



OPEN ACCESS

EDITED BY

Jun Li,
Lake Superior State University,
United States

REVIEWED BY

Fan Wang,
Shantou University, China
Weilin Wang,
Dalian Ocean University, China

*CORRESPONDENCE

Anlong Xu
lssxl@mail.sysu.edu.cn
Shengfeng Huang
hshengf2@mail.sysu.edu.cn

SPECIALTY SECTION

This article was submitted to
Comparative Immunology,
a section of the journal
Frontiers in Immunology

RECEIVED 17 June 2022

ACCEPTED 08 September 2022

PUBLISHED 07 October 2022

CITATION

Li J, Liu S, Zhang Y, Huang Q,
Zhang H, OuYang J, Mao F, Fan H,
Yi W, Dong M, Xu A and Huang S
(2022) Two novel mollusk short-form
ApeC-containing proteins act as
pattern recognition proteins
for peptidoglycan.
Front. Immunol. 13:971883.
doi: 10.3389/fimmu.2022.971883

COPYRIGHT

© 2022 Li, Liu, Zhang, Huang, Zhang,
OuYang, Mao, Fan, Yi, Dong, Xu and
Huang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Two novel mollusk short-form ApeC-containing proteins act as pattern recognition proteins for peptidoglycan

Jin Li^{1,2}, Shumin Liu^{1,2}, Yang Zhang³, Qiuyun Huang³,
Hao Zhang^{1,2}, Jihua OuYang^{1,2}, Fan Mao³, Huiping Fan^{1,2},
Wenjie Yi³, Meiling Dong¹, Anlong Xu^{1,4*}
and Shengfeng Huang^{1,2*}

¹Key Laboratory of Biocontrol, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Key Laboratory of Pharmaceutical Functional Genes, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, ²Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, ³Chinese Academy of Sciences Key Laboratory of Tropical Marine Bio-Resources and Ecology and Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China, ⁴School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China

The Apextrin C-terminal (ApeC) domain is a new protein domain largely specific to aquatic invertebrates. In amphioxus, a short-form ApeC-containing protein (ACP) family is capable of binding peptidoglycan (PGN) and agglutinating bacteria *via* its ApeC domain. However, the functions of ApeC in other phyla remain unknown. Here we examined 130 ACPs from gastropods and bivalves, the first and second biggest mollusk classes. They were classified into nine groups based on their phylogenetics and architectures, including three groups of short-form ACPs, one group of apextrins and two groups of ACPs of complex architectures. No groups have orthologs in other phyla and only four groups have members in both gastropods and bivalves, suggesting that mollusk ACPs are highly diversified. We selected one bivalve ACP (CgACP1; from the oyster *Crossostrea gigas*) and one gastropod ACP (BgACP1; from the snail *Biomphalaria glabrata*) for functional experiments. Both are highly-expressed, secreted short-form ACPs and hence comparable to the amphioxus ACPs previously reported. We found that recombinant CgACP1 and BgACP1 bound with yeasts and several bacteria with different affinities. They also agglutinated these microbes, but showed no inhibiting or killing effects. Further analyses show that both ACPs had high affinities to the Lys-type PGN from *S. aureus* but weak or no affinities to the DAP-type PGN from *Bacillus subtilis*. Both recombinant ACPs displayed weak or no affinities to other microbial cell wall components, including lipopolysaccharide (LPS), lipoteichoic acid (LTA), zymosan A, chitin, chitosan and cellulose, as well as to several PGN moieties, including muramyl dipeptide (MDP), N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc). Besides, CgACP1 had the highest expression in the gill and could be greatly up-regulated quickly after bacterial challenge. This is reminiscent of the

amphioxus ACP1/2 which serve as essential mucus lectins in the gill. Taken together, the current findings from mollusk and amphioxus ACPs suggest several basic common traits for the ApeC domains, including the high affinity to Lys-type PGN, the bacterial binding and agglutinating capacity, and the role as mucus proteins to protect the mucosal surface.

KEYWORDS

ApeC, ACP, microbial binding, peptidoglycan, mollusca, *Crossostrea gigas*, *Biomphalaria glabrata*

Introduction

The Apextrin C-terminal (ApeC) domain is a novel protein domain first discovered in the cephalochordate amphioxus (1–3), a basal living chordate lineage (4). ApeC was later found to be widely distributed in at least fifteen invertebrate phyla, most of which are marine and freshwater invertebrates, including ascidians, cephalochordates, echinoderms, hemichordates, mollusks, cnidarians, etc (5). However, ApeC was not found in vertebrates and nor in the major arthropod lineages including insects and crustaceans, which may make ApeC largely an aquatic invertebrate-specific protein domain (5). ApeC is a large domain with about 200 amino acids, including eight conserved cysteines and three conserved DEXD motifs, and is mostly present in secreted and membrane-bound proteins (5). ApeC are found to form different domain combinations with other domain types (5), making it one of the versatile and promiscuous domains like those immunoglobulin (IG) and C-type lectin domains.

Thus far, the functional information of ApeC domains came from a unique ApeC-containing proteins (ACPs) family in amphioxus. Four members (ACP1, 2, 3 and 5) of this family were investigated (2, 6). Despite the sequence and structural differences, these ACPs adopt a core architecture: a leading signal peptide (SP) and a C-terminal ApeC domain. They could agglutinate bacteria by using the ApeC domain to bind with the cell wall component peptidoglycan (PGN), but they showed no inhibiting or killing effects on bacteria. ACP1 and ACP2 are mainly expressed in the gill and skin, while ACP3 and ACP5 are concentrated in the gut. Secreted ACP1 is indispensable for the anti-bacterial mucosal immune responses in the gill, while ACP2 and ACP3 can regulate the TRAF6-NF- κ B pathways when present in cytosol. Therefore, this amphioxus ACP family have dual roles, as extracellular lectins for PGN and as intracellular immune regulators (2, 6).

In other phyla, the functions of ApeC domain have not been directly investigated, but there are evidence suggesting that many ACPs might have important roles in development,

immunity and stress resistance. In the sea chitin *Heliocidaris erythrogramma*, an ACP called apextrin is a secreted protein with a unique MACPF/perforin-ApeC domain structure (7–9). During embryogenesis, this protein was found to be highly expressed and concentrated in the apical extracellular matrix of ectoderm columnar cells, hence was suggested to be involved in apical cell adhesion. Another apextrin, from the clam *Ruditapes philippinarum*, exhibited an inhibitory effect on gram-positive bacteria probably due to its MACPF/perforin domain, but the role of its ApeC domain was not examined (10). In mussel *Mytilus galloprovincialis*, two ACPs called apelB and apelP, were also proposed to have a role in embryogenesis because their expression could be increased thousands of times after fertilization and reached a peak at the blastula and trochophore stages (11). In the oyster *Crossostrea gigas*, an ACP was found to be highly up-regulated under hypoxic conditions, suggesting a role in the anti-stress responses (12). In terms of immune roles, several ACP genes from sea urchins, oysters, mussels, clams, and brachiopods exhibited high expression in the immune cells and tissues and could be up-regulated in response to bacterial challenge (11, 13–18). Moreover, a recent study found that ACPs existed in the pedal mucus secreted by limpets, probably being involved in immunity and glue-like adhesion (19).

With approximately 200 thousand extant species, mollusks comprise one of the largest animal phyla, second only to the arthropods (20) and accounting for roughly 10% of the animal biomass (21). Mollusks are heterogeneous in size, morphology, adapting to different habitats (terrestrial, fresh water and marine) and feeding behaviors, hence reflecting its massive radiation since the Cambrian era (22). Bivalvia and Gastropoda represent the two largest mollusk classes. Bivalves are restricted to aquatic life, accounting for about 14% of mollusks, while gastropods comprise over 80% of mollusks, adapted to both aquatic and terrestrial life (23). Previous studies showed that bivalves and gastropods have a large number of ACPs (5). Some mollusk ACPs appear to play roles in different processes including embryogenesis, anti-stress

responses, mucus-based adhesion, and mucosal and humoral immunity (11, 12, 15–19).

As a new protein domain, the functions of ApeC remains largely elusive. Previous analysis of an amphioxus ACP family suggests that its members serve as novel pattern-recognition proteins and its ApeC domains are responsible for the carbohydrate-binding activity (2, 6). It is not known if these functional properties could be applied to other ACP families and other phyla. After all, ApeC sequences are highly diverged and no ACP orthologs could be found between different phyla (5). On the other side, mollusks have great ecological and economical significance, and coincidentally possess many ACPs appearing to have relevant functions. Therefore, this study is focused on the mollusk ACPs. First, we analyzed their composition, species distribution, phylogenetics, domain architectures and sequence evolution. Then, we chose two highly-expressed, secreted short-form ACPs, one from bivalves and the other from gastropods, for molecular functional experiments. Finally, we compare the functional traits between these two mollusk ACPs and the previously studied amphioxus ACPs. Our findings could provide new insights into the basic ApeC functions, which may help the future functional studies of the ACP members in mollusks and other invertebrate phyla.

