Thesis for the Degree of Master of Engineering

Molecular characterization and expression of two types of crustin-like proteins in the Morotoge shrimp, *Pandalopsis japonica*

by

Ah Ran Kim

Interdisciplinary Program of Biomedical Engineering

The Graduate School

Pukyong National University

February 2014

Molecular characterization and expression of two types of crustin-like proteins in the Morotoge shrimp, *Pandalopsis japonica*

(물렁가시붉은새우의 2 가지 유형의 crustin 유사 단백질의 분자적인 특성과 발현에 관한 연구)

Advisor: Prof. Hyun-Woo Kim

By

Ah Ran Kim

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Engineering

In Interdisciplinary Program of Biomedical Engineering,

The graduate School,

Pukyong National University

February 2014

Molecular characterization and expression of two types of crustin-like proteins in the Morotoge shrimp, *Pandalopsis japonica*

A dissertation

by

Ah Ran Kim

Approved by:

_____

(Chairman) Dr. Hak Jun Kim

_____ _____

(Member) Dr. Yoon Kwon Nam **(Member) Dr. Hyun-Woo Kim**

February 21, 2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Molecular characterization and expression of two types of crustin-like proteins in the Morotoge shrimp, *Pandalopsis japonica*

Ah Ran Kim

Interdisciplinary Program of Biomedical Engineering, The Graduate School

Pukyong National University

Abstract

Crustins are cysteine-rich cationic antimicrobial peptides (AMPs) found in decapod crustaceans. Total nine crustin genes (Paj-CrusIa, Ib, Ic, Id, Ie, If, IIa, IIb and IIc) were identified in the morotoge shrimp, *Pandalopsis japonica* as result of bioinformatic analysis and PCR-based cloning strategy. Deduced amino acid sequences of isolated Paj-Crus genes ranged from 99 to 178 amino acid residues (10.6–17.8 kDa). Sequence analysis of nine isolated Paj-Crus genes and 100 different crustins from various decapod crustaceans revealed that a splice site and KXXXCP motif within the WAP domain may be the main criteria for classifying type I and II crustins, suggesting that the two types of crustin genes may have been generated by different processes. We also identified three intron-less crustin I genes (Paj-Crus Id, Ie and If) for the first time, which may have been generated by gene

duplication. The tissue distribution profiles showed that Paj-CrusI genes were expressed predominantly in the gill and epidermis, whereas Paj-CrusII genes were expressed ubiquitously, suggesting that the two types of crustins may play different roles in various tissues or under different physiological conditions. In order to know the functional differences between two types of crustins, four recombinant crustins (Paj-CrusIa, Ib, IIa, IIc) were expressed in *E. coli* and then purified with his-tag column. All four purified recombinant crustins showed no antimicrobial activity against both Gram-positive and Gram-negative bacteria.

INTRODUCTION

Generally, the crustacean immune system is considered to lack a specific adaptive immune system as in vertebrates and likely relies mostly on innate immune responses (Rosa et al., 2007). Antimicrobial peptides (AMPs) play an important role in innate immunity in crustaceans, and fifteen AMP family members have been identified (Rosa and Barracco, 2010). Crustins are small cationic AMPs (7–14 kDa), which contain high numbers of cysteine and proline residues and a whey acidic protein (WAP) domain at the carboxyl terminus. They show their highest antimicrobial activity against mostly gram-positive bacteria (Amparyup et al., 2008; Relf et al., 1999; Smith et al., 2008; Zhang et al., 2007), but also against some gram-negative bacteria (Amparyup et al., 2008).

Since the first discovery of a crustin family member from hemocytes of *Carcinus maenas* (Relf et al., 1999), homologous sequences have been identified in most decapod crustaceans, including Dendrobranchiata (Antony et al., 2010; Jia et al., 2008; Rojtinnakorn et al., 2002; Rosa et al., 2007; Sun et al., 2010; Supungul et al., 2004; Vatanavicharn et al., 2009), Caridea (Dai et al., 2009), Anomura (Sperstad et al., 2009), Achelata (Pisuttharachai et al., 2009), Astacidea (Christie et al., 2007; Hauton et al., 2006; Jiravanichpaisal et al., 2007) and Brachyura (Afsal et al., 2011; Imjongjirak et al., 2009; Mu et al., 2011; Sperstad et al., 2009; Yue et al., 2010). As various crustin-like sequences accumulated, Smith *et al.* (2008) classified them into three types (type I, II and III) based on differences in primary

structure. Besides the commonly conserved motifs in all three types of crustins, including an N-terminal signal peptide sequence, cysteine-containing region and a C-terminal WAP domain, an additional glycine-rich region was identified between the signal peptide sequence and cysteine-rich region of type II crustins. Type III crustins contain a unique, short proline/arginine-rich region between the signal peptide sequence and WAP domain instead of a glycine-rich or cysteine-containing region.

Despite development of a number of methods of classification, none use the highly variable N-terminal region of each crustin. Additionally, functional differences among the three types of crustins are not clear. In this study, we identified six genes in addition to the previous three crustin genes from *Pandalopsis japonica* (Kim et al., 2012). Their structure was compared with other crustins from various decapods and identified several characteristics to distinguish type I from type II crustins. Expression analysis also revealed the correlation between gene organization and expression profile, suggesting that the two types of crustin genes have been generated by different processes and so their biological function may also differ. We also constructed and purified the recombinant Paj-CrusIb and Paj-CrusIIc proteins using bacterial expression system and their antimicrobial activities were measured.

MATERIALS AND METHODS

Cloning of six full-length cDNAs encoding crustin from *Pandalopsis japonica*

To isolate partial contigs with similarity to crustin, sequence similarity analysis was performed using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast) using sequences from *Panulirus japonicus* (ACU25384) and *Penaeus monodon* (ABW82154) as bait against the cDNA database, which was generated from the transcriptome of neuronal tissues (brain, X-organ/sinus gland (XO/SG) and thoracic and abdominal ganglia) of *Pandalopsis japonica* (Kim et al., 2012).

Live animals were purchased from a local seafood market in Busan, Korea. Before sacrificing the animals, hemocytes were harvested as described previously (Relf et al., 1999). Immediately after sacrificing the animals, tissues were collected, frozen by placement into liquid nitrogen and stored in a deep freezer (-70°C) until use. Total RNA was extracted from various tissues (*i.e.*, brain, abdominal ganglia, thoracic ganglia, flexor muscle, extensor muscle, heart, gill, epidermis, hepatopancreas, gonad and hemocytes) using RNAiso Plus reagent according to the manufacturer's protocol (Takara Bio Inc., Japan). To avoid contamination by genomic DNA, all extracted total RNA was treated with RNase-free DNase I (Takara Bio Inc., Japan) at 37°C for 30 min. RNA quantity, purity and integrity were measured spectrophotometrically at 260 nm (Nanodrop Technologies, Inc., USA) and confirmed by electrophoresis on 1% agarose gels. Isolated total RNAs were stored at -70°C until use.

cDNA was synthesized in a reaction containing reverse transcriptase and oligo-dT primer. A 15-μl mixture of 5-μg total RNA, 1 μl of 20 μM oligo-dT primer and 4-μl dNTPs (2.5mM/ μL) was heated to 70°C for 5 min and chilled on ice for 2 min. First-strand buffer (5×, 5 μl), 2 μl of 0.1 M DTT and RNaseOUT (Invitrogen Co., USA) were added to the mixture, which was then incubated at 37°C for 2 min. Finally, 1 μl M-MLV reverse transcriptase (Invitrogen Co., USA) was added and the mixture was incubated at 37°C for 50 min. Synthesized cDNA was quantified spectrophotometrically and stored at -20°C.

The full-length cDNA sequence of each contig was obtained by the 5' and 3' rapid amplification of cDNA ends (RACE) strategy, as described previously (Jeon et al., 2010). All primers used in this experiment were designed using the IDTSciTools program (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/) and were synthesized by Bioneer Corporation, Korea. Briefly, cDNA was synthesized using a reverse primer (Table. 1). The 3'end of the full-length mRNA was identified by 3'RACE, as described previously (Jeon et al., 2010). The 5'end of full-length mRNA was identified by 5'RACE. The amplified PCR products were analyzed by electrophoresis in a 1.5% agarose gel and stained with ethidium bromide. The PCR products with the expected sizes were purified using Gel Extraction Kit (Bioneer Co., Korea) and ligated into TA plasmid vector with the pGEM-T Easy Cloning Kit (Promega, USA). Vectors containing cloned inserts were transformed into *E. coli* DH5α and incubated overnight at 37°C. Positive

clones were then screened and sequenced in both directions using M13 forward and reverse primers (Table. 1). Sequences from RACE were assembled with original contigs by computer-aided homology analysis and the full-length of each sequence was reconfirmed by RT-PCR using two sequence-specific primers, targeted at each end of the full sequence. RT-PCR results confirmed that all six sequences were real transcripts.

Determination of six Paj-Crus gene sequences

To identify differences in the gene organization of the Paj-Crus genes, the genomic DNA sequences of the six obtained crustin genes were analyzed. Genomic DNA was extracted from abdominal flexor muscle tissue using an AccuPrep® Genomic DNA extraction kit according to the manufacturer's protocol (Bioneer Co., Korea). Sequence-specific primers were designed to amplify the whole ORF sequence with the 5' and 3' UTR of each crustin gene (Table. 1). The first PCR was carried out with a total volume of 30-μl reaction mixture containing 100 ng of genomic DNA as a template, 1 μl of 10 μM each forward and reverse primer (Table. 1), 10× Ex Taq buffer, 2 μl of dNTP mixture (2.5 mM each) and 0.5 μl of Ex Taq DNA polymerase (5 U/μl) (Takara Bio Inc., Japan). PCR conditions were 94°C for 5 min followed by 35 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 2 min and a final cycle of 72°C for 5 min. Nested PCR was carried out using 1 μl of the first PCR product as a template and the PCR conditions were the same as those for

the first PCR. Sequence confirmation procedures for genomic DNA PCR products were as for the cloning of RACE products.

