Molecular characterization and phylogenetic analysis of an anti-lipopolysaccharide factor from the crucifix crab, Charybdis feriatus

Abstract: Antimicrobial peptides are small cationic, gene-encoded, amphipathic, host defense peptides with a ubiquitous distribution in all living kingdoms. They are <10 kDa in size, with 15–100 amino acids having a net positive charge of +2 to +9. Anti-lipopolysaccharide factor (ALF) is a cationic antimicrobial peptide which constitutes one of the key effector molecules in the innate immune system of crustaceans, and is capable of binding and neutralizing lipopolysaccharides. In the present study, an ALF homolog (Charybdis feriatus [Cf]-ALF1)-encoding cDNA sequence from the hemocytes of the crucifix crab, C. feriatus, was cloned, identified, and characterized. The deduced peptide of CF-ALF1 encoded for a 123 amino acid peptide with a 97 residue mature peptide (11.16 kDa) that had a net charge of +10. Two conserved cysteine residues and a putative lipopolysaccharide binding domain were observed in the CF-ALF1 mature peptide. BLAST analysis of CF-ALF1 nucleotides showed a 99% similarity to Scylla serrata. The spatial structure of CF-ALF1 was composed of three α-helices packed against a four-strand β-sheet. Two of these helices were linked by a disulfide bond to form an amphipathic loop similar to the structure of anti-lipopolysaccharide factor isoform 3 from Penaeus monodon (ALF-Pm3). All these features suggest that CF-ALF1 could play a significant role in the innate immune defense mechanism of C. feriatus.

Keywords: antimicrobial peptides, crab, innate immunity, Charybdis feriatus, anti-lipopolysaccharide factor

Introduction
Crustaceans represent one of the most abundant groups of animals, inhabiting both aquatic and terrestrial areas. By virtue of their diversity and abundance, they have been considered as a potential source of bioactive compounds. The lack of an adaptive immune system in crustaceans has led to the evolution of potent compounds, including antimicrobial peptides (AMPs) that are used as defense against pathogens. Anti-lipopolysaccharide factor (ALF) represents one of the evolutionarily-conserved cationic AMPs distributed among horseshoe crabs and crustaceans which can bind against lipopolysaccharides (LPS) by neutralizing the lipid-A moiety of LPS.1

The first ALF was isolated from the amebocytes of two kinds of horseshoe crab, Tachypleus tridentatus and Limulus polyphemus.2 Since then, numerous ALF homologs have been identified and characterized from decapod crustaceans, including penaeid shrimp,3–6 nonpenaeid shrimp,7 lobsters,8,9 crabs, and crayfish.10–13 ALFs belong to the group of single-domain AMPs with a signal peptide at the N-terminal region followed by a conserved LPS-binding domain.14 Different ALF isoforms and variants are either encoded by distinct genes or formed as a consequence of alternative mRNA splicing.15

Correspondence: Rosamma Philip
Department of Marine Biology, Microbiology and Biochemistry, School of Marine Sciences, Cochin University of Science and Technology, Fine Arts Avenue, Kochi – 682 016, Kerala, India
Tel +91 484 236 8120
Fax +91 484 238 1120
Email rosammap@gmail.com
Functional characterization of ALFs has been conducted by chemically synthesizing the LPS domain region, 16-22 and by recombinant expression of mature peptides.9,13,23-28 Biological activity reported so far for ALFs include antibacterial, 6,9,20 antifungal,6,27 antiviral,29,30 immunomodulatory,31 antitumor,32 anti-inflammatory, and antiprotozoal activity.32,33 As per the previous studies, the structure of ALF was elucidated using nuclear magnetic resonance (NMR) spectroscopy, with the finding that it was composed of three α-helices packed against a four-stranded β-sheet.34,35

Thus far, there are several studies concerning the molecular and phylogenetic characterization, gene expression analysis, and structural and functional analysis of ALFs from decapod crustaceans and chelicerates. However, there is little research on AMPs from the crucifix crab, Charybdis feriatus. C. feriatus, a member of the family Portunidae, is one of the important species of commercial crab sold in markets across East Asia, where it commands substantially higher premium prices than Portunus spp. (typically US$8–$15 per kg). Also is it a promising aquaculture species because of its meat quality, taste and size. Thus far, there are several studies concerning the molecular and phylogenetic characterization, gene expression analysis, and structural and functional analysis of ALFs from decapod crustaceans and chelicerates. However, there is little research on AMPs from the crucifix crab, Charybdis feriatus. C. feriatus, a member of the family Portunidae, is one of the important species of commercial crab sold in markets across East Asia, where it commands substantially higher premium prices than Portunus spp. (typically US$8–$15 per kg). Also is it a promising aquaculture species because of its meat quality, taste and size.

The present study focused on identifying an antimicrobial peptide from the crucifix crab, C. feriatus.

Materials and methods

Animals and sample collection

Live and healthy adult crucifix crabs, C. feriatus, were collected from Cochin estuary along Vypeen, Kerala, India. To ensure survival of the crabs during their transport to the laboratory, adequate ventilation was provided. The hemolymph was extracted from the base of the crabs’ abdominal appendages using DEPC-treated, RNase-free capillary tubes that had been rinsed with precooled anticoagulant solution (RNase-free 10% sodium citrate, pH 7.0). The hemolymph was then homogenized in TRI reagent (Sigma-Aldrich Co, St Louis, MO, USA) and kept at −20°C until ready for total RNA isolation.

Total RNA isolation and first-strand cDNA synthesis

Total RNA was isolated from hemocytes using TRI reagent, as per the manufacturer’s instructions. Qualitative analysis of RNA was performed using agarose gel electrophoresis with a 0.8% agarose gel, as well as calculating the ratio of absorbance at 260 nm and 280 nm. RNA was quantified by spectrophotometry at 260 nm (Hitachi U-2900, Hitachi Ltd, Tokyo, Japan), and RNA samples with absorbance ratios (A260:A280) >1.8 were used for the present work. First-strand cDNA was generated in a 20 µL reaction volume containing 5 µg total RNA, 1× RT buffer, 2 mM dNTP, 2 mM oligo dT20, 20 U of RNase inhibitor, and 100 U of MMLV reverse transcriptase (New England Biolabs, Ipswich, MA, USA). The reaction was conducted at 42°C for 1 hour, followed by an inactivation step at 85°C for 15 minutes. Gene-specific primers β-actin-F (5′-CTTGTG-GTTGACAATGGCTCCG-3′) and β-actin-R (5′-TGGTT-GAAGGAGTGCCACGCTC-3′) were used to amplify the housekeeping gene β-actin (520 bp), which was used as an internal control to verify the reverse transcription polymerase chain reaction (RT-PCR).

PCR amplification and TA-Cloning

PCR amplification of 1 mL of cDNA was performed in a 25 µL reaction volume containing 1× standard Taq buffer (10 mM Tris HCl, 50 mM KCl, pH 8.3), 3.5 mM MgCl2, 200 mM dNTPs, 0.4 mM each primer, and 1 U Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA). Amplification was performed using the primers, ALF-F (5′-GGACAGAAGAAACATTGAGGACGACGCAGCA-3′) and ALF-R (5′-GGAAATCAAAAACATCCATTACAGGTCA-3′).11 The thermal profile used was 95°C for 2 minutes followed by 35 cycles of 94°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds; final extension was at 72°C for 10 min. The PCR product was analyzed by electrophoresis in 1% agarose gels in TBE buffer, stained with ethidium bromide, and visualized using a gel documentation unit (Syngene, Synoptics Ltd, Cambridge, UK).

The purified PCR product was ligated into the pGEM®-T Easy Vector and transformed using DH5α-competent Escherichia coli cells, as per the manufacturer’s protocol (pGEM-T Easy TA Cloning Kit, Promega Corporation, Fitchburg, WI, USA). Transformed bacteria were cultured in Luria Bertani agar containing ampicillin, IPTG, and X-gal, at 37°C for 16 hours, and the recombinant clones with the inserts were selected by blue-white screening. Recombinant white colonies were selected and streaked onto fresh ampicillin plates, and screened by colony PCR using vector-specific primers (T7 F and Sp6 R) and the ALF sequence-specific primers. For vector-specific primers, the thermal profile used was 95°C for 3 minutes followed by 35 cycles of 94°C for 15 seconds, 57°C for 30 seconds, and 72°C for 30 seconds; final extension was at 72°C for 10 minutes. Amplicons were examined on 1% agarose gels in TBE buffer, stained with ethidium bromide, and visualized using a gel documentation unit. The potentially positive recombinant clones were selected for plasmid isolation as per the manufacturer’s protocol (Genelute Hp Plasmid Miniprep Kit Sigma-Aldrich Co, St Louis, MO, USA). Isolated plasmids were analyzed on a 0.8% agarose gel, and gene insertion was confirmed by PCR.
using vector-specific and insert-specific primers. Selected recombinant plasmids were sequenced using T7 F and Sp6 R primers with the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit on an ABI PRISM 377 DNA Sequencer (Thermo Fisher Scientific, Waltham, MA, USA) at SciGenom Sequencing Facility, India.