Materials and methods

Animals, pathogen challenge and hemocyte preparation

The Pacific oysters *Crossostrea gigas* (two ages with an average shell length of 100 mm) were from Qingdao, Shandong Province, China, and were maintained under previously described conditions (24), 22–25°C in a tank with circulating seawater for two weeks prior to the experiment. Oysters were fed twice daily on *Tetraselmis Suecica* and *Isochrysis Galbana*. All experimental manipulations were performed in accordance with local guidelines for the care and use of laboratory animals.

For monitoring the gene expression changes after immune stimulation, bacterial challenged oysters were prepared as previously described (25, 26). Briefly, *Vibrio parahaemolyticus* was cultured in LB (Luria-Bertani) broth at 37°C to OD₆₀₀=0.6–0.8 (they grew faster at 37°C and remained healthy), and then centrifuged at 800×g for 10 min at 4°C. After being washed for 3 times in PBS (pH 7.4), bacterial pellet was resuspended in PBS to an adjusted density of OD₆₀₀ = 1.0. Oysters in the challenged group were injected with a 100 µl suspension into the adductor muscle, while oysters in the control group were injected with an equal volume of PBS. After injection, oysters were returned to separate tanks for subsequent sampling.

Oyster hemolymph was extracted from the posterior adductor muscle of the oysters by using a medical-grade

syringe (0.45×15.5 mm) and hemolymph from each individual was counted toward one sample. Hemolymph was collected immediately by centrifugation at 200×g for 10 min, then the supernatant was sterilized through a 0.22 µm filter for subsequent analyses. For each time point after immune stimulation, three individuals were sampled and the sampled materials were pooled together for later use.

RNA isolation and cDNA synthesis

Total RNA was extracted from frozen tissue with TRIzol reagent (Invitrogen) according to the manufacturer's directions and precipitated with isopropyl alcohol. After quality checks using a spectrophotometer and agarose gel electrophoresis, the RNA was reverse-transcribed to synthesize the first-strand cDNA using the PrimeScript First-strand cDNA Synthesis Kit (Takara) using the oligo d(T) primer following the manufacturer's protocol. Then the cDNA was stored at -20°C.

Cloning and synthesizing of the ACP cDNA

Mollusk ACP sequences were obtained by conducting homology-based searches on the National Center for Biotechnology Information database (NCBI, <http://www.ncbi.nlm.nih.gov/genbank/>) using the amphioxus ApeC sequences as baits. Pacific oyster CgACP1 was identified in the *Crossostrea gigas* genome and the accession is XP_011420902 (or XP_034306861). Gastropod snail BgACP1 was identified in the *Biomphalaria glabrata* genome and the accession is XP_013078768. To obtain the complete cDNA sequence of CgACP1, a pair of gene-specific primers (forward primer: 5'-TTTAACGGTCTAAAGGTCCTG-3'; reverse primer: 5'-AAGAGATTACACATTTGCCT-3') was used. As for the cDNA sequence of BgACP1, it was directly synthesized based on a manually curated sequence (namely, the consensus sequence of multiple sequences and RNA-seq reads). Finally, the cDNA fragments were cloned into the pGEX-Teasy vector (Promega) and verified by sequencing at The Beijing Genomics Institute (BGI).

Bioinformatic analysis

The protein domain and signal peptide were predicted with the Simple Modular Architecture Research Tool (SMART, <http://smart.embl-heidelberg.de/>) and SignalP (<http://cbs.dtu.dk/services/SignalP>) software, respectively. The isoelectric point (pI) and m.w. were speculated on ExPASy Web site (<http://www.expasy.org/tools/>). BLASTP was performed to analyze the sequence identities. Multiple sequence alignments

of ApeC domains from different species were analyzed using the default parameters of MEGA-X (27) by using ClustalW algorithm and were manually corrected using GeneDoc software (28). The Neighbor-joining phylogenetic tree (29) was built using MEGA-X (27) with the JTT matrix-based method (30) and handling gaps by pairwise deletion. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

Real-time quantitative RT-PCR

Real-time quantitative RT-PCR (QPCR) was performed on the Roche LightCycler 480 using SYBR[®] Green Realtime PCR Master Mix (Toyobo) according to the manufacturer's protocol. The reaction volume was 10 μ l, which containing 1 μ l template cDNA, a primer concentration of 0.5 μ M and 5 μ l of 2 \times SYBR Green Mix. The PCR program is 95°C for 1 min, followed by 40 cycles of 95°C for 15 s, 60°C for 15 s, and 85°C for 20 s. The cycle threshold values were calculated by the Roche LightCycler 480 software. The expression levels of different tissues were calculated using $2^{-\Delta C_t}$, and the fold change after challenge were calculated using $2^{-\Delta\Delta C_t}$ method based on the cycle threshold values. Reaction of each sample was performed in triplet using *gapdh* mRNA as the internal control. *CgACP1* primers for QPCR were 5'-GCCGTGCTAAGTCTCT-3' (forward) and 5'-TCCAGTGTCCCTCATTATCC-3' (reverse), and *gapdh* primers for QPCR were 5'-CTTCCGCGTACCAGTTCCA-3' (forward) and 5'-GCTGCTTCGCTTGTCTCCAC-3' (reverse).

Transcriptome analysis

Three mollusk species, including *Aplysia californica*, *Biomphalaria glabrata* and *Crossostrea gigas*, have high-quality genome assemblies and abundant transcriptome data deposited in the NCBI server. The accumulated expression levels (total RNA-seq raw read counts) of their short-form ACP genes could be queried through NCBI's Genes of Genomes in the Reference Sequence Collection (<https://www.ncbi.nlm.nih.gov/gene/>). For comparison, we also analyzed the short-form ACP genes in six species from other four phyla, including *Echinodermata*, *Hemichordata*, *Cephalochordata* and *Cnidaria*.

To analyze the expression of *CgACP1* in different tissues and at different time points in the gill after immune stimulation, we downloaded both the pacific oyster (*Crossostrea gigas*) reference genome (assembly oyster_v9: GCF_000297895) and its corresponding gene annotation files from NCBI's genome website. The RNA-seq data in different tissues were downloaded from NCBI's databases (BioProject PRJNA146329) (31). The Transcriptome data of the Pacific

oyster under *Vibrio* (an equal mixture of *V. anguillarum*, *V. tubiashii*, *V. aestuarianus*, *V. alginolyticus*) challenge at different time points were also downloaded from NCBI's databases (BioProject PRJNA194079) (32). The Raw RNA-seq data was cleaned using Trimmomatic v.0.36 (33) with default parameters for quality control. The clean data was mapped to the oyster reference genome using Hisat2 v.2.1.0 (34) under default parameters. The fragments per kilobase of exon model per million mapped fragments (FPKM) of each gene was calculated.

Preparation of recombinant proteins

Recombinant Mollusk ACP proteins were expressed in two different processes. PCR fragments encoding mature amino acids 21-435 aa of BgACP1 and 21-239 aa of CgACP1 were amplified by recombinant primers from the pGEM-T Easy vector (Promega) carrying full-length cDNA and inserted into plasmid pET32a (+) by recombinant method mentioned in ClonExpress[®]II One Step Cloning Kit (Vazyme). BgACP1 recombinant primers were 5'-gctgatcggatccTCATTCTTGGTGACCATCAACCC-3' (forward) and 5'-cgaattcggatccgaAACTCGTTCTTCGTAATAGCACAGC-3' (reverse); CgACP1 recombinant primers were 5'-gctgatcggatccGTCGGATGGCCTTCTGGAA-3' (forward) and 5'-cgaattcggatccga-ACTGTAGTAGCAGTAGTAGAGACGATGGT (reverse). Recombinant plasmids were confirmed by DNA sequencing to construct the expression vector. Then, the expression plasmids were introduced into *E. coli* BL21 (DE3) and 0.1mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to induce the protein expression. TRX-BgACP1 was soluble and expressed in the supernatant but TRX-CgACP1 was insoluble and expressed as inclusion body. For TRX-BgACP1, the supernatant of cell lysates was applied to a Ni²⁺-chelating Sepharose column (GE Healthcare), pooled by elution with 250 mM imidazole. The purified recombinant TRX-BgACP1 was put into dialysis bags and dialyzed in PBS buffer at 4°C for 12 hours three times, and concentrated by filtration through an Ultrafree centrifugal filter device (Millipore).

TRX-CgACP1 fusion protein in inclusion body was denatured in 8 M urea using a Ni²⁺-chelating Sepharose column and eluted with 250 mM imidazole. Purified TRX-CgACP1 in 8M urea was renatured by three dialysis steps as previously described (35) with certain modifications, first in 3 M urea in PBS buffer containing 150 mM NaCl, pH 7.4, 2 mM reduced glutathione, 0.02 mM oxidized glutathione, 5% glycerol, and 0.05% Tween-20, then in 1 M urea in the same buffer, and finally in the buffer without urea, glutathione and Tween-20. Each dialysis step was performed for at least 12 hours at 4°C. Soluble fusion protein was concentrated by filtration through an Ultrafree centrifugal filter device (Millipore). The protein concentration was determined using the Pierce[™] BCA Protein Assay Kit (Thermo) according to the manufacturer's protocol.

