Fish & Shellfish Immunology 35 (2013) 1176–1184

Contents lists available at ScienceDirect

Fish & Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Differences in gene organization between type I and type II crustins in the morotoge shrimp, *Pandalopsis japonica*

Bokwang Kim [a,1], Meesun Kim [a,1], Ah Ran Kim [a], Myunggi Yi [b], Jung-Hwa Choi [c], Hyun Park [d], Wongyu Park [a], Hyun-Woo Kim [a,*]

[a] Department of Marine Biology, Pukyong National University, Busan 608-737, Republic of Korea
[b] Department of Biomedical Engineering, Pukyong National University, Busan 608-737, Republic of Korea
[c] Fisheries Resources Research Division, National Fisheries Research and Development Institute, Busan 619-902, Republic of Korea
[d] Korea Polar Research Institute, Korea Ocean Research and Development Institute, Incheon 406-840, Republic of Korea

ARTICLE INFO

Article history:
Received 1 May 2013
Received in revised form
15 July 2013
Accepted 15 July 2013
Available online 23 July 2013

Keywords:
Antimicrobial peptides
Crustin
Whey acidic protein domain
Decapods
Pandalopsis japonica

ABSTRACT

Crustins are cysteine-rich cationic antimicrobial peptides (AMPs) found in decapod crustaceans. Six novel crustin genes (Paj-CrusIc, Id, Ie, If, IIb and IIc) were identified in the morotoge shrimp, *Pandalopsis japonica*. 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. Differing from previous results, hemocyte-specific crustin was not isolated from *Pandalopsis japonica*. This study showed that both types of crustin genes (types I and II) exist in decapod crustaceans and their primary structure and expression profiles differ from each other, suggesting that they may play different biological roles. This will help to extend our knowledge of the crustacean innate immune response, which will provide important basic information of shrimp immunity against various pathogens.

© 2013 Elsevier Ltd. All rights reserved.

1. 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 [1]. Antimicrobial peptides (AMPs) play an important role in innate immunity in crustaceans, and fifteen AMP family members have been identified [2]. 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 [3–6], but also against some gram-negative bacteria [3].

Since the first discovery of a crustin family member from hemocytes of *Carcinus maenas* [4], homologous sequences have been identified in most decapod crustaceans, including Dendrobranchiata [1,7–12], Caridea [13], Anomura [14], Achelata [15], Astacidea [16–18] and Brachyura [14,19–22]. As various crustin-like sequences accumulated, Smith et al. [5] 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

* Corresponding author. Tel.: +82 51 629 5926; fax: +82 51 629 5930.
E-mail address: kimhw@pknu.ac.kr (H.-W. Kim).
[1] Contributed equally to this work.

1050-4648/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsi.2013.07.031

crustins are not clear. In this study, we identified six genes in addition to the previous three crustin genes from *Pandalopsis japonica* [23]. 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.

2. Materials and methods

2.1. Cloning of six full-length cDNAs encoding crustin from Pandalposis 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 *P. japonica* [23].

Live animals were purchased from a local seafood market in Busan, Korea. Before sacrificing the animals, hemocytes were harvested as described previously [4]. 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 [24]. 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 (Supplementary data). The 3′ end of the full-length mRNA was identified by 3′ RACE, as described previously [24]. 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 *Eriocheir coli* DH5α and incubated overnight at 37 °C. Positive clones were then screened and sequenced in both directions using M13 forward and reverse primers (Supplementary data). 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.

2.2. 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 (Supplementary data). 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 (Supplementary data), 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 s, 60 °C for 30 s, 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.

2.3. 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 (Supplementary data) 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 (200 ng), 1 μl of 10 μM gene-specific forward primer and reverse primer, respectively (Supplementary data), 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 s, 60 °C for 30 s, 72 °C for 30 s 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 (Supplementary data). The PCR conditions were 94 °C for 1 min, followed by 40 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s using DNA Engine Choromo4 Real-time Detector (Bio-Rad

