductor muscle was still beyond our expectation. It is suspected that CgPGRPS2 may be involved in the fundamental physiological processes of oysters. Consistent with the expression profile of HcPGRPS1 (Yang et al., 2013), HcPGRP1 and HcPGRP1a (Ye et al., 2014) from *Hyriopsis cumingi* and SgPGRP-S1 from *Solen grandis* (Wei et al., 2012), CgPGRPS4 was found with the highest expression level in hepatopancreas (Fig.3B). And the relative higher mRNA expression level of both CgPGRPS2 and CgPGRPS4 was observed in haemocytes (Fig.3), similar to AiPGRP (Ni et al., 2007). As haemocytes and hepatopancreas comprise the major component of the non-specific defense mechanisms and are involved in a series of cellular immune reactions in molluscs (Song et al., 2010; Mello et al., 2012; Wang et al., 2014), the high expression of CgPGRPS2 and CgPGRPS4 in hepatopancreas and haemocytes indicated their involvement in the cellular immune response of oyster.

Both of CgPGRPS2 and CgPGRPS4 proteins were mainly observed in the cytoplasm of oyster haemocytes (Fig.5). The intracellular localization of protein with signal peptide was also observed in the splice variants of gcPGRP6 from *Ctenopharyngodon idella* (Yu et al., 2014), and the possible explanation was that the

secretion of these proteins was related to the invasion of pathogens or to some specific cell types. The different tissue distribution CgPGRPS2 and CgPGRPS4 mRNA transcripts and similar cellular localization of their proteins were speculated to be related with their potential functions.

Though PGRPs were originally described as PGN-binding and recognition proteins, some of them could bind various PAMPs or microbes to play indispensable roles in the innate immune system (Dziarski, 2004). For example, rPGRP-S from *B. japonicum* could bind Dap-type PGN, Lys-type PGN, chitin, and microorganisms including *E. coli*, *S. aureus* and *Pichia pastoris* (Yao et al., 2012), rCPGRPs from *Camelus dromedarius* could bind LPS, LTA and fatty acids (Sharma et al., 2012, 2013a and 2013b), and gcPGRP6 splice variants from *C. idella* were able to bind microbial PAMPs including Lys-type PGN, DAP-type PGN, GLU, Man, and microorganisms including *Streptococcus dysgalactiae*, *Flavobacterium columnare* and *S. cerevisiae* (Yu et al., 2014). In the present study, both rCgPGRPS2 and rCgPGRPS4 showed high binding affinity to Lys-type PGN and Gram-positive bacteria *S. aureus* in a dose-dependent manner (Fig.6C and D, Fig.7). These results were consistent with most of the PGRPs (Dziarski, 2004; Yang et al., 2010; Yao et al., 2012; Yu et al., 2014). It was reported that rCgPGRP-S1S could bind DAP-type PGN specifically (Iizuka et al., 2014), indicating the functional differentiation of oyster PGRP genes in the PGN recognition. Moreover, rCgPGRPS2 and rCgPGRPS4 could bind LPS and Gram-negative bacteria including *E. coli* and *V. anguillarum* (Fig.6A and B, Fig.7), which was similar to rCfPGRP-S1 (Yang et al., 2010), rPGRP-S (Yao

et al., 2012), rCPGRPs (Sharma et al., 2013b) and rCgPGRP-S1S (Iizuka et al., 2014).

And rCgPGRPS2 exhibited a higher binding affinity to LPS than rCgPGRPS4.

RCgPGRPS2 also showed a high binding affinity to Man (Fig.6E), as well as to

fungus *Y. lipolytica* (Fig.7), while rCgPGRPS4 showed no binding activity to them

(Fig.7). There was some controversy on the binding activities of PGRPs to fungus and

their PAMPs. For example, rPGRP-S could bind *Pichia pastoris* (Yao et al., 2012),

and gcPGRP6 splice variants were able to bind GLU, Man, and *S. cerevisiae* (Yu et

al., 2014), while rCfPGRP-S1 did not bind GLU (Yang et al., 2010). The present

results indicated that rCgPGRPS2 exhibited broader and stronger binding activities

towards various PAMP and bacteria than rCgPGRPS4, while rCgPGRPS4 seems

preferential binding to PGN. In brief, the difference in pattern recognition capacity

and microbial binding activities further confirmed the hypothesis that CgPGRPS2 and

CgPGRPS4 might play different roles in the fundamental physiological processes and

innate immune response of oysters.