Microbial binding assays

The bacteria including *Staphylococcus aureus*, *Enterococcus faecalis*, *Escherichia coli*, *Vibrio anguillarum*, *Vibrio parahaemolyticus*, *Acinetobacter calcoaceticus* and *Saccharomyces cerevisiae* were inoculated and cultured as described (6). The cells were washed and resuspended in PBS buffer. Approximately 2×10^6 microbes were incubated with 1 μg of purified recombinant proteins in PBS by gentle orbital rotation overnight at 4°C. Microbes were pelleted and washed five times with 1 ml of PBST (0.05% Tween-20 in PBS). The washed pellets were then suspended with reducing sample buffer quickly and denatured by heating at 100°C for 10 min. The binding proteins were validated by Western blot with an anti-His mouse monoclonal antibody (Sigma).

Microbial aggregation assays

Fluorescein isothiocyanate (FITC, sigma) was used to label bacteria. Microbes collected from liquid cultures were suspended in 1 ml PBS and mixed with 50 μl fluorescein isothiocyanate (FITC) (sigma, 10 mg/ml in DMSO). The reaction was gently agitated at room temperature in the dark for 3h and then washed five times with PBS. FITC-labeled *S. aureus* (2×10^8 cells/ml), *E. faecalis* (2×10^8 cells/ml), *E. coli* (8×10^8 cells/ml), *V. anguillarum* (2×10^8 cells/ml) or *S. cerevisiae* (3×10^7 cells/ml) were mixed with 5 μg TRX fusion proteins in PBS and incubated at room temperature in the dark for 2h, respectively. The agglutinating reaction was examined immediately under fluorescence microscopy (Carl Zeiss).

Antimicrobial activity assays

The growth curves of *S. aureus* and *E. coli* cultured with recombinant ACP proteins were tested as follows. Two single colonies were picked up and transferred into 1 ml of Luria-Bertani (LB) respectively and grown to mid-log phase. A volume of 50 μl of cell suspension was mixed with 200 μg purified recombinant proteins or TRX and added to 1 mL broth. Each example was incubated with 200 rpm at 37°C and the OD600 was measured every hour.

For colony counts assays, our procedures paralleled previously reported (36). *S. aureus* were suspended and experiments were performed in PBS buffer (pH 7.4). Incubate mixtures contained 10^5 bacterial CFU and 20 μg ACP proteins or TRX (as negative control) in 100 μl PBS buffer. After incubation for 3 hours at 37°C, 900 μl of the same buffer was added, and two additional serial 10-fold dilutions were made with that buffer. 100 μg portions from the three serial dilutions were spread over LB agar plates. After incubation for 12 hours, colony forming units (CFUs) were counted.

Binding assays of ACPs with the components of microbes

ELISA was used to analyze the binding of ACP proteins with soluble microbial cell wall components as previously described (37). In brief, a total of 20 μg of peptidoglycan (PGN) from *S. aureus* (Sigma), PGN from *B. subtilis* (Sigma), lipopolysaccharide (LPS, Sigma), lipoteichoic acid (LTA, Sigma), Zymosan A (Sigma), Chitin (Sigma), Chitosan (Sigma), Cellulose (Sigma), muramyl dipeptide (MDP), N-acetylglucosamine (GlcNAc, Sigma) and N-acetylmuramic acid (MurNAc, Sigma) in PBS (PGN, Zymosan A, Chitin, Chitosan and Cellulose were ultra-sonically solubilized) were used to coat a 96-well microplate (Corning 96-well Clear Polystyrene High-Bind Strip well Microplate) for 3 h at 37°C. Nonspecific binding to the wells was prevented by the addition of PBST (0.05% Tween 20 in PBS) containing 10% (wt/vol) skimmed milk overnight at 4°C. Several concentrations of ACP proteins were then added to the well and the mixtures were incubated for 2 h at 37°C. Binding proteins were detected with anti-His mAb (sigma) diluted 10000-fold for 1h at room temperature, followed by an hour incubation with a 5000-fold dilution of HRP-labeled anti-mouse IgG. Between each incubation step, unbound protein, mAb, or HRP-labeled anti-mouse IgG was washed off three times. Incubate with 100 μl TMB Substrate Solution (Thermo) at 25°C for 15 minutes and stop reaction. Add 50 μl of 2M sulfuric acid to each well to stop the reaction, and the absorbance was read at 450 nm. The assay was repeated at least three times.

Statistical analysis

Quantitative data are presented as the mean \pm SD and compared statistically by two-tailed Student *t* test. In all cases, differences of $p < 0.05$ were considered significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ indicate statistical differences. All experiments were repeated at least three times.

Results

Phylogenetic analysis of mollusk ACPs

Here we identified 130 non-redundant mollusk ACP sequences from different online databases. One third of these ACP sequences are contributed by mollusks (Figure 1A and Supplementary Table 1). Each mollusk species has 1-25 ACP sequences, and on average bivalves appear to have more ACPs than gastropods. Particularly, bivalve *Mytilus galloprovincialis* had the largest number of ACPs (up to 25 ACPs), followed by bivalve *Crassostrea gigas* (21 ACPs). As comparison, amphioxus and cnidarians have 2-34 ACPs per species. We conducted a phylogenetic analysis of these sequences using the Neighbor-

Joining (NJ) method and the Maximum Likelihood (ML) method (Figure 1B and Supplementary Figure 1). Based on the phylogenetic trees, mollusk ACPs could be classified into nine groups. None of these groups have orthologs in other phyla, which is consistent with the previous report that ACPs are rapid evolving and highly diversified (5). As comparison, bivalve ACPs have members in eight groups, while gastropod ACPs are only present in five groups. And only four groups (2, 3, 7 and 8) are conserved in both bivalves and gastropods. Moreover, the group 1 is the largest group and contains only bivalve ACPs, hence indicating a bivalve-specific family expansion and diversification. These findings suggest that mollusk ACPs are more diversified and more actively duplicated in bivalves than in gastropods.

Domain architectures of mollusk ACPs

Mollusk ACPs can be classified into at least nine protein architectures, each corresponding to one phylogenetic group (Figure 1B). It suggests that these architectures have their own origins and evolve independently. The group 1, 2 and 4 are secreted short-form ACPs, which features a single C-terminal ApeC domain and no other discernable known domains or motifs. On the other hand, the group 8 and 9 are highly complex proteins, containing more than ten domains or motifs, most of which are often present in extracellular proteins and matrix proteins. The other four groups have medium-sized architectures. Among them, the group 6 features a dual-ApeC module. This architecture is also found in other phyla, including Cnidaria, Tardigrada, Echinodermata and Platyhelminthes (5), though they share no orthologous relationships. Unlike other short repetitive motifs, ApeC is a large domain with ~200 aa, therefore we suspect that this dual-ApeC module may represent a conserved but unknown functional mode. As comparison, versatile immunoglobulin (IG) and C-type lectin domains are often present in tandem repeats. Moreover, the group 3 contains a special group of ACPs called apextrins. This subfamily adopts a highly conserved architecture (signal_peptide + MACPF/perforin+ApeC), and exist in at least four other phyla, including Cnidaria, Tardigrada, Hemichordata and Echinodermata, although no orthologs can be found between phyla. Notably, it has been suggested that in sea urchins and clams, apextrins participated in embryogenesis and immunity, respectively (7–10). Taken together, mollusk ApeC domains are present in various protein architectures.

Characterization of two short-form ACPs from oysters and snails

To understand the basic functions of mollusk ApeC domains, we chose two ACPs for experiments based on these criteria: having a short-form architecture and the high expression level.

These criteria were previously used to identify the amphioxus ACP1 and ACP2 (2). The first ACP (CgACP1) is chosen from the pacific oyster *C. gigas*, which is a representative bivalve, a famous economic species and a model organism for the mollusk immune and anti-stress systems (38). CgACP1 belongs to the group 1 of ACPs, which contains the shortest form of ACPs ever known and has no homologs from the gastropods (Figures 1B, 2A). For the purpose of comparison, we also selected a gastropod ACP (BgACP1) from the freshwater snail *Biomphalaria glabrata*, which is the intermediate host for the pathogen of human intestinal schistosomiasis (39). BgACP1 belongs to the group 4 ACPs, which has a short-form architecture and contains no homologs from the bivalves (Figures 1B, 2A). According to the accumulated RNA-seq statistics in NCBI's gene portal, both CgACP1 and BgACP1 have the highest expression among all ACP genes in their own species (Figure 2B and Supplementary Table 2). As comparison, amphioxus ACP1 also have the highest expression among all amphioxus ACPs (2).

We cloned the full-length cDNA of CgACP1. It encodes a 219 aa protein, which contains only a signal peptide and an ApeC domain, hence is the shortest and simplest ACP ever found (Figure 2A; full sequence shown in Supplementary Figure 2A). The ApeC domain of CgACP1 has about 47% similarity with the previously reported amphioxus ACP1/2/3. As for BgACP1, we did not have access to the animal of *B. glabrata*, hence its coding DNA was directly synthesized. BgACP1 has 435 aa, containing a signal peptide, a short middle region and a C-terminal ApeC domain (Figure 2A; full sequence shown in Supplementary Figure 2B). The length and structure of BgACP1 is similar to amphioxus ACP1/2/3, but it shares only 35% identity with amphioxus ACP1/2/3 in the ApeC domain. We created an ApeC-based multiple alignment for a selected set of short-form ACPs from different phyla (Figure 2C). It shows that both CgACP1 and BgACP1 have all eight conserved cysteines, which is the most prominent feature of the ApeC domains in all animals. In addition, CgACP1 preserve all three DXED motifs, while BgACP1 has mutations in all three DXED motifs.