Expression analysis of six Paj-Crus genes

The mRNA expression of the six Paj-Crus genes in various tissues, including brain, abdominal ganglia, thoracic ganglia, flexor muscle, extensor muscle, heart, gill, epidermis, hepatopancreas, gonad and hemocytes, were determined by end-point RT-PCR. Total RNA from various tissues was extracted and quantified as described above. cDNA was synthesized as described for the cloning of crustin cDNA, except for random hexamers, which were used as primers for reverse transcription.

Pairs of Paj-Crus sequence-specific primers (Table. 1) were used to amplify products of the six crustin genes with predicted sizes of 370 bp for Paj-CrusIc, 261 bp for Paj-CrusId, 231 bp for Paj-CrusIe, 302 bp for Paj-CrusIf, 283 bp for Paj-CrusIIb and 285 bp for Paj-CrusIIc. Primers were designed to target each exon to avoid possible genomic contamination. 18s rRNA primers were used as an internal control. Each reaction (30 μl) contained 2 μl of cDNA (200ng), 1 μl of 10 μM gene-specific forward primer and reverse primer, respectively (Table. 1), 0.5 μl of Takara Taq DNA polymerase (5 U/μl) (Takara Bio Inc., Japan), 2 μl of dNTP mixture (2.5 mM each) and 3 μl of 10× PCR buffer. PCR conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec

and a final cycle of 72°C for 5 min. After 30 reaction cycles, the expected PCR products were detected by agarose gel electrophoresis, as described for the cloning of crustin cDNA. The sequence-specific primers for each target gene were designed using IDTSciTools (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/).

To compare the copy numbers of each crustin transcript, quantitative PCR was performed with hemocyte, gill and epidermis tissues. The method for total RNA extraction and first-strand cDNA synthesis from different tissues was the same as that used for tissue distribution analysis. PCR amplifications were carried out in a total volume of 20 μl containing 10 μl of 2× SYBR Green premix Ex Taq™ (Takara Bio Inc., Japan), 3 μl of 300-ng cDNA template and 1 μl of 10 μM of each gene-specific forward and reverse primer (Table. 1). The PCR conditions were 94°C for 1 min, followed by 40 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 30 sec using DNA Engine Choromo4 Real-time Detector (Bio-Rad Laboratories, Inc., USA). Specific primers (Table. 1) were used to amplify an 18S rRNA gene fragment as an internal control. Standard curves were constructed to confirm the efficiency of primers and copy numbers were normalized to the 18S rRNA copy number according to the equation: (actual copy numbers of Paj-CrusIc, Id, Ie, If, IIb and IIc / actual copy number of 18S rRNA) × 10^{10}.

Molecular modeling and computational analysis of the six Paj-Crus genes

The open reading frame (ORF) of each cDNA was identified and the deduced

amino acid sequence was determined by ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). Multiple amino acid sequence alignment was performed using ClustalW software (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and represented by the GeneDoc program. The signalP 3.0 program was used to predict the presence and location of signal peptide (http://www.cbs.dtu.dk/services/SignalP/). Molecular structures of Paj-CrusIa and Paj-CrusIIb were modeled by Rosetta (Leaver-Fay et al., 2011). In the modeling we assumed that the four disulfide bonds formed by the eight conserved cysteine residues in the WAP domain were fixed according to known structures (Hu et al., 2005; Torres et al., 2003; Tsunemi et al., 1996). For the four cysteine residues in the Cys-containing region, models were constructed based on all 10 possible disulfide bond formations. We modeled 1) no disulfide bond formation in the Cys-containing region, 2) one disulfide bond formation with six possible combinations: bond between the first and the second Cysteine residues (called 1-2), bond between the first and the third Cys (1-3), 1-4, 2-3, 2-4 and 3-4 and 3) two disulfide bond formations in three possible combinations: 1-2 3-4, 1-3 2-4 and 1-4 2-3.

Expression and purification of recombinant Paj-Crus proteins

cDNA encoding the mature peptide region of each crustin (Paj-CrusIa, Ib, IIa, IIc) was amplified by PCR using the sequence-specific primers (Table 1). The

amplified PCR product of each crustin was confirmed by electrophoresis on 1% agarose gels and purified using Gel Extraction Kit (Bioneer Co., Korea). The purified PCR products were digested and cloned into the pET28a vector (Clontech Laboratories, Inc., USA) at the NdeI/XhoI sites. Recombinant plasmids were transformed into *E. coli* BL21 (DE3) and inserted crustin genes were reconfirmed by the commercial DNA sequencing service (Macrogen, Korea). In order to express each recombinant protein, the transformed cells were cultured at 20℃ in LB medium up to 0.8 in OD_{600}. 1mM IPTG (isopropyl-β-D-thiogalactopyranoside) was added to the culture for the protein induction. After overnight culture, cells were harvested, resuspended in 1X phosphate-buffered saline, pH 7.4 (PBS) and sonicated for 5 min. The lysates were centrifuged and both soluble supernatant and insoluble pellet fractions were analyzed by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

As result of SDS-PAGE, I identified all three crustins in both soluble and insoluble fraction (Fig. 11). Soluble fractions were purified using Ni-NTA agarose column (Qiagen, Germany) according to manufacturer's instructions (Clontech Laboratories, Inc., USA). For concentration, eluted protein was concentrated using centrifugal filter device (EMD Millipore Co., USA) at 4℃ for 1 hour. The concentration of the purified recombinant proteins were measured by both the Bradford's assay (Bradford, 1976) and Western blot. The purified proteins were

separated by the SDS-PAGE and electro-transferred to a PVDF membrane (EMD Millipore Co., USA) at 30V for 1h. Transferred membrane was washed with TBS-T (TBS with Tween-20) and incubated in a blocking buffer (TBS-T with 1.0% BSA) at 25℃ for overnight. The membrane was incubated with primary antibodies, 1ng of anti-His (Cell biolabs, Inc., USA) at 25℃ for 1h and secondary antibody, anti-mouse IgG (EMD Millipore Co., USA) at 25℃ for 1h. The recombinant proteins were developed using the Immunoblotting Kit (EMD Millipore Co., USA).

Antimicrobial activity assay

The antimicrobial activity of purified recombinant Paj-CrusIb and Paj-CrusIIc, was determined by the liquid growth inhibition assay (Supungul et al., 2008). Antimicrobial activities of crustins were tested against four Gram-positive bacteria (*Bacillus subtilis, Enterococcus faecalis, Listeria monocytogenes, Staphylococcus aureus*) and four Gram-negative bacteria (*Escherichia coli, Pseudomonas aeruginosa, Vibrio anguillarum, Vibrio parahaemolyticus*). The mid-logarithmic growth phase cultures of bacteria were diluted with poor broth (1% bactotryptone, 0.5% NaCl, pH 7.5) to an OD_{600} of 0.001. The two-fold serially diluted proteins and negative control (PBS, pH 7.4) was inoculated with diluted bacteria cultures. After vigorous shaking overnight, bacterial growth was measured at OD_{630}.

Table 1. Primers used for the Paj-Crus gene study.

Primer	Sequence (5'-3')	Description
linker oligo-dT	GCACGCTGTGAATGCTGCGACTACGATTTTTTT TTTTTTTTTT	cDNA synthesis
linker R1	ACGCTGTGAATGCTGCGACTAC	First reverse primer for 3'RACE
linker R2	GCTGTGAATGCTGCGACTACGA	Second reverse primer for 3'RACE
Paj-crus Ic F1	GCTGACGAGCATCTAGGAGAATGC	First forward primer for 3'RACE
Paj-crus Ic F2	CGAGCATCTAGGAGAATGCCCCAAG GGA	Second forward primer for 3'RACE
Paj-crus Id F1	GGTTTTGGGAACTACATCTGCTG	First forward primer for 3'RACE
Paj-crus Id F2	CTACATCTGCTGTGATGATCATCCTGG	Second forward primer for 3'RACE
Paj-crus Ie F1	CTCTTGGTCAGAAGTCGCCTCATCTC	First forward primer for 3'RACE
Paj-crus Ie F2	GTCGCCTCATCTCCTGGAATCATAT	Second forward primer for 3'RACE Forward primer for genomic DNA PCR Forward primer for RT-PCR & real-time PCR
Paj-crus If F1	TTGGCAACTACATCTGCTGCGATG	First forward primer for 3'RACE
Paj-crus If F2	CTGCGATGACCATCCTGGTCAATG	Second forward primer for 3'RACE
Paj-crus IIb F1	ATGTTCACCAAGGCGTTTATGTCCG	First forward primer for 3'RACE Forward primer for RT-PCR & real-time PCR
Paj-crus IIb F2	CGTTTATGTCCGTGGCCTTTATGG	Second forward primer for 3'RACE
Paj-crus IIc F1	GGCTTCGGAGGTGGATTGGG	First forward primer for 3'RACE
Paj-crus IIc F2	GGAGGTGGATTGGGTCCCAG	Second forward primer for 3'RACE
Paj-crus Ic RT F1	CGGTCAACTCTCTGAAGAAGTGTCTC	Forward primer for RT-PCR & real-time PCR
Paj-crus Ic RT R1	CAAGTGCTGTTTCGACACCTGTCT	Reverse primer for RT-PCR & real-time PCR
Paj-crus Id RT F1	CTTGACTCGAAAGTCGCCTCTCTC	Forward primer for RT-PCR & real-time PCR
Paj-crus Id RT R1	CATTTAGACCAGTTTGTCTGGGTCTTCC	Reverse primer for RT-PCR & real-time PCR
Paj-crus Ie RT R1	AAGTGTCCTGCTGTTCGACCAATC	Reverse primer for RT-PCR & real-time PCR
Paj-crus IIb RT R1	CATTGGAGGTGTTGGGGCCATC	Reverse primer for RT-PCR & real-time PCR
Paj-crus IIc RT F1	ATCATCGTTGCGTCTACAACATC	Forward primer for RT-PCR & real-time PCR Forward primer for genomic DNA PCR
Paj-crus IIc RT R1	GTCCTCCAGTTCGACCAACATGTC	Reverse primer for RT-PCR & real-time PCR
Paj-crus Ic gF1	CGGTCAACTCTCTGAAGAAGTGTCTC	Forward primer for genomic DNA PCR
Paj-crus Ic gR1	GGTGTAAACTAAATCGAGTTTTCTATCATTTGT GTG	Reverse primer for first genomic DNA PCR
Paj-crus Ic gR2	TCATCACACCTGCAAATCCGTCCT	Reverse primer for second genomic DNA PCR
Paj-crus Id gF1	GCTCAGTCGCACTTGACTCGAAAG	Forward primer for genomic DNA PCR
Paj-crus Id gR1	GACGGGTAAGCCTCTATTTTGATCGTT	Reverse primer for first genomic DNA PCR
Paj-crus Id gR2	CAGACGGGTAAGCCTCTATTTTGATCG	Reverse primer for second genomic DNA PCR
Paj-crus Ie gR1	CGAAATGGGAAGCTCACAAACTGTATAC	Reverse primer for first genomic DNA PCR
Paj-crus Ie gR2	CACTGATATGAGGATAACTTCACATCGAAATG	Reverse primer for second genomic DNA PCR
Paj-crus If gF1	CCAAGACTCAGTTTCACCTGATCCG	Forward primer for genomic DNA PCR