According to the transcriptional analysis after LPS, PGN and GLU challenge, it

was learned that the mRNA expression of CgPGRPS2 were regulated by LPS, PGN

and GLU, and was significant up-regulated by LPS. The mRNA expression profile of

CgPGRPS2 were consistent with toll-like receptor 2, toll-like receptor 13, MyD88

and nuclear factor NF- κ B p105 subunit-like, indicating that CgPGRPS2 might be

involved in the activation of the *C. gigas* Toll pathway after LPS challenge (Fig.S3)

(Michel et al., 2001; Lemaitre et al., 2007). Moreover, CgPGRPS2 shared similar

mRNA expression profile with peroxinectin (CGI_10028272), indicating that it might

be involved in the processes of cell adhesion and degranulation (Johansson et al., 1995; Sritunyalucksana et al., 2001). However, there was no significant differential expression of CgPGRPS4 after LPS, PGN and GLU challenge (Fig.S2). On the contrary, the mRNA transcripts of CgPGRPS4 showed high expression level from the early D-shaped larvae to juvenile according to the transcriptional analysis by Zhang et al. (Zhang et al., 2012), while that of CgPGRPS2 could be hardly detected during the whole embryonic development stages (Fig.S4). Combined transcriptional analysis with the PAMPs and microbes binding activities of CgPGRPS4, it was speculated that CgPGRPS4 might exhibit some direct immune activity during the larval development.

In conclusion, two PGRPs with similar structure were cloned from oyster *C. gigas*, and they have highly homology, similar but not identical phylogenesis, yet different in expression profile and pattern recognition capacity. The CgPGRPS2 and CgPGRPS4 mRNA transcripts were constitutively expressed in all the detected tissues with the highest expression level in adductor muscle and hepatopancreas, respectively. Both CgPGRPS2 and CgPGRPS4 proteins appeared mainly in the cytoplasm. The rCgPGRPS2 protein could bind three typical PAMPs (LPS, PGN and Man) and three kinds of microorganisms including *E. coli*, *V. anguillarum*, *S. aureus* and *Y. lipolytica* efficiently, while rCgPGRPS4 showed a low binding affinity to LPS and failed to bind Man and *Y. lipolytica*. All these results indicated that CgPGRPS2 and CgPGRPS4 showed difference in gene expression patterns and PAMP binding activities, especially in binding of Man. As the knowledge of the binding mechanism of PGRP to LPS and Man is still limited, their different binding activities to Man

needs to be further studied.

Acknowledgements

We are grateful to all the laboratory members for their technical advice and helpful discussions. This research was supported by grants from National Science Foundation of China (Nos. 41306158, 31530069), Dalian high level talent innovation support program (2015R020), High Technology Project (863 Program, No. 2014AA093501) from the Chinese Ministry of Science and Technology, and the Research Foundation for Talented Scholars in Dalian Ocean University (to L. S.).

References

- Basbous, N., Coste, F., Leone, P., Vincentelli, R., Royet, J., Kellenberger, C., Roussel, A., 2011. The *Drosophila* peptidoglycan-recognition protein LF interacts with peptidoglycan-recognition protein LC to downregulate the Imd pathway. *EMBO reports* 12, 327-333.
- Beutler, B., 2004. Innate immunity: an overview. *Mol. Immunol.* 40, 845-859.
- Bischoff, V., Vignal, C., Duvic, B., Boneca, I.G., Hoffmann, J.A., Royet, J., 2006. Down-regulation of the *Drosophila* immune response by peptidoglycan-recognition proteins SC1 and SC2. *Plos Pathog.* 2, 139-147.
- Cao, A., Mercado, L., Ramos-Martinez, J.I., Barcia, R., 2003. Primary cultures of hemocytes from *Mytilus galloprovincialis* Lmk.: expression of IL-2R α subunit. *Aquaculture* 216, 1-8.

485 Cheng, S., Zhan, W., Xing, J., Sheng, X., 2006. Development and characterization of
486 monoclonal antibody to the lymphocystis disease virus of Japanese flounder
487 *Paralichthys olivaceus* isolated from China. J. Virol. Methods 135, 173-180.

488 Choe, K.M., Werner, T., Stöven, S., Hultmark, D., Anderson, K.V., 2002.
489 Requirement for a peptidoglycan recognition protein (PGRP) in relish activation
490 and antibacterial immune responses in *Drosophila*. Science 296, 359-362.