The expression patterns of CgACP1

CgACP1 has been shown to be the highest expressed ACP in pacific oysters (Figure 2B and Supplementary Table 2). Here its expression in different adult tissues was further evaluated. Using the public available transcriptome data (32, 33), we showed that CgACP1 was predominantly present in the gill and labial palp, at an expression level even higher than that of beta-actin (Figure 3A and Supplementary Table 3). Note that the gill and labial palp are adjacent to each other and comprise the filter feeding organ. This expression pattern was confirmed by quantitative RT-PCR assays (Figure 3B). Transcriptome data also indicate that the expression of CgACP1 in the gill

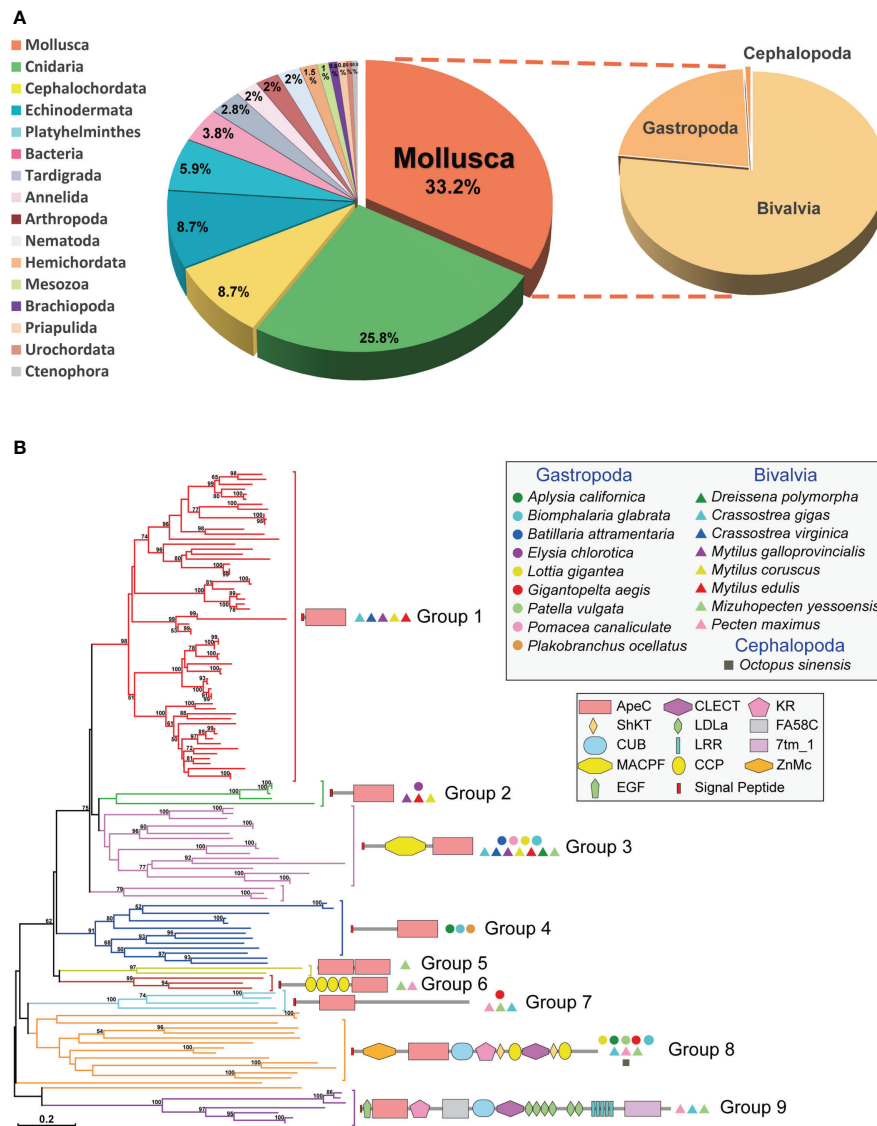


FIGURE 1 The distribution, phylogenetics and domain architectures of 130 mollusk ACPs. **(A)** The proportion of ACPs in different phyla. **(B)** A phylogenetic tree of all 130 mollusk ACPs based on the ApeC domain. The tree was constructed using the neighbor-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The protein architectures, grouping and source species are shown in corresponding positions. A Maximum Likelihood tree based on the same alignment is also provided in [Supplementary Figure 1](#). A full list of these mollusk ACPs is provided in [Supplementary Table 1](#).

could be greatly up-regulated shortly after being challenged by a mixture of *Vibrio* bacteria (Figure 3C). On the other hand, the hemolymph and the abductor muscle, had the lowest expression of CgACP1, approximately 1/50 of the expression in the gut (Figure 3A and Supplementary Table 3). Quantitative RT-PCR assays further showed that even after *Vibrio* bacterial stimulation, the expression of CgACP1 in the hemolymph was only transiently up-regulated (Figure 3D). This is surprising because the hemolymph contains mainly the immunocompetent hemocytes. Anyway, these findings suggest

that CgACP1 should have a more important role in the gill than in the hemocytes.

Recombinant CgACP1 and BgACP1 bound and aggregated different microbes

As secreted simple-structured proteins (especially CgACP1), both CgACP1 and BgACP1 might correspondingly have a

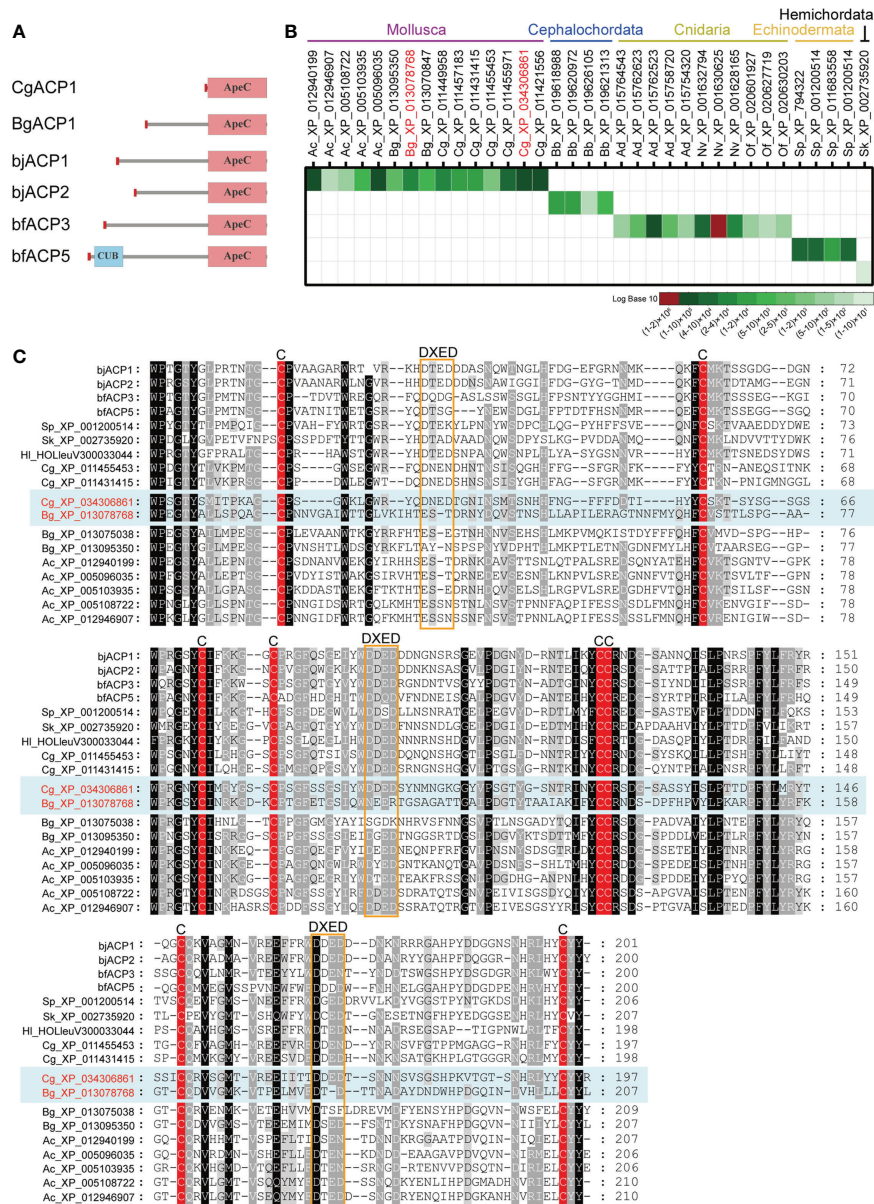


FIGURE 2

Comparison of the amino acid sequences and the global expression levels of a set of representative ACPs. **(A)** The domain architectures of the oyster CgACP1 and the snail BgACP1 compared with the amphioxus ACP1/2/3/5. **(B)** The accumulated expression level of a set of representative ACPs based on the RNA-seq data from NCBI's gene browser. The accession for CgACP1 and BgACP1 are highlighted in red. The raw expression data are shown in [Supplementary Table 2](#). **(C)** A multiple alignment of the ApeC domains of several short-form ACPs selected from different aquatic invertebrates. The conserved cysteine residues and the DXED motifs are highlight by red color and orange boxes, respectively. The accession of CgACP1 and BgACP1 are also highlighted in red. Ac, *Aplysia californica*; Bg, *Biomphalaria glabrata*; Cg, *Crossostrea gigas*; bb, *Branchiostoma belcheri*; bj, *Branchiostoma japonicum*; bf, *Branchiostoma floridae*; Hl, *Holothuria leucospilota*; Ad, *Acropora digitifera*; Nv, *Nematostella vectensis*; Of, *Orbicella faveolata*; Sk, *Saccoglossus kowalevskii*; Sp, *Strongylocentrotus purpuratus*.