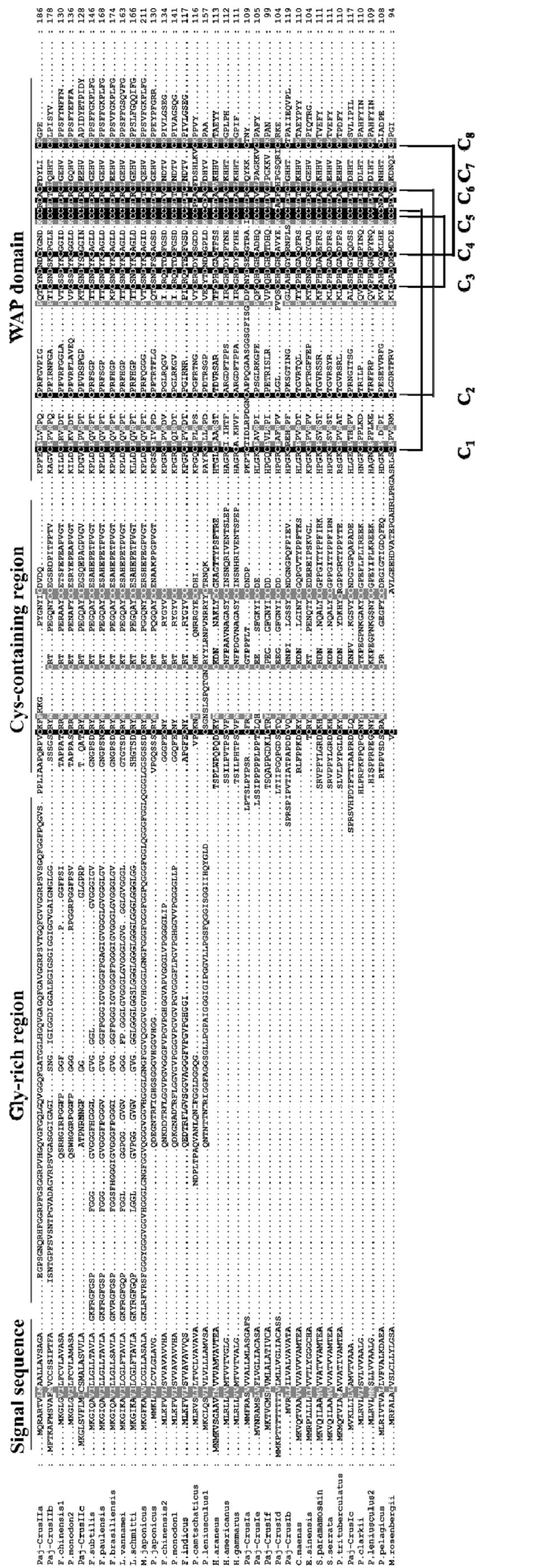

Fig. 1. Amino acid sequence multiple alignment of Paj-CrusIa, Ib, Ic, Id, Ie, If, IIa, IIb and IIc with other crustacean crustin family genes. Black boxes indicate similarities in amino acid sequences. GenBank accession numbers: *Carcinus maenas*, CAD20734; *Eriocheir sinensis*, ACR77767; *Fenneropenaeus 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.

Table 1
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	1102 bp	4	Kim et al. [23]
Paj-CrusIb	JQ004016	119	I	452 bp	3	Kim et al. [23]
Paj-CrusIc	KC608994	117	I	557 bp	4	This study
Paj-CrusId	KC608995	104	I	812 bp	1	This study
Paj-CrusIe	KC608996	105	I	623 bp	1	This study
Paj-CrusIf	KC608997	99	I	538 bp	1	This study
Paj-CrusIIa	JQ004017	186	II	663 bp	3	Kim et al. [23]
Paj-CrusIIb	KC608998	178	II	781 bp	2	This study
Paj-CrusIIc	KC608999	128	II	583 bp	2	This study

Laboratories, Inc., USA). Specific primers (Supplementary data) 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) $\times 10^{10}$.

2.4. 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 [25]. 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 [26–28]. 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.

3. Results

3.1. 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

Table 2
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

identified (Fig. 1 and Supplementary data). 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 (Supplementary data). Each crustin cDNA was named by the primary structure proposed by Smith et al. [5] and the previous nomenclature of crustins in *P. japonica* [23]. Differing from previous studies [3,20], 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 (~50–60 residues) (Fig. 1).