491 Coteur, G., Mellroth, P., De Lefortery, C., Gillan, D., Dubois, P., Communi, D.,
492 Steiner, H., 2007. Peptidoglycan recognition proteins with amidase activity in early
493 deuterostomes (Echinodermata). Dev. Comp. Immunol. 31, 790-804.

494 Dai, L.S., Qian, C., Wang, L., Wei, G.Q., Liu, Q.N., Sun, Y., Zhang, C.F., Li, J., Liu,
495 D.R., Zhu, B.J., Liu, C.L., 2015. Characterization and function of a short
496 peptidoglycan recognition protein from the Chinese oak silkworm, *Antheraea*
497 *pernyi*. J. Asia-Pac. Entomol. 18, 701-707.

498 Dziarski, R., 2004. Peptidoglycan recognition proteins (PGRPs). Mol. Immunol. 40,
499 877-886.

500 Guan, R., Wang, Q., Eric, J.S., Roy, A.M., 2005. Crystal structure of human
501 peptidoglycan recognition protein S (PGRP-S) at 1.70 Å resolution. J. Mol. Biol.
502 347, 683-691.

503 Iizuka, M., Nagasaki, T., Takahashi, K.G., Osada, M., Itoh, N., 2014. Involvement of
504 Pacific oyster CgPGRP-S1S in bacterial recognition, agglutination and granulocyte
505 degranulation. Dev. Comp. Immunol. 43, 30-34.

506 Itoh, N., Takahashi, K.G., 2008. Distribution of multiple peptidoglycan recognition

507 proteins in the tissues of Pacific oyster, *Crassostrea gigas*. Comp. Biochem. Phys.
 508 B. 150, 409-417.

509 Itoh, N., Takahashi, K.G., 2009. A novel peptidoglycan recognition protein containing
 510 a goose-type lysozyme domain from the Pacific oyster, *Crassostrea gigas*. Mol.
 511 Immunol., 2009, 46: 1768-1774.

512 Janeway, C., Medzhitov, R., 2002. Innate immune recognition. Sci STKE. 20, 197.

513 Jia, Z., Zhang, T., Jiang, S., Wang, M., Cheng, Q., Sun, M., Wang, L., Song, L., 2015.
 514 An integrin from oyster *Crassostrea gigas* mediates the phagocytosis toward *Vibrio*
 515 *splendidus* through LPS binding activity. Dev. Comp. Immunol. 53, 253-264.

516 Johansson, M.W., Lind, M.I., Holmblad, T., Thornqvist, P.O., Söderhäll, K., 1995.
 517 Peroxinectin, a novel cell adhesion protein from crayfish blood.
 518 Biochem. Bioph. Res. Co. 216, 1079-1087.

519 Kang, D., Liu, G., Lundström, A., Gelius, E., Steiner, H., 1998. A peptidoglycan
 520 recognition protein in innate immunity conserved from insects to humans. Proc.
 521 Natl. Acad. Sci. USA 95, 10078-10082.

522 Kim, M.S., Byun, M.J., Oh, B.H., 2003. Crystal structure of peptidoglycan
 523 recognition protein LB from *Drosophila melanogaster*. Nat. Immunol. 4, 787-793.

524 Lee S.Y., Söderhäll K., 2001. Characterization of a pattern recognition protein, a
 525 Masquerade-like protein, in the freshwater crayfish *Pacifastacus leniusculus*. J.
 526 Immunol. 166:7319-7326.

527 Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*.
 528 Annu. Rev. Immunol. 25, 697-743.

529 Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J.H., Caroff, M., Lee, W.J.,
530 Mengin-Lecreulx, D., Lemaitre B., 2003. The *Drosophila* immune system detects
531 bacteria through specific peptidoglycan recognition. *Nat. Immunol.* 4, 478-484.

532 Liu, C., Xu, Z., Gupta, D., Dziarski, R., 2001. Peptidoglycan recognition proteins: a
533 novel family of four human innate immunity pattern recognition molecules. *J. Biol.*
534 *Chem.* 276, 34686-34694.

535 Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using
536 real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 251, 402-408.

537 Li X, Wang S, Qi J, Echtenkamp, S.F., Chatterjee, R., Wang, M., Boons, G.J.,
538 Dziarski, R., Gupta, D., 2007. Zebrafish peptidoglycan recognition proteins are
539 bactericidal amidases essential for defense against bacterial infections. *Immunity*
540 27, 518-529.