simple, straightforward function such as recognizing microbes in the extracellular space. To investigate their molecular functions, we prepared and purified recombinant His-tagged TRX-CgACP1 and TRX-BgACP1 fusion proteins (Figure 4A). These recombinant proteins were then incubated with several microbes, including two gram-positive bacteria (*S. aureus* and

E. faecalis), four gram-negative bacteria (*E. coli*, *V. anguillarum*, *V. parahaemolyticus*, *A. calcoaceticus*) and a yeast (*S. cerevisiae*). After incubation, the microbial pellets were assessed by Western Blot using anti-His monoclonal antibodies (mAb), which shows that both TRX-CgACP1 and TRX-BgACP1 could bind with all selected microbes in different affinities (Figure 4B).

We then analyzed if the recombinant proteins could form multimers, as suggested by the conserved cysteines in the ApeC domain (Figure 2C). The intermolecular disulfide bonds could be broken down in a strong reducing condition. Indeed, we showed that in the reducing condition, both recombinant TRX-CgACP1 and TRX-BgACP1 were reduced to a single band of monomers, which are corresponding to their respectively molecular weights (Figures 4C, D). However, under a non-reducing condition in which disulfide bonds might be preserved, we observed the bands of monomers and dimers, as well as the high-weight smears which were supposed to be polymers (Figures 4E, F). This suggests that both recombinant TRX-CgACP1 and TRX-BgACP1 were capable of forming multimers through disulfide bonds.

To further investigate whether the microbe-binding and multimerizing ability of BgACP1 and CgACP1 could allow them to aggregate the microbes, we incubated the recombinant proteins with FITC-labeled *S. aureus*, *E. faecalis*, *E. coli*, *V. anguillarum* and yeast *S. cerevisiae*, and then assessed the microbial aggregation by using fluorescence microscopy. The results show that the addition of TRX-CgACP1 and TRX-BgACP1 could cause observable aggregation of all the tested

microbes. Quantification of the diameter of green microbial puncta confirmed the significance of the agglutination effects (Figures 5A–E). These results suggest that both CgACP1 and BgACP1 have broad-spectrum microbial binding and agglutinating capacities.

Recombinant CgACP1 and BgACP1 showed no inhibition on bacterial growth

Some secreted short-form bacteria-binding lectins, such as the human C-type lectin REG3 (40), are able to kill bacteria. To determine whether recombinant CgACP1 and BgACP1 could kill bacteria or interfere their growth, we monitored the growth curves of *S. aureus* and *E. coli* in the presence of the recombinant proteins. Antibiotic ampicillin was used as the positive control. The results showed that both TRX-CgACP1 and TRX-BgACP1 fusion protein had no inhibiting effects on the examined bacteria (Figures 6A, B). Moreover, the growth curves and the colony counting assays confirmed that both recombinant proteins could not kill bacteria *S. aureus* or *E. coli* (Figures 6C, D). As comparison, recombinant amphioxus ACP1/2/3/5 also showed

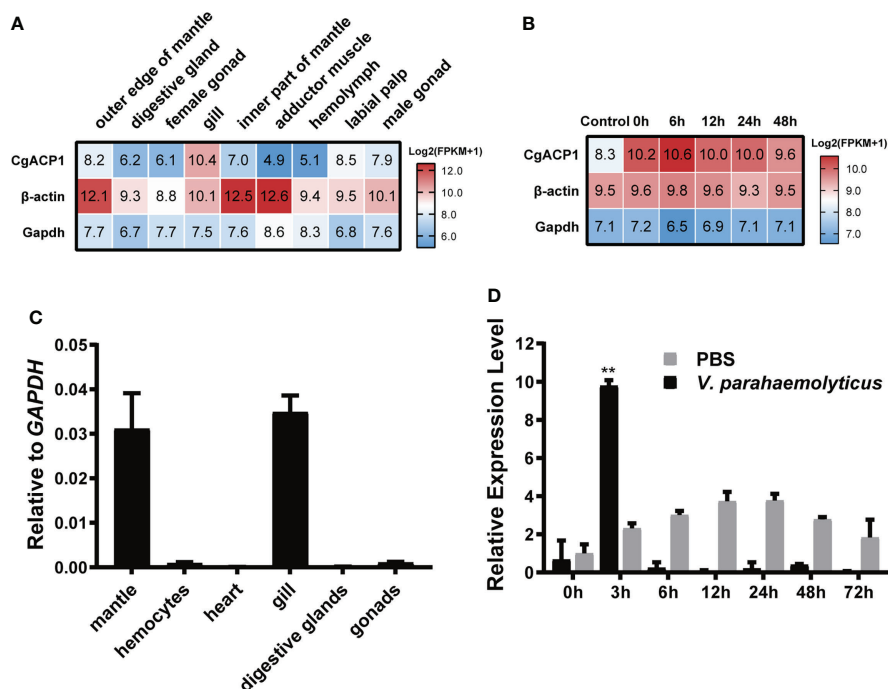


FIGURE 3

The expression patterns of the CgACP1 gene. (A) The CgACP1 mRNA expression in different tissues based on the RNA-seq data from NCBI BioProject PRJNA146329 (31). The raw expression data are provided in Supplementary Table 3. (B) The CgACP1 mRNA expression in the gill before and after *Vibrio* challenge based on the RNA-seq data from NCBI BioProject PRJNA194079 (32). The raw expression data are provided in Supplementary Table 4. (C) Real-time quantitative RT-PCR analysis of the relative mRNA expression of CgACP1 in different tissues. (D) Real-time quantitative RT-PCR analysis of the relative mRNA expression of CgACP1 in the hemocytes before and after *Vibrio* challenge. The qPCR-based expression data were shown as a ratio to the gapdh mRNA expression and were plotted as the mean \pm SD. ** $p < 0.01$ versus mRNA level at 0 h.

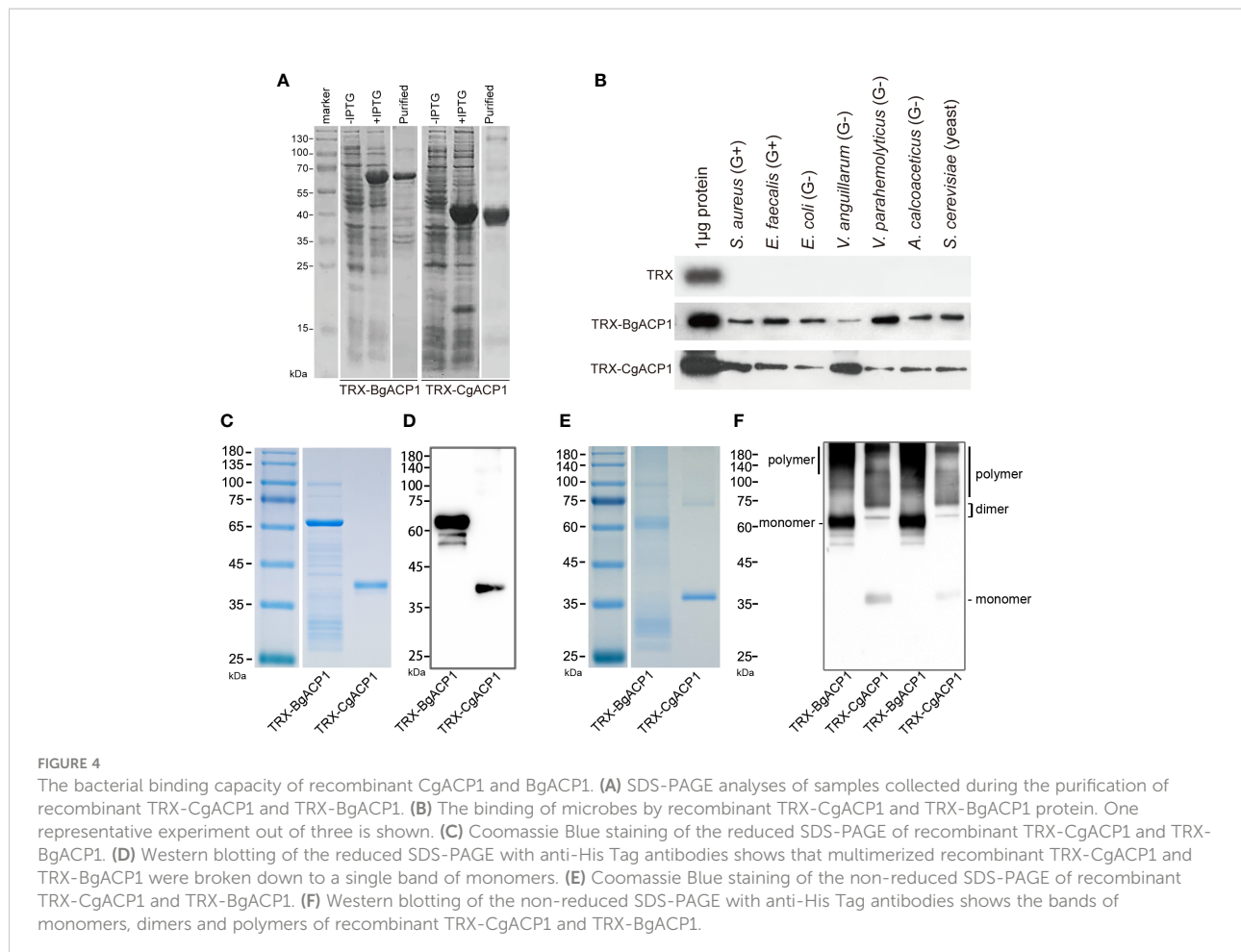
no bactericidal activity (2, 6). We concluded that both CgACP1 and BgACP1 have microbial binding and aggregation activities, but might not have bactericidal and bacteriostatic ability.