Paj-crus If gR1	CACTTGGGCATTGCTTTCAAGGTC	Reverse primer for first genomic DNA PCR
Paj-crus If gR2	GAGAGACACTCTCTTCACACATGTCAAAG	Reverse primer for second genomic DNA PCR
Paj-crus IIb gF1	CCAATAGCACCAGCAATCGCAGATAC	Forward primer for genomic DNA PCR
Paj-crus IIb gR1	GCAAATGAAGTGCAAGTTATTGACCAAG	Reverse primer for first genomic DNA PCR
Paj-crus IIb gR2	GTCATATCAATTGCAAATGAAGTGCAAG	Reverse primer for second genomic DNA PCR
Paj-crus IIc gR1	CAATAGTAAATAGACAGCCTTGGTTGAGC	Reverse primer for first genomic DNA PCR
Paj-crus IIc gR2	GTTGTTCCTCGTGAATCAAGTAATTG	Reverse primer for second genomic DNA PCR
Paj-crus Ia NdeI	GAGATATACATATGCTCCCCGCG	Forward primer for inserting NdeI site
Paj-crus Ia XhoI	CAACGTCTCGAGTTAGTGGTGGTGGTGGTGGT GGTAGTTAGTACACTTCTTATA	Reverse primer for inserting XhoI site and His-tag
Paj-crus Ib NdeI	GAGATATACATATGTCCCCCCGGT	Forward primer for inserting NdeI site
Paj-crus Ib XhoI	CAACGTCTCGAGTTAGTGGTGGTGGTGGTGGT GAAGAGGAACTTGTTCAATGAT	Reverse primer for inserting XhoI site and His-tag
Paj-crus IIa NdeI	GAGATATACATATGGAAGGACCTTCAGG	Forward primer for inserting NdeI site
Paj-crus IIa XhoI	CAACGTCTCGAGTCAGTGGTGGTGGTGGTGGT GCTCAGGCCCTTTGCAGATGAG	Reverse primer for inserting XhoI site and His-tag
Paj-crus IIc NdeI	GAGATATACATATGGCAACACCCAACA	Forward primer for inserting NdeI site
Paj-crus IIc XhoI	CAACGTCTCGAGTCAGTGGTGGTGGTGGTGGT GGTAATCAATAGGTGTTTC	Reverse primer for inserting XhoI site and His-tag
18S rRNA F	ATGAGAGTGCTCAAAGCAGGCTACTC	Forward primer for RT-PCR & Real-time PCR
18S rRNA R	GGCGAATCGCTAGTCAGCATCGTT	Reverse primer for RT-PCR & Real-time PCR
M13F(-40)	CAGGAAACAGCTATGAC	Vector forward primer for PCR screening & sequencing
M13R(-20)	GTAAAACGACGGCCAG	Vector reverse primer for PCR screening & sequencing

RESULTS

Cloning of cDNAs encoding six crustins

As a result of the bioinformatic similarity analysis and the PCR-based cloning strategy, six novel cDNAs encoding crustin were identified (Fig. 1, 2, 3, 4, 5, 6 and 7). A single transcript of each cDNA sequence and open reading frame (ORF) were confirmed by RT-PCR using primers for each end of the sequence (Table. 1). Each crustin cDNA was named by the primary structure proposed by Smith *et al*. (2008) and the previous nomenclature of crustins in *Pandalopsis japonica* (Kim et al., 2012). Differing from previous studies (Amparyup et al., 2008; Imjongjirak et al., 2009), we did not consider phylogenetic analysis for classification mainly due to the high degree of diversity in length and sequence within the N-terminal region and the very small WAP domain (~50 to 60 residues) (Fig. 7).

Since three different crustin genes (two type Is and one type II) were previously identified from this species (Kim et al., 2012), we named the new crustin genes according to the previous rule. However, In order not to raise any confusion from two-lettered nomemclature, we used three letters (first two for genus and third for species) for each crustin genes (see insect Biochemistry 18:785-787,1988). Therefore, we changed three previous crustin gene names in *Pandalopsis japonica* from Pj-crus Ia, Ib and II to Paj-CrusIa, Ib and IIa, respectively. The first cDNA had 557 bp and encoded 117 amino acid residues and showed 73% identity to Paj-CrusIb from *Pandalopsis japonica* (AFN80342). The cDNA was named Paj-CrusIc

(GenBank accession no; KC608994), because it showed the typical domain organization of type I crustins (Fig. 7). The other three crustins also exhibited high similarity in domain organization to those of type I crustins from Pleocyamata, and were named Paj-CrusId, Ie and If (GenBank accession no; KC608995, KC608996, KC608997, respectively). Paj-CrusId (812 bp), Paj-CrusIe (623 bp) and Paj-CrusIf (538 bp) encoded proteins with 104, 105 and 99 residues, respectively (Fig. 7 and Table 2). The other two crustins (Paj-CrusIIb and IIc) showed typical type II domain organizations, which comprise a glycine-rich region, a cysteine-containing region and a WAP domain (Fig. 7 and Table 2). The cDNAs of Paj-CrusIIb (781 bp) and Paj-CrusIIc (583 bp) encoded deduced proteins with 178 and 128 residues, respectively (GenBank accession no; KC608998, KC608999) and showed the highest sequence identities (59% and 75%) to crustins from *Macrobrachium rosenbergii* (AFO68120) and *Panulirus japonicas* (ACU25385).

To compare the primary structure of Paj-Crus to other crustins from various decapod crustaceans, multiple alignment was performed (Fig. 7). As in other secretory proteins, all six crustins from *Pandalopsis japonica* contained conserved signal peptide sequences with lengths of 16 to 26 residues (Fig. 7). A glycine-rich region was identified only in type II crustins, including Paj-CrusIIs, whereas a cysteine-containing region and WAP domain were highly conserved at the carboxyl-terminus of all crustins. Twelve conserved cysteine residues within the cysteine-containing region (four cysteine residues) and WAP domain (eight

cysteine residues) are representative of crustins (Fig. 7). To estimate the effects of disulfide bonds on structure, the number of residues between cysteine residues within the regions were analyzed (Table 3). Paj-Crus peptides exhibited 12 well-conserved cysteine residues, except in Paj-CrusIIa, which lacks two cysteine residues within the cysteine-containing region. With several exceptions, overall cysteine spacing was conserved: C_1-X_3-C_2-X_{7-9}-C_3-C_4-X_{6-16}-C_5-X_{6-9}-C_6-X_{7-19}-C_7-X_5-C_8-X_{5-6}-C_9-C_{10}-X_3-C_{11}-X_{5-7}-C_{12}, where X represents any amino acid residue (Table 3). The highest variable sites are C_4-C_5, which is located between a cysteine containing region and WAP domain, and C_6-C_7, which are the third and fourth cysteines within the WAP domain (Table 3), suggesting that these two sites make the greatest contributions to diversity in Paj-Crus. To date, the three dimensional structure of crustin has not been elucidated and the effects of the cysteine-containing and glycine-rich regions on crustin structure are not clear. To estimate the structural differences between the two types of crustins, we constructed representative models of both type I (Paj-CrusIa) and type II (Paj-CrusIIb) crustins (Fig. 8). In the lowest score models, formation of one disulfide bond in the Cys-containing region was recorded in both types of crustins. However, bond formation differed in that the second and third Cys formed a disulfide bond (2-3) in Paj-CrusIa, while the first and third Cys formed a disulfide bond in Paj-CrusIIb (Fig. 8). Therefore, the highly flexible glycine-rich region of type II crustins may affect the overall structure of the cysteine-containing region, including disulfide bond

formation.

Fig. 1.

