

Cloning and molecular modeling of *Litopenaeus vannamei* (Penaeidae) C-type lectin homologs with mutated mannose binding domain-2

F.H.F. Costa¹, N.S.M.S. Valença², A.R.B.P. Silva³, G.A. Bezerra^{4,5},
B.S. Cavada⁵ and G. Rádis-Baptista^{2,6}

¹Departamento de Engenharia de Pesca, Universidade Federal do Ceará, Fortaleza, CE, Brasil

²Departamento de Bioquímica, Universidade Federal de Pernambuco, Recife, PE, Brasil

³Departamento de Imunogenética, Instituto Butantan, São Paulo, SP, Brasil

⁴Institute of Molecular Biosciences, Karl-Franzens University of Graz, Graz, Austria

⁵Laboratório de Moléculas Bioativas, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará, Fortaleza, CE, Brasil

⁶Instituto de Ciências do Mar, Universidade Federal do Ceará, Fortaleza, CE, Brasil

Corresponding author: G. Rádis-Baptista

E-mail: gandhi.radis@ufc.br

Genet. Mol. Res. 10 (2): 650-664 (2011)

Received August 3, 2010

Accepted November 12, 2010

Published April 19, 2011

DOI 10.4238/vol10-2gmr999

ABSTRACT. C-type lectins are animal proteins that contain at least one carbohydrate recognition domain (CRD) capable of mediating sugar and calcium binding. Carbohydrate recognition is directly required for some biological functions, including the innate immune response. We cloned two novel C-type lectin (CTL) precursors from the commercial marine shrimp *Litopenaeus vannamei*. The cloned cDNAs encompass ORFs of 1044 nucleotides and encode highly similar two-

domain polypeptides of 347 residues. The predicted proteins, LvCTL-br1 and -br2, contain the consensus triad that recognizes galactose (-GlnProAsp-) in CRD1 but also contain a mutated mannose-binding site (-GluProAsn-) in the second domain (CRD2). Phylogenetic analysis of LvCTL-br1 and -br2 and hundreds of CTL-like domain-containing proteins have allowed grouping of penaeid shrimp CTLs into three functional clusters. Reverse transcription coupled to PCR indicated that LvCTL-br1 expression is induced in shrimp gills upon IHHNV infection. Computational molecular modeling of LvCTL-br1 and -br2 revealed that three amino acid substitutions in CRD1 occur near the sugar binding site. Also, the 3-D models show a long loop of LvCTL-br1 CRD2 that might accommodate complex sugars. The structural data, evolutionary history and functional analysis support the hypothesis that gene duplication and accelerated evolution have caused functional diversification of penaeid shrimp C-type lectins.

Key words: C-type lectin-like domain; Invertebrate C-type lectin; Innate immune-related gene; Molecular phylogenetic analysis; Molecular modeling; Penaeid shrimp

INTRODUCTION

Lectins are proteins or glycoproteins, mostly without catalytic activity, with one or more conserved carbohydrate recognition domains (CRDs) capable of discriminating and reversely binding to mono- and oligosaccharides in solution or on the surface of cells. Distinct families of lectins are recognized based on the structures of their CRDs. For instance, L-type lectins are characterized by a β -sandwich, with more than 230 amino acid residues, which has the ability to bind various types of sugars and to sort proteins in the endoplasmic reticulum of animal cells (Dodd and Drickamer, 2001). C-type lectins encompass another significant group of modular proteins, in which the CRDs (>115 amino acids), with one or more mixed α/β -folds, compose the most diverse family of animal lectins. The CRDs of C-type lectins specifically recognize a variety of sugars in a Ca^{2+} -dependent basis, whereas other adjacent domains of the lectin molecule respond to distinct biological activities, such as cell adhesion, glycoprotein clearance and pathogen neutralization (Drickamer, 1993, 1995; Weis and Drickamer, 1996; Dodd and Drickamer, 2001).

A number of C-type lectins from vertebrates and invertebrates act as pattern recognition receptors (PRRs) by binding to pathogen-associated molecular patterns and activating the innate host defense systems. For instance, in vertebrates, transmembrane proteins that possess C-type lectin-like domains (CTLs), such as the mannose-binding receptor, DCL-1, dectin-1; dendritic cell-specific ICAM grabbing non-integrin receptor (DC-SIGNR), as well as soluble collectins that have CTLs in their structure, are all involved in phagocytosis (Kerrigan and Brown, 2009).

In invertebrates such as *Anopheles gambiae*, C-type lectins are also included among the various PRR proteins, and they contribute to the insect's defense against pathogens (Christophides et al., 2004; Schnitger et al., 2009). In silkworms, three homologs were found to be

expressed in the testis and ovary that act as primary PRRs in *Bombyx mori* cellular immunity (Takase et al., 2009). Additionally, C-type lectins with two CRDs arranged in tandem function as pattern recognition receptors in the cotton bollworm, promoting hemocyte phagocytosis of pathogens, protecting the insect from bacterial infection (Tian et al., 2009). In marine invertebrates, C-type lectins and CTLD-containing proteins have been identified in several species of crustaceans, including crabs (Kong et al., 2008) and shrimp (Luo et al., 2006; Liu et al., 2007), as well as in barnacles (Matsubara et al., 2007), cnidarians (Wood-Charlson and Weis, 2009) and mollusks (Yamaura et al., 2008; Zhang et al., 2009a).

The penaeid shrimp include species that are commercially cultivated, such as *Litopenaeus vannamei* and *Penaeus monodon*. Due to their worldwide production and their aquatic nature, these organisms suffer from viral and bacterial infections that are often fatal. Like all invertebrates, penaeid shrimp lack adaptive immunity, and thus rely on their innate immune system, partially consisting of CTLD-containing PRRs, for any defense against pathogens (Hofmann et al., 1999).

In penaeid shrimp, C-type lectins have been shown to function as antimicrobial and antiviral proteins (Sun et al., 2008b; Zhao et al., 2009), proteins of pattern recognition (Ji et al., 2009; Zhang, 2009b), proteins of pathogen clearance (Wang et al., 2009b), and they have been shown to participate in the innate immune response (Ma et al., 2008; Wang et al., 2009a). It has also been documented that in the hemolymph and in other tissues of penaeid shrimp, the expression of C-type lectins is regulated at the transcriptional and translational levels after bacterial and viral infection (Leu et al., 2007; Ma et al., 2007; Soonthornchai et al., 2010).