541 Lu, X., Wang, M., Qi, J, Wang, H., Li, X., Gupta, D., Dziarski, R., 2006.
542 Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J.*
543 *Biol. Chem.* 281: 5895-5907.

544 Maillet, F., Bischoff, V., Vignal, C., Hoffmann, J., Royet, J., 2008. The *Drosophila*
545 peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK
546 pathway activation. *Cell Host Microbe* 3, 293-303.

547 Mello, D.F., de Oliveira, E.S., Vieira, R.C., Simoes, E., Trevisan, R., Dafre, A.L.,
548 Barracco, M.A., 2012. Cellular and transcriptional responses of *Crassostrea gigas*
549 hemocytes exposed in vitro to brevetoxin (PbTx-2). *Mar. Drugs* 10, 583-597.

550 Mellroth, P., Karlsson, J., Steiner, H., 2003. A scavenger function for a *Drosophila*

551 peptidoglycan recognition protein. J. Biol. Chem. 278, 7059-7064.

552 Mellroth, P., Steiner, H., 2006. PGRP-SB1: An N-acetylmuramoyl L-alanine amidase
553 with antibacterial activity. Biochem. Bioph. Res. C. 350, 994-999.

554 Michel, T., Reichhart, J.M., Hoffmann, J.A., Royet, J., 2001. Drosophila Toll is
555 activated by Gram-positive bacteria through a circulating peptidoglycan recognition
556 protein. Nature 414, 756-759.

557 Ni, D., Song, L., Wu, L., Chang, Y., Yu, Y., Qiu, L., Wang, L., 2007. Molecular
558 cloning and mRNA expression of peptidoglycan recognition protein (PGRP) gene
559 in bay scallop (*Argopecten irradians*, Lamarck 1819). Dev. Comp. Immunol. 31,
560 548-558.

561 Royet, J., Gupta, D., Dziarski, R., 2011. Peptidoglycan recognition proteins:
562 modulators of the microbiome and inflammation. Nat. Rev. Immunol. 11, 837-851.

563 Rämets, M., Manfrulli, P., Pearson, A., Mathey-Prevot, B., Ezekowitz, R.A.B., 2002.
564 Functional genomic analysis of phagocytosis and identification of a *Drosophila*
565 receptor for *E-coli*. Nature 416, 644-648.

566 Sharma P., Dube D., Sinha M., Dey S., Kaur P., Sharma S., Singh T.P., 2012.
567 Structural basis of heparin binding to camel peptidoglycan recognition protein-S.
568 Int. J. Biochem. Mol. Biol. 3, 86-94.

569 Sharma, P., Yamini, S., Dube, D., Singh, A., Mal, G., Pandey, N., Sinha, M., Singh,
570 A.K., Dey, S., Kaur, P., Mitra, D.K., Sharma, S., Singh, T.P., 2013a. Structural basis
571 of the binding of fatty acids to peptidoglycan recognition protein, PGRP-S through
572 second binding site. Arch. Biochem. Biophys. 529, 1-10.

573 Sharma, P., Dube, D., Sinha, M., Yadav, S., Kaur, P., Sharma, S., Singh, T.P., 2013b.

574 Structural insights into the dual strategy of recognition by peptidoglycan

575 recognition protein, PGRP-S: structure of the ternary complex of PGRP-S with

576 lipopolysaccharide and stearic acid. PLoS ONE 8, e53756.

577 Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano,

578 M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement

579 of Protein Using Bicinchoninic Acid. Anal. Biochem. 150, 76-85.

580 Song, L., Wang, L., Qiu, L., Zhang, H., 2010. Bivalve immunity. In: Söderhäll K (eds)

581 Invertebrate Immunity. Springer-Verlag, New York, USA, pp 44-65.

582 Sritunyalucksana, K., Wongsuebsantati, K., Johansson, M.W., Söderhäll, K., 2001.

583 Peroxinectin, a cell adhesive protein associated with the proPO system from the

584 black tiger shrimp, *Penaeus monodon*. Dev. Comp. Immunol. 25, 353-363.

585 Steiner, H., 2004. Peptidoglycan recognition proteins: on and off switches for innate

586 immunity. Immunol. Rev. 198, 83-96.

587 Su, J., Ni, D., Song, L., Zhao, J., Qiu, L., 2007. Molecular cloning and

588 characterization of a short type peptidoglycan recognition protein (CfPGRP-S1)

589 cDNA from Zhikong scallop *Chlamys farreri*. Fish Shellfish Immunol. 23,

590 646-656.

591 Sun, Q., Sun, L., 2015. A short-type peptidoglycan recognition protein from tongue

592 sole (*Cynoglossus semilaevis*) promotes phagocytosis and defense against bacterial

593 infection. Fish Shellfish Immunol. 47, 313-320.

594 Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T., Kurata,

595 S., 2002. Overexpression of a pattern-recognition receptor, peptidoglycan
 596 recognition protein-LE, activates imd/relish-mediated antibacterial defense and the
 597 prophenoloxidase cascade in *Drosophila* larvae. Proc. Natl. Acad. Sci. USA 99,
 598 13705-13710.

599 Tao, Y., Yang, Z., Zhang, X., Wu, H., 2014. Molecular cloning and mRNA
 600 expression of the peptidoglycan recognition protein gene HcPGRP1 and its isoform
 601 HcPGRP1a from the freshwater mussel *Hyriopsis cumingi*. Genet. Mol. Biol. 37,
 602 508-517.

603 Tydell, C.C, Yuan, J., Tran, P., Selsted, M.E., 2006. Bovine peptidoglycan recognition
 604 protein-S I antimicrobial activity, localization, secretion and binding properties. J.
 605 Immunol. 176, 1154-1162.

606 Udompetcharaporn, A., Junkunlo, K., Senapin, S., Roytrakul, S., Flegel, T.W.,
 607 Sritunyalucksana, K., 2014. Identification and characterization of a QM protein as a
 608 possible peptidoglycan recognition protein (PGRP) from the giant tiger shrimp
 609 *Penaeus monodon*. Dev. Comp. Immunol. 46, 146-154.

610 Wang, Z., Li, X., Cocklin, R.R., Wang, M., Wang, M., Fukase, K., Inamura, S.,
 611 Kusumoto, S., Gupta, D., Dziarski, R., 2003. Human peptidoglycan recognition
 612 protein-L is an N-acetylmuramoyl-L-alanine amidase. J. Biol. Chem. 278,
 613 49044-49052.

614 Wang, M., Liu, L.H., Wang, S., Li, X., Lu, X., Gupta, D., Dziarski, R., 2007. Human
 615 peptidoglycan recognition proteins require zinc to kill both Gram-positive and
 616 Gram-negative bacteria and are synergistic with antibacterial peptides. J.

617 Immunol.178, 3116-3125.

618 Wang, W., Wu, X., Liu, Z., Zheng, H., Cheng, Y., 2014b. Insights into
619 hepatopancreatic functions for nutrition metabolism and ovarian development in
620 the crab *Portunus trituberculatus*: gene discovery in the comparative transcriptome
621 of different hepatopancreas stages. PLoS ONE 9, e84921.

622 Wei, X., Yang, J., Yang, D., Xu, J., Liu, X., Yang, J., Fang, J., Qiao, H., 2012.
623 Molecular cloning and mRNA expression of two peptidoglycan recognition protein
624 (PGRP) genes from mollusk *Solen gradis*. Fish Shellfish Immunol. 32, 178-185.

625 Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H., Hultmark, D., 2000. A family
626 of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*.
627 Proc. Natl. Acad. Sci. USA 97, 13772-13777.

628 Yang, J., Wang, W., Wei, X., Qiu, L., Wang, L., Zhang, H., Song, L., 2010.
629 Peptidoglycan recognition protein of *Chlamys farreri* CfPGRP-S1 mediates
630 immune defenses against bacterial infection. Dev. Comp. Immunol. 34, 1300-1307.

631 Yang, Z., Li, J., Wu, H., Li, Y., Wang, X., 2013. Molecular cloning and functional
632 characterization of a short peptidoglycan recognition protein (HcPGRPS1) from the
633 freshwater mussel, *Hyriopsis cumingi*. Mol. Immunol. 56, 729-738.

634 Yao, F., Li, Z., Zhang, Y., Zhang, S., 2012. A novel short peptidoglycan recognition
635 protein in amphioxus: Identification, expression and bioactivity. Dev. Comp.
636 Immunol. 38, 332-341.

637 Yoshida, H., Kinoshita, K., Ashida, M., 1996. Purification of a peptidoglycan
638 recognition protein from hemolymph of the silkworm, *Bombyx mori*. J. Biol. Chem.

639 271, 13854-13860.

640 Yu, X., Gan, H., Kanost M.R., 1999. Immulectin, an inducible C-type lectin from an
641 insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. Insect
642 Biochem. Mol. Biol. 29, 585-597.

643 Yu, Z., Li, J., Xue, N., Nie, P., Chang, M., 2014. Expression and functional
644 characterization of PGRP6 splice variants in grass carp *Ctenopharyngodon idella*.
645 Dev. Comp. Immunol. 47, 264-274.

646 Yu, Y., Yu, Y., Huang, H., Feng, K., Pan, M., Yuan, S., Huang, S., Wu, T., Guo, L.,
647 Dong, M., Chen, S., Xu, A., 2007. A short-form C-type lectin from amphioxus acts
648 as a direct microbial killing protein via interaction with peptidoglycan and glucan. J.
649 Immunol. 179, 8425-8434.

650 Zhang, Y., Yu, Z., 2013. The first evidence of positive selection in peptidoglycan
651 recognition protein (PGRP) genes of *Crassostrea gigas*. Fish Shellfish Immunol. 34,
652 1352-1355

653 Zhang, H., Song, X., Wang, L., Kong, P., Yang, J., Liu, L., Qiu, L., Zhang, Y., Qiu, L.,
654 Song, L., 2011. AiCTL-6, a novel C-type lectin from bay scallop *Argopecten*
655 *irradians* with a long C-type lectin-like domain. Fish Shellfish Immunol. 30, 17-26.

656 Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X.,
657 Qi, H., Xiong, Z., Que, H., Xie, Y., Peter, W.H.H., Jordi, P., Zhu, Y., Wu, F., Chen,
658 Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z.,
659 Zhu, Q., Feng, Y., Andrew, M., Dennis, H., Xu, Z., Liu, Y., Tomislav, D. Du, Y.,
660 Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W.,

Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y.,
Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M.,
Patrick, M. G., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, S.,
Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, J., Wang, Q., Christian,
E.W.S., Wang, H., Li, N., Qian, L., Zhang, G., Li, Y., Yang, H., Liu, X., Wang, J.,
Yin, Y., Wang, J., 2012. The oyster genome reveals stress adaptation and
complexity of shell formation. *Nature* 490, 49-54.

Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J. Zhang, G., 2015. Massive
expansion and functional divergence of innate immune genes in a protostome.
SCI REP-UK. 5: 8693.

Zhou, Z., Wang, M., Zhao, J., Wang, L., Gao, Y., Zhang, H., Liu, R., Song, L., 2013.
The increased transcriptional response and translocation of a Rel/NF- κ B
homologue in scallop *Chlamys farreri* during the immune stimulation. *Fish*
Shellfish Immunol. 34, 1209-1215.

Figure legends

Fig. 1. Domain organization (A) and phylogenetic tree (B) of CgPGRPS2, CgPGRPS4 and other PGRPs from *Homo sapiens*, *Danio rerio*, *Drosophila melanogaster* and *C. gigas*. Sequence information of 4 HsPGRPs, 5 DrPGRPs, 17 DmPGRPs and 9 CgPGRPs are described in Table S1. The tree is constructed by the neighbor-joining (NJ) algorithm using the MEGA4.1 beta software based on the multiple sequence alignment by ClustalW. The reliability of the branching was tested

by bootstrap re-sampling (1000 pseudo-replicates). The scale bar corresponds to 0.05 estimated amino-acid substitutions per site.

Fig. 2. Multiple alignment of the amino acid sequences for amidase/PGRP domains. Amino acid residues that are conserved in at least 60% sequences are shaded in dark, and similar amino acids are shaded in gray. Catalytic residues responsible for amidase activity and specific PGN recognition activity were indicated by inverted triangle (▼) and dots (●), respectively. The possible substrate binding sites were marked by asterisks.

Fig. 3. Tissue distribution of CgPGRPS2 (A) and CgPGRPS4 (B) mRNA transcripts detected by qRT-PCR. The transcript level in haemocytes, hepatopancreas, mantle, gonad, heart, muscle and gill of six adult oysters is normalized to that of hepatopancreas. The EF1- α gene was used as an internal control to calibrate the cDNA template for each sample. Vertical bars represent mean $\pm$ S.D. ($n \geq 4$), and bars with different characters were significantly different ($p < 0.05$).