Since three different crustin genes (two type Is and one type II) were previously identified from this species [23], we named the new crustin genes according to the previous rule. However, in order not to raise any confusion from two-lettered nomenclature, 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 *P. 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 *P. japonica* (AFN80342). The cDNA was named Paj-CrusIc (GenBank accession no: KC608994), because it showed the typical domain organization of type I crustins (Fig. 1). 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 resides, respectively (Fig. 1 and Table 1). 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. 1 and Table 1). 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. 1). As in other secretory proteins, all six crustins from *Pandalopsis japonica* contained conserved signal peptide sequences with lengths of 16–26 residues (Fig. 1). 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. 1). To estimate the effects of disulfide bonds on structure, the number of residues between cysteine residues within the regions were analyzed (Table 2). 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 2). 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 2), 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

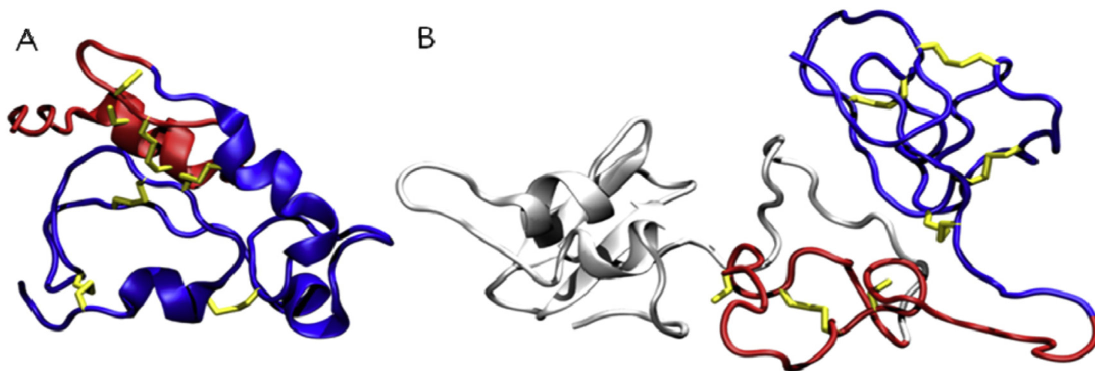

Fig. 2. Structural models of Paj-CrusIa (A) and Paj-CrusIIb (B) are represented by ribbons. The Gly-rich region, Cys-containing region and WAP domain are colored white, red and blue, respectively. Conserved cysteine residues and plausible disulfide bonds are represented by yellow sticks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

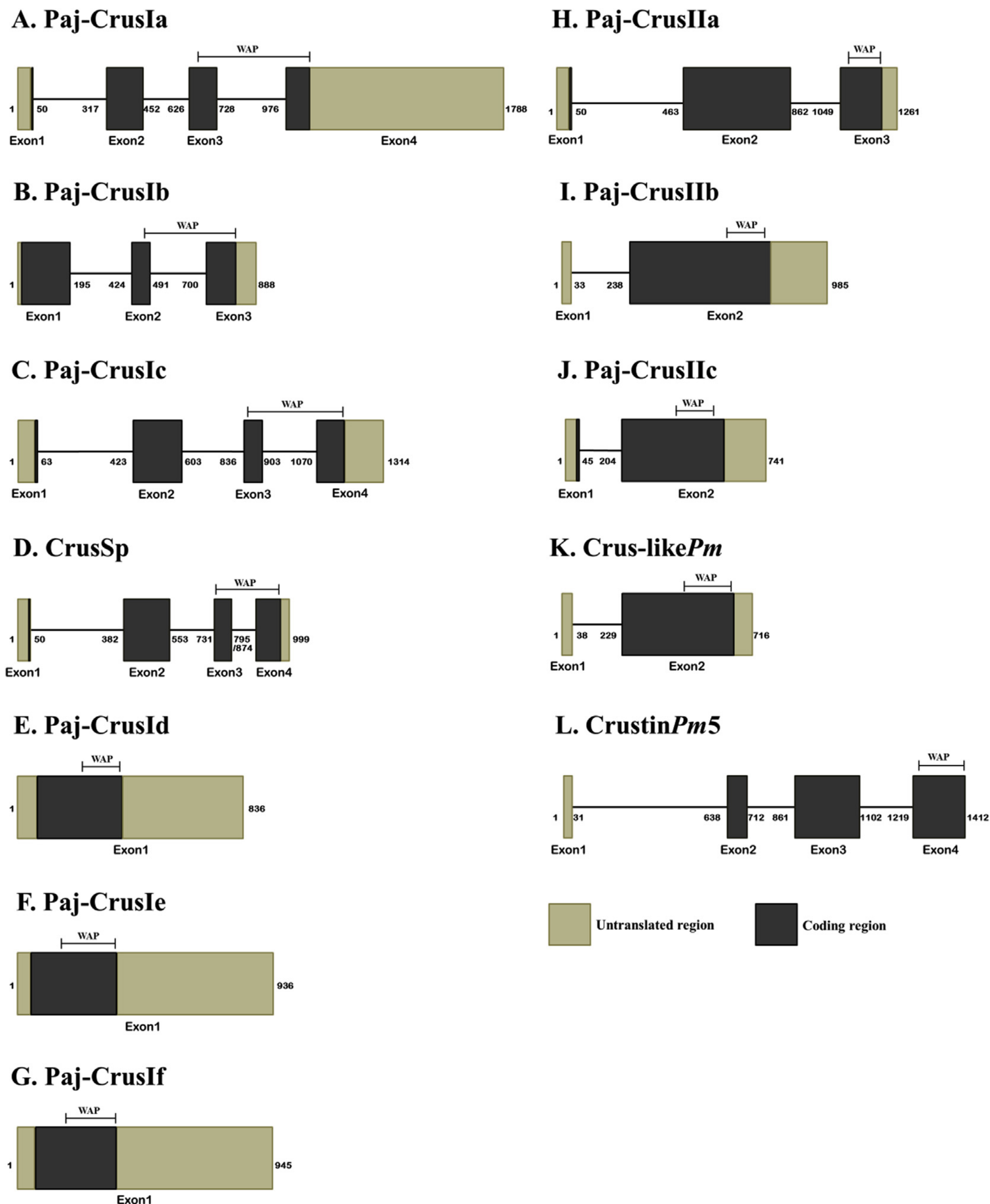

Fig. 3. Schematic representation of the genomic organization of Paj-CrusIc, Id, Ie, If, IIb and Paj-CrusIIc. The boxes represent exons and the dark grey boxes indicate ORFs. UTRs are shown as light grey boxes and exons are connected by introns (lines). Paj-CrusIa (JQ004015), Paj-CrusIb (JQ004016), Paj-CrusIIa (JQ004017), Crustin*Pm*5 (FJ380049), Crus-like*Pm* (EF654659) and Crus*Sp* (EU161288) are crustin genes from *Pandalopsis japonica*, *P. monodon* and *S. paramamosain*.

type II (Paj-CrusIIb) crustins (Fig. 2). 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. 2). 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.

3.2. 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* [29], *Pandalopsis japonica* [23], *Scylla paramamosain* Ref. [20] and *Penaeus monodon* [3,12] (Fig. 3). All analyzed crustin genes followed the typical splicing GT/AG rule [30]. 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. 3C). 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. 3A) and CrusSp (Fig. 3D) from *S. paramamosain* (GenBank Number; EU161288). Although four

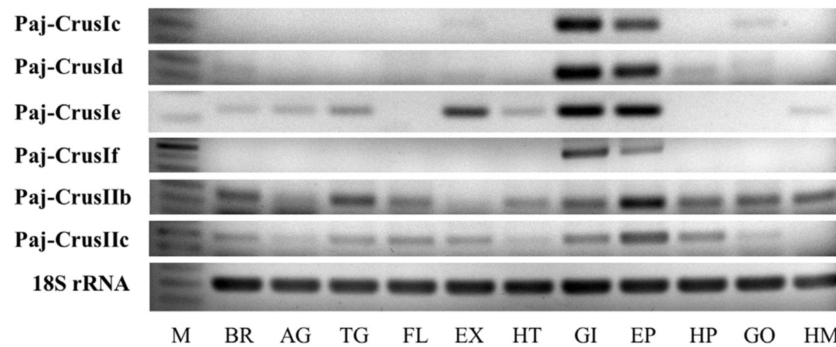

Fig. 4. Expression of Paj-CrusIc, Id, Ie, If, IIb and IIc in various tissues. End-point RT-PCR was carried out for 40 cycles and products were separated on a 1.5% agarose gel. 18S rRNA was used as the control. M, size marker; BR, brain; AG, abdominal ganglia; TG, thoracic ganglia; FL, flexor muscle; EX, extensor muscle; HT, heart; GI, gill; EP, epidermis; HP, hepatopanceas; GO, gonad; HM, hemocyte.