The cloned and purified penaeid shrimp C-type lectins fall into two structurally related classes of proteins: those containing a single CTLD with a spacer region between the signal peptide (Ma et al., 2007; Sun et al., 2008a; Zhang et al., 2009b), and those with two CRDs disposed in tandem with spacer regions between the signal peptide and the CRD1 that intervenes between CRD1 and CRD2 (Ma et al., 2008; Zhang et al., 2009b). These shrimp C-type lectins were shown to recognize and bind with high affinity to galactose, mannose, N-acetyl glucosamine, N-acetyl galactosamine, and N-acetyl mannosamine, as well as to disaccharides, such as maltose and trehalose, to some degree (Rittidach et al., 2007; Ma et al., 2008; Zhang et al., 2009c). Although all C-type lectin CRDs self-recognize and bind specifically to carbohydrates in a Ca^{2+} -dependent fashion, the global architecture of a given C-type lectin results from the combination of conjugated domains, and this modular disposition of CTLD-containing proteins determines the diversity of lectin activities (Drickamer, 1995).

We cloned, made a phylogenetic analysis and molecularly modeled two divergent *L. vannamei* C-type lectin cDNA precursors that possess a mutated mannose carbohydrate recognition domain-2.

MATERIAL AND METHODS

Preparation of mRNAs from the *L. vannamei* hepatopancreas and first-strand cDNA synthesis

Most of the experimental procedures using recombinant DNA technology were performed according to current protocols in Molecular Biology (Ausubel et al., 1995). Frozen hepatopancreas of 10 *L. vannamei* juveniles (not exposed to microbial or viral infection) were finely

powdered in a mortar and pestle under liquid nitrogen, and total RNA was purified using the Trizol reagent (Invitrogen, Carlsbad, CA, USA), according to manufacturer instructions. The quality and yield of total RNA were verified by analyzing the integrity of 28S and 18S rRNA using denaturing agarose gel electrophoresis and spectrophotometrically using the 260/280 nm ratio.

Poly(A⁺)-RNA was purified from total RNA using a complex of oligo(dT)-biotin and streptavidin-MagneSphere paramagnetic particles (PolyAtract System, Promega, Madison, WI, USA). The mRNAs were quantified and used for cDNA synthesis. ImProm-II reverse transcriptase (Promega) was used to prepare the full-length cDNA, as follows: mRNA (1.2 µg) from *L. vannamei* hepatopancreas was mixed with 500 ng oligo-dT primer (Promega), in a final volume of 14 µL, heated to 70°C for 10 min to disrupt possible secondary structures, and quickly cooled on ice. Thereafter, a total of 100 U reverse transcriptase was combined with 1 mM deoxynucleoside triphosphate (dNTPs), 40 U RNase inhibitor, and RNase-free water to make a final reaction volume of 25 µL. The reverse transcription was conducted at 42°C for 60 min and finally heated to 70°C for 15 min.

Molecular cloning of hepatopancreatic *L. vannamei* C-type lectin homologs

For molecular cloning and structural studies of hepatopancreatic *L. vannamei* C-type lectin, gene specific primers were synthesized based on previously reported LvLT sequences (Ma et al., 2007), and used for high-fidelity polymerase chain reaction (PCR) amplification, as follows: each reaction, in a final volume of 50 µL, consisted of 2.5 U Platinum Taq DNA polymerase, 2.0 mM of MgCl₂, dNTPs, and 0.2 µM of each forward (5'-CGAGTTACCTGGAATC GAACCA-3') and reverse primer (5'-TCATCCAGTATAGACACACAGT-3'), in high-fidelity PCR buffer (60 mM Tris-SO₄, pH 8.4, 18 mM ammonium sulfate, 2.5 mM MgSO₄). The PCR conditions included an initial denaturation step at 95°C for 5 min, followed by 32 cycles of denaturation at 95°C for 1 min, annealing at 58°C for 1 min, and extension at 72°C for 1 min. A final step at 72°C for 8 min was performed to terminate DNA synthesis. The efficiency of gene amplification was confirmed by agarose gel electrophoresis, after staining with ethidium bromide.

The cloned hepatopancreatic *L. vannamei* C-type lectin cDNA precursors were sequenced by the dideoxy chain termination method, using Dye terminator chemistry (DYEnamic ET Dye Terminator kit, GE Healthcare, Piscataway, NJ, USA) and the MegaBACE 750 DNA Analysis System (GE Healthcare). Each clone was sequenced twice with both sense and anti-sense 5'- and 3'-flanking primers and the contig sequence was assembled with the SeqMan software (Lasergene, DNASTar Inc., Madison, WI, USA). The annotated nucleotide sequences were submitted to GenBank (<http://www.ncbi.nlm.nih.gov/>), and they received the accession Nos. GU206551 (LvCTL-br1, in which '-br' indicates Brazil) and GU206552 (LvCTL-br2).

Molecular evolutionary analysis and sequence phylogeny

The *L. vannamei* C-type lectin cDNAs were *in silico* translated, and the predicted amino acid sequences were compared to all entries in the GenBank database using the BLASTp program (Altschul et al., 1997). One hundred and fifteen C-type lectin-like domain-containing protein predicted sequences (CTLDcps) were selected to study the evolutionary relationships. Sequence alignment, phylogenetic and molecular evolutionary analyses were conducted using

SeaView 4.2 (Gouy et al., 2010). The sequences were multi-aligned against a profile of pre-aligned sequences of *L. vannamei* shrimp CTL-br1, CTL-br2 and the two separated CRDs that compose the two-domain CTL-Br, using the algorithm MUSCLE 3.6 (Edgar, 2004). The cladogram tree was reconstructed using the BioNJ tree-building algorithm on the deduced amino acid sequences (Gascuel, 1997), and the reliability and statistical support of the clade were tested by bootstrap re-sampling of 100 pseudo-replicates. Comparison of the predicted domains, repeats, motifs, and features of lectins from different animals was performed on the SMART (Sample Modular Architecture Research Tool) web site (<http://smart.embl-heidelberg.de/>).

Expression analysis of LvCTL-br1 in shrimp gills upon IHHNV infection

Litopenaeus vannamei suffering from infectious by hypodermal and hematopoietic necrosis (IHHN) disease were obtained from a local farm. The infection caused by the IHHN virus (IHHNV) was confirmed by PCR as recommended by the World Organization for Animal Health (OIE), according to Nunan et al. (2001). Gills of shrimp at stage 2 and 3 of infection, as diagnosed by gross signal of disease and PCR analysis, were surgically removed and processed for total RNA purification by means of Wizard SV total RNA purification systems (Promega), according to manufacturer instructions. For cDNA synthesis, 1 µg of each DNase I treated-total RNA purified from a shrimp gill was mixed with 500 ng random primers (Promega), in a final volume of 10 µL, heated to 70°C for 10 min, to disrupt possible secondary structures, and quickly cooled on ice. Thereafter, 100 U ImProm-II reverse transcriptase (Promega) was combined with 1 mM of each dNTPs, 2 mM MgSO₄, 1 mM DTT, 20 U recombinant RNase inhibitor, and nuclease-free water to make a final volume of 20 µL. The reverse transcription was conducted at 42°C for 60 min and terminated by heating the reaction at 70°C for 15 min. The samples were diluted 10 times with TE (10 mM Tris-HCl, pH 8.5, 1 mM EDTA) and aliquots were used in a conventional PCR with specific primers for *L. vannamei* β-actin (AF300705.2) and for LvCTL-br1 CRD2 (CTLbr1_GTR-SE: 5'-ATCCAGGAACCCGATGGAGGA-3' and CTLbr1_CRD2-AS: 5'-TTATCCAGTATAGACACACAGTGGAT-3', corresponding to a nucleotide segment starting from the -GTR- moiety and ending at the carboxyl-terminus of CRD2). The RT-PCR products were separated by agarose gel electrophoresis and visualized by ethidium bromide staining.

Molecular modeling

The CTL-br1 3-D model was calculated by homology modeling with the Yasara program (Krieger et al., 2002). The template search was performed using built-in PSI-BLAST (Altschul et al., 1997). Approximately 20 hits were identified and obtained from the PDBfinder2_database (Hooft et al., 1996a). The quality of the retrieved structures was analyzed according to WHAT CHECK (Hooft et al., 1996b). Eighteen models were generated based on the top five ranked templates, and alternative alignments were used when necessary. After the side-chains were built, optimized and fine-tuned, the newly modeled parts were subjected to a combined steepest descent and simulated annealing minimization. Then, a full unrestrained simulated annealing minimization was performed for the entire molecule. Only the segments ranging from residues 37 to 195 (CRD1) and 198 to 347 (CRD2) were used for the 3-D structure calculation. The best CRD1 CTL-br1 model

(residues 37 to 195) was generated from human tetranectin, a plasminogen kringle 4-binding protein, sharing 25% identity with the covered region (PDB code: 1htn, X-ray solved at 2.8 Å resolution). The final model was evaluated and validated by the Yasara twinset.

The second domain of the CTL-br1 (CRD2, residues ranging from position 198 to 347) was separately modeled according to the same procedure described above. The best model was obtained from the dendritic cell-specific ICAM grabbing non-integrin (DC-SIGNR) in complex with GlcNAc2Man3. This template and CTL-br1 CRD2 share 32% identity with respect to the covered region (PDB code: 1k9J, X-ray solved at 1.9 Å resolution).

RESULTS AND DISCUSSION

Molecular cloning of C-type lectin cDNA precursors from the hepatopancreas of *L. vannamei*

Based on reversed transcription coupled to high-fidelity PCR, which was conducted with gene-specific oligonucleotides, a product of 1044 bp was amplified from the *L. vannamei* hepatopancreas. The amplicon was cloned into a plasmid vector and propagated in *Escherichia coli*. The automatic sequence analysis of 30 independent clones generated sets of nucleotide sequences that encompassed two open reading frames. Prediction of amino acid sequences revealed that both C-type lectin cDNA precursors encode two highly similar proteins, LvCTL-br1 and LvCTL-br2, with 347 amino acid residues. These proteins differ by only five residues at positions 29, 91, 109, 150, and 330, where the following substitutions were observed: Leu²⁹ → Ser²⁹, Ala⁹¹ → Thr⁹¹, Gly¹⁰⁹ → Asp¹⁰⁹, Ala¹⁵⁰ → Thr¹⁵⁰, Leu³³⁰ → Phe³³⁰. These substitutions were caused by the following transitions: cytosine → thymine and adenosine → guanosine (Figure 1). The estimated molecular masses and pI values of each cDNA precursor are 38.4 kDa and 4.93 for LvCTL-br1, and 38.5 kDa and 4.87 for LvCTL-br2, respectively.

Direct comparison of LvCTL-br1 and -br2 with their closest homolog, C-type lectin from *L. vannamei* (LvLT, GenBank DQ871245), showed a similarity of 81% based on amino acid identity. LvCTL-br1 and -br2 precursors, similar to LvLT, encompass a peptide signal (residues 1 to 20), two carbohydrate recognition domains (CRD1 and CRD2) disposed in tandem, four repetitive internal sequences (-EGVWV- and -GEA/VVPLGTPFW-, in CRD1, and -EGTWV- and -GEPVPMGTPFW-, in CRD2), and conservation of both position and number of cysteine residues (Figure 1). A relevant contrast between LvCTL-br1/-br2 and LvLT is present in the second CRD. The amino acid sequence in the first CRD1, namely -GlnProAsp- (-QPD-), which is known to bind galactose, is conserved in these sequences, but the tripeptide (-GlnProAsn-, -EPN-) into the CRD2 in LvLT, which is specific for mannose, was replaced by the triad -GlyThrArg- (-GTR-). These annotated LvCTL-br1 and -br2 precursors can be queried by the GenBank accession Nos. GU206551 and GU206552.

Lectins are modular proteins in which the CTDL can be connected to other CTDLs or to several kinds of domains and modules, resulting in distinct biological activity of the members of a given lectin family (Dood and Drickamer, 2001). Thus, it would be expected that the mutated mannose binding site detected in the second domain of LvCTL-br1 and -br2 could correspond to a quickly evolving C-type lectin homolog, in which the first domain could preserve its galactose binding property and the second domain could show an unrelated biological activity.

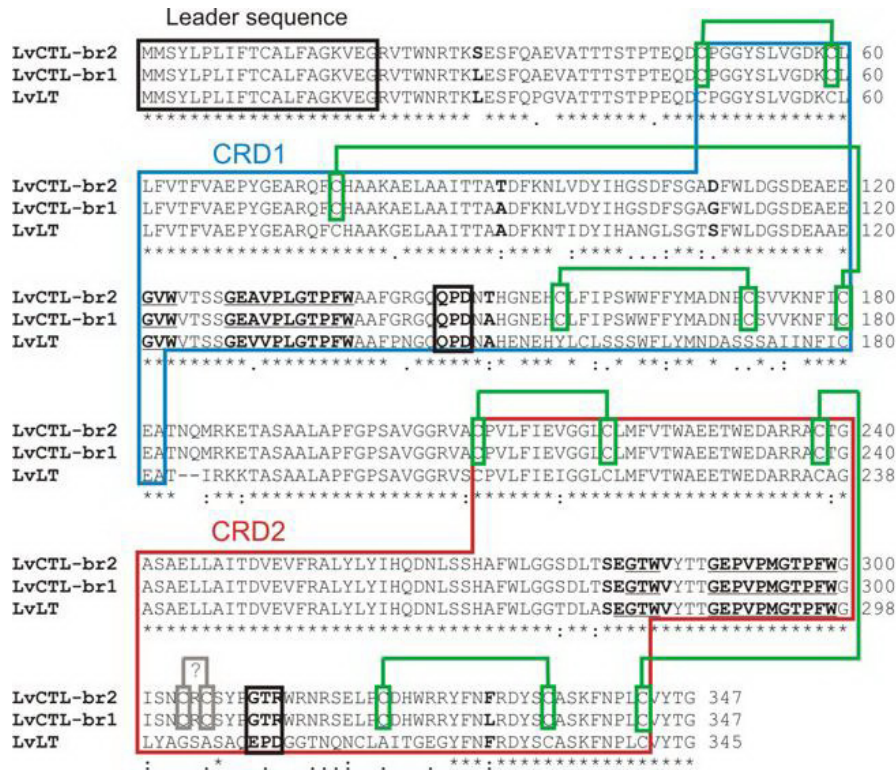


Figure 1. Comparison of the deduced amino acid sequences and major structural features of CTLbr-1, CTLbr-2 and CTL from the hepatopancreas of *Litopenaeus vannamei*. LvCTLbr-1, CTLbr-2 (this study) and LvCTL (Ma et al., 2007) are represented. The black open rectangle at the beginning of the sequences corresponds to the leader sequence. The carbohydrate recognition domain-1 and -2 (CRD1 and CRD2) are shown in the blue and red open boxes, respectively. The conserved S-S bonds (cysteine pairs 48-59, 77-180, 156-172, 209-220, 238-343, and 321-335) are highlighted in green. An extra pair of cysteine residues is shown in the light gray boxes. The repetitive stretches of amino acids are in bold and underlined. The peptide triads of both CRDs (-QPD- and -GTR-) are boxed in black. The similarity between LvCTLbr-1 and CTLbr-2 is 98%, and that between LvCTLbr-1 or CTLbr-2 and LvCTL is 81%.

Molecular evolutionary analysis of LvCTL-br1 and -br2

The shrimp C-type lectin members with which LvCTL-br1 and -br2 share the highest sequence similarity, 80 and 60%, respectively, are from the hepatopancreas of *L. vannamei* (Q009U0_LITVA), which grows in Asia, and from *Penaeus semisulcatus* (Q009U2_PENSE). However, as pointed out above, the crucial residues that create the carbohydrate binding motif are replaced by nonsynonymous substitutions in the CRD2 of LvCTL-br1 and -br2, as determined by sequence comparison.

A noticeable fact is that C-type lectins from several species of penaeid shrimp are either C-type lectins with a single-CRD or with two CRD domains arranged in tandem. Thus, a comparison of amino acid sequences of cloned or purified C-type lectin homologs from *Fenneropenaeus chinensis*, *F. merguensis*, *L. vannamei*, *P. monodon*, and *P. semisulcatus*

shows that single-domain penaeid shrimp CTLs align with the second domain of two CRD-containing CTLs. Thus, single-CRD CTL amino acid sequences align with the second CRD of two-domain CTLs, while the triad QPD - a characteristic of the CRD - is invariably present. Moreover, the triad EPD/N appears mainly in the CRD2 of two-domain CTLs, despite being present in one homolog of a single-domain CTL from *P. chinensis* (data not shown). In CRD2 of LvCTL-br1 and -br2, this carbohydrate binding site was replaced by the unusual triad -GTR-. The fact that single-domain CTLs with the -QPD- triad align with the CRD2 of two-domain shrimp CTLs might indicate that some part of the molecule is under evolutionary constraint, whereas other parts are relatively free to evolve.

To have a wider view of how these C-type lectins have evolved, a phylogenetic tree was constructed with 115 CTLDcps, from 10 phyla and sub-phyla, based on a multi-alignment against a profile of pre-aligned sequences of *L. vannamei* shrimp CTL-br1, CTL-br2 and their two individually separated CRDs (Figure 2).

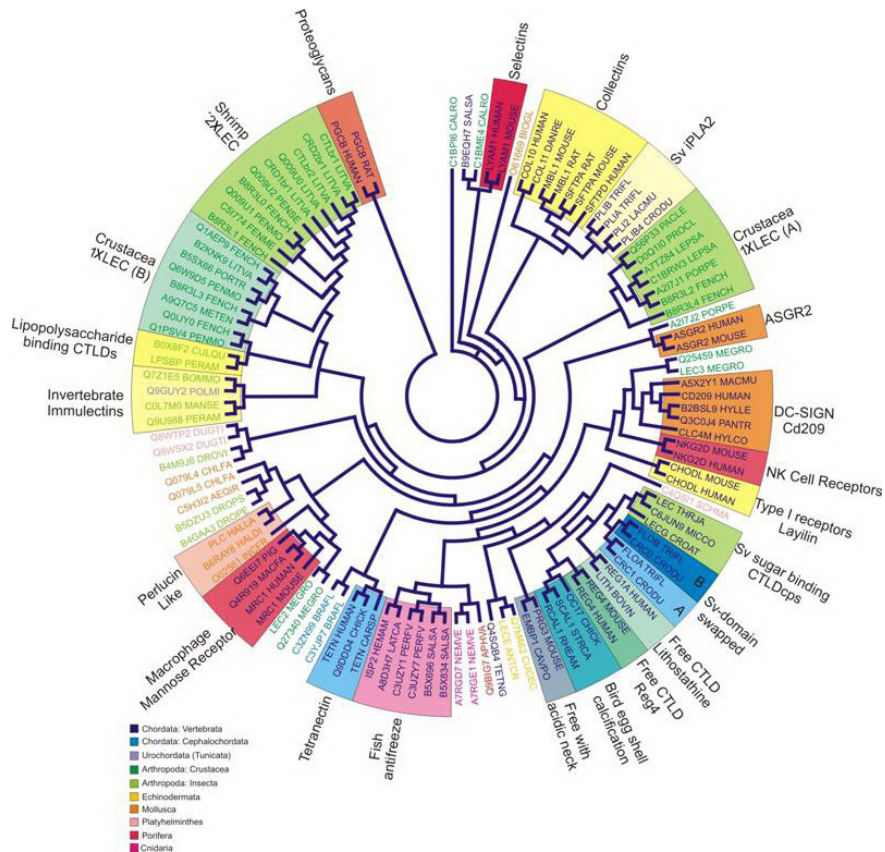


Figure 2. Evolutionary history of representative vertebrate and invertebrate C-type lectin-like domain-containing proteins deduced sequences (CTLDcps). The sequence codes are according to the following criteria: NAME_SPECIES [organism] CTLD protein name (number of amino acids and LEC domains) |UNIPROT ACCESSION NUMBER|. The unrooted phylogenetic tree encompasses 10 phyla and sub-phyla, which are highlighted with different colors. Shrimp CTLs are clustered as those that have a single domain (crustacean 1XLEC [A] and 1XLEC [B]), and those with two in tandem domains (shrimp 2XLEC).

These sequences were clustered according to the spatial architecture of vertebrate CTLDcps and others (Drickamer, 1993; Zelensky and Gready, 2005). Phylogenetic analysis of proteins containing C-type lectin domains from vertebrates and invertebrates clearly grouped the shrimp CTLDcps into three functional clusters (Figure 2). 1) Prawn and shrimp lectins with dual (in tandem) Carbohydrate Recognition Domains (denominated “Crustacea 2XLEC” and including the following entries: B8R3L1, C5I774, B8R3L0, Q009U2, Q009U0, and the two sequences of *L. vannamei* from this study). 2) Shrimp and crab lectins with a single-CRD (named “Crustacea 1XLEC [B]”, with the entries Q1PSV4, Q0IJY0, A9Q7C5, B8R3L3, Q6W9D5, B5SX66, B2KNK9, Q1AEP9) are clustered in one large group related to insect lipopolysaccharide binding CTLDs and immunoelectins. This cluster is related to the subgroups of CTLDcps with single- or dual-CRDs, which split from a common ancestor molecule, and includes sequences from the salmon louse Copepoda (*Lepeophtheirus salmonis*), the red swamp crayfish (*Procambarus clarkii*), the blue swimmer crab (*Portunus pelagicus*), and the shrimp (*F. chinensis*). 3) The cluster “1XLEC [A]”, composed by the entries Q56P33, D0Q1I0, A7TZ84, C1BRW3, A2I7J1, B8R3L2, and B8R3L4, formed another group of single-CRDs, which are related to zymogen granule membrane proteins from Atlantic salmon (*Salmo salar* B9EQH7) and Copepod Vesicular integral-membrane proteins (*Caligus rogercresseyi* C1BPI6), as well as to selectins and collectins.

In summary, the branch “Crustacea 1XLEC [A]” includes the cluster of Copepoda, Crab and Shrimp of a single-CTLdcp (Q56P33_PACLE; D0Q1I0_PROCL; A7TZ84_LEPSA; C1BRW3_LEPSA; B8R3L2_FENCH). These CTLDcps are probably involved in cellular adhesion and developmental regulation, or they are humoral sugar-binding proteins with suggested involvement in innate immunity. The other Crustacean cluster, named “Crustacea 1XLEC [B]”, encompasses the cluster of crab and shrimp CTLcps with a single-CTL domain. The group “Shrimp 2XLEC” comprises sequences with two in tandem CRDs: B8R3L1_FENCH; C5I774_FENME; Q009U1_PENMO; B8R3L0_FENCH; Q009U2_PENSE; Q009U0_LITVA; CTLbr2_LITVA; CTLbr1_LITVA, in which the CRD1 from CTL-br1_LITVA sequence is the basal ancestral domain that is similar to other shrimp single-domain CTLDcps. Moreover, the entries Q1PSV4_PENMO, Q0IJY0_FENCH, A9Q7C5_METEN, B8R3L3_FENCH, Q6W9D5_PENMO, B5SX66_PORTR, B2KNK9_LITVA, and Q1AEP9_FENCH are found in the hepatopancreas of distinct shrimp species, showing the importance of C-type lectins in the initial detection of microbial infection and their role in the innate immune response.

As shown in Figure 2, phylogenetic relationship analysis strongly suggests that penaeid shrimp C-type lectins containing two CRDs originated by duplication of single-domain C-type lectin genes and the sequences corresponding to one of these two domains quickly became free to evolve, as observed for the LvCTL-br1 and -br2 homologs. In fact, a corresponding phenomenon called positive Darwinianism or accelerated evolution is known to operate in animal toxin genes, such as those belonging to the families of snake venom phospholipases A₂ (Ohno et al., 2003) and C-type lectins (Ogawa et al., 2005). Thus far, such modular arrangements of proteins and their genes seem to offer flexible templates for the creation of functional and structural diversity of some active families of proteins.

Expression of LvCTL-br1 in shrimp gills upon IHNV infection

The closest LvCTL-br1 and -br2 homolog, namely LvLT (Q009U0_LITVA), is exclu-

sively expressed in the hepatopancreas of *L. vannamei* and is subjected to early gene down-regulation after WSSV infection (Ma et al., 2007). To determine whether LvCTL-br1 (or -br2) are responsive to viral infection, we initially analyzed the expression of LvCTL-br1 in the gills of *L. vannamei* chronically affected by IHHNV. The IHHNV is the viral causative agent of infection with hypodermal and hematopoietic necrosis (IHHN) disease. This virus affects all life stages of penaeid shrimp and causes massive mortality (>90%). Because WSSV is not the prevalent virus in Brazilian shrimp farms, neutralization of IHHNV by LvCTL-br1 offers a good investigative model to improve our knowledge about genes and proteins that participate in the shrimp innate immune system.

As shown in Figure 3, despite being expressed in the hepatopancreas of uninfected *L. vannamei*, similar to LvLT of *L. vannamei* growing in Asia, LvCTL-br1 (or -br2) are inductively expressed in the gills of shrimps severely infected by IHHNV (stages 2 and 3 of the IHHN disease). This pattern of expression contrasts with the LvLT expression pattern, which is restricted specifically to the hepatopancreas.

Increased levels of LvCTL-br1 transcripts in the shrimp gills in response to viral infection indicate that these proteins are functionally active, constituting novel two-domain C-type lectin members that participate in the innate immune response of penaeid shrimp against viruses.

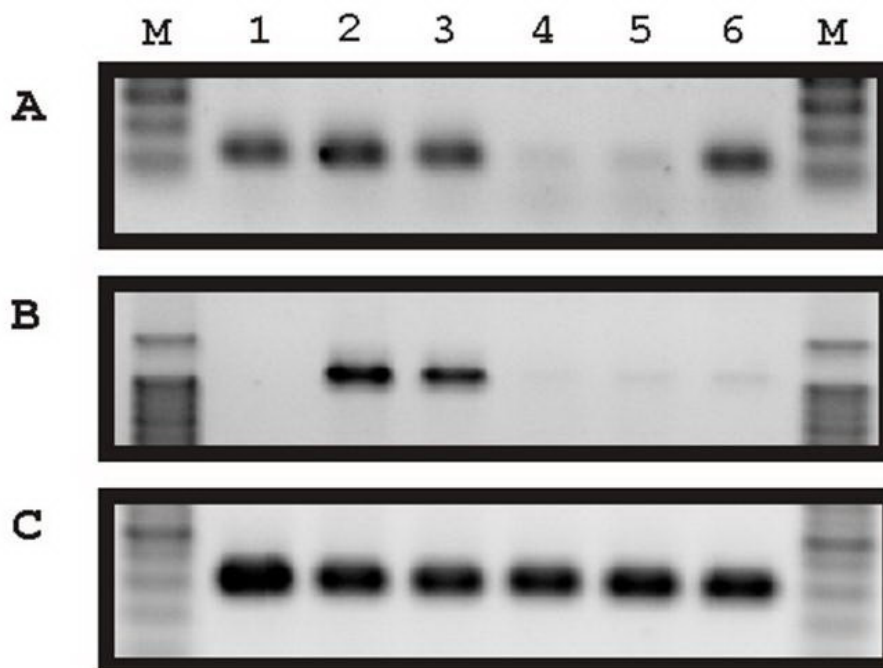


Figure 3. Induced expression of LvCTL-br1/-br2 in the gill of IHHNV infected *Litopenaeus vannamei*. Lane 1 = *L. vannamei* hepatopancreas; lanes 2 to 6 = *L. vannamei* gills from shrimp with distinct grade of IHHNV infection: severely affected (lanes 2 and 3) and moribund shrimp with initial symptoms of IHHN disease (lanes 4 to 6); M = 100-bp ladder molecular marker. A. LvCTL-br1 expression. B. IHHNV positive samples. C. Reference gene (*L. vannamei* β -actin). Note that a sense primer for the GTR-domain was used in this analysis (5'-ATCCAGGAACCCGATGGAGGA-3'), hence both LvCTL-br1 and -br2 transcripts are detected.

Molecular modeling

Because LvCTL-br1 seems to be a novel active C-type lectin, a comparison between molecular models is desirable to understand how the mutated mannose binding domain-2 might influence the affinity and capacity of both domains to bind oligosaccharides. Until now, no crystal structures of any shrimp C-type lectins have been solved, but inferences can still be made about the mechanism of interaction between sugars and residues in the binding site of a carbohydrate domain. This information can help to reveal the precise way in which a given interaction occurs. Thus, we have performed homology modeling of the LvCTL-br1 and -br2 protein homologs in order to identify where the mutated residues lie in the respective 3-D structures and how they might affect the ligand binding.

The models reveal the typical CTLDs constituted by α -helix chains flanked by β -sheets and interlaced by loops. Superimposition of CRD1 and CRD2 of LvCTL-br1 clearly shows the presence of a longer loop in CRD2 (Figure 4), where the mutated mannose binding site -GTR- triad is located (Figure 5). Additionally, the longer loop in the second domain lies in a region where the accommodation of complex carbohydrates would not be affected (Figure 5A and B). The observation of a long loop in CRD2 could explain why there is no crystal structure available for some C-type lectins. Long loops likely prevent the crystallization process by increasing the vibration scattering, which can interfere with the crystallization process. Moreover, the superimposition between the first domain of LvCTL-br2 and DC-SIGNR in complex with GlcNAc2Man3 (PDB code: 1K9J) reveals that two of its mutated residues in the sequence are close to the active site (data not shown).

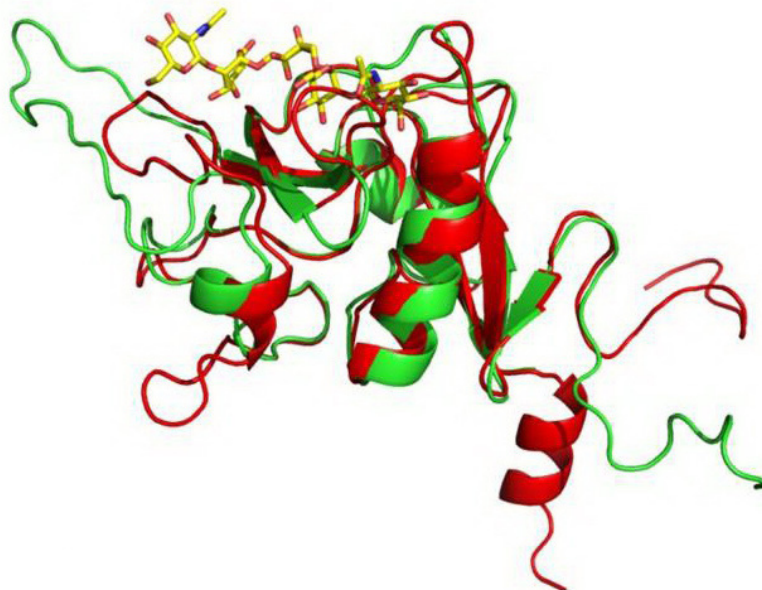


Figure 4. Superimposition of 3-D structures of LvCTLbr-1 carbohydrate recognition domain-1 and -2. The CRD1 of LvCTL-br1 is represented in red, whereas the CRD2 is in green. The long loop in CRD2 and the superimposed sugar binding sites are shown.

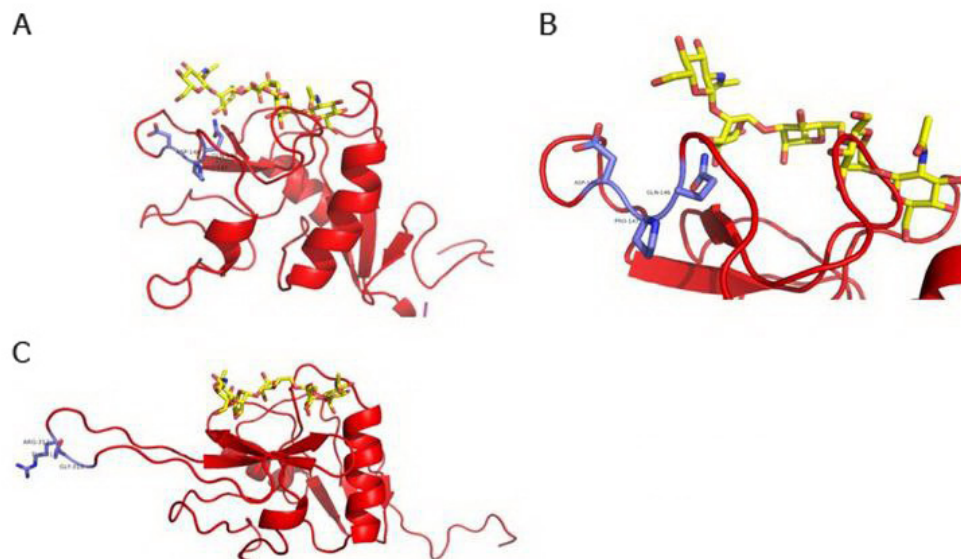


Figure 5. Tri-dimensional structures of LvCTLbr-1 carbohydrate recognition domain-1 and -2. **A.** and **B.** A 3-D model of CRD1 showing the triad -QPD- of the sugar binding site. **C.** The overall structure of CRD2, presenting the -GTR- triad in the long loop.

In fact, by analyzing the replacements of threonine¹⁵⁰ for alanine in the CRD1 of LvCTL-br1 and aspartic acid¹⁰⁹ for glycine, it is evident that the number of hydrogen bonds is less in the CRD1 of LvCTL-br2 than the CRD1 of LvCTL-br1 (Figure 6). This smaller number of bonds cause a weaker chemical interaction between residues of the LvCTL-br2 CRD1 binding site and its sugar partner.

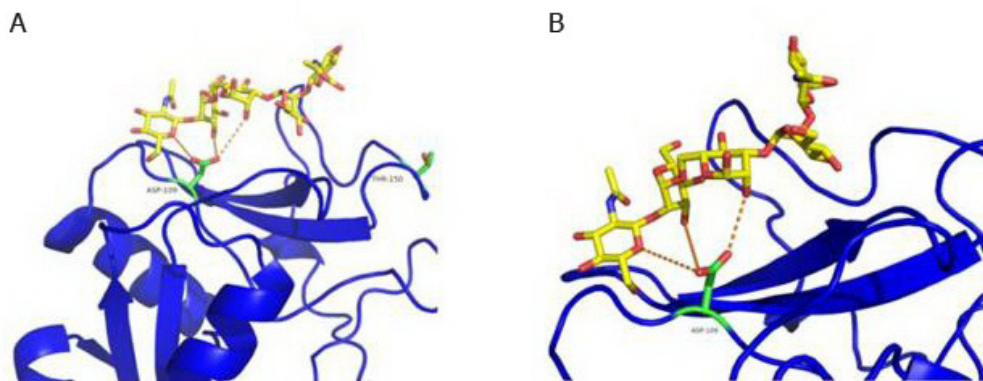


Figure 6. Partial tri-dimensional structure of LvCTL-br1 CRD1 in complex with GlcNAc₂Man₃. **A.** and **B.** The pattern of hydrogen bonding between the complex sugar GlcNAc₂Man₃ and the Asp109 from the triad -QPD- in the carbohydrate binding site is highlighted.

Concerning the second domain (CRD2) of LvCTL homologs (isoforms), the single amino acid difference is located at position 141, where the replacement of a leucine residue in LvCTL-br1 for a phenylalanine in CTL-br2 is observed. Although this substitution is located close to the binding site, local structural conformational changes were not significantly detectable (data not shown). Such amino acid replacement is more likely to affect the affinity of the proteins for their specific carbohydrates, but this prediction deserves detailed biochemical investigation. Furthermore, the assumption that changes at key positions are responsible for the wide differences in sugar affinity and specificity displayed by a handful of lectin family members is well described (Drickamer, 1995; Dodd and Drickamer, 2001).

CONCLUSIONS

Taking all these findings into account, the structural features described here for the first and second domains of these novel *L. vannamei* C-type lectin CRDs, observed using current 3-D models, together with the fact that such proteins are induced upon viral infection and cluster in a group of functionally related members, provide some clues about how *Litopenaeus vanamei* CTLs have evolved to bind complex sugars as part of the intricate invertebrate-pathogen interaction process.

ACKNOWLEDGMENTS

We are indebted to the National Brazilian Council for Scientific and Technological Development (CNPq) for the financial support of this study. We also wish to convey our gratitude to Dr. Georg Steinkellner (Centre of Molecular Biosciences, University of Graz, Austria) for the use and support of the Yasara program. We also thank Dr. Bruno Matias (Department of Biochemistry, Federal University of Ceará, Brazil) for critical suggestions on molecular structural analysis. B.S. Cavada and G. Rádís-Baptista are members of the Scientific Committee of CNPq.

REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Ausubel FM, Brent R, Kingston RE, Moore DD, et al. (1995). *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., New York.
- Christophides GK, Vlachou D and Kafatos FC (2004). Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*. *Immunol. Rev.* 198: 127-148.
- Dodd RB and Drickamer K (2001). Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* 11: 71R-79R.
- Drickamer K (1993). Evolution of Ca²⁺-dependent animal lectins. *Prog. Nucleic Acids Res. Mol. Biol.* 45: 207-232.
- Drickamer K (1995). Increasing diversity of animal lectin structures. *Curr. Opin. Struct. Biol.* 5: 612-616.
- Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.
- Gascuel O (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14: 685-695.
- Gouy M, Guindon S and Gascuel O (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27: 221-224.
- Hoffmann JA, Kafatos FC, Janeway CA and Ezekowitz RA (1999). Phylogenetic perspectives in innate immunity. *Science*