Fig. 4. SDS-PAGE and Western-blot analysis of rCgPGRPS2 (A) and rCgPGRPS4 (B). M: protein molecular standard; Lane 1: negative control for rCgPGRPS2 and rCgPGRPS4 (without induction); Lane 2: induced rCgPGRPS2 and rCgPGRPS4; Lane 3: purified rCgPGRPS2 and rCgPGRPS4; Lane 4: western-blot of rCgPGRPS2 and rCgPGRPS4.

705

706 **Fig. 5. Cellular localization of CgPGRPS2 (A) and CgPGRPS4 (B) in oyster**
707 **haemocytes.** After the haemocytes were incubated adhering to the cell culture dish,
708 immunohistochemistry was performed to detect the expression of CgPGRPS2 and
709 CgPGRPS4. A1, B1: The nucleuses of haemocytes were stained by DAPI (blue); A2,
710 B2: Bright field; A3, B3: CgPGRPS2 and CgPGRPS4 (green); A4, B4: Overlay 1, 2
711 and 3. (For interpretation of the references to color in this figure legend, the reader is
712 referred to the web version of this article.)

713

714 **Fig. 6. The binding activity of rCgPGRPS2 and rCgPGRPS4 to different PAMPs.**
715 A, C, E: rCgPGRPS2 exhibited high binding affinity to LPS ($K_d = 3.26 \times 10^{-6}$ M),
716 PGN ($K_d = 3.84 \times 10^{-6}$ M) and Man ($K_d = 3.31 \times 10^{-6}$ M), respectively. B, D:
717 rCgPGRPS4 showed low binding affinity to LPS ($K_d = 8.76 \times 10^{-6}$ M) and high
718 binding affinity to PGN ($K_d = 4.78 \times 10^{-6}$ M). Data are shown as the mean $\pm$ S.D. (N
719 = 3).

720

721 **Fig. 7. Binding activity of rCgPGRPS2 (A) and rCgPGRPS4 (B) to different**
722 **microbes.** Lane 1: *Escherichia coli* incubated with rCgPGRPS2 and rCgPGRPS4;
723 lane 2: *Vibrio anguillarum* incubated with rCgPGRPS2 and rCgPGRPS4; lane 3:
724 *Staphylococcus aureus* incubated with rCgPGRPS2 and rCgPGRPS4; lane 4:
725 *Yarrowia lipolytica* incubated with rCgPGRP2; lane 5: purified rCgPGRPS2 and
726 rCgPGRPS4 as positive control.

Table 1. Primers used in this study.

Primer name	Sequence (5'-3')
Primers for gene clone	
Oligo(dT)-adaptor	GGCCACGCGTCGACTAGTACT ₁₇
CgPGRPS2-F	ATGAGGATTGAAGCTCTTTATTTTGC
CgPGRPS2-R	TTATACACAGTATATGTCTGGTTTCCACAG
CgPGRPS4-F	ATGTTGATTGCTTTATTTTTTTCTTGC
CgPGRPS4-R	TTAAGGATGTGTGTGGCTGTAATGA
Primers for RT PCR	
CgPGRPS2-RTF	GAGCACGGAGACCCAAAAGTTC
CgPGRPS2-RTR	CCGATGGACTGGCTATTGTAACC
CgPGRPS4-RTF	TACAGCCGTGGACATTGGAGG
CgPGRPS4-RTR	GCGGTATGGTGGATGAAGAACA
CgEF1- α -RTF	AGTCACCAAGGCTGCACAGAAAG
CgEF1- α -RTR	TCCGACGTATTCTTTGCGATGT
Primers for protein recombination	
CgPGRPS2-ReF	AACGGGATCCTGTTCCAATGTACGGATCTA
CgPGRPS2-ReR	TTACTCGAGTTATACACAGTATATGTCTGG
CgPGRPS4-ReF	AACGGGATCCAGTGATGCTCCTTGTCATAACATT
CgPGRPS4-ReR	TTACTCGAGTTAAGGATGTGTGTGGCTGTAATG

Figure1

A

CgPGRPS2 (AB425337.1)

CgPGRPS4 (EKC26200)

B

Figure2

Figure3

Figure4

Figure5

A

B

Figure6

Figure7

A

1

2

3

4

5

B

1

2

3

4

5