Recombinant CgACP1 and BgACP1 specifically recognized peptidoglycan

Previous studies have shown that some amphioxus ApeC domains showed high binding affinity to certain microbial cell wall components (2, 6). To determine which microbial cell wall components could be recognized by CgACP1 and BgACP1, here we implemented ELISA assays as previously described (37). The results show that both recombinant CgACP1 and BgACP1 had much higher binding affinity with peptidoglycan (PGN) than with lipopolysaccharide (LPS), lipoteichoic acid (LTA), Zymosan A, chitin, chitosan and cellulose (Figures 7A, B). PGN have two major structural types, namely, the Lys-type and the DAP-type. PGN from *S. aureus* is Lys-type while PGN from *B. subtilis* is DAP-type. Our ELISA assays further showed that both recombinant ACPs had high affinity to the Lys-type PGN

from *S. aureus* but have trivial or even no affinity to DAP-type PGN from *B. subtilis* (Figures 7A, B).

PGN is a macromolecule composed of several heteropolysaccharides alternately linked by N-acetylglucosamine (GlcNAc), N-acetylmuramic acid (MurNAc) and cross-linked with peptides of different compositions. The PGN breakdown products could stimulate an inflammatory response. The products include GlcNAc and several dipeptide-containing motifs, such as muramyl dipeptide (MDP; MurNAc-L-Ala-D-isoGln) (41). MDP is the minimal bioactive moiety of PGN and the agonist of NOD2, which can activate the NOD2-mediated NF- κ B and NLRP1 inflammasome (42–44). Here we showed that although recombinant CgACP1 and BgACP1 could bind with PGN, they showed no significant affinities to any of the PGN moieties, including MDP, GlcNAc and MurNAc (Figures 7C, D). We speculated that either the optimal affinity of CgACP1 and BgACP1 required a special conformation of PGN, or the recombinant CgACP1 and BgACP1 might not preserve certain binding ability in their native forms. Anyway, the binding specificities of CgACP1 and BgACP1 have both similarities and dissimilarities with those of amphioxus ACP1/2/3/5 (6).



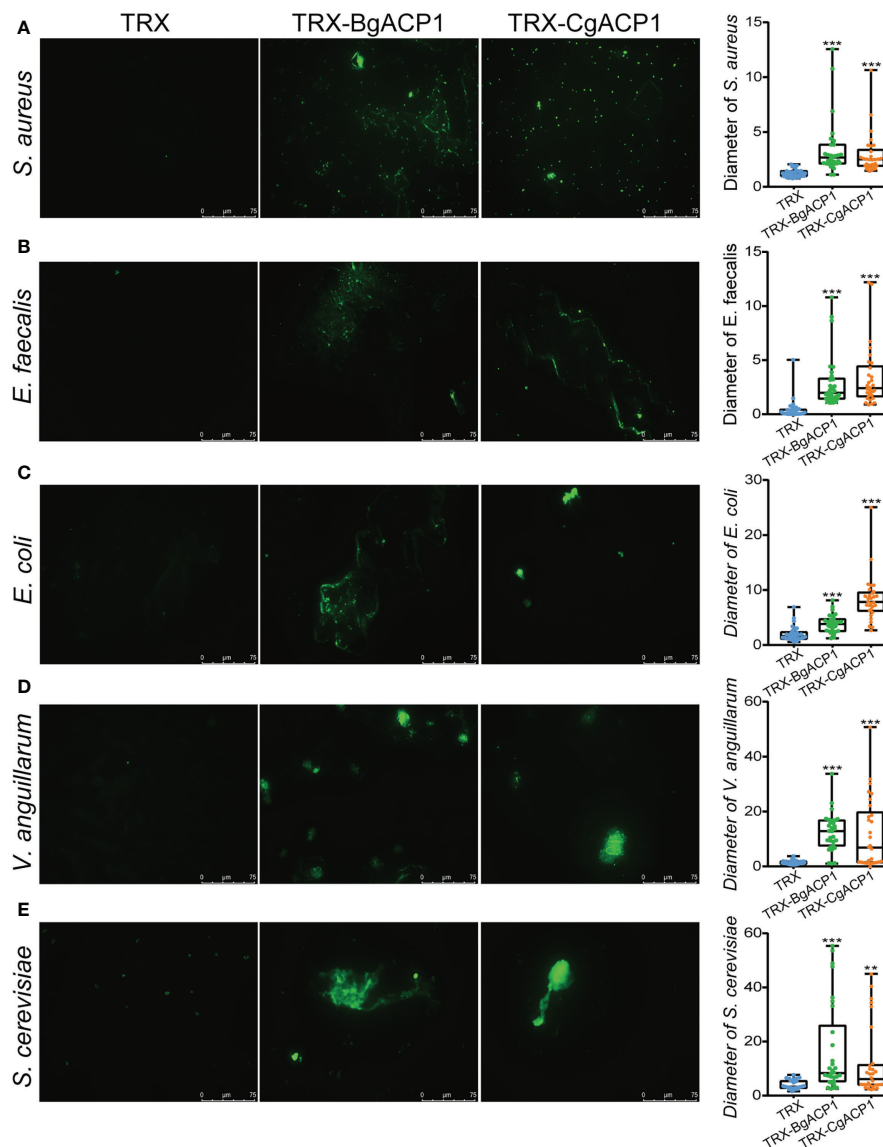


FIGURE 5

The bacterial agglutinating capacity of recombinant CgACP1 and BgACP1. (A–E) Agglutination of the different microbes by TRX-CgACP1 and TRX-BgACP1. The diameters of green puncta in microbial aggregation assays were measured by boxplot and shown on the right panels. Boxplot explanation: upper horizontal bar outside box, maximum value; lower horizontal bar outside box, minimum value; upper horizontal line of box, 75th percentile; lower horizontal line of box, 25th percentile; horizontal bar within box, median; ** $p < 0.01$, *** $p < 0.001$ versus TRX control.

Discussion and conclusions

Distribution, phylogenetics and diversity of the mollusk ACPs

ApeC is an invertebrate-specific protein domain and predominantly present in marine and freshwater invertebrates (5). Among those ApeC domain-containing phyla, *Mollusca* is the largest one. So far, a large number of ACPs (approximately one third of the known ACPs) are from mollusks, most of which

are from gastropods and bivalves, the first and the second largest mollusk classes, respectively. Mollusk ACPs could be classified into at least nine phylogenetic groups. Mollusk ACPs have no orthologs in other phyla, and also appear to be more diversified and more actively duplicated in bivalves than in gastropods. Similar to the situations in amphioxus and other phyla, mollusk ApeC domains are present in various protein architectures, making them the versatile and promiscuous domains like the IG and C-type lectin domains (45, 46). In fact, each mollusk phylogenetic group has its own special protein architecture.