Characterization of nucleotide and amino acid sequences
GeneTool software (Syngene, Synoptics Ltd, Cambridge, UK) was used for the nucleotide sequence analysis and assembly. The cDNA sequences acquired were translated using the Expert Protein Analysis System (ExPASy) (http://web.expasy.org/tools/translate/) to obtain their amino acid sequences. Homology searches of nucleotide and deduced amino acid sequences were accomplished using the BLASTn and BLASTp algorithms of the NCBI (http://www.ncbi.nlm.nih.gov/blast). Presumed signal peptide region was identified using the online program SignalP (http://www.cbs.dtu.dk/services/SignalP/). The transmembrane region within the mature peptide was predicted using the TMPred tool of ExPASy (http://embnet.vital-it.ch/software/TMPRED_form.html). Physicochemical parameters of the protein sequence were predicted using the online protein analysis tool, ExPASy ProtParam (http://web.expasy.org/protparam).

Phylogenetic analysis of C. feriatus ALF1
Multiple sequence alignment of C. feriatus (Cf)-ALF1’s amino acid sequence was performed on ClustalW in BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html) for the comparison and alignment with other crustacean ALF-encoding sequences. The phylogenetic tree was constructed based on the neighbor-joining method, using Molecular Evolutionary Genetics Analysis (http://www.megasoftware.net/) version 6.

Structure analysis of Cf-ALF1
The spatial structure of MnALF was established with PyMOL (https://www.pymol.org/), using the SWISS-MODE (http://swissmodel.expasy.org/) prediction algorithm, which is based on homology modeling. The helical property of Cf-ALF1 was predicted using HeliQuest (http://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParamsV2.py). The cDNA sequence of Cf-ALF1 was converted to the corresponding RNA sequence (Attotron Biosensor Corporation, http://www.attotron.com/) and submitted to RNAStructure WebServer (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html) to envisage the RNA structure with minimum free energy.

Prediction of Cf-ALF1 antimicrobial peptide properties
The antimicrobial activity of Cf-ALF1 was predicted using the Antimicrobial Peptide Database’s (APD2) calculation and prediction tool (http://aps.unmc.edu/AP/main.php). The amphipathicity and antimicrobial activity of the LPS domain was interpreted using HeliQuest.

Results and discussion
In the present study, an AMP belonging to the ALF family (herein referred to as Cf-ALF1) was identified and characterized from the hemocytes of C. feriatus. A 524 bp fragment of cDNA encoding an ORF of 123 amino acids was obtained from the hemocytes of C. feriatus (Figure 1). The Cf-ALF1 nucleotide and deduced amino acid sequences were deposited in the GenBank sequence database (GenBank ID: KP688577). BLAST analysis of the 372 bp of cDNA sequence, as well as the homology search by BLASTn and BLASTp of the deduced amino acid sequence of Cf-ALF1, revealed that it belongs to the DUF3254 superfamily. This superfamily includes all the family of ALFs. BLASTn analysis of the Cf-ALF1 nucleotide sequence showed that it exhibited a 99% similarity to an ALF isoform from Scylla serrata (SsALF) (GenBank ID: ACH87655.1), a 97% similarity to an S. paramamosain ALF (GenBank ID: EF207786.1), a 96% similarity to a P. pelagicus ALF (GenBank ID: JQ899452.1), an 87% similarity to a P. trituberculatus ALF (GenBank ID: GQ165621.2), and a 72% similarity to an Eriocheir sinensis ALF (GenBank ID: HQ850572.1). SignalP analysis of the peptides determined the existence of a signal peptide composed of 26 amino acids and containing an exceedingly hydrophobic core region. The presence of a putative cleavage site after the 26th amino acid (CEA Y) demarcated the highly cationic mature peptide of 97 amino acid residues.

Based on the ExPASy ProtParam analyses, Cf-ALF1 had an estimated molecular weight of 14.025 kDa, a net charge of +11, and a theoretical isoelectric point of 9.78. Comparatively, the mature peptide region measured 11.172 kDa, +10, and 10.01, respectively. Cationicity of the mature peptide was mainly attributed to the 18 positive amino acid residues (twelve lysine residues and six arginine residues), against eight negative residues (four aspartic acid residues and four glutamic acid residues). The estimated half-life of the mature peptide of Cf-ALF1 was found to be 0.8 hours (mammalian reticulocytes, in vitro),
The predicted instability index was computed to be 34.37, and thus classified the protein as stable. The presence of a transmembrane helical region was determined in the mature peptide region, starting from L_{43}SGL WHSDTVDFMGHT_{59}, which may assist in the peptide and bacterial membrane interaction.

The multiple protein sequence alignment of Cf-ALF1 (with representatives of each crustacean’s ALFs) revealed the existence of conserved sequence features (Figure 2), ie, cysteine residues (C_{55}–C_{76}) that established a disulfide bridge with a highly conserved sequence. This sequence contained a cluster of seven positively charged amino acid residues within the disulfide loop, forming the functional LPS domain.

![Figure 1](image1.png)

**Figure 1** Nucleotide and deduced amino acid sequence of the ALF isoform from the hemocyte mRNA transcripts of Charybdis feriatus-ALF1 (GenBank ID: KP688577).

**Notes:** Gray-colored highlighted portion specifies the 25 amino acid signal peptide. The bioactive mature peptide is highlighted in turquoise, and the underlined region within the mature peptide is the LPS binding domain.

**Abbreviations:** ALF, anti-lipopolysaccharide factor; LPS, lipopolysaccharides.

![Figure 2](image2.png)

**Figure 2** Multiple alignment of amino acid sequence of the Charybdis feriatus (Cf)-ALF1 (GenBank ID: KP688577) with other crustacean and limulid ALFs obtained using BioEdit software: Macrobrachium rosenbergii ALF (Mr) (GenBank ID: AE84102.1), Fenneropenaeus indicus ALF (Fi) (GenBank ID: AEP84102.1), Homarus americanus ALF (Hm) (GenBank ID: ACH87635.1), Procambarus clarkii ALF (Pc) (GenBank ID: ACH87655.1), and Tachypleus tridentatus (Tt) (GenBank ID: AAK00651.1).

**Notes:** The LPS binding domains are enclosed within a square. The conserved residues are highlighted with uniform background colors.

**Abbreviations:** ALF, anti-lipopolysaccharide factor; LPS, lipopolysaccharides.
An anti-lipopolysaccharide factor from the crucifix crab

it differs from SsALF by only one amino acid residue in the signal peptide, whereas the mature peptide region was found to be highly conserved.

The level of antimicrobial activity of Cf-ALF1 predicted by Antimicrobial Peptide Database 2 revealed Cf-ALF1 as a potent AMP with a protein-binding potential (Boman index) and Wimley–White whole-residue hydrophobicity of 2.1 kcal/mol and 15.8 kcal/mol, respectively. In silico analysis of the anticipated LPS binding domain of Cf-ALF1 (C<sub>55</sub>HRRKPFRKFKLYHEGKFWC<sub>70</sub>) based on amphipathicity, using HeliQuest, showed that Cf-ALF1 is composed of 22 amino acids. Polar residues comprised 45.45% of these amino acids, with the remaining amino acids being nonpolar residues, thus establishing a hydrophobic face with IWLP residues and a hydrophobicity of 0.342 H with a +7 net charge. Such properties could assist in the binding of the LPS domain with the highly negatively-charged lipid A moiety of LPS, and therefore confirm the antimicrobial activity (Figure 3).

Recombinant expression of SsALF has previously been carried out using a bacterial expression system, and it showed a stronger inhibition of Gram-negative bacterial growth than that of Gram-positive bacteria. Using homology modeling with the template recombinant-ALF-Pm3 (SMTL id: 2job.1) NMR structure (chosen for its shared similarity of 38.61%), prediction of the spatial 3D structure of Cf-ALF1 demonstrated the presence of three α-helices packed against four strands of β-sheets (Figure 4). Somboonwiwat et al studied and elucidated the functional role of recombinant-ALF-Pm3 expressed in the Pichia pastoris system, and demonstrated a comprehensive spectrum of activity against the filamentous fungi, and both Gram-positive and Gram-negative bacteria. In the current study, the predicted RNA sequence based on the Cf-ALF1 cDNA sequence, as well as its ideal secondary structure, was used to identify the amphipathicity of the LPS binding domain. The Amino and Carboxy terminal ends are mentioned as N and C, respectively. The expected hydrophobic face IWLP is shown in the red circle.

Abbreviations: ALF, anti-lipopolysaccharide factor; LPS, lipopolysaccharides.

[Figure 3: Helical wheel diagram of Charybdis feriatus-ALF1 (GenBank ID: KP688577) LPS domain predicted using the HeliQuest online tool. Notes: The structure was built to identify the amphipathicity of the LPS binding domain. The Amino and Carboxy terminal ends are mentioned as N and C, respectively. The expected hydrophobic face IWLP is shown in the red circle.]

[Figure 4: Spatial structure of Charybdis feriatus-ALF1 (GenBank ID: KP688577). Notes: Structure created with the PyMOL software, using the pdb data generated by SWISS-MODEL server. The NMR structure of rALF-Pm3 (SMTL id: 2job.1) was used as template for the data generation. The diagrammatic representation (A) and the spatial structure (B) are presented. Abbreviations: ALF, anti-lipopolysaccharide factor; NMR, nuclear magnetic resonance; rALF-Pm3, recombinant-ALF-Pm3.]
structure with minimum free energy, was found to consist of both double-stranded and looped regions (Figure 5).

Analysis of the phylogenetic tree constructed using the neighbor-joining method clearly showed that ALF sequences were clustered according to species (Figure 6).

The phylogenetic tree revealed that ALFs including Cf-ALF1 and SsALF shared a common ancestral origin and could have subsequently diverged in the course of evolution. The ALF family of AMPs were found to be evolutionarily conserved, with a diversity of ALF isoforms exhibiting the same
physicochemical properties that formed one of the main AMPs involved in the removal of pathogenic microorganisms. The phylogenetic tree could be divided into three groups: Group I consisted of shrimp ALFs, including both penaeid and nonpenaeid shrimp ALFs; Group II comprised all crustacean ALFs; and Group III included crab ALFs. Cf-ALF1 was grouped along with SsALF and other portunid crab ALFs in Group III.

Conclusion
The present study reports the identification, cloning, molecular, and phylogenetic characterization of an isoform of ALF from the crucifix crab, *C. feriatus*, exhibiting an 100% identity to SsALF. The occurrence of sequence variation mainly in the mature peptide could be an important factor that has led to the functional divergence of ALF isoforms from limulids to decapod crustaceans. *C. feriatus* is now an emerging species in the aquaculture market, due to the lower quantities in wild catch and the requirement to meet the demands of the market. Also, *C. feriatus* has been suggested as one of the appropriate species for stock enhancement and aquaculture in the international forum on the culture of portunid crabs.  

To date, there is little research regarding the physiological processes involved in the immune system or the host defense peptides from *C. feriatus*. As there are currently restrictions in antibiotic usage due to the emergence of multidrug resistant strains, the present study is of great significance, particularly in the aquaculture system because *C. feriatus* has shown its ability to combat pathogenic microorganisms. Since recombinant expression of SsALF has already been carried out using a bacterial expression system, further work with Cf-ALF1 could include heterologous expression with post-translational modification, using a eukaryotic system such as Baculovirus-mediated insect cell line expression, and characterization of biological activity. These approaches would shed light on the possibilities of the Cf-ALF1 peptide as a therapeutic candidate in aquaculture and medicine. Bioprospecting for AMPs and other host defense molecules from potential aquaculture species would contribute to the understanding of crab immunology, and their potential role in prevention of microbial diseases in wild and domestic animals.

Acknowledgments
The authors are grateful to the Director of the Centre for Marine Living Resources and Ecology (CMLRE), and the Ministry of Earth Sciences (MoES), Govt. of India for the research grant (MoES/CMLRE/10-MLR-TD/12/2012) and scientific support for the work. Authors would also like to thank the Coordinator of the National Centre for Aquatic and Animal Health (NCAAH) for scientific support. KS Sruthy gratefully acknowledges Kerala State Council for Science, Technology and Environment (KSCSTE) for the Award of Fellowship.

Disclosure
The authors report no conflicts of interest in this work.

References