The nucleotides and deduced amino acid sequences of the Paj-CrusIc gene. Four exons (capitals) are interrupted by three introns (lower case). The putative signal sequence is underlined, ORF is in bold and polyadenylation signal is underlined and in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

Fig. 2.

The nucleotides and deduced amino acid sequences of the Paj-CrusId gene. The exon is written in capitals. The putative signal sequence is underlined, ORF is in bold and the polyadenylation signal is underlined and in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

Fig. 3.

The nucleotide and deduced amino acid sequences of the Paj-CrusIe gene. The exon is written in capitals. The putative signal sequence is underlined, ORF is in bold and the polyadenylation signal is underlined and in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

Fig. 4.

The nucleotides and deduced amino acid sequences of the Paj-CrusIf gene. The exon is written in capitals. The putative signal sequence is underlined and the ORF is in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

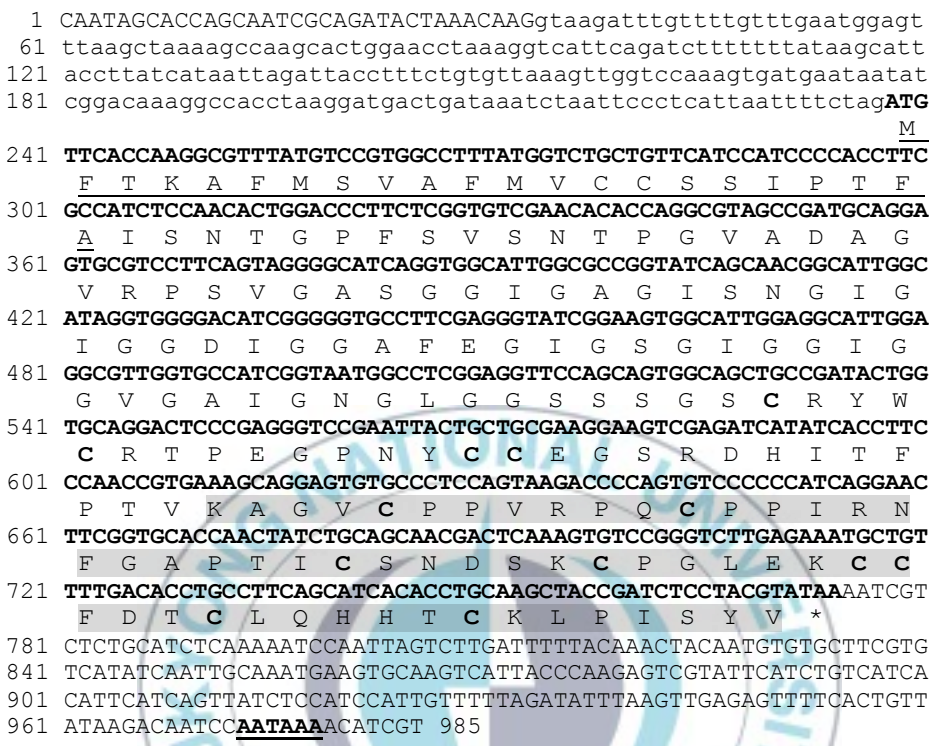

Fig. 5.

The nucleotides and deduced amino acid sequences of the Paj-CrusIIb gene. Two exons (capitals) are interrupted by one intron (lower case). The putative signal sequence is underlined, ORF is in bold and the polyadenylation signal is underlined and in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

Fig. 6.

The nucleotides and deduced amino acid sequences of the Paj-CrusIIc gene. Two exons (capitals) are interrupted by one intron (lower case). The putative signal sequence is underlined, ORF is in bold and the polyadenylation signal is underlined and in bold. The stop codon (TAA) is indicated by an asterisk. 12 cysteines forming six putative disulfide bridges are in bold and the WAP domain is shaded.

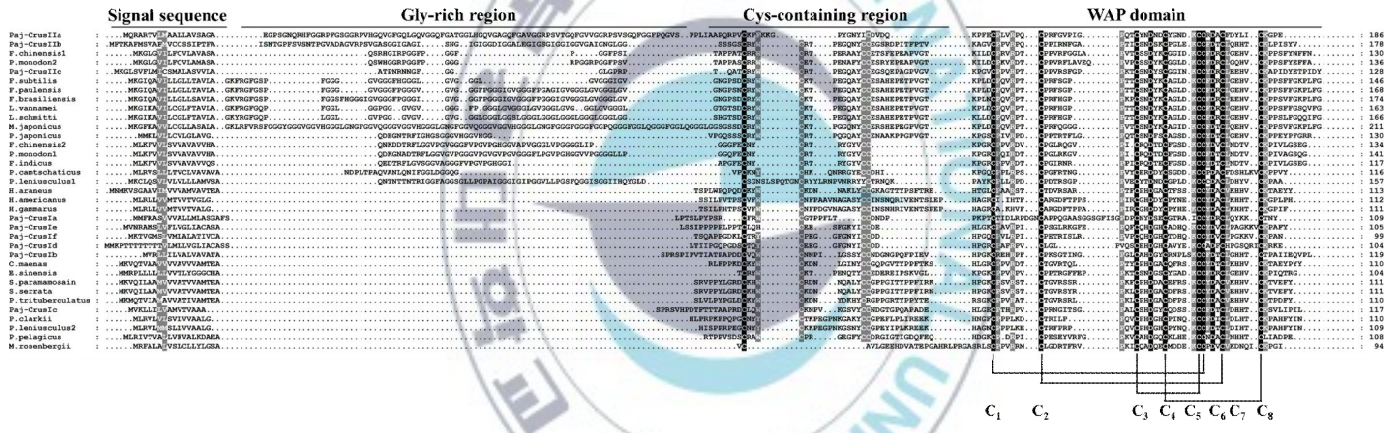
Signal sequence
Gly-rich region
Cys-containing region
WAP domain
C1
C2
C3
C4
C5
C6
C7
C8

Fig. 7. Amino acid sequence multiple alignment of Paj-CrusIa, Ib, Ic, Id, Ie, If, IIa, IIb and IIc with other crustacean crustin family genes. Black boxes indicate similarities in amino acid sequences. GenBank accession numbers: *Carcinus maenas*, CAD20734; *Eriocheir sinensis*, ACR77767; *Fenneropenaeus chinensis*, AAX63903, AAZ76017; *Farfantepenaeus paulensis*, ABM63361; *Farfantepenaeus brasiliensis*, ABQ96197; *Farfantepenaeus subtilis*, ABO93323; *Fenneropenaeus indicus*, ACV84092; *Homarus americanus*, ABM92333; *Homarus gammarus*, CAH10349; *Hyas araneus*, ACJ06763; *Litopenaeus schmitti*, ABM63362; *Litopenaeus vannamei*, AAS59736; *Macrobrachium rosenbergii*, ABQ41252; *Marsupenaeus japonicus*, BAD15064; *Pacifastacus leniusculus*, ABP88044, ABP88043; *Pandalopsis japonica*, JQ004015, JQ004016, JQ004017, KC608994, KC608995, KC608996, KC608997, KC608998, KC608999; *Paralithodes camtschaticus*, ACJ06765; *Panulirus japonicus*, ACU25382; *Penaeus monodon*, ABV25094, ABW82154; *Portunus pelagicus*, ABM65762; *Portunus trituberculatus*, ACO07303; *Procambarus clarkii*, AEB54630; *Scylla serrata*, ADW11096; *Scylla paramamosain*, ABY20728.

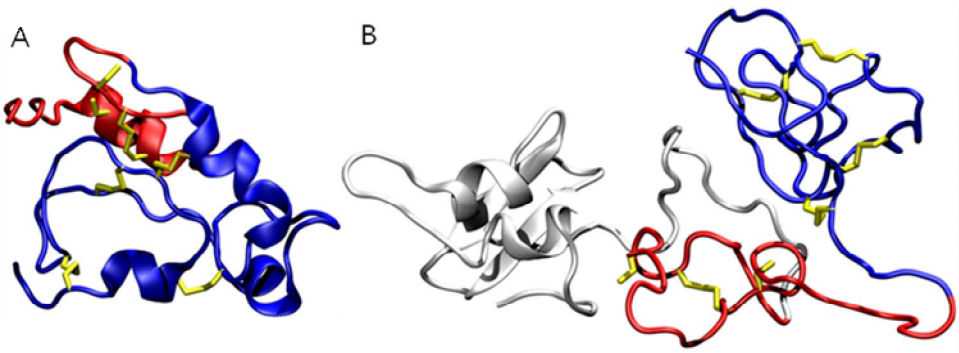

Fig. 8. Structural models of Paj-CrusIa (A) and Paj-CrusIIb (B) are represented by ribbons. The Gly-rich region, Cys-containing region and WAP domain are colored white, red and blue, respectively. Conserved cysteine residues and plausible disulfide bonds are represented by yellow sticks.

Table 2. Summary of characterized crustins in *Pandalopsis japonica.*

Peptides	Accession No.	Coded amino acids	Crustin Type	cDNA length	Number of exons	Reference
Paj-CrusIa	JQ004015	109	I	1102bp	4	Kim *et al*., 2012
Paj-CrusIb	JQ004016	119	I	452bp	3	Kim *et al*., 2012
Paj-CrusIc	KC608994	117	I	557bp	4	This study
Paj-CrusId	KC608995	104	I	812bp	1	This study
Paj-CrusIe	KC608996	105	I	623bp	1	This study
Paj-CrusIf	KC608997	99	I	538bp	1	This study
Paj-CrusIIa	JQ004017	186	II	663bp	3	Kim *et al*., 2012
Paj-CrusIIb	KC608998	178	II	781bp	2	This study
Paj-CrusIIc	KC608999	128	II	583bp	2	This study

Table 3. Number of amino acid residues between cysteines in the cysteine-containing region and WAP domain (cysteine-containing region; C_1-C_4, WAP domain; C_5-C_{12}).

Peptides	Number of amino acid residues between cysteines										
	C_1-C_2	C_2-C_3	C_3-C_4	C_4-C_5	C_5-C_6	C_6-C_7	C_7-C_8	C_8-C_9	C_9-C_{10}	C_{10}-C_{11}	C_{11}-C_{12}
Paj-CrusIa	3	7	0	8	9	19	5	5	0	3	5
Paj-CrusIb	3	9	0	16	6	11	5	6	0	3	5
Paj-CrusIc	3	9	0	16	6	11	5	5	0	3	5
Paj-CrusId	3	9	0	6	6	7	5	5	0	3	7
Paj-CrusIe	3	8	0	6	6	12	5	5	0	3	7
Paj-CrusIf	3	9	0	6	6	11	5	5	0	3	6
Paj-CrusIIa	12	-	-	8	6	11	5	5	0	3	5
Paj-CrusIIb	3	8	0	16	6	11	5	5	0	3	5
Paj-CrusIIc	3	8	0	16	6	11	5	5	0	3	5

Organization of Paj-Crus genes

To determine differences in gene organization, the genomic DNA sequences of Paj-Crus genes were compared with other currently known crustin genes, including *Portunus trituberculatus* (Cui et al., 2012), *Pandalopsis japonica* (Kim et al., 2012), *Scylla paramamosain* (Imjongjirak et al., 2009) and *Penaeus monodon* (Amparyup et al., 2008; Vatanavicharn et al., 2009) (Fig. 9). All analyzed crustin genes followed the typical splicing GT/AG rule (Shapiro and Senapathy, 1987). Introns were identified only in three Paj-Crus genes (Paj-CrusIc, IIb and IIc), whereas no intron was identified in the other three genes (Paj-CrusId, Ie and If). Paj-CrusIc (1285 bp) consisted of four exons and three introns (Fig. 9C). The first exon/intron boundary was between the first and second amino acid residue within the signal peptide sequence. The second splice site was in the cysteine-containing region and the third site was within the WAP domain. The same gene organization was also identified in Paj-CrusIa (Fig. 9A) and CrusSp (Fig. 9D) from *Scylla paramamosain* (GenBank Number; EU161288). Although four exons and three introns were also identified in Crustin*Pm*5 (GenBank Number; EF654659), the splice sites were different (Fig. 9L). Although one intron was identified in both Paj-Crus IIb and IIc (Fig. 9I and J), its location differed between the two genes. The splice site was outside the ORF in Paj-CrusIIb, which is similar to Crus-like*Pm* (GenBank Number: FJ380049). Gene organization of Paj-CrusIIc was unique in that splicing occurs between the second and third residue within the signal peptide sequence

(Fig. 9J). The previously identified Paj-CrusIIa also differed from other type II crustins because of the presence of two introns (Fig. 9H). Although type II crustins had diverse gene organization in which the number and location of splicing varies among individual crustins (Fig. 9), one common characteristic of the type II crustin gene is the absence of a splice site within the WAP domain. Interestingly, three intron-less crustins (Paj-CrusId, Ie and If) were identified, but have not been identified in other decapod species (Fig. 9D, 9K and 9L). Although their gene structure is unique, all belong to type I crustins with a cysteine-containing region and WAP domains according to the classification of Smith *et al*. (Smith et al., 2008). No type II crustin gene without introns has been identified in *Pandalopsis japonica*.

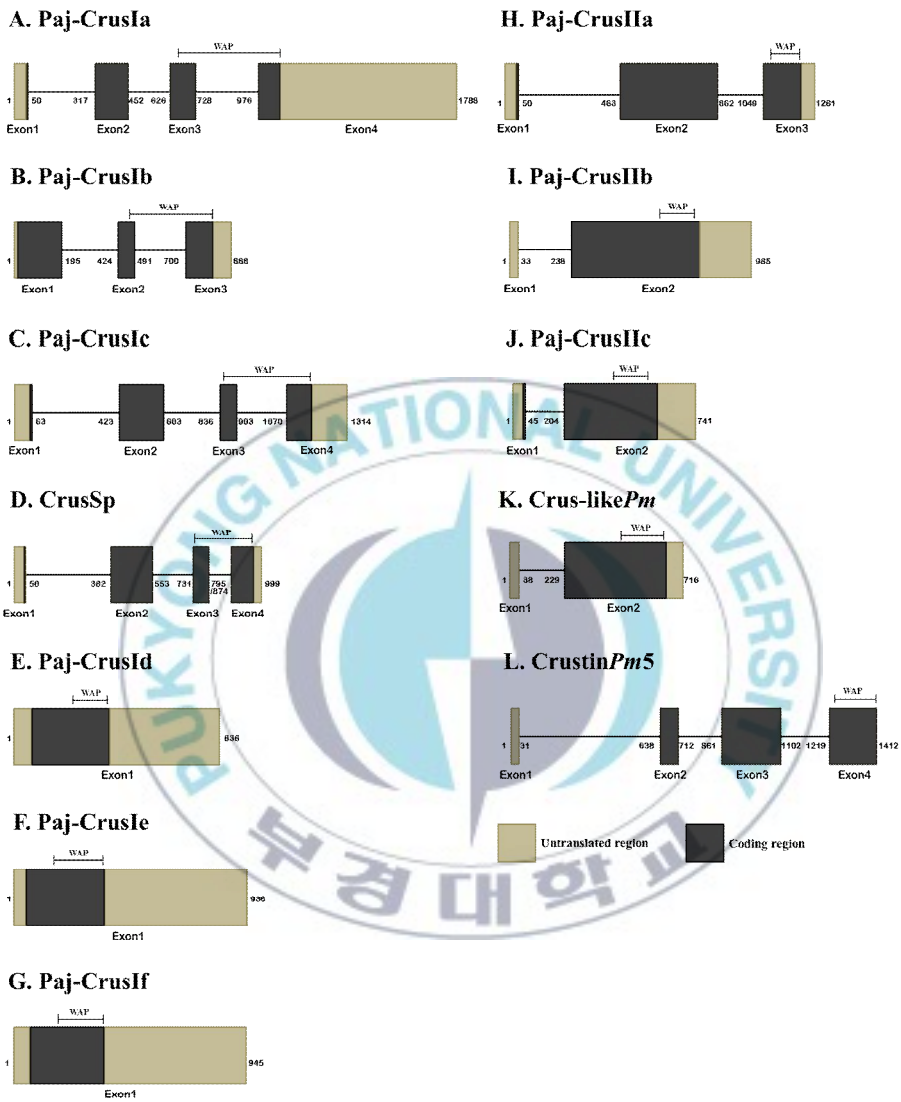

Fig. 9. Schematic representation of the genomic organization of Paj-CrusIc, Id, Ie, If, IIb and Paj-CrusIIc. The boxes represent exons and the dark grey boxes indicate ORFs. UTRs are shown as light grey boxes and exons are connected by introns (lines). Paj-CrusIa (JQ004015), Paj-CrusIb (JQ004016), Paj-CrusIIa (JQ004017), Crustin*Pm*5 (FJ380049), Crus-like*Pm* (EF654659) and Crus*Sp* (EU161288) are crustin genes from *Pandalopsis japonica*, *P. monodon* and *S. paramamosain* .

Expression analysis of Paj-Crus genes

To determine the relationship between gene organization and expression profile, RT-PCR was performed (Fig. 10). As in Paj-CrusIa and Ib (Kim et al., 2012), Paj-CrusIc was expressed predominantly in the gill and epidermis (Fig. 10). Intron-less Paj-CrusId and If also showed similar tissue-distribution profiles. Type II crustins, including Paj-CrusIIb and IIc, exhibited ubiquitous expression patterns, as with Paj-CrusIIa (Kim et al., 2012). Previous results showed that a major production site for crustin genes was the hemocyte (Amparyup et al., 2008; Donpudsa et al., 2010; Relf et al., 1999; Sperstad et al., 2009). We could not identify any crustin gene from *Pandalopsis japonica* that was expressed mainly in the hemocyte (Fig. 10). Similar results were also reported by us previously (Kim et al., 2012). To determine differences in the expression level of each crustin gene, quantitative PCR was performed (Table 3). As we expected in end-point RT-PCR, the expression level of Paj-CruIs is an order of magnitude higher than those of Paj-CrusIIs in the gill and epidermis. Transcripts of type I crustins in both the gill and epidermis are several hundred to thousand fold higher than in hematocytes, whereas those of type II crustins were less than 50 fold greater than or similar to those in hematocytes. Transcripts of Paj-CrusIc, Id, Ie, IIb and IIc in the gill were 482.45, 196.49, 489.95, 1.03 and 5.06 fold higher than in hemocytes, respectively, and those in the epidermis were 811.21, 729.25, 1425.87, 3.05 53.79 fold higher than in hemocytes, respectively (Table 3.).

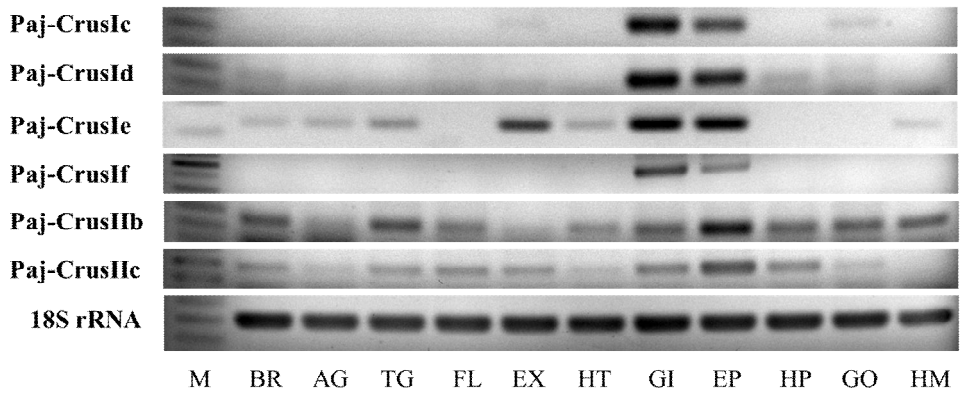

Fig. 10. Expression of Paj-CrusIc, Id, Ie, If, IIb and IIc in various tissues. End-point RT-PCR was carried out for 40 cycles and products were separated on a 1.5% agarose gel. 18S rRNA was used as the control. M, size marker; BR, brain; AG, abdominal ganglia; TG, thoracic ganglia; FL, flexor muscle; EX, extensor muscle; HT, heart; GI, gill; EP, epidermis; HP, hepatopanceas; GO, gonad; HM, hemocyte.

Table 4. Relative copy number of Paj-CrusIc, Paj-CrusId, Paj-CrusIe, Paj-CrusIIb and Paj-CrusIIc.

	Mean relative copy number	
Peptide	Gill / Hemocyte	Epidermis / Hemocyte
Paj-CrusIc	97394.80/ 201.88 ($P < 0.001$)	163763.81 / 201.88 ($P < 0.001$)
Paj-CrusId	13954.75 / 71.02 ($P < 0.001$)	51791.97 / 71.02 ($P < 0.01$)
Paj-CrusIe	23435.56 / 47.83 ($P < 0.001$)	68203.53 / 47.83 ($P < 0.001$)
Paj-CrusIIb	1529.25 / 1480.39 ($P > 0.05$)	4513.36 / 1480.39 ($P < 0.05$)
Paj-CrusIIc	34347.80 / 6787.78 ($P < 0.01$)	365085.90 / 6787.78 ($P < 0.001$)

Expression and purification of four crustin proteins

Expression of each recombinant crustins from *P. japonica* was determined by SDS-PAGE and Western blot (Fig. 11 and 12). Majority of proteins of Paj-CrusIa and Paj-CrusIIa were identified in the pellet fraction, whereas Paj-CrusIb and Paj-CrusIIc were identified both in pellet and supernatant fractions (Fig. 11). In order to confirm the proper size of recombinant proteins, western blot was performed with the soluble fraction (Fig. 12). Expected size of recombinant proteins were identified in two crustins, Paj-CrusIb (12 kDa) and Paj-CrusIIc (13 kDa). Two positive bands (14 kDa and 18 kDa) were detected in Paj-CrusIIa and little band was detected in Paj-CrusIa (Fig. 12). From the result, we identified only two crustins were properly expressed and they were used for further antimicrobial activities.

Antimicrobial activity assay

Compared with negative control, no conceivable antimicrobial activity was detected in neither Paj-CrusIb nor Paj-CrusIIc up to 40μM (Table 5).

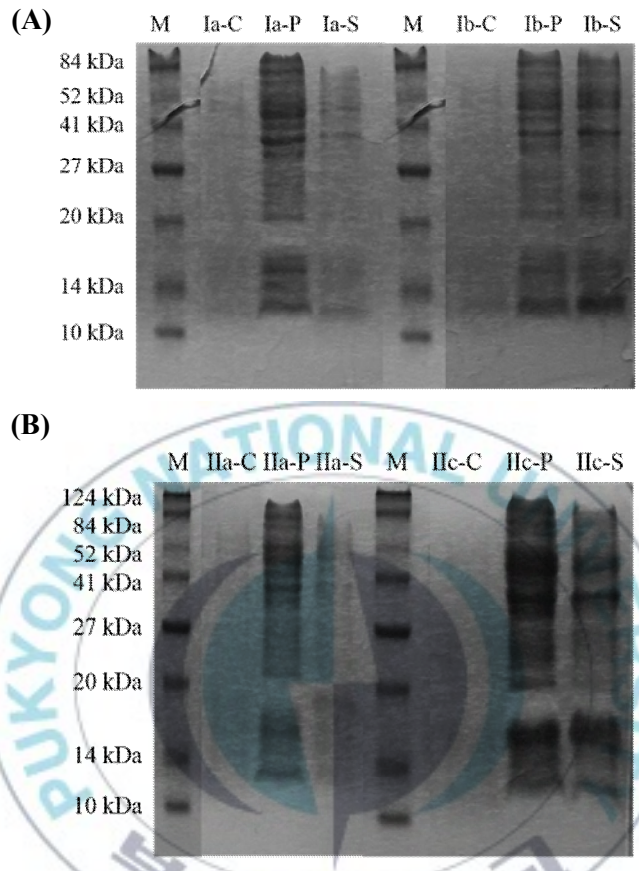

Fig. 11. SDS-PAGE analysis of the expressed crustins using *E. coli* system. (A) The lane Ia-C, Ib-C, Uninduced Paj-CrusIa, Ib ; lane Ia-P, Ib-P, the inclusion fraction of induced recombinant Paj-CrusIa, Ib ; lane Ia-S, Ib-S, the supernatant fraction of induced recombinant Paj-CrusIa, Ib. (B) The lane IIa-C, IIc-C, uninduced Paj-CrusIIa, IIc ; lane IIa-P, IIc-P, the inclusion fraction of induced recombinant Paj-CrusIIa, IIc ; lane IIa-S, IIc-S, the supernatant fraction of induced recombinant Paj-CrusIIa, IIc.

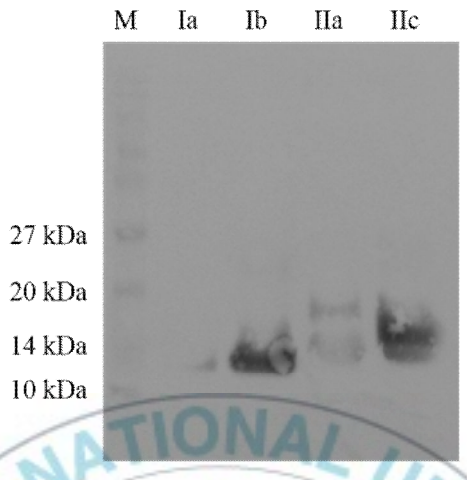

Fig. 12. Western blotting of purified protein. The protein detected with the anti-His antibodies and the second antibodies conjugated to horseradish peroxidase. M, Marker; Lane 1, Recombinant Paj-CrusIa ; Lane 2, Recombinant Paj-CrusIb ; Lane 3, Recombinant Paj-CrusIIa ; Lane 4, Recombinant Paj-CrusIIc.

Table 5. Antimicrobial activity assay of Paj-CrusIb, Paj-CrusIIc.

Microorganisms	MIC (μM)	
	Paj-CrusIb	Paj-CrusIIc
Negative Control	> 40 μM	> 40 μM
Gram-positive bacteria		
Bacillus subtilis	> 40 μM	> 40 μM
Enterococcus faecalis	> 40 μM	> 40 μM
Listeria monocytogenes	> 40 μM	> 40 μM
Staphylococcus aureus	> 40 μM	> 40 μM
Gram-Negative bacteria		
Escherichia coli	> 40 μM	> 40 μM
Pseudomonas aeruginosa	> 40 μM	> 40 μM
Vibrio anguillarum	> 40 μM	> 40 μM
Vibrio parahaemolyticus	> 40 μM	> 40 μM

MIC (Minimum inhibitory concentration) are expressed as the interval a-b, a is the highest concentration that microorganisms are growing and b is the lowest concentration that cause 100% growth inhibition.

DISCUSSION

From the results of the bioinformatic analysis and typical cloning strategy, a total of nine crustin genes (six type Is and three type IIs) have been identified in *Pandalopsis japonica*. In fact, we identified several additional partial contigs from our database, which exhibit high similarity to crustins (data not shown), and without concrete evidence of expression by RT-PCR these contigs were excluded. In the GenBank database, 20 cDNAs and two full or partial genomic DNA sequences were identified in *Penaeus monodon* and five cDNAs and eleven full or partial genomic DNA sequences were also found in *Portunus trituberculatus*, suggesting similar numbers of crustin genes may exist in *Pandalopsis japonica*. Although the biological implications of the various crustins in each species are not clear, their structural variation may contribute to the response to various pathogens. The effects of structural differences on bacterial specificity are not known and so further investigation is necessary.

More than 100 nucleotide sequences encoding crustin homologs have been identified in the GenBank database, which we classified them as suggested by Smith *et al*. (Smith et al., 2008). Type I crustins were found mostly in the Pleocyemata (Table 4), including infraorder Brachyura and Astacidae. Exceptionally, two type I-like crustins (FJ853148, EU500912) were identified in the Dendrobranchiata *Fenneropenaeus chinensis*, which is the only report of type I crustins in Dendrobranchiata (Sun et al., 2010).

Type II crustins were identified in Dendrobranchiata, mostly Penaeids (Table 4). However, type II crustins were also identified in Pleocyemata, including *Pandalopsis japonica, Paralithodes camtschaticus* (EU921643) and *Panulirus japonicus* (FJ797420, FJ797418, FJ797417, FJ797419), suggesting that type II crustins may exist in all decapod crustaceans. We failed to identify type III crustins in *Pandalopsis japonica* and have no evidence of type III crustin genes in Pleocyemata. Currently, type III crustin cDNAs have been identified only in the Dendrobranchiata, including *Penaeus monodon* (Supungul et al., 2004), *Litopenaeus vannamei* (Jiménez-Vega et al., 2004), *Marsupenaeus japonicus* (Rojtinnakorn et al., 2002) and *Fenneropenaeus chinensis* (Jia et al., 2008).

As new sequences accumulate from both Dendrobranchiata and Pleocyemata, more crustins that do not fall into the Smith *et al*. classification may be identified. Crustins from *M. rogenbergii* (FJ429308, EF364560), *E. sinensis* (EU183311, GQ200833) and *P. leniusculus* (EF523612, EF523614) showed the different cysteine spacing patterns from those suggested by Smith *et al*. (2008). Crustin from *P. trituberculatus* (FJ612106) is not a typical type I or II crustin and that from *F. chinensis* (DQ097704), which is an alternative splicing form of DQ097703, is missing cysteine residues within the cysteine-containing region and WAP domain. Although the previous classification (types I, II, and II) by Smith *et al*. (2008) based on differences in primary structure is acceptable, some crustins were difficult to classify due to various exceptional sequences, as described above. Previously,

we proposed that splice sites within the WAP domain may be the key to distinguishing type I from type II crustins (Kim et al., 2012). From the comparative analysis of nine crustin genes from *Pandalopsis japonica* and more than 100 different nucleotide sequences from other decapods, we identified two keys to distinguishing type I from type II crustins. The presence of a splice site in the WAP domain is the first key for type I crustin identification. However, it cannot be used for intron-less crustins, such as Paj-CrusIe, which may have been generated independently after the evolution of the two types of crustins. The second key is the conserved KXXXCP residues, which includes first one of eight conserved cysteine residues in the WAP domain of type II crustins (Fig. 7) (Ranganathan et al., 1999). Although it is well conserved in the most proteins with WAP domain, the biological implications of the conserved motif has not been elucidated (Ranganathan et al., 1999).

Although WAP domains were identified in both types of crustins, their generation mechanism appears to differ. Insertion or deletion of intron is one of major processes in generation of new genes (Jeffares et al., 2006; William Roy and Gilbert, 2006). As shown in Fig. 8, WAP domains were generated by two exons in type I crustin, whereas the whole WAP domain was encoded within a single exon of type II crustins. Multiple alignment data showed that KXXXCP residues are well conserved in the WAP domain of all type II crustins. Differences in splice sites and KXXXCP residues suggest that generation of the WAP domain differed

between the two types of crustins. Since type I crustins were identified only in Pleocyemata and not in Dendrobranchiata, including penaeid shrimps, this suggests that type I crustins may be newly evolved in Pleocyemata and intron-less crustins may have been generated by typical gene duplication events (Zhang, 2003). Type III crustins were identified only in Dendrobranchiata, suggesting that they are unique to the suborder. In this study, we identified major differences in the two types of crustins in terms of gene organization, primary structure and expression profile. However, further biochemical and structural analyses are necessary to better understand types I and II crustins.

Gene organization and expression profile showed a strong relationship in that type I crustins are expressed strongly in the epidermis and gill, whereas type II crustins are expressed ubiquitously (Fig. 9). Unfortunately, we could not identify a crustin gene that is strongly expressed in hemocytes. Additionally, we could not find any differences in the nine crustin genes from *Pandalopsis japonica* and the previously known hemocytic crustins in terms of primary structure and gene organization. This result suggests that hemocyte-specific crustins may not exist outside the epidermal (type I) or ubiquitous (type II) crustins. In addition, models of the two types of crustins predict that their amino-terminal regions may be significantly different, including disulfide bond formation. In particular, the highly flexible and variable glycine-containing region of type II crustins may contribute significantly to overall crustin structure. Structural and expressional differences

suggest functional or biological differences between the two types of crustins. Although whey acidic protein (WAP) was first discovered as the main protein in milk (Hennighausen and Sippel, 1982), proteins with the WAP domain exhibit a variety of functions, including elastase or proteinase inhibition, anti-inflammation, antimicrobial, serine-proteinase inhibition, inhibition of cell growth and proangiogenesis (Ali et al., 2002; Bouchard et al., 2006; Hiemstra, 2002; McAlhany et al., 2003; Moreau et al., 2008; Richardson et al., 2001; Rowley et al., 1995; Schalkwijk et al., 1999; Tamechika et al., 1996). Currently, no functional difference has been identified between the two types of crustin and so further investigation is necessary.

In this study, we failed to detect any antimicrobial activity of recombinant Paj-Crus proteins. Although their crystal structure has not been identified, all crustins contain a cystein-rich WAP domain which harbors the conserved eight cysteine residues forming four disulfide bonds (Smith et al., 2008). Failure of functional expression of Paj-Crus proteins may have come from the bacterial expression system, which lacks the sophisticated system for eukaryotic protein expression (Sahdev et al., 2008). Expression of eukaryotic protein using the bacterial cell systems sometimes inactive mainly due to the misfolding (Sahdev et al., 2008). Different from our results, several crustins exhibited activities using *E. coli* system (Amparyup et al., 2008; Donpudsa et al., 2010; Krusong et al., 2012; Sperstad et al., 2009; Supungul et al., 2008). Major difference between previous studies and our

experiment is the process of refolding from inclusion body. We just concentrated soluble form of crustins and did not undergone refolding or denaturing process, desalting process. Until now, there has been no report about the expression of crustins using eukaryotic expression system. Further study should be made to know difference in antimicrobial activity between bacterial and eukaryotic expression system.

Table 6. Crustin sequences in GenBank database

	Species	GenBank accession number		Species	GenBank accession number
Type I crustins	*Carcinus maenas*	AJ427538	Type II crustins	*Farfantepenaeus brasiliensis*	EF601055
	Eriocheir sinensis	EU183310		*Farfantepenaeus paulensis*	EF182747
		FJ974138		*Farfantepenaeus subtilis*	EF450744
		GQ200832		*Fenneropenaeus chinensis*	AY871268
	Homarus americanus	EF193003			DQ097703
	Homarus gammarus	AJ786653			FJ853147
	Hyas araneus	EU921641		*Fenneropenaeus indicus*	GQ469987
		EU921642		*Litopenaeus schmitti*	EF182748
	Pacifastacus leniusculus	EF5236143		*Litopenaeus setiferus*	AF430077
	Portunus pelagicus	EF120999			AF430078
		JQ965930			AF430079
	Portunus trituberculatus	FJ467931		*Litopenaeus vannamei*	AY488492
		FJ612108			AY488494
		GU373914			AY488495
		JQ728424			AY488496
		JQ728425		*Marsupenaeus japonicus*	AB121740
		JQ728429			AB121741
		JQ728433			AB121742
		JQ728435			AB121743
	Procambarus clarkii	GQ301202			AB121744
		HQ414551		*Penaeus monodon*	EF654658
	Scylla paramamosain	EU161287			EF654659
		EU161288			EU103630
	Scylla serrata	HQ638025			FJ380049
	Scylla tranquebarica	JQ753312			FJ539174
					FJ539175
					FJ539176
					FJ539177
					FJ539178
					FJ686014
					FJ686015
					GQ334395
					GU299808
					HM034319
					JX912161

REFERENCES

Afsal, V., Antony, S.P., Sathyan, N., Philip, R., 2011. Molecular characterization and phylogenetic analysis of two antimicrobial peptides: Anti-lipopolysaccharide factor an d crustin from the brown mud crab, *Scylla serrata*. Results in Immunology 1, 6-10.

Ali, M.F., Lips, K.R., Knoop, F.C., Fritzsch, B., Miller, C., Conlon, J.M., 2002. Ant imicrobial peptides and protease inhibitors in the skin secretions of the crawfish fro g, *Rana areolata*. Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics 1601, 55-63.

Amparyup, P., Kondo, H., Hirono, I., Aoki, T., Tassanakajon, A., 2008. Molecular cl oning, genomic organization and recombinant expression of a crustin-like antimicrobi al peptide from black tiger shrimp, *Penaeus monodon*. Molecular immunology 45, 1 085-1093.

Antony, S.P., Bright Singh, I., Philip, R., 2010. Molecular characterization of a crust in-like, putative antimicrobial peptide, Fi-crustin, from the Indian white shrimp, *Fenn eropenaeus indicus*. Fish and Shellfish Immunology 28, 216.

Bouchard, D., Morisset, D., Bourbonnais, Y., Tremblay, G.M., 2006. Proteins with w hey-acidic-protein motifs and cancer. The lancet oncology 7, 167-174.

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

Christie, A.E., Rus, S., Goiney, C.C., Smith, C.M., Towle, D.W., Dickinson, P.S., 20 07. Identification and characterization of a cDNA encoding a crustin-like, putative a ntibacterial protein from the American lobster, *Homarus americanus*. Molecular imm unology 44, 3333-3337.

Cui, Z., Song, C., Liu, Y., Wang, S., Li, Q., Li, X., 2012. Crustins from eyestalk c DNA library of swimming crab, *Portunus trituberculatus*: Molecular characterization, genomic organization and expression analysis. Fish & Shellfish Immunology.

Dai, Z.-M., Zhu, X.-J., Yang, W.-J., 2009. Full-length normalization subtractive hybri dization: a novel method for generating differentially expressed cDNAs. Molecular bi otechnology 43, 257-263.

Donpudsa, S., Rimphanitchayakit, V., Tassanakajon, A., Söderhäll, I., Söderhäll, K., 2010. Characterization of two crustin antimicrobial peptides from the freshwater cray fish, *Pacifastacus leniusculus*. Journal of invertebrate pathology 104, 234-238.

Hauton, C., Brockton, V., Smith, V., 2006. Cloning of a crustin-like, single whey-aci dic-domain, antibacterial peptide from the haemocytes of the European lobster, *Homa rus gammarus*, and its response to infection with bacteria. Molecular immunology 4 3, 1490-1496.

Hennighausen, L.G., Sippel, A.E., 1982. Mouse whey acidic protein is a novel mem ber of the family of 'four-disulfide core'proteins. Nucleic acids research 10, 2677-26 84.

Hiemstra, P., 2002. Novel roles of protease inhibitors in infection and inflammation. Biochemical Society Transactions 30, 116-119.

Hu, Y., Sun, Z., Eaton, J.T., Bouloux, P.M., Perkins, S.J., 2005. Extended and flexib le domain solution structure of the extracellular matrix protein anosmin-1 by X-ray scattering, analytical ultracentrifugation and constrained modelling. Journal of molecul ar biology 350, 553-570.

Imjongjirak, C., Amparyup, P., Tassanakajon, A., Sittipraneed, S., 2009. Molecular cl

oning and characterization of crustin from mud crab *Scylla paramamosain*. Molecula r biology reports 36, 841-850.

Jeffares, D.C., Mourier, T., Penny, D., 2006. The biology of intron gain and loss. T rends in Genetics 22, 16-22.

Jeon, J.-M., Lee, S.-O., Kim, K.S., Baek, H.-J., Kim, S., Kim, I.-K., Mykles, D.L.,

Kim, H.-W., 2010. Characterization of two vitellogenin cDNAs from a Pandalus shri mp (*Pandalopsis japonica*): Expression in hepatopancreas is down-regulated by endos ulfan exposure. Comparative Biochemistry and Physiology Part B: Biochemistry and

Molecular Biology 157, 102-112.

Jia, Y.-P., Sun, Y.-D., Wang, Z.-H., Wang, Q., Wang, X.-W., Zhao, X.-F., Wang, J.-X., 2008. A single whey acidic protein domain (SWD)-containing peptide from flesh y prawn with antimicrobial and proteinase inhibitory activities. Aquaculture 284, 246 -259.

Jiménez-Vega, F., Yepiz-Plascencia, G., Söderhäll, K., Vargas-Albores, F., 2004. A sin gle WAP domain-containing protein from *Litopenaeus vannamei* hemocytes. Biochemi cal and biophysical research communications 314, 681-687.

Jiravanichpaisal, P., Lee, S.Y., Kim, Y.-A., Andrén, T., Söderhäll, I., 2007. Antibacte rial peptides in hemocytes and hematopoietic tissue from freshwater crayfish, *Pacifas tacus leniusculus*: Characterization and expression pattern. Developmental & Compara tive Immunology 31, 441-455.

Kim, M., Jeon, J.-M., Oh, C.-W., Kim, Y.M., Lee, D.S., Kang, C.-K., Kim, H.-W., 2012. Molecular characterization of three crustin genes in the Morotoge shrimp, *Pan*

dalopsis japonica. Comparative Biochemistry and Physiology Part B: Biochemistry a nd Molecular Biology 163, 161–171.

Krusong, K., Poolpipat, P., Supungul, P., Tassanakajon, A., 2012. A comparative stud y of antimicrobial properties of crustinPm1 and crustinPm7 from the black tiger shri mp Penaeus monodon. Developmental and comparative immunology 36, 208-215.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kauf man, K., Renfrew, P.D., Smith, C.A., Sheffler, W., 2011. ROSETTA3: an object-orie nted software suite for the simulation and design of macromolecules. Methods Enzy mol 487, 545-574.

McAlhany, S.J., Ressler, S.J., Larsen, M., Tuxhorn, J.A., Yang, F., Dang, T.D., Rowl ey, D.R., 2003. Promotion of angiogenesis by ps20 in the differential reactive strom a prostate cancer xenograft model. Cancer research 63, 5859-5865.

Moreau, T., Baranger, K., Dadé, S., Dallet-Choisy, S., Guyot, N., Zani, M.-L., 2008. Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SL PI), two serine protease inhibitors of the chelonianin family. Biochimie 90, 284-295.

Mu, C., Zheng, P., Zhao, J., Wang, L., Qiu, L., Zhang, H., Gai, Y., Song, L., 2011. A novel type III crustin (Crus *Es*2) identified from Chinese mitten crab *Eriocheir sinensis*. Fish & Shellfish Immunology 31, 142-147.

Pisuttharachai, D., Fagutao, F.F., Yasuike, M., Aono, H., Yano, Y., Murakami, K., K ondo, H., Aoki, T., Hirono, I., 2009. Characterization of crustin antimicrobial protein s from Japanese spiny lobster *Panulirus japonicus*. Developmental & Comparative I mmunology 33, 1049-1054.

Ranganathan, S., Simpson, K.J., Shaw, D.C., Nicholas, K.R., 1999. The whey acidic

protein family: a new signature motif and three-dimensional structure by comparativ e modeling. Journal of Molecular Graphics and Modelling 17, 106-113.

Relf, J.M., Chisholm, J.R., Kemp, G.D., Smith, V.J., 1999. Purification and character ization of a cysteine-rich 11.5-kDa antibacterial protein from the granular haemocyte s of the shore crab, *Carcinus maenas*. European Journal of Biochemistry 264, 350-3 57.

Richardson, R.T., Sivashanmugam, P., Hall, S.H., Hamil, K.G., Moore, P.A., Ruben, S.M., French, F.S., O'Rand, M., 2001. Cloning and sequencing of human *Eppin*: A novel family of protease inhibitors expressed in the epididymis and testis. Gene 270, 93-102.

Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y., Aoki, T., 2002. Gene expressio n in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. Fish & shellfish immunology 13, 69-83.

Rosa, R., Barracco, M., 2010. Antimicrobial peptides in crustaceans. Invertebr Surviv J 7, 262-284.

Rosa, R.D., Bandeira, P.T., Barracco, M.A., 2007. Molecular cloning of crustins fro m the hemocytes of Brazilian penaeid shrimps. FEMS microbiology letters 274, 287-290.

Rowley, D.R., Dang, T.D., Larsen, M., Gerdes, M.J., McBride, L., Lu, B., 1995. Pu rification of a novel protein (ps20) from urogenital sinus mesenchymal cells with gr owth inhibitory properties in vitro. Journal of Biological Chemistry 270, 22058-2206 5.

Sahdev, S., Khattar, S.K., Saini, K.S., 2008. Production of active eukaryotic proteins

through bacterial expression systems: a review of the existing biotechnology strategies. Molecular and cellular biochemistry 307, 249-264.

Schalkwijk, J., Wiedow, O., Hirose, S., 1999. The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. Biochemical Journal 340, 569.

Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic acids research 15, 7155-7174.

Smith, V.J., Fernandes, J.M., Kemp, G.D., Hauton, C., 2008. Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. Developmental & Comparative Immunology 32, 758-772.

Sperstad, S.V., Haug, T., Paulsen, V., Rode, T.M., Strandskog, G., Solem, S.T., Styrvold, O.B., Stensvåg, K., 2009. Characterization of crustins from the hemocytes of the spider crab, *Hyas araneus*, and the red king crab, *Paralithodes camtschaticus*. Developmental & Comparative Immunology 33, 583-591.

Sun, C., Du, X.-J., Xu, W.-T., Zhang, H.-W., Zhao, X.-F., Wang, J.-X., 2010. Molecular cloning and characterization of three crustins from the Chinese white shrimp, *Fenneropenaeus chinensis*. Fish & Shellfish Immunology 28, 517-524.

Supungul, P., Klinbunga, S., Pichyangkura, R., Hirono, I., Aoki, T., Tassanakajon, A., 2004. Antimicrobial peptides discovered in the black tiger shrimp *Penaeus monodon* using the EST approach. Diseases of aquatic organisms 61, 123-135.

Supungul, P., Tang, S., Maneeruttanarungroj, C., Rimphanitchayakit, V., Hirono, I., Aoki, T., Tassanakajon, A., 2008. Cloning, expression and antimicrobial activity of cru

stinPm1, a major isoform of crustin, from the black tiger shrimp Penaeus monodon. Developmental and comparative immunology 32, 61-70.

Tamechika, I., Itakura, M., Saruta, Y., Furukawa, M., Kato, A., Tachibana, S., Hiros e, S., 1996. Accelerated evolution in inhibitor domains of porcine elafin family me mbers. Journal of Biological Chemistry 271, 7012-7018.

Torres, A.M., Wong, H.Y., Desai, M., Moochhala, S., Kuchel, P.W., Kini, R.M., 200 3. Identification of a Novel Family of Proteins in Snake Venoms. Purification and st ructural characterization of nawaprin from *naja nigricollis* snake venom. Journal of Biological Chemistry 278, 40097-40104.

Tsunemi, M., Matsuura, Y., Sakakibara, S., Katsube, Y., 1996. Crystal structure of a n elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determi ned at 1.9 Å resolution. Biochemistry 35, 11570-11576.

Vatanavicharn, T., Supungul, P., Puanglarp, N., Yingvilasprasert, W., Tassanakajon, A., 2009. Genomic structure, expression pattern and functional characterization of crusti n *Pm* 5, a unique isoform of crustin from *Penaeus monodon*. Comparative Biochemi stry and Physiology Part B: Biochemistry and Molecular Biology 153, 244-252.

William Roy, S., Gilbert, W., 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7, 211-221.

Yue, F., Pan, L., Miao, J., Zhang, L., Li, J., 2010. Molecular cloning, characterizatio n and mRNA expression of two antibacterial peptides: Crustin and anti-lipopolysacch aride factor in swimming crab *Portunus trituberculatus*. Comparative Biochemistry a nd Physiology Part B: Biochemistry and Molecular Biology 156, 77-85.

Zhang, J., 2003. Evolution by gene duplication: an update. Trends in ecology & evo

lution 18, 292-298.

Zhang, J., Li, F., Wang, Z., Xiang, J., 2007. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. Journal of biot echnology 127, 605-614.