- 284: 1313-1318.
- Hoofst RW, Sander C, Scharf M and Vriend G (1996a). The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. *Comput. Appl. Biosci.* 12: 525-529.
- Hoofst RW, Vriend G, Sander C and Abola EE (1996b). Errors in protein structures. *Nature* 381: 272.
- Ji PF, Yao CL and Wang ZY (2009). Immune response and gene expression in shrimp (*Litopenaeus vannamei*) hemocytes and hepatopancreas against some pathogen-associated molecular patterns. *Fish Shellfish Immunol.* 27: 563-570.
- Kerrigan AM and Brown GD (2009). C-type lectins and phagocytosis. *Immunobiology* 214: 562-575.
- Kong HJ, Park EM, Nam BH, Kim YO, et al. (2008). A C-type lectin like-domain (CTLN)-containing protein (PtLP) from the swimming crab *Portunus trituberculatus*. *Fish Shellfish Immunol.* 25: 311-314.
- Krieger E, Koraimann G and Vriend G (2002). Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. *Proteins* 47: 393-402.
- Leu JH, Chang CC, Wu JL, Hsu CW, et al. (2007). Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*. *BMC Genomics* 8: 120.
- Liu YC, Li FH, Dong B, Wang B, et al. (2007). Molecular cloning, characterization and expression analysis of a putative C-type lectin (Fclectin) gene in Chinese shrimp *Fenneropenaeus chinensis*. *Mol. Immunol.* 44: 598-607.
- Luo T, Yang H, Li F, Zhang X, et al. (2006). Purification, characterization and cDNA cloning of a novel lipopolysaccharide-binding lectin from the shrimp *Penaeus monodon*. *Dev. Comp. Immunol.* 30: 607-617.
- Ma TH, Tiu SH, He JG and Chan SM (2007). Molecular cloning of a C-type lectin (LvLT) from the shrimp *Litopenaeus vannamei*: early gene down-regulation after WSSV infection. *Fish Shellfish Immunol.* 23: 430-437.
- Ma TH, Benzie JA, He JG and Chan SM (2008). PmLT, a C-type lectin specific to hepatopancreas is involved in the innate defense of the shrimp *Penaeus monodon*. *J. Invertebr. Pathol.* 99: 332-341.
- Matsubara H, Nakamura-Tsuruta S, Hirabayashi J, Jimbo M, et al. (2007). Diverse sugar-binding specificities of marine invertebrate C-type lectins. *Biosci. Biotechnol. Biochem.* 71: 513-519.
- Nunan LM, Arce SM, Staha RJ and Lightner DV (2001). Prevalence of infectious hypodermal and hematopoietic necrosis virus (IHHNV) and white spot syndrome virus (WSSV) in *Litopenaeus vannamei* in the Pacific Ocean of the coast of Panama. *J. World Aquaculture Soc.* 32: 330-334.
- Ogawa T, Chijiwa T, Oda-Ueda N and Ohno M (2005). Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. *Toxicon* 45: 1-14.
- Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, et al. (2003). Molecular evolution of myotoxic phospholipases A2 from snake venom. *Toxicon* 42: 841-854.
- Rittidach W, Pajit N and Utarabhand P (2007). Purification and characterization of a lectin from the banana shrimp *Fenneropenaeus merguensis* hemolymph. *Biochim. Biophys. Acta* 1770: 106-114.
- Schnitger AK, Yassine H, Kafatos FC and Osta MA (2009). Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. *J. Biol. Chem.* 284: 17616-17624.
- Soonthornchai W, Rungrasamee W, Karoonuthaisiri N, Jarayabhand P, et al. (2010). Expression of immune-related genes in the digestive organ of shrimp, *Penaeus monodon*, after an oral infection by *Vibrio harveyi*. *Dev. Comp. Immunol.* 34: 19-28.
- Sun J, Wang L, Wang B, Guo Z, et al. (2008a). Purification and characterization of a natural lectin from the plasma of the shrimp *Fenneropenaeus chinensis*. *Fish Shellfish Immunol.* 25: 290-297.
- Sun YD, Fu LD, Jia YP, Du XJ, et al. (2008b). A hepatopancreas-specific C-type lectin from the Chinese shrimp *Fenneropenaeus chinensis* exhibits antimicrobial activity. *Mol. Immunol.* 45: 348-361.
- Takase H, Watanabe A, Yoshizawa Y, Kitami M, et al. (2009). Identification and comparative analysis of three novel C-type lectins from the silkworm with functional implications in pathogen recognition. *Dev. Comp. Immunol.* 33: 789-800.
- Tian YY, Liu Y, Zhao XF and Wang JX (2009). Characterization of a C-type lectin from the cotton bollworm, *Helicoverpa armigera*. *Dev. Comp. Immunol.* 33: 772-779.
- Wang XW, Xu WT, Zhang XW, Zhao XF, et al. (2009a). A C-type lectin is involved in the innate immune response of Chinese white shrimp. *Fish Shellfish Immunol.* 27: 556-562.
- Wang XW, Zhang XW, Xu WT, Zhao XF, et al. (2009b). A novel C-type lectin (FcLec4) facilitates the clearance of *Vibrio anguillarum* in vivo in Chinese white shrimp. *Dev. Comp. Immunol.* 33: 1039-1047.
- Weis WI and Drickamer K (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* 65: 441-473.
- Wood-Charlson EM and Weis VM (2009). The diversity of C-type lectins in the genome of a basal metazoan, *Nematostella vectensis*. *Dev. Comp. Immunol.* 33: 881-889.
- Yamaura K, Takahashi KG and Suzuki T (2008). Identification and tissue expression analysis of C-type lectin and galectin in the Pacific oyster, *Crassostrea gigas*. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 149: 168-175.
- Zelensky AN and Gready JE (2005). The C-type lectin-like domain superfamily. *FEBS J.* 272: 6179-6217.

- Zhang H, Wang H, Wang L, Song X, et al. (2009a). A novel C-type lectin (Cflec-3) from *Chlamys farreri* with three carbohydrate-recognition domains. *Fish Shellfish Immunol.* 26: 707-715.
- Zhang XW, Xu WT, Wang XW, Mu Y, et al. (2009b). A novel C-type lectin with two CRD domains from Chinese shrimp *Fenneropenaeus chinensis* functions as a pattern recognition protein. *Mol. Immunol.* 46: 1626-1637.
- Zhang Y, Qiu L, Song L, Zhang H, et al. (2009c). Cloning and characterization of a novel C-type lectin gene from shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 26: 183-192.
- Zhao ZY, Yin ZX, Xu XP, Weng SP, et al. (2009). A novel C-type lectin from the shrimp *Litopenaeus vannamei* possesses anti-white spot syndrome virus activity. *J. Virol.* 83: 347-356.