exons and three introns were also identified in Crustin*Pm*5 (GenBank Number; EF654659), the splice sites were different (Fig. 3L). Although one intron was identified in both Paj-Crus IIb and IIc (Fig. 3I 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. 3J). The previously identified Paj-CrusIIa also differed from other type II crustins because of the presence of two introns (Fig. 3H). Although type II crustins had diverse gene organization in which the number and location of splicing varies among individual crustins (Fig. 3), one common characteristic of the type II crustin gene is the absence of a splice site within the WAP domain. Interestingly, three intronless crustins (Paj-CrusId, Ie and If) were identified, but have not been identified in other decapod species (Figs. 3D, 2K and L). 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. [5]. No type II crustin gene without introns has been identified in *Pandalopsis japonica*.

3.3. Expression analysis of Paj-Crus genes

To determine the relationship between gene organization and expression profile, RT-PCR was performed (Fig. 4). As in Paj-CrusIa and Ib [23], Paj-CrusIc was expressed predominantly in the gill and epidermis (Fig. 4). 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 [23]. Previous results showed that a major production site for crustin genes was the hemocyte [3,4,14,31]. We could not identify any crustin gene from *Pandalopsis japonica* that was expressed mainly in the hemocyte (Fig. 4). Similar results were also reported by us previously [23]. 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- and 53.79-fold higher than in hemocytes, respectively (Table 3).

Table 3
Relative copy number of Paj-CrusIc, Paj-CrusId, Paj-CrusIe, Paj-CrusIIb and Paj-CrusIIc.

Peptide	Mean relative copy number	
	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 ($P < 0.001$)

4. 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. [5]. 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 [10].

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* [11], *Litopenaeus vannamei* [32], *Marsupenaeus japonicus* [9] and *Fenneropenaeus chinensis* [8].

Table 4
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

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 *Macrobrachium rogenbergii* (FJ429308, EF364560), *Eriocheir sinensis* (EU183311, GQ200833) and *Pacifastacus leniusculus* (EF523612, EF523614) showed the different cysteine spacing patterns from those suggested by Smith et al. [5]. 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. [5] 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 [23]. 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 intronless 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. 1) [33]. Although it is well conserved in the most proteins with WAP domain, the biological implications of the conserved motif has not been elucidated [33].

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 [34,35]. As shown in Fig. 2, 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 [36]. 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. 3). 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 [37], 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 [38–46]. Currently, no functional difference has been identified between the two types of crustin and so further investigation is necessary.

Acknowledgments

This work was supported by a Functional Genomics on Polar Organisms grant (PE13020) funded by the Korean Polar Research Institute (KOPRI).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fsi.2013.07.031.

References

[1] Rosa RD, Bandeira PT, Barracco MA. Molecular cloning of crustins from the hemocytes of Brazilian penaeid shrimps. FEMS Microbiology Letters 2007;274:287–90.
[2] Rosa R, Barracco M. Antimicrobial peptides in crustaceans. Invertebrate Survival Journal 2010;7:262–84.
[3] Amparyup P, Kondo H, Hirono I, Aoki T, Tassanakajon A. Molecular cloning, genomic organization and recombinant expression of a crustin-like antimicrobial peptide from black tiger shrimp, *Penaeus monodon*. Molecular Immunology 2008;45:1085–93.
[4] Relf JM, Chisholm JR, Kemp GD, Smith VJ. Purification and characterization of a cysteine-rich 11.5-kDa antibacterial protein from the granular haemocytes of the shore crab, *Carcinus maenas*. European Journal of Biochemistry 1999;264:350–7.
[5] Smith VJ, Fernandes JM, Kemp GD, Hauton C. Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. Developmental & Comparative Immunology 2008;32:758–72.
[6] Zhang J, Li F, Wang Z, Xiang J. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. Journal of Biotechnology 2007;127:605–14.
[7] Antony SP, Bright Singh I, Philip R. Molecular characterization of a crustin-like, putative antimicrobial peptide, Fi-crustin, from the Indian white shrimp, *Fenneropenaeus indicus*. Fish and Shellfish Immunology 2010;28:216.
[8] Jia Y-P, Sun Y-D, Wang Z-H, Wang Q, Wang X-W, Zhao X-F, et al. A single whey acidic protein domain (SWD)-containing peptide from fleshy prawn with antimicrobial and proteinase inhibitory activities. Aquaculture 2008;284:246–59.
[9] Rojtinnakorn J, Hirono I, Itami T, Takahashi Y, Aoki T. Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. Fish & Shellfish Immunology 2002;13:69–83.
[10] Sun C, Du X-J, Xu W-T, Zhang H-W, Zhao X-F, Wang J-X. Molecular cloning and characterization of three crustins from the Chinese white shrimp, *Fenneropenaeus chinensis*. Fish & Shellfish Immunology 2010;28:517–24.
[11] Supungul P, Klinbunga S, Pichyangkura R, Hirono I, Aoki T, Tassanakajon A. Antimicrobial peptides discovered in the black tiger shrimp *Penaeus monodon* using the EST approach. Diseases of Aquatic Organisms 2004;61:123–35.
[12] Vatanavicharn T, Supungul P, Puanglarp N, Yingvilasprasert W, Tassanakajon A. Genomic structure, expression pattern and functional characterization of crustin *Pm* 5, a unique isoform of crustin from *Penaeus monodon*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2009;153:244–52.
[13] Dai Z-M, Zhu X-J, Yang W-J. Full-length normalization subtractive hybridization: a novel method for generating differentially expressed cDNAs. Molecular Biotechnology 2009;43:257–63.
[14] Sperstad SV, Haug T, Paulsen V, Rode TM, Strandskog G, Solem ST, et al. Characterization of crustins from the hemocytes of the spider crab, *Hyas araneus*, and the red king crab, *Paralithodes camtschaticus*. Developmental & Comparative Immunology 2009;33:583–91.
[15] Pisuttharachai D, Fagutao FF, Yasuike M, Aono H, Yano Y, Murakami K, et al. Characterization of crustin antimicrobial proteins from Japanese spiny lobster *Panulirus japonicus*. Developmental & Comparative Immunology 2009;33:1049–54.
[16] Christie AE, Rus S, Goiney CC, Smith CM, Towle DW, Dickinson PS. Identification and characterization of a cDNA encoding a crustin-like, putative antibacterial protein from the American lobster, *Homarus americanus*. Molecular Immunology 2007;44:3333–7.
[17] Hauton C, Brockton V, Smith V. Cloning of a crustin-like, single whey-acidic-domain, antibacterial peptide from the haemocytes of the European lobster, *Homarus gammarus*, and its response to infection with bacteria. Molecular Immunology 2006;43:1490–6.
[18] Jiravanichpaisal P, Lee SY, Kim Y-A, Andrén T, Söderhäll I. Antibacterial peptides in hemocytes and hematopoietic tissue from freshwater crayfish, *Pacifastacus leniusculus*: characterization and expression pattern. Developmental & Comparative Immunology 2007;31:441–55.
[19] Afsal V, Antony SP, Sathyan N, Philip R. Molecular characterization and phylogenetic analysis of two antimicrobial peptides: anti-lipopolysaccharide factor and crustin from the brown mud crab, *Scylla serrata*. Results in Immunology 2011;1:6–10.
[20] Imjongjirak C, Amparyup P, Tassanakajon A, Sittipraneed S. Molecular cloning and characterization of crustin from mud crab *Scylla paramamosain*. Molecular Biology Reports 2009;36:841–50.
[21] Mu C, Zheng P, Zhao J, Wang L, Qiu L, Zhang H, et al. A novel type III crustin (Crus *Es*2) identified from Chinese mitten crab *Eriocheir sinensis*. Fish & Shellfish Immunology 2011;31:142–7.