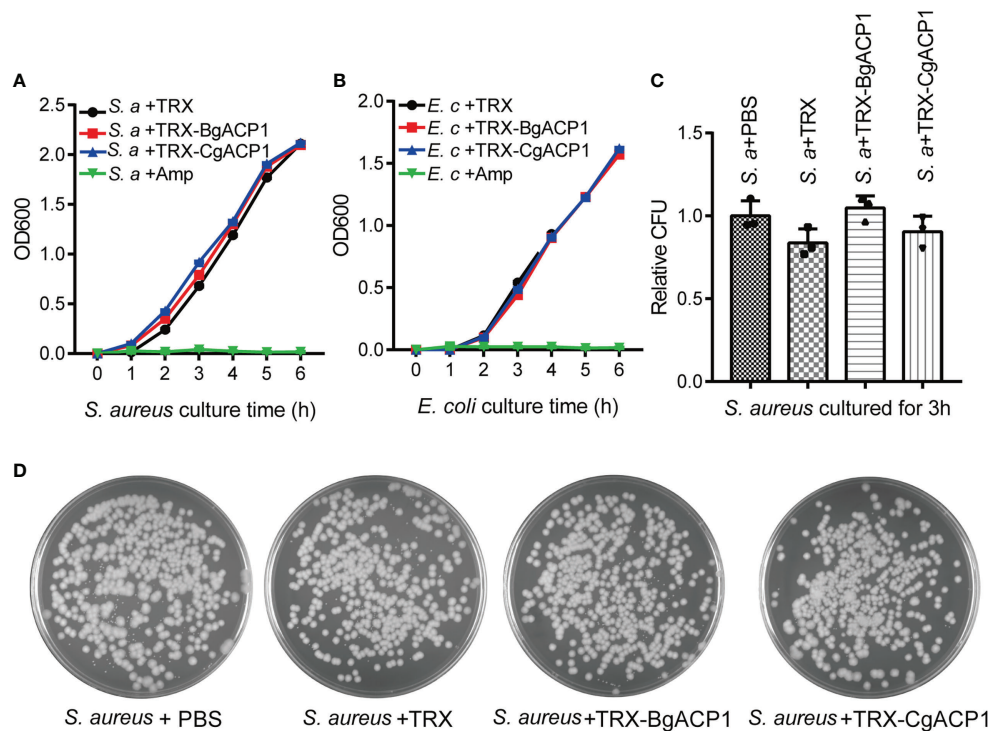


FIGURE 6

Recombinant CgACP1 and BgACP1 showed no inhibiting or killing activity to bacteria *S. aureus* and *E. coli*. Growth curves of *S. aureus* (A) and *E. coli* (B) in the presence of TRX-CgACP1 or TRX-BgACP1. 200 μ g/ml of each ACP, Ampicillin (Amp) or TRX (as control) were incubated with the indicated bacteria and the OD₆₀₀ was measured every 1 hour after starting the culture (mean \pm SD, n = 3). (C) Mollusk ACPs did not affect the colony forming unit (CFU) of *S. aureus*. 200 μ g/ml of each ACP or TRX (control) were incubated with the *S. aureus* for 3 hours and the numbers of bacteria were determined by colony counts. Relative CFU represents the ratio of CFU incubated with TRX/ACP and incubated with PBS. One of the three petri dishes used in (C) is shown in (D).

Nevertheless, these sequence and structural diversities might reflect the diverse roles of ACPs in mollusks.

Molecular functions of the mollusk ApeC domains

So far, the functions of ApeC were only investigated in several highly-expressed, secreted short-form ACPs from the amphioxus (2, 6). To facilitate cross-phylum comparison and to guarantee proper representation, here we selected two highly-expressed, secreted short-form mollusk ACPs for function investigation, with one from bivalves and the other from gastropods. We observed that both recombinant mollusk ACPs (CgACP1 and BgACP1) could bind and aggregate different microbes in different affinities. As comparison, amphioxus ACP1/2/3/5 also showed similar bacterial binding and aggregating capacities, though with different binding spectrum and affinity (2, 6). Moreover, both the two mollusk ACPs and the four amphioxus ACPs exhibited no inhibiting or killing effects on the tested microbes (2, 6).

In addition, we found that recombinant CgACP1 and BgACP1 could specifically bind to Lys-type PGN from *S. aureus* with high affinity, but had trivial or no affinity to DAP-type PGN from *B. subtilis* (Figure 6). As comparison, amphioxus ACP1 and ACP2 showed high affinity to Lys-type PGN and relatively low affinity to DAP-type PGN (2), whereas amphioxus ACP3 and ACP5 showed high affinity to Lys-type PGN but nearly no affinity to DAP-type PGN (6). Despite these difference, Lys-type PGN emerges to be the primary specific binding target of all examined ApeC domains so far.

Although Lys-type PGN had been identified as the main binding target of CgACP1 and BgACP1, both proteins could actually bind with G-negative bacteria and yeast which are supposed to have no Lys-type PGN. This inconsistency might be caused by several factors. First, these ACPs showed certain low level or background level of affinities to DAP-type PGN and other tested substances (Figure 7). Second, ApeC domains may have affinity to other unknown cell wall components. Third, there are differences in terms of the density of the target sites and their accessibility in the cell wall of different microbes, and different ACPs may have different ability to penetrate into the

cell wall. Therefore, even if an ACP could recognize a cell wall component, they might bind and aggregate different bacteria with different strength. Fourth, the binding assays might not be conducted in the optimal conditions for CgACP1 and BgACP1. Nevertheless, the mechanisms underlying the binding specificity of ApeC domains require further research efforts.

The role of short-form ACPs in the mucosal surface of the gill

Mollusk ACPs have been suggested to play important roles in immunity, development and stress resistance (10–12, 15–18). ACPs were even discovered in the pedal mucus of the limpets, suggesting a role in the glue-like adhesion (19). However, so far, there are no direct evidence to support any immune functions for the mollusk ApeC domains (10, 11, 15, 17). On the other hand, there are direct evidence from the amphioxus. Previous studies showed that the amphioxus ACP1 had the highest expression among all amphioxus ACPs and was mainly expressed in the gill, which is the filter feeding organ of amphioxus. ACP1 could be further up-regulated by hundreds of times in response to bacterial infection, and has been shown to act as an essential mucus lectin to defend the gill surface (2, 3, 47). In the normal physiological condition, if amphioxus ACP1 protein was neutralized, it would cause the disintegration of the gill due to bacterial invasion (2).

As comparison, here we showed that in pacific oyster, CgACP1 had the highest expression level of all oyster ACPs. CgACP1 was also predominantly expressed in the gill and labial palp in a very high level which is comparable to that of the beta-actin. This expression level could be greatly up-regulated quickly after bacterial infection. Remarkably, the gill and the labial palp are adjacent to each other and comprise the filter feeding organ for oysters, which is contacting the water environment, and continuously and non-selectively sieving the running water for small particles, including algae, bacteria, other food particles and even pathogens. In addition, CgACP1 and amphioxus ACP1 are both typical short-form ACPs and have similar binding properties. In the sense, CgACP1 could play a similar role in the gill and in the filter feeding process as the amphioxus ACP1.

Basic functional properties of the ApeC domains

In this study, we provided a survey of the mollusk ACPs in terms of their species distribution, composition, phylogenetics, structures and expression patterns. Based on this survey, we selected two highly-expressed, secreted short-form ACPs as the representative mollusk ACPs for the further functional survey. By comparing the findings in this study and in the previous studies of amphioxus ACP1/2/3/5, and by referring to the indirect evidence from other related studies, we could identify several functional traits of the ApeC

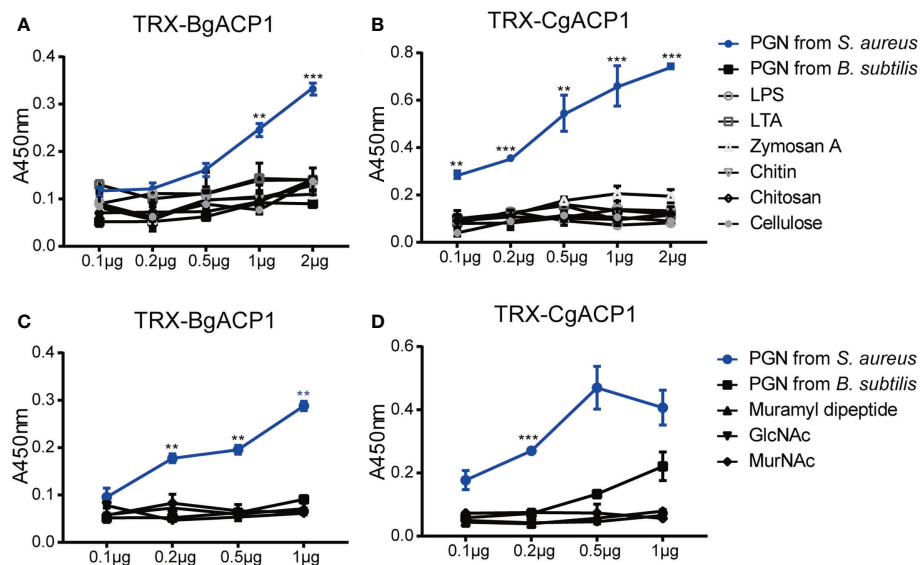


FIGURE 7

The carbohydrate-binding specificity and affinity of recombinant CgACP1 and BgACP1. (A, B) ELISA assays of the interaction between recombinant fusion TRX-CgACP1 and TRX-BgACP1 to different bacterial cell wall components. (C, D) ELISA assays of the interaction between recombinant fusion ACPs and several moieties of PGN. GlcNAc, N-acetylglucosamine; MurNAc, N-acetylmuramic acid. ** $p < 0.01$, *** $p < 0.001$. Three biological replicates were conducted for each experiment, and three technical replicates were performed. One of the representative results was shown here.

domains. These traits have been shown to be conserved across different phyla, including the high affinity to Lys-type PGN, the bacterial binding and agglutinating capacity, and the role as mucus lectin-type pattern-recognition proteins in the mucosal surface. In conclusion, this may not only extend our understanding of the mollusk's immune diversity, but guide our future researches of the ACP functions in mollusks as well as in other animal clades.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

Author contributions

SH, JL and AX designed the study. JL performed the experiments. JL, SL, YZ, FM, HF and QH analyzed the data. HZ, JO, WY, FM and MD also conducted some experiments and provided reagents, materials and some data explanation. JL, SH and AX drafted the manuscript. All authors discussed and agreed on the results and approved the manuscript.

Funding

This work was supported by the project from Southern Marine Science and Engineering Guangdong Laboratory

References

- Huang S, Chen Z, Yan X, Yu T, Huang G, Yan Q, et al. Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun* (2014) 5:5896. doi: 10.1038/ncomms6896
- Huang G, Huang S, Yan X, Yang P, Li J, Xu W, et al. Two apextrin-like proteins mediate extracellular and intracellular bacterial recognition in amphioxus. *Proc Natl Acad Sci USA* (2014) 111:13469–74. doi: 10.1073/pnas.1405414111
- Huang G, Liu H, Han Y, Fan L, Zhang Q, Liu J, et al. Profile of acute immune response in Chinese amphioxus upon staphylococcus aureus and vibrio parahaemolyticus infection. *Dev Comp Immunol* (2007) 31:1013–23. doi: 10.1016/j.dci.2007.01.003
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* (2006) 439:965–8. doi: 10.1038/nature04336
- Li Y, Li J, Yan X, Chen S, Wu C, Huang H, et al. Broad distribution, high diversity and ancient origin of the ApeC-containing proteins. *Mol Phylogenet Evol* (2021) 155:107009. doi: 10.1016/j.ympev.2020.107009
- Li J, Li Y, Fan Z, Chen S, Yan X, Yue Z, et al. Two amphioxus ApeC-containing proteins bind to microbes and inhibit the TRAF6 pathway. *Front Immunol* (2021) 12:715245. doi: 10.3389/fimmu.2021.715245
- Haag ES, Raff RA. Isolation and characterization of three mRNAs enriched in embryos of the direct-developing sea urchin *heliocidaris erythrogramma*: Evolution of larval ectoderm. *Dev Genes Evol* (1998) 208:188–204. doi: 10.1007/s004270050173
- Haag ES, Sly BJ, Andrews ME, Raff RA. Apextrin, a novel extracellular protein associated with larval ectoderm evolution in *heliocidaris erythrogramma*. *Dev Biol* (1999) 211:77–87. doi: 10.1006/dbio.1999.9283
- Minsuk SB, Raff RA. Pattern formation in a pentamerous animal: Induction of early adult rudiment development in sea urchins. *Dev Biol* (2002) 247:335–50. doi: 10.1006/dbio.2002.0704
- Jiang K, Nie H, Yin Z, Yan X, Li Q. Apextrin from *ruditapes philippinarum* functions as pattern recognition receptor and modulates NF- κ B pathway. *Int J Biol Macromolecules* (2022) 214:33–44. doi: 10.1016/j.ijbiomac.2022.06.041
- Estevez-Calvar N, Romero A, Figueras A, Novoa B. Involvement of pore-forming molecules in immune defense and development of the Mediterranean mussel (*Mytilus galloprovincialis*). *Dev Comp Immunol* (2011) 35:1017–31. doi: 10.1016/j.dci.2011.03.023
- David E, Tanguy A, Pichavant K, Moraga D. Response of the pacific oyster *crassostrea gigas* to hypoxia exposure under experimental conditions. *FEBS J* (2005) 272:5635–52. doi: 10.1111/j.1742-4658.2005.04960.x
- Dheilly NM, Haynes PA, Bove U, Nair SV, Raftos DA. Comparative proteomic analysis of a sea urchin (*Heliocidaris erythrogramma*) antibacterial response revealed the involvement of apextrin and calreticulin. *J Invertebr Pathol* (2011) 106:223–9. doi: 10.1016/j.jip.2010.09.008
- Gerdol M, Luo YJ, Satoh N, Pallavicini A. Genetic and molecular basis of the immune system in the brachiopod *lingula anatina*. *Dev Comp Immunol* (2018) 82:7–30. doi: 10.1016/j.dci.2017.12.021

(Zhuhai) (SML2021SP304), National Key R&D Program of China (2018YFD0900503), National Science Foundation of China (NNSF) Projects (31872595, 32073002 and 31722052), projects from Guangdong and Guangzhou (2021A1515012380 and 2020B1212060031), and Fundamental Research Funds for the Central Universities (Sun Yat-sen University; 22lgj09).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.971883/full#supplementary-material>

15. McDowell IC, Modak TH, Lane CE, Gomez-Chiarri M. Multi-species protein similarity clustering reveals novel expanded immune gene families in the eastern oyster *Crassostrea virginica*. *Fish Shellfish Immunol* (2016) 53:13–23. doi: 10.1016/j.fsi.2016.03.157
16. Moreira R, Balseiro P, Forn-Cuni G, Milan M, Bargelloni L, Novoa B, et al. Bivalve transcriptomics reveal pathogen sequences and a powerful immune response of the Mediterranean mussel (*Mytilus galloprovincialis*). *Mar Biol* (2018), 165:61. doi: 10.1007/s00227-018-3308-0
17. Jiang K, Yin Z, Zhang Y, Xu Q, Yu Y, Cong W, et al. Genome-wide investigation and expression analysis of MACPF gene family reveals its immune role in response to bacterial challenge of Manila clam. *Genomics* (2021) 113:1136–45. doi: 10.1016/j.ygeno.2021.02.013
18. Gerdol M, Venier P. An updated molecular basis for mussel immunity. *Fish Shellfish Immunol* (2015) 46:17–38. doi: 10.1016/j.fsi.2015.02.013
19. Kang V, Lengerer B, Wattiez R, Flammang P. Molecular insights into the powerful mucus-based adhesion of limpets (*Patella vulgata* L.). *Open Biol* (2020) 10:18. doi: 10.1098/rsob.200019
20. Glaubrecht M. Phylogeny and evolution of the mollusca. *Science* (2008) 320:1014–5. doi: 10.1126/science.1158066
21. Bar-On YM, Phillips R, Milo R. The biomass distribution on earth. *Proc Natl Acad Sci USA* (2018) 115:6506–11. doi: 10.1073/pnas.1711842115
22. Gerdol M. Immune-related genes in gastropods and bivalves: A comparative overview. *Isj-Invertebrate Survival J* (2017) 14:103–18.
23. Nicol D. The number of living species of molluscs. *Systematic Biol* (1969) 18:251–4. doi: 10.2307/2412618
24. Lin Y, Mao F, Wong NK, Zhang XY, Liu KN, Huang MW, et al. Phagocyte transcriptomic analysis reveals focal adhesion kinase (FAK) and heparan sulfate proteoglycans (HSPGs) as major regulators in anti-bacterial defense of *Crassostrea hongkongensis*. *Front Immunol* (2020), 11:416. doi: 10.3389/fimmu.2020.00416
25. Zhang XY, Mao F, Wong NK, Bao YB, Lin Y, Liu KN, et al. CLIC2 alpha chloride channel orchestrates immunomodulation of hemocyte phagocytosis and bactericidal activity in *Crassostrea gigas*. *Iscience* (2020) 23:101328. doi: 10.1016/j.isci.2020.101328
26. Mao F, Liu K, Wong NK, Zhang X, Yi W, Xiang Z, et al. Virulence of *Vibrio alginolyticus* accentuates apoptosis and immune rigor in the oyster *Crassostrea hongkongensis*. *Front Immunol* (2021) 12:746017. doi: 10.3389/fimmu.2021.746017
27. Kumar S, Stecher G, Li M, Niyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* (2018) 35:1547–9. doi: 10.1093/molbev/msy096
28. Nicholas K, Nicholas HB. GeneDoc: A tool for editing and annotating multiple sequence alignments. (1997).
29. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* (1987) 4:406–25. doi: 10.1093/oxfordjournals.molbev.a040454
30. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* (1992) 8:275–82. doi: 10.1093/bioinformatics/8.3.275
31. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* (2012) 490:49–54. doi: 10.1038/nature11413
32. Zhang L, Li L, Guo X, Litman GW, Dishaw LJ, Zhang G. Massive expansion and functional divergence of innate immune genes in a protostome. *Sci Rep* (2015) 5:8693. doi: 10.1038/srep08693
33. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* (2014) 30(15):2114–20. doi: 10.1093/bioinformatics/btu170
34. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SA-O. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nat Protoc* (2016) 11(9):1650–67. doi: 10.1038/nprot.2016.095
35. Yu XQ, Tracy ME, Ling E, Scholz FR, Trenczek T. A novel c-type immunlectin-3 from *Manduca sexta* is translocated from hemolymph into the cytoplasm of hemocytes. *Insect Biochem Mol Biol* (2005) 35:285–95. doi: 10.1016/j.ibmb.2005.01.004
36. Selsted ME, Szklarek D, Lehrer RI. Purification and antibacterial activity of antimicrobial peptides of rabbit granulocytes. *Infection Immun* (1984) 45:150–4. doi: 10.1128/iai.45.1.150-154.1984
37. Yu Y, Yu Y, Huang H, Feng K, Pan M, Yuan S, et al. A short-form c-type lectin from amphioxus acts as a direct microbial killing protein via interaction with peptidