

Elsevier Editorial System(tm) for Developmental & Comparative Immunology
Manuscript Draft

Manuscript Number: DCI-D-07-00179R2

Title: Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans

Article Type: Review Article

Keywords: antibacterial peptides, crustins, whey acidic protein domain, four-disulphide core, innate defence effectors, invertebrate immunity, defensins

Corresponding Author: Dr. Valerie Smith,

Corresponding Author's Institution: Univ. of St. Andrews

First Author: Valerie J Smith, B.Sc., Ph.D.

Order of Authors: Valerie J Smith, B.Sc., Ph.D.; Jorge M Fernandes, B.Sc., PhD; Graham D Kemp, B.Sc., Ph.D; Chris Hauton, B.Sc. PhD

Abstract: Crustins are antibacterial proteins of ca 7-14 kDa with a characteristic four disulphide core-containing whey acidic protein (WAP) domain, expressed by the circulating haemocytes of crustaceans. Over 50 crustin sequences have been now reported from a variety of decapods, including crabs, lobsters, shrimp and crayfish. Three main types seem to occur but all possess a signal sequence at the amino terminus and a WAP domain at the carboxyl end. Differences between types lie in the structure of the central region. Those crustins purified as the native protein or expressed recombinantly all kill Gram-positive bacteria, and gene studies have shown that they are constitutively expressed, often at high levels, but show no consistent patterns of change in expression following injection of bacteria. This variable response to infection is enigmatic but indicates that these proteins could perform additional functions, perhaps as immune regulators in recovery from wounding, trauma or physiological stress.

Prof Kenneth Soderhall
Co-Editor in Chief
Developmental and Comparative Immunology

Dear Kenneth

I attach the revised manuscript for the invited review, *Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans*, for which all the referees comments have been addressed. Below is a summary of changes made.

1. Two of the reviewers have issues with the SS tree (formerly Fig 2), with reviewer 2 making the point that the functional WAP domain will more usefully indicate relationships rather than the SS. We do not concur with the comment that the signal sequence (SS) tree is unnecessary but do accept that its placement in the article could be better and that it could be more fully discussed. Importantly, the SS radiation tree did not drive the 3-type classification, rather it was led by the overall domain organisation of the various types that we judged, from our starting definition, to be within the crustin 'family'. The region of variation is mainly in the 'central' region of the molecule, ie between the SS and WAP domain. We constructed the SS and WAP domain trees to test if the suggested classification was supported, independently by phylogenies within these regions. We believe that they do and the inclusion of both trees, rather than just a tree for the WAP domain, makes the case for the classification of crustins into Types I, II or III, even stronger. We have therefore retained the SS radiation tree but discuss it more fully in the section on 'Relationships' (new pages 11 & 12) rather than its former position. Fig 2 therefore is now designated Fig 6. The WAP radiation tree is now Fig 7.

We prefer not to add species identifiers to the as these will make it very 'busy' and untidy. We feel that Accession numbers are more appropriate than species names. Please note we have slightly modified the annotations by leaving ambiguous sequences out of the clusters marked by dotted or dashed lines.

2. Referee 1 correctly points out that the argument (beginning in the last paragraph on pg 6 and continued on pg 7 in the original m/s) that the signal sequence is probably not involved in crustin transport is rather weak. We have removed this part of the review, leaving only the comment that the role of the signal sequences is as yet unclear.

3. Likewise, we have removed the discussion about the possible role of the signal sequence in increasing crustin stability presented in the second paragraph of pg 7. Comparison of predicted stability indices for all crustin cDNA sequences and signal regions did not offer sufficient support for this hypothesis at present.
4. We accept Referee 1's comment that the use of the word 'convincing' is inappropriate in the sentence that about the WAP domain in the *C. finmarchicus* EST sequence. We have changed the word 'convincing' to 'putative' (see new page 11).
5. We have re-written the section on Patterns of Expression (new pages 12 & 17) where we discuss the synthesis (apparent lack of) of crustins in the hepatopancreas. We trust that this is now clearer.
6. We have altered the remark in the Concluding Remarks about 'more expression studies'. This has been changed to more 'directed functional and proteomic studies' (new page 20). We did not want to lengthen the article further by detailing precisely which future studies are needed as many such hints are dropped throughout the article. In short we need to know much, much more about the natural *proteins* and what functions, other than antibacterial activity, that they perform.
7. We have altered the wording of the paragraph, queried by Ref 2, about the antibacterial activity of crustins being restricted to Gram-positive bacteria. Whilst so far all of the few studies of the antibacterial properties of crustin proteins have demonstrated activity primarily against Gram positive bacteria, we accept that there are too few studies so far to be proscriptive about this. Page 14 (top) of the revised version now includes the comment '*it is also conceivable that broader specificity to deal with a wider range of pathogens might be achieved through either synergism with other AMPS in the host or through sequence diversification especially in the isoforms*'. We trust this is acceptable.
8. We do not agree with Ref 3 in his / her assertion that the Type III crustins are unrelated to the other types. We wonder if this Ref has overlooked the definition of a crustin too (m/s page 4) where we explicitly state that having a WAP domain is not enough to be considered a crustin. This definition is quite clear and doesn't need changing, in our view. Please note that the other two Referees do not seem have a problem with it. Our definition of crustins clearly excludes consideration of other structures, such as the mouse WAP domain protein and the SLPIs. However, in case other readers might be confused about this we have included an additional figure (new Fig 3) in which a phylogenetic tree has been constructed to show the affinity between the three types of crustin with outgroups. Our case, we believe, is supported by the un-rooted radiation trees, shown in Figures 6 and 7. We hope this strengthens our case to include the SWDs and chelonianins etc as true (Type III) crustins.

9. We have updated the text, Table 1 and Figs 6 and 7 with new sequence data and new publications published after the initial submission. We have further noted an error on the Genbank sequence of one crustin described in our paper, consulted with the author of this sequence paper and redrawn our phylogenetic and radiation trees using the corrected sequence.

I hope this addresses all the referees' comments and that the m/s is now suitable for publication in DCI.

Very kind regards

Val

We hope this improves the m/s and that it is now in a form acceptable for publication in DCI.

Kind regards

Val

Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans

Valerie J. Smith¹, Jorge M. O. Fernandes²,

Graham D. Kemp³ & Chris Hauton⁴

¹ Comparative Immunology Group, Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife, Scotland KY16 8LB, UK

²Department of Fisheries and Natural Sciences, Bodø University College, No-8049 Bodø, Norway

³Biomolecular Sciences Centre, University of St Andrews, St Andrews, Fife, Scotland KY16 9ST

⁴School of Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, European Way, Southampton, Hants, SO14 3ZH, UK

Author for correspondence: V.J. Smith. E-mail vjs1@st-andrews.ac.uk

Keywords: antibacterial peptides, crustins, whey acidic protein domain, four-disulphide core, innate defence effectors, invertebrate immunity, defensins

Abstract

Crustins are antibacterial proteins of ca 7-14 kDa with a characteristic four disulphide core-containing whey acidic protein (WAP) domain, expressed by the circulating haemocytes of crustaceans. Over 50 crustin sequences have been now reported from a variety of decapods, including crabs, lobsters, shrimp and crayfish. Three main types seem to occur but all possess a signal sequence at the amino terminus and a WAP domain at the carboxyl end. Differences between types lie in the structure of the central region. Those crustins purified as the native protein or expressed recombinantly all kill Gram-positive bacteria, and gene studies have shown that they are constitutively expressed, often at high levels, but show no consistent patterns of change in expression following injection of bacteria. This variable response to infection is enigmatic but indicates that these proteins could perform additional functions, perhaps as immune regulators in recovery from wounding, trauma or physiological stress.

Introduction

The Crustacea is the largest, most conspicuous and, arguably, the most important group of marine or aquatic arthropods in terms of their biomass and ecological or economic value. Crustaceans have been popular experimental animals in nearly all aspects of biology, but it is decapods that attract most attention in relation to their immune responses because of their huge commercial importance and the need to control disease outbreaks in shellfish aquaculture. Given the ecological and economic significance of crustaceans and the amount

of knowledge that has been gained about their defence responses, it is very surprising that it has taken so long for a body of work to build up on the presence of the low molecular weight defence effectors, more popularly known as antimicrobial peptides (AMPs), in these animals. To date, over 800 AMPs from eukaryotic organisms have now been reported in the literature or lodged on databases (see, for example:

<http://www.bbcm.univ.trieste.it/tossi/pag1.html>, <http://public-1.cryst.bbk.ac.uk/peptaibol/home.shtml>, <http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC/>) yet the majority are

still derived from mammals, insects or amphibians. So far less than 10% of all known animal AMPs are from crustaceans, with nearly all of these found in decapods.

Whilst numerous papers have described antimicrobial activities of crustacean blood, tissues or body fluids [see for examples 1-3], it was not until 1995 that the first crustacean AMP was isolated and characterised. This was a 6.5 kDa proline-rich cationic peptide, purified from the haemocytes of the shore crab, *Carcinus maenas*, with activity against both Gram-positive and Gram-negative bacteria [4]. Soon after reports began to emerge of antibacterial proteins from other species, notably those from shrimp [5]. The number is now rising steadily with over 70 from a variety of species presently listed on databases or described in the literature. They comprise a diverse collection of proteins with two main groups; the penaeidins and crustins. The penaeidins are cationic antimicrobial peptides of 5-7 kDa, characterised by a proline-rich amino-terminal domain and a cysteine-rich carboxyl terminus domain. They seem to be confined to the Dendrobranchiata and have been reviewed elsewhere [6-8]. The crustins, on the other hand, are ca 7-14 kDa, cysteine-rich and occur more widely across the Decapoda, having been found in both Pleocyemata and Dendrobranchiata.

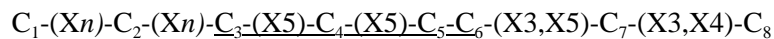
The first crustin to be found was an 11.5 kDa cationic and hydrophobic protein isolated from the haemocytes of the shore crab, *Carcinus maenas* [9]. This protein was later designated carcinin [10] and judged to be a crustin-type molecule following the determination of its full coding cDNA and its recombinant expression *in vitro* [11]. Carcinin remains the only crustin to have been isolated in its native form from crustacean haemolymph, cloned and expressed recombinantly *in vitro*. The name ‘crustin’ was originally invoked to describe a cys-rich gene, with high sequence homology to carcinin, in the shrimp, *Litopenaeus vannamei* [12]. However many more of these cys-rich carcinin-like genes have been subsequently detected in crustaceans and it is becoming clear that they are expressed by many, making the term more generic. To date some 50, including isoforms and ESTs, have been found in 20 crustacean species including crabs, lobsters, shrimp, prawns and crayfish (Table 1). As crustins are relative newcomers to the abecedarary of invertebrate immune molecules, there is some confusion as to what constitutes a member of the crustin group, what biological effects they have and how they respond to non-self challenge. Accordingly, the present review is aimed at bringing together the available data on this group of proteins, summarising their characteristics and diversity and discussing their biological roles.

Definition of a crustin

There is no existing universally accepted definition of a crustin but the present review considers them to be cationic cysteine rich-antibacterial polypeptides of ca 7-14 kDa, with an isoelectric point usually in the range of 7.0-8.7, present in crustaceans that contain one whey acidic protein (WAP) domain at the carboxyl terminus. This domain has eight cysteine residues in a conserved arrangement that forms a tightly packed

structure described on PROSITE as a four-disulphide core (4DSC). The terms ‘WAP-domain’ and ‘4DSC’ are now coming to be used synonymously and this review does not make a distinction between them.

The 4DSC motif is not unique to crustins. The term ‘whey acidic protein’ is derived from the name given to a family of proteins, originally discovered in the whey fraction of mammalian milk. Whilst all these milk proteins are characterised by possession of two WAP domains, each comprising 50 aa [13], numerous other non-milk whey acidic proteins are now also known and these may have one or more 4DSCs. Amongst these non-milk whey acidic proteins are small secretory proteins with protease inhibitory properties or regulatory functions in growth, tissue differentiation or regulation and may sometimes be expressed in certain cancer states [14, 15]. A few are antibacterial [16,17]. Particularly well-known WAP-domain containing proteins in mammals are antileukoproteinases, elafins, and trappins. Analysis of numerous whey acidic proteins from vertebrates reveals a high degree of similarity between the WAP domain structures and Ranganathan et al. [13] have proposed that the PROSITE definition of the domain structure be modified to the following identifying motif which the positions of the eight conserved cysteines (C₁₋₈) are as follows:



X indicates any residue and Xn is a stretch of n residues. The signature motif of the central four cysteines that form the basis of the 4DSC is underlined.

The definition of a crustin offered here complies with this format and thereby excludes the many cys-rich crustacean antimicrobial proteins that lack a WAP-domain, for

example defensins or penaeidins. Importantly, it embraces some crustacean molecules that have not so far been specifically designated as crustins or for which an antibacterial function has yet to be established. Moreover, it sets crustins within their own category of WAP-containing proteins from those found in other taxa [14, 18-22] and distinguishes them from those that have a WAP domain at the N terminus or more than one at the C terminus, even though there may be some functional equivalence between these proteins in terms of their role in inflammation and /or antimicrobial properties [16, 23,24, 25].

Crustin structure

All crustins described to date possess a leader / signal sequence at the N terminus and the WAP domain at the carboxyl end (Fig 1). The putative signal sequence at the N terminus comprises ca 16-24 aa which in many species shows strong representation of valine residues (Table 1). Across the Decapoda as a whole, though, the signal region is not markedly conserved, at least compared to the situation with other AMPs, such as cathelicidins and defensins, where the signal regions at the N terminus are often highly similar between species, making the C terminus, which constitutes the bioactive ‘mature’ protein, the region of greatest diversification [26, 27].

The cleavage site, which marks the end of the signal sequence, at least as predicted by software programmes, is usually between alanine and glycine, although in some crustins it lies between glycine and glutamine, alanine or threonine. Analysis of genomic and recombinantly expressed carcinin has revealed that the signal and mature sections are probably not encoded by separate exons [11]. It is unclear if the signal sequence is directing trans-membrane transportation of the protein, as in insects and mammals, or if

the mature protein is released from the haemocytes through regulated exocytosis, as seems to be the case with the penaeidins [28,29]. Many immune proteins are released from crustacean haemocytes by exocytosis upon non-self challenge [30, 31], so it is plausible that crustins might be liberated into the body fluids in the same way.

The WAP domain, in contrast to the signal sequence, is highly conserved between species and in several crustins, especially those from shrimp, aspartic acid and lysine residues are positioned as follows:

-C-XX-D-XX-C-XXXD-K-CC-X-D-

This arrangement, however, does not hold true in every case. For example, glycine substitutes for aspartic acid after the first cysteine in a *Litopenaeus vannamei* EST (AAS57715), and serine replaces lysine before the double cysteines in the *Callinectes sapidus* EST (CV462984). The crayfish, *Pacifastacus leniusculus*, is unusual in expressing three quite diverse crustins, one of which *Plcrustin3* has an extra cysteine immediately after C₆ in the 4DSC, while another (*Plcrustin2*) lacks the one at position seven [32]. This means that according to the formula proposed by Ranganathan *et al* [13] (above), and used in the present review for convenience, *Plcrustin2* cannot be regarded as a ‘true’ crustin’ because it does not have a ‘proper’ 4DSC. On the basis of the less rigorous PROSITE definition, however, *Plcrustin2* can be accepted as a WAP domain-containing protein, thereby qualifying it as a crustin by slightly different criteria. Faced with this dilemma, we have chosen to include *Plcrustin2* in this review of crustins, in order to shed more light on the diversity of these molecules and stimulate interest in researching them further. Certainly the absence of C₇ in *Plcrustin2* could make the molecule somewhat unstable because the disulphide bridge between C₂ and C₇ cannot

form, which may affect its biological activities. It would be very interesting to determine if *P*lcrustin2 has the same functional properties as more ‘conventional’ crustins. Other differences in the WAP domain of crustins could well come to light as more crustin-like sequences are determined for a wider range of crustacean species. Despite small differences in the sequences, the domain, itself, seems to form a tightly coiled structure enclosing two β sheets and a helical segment. Structural models of the WAP domain from three different crustins [33, 34] suggest that the tertiary structure of this part of the molecule is well conserved between decapod species (Fig 1).

Crustin Types

The region between the signal sequence and WAP domain is variable but conforms to one of a small number of distinct structural patterns with regard to the presence or absence of other domains. The arrangements of these are largely, but not entirely, conserved within taxonomic sub-groups of arrangements within the Decapoda. At least three main sub-groups appear to exist and we propose that they should be designated Types I-III (Fig 2) at least until future research, especially from non-decapod taxa, renders this classification redundant.

In Type I crustins, the region that lies between the signal sequence and the WAP domain is of variable length and cysteine-rich but has rarely more than six of these residues.

These cannot form a full 4DSC configuration and the region might be loosely thought of as resembling an incomplete 4DSC (Fig 1). A suitable consensus framework for this region might be:

-C-X(3)-C-X(8-12)-C-C-X(16-17)-C-X(6)-C-X(9-10)-

These types of crustins are present mainly in crabs, lobsters and crayfish (the Pleocyemata) [11, 32, 35-37] (Table 1). We designate them as Type I as they all show a similar domain arrangement as carcinin, the first crustacean WAP domain containing AMP to be found [9].

Type II crustins, on the other hand, possess not only a cys-rich region but also a long gly-rich domain of approximately 40-80 amino acids adjacent to the signal region (Fig 1). The number of glycines varies between species but is usually between 20 and 50, and in shrimp is often arranged as repeat VGGGLG motifs that vary in number from 5 to 8 [12, 38, 40-44]. Type II crustins were the second type to be described [12] and occur mainly in shrimp (Dendrobrachiata). Recently, however, one has been reported in *Penaeus monodon* ([EF523614](#)) and while it contains some 22 glycine residues it does not show the same repeat VGGGLG motif as some other shrimp [43]. The glycines, being small amino acids, might render the gly-rich region flexible and/or allow tight bends in the structure, but, as yet, it is unclear why this gly-rich region occurs so frequently in shrimp crustins and so rarely in the Pleocyemata or what functional properties it confers on the mature proteins.

A third group of WAP domain-containing proteins from decapods resemble crustins but lack not only the gly-rich domain of the Type II molecules but also the cys-rich region present in both Type I and Type II (Fig 1). Only a few of these are known and so far reports of them are confined to the shrimps, *Penaeus monodon* [40], *L. vannamei* [45], *M. japonicus* [46] and *F. chinensis* ([EF216349](#)) (Table 1). In the literature these types of

proteins have been termed, not crustins, but single whey domain (SWD) proteins [45], chelonianin-like proteins [39] or antileukoproteinase-like proteins [46]. However, we believe but these terms are confusing and potentially misleading as the presence of the single WAP domain at the C terminal end unites them with the crustins. Indeed, one EST from *Marsupenaeus japonicus* (AU175528) that is described by the authors as chelonianin-like [46], in fact has no WAP domain by the criteria of either Ranganathan et al. [13] or PROSITE, and therefore does not qualify as a crustin. Figure 3 shows a phylogenetic tree based on the similarity of the full coding sequence of all known crustins and the crustacean SWDs. Importantly, these cluster into three distinct but related clades. The Type I crustins form one, the Type IIs a second and the SWDs form a third, more distantly related, one (Fig 3). This SWD cluster further shows affinity to some non-crustacean WAP-containing proteins (Fig 3), but as their structure fits the definition of a crustin (above) we propose that these small proteins be classed as Type III crustins, at least until further data, crucially if they have antibacterial properties, is presented to the contrary.

Isoforms

Molecular studies have revealed that some decapod species express more than one type of crustin that differ from one another only by a few amino acids, usually 1-4. The potential number of variants of the encoded protein can thus, at least theoretically, be quite large. For example, evidence from cDNA libraries has revealed that multiple isoforms exist in *P. monodon* [40, 41] *Litopenaeus setiferus* [12] and *M. japonicus* [47] (Fig 4). Shaffer et al (2004, unpublished) have also lodged on databases some five ESTs from *C. sapidus* that resemble *M. japonicus* crustin sequences.

An interesting situation exists in *L. vannamei* for which Bartlett et al. [12] have reported the occurrence of six possible crustin variants while Vargas-Albores et al. [38] have discovered two (Fig 4). Possibly many other isoforms exist which have yet to be recorded. Depending on the amino acids involved and their location within the protein, isoforms may vary slightly in biological activity and undergo positive selection if any confer survival advantages to host under local conditions. Interestingly, these two studies used shrimps from geographically separate regions, namely South Carolina, USA and Mexico respectively. Farmed shrimp come from particular, usually locally supplied, broodstocks and may have more restricted gene pools than wild shrimp. Such inbred stocks could conceivably show marked differences in the repertoire of isoforms expressed between each other and wild types. Working with crabs from a single wild population on the North East coast of Scotland, Brockton *et al.* [11] have found that carcinin exhibits multiple isoforms in mRNA transcripts (Fig 4). In addition, analyses of genomic DNA has further revealed that sequences can differ even between as few as two crabs by non-synonymous nucleotide substitutions [11]. As they have probably not arisen by somatic mutation within individuals, isoform frequency could well be population related.

WAP-domain containing molecules in non-decapod crustaceans

Reports of crustins from crustaceans outside the Decapoda are extremely rare, but at least three crustacean species have been found to express ESTs that possess a putative WAP-like domain. They include two from the freshwater amphipod, *Gammarus pulex* (EH68909 and EH273178) and one from the copepod, *Calanus finmarchicus*

([EL585729](#)). The two *G. pulex* ESTs are identical to each other apart from an extra GGFGSG- sequence at the N terminus of [EH68909](#). Despite being incomplete these sequences show many characteristics of Type II crustins in possessing a distinct gly-rich region of 41 aa that matches well to those of shrimp (Fig 5a). A cys-rich region leads into the WAP domain, which itself shows the same substitution (proline for methionine) after the second cysteine as other crustins. Like the *C. sapidus* crustin EST, the *G. pulex* ESTs have an arginine in place of a lysine before the double cysteine but have a histidine in place of aspartic acid after the first (Fig 5a). The *G. pulex* ESTs are thus very strong candidates for membership of the crustin group and seem to represent Type II crustins, although their status and categorisation await determination of their full cDNAs and functional properties. The *C. finmarchicus* EST sequence has a putative WAP domain (Fig 5b) and an extensive cys-rich region next to this on the N terminal side, but in other respects does not conform well to the expected size and domain arrangement of other crustins. An EST from the brine shrimp, *Artemia franciscana* ([Q1W1H4](#)) is cys-rich and displays at least one 4DSC (Fig 5b) but otherwise has relatively poor similarity to other crustin sequences and thus cannot at present be considered to be a crustin without further molecular and functional characterisation.

Relationships

The three types of crustin, described above, are distinguished on the basis of the domain structure between the signal sequence and the WAP domain. While it is a useful way to differentiate between types, alone it tells us little about the relationships between them. From Table 1 it seems that there is a tendency for types to track to taxon, for example Type IIs are generally present in shrimp, but this does not hold true across the board. The

presence of both Type I and Type II crustins in the crayfish, *P. leniusculus*, for example shows that crustin type is not entirely taxonomically entrained. To examine the relationship of the various crustins to one another we undertook un-rooted radiation tree analyses of the signal sequences and the WAP domains of the known decapod crustins for which appropriate sequence information is available. After aligning the sequences with ClustalW, Bayesian phylogenetic reconstructions were performed using average mixed models of amino acid substitution and run for 500,000 generations.

Fig 6 shows the tree generated for the signal sequences. Despite the relatively small number of sequences available, the sequences do tend to cluster into groups. The SWD and chelonianins form a strong affiliation as one cluster while the shrimp crustins form another, albeit less well supported, group of Type II crustins and the Type Is comprise a third, diverse, cluster (Fig 6). Interestingly, the Type IIs actually contain two sub-clusters, which may reflect sequence diversity between species or stocks from geographically separate areas. Similar sub-clustering occurs with the signal sequences of the Type I crustins (Fig 6). One sub-cluster contains the lobster (AJ786653; EF193003) and *P. pelagica* (EF120999) crustins, PET-15 (AY340636) and *P. leniusculus* crustin 2 (EF523612) that affiliates with the signal sequence of the *C. pugilator* crustin (DW176879) (Fig 6). The second sub-cluster of Type Is contains carcinin and *P. leniusculus* crustin1 (EF523612) (Fig 6). Curiously the signal sequence of the *P. monodon* crustin (EF654658) branches ambiguously between the carcinin cluster and one of the Type II sub-clusters (Fig 6). Thus the signal sequences track reasonably well to crustin type independently of the adjacent domain structure and irrespective of the taxonomic status

of the host with sub-clustering possibly reflecting geographically separate populations of crustaceans under study.

The un-rooted radiation tree constructed for the WAP domains also reveal at least three quite distinct clusters that contain Type I, II and III crustins respectively (Fig 7), but sub-clustering within each is less marked than with the signal sequences. Unusually the carcinin isoforms appear to cluster separately (Fig 7) to the other Type Is but, unlike many of the shrimps which often come from farmed populations from different brood-stocks around the world, studies on *C. maenas* were made on a single population of wild-caught crabs on the East coast of Scotland. Interestingly, the WAP domains of the two *G. pulex* ESTs cluster within the shrimp WAPs, thus further supporting their candidacy as Type II crustins (Fig 7).

Overall, it would appear that evaluations of the domain organisations support the existence of three crustin types, at least in the decapods studied to date. These three types are not distributed purely according to the taxonomic status of the host and one species may express more than one type of crustin gene. It remains to be tested if the mature proteins of the three types have different functions, sites of expression or antibacterial specificities.

In terms of the relationship of crustins to other animal AMP families, crustins do share several similarities to the defensins, especially the β -defensins of insects and chelicerates, by virtue of their overall gene organisation, abundance of cysteine residues, β -sheets, and the possession of a signal sequence of 16 to 24 aa [48-51]. In contrast to defensins,

however, crustins lack strong conservation of the amino terminus and do not possess an anionic propiece. Instead their unique cysteine framework points to them comprising a group distinct from the defensins, albeit possibly sharing a common ancestral gene with them.

Antibacterial properties

Crustins are widely regarded as antimicrobial effectors, yet there have been surprisingly few studies of their antibacterial properties *in vitro*. Carcinin, which was purified from the haemocytes of *C. maenas* on the basis of its ability to inhibit the growth of bacteria, appears to be active against Gram-positive but not Gram-negative bacteria [9].

Susceptible strains include the lobster pathogen, *Aerococcus viridans* var *homari*, two strains of marine *Planococcus* spp. and a salt tolerant strain of *Micrococcus luteus* [9].

The crustins purified from the spider crab, *Hyas araneus* also seem to kill Gram-positive bacteria, most strongly showing an MIC value of ca 3 μ M against *Corynebacterium glutamicum* [52]. A recombinant crustin from *F. chinensis* [42, 53] similarly targets Gram-positive bacteria with MIC values ranging from 2 to 8 μ M for strains including *Staphylococcus aureus*, *M. luteus*, and three species of *Bacillus*, but no detectable activity against Gram-negative bacteria or fungi [42, 53]. One of the *P. monodon* crustins (CD766060), expressed as a recombinant protein, similarly shows activity only against Gram-positive bacteria with particularly strong activity against *S. aureus* and *Streptococcus iniae* but has no effect on *A. viridans* or *M. luteus* [41]. Unusually, however, another recombinant Type II crustin (EF654658) from the same species has recently been reported to show strong bactericidal activity not only against Gram-positive bacteria, such as including *A. viridan*, but also against *E. coli* 363 and the pathogen,

Vibrio harveyi, both Gram-negative bacteria [43]. This *P. monodon* protein is the only crustin so far claimed to kill Gram-negative bacteria. As yet no studies have been made of the spectra of antibacterial activities of the Type III molecules and this is urgently required to confirm their status as members of the crustin family. It would be very interesting to see if they too kill Gram-positives and have any effects at all on Gram-negatives. Thus whilst several crustins clearly have antibacterial effects primarily against Gram-positive bacteria, broader specificity to deal with a wider range of pathogens can occur and seems to be achieved through sequence variations. The extent to which single nucleotide substitutions, such as occurs in isoforms, alter function remains largely untested, as does the degree to which crustins synergise with other AMPs in the host. Indeed, given that Gram-negative bacteria tend to dominate the prokaryotic microflora of the marine environment, it seems counter-intuitive for such prevalent antibacterial proteins in decapods to have restricted spectra of activity.

Unfortunately, the mechanism of action of crustins on susceptible bacteria is unknown. Many animal AMPs, especially those with amphipathic structures, are known to kill their bacterial targets by depolarising and permeabilizing the outer bacterial cell wall, often by either a 'barrel-stave' or 'carpet' mechanism [54, 55] although others may disrupt cell metabolism or interfere with DNA synthesis [56]. As crustins all share a common WAP domain, this part of the molecule must play a key role in its antibacterial effects. Indeed in *F. chinensis*, a full length cDNA has been found that occurs alongside the Type II crustin and, while it is identical to it in possessing a signal sequence and a gly-rich region, it has no WAP domain and does not exhibit any antibacterial effect [42]. Moreover, the antibacterial activity of a WAP-containing protein from snake venom has

been shown to lose its bactericidal effects upon reduction and alkylation of the WAP cysteine residues [57]. Thus it would appear that tertiary structure of the 4DSC is essential for microbicidal activity. In mammals, the WAP domain is associated with serine protease inhibition and exerts its effects by inserting its inhibitory loop into the active site pocket of the target protease and interfering with its catalytic residues [16, 17, 23]. Protease inhibitor activity is believed to be characterised by a methionine residue adjacent to the second cysteine in the 4DSC [58]. In those vertebrate WAP-domain containing proteins that have antibacterial properties, however, this is replaced by a cationic or hydrophobic amino acid, indicating that methionine in this position dictates bioactivity [24, 55]. Methionine is also seldom present in the WAP domain of crustins, so that they too are probably all bactericidal.

Patterns of expression

Evidence of the role of crustins as direct antimicrobial defence effectors has been sought from studies of their expression in different tissues and following experimental infection. Certainly most crustins seem to be constitutively expressed by the blood cells [9, 11, 31, 36, 38, 41, 43, 45, 59] often at very high levels [36, 59]. The proteins seem to be synthesised in the granular haemocytes, at least from the few studies that have studied different haemocyte populations [9, 11]. Transcripts of crustin-encoding genes have also been observed in gills, heart and, intestine [41, 42, 59] but as these tissues are highly vascularised, it is not clear if the signal from these organs is due primarily to the haemocytes. Surprisingly, an EST study by Gross et al [61] of the haemocytes and hepatopancreas from *L. vannamei* and *L. setiferus* found that while multiple copies of crustin genes were expressed by haemocytes, none were expressed in the hepatopancreas.

By contrast, two crustin-like transcripts have been identified from regenerating tissues of decapods; one, PET-15, was identified from regenerating epithelial tissue in the olfactory organ of the spiny lobster, *Panulirus argus* [35] while the other ([DW176897](#)) is expressed in regenerating limbs of the fiddler crab, *Celuca pugilator* [62]. It is unknown if crustin-like genes are expressed during tissue regeneration or wound repair in other species but, irrespective of whether or not crustins are produced by multiple tissues, one would expect them, on account of their unquestionable haemocytic origin, to be synthesised in developing or maturing haemocytes in the haemopoietic tissue. However, only one research group [32] has made a study of AMP expression in this organ and they have found that the host, the crayfish, *P. leniusculus*, expresses constitutively only one of its three crustins (the 'atypical' *Plcrustin2*) in this tissue.

Patterns of expression during development are also very poorly understood, as most studies of crustins have been performed only on late stage postlarvae or adults. However, Hauton et al. [63] have shown that lobster, *H. gammarus*, postlarvae at both stage IV and stage VI express crustin genes with levels of expression similar at both developmental stages. Larvae of the shrimp, *P. monodon*, have been further reported to express a crustin transcript at high levels at all stages of development from nauplii stage IV through to juveniles [64]. This concurs with a recent finding that in shrimp, at least, discernable granular haemocytes appear at the nauplii stage of ontogenesis [65]. It is possible that crustins are also expressed at earlier life stages as one of the present authors (Smith V.J, unpublished) has obtained preliminary evidence, from Northern and dot blot analyses, that crustin-like transcripts are expressed in *Macrobrachium rosenbergii* eggs from 1-14 days post fertilization and through all nine stages of larval development to postlarvae and

adults. Unfortunately, no studies have tracked the sites of tissue expression during development so we do not know where the proteins are synthesised and stored in the early life stages of these animals.

The response of crustin expression to bacterial challenge is also enigmatic and often does not follow the pattern expected for immune genes and other AMPs from suppressive subtraction hybridisation studies following challenge by known pathogens [46, 61, 66]. For example, expression of *L. vannamei* Type II crustin is down-regulated at ca 12-24 h following injection of *Vibrio alginolyticus* [38] whereas the Type III transcript in this species is up-regulated by the same bacterium at 3 and 6 h post treatment but down-regulated at 24 h [45]. In one of the *P. monodon* Type II crustins (EF654658), expression following injection at 24 h previously of the Gram-negative pathogen, *V. harveyi*, increases approximately 5 fold, at least compared to the housekeeping gene, but returns to normal levels by 72 h [43]. The crayfish, *P. leniusculus*, also shows up-regulation of expression of Type I *Plcrustin1* and Type II *Plcrustin3* in both the haemocytes and the haemopoietic tissue after challenge with the Gram-negative bacteria *Escherichia coli* or *Acetivobacter* spp but *Plcrustin 3* shows no change [32] in response to these micro-organisms. A few authors have observed an unexpected response of down-regulation of crustin transcripts after bacterial challenge, usually with Gram-negative bacteria. Examples include the Type II crustin from *P. monodon* [41] and the Type I from *H. gammarus* [36]. To complicate the picture further, the *H. gammarus* crustin is up-regulated, at least initially, after challenge with the Gram-positive lobster pathogen, *Aerococcus viridans* var *homari*, although expression subsequently goes down from 9-36 h post treatment with this pathogen [36]. Unusually, carcinin expression in *C. maenas* (a

Type I crustin) remains unchanged after injection of the Gram-positive bacterium, *Planococcus citreus* for at least 48 h but is down-regulated ca four fold at 84 h [60].

Clearly no consistent pattern emerges but it should be borne in mind that little heed has been paid to either the half life-of the mature protein in the haemolymph or any differential responses between crustin isoforms within the same species. One exception to this is a study by Vargas Albores et al. [38] on *L. vannamei*, where the I-crustin isoform was seen to be down regulated 12-24 h after challenge with *V. alginolyticus* whereas the P-crustin isoform remained unchanged. Furthermore, different studies have used a variety of bacterial strains, administered at widely different doses and the responses in some cases have been measured only semi-quantitatively and usually only on small numbers of experimental animals. It is likely that comparing the response to Gram-negative bacteria with that to Gram-positives might be too simplistic, as factors such as surface configuration of the test micro-organism or its pathogenicity for the host might be more important.

One paper that has attempted to measure the response of crustins to different doses of injected LPS has shown that crustin mRNA levels in *L. vannamei* significantly decrease between 4 and 72 h [67]. Unfortunately, however, in this paper no account seems to have been taken of the reduction in the size of the circulating haemocyte population which is known to occur in decapods following such non-self onslaught [30, 31, 68, 69], so the effect reported for *L. vannamei* [67] may be smaller than it appears initially. Thus we still need to know how expression of this group of AMPs responds to pathogen associated molecular patterns because some, especially glucans, are key components of compounds

marketed to the aquaculture sector as dietary supplements for improving survival against transmissible microbial diseases on account of the purported immune stimulating or enhancing properties of their constituents. Whilst positive effects have been claimed for such compounds in improving the antibacterial defences of decapods [70], significant enhancement of crustin expression was not seen in *H. gammarus* postlarvae treated with some commercial immunostimulants [63].

Concluding comments

Because crustins have antibacterial properties, they must contribute to a greater or lesser extent in defence against bacterial infections, at least in decapods. However, the variable pattern in their expression after bacterial challenge is unlike that known for other arthropod AMPs [71, 72] and makes these molecules highly enigmatic. It begs the question if their only biological role in the decapod host is as direct antimicrobial effectors. Their haemocytic location points to an association with host defence and whilst immunity is a major preoccupation for these cells, it is not the only one. Certainly, those, such as carcinin, that are expressed at very high levels must play an important role for the host but whether the function of these proteins is solely for bacteriostasis or for some other aspect of homeostasis remains to be determined. The expression of two Type I crustins in regenerating tissues [35, 62] is suggestive of a possible role for crustins in recovery from trauma or response to physiological stress or as negative regulators of other host-defence factors. This notion is supported by the observations of Brockton and Smith [60] who found that carcinin expression in *C. maenas* is up-regulated threefold in crabs held at 5 °C and 20 °C, compared to those maintained at ambient seawater temperatures of 10-15 °C. Thus, carcinin expression does not follow the pattern expected

of a thermally regulated metabolic response and thus must be a reaction to the stress of encountering temperatures at the extremes of its usual winter and summer range. Since infection, injury and stress are common bedfellows and recovery involves numerous whole body compensatory processes, often mediated through the haemocytes, crustins or crustin-like molecules may well be up-regulated by a variety of triggers. Clearly, to resolve the enigmas about crustins, more directed functional and proteomic studies need to be undertaken, especially with respect to the diversity of bioactivities of the natural proteins, as well as considering how expression, processing and bioactivities relate to different types of threats to homeostatic integrity.

References

- [1] Acton RT, Weinheimer PF, Evans E.E. A bactericidal system in the lobster *Homarus americanus*. J Invertebr Pathol 1969; 13: 463-464.

- [2] Stewart JE, Zwicker BM. Induction and natural bactericidal activities of the lobster *Homarus americanus*: Products of haemocyte:plasma interactions. Can J Microbiol 1972; 18: 1499-1511.

- [3] Chisholm JRS, Smith VJ. Antibacterial activity in the haemocytes of the shore crab, *Carcinus maenas* (L.). J Mar Biol Ass UK. 1992; 72: 529-542.

- [4] Schnapp D, Kemp GD, Smith VJ. Purification and characterization of a proline-rich antibacterial peptide, with sequence similarity to bactenecin 7, from the haemocytes of the shore crab, *Carcinus maenas*. Eur J Biochem 1996; 240: 532-539.

- [5] Destoumieux D, Bulet P, Loew D, van Dorsellaer A, Rodriguez J, Bachère E. Penaeidins: a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). J Biol Chem 1997; 272: 28398–28406.

- [6] Bachère E, Destoumieux D, Bulet P. Penaeidins, antimicrobial peptides of shrimp: a comparison with other effectors of innate immunity. *Aquaculture* 2000; 191:71-88.
- [7] Bachère E, Gueguen Y, Gonzalez M, Lorget J, Garnier J, Romestand B. Insights into the antimicrobial defence of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas*. *Immunol Rev* 2004; 198, 149-168.
- [8] Gueguen Y, Garniera J, Robert L, Lefranc M-P et al. PenBase, the shrimp antimicrobial peptide penaeidin database: Sequence-based classification and recommended nomenclature. *Dev Comp Immunol* 2006; 30: 283–288.
- [9] Relf JM, Chisholm JRS, Kemp GD, Smith VJ. Purification and characterisation of a cysteine-rich 11.5 kDa antibacterial peptide from the granular haemocytes of the shore crab, *Carcinus maenas*. *Eur J Biochem* 1999; 264: 1-9.
- [10] Smith VJ, Chisholm JSR. Antimicrobial proteins in crustaceans. In: Phylogenetic Perspectives on the Vertebrate Immune System. Beck G, Sugumaran M, Cooper EL ed. Kluwer Academic/Plenum Publ. Adv. Exp Med Biol 2001; 484: 95-112.
- [11] Brockton V, Hammond JA, Smith VJ. Gene characteristics, isoforms and recombinant expression of carcinin, an antibacterial protein from the shore crab, *Carcinus maenas*. *Mol Immunol* 2007; 44: 943-949.
- [12] Bartlett TC, Cuthbertson BJ, Shepard EF, Chapman RW, Gross PS, Warr GW. Crustins, homologues of an 11.5-kDa antibacterial peptide from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Mar. Biotechnol* 2002; 4: 278-293.
- [13] Ranganathan S, Simpson KJ, Shaw DC, Nicholas K. The whey acidic protein family: A new signature motif and three-dimensional structure by comparative modeling. *J Mol Graph Model* 1999; 17: 106–113.
- [14] 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. *Biochem J* 1999; 340: 569-577.

[15] Bouchard D, Morisset D, Bourbonnais Y, Tremblay GM. Proteins with whey-acidic-protein motifs and cancer. *Lancet Oncol* 2006; 7: 167–74.

[16] Hiemstra PS, Massen RJ, Stolk J, Heinzl-Wieland R, Staffens GJ, Dijkman JH. Antibacterial activity of antileukoprotease. *Infect Immunol* 1996; 64: 4520-4524.

[17] Sallenave JM. Antimicrobial activity of antiproteases. *Biochem Soc. Trans* 2002; 30: 111-115.

[18] Seemuller U, Arnhold M, Fritz H, Wiedenmann K, Machleidt W, Heinzl R, Appelhans H, Gassen Hg, Lottspeich F. The acid-stable proteinase inhibitor of human mucous secretions (HUSI-I, antileukoprotease): Complete amino acid sequence as revealed by protein and cDNA sequencing and structural homology to whey proteins and Red Sea turtle proteinase inhibitor. *FEBS Lett* 1986; 199: 43-48.

[19] Talas-Ogras T. Screening antimicrobial activities of basic protein fractions from dry and germinated wheat seeds. *Bio Plant* 2004; 48: 583-588.

[20] Andrenacci D, Grimaldi MR, Panetta V, Riano, E, Rugarli EI, Graziani F. Functional dissection of the *Drosophila* Kallmann's syndrome protein DmKal-1. *BMC Genetics* 2006; 7:47. doi:10.1186/1471-2156-7-47.

[21] Nile CJ, Townes CL, Hirst BH, Hall J. The novel avian protein, AWAK, contains multiple domains with homology to protease inhibitory modules. *Mol Immunol* 2006; 43: 388–394.

[22] Treccani L, Mann KH, Heinemann F, Fritz M. Perlwapin, an abalone nacre protein with three four-disulfide core (WAP) domains, inhibits the growth of calcium carbonate crystals. *Biophys J BioFAST*, <http://www.biophysj.org>.

[23] Hiemstra PS. Novel roles of protease inhibitors in infection and inflammation. *Biochem Soc Trans* 2002; 30: 116-120.

[24] Hagiwara K, Kikuchi T, Endo Y, Huqun Usui K, Takahashi M, Shibata, N, Kusakabe T, Xin

H, Hoshi S, Miki M, Inooka N, Tokue Y, Nukiwa T. Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif .J Immunol 2003; 170: 1973-1979.

[25] Jimenez-Vega F, Vargas-Albores F. A secretory leucocyte proteinase inhibitor (SLPI)-like protein from *Litopenaeus vannamei*. Fish Shellfish Immunol 2007; 23: 1119-1126.

[26] Lehrer RI, Ganz T. Antimicrobial peptides in mammalian and insect host defence. Curr Opin Immunol 1999;11: 23–27.

[27] Tomasinsig L, Zanetti, M. The cathelicidins: structure, function and evolution. Curr Prot Pept Sci 2005; 6:23-34.

[28] Destoumieux D, Muñoz M, Cosseau C, Rodriguez J, Bulet P, Comps M, Bachère E. Penaeidins, antimicrobial peptides with chitin binding activity, are produced and stored in shrimp granulocytes and released after microbial challenge. J Cell Sci 2000; 113: 461-469.

[29] Munoz M, Vandenbulcke F, Saulnier D, Bachère E. Expression and distribution of penaeidin antimicrobial peptides are regulated by haemocyte reactions in microbial challenged shrimp. Eur J Biochem 2002; 269: 2678-2689.

[30] Söderhäll K, Cerenius L. Crustacean immunity. Annu Rev Fish Dis 1992; 2:3-23.

[31] Söderhäll K, Cerenius L. Role of prophenoloxidase-activating system in invertebrate immunity. Curr Opin Immunol 1998;10: 23-28.

[32] Jiravanichpaisal P, Lee SY, Kim Y-A, Tove André T, Söderhäll I. Antibacterial peptides in hemocytes and hematopoietic tissue from freshwater crayfish *Pacifastacus leniusculus*: Characterization and expression pattern. Dev Comp Immunol 2007; 31: 441–455.

[33] Schwede T, Kopp J, Guex, N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003; 31: 3381-3385.

- [34] Francart C, Dauchez M, Alix AJP, Lippens G. Solution structure of R-elafin, a specific inhibitor of elastase. *J Mol Biol* 1997; 268: 666-667.
- [35] Stoss TD, Nickell MD, Hardin D, Derby CD, McClintock TS. Inducible transcript expressed by reactive epithelial cells at sites of olfactory sensory neuron proliferation. *J. Neurobiol* 2003; 58: 355-368.
- [36] Hauton C, Brockton V, Smith VJ. 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. *Mol Immunol* 2006; 43: 1490-1496.
- [37] Christie AE, Rus S, Goiney CC, Smith C, Towle DW, Dickinson PS. Identification and characterization of a cDNA encoding a crustin-like putative antibacterial protein from the American lobster *Homarus americanus*. *Mol Immunol* 2007; 44: 3333-3337.
- [38] Vargas-Albores F, Yepiz-Plascencia G, Jimenez-Vega F, Avila-Villa A. Structural and functional differences of *Litopenaeus vannamei* crustins. *Comp Biochem Physiol B-Biochem Mol Biol* 2004; 138: 415-422.
- [39] Chen JY, Chuang H, Pan CY, Kuo CM. cDNA sequence encoding an antimicrobial peptide of chelonianin from the tiger shrimp *Penaeus monodon*. *Fish Shellfish Immunol* 2005; 18: 179-183.
- [40] 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. *Dis Aq Org* 2004; 6: 123-135.
- [41] Supungul P, Tang S, Maneeruttanarungroi C, Timphanitchayakit V, Hirono I, Aoki T, Tassanajon A. Cloning, expression and antimicrobial activity of crustinPm1, a major isoform of crustin, from the black tiger shrimp *Penaeus monodon*. *Dev Comp Immunol* 2007; doi:10.1016/j.dci.2007.04.004
- [42] Zhang J, Li F, Wang Z, Xiang J. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. *J Biotech* 2007; 127: 605-614.

- [43] Amparyup P, Kondo H, Hirono I, Aoki T, Tassanakajon A. Molecular cloning, genomic organisation and recombinant expression of a crustin like antimicrobial peptide from black tiger shrimp *Penaeus monodon*. *Mol Immunol* 2008; 45: 1085-1093.
- [44] Rosa RD, Bandeira PT, Barracco MA. Molecular cloning of crustins from the hemocytes of Brazilian penaeid shrimps. *FEMS Microbiol Lett* 2007; 274: 287-290.
- [45] Jimenez-Vega F, Yepiz-Plascencia G, Soderhall K, Vargas-Albores F. A single WAP domain-containing protein from *Litopenaeus vannamei* hemocytes. *Biochem Biophys Res Commun* 2004; 314: 681-687.
- [46] Rojtinnakorn J, Hirono, I, Itami T, Takahashi Y, Aoki T. Gene expression in haemocytes of kuruma shrimp, *Penaeus japonicus* in response to infection with WSSV by EST approach. *Fish Shellfish Immunol* 2002; 13: 69-83.
- [47] Rattanachai A, Hirono I, Ohira T, Takahashi Y Aoki T. Cloning of kuruma prawn *Marsupenaeus japonicus* crustin-like peptide cDNA and analysis of its expression. *Fish Sci* 2004; 70: 765-771.
- [48] Selsted ME. Purification, primary structures and antibacterial activities of beta defensins, a new family of antimicrobial peptides from bovine neutrophils. *J Biol Chem* 1993; 268: 6641-6648.
- [49] Kawabata S, Nagayama R, Hirata M, Shigenaga T, Agarwala KL, Saito T, Cho J, Nakajima H, Takagi T, Iwanaga S. Tachycidin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. *J Biochem* 1996; 120: 1253-1260.
- [50] Froy O, Gurevitz M. Arthropod and mollusk defensins- evolution by exon shuffling. *Trends Genetics* 2003; 19: 684-687.
- [51] Ganz T. Defensins: Antimicrobial peptides of innate immunity. *Nature Rev Immunol* 2003; 3: 710-720.

- [52] Haug T, Stensvåg K, Vasskog T, Sandsdalen, Styrvold O. Antibacterial peptides isolated from the haemocytes of the small spider crab, *Hyas araneus* (Abst.) 10th Int Congress of ISDCI, Charleston, SC, USA. 2006.
- [53] Zhang J, Li F, Wang Z, Xiang J. Expression, purification, and characterization of recombinant Chinese shrimp crustin-like protein (CruFc) in *Pichia pastoris*. *Biotech Lett* 2007; 29: 813-817
- [54] Huang HW. Action of antimicrobial peptides: two-state model. *Biochem* 2002; 39: 8347-8352.
- [55] Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature* 2002; 415: 389-395.
- [56] Devine DA, Hancock RE. Cationic peptides: distribution and mechanisms of resistance. *Curr Pharm Res* 2002; 8:703-714.
- [57] Nair DG, Fry BG, Alewood P, Kumar PP, Kini RM. Antimicrobial activity of omwaprin, a new member of the waprin family of snake venom proteins. *Biochem J* 2007; 402: 93-104.
- [58] Ota Y, Shimoya K, Zhang Q, Moriyama A, Chin A, Murata Y. The expression of secretory leukocyte protease inhibitor (SLPI) in the fallopian tube: SLPI protects the acrosome reaction of sperm from inhibitory effects of elastase. *Hum Reprod* 2002; 17: 2517– 2522.
- [59] Imjongirak C, Amparyup P, Tassanakajon A, Sittipraned S .Cloning of a carcinin-like single whey acidic domain antimicrobial peptide from hemocytes of the mud crab *Scylla partamamosan*, *Proc 32nd Congress Sci Technol Thailand*. Article B4B0038.
http://www.scisoc.or.th/stt/32/sec_b/paper/stt32_B4_B0038.pdf
- [60] Brockton V, Smith V.J. Change in crustin expression following bacterial injection and temperature change in the shore crab, *Carcinus maenas*. Submitted.
- [61] Gross PS, Bartlett TC, Browdy CL, Chapman RW, Warr GW. Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. *Dev Comp Immunol* 2001;

25: 565–577.

[62] Durica DS, Kupfer D, Najar F, Lai H, Tang Y, Griffin K, Hopkins PM, Roe B. EST library sequencing of genes expressed during early limb regeneration in the fiddler crab and transcriptional responses to ecdysteroid exposure and limb bud explants. *Int Comp Biol* 2006; 46: 948-969.

[63] Hauton C, Brockton V, Smith VJ. *In vivo* effects of immunostimulants on gene expression and disease resistance in lobster (*Homarus gammarus*) post-larval stage VI (PLVI) juveniles. *Mol Immunol* 2007; 44: 443-450.

[64] Jiravanichpaisal P, Puanglarp N, Petkon S, Donnuea S, Söderhäll I, Söderhäll K. Expression of immune-related genes in larval stages of the giant tiger shrimp, *Penaeus monodon*. *Fish Shellfish Immunol* 2007; 23: 815-824.

[65] Lin Y, Zhan W, Li Q, Zhang Z, Wei X, Sheng X. Ontogenesis of haemocytes in shrimp (*Fenneropenaeus chinensis*) studied with probes of monoclonal antibody. *Dev Comp Immunol* 2007; 31: 1073-1081.

[66] Lorgeril J, Saulnier, D, Janech MG, Guguen Y, Bachere E. Identification of genes that are differentially expressed in hemocytes of the Pacific blue shrimp (*Litopenaeus stylostris*) surviving an infection with *Vibrio penaeicida*. *Physiol Genomics* 2005; 21: 174-183.

[67] Okumura T. Effects of lipopolysaccharide on gene expression of antimicrobial peptides (penaeidins and crustin), serine proteinase and prophenoloxidase in haemocytes of the Pacific white shrimp, *Litopenaeus vannamei*. *Fish Shellfish Immunol* 2007; 22: 68-76.

[68] Söderhäll, K, Smith VJ, Johansson M. Exocytosis and phagocytosis by isolated haemocyte populations of crustaceans: evidence for cell co-operation in the cellular defence reactions. *Cell Tiss Res* 1986; 245: 43-49.

[69] Lorenzon S, De Guarrini, S, Smith VJ, Ferrero EA. Effects of LPS injection on circulating haemocytes in crustaceans *in vivo*. *Fish Shellfish Immunol* 1999; 9: 31-50.

[70] Raa J. The use of immunostimulatory substances in fish and shellfish farming. *Rev Fish Sci* 1996; 4: 229-288.

[71] Bulet P, Hetru C, Dimarcq JL, Hoffmann D. Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol* 1999; 23: 329-344.

[72] Engstrom Y. 1999. Induction and regulation of antimicrobial peptides in *Drosophila*. *Dev Comp Immunol* 1999; 23: 345-358.

[73] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press 2005: 571-607.

Figure legends

Fig. 1. Schematic representation (not to scale) of the domain organisation of the three main crustin types from decapods. Signal sequence (ss).

Fig. 2. Structural models WAP domain in various decapod crustins created with Swiss-Model Server [33] using the elafin WAP domain of vertebrates [34] as a template.

Fig. 3. Phylogenetic tree of whole protein sequences, highlighting the SWDs (box C) and their relationship to Type I and Type II crustins (boxes A and B respectively) and other single WAP domain containing proteins. Phylogenetic analysis was performed with *Mr Bayes* using an average mixed model of amino acid evolution. This consensus tree was derived from the last 9,000 trees after convergence. Bootstrap values corresponding to Bayesian posterior probabilities are indicated at the tree nodes. Sequences in italics are from non-crustacean species. X04502 *Homo sapiens* secretory leucocyte protease inhibitor (SLPI); P03973 *H. sapiens* antileukoprotease (ALPI); AF151982 *Rattus norvegicus* SLPI; P97430 *Mus musculus* ALP; P19957 *H. sapiens* elafin; Q29125 *Sus scrofa* elafin; AF276975 *M. musculus* single WAP motif protein 2; P00993 *Caretta caretta* chelonianin; AF037272 *R. norvegicus* WAP domain-containing protein; Q9H1F0 *H. sapiens* WAP domain-containing protein.

Fig. 4. Schematic representations (not to scale) of the amino acid substitutions in crustin isoforms from *L. vannamei* and *C. maenas*.

Fig. 5. Typical sequences of non-decapod crustaceans compared to representative decapod crustins. (a) Alignment of the amphipod crustin-like EST with representative Type II crustins. (b) Alignment of the WAP domain of *C. finmarchicus* with representative Type I, II and III crustins.

Fig. 6. Unrooted radiation tree of the signal sequences of known decapod crustins. After aligning the peptide sequences with ClustalW, a Bayesian phylogenetic reconstruction was performed using an average mixed model of amino acid substitution and run for 1,000,000 generations. Supporting Bayesian posterior probabilities, based on the last 9,000 trees, are shown at the tree nodes. Details of the origin of the sequences including GenBank accession numbers are given in Table 1. Group A denotes sequences isolated from shrimp. The large solid triangle includes sequences from *L. vannamei*, *L. setiferis*, *L. schmitti*, *F. brasillensis*, *F. chinensis*, *F paulensis* and *F. subtilis* (see Table 1). Group B identifies shrimp SWD and chelonianin sequences.

Fig 7. Unrooted radiation tree of crustins based on similarity of the WAP domains produced by Bayesian inference using an average mixed model of amino acid evolution. The percentages at the tree nodes correspond to Bayesian posterior probabilities, calculated from the last 900,000 generations after a 10% burn-in phase. Details of the origin of the sequences including GenBank accession numbers are given in Table 1. Group A denotes sequences isolated from shrimp *L. vannamei*, *L. setiferis*, *L. schmitti*, *F. brasillensis*, *F. chinensis*, *F paulensis* and *F. subtilis*. Group B identifies shrimp SWD and chelonianin sequences.

Table 1. Summary of known crustin or crustin-like sequences

Species	Common name	Accession No	Sequence type	Coded amino acids	Mature protein molecular mass (kDa)	PI of mature protein	Signal sequence amino acids	Crustin type	Number of isoforms	Key reference
Decapoda: Brachyura										
<i>Carcinus maenas</i> (carcinin)	Shore crab	AJ427538	cDNA	111	11.5 *; 12.2	8.7	21 Val-rich	I	6	[11]
<i>Hyas areneus</i>	Spider crab	Not available	Partial aa	90	10.1 & 10 *		23	I	2	[52]
<i>Callinectes sapidus</i>	Blue crab	CV462984	EST	70				I		Shafer <i>et al.</i> , 2004 (unpubl)
<i>Celuca pugilator</i>	Fiddler crab	DW176897	EST	90				I		Tang <i>et al.</i> , 2005 (unpubl)
<i>Scylla paramamosain</i>	Mud crab	Not available	cDNA	111			21	I	7	[59]
<i>Portunus pelagicus</i>	Sand crab	EF120999	cDNA	108	9.8	5.7	20	I		Kuballa & Elizur (unpubl)
Decapoda: Astacidea										
<i>Homarus gammarus</i>	European lobster	AJ786653	cDNA	111	10.5	7.0	17 Val-rich	I		[36]
<i>Homarus americanus</i>	American lobster	EF193003	cDNA	112	10.6	6.9	16 Val-rich	I		[37]
<i>Pacifastacus leniusculus</i> (Plcrustin1)	Signal crayfish	EF523612	cDNA	118,	11.7	6.9	16 Val-rich	I		[32]
<i>Pacifastacus leniusculus</i> (Plcrustin2)	Signal crayfish	EF523613	cDNA	109	10.0	8.3	16 Val-rich	I		[32]
<i>Pacifastacus leniusculus</i> (Plcrustin3)	Signal crayfish	EF523614	cDNA	157	12.3	8.7	19	II		[32]
Decapoda: Palinura										
<i>Panulirus argus</i> (PET-15)	Spiny lobster	AY340636	cDNA	95	8.8	7.0	18	I		[35]
Decapoda: Penaeoidea										
<i>Litopenaeus vannamei</i> (crustin)	White shrimp	AF430071-6	cDNAs	163	13.4-14.7	7.6-8.0	18	II	6	[12]
<i>Litopenaeus vannamei</i> (crustin)	White shrimp	AY488494	cDNA	163	14.3	7.6	18	II		[38]
<i>Litopenaeus vannamei</i> (SWD)	White shrimp	AY465833; AY464465	cDNAs	93	7.3		24 Val-rich	III		[45]
<i>Litopenaeus vannamei</i> (chelonianin)	White shrimp	BE 188538	EST	77			24 Val rich	III		[61]
<i>Litopenaeus setiferus</i>	White shrimp	AF430077	cDNA	123	11.2	8	19	II?		[12]
<i>Litopenaeus setiferus</i>	White shrimp	AF430078	cDNA	170	16.1	7.6	18	II	3	[12]
<i>Litopenaeus setiferus</i>	White shrimp	AF430079	cDNA	141	14.6	8.6	19			[12]
<i>Litopenaeus schmitti</i>	Brazilian shrimp	EF182748	cDNA	166	14.5	7.6	19	II		[44]
<i>Penaeus monodon</i> (chelonianin)	Tiger shrimp	Not on database	cDNA	92	5.4	5.0	24 Val-rich	III		[39]
<i>Penaeus monodon</i>	Tiger prawn	EF654658	cDNA	141	12.5	7.8	17 Val-rich	II		[43]

<i>Penaeus monodon</i> (crustin)	Tiger prawn	CD766060	cDNA	145	13.7	8.5	18	II		[42]
<i>Marsupenaeus japonicus</i> (crustin)	Kuruma prawn	AB121740-4	cDNA	145	16.8	8.3	17	II	5	[42]
<i>Fenneropenaeus chinensis</i> (crustin)	Chinese prawn	AY871268	cDNA	130	12.3: 11.8 *	8.5	17 Val-rich	II	2	[43]
<i>Fenneropenaeus chinensis</i> (SWD)	Chinese prawn	EF216349	cDNA	90	7.0	8.5	24 Val-rich	III		Jia <i>et al.</i> , (unpubl)
<i>Farfantepenaeus brasiliensis</i>	Pink spotted shrimp	EF601055	cDNA	70	15.3	8.3	18	II		[44]
<i>Farfantepenaeus paulensis</i>	Pink shrimp	EF182747	cDNA	168	14.7	8.3	18	II		[44]
<i>Farfantepenaeus subtilis</i>	Brown shrimp	EF450744	cDNA	146	13.0	8.0	18	II		[44]

Amphipoda

<i>Gammarus pulex</i>	Amphipod	EH273178, EH268909	EST	>156aa				II?	2	Sambles <i>et al.</i> , (unpubl)
-----------------------	----------	--------------------	-----	--------	--	--	--	-----	---	----------------------------------

Molecular mass values are for the mature protein after removal of the signal sequence as predicted by ProtParam [73]. Those marked with an asterisk are measured values of purified native or recombinant proteins. PI values were predicted for the mature protein using the ProtParam server. Coded amino acids are for the immature protein. Signal sequences were predicted by the SignalP 3.0 server. See text for definition of crustin types. Unpublished citations are derived from databases.

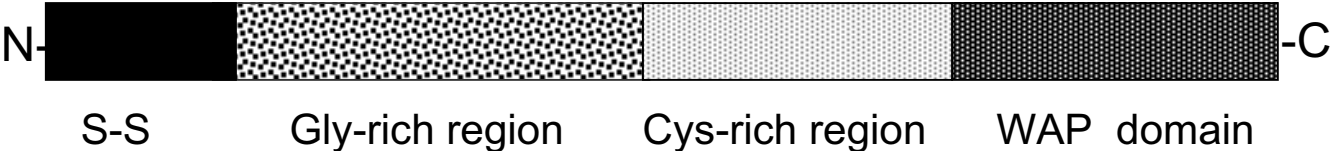
Figure 1

Type I



Eg Crab, lobster, crayfish and other crustins in the Pleocyemata

Type II



Eg Shrimp crustins

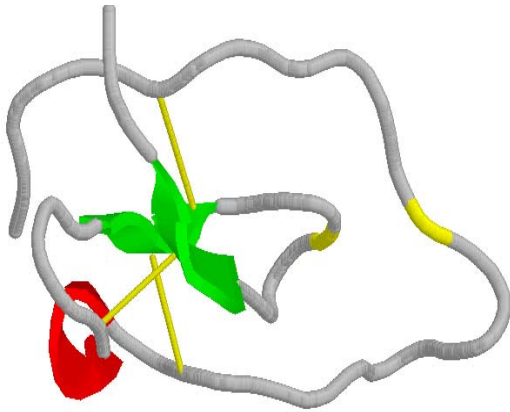
Type III



Eg SWD and chelonianin-like crustins

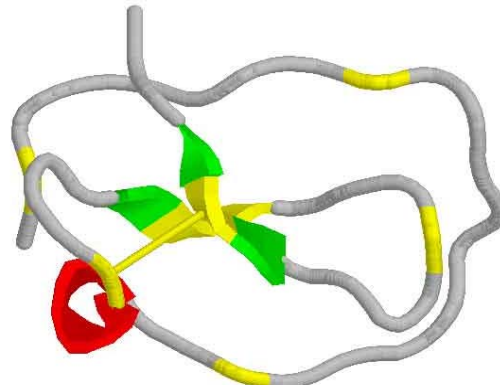
Fig 2

Type I



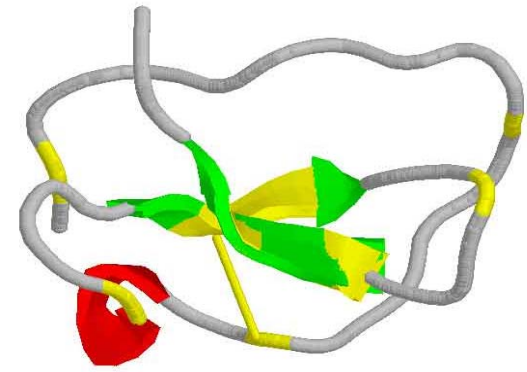
carcinin

Type II



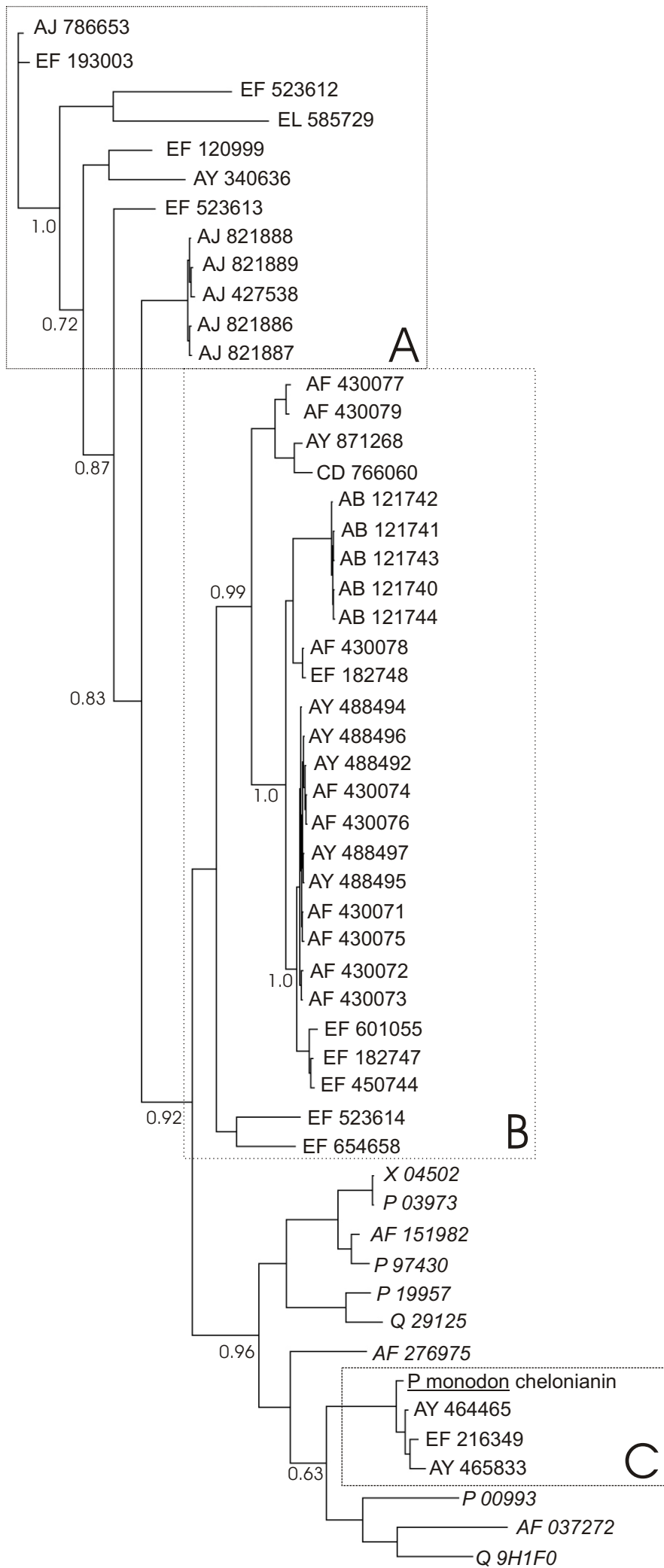
L. vannamei crustin

Type III



L. Vannamei SWD

Figure(s)

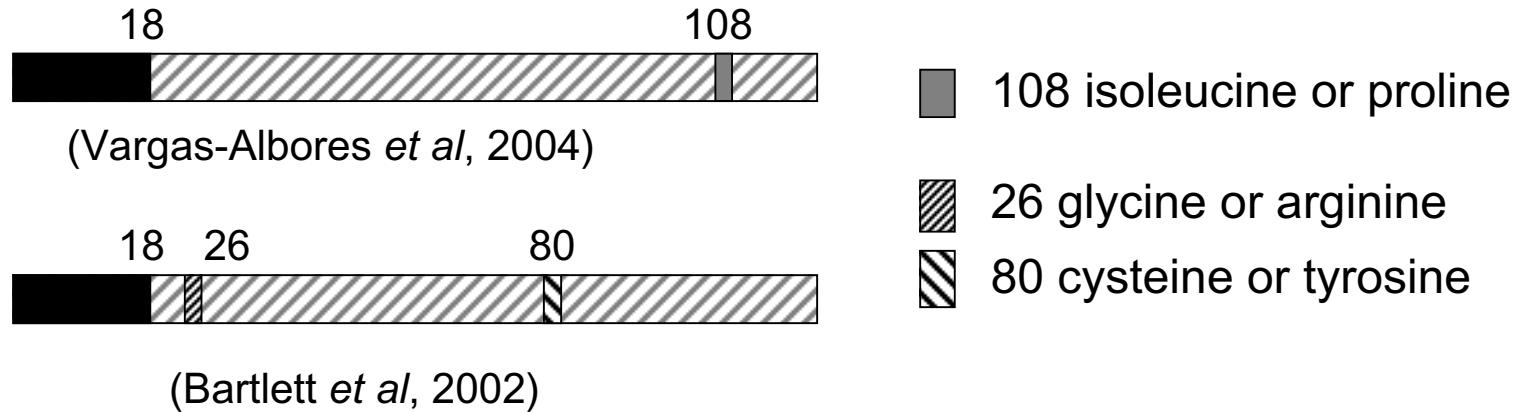


0.2

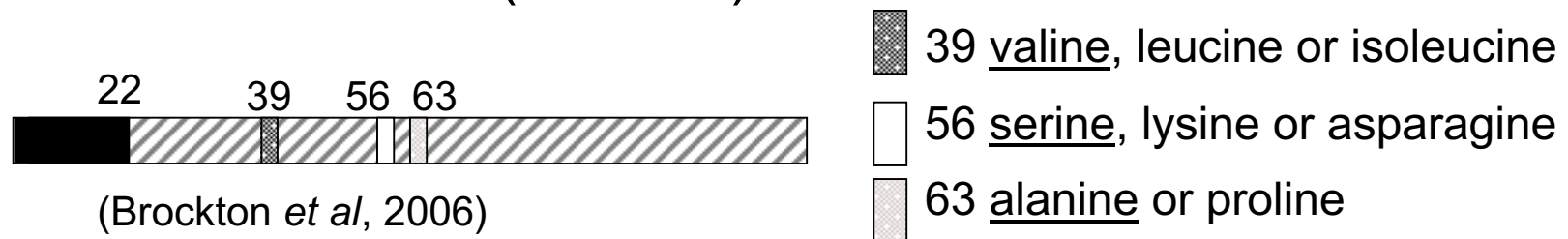
Figure 3

Figure 4

Litopenaeus vannamei crustin



Carcinus maenas (carcinin)



18-22 aa signal peptide

G.pulex EH268909GGFSSGGAGYGCNVGGGFPAQAVGGFSSGGFPG	33
L. vannamei AY488494MKGLKAVILCGLFTAVLAGKFRGFGQPFGG.LGGPGGGV...GV	40
L. setiferus AF430077MKGLGVILCCVLAIVPAHAG	20
L. schmitti EF182748MKGLKAVILCGLFTAVLAGKYRGGFQPLGG.LGVPPGGGVGV.GV	42
Fe. chinensis AY871268MKGLGVILFCVLAIVASQSRH.....G.....IR	24
Fa. subtilis EF450744MKGLQAVIILGLLITAVLAGKFRGFGSPFG...GG.....GV	33
Pe. monodon CD766060MKGLGVILFCVLAIVASQSWH.....G.....GR	24
M. japonicus AB121740MKGFKAVVLCSSLASALAGKLR.FVRSFGGYGGGVGGVHGGGL	43

G.pulex EH268909	GGCFQ...CAGINQFPAGGHHGGGFIQPPGGFPGGVVGPPSG.....	72
L. vannamei AY488494	GGCFP...CGGLG...VGGGLG...VGGGLGVGGGLGVGGGL.....	73
L. setiferus AF430077	PGCFP...GG...VPCR.....FPSAT.....	36
L. schmitti EF182748	GGGL...CGGLGG.SLGGGLGGG...LGGGLGGGLGGGLGG.....	76
Fe. chinensis AY871268	PGCFP...GG...FPGG.....FPSIT.....	40
Fa. subtilis EF450744	GGCFH...CGGLG...VGGGLG...VGGGICVG.....	57
Pe. monodon CD766060	PGCFP...CGGRPGGFPGGGRPG...RPGGFPSVI.....	54
M. japonicus AB121740	GNCFGGVQGGVGG.VHGGGLNG...FGGFGGGFGGPGGGFGLQGG	89

G.pulex EH268909GIGSSNCRYWCKTPEGQAYCCENNOQAPKHPVAALIQKE	111
L. vannamei AY488494GIGISDCRYWCKTPEGQAYCCES.AHEPETPVG...TKP	108
L. setiferus AF430077APPATCRRWCKTPEQAYCCET.IFEPEAPVG...TKP	70
L. schmitti EF182748SHGISDCRYWCKTPEGQAYCCES.AHEPETPVG...TKL	111
Fe. chinensis AY871268APPATCRRWCKTPEQAYCCET.SFEPEAPVG...TKI	74
Fa. subtilis EF450744NGPISDCRYWCKTPEGQAYCCES.AHEPETPVG...TKP	91
Pe. monodon CD766060APPASCRRWCKTPEAFYCCES.RYEPEAPVG...TKI	88
M. japonicus AB121740	GFGGLQGGGLGSSGSSDCRYWCKTPEGQAYCCER.SHEPEGPVG...TKP	135

G.pulex EH268909	GECPVRAECPRLRFGGP..PQICSKHGDGYGTERCCFDRCLNEHVCKEP	159
L. vannamei AY488494	LDCPQVRPTCP..RFHGPP..PITICSNDYKCAGLDKCCFDRCLGEHVCKEP	154
L. setiferus AF430077	LDCPQVRPTCP..RFHGPP..VITCSSDYKCGGVDKCCFDRCLGEHVCKEP	116
L. schmitti EF182748	LDCPQVRPTCP..RFHGPP..PITICSNDYKCAGLDKCCFDRCLGEHVCKEP	157
Fe. chinensis AY871268	LDCPVRDTCPPVRFGGLA.PVITCSSDYKCGGIDKCCFDRCLGEHVCKEP	123
Fa. subtilis EF450744	LDCPQVRPTCP..RFSGP..PITICSNDYKCAGLDKCCFDRCLGEHVCKEP	137
Pe. monodon CD766060	LDCPKVRDTCPPVRFVAVEQVPCSSDYKCGGLDKCCFDRCLGQHVCKEP	138
M. japonicus AB121740	LDCPQVRPTCP..RFQGGG.PVITCSNDYKCAGLDKCCFDITCLQEHVCKEP	182

G.pulex EH268909	.LYG.....	162
L. vannamei AY488494	SFFGSOVFG.....	163
L. setiferus AF430077	SFYSEPTKP.....	126
L. schmitti EF182748	SLFGQQ.....	163
Fe. chinensis AY871268	SFYNFENHV.....	132
Fa. subtilis EF450744	SFFGKPLFGLG.....	148
Pe. monodon CD766060	SFYEFFA.....	145
M. japonicus AB121740	SVFGKPLFGI.....	192

Figure 5a

A. franciscana Q1W1H4	C.....C.KN..GCT...CAPNCRCAQDCKCCCKGCE..C.KSNPFC	32
G. pulex EH273178	CP...PVRAEC..PLRIFGGPPQICSKHGDCYGTIERCCF..DR..CLNEHV..C	43
C. finmarchicus EL585729	C...TSDSD..CNDSDKC..PGGICNIDGICSYPDCCCKSDSDCDHFDGMC	44
H. gammarus AJ786653	C...AKHVFCAARGDFTPPAIPRCGHDDYCPYHEKCCY..DA..CLKHHT..C	43
L. vannamei AF430071	CP...QVRPTC...PRFHGPPITCSNDYKCAGLDKCCF..DR..CLGEHV..C	41
L. vannamei SWD AY465833	CPDTSDVIGPC...VITERN...CLSDSCAPGQKCCPIG..CGRE...C	39

Figure 5b

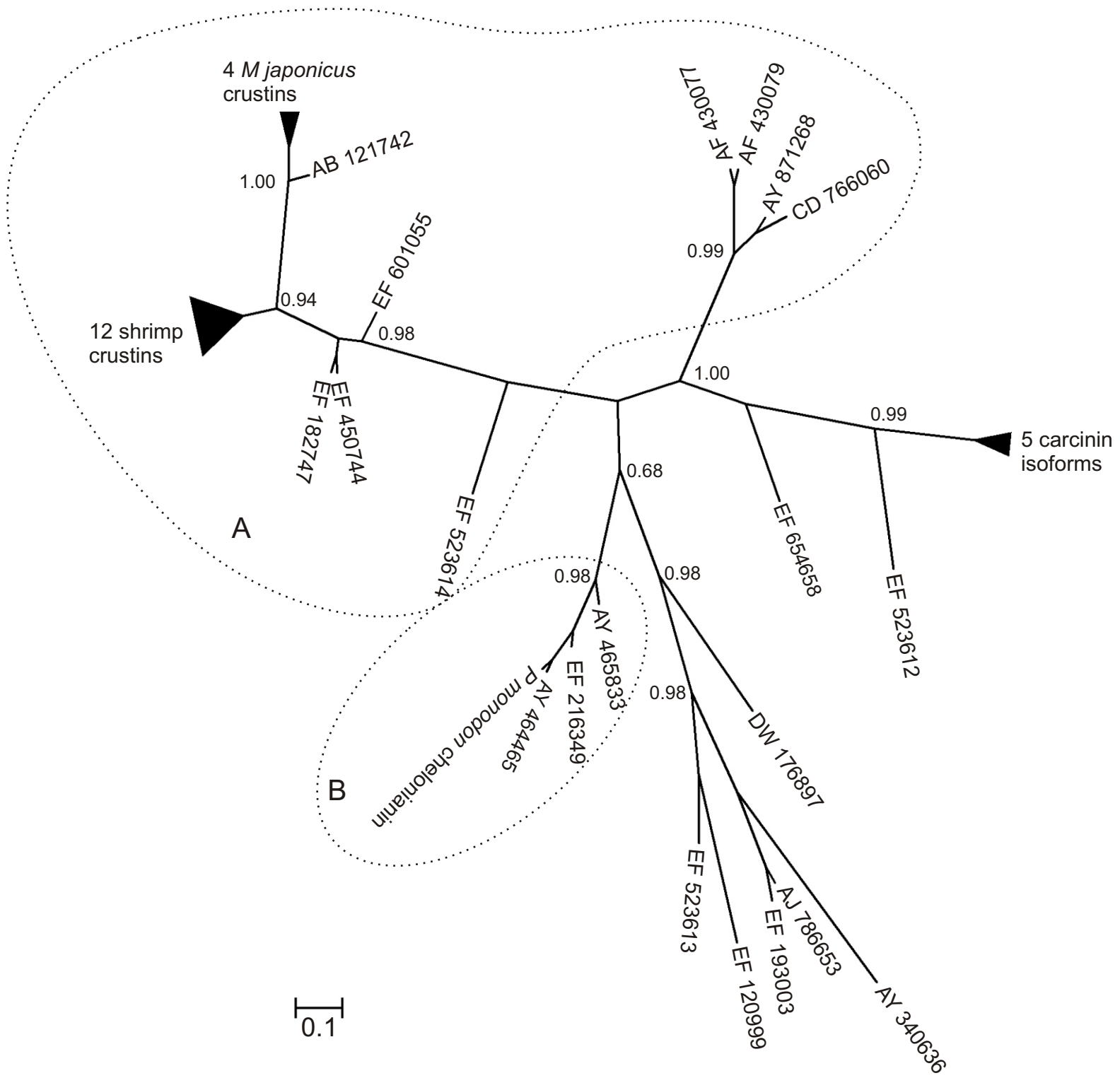


Figure 6

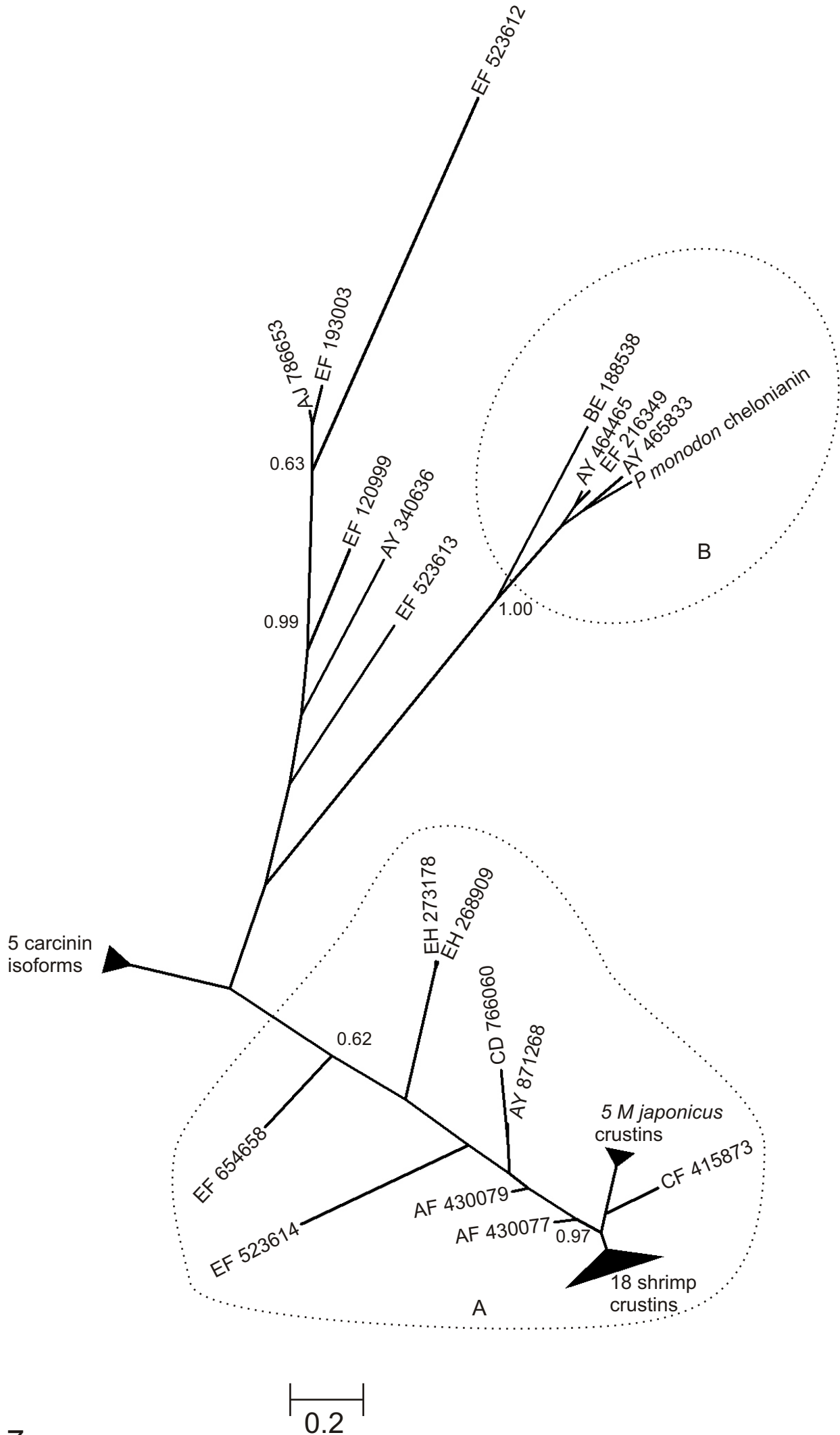


Figure 7