[22] Yue F, Pan L, Miao J, Zhang L, Li J. Molecular cloning, characterization and mRNA expression of two antibacterial peptides: crustin and anti-lipopolysaccharide factor in swimming crab *Portunus trituberculatus*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2010;156:77–85.
[23] Kim M, Jeon J-M, Oh C-W, Kim YM, Lee DS, Kang C-K, et al. Molecular characterization of three crustin genes in the Morotoge shrimp, *Pandalopsis japonica*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2012;163:161–71.
[24] Jeon J-M, Lee S-O, Kim KS, Baek H-J, Kim S, Kim I-K, et al. Characterization of two vitellogenin cDNAs from a Pandalus shrimp (*Pandalopsis japonica*): expression in hepatopancreas is down-regulated by endosulfan exposure. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2010;157:102–12.
[25] Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in Enzymology 2011;487:545–74.
[26] Hu Y, Sun Z, Eaton JT, Bouloux PM, Perkins SJ. Extended and flexible domain solution structure of the extracellular matrix protein anosmin-1 by X-ray scattering, analytical ultracentrifugation and constrained modelling. Journal of Molecular Biology 2005;350:553–70.
[27] Torres AM, Wong HY, Desai M, Moochhala S, Kuchel PW, Kini RM. Identification of a novel family of proteins in snake venoms. Purification and structural characterization of nawaprin from *naja nigricollis* snake venom. Journal of Biological Chemistry 2003;278:40097–104.
[28] Tsunemi M, Matsuura Y, Sakakibara S, Katsube Y. Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determined at 1.9 Å resolution. Biochemistry 1996;35:11570–6.
[29] Cui Z, Song C, Liu Y, Wang S, Li Q, Li X. Crustins from eyestalk cDNA library of swimming crab, *Portunus trituberculatus*: molecular characterization, genomic organization and expression analysis. Fish & Shellfish Immunology 2012.
[30] Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Research 1987;15:7155–74.
[31] Donpudsa S, Rimphanitchayakit V, Tassanakajon A, Söderhäll I, Söderhäll K. Characterization of two crustin antimicrobial peptides from the freshwater crayfish, *Pacifastacus leniusculus*. Journal of Invertebrate Pathology 2010;104:234–8.
[32] Jiménez-Vega F, Yepiz-Plascencia G, Söderhäll K, Vargas-Albores F. A single WAP domain-containing protein from *Litopenaeus vannamei* hemocytes. Biochemical and Biophysical Research Communications 2004;314:681–7.
[33] Ranganathan S, Simpson KJ, Shaw DC, Nicholas KR. The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. Journal of Molecular Graphics and Modelling 1999;17:106–13.
[34] Jeffares DC, Mourier T, Penny D. The biology of intron gain and loss. Trends in Genetics 2006;22:16–22.
[35] William Roy S, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. Nature Reviews Genetics 2006;7:211–21.
[36] Zhang J. Evolution by gene duplication: an update. Trends in Ecology & Evolution 2003;18:292–8.
[37] Hennighausen LG, Sippel AE. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. Nucleic Acids Research 1982;10:2677–84.
[38] Ali MF, Lips KR, Knoop FC, Fritzsch B, Miller C, Conlon JM. Antimicrobial peptides and protease inhibitors in the skin secretions of the crawfish frog, *Rana areolata*. Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics 2002;1601:55–63.
[39] Bouchard D, Morisset D, Bourbonnais Y, Tremblay GM. Proteins with whey-acidic-protein motifs and cancer. The Lancet Oncology 2006;7:167–74.
[40] Hiemstra P. Novel roles of protease inhibitors in infection and inflammation. Biochemical Society Transactions 2002;30:116–9.
[41] McAlhany SJ, Ressler SJ, Larsen M, Tuxhorn JA, Yang F, Dang TD, et al. Promotion of angiogenesis by ps20 in the differential reactive stroma prostate cancer xenograft model. Cancer Research 2003;63:5859–65.

[42] Moreau T, Baranger K, Dadé S, Dallet-Choisy S, Guyot N, Zani M-L. Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family. Biochimie 2008;90:284–95.
[43] Richardson RT, Sivashanmugam P, Hall SH, Hamil KG, Moore PA, Ruben SM, et al. Cloning and sequencing of human *Eppin*: a novel family of protease inhibitors expressed in the epididymis and testis. Gene 2001;270:93–102.
[44] Rowley DR, Dang TD, Larsen M, Gerdes MJ, McBride L, Lu B. Purification of a novel protein (ps20) from urogenital sinus mesenchymal cells with growth inhibitory properties in vitro. Journal of Biological Chemistry 1995;270:22058–65.
[45] Schalkwijk J, Wiedow O, Hirose S. The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. Biochemical Journal 1999;340:569.
[46] Tamechika I, Itakura M, Saruta Y, Furukawa M, Kato A, Tachibana S, et al. Accelerated evolution in inhibitor domains of porcine elafin family members. Journal of Biological Chemistry 1996;271:7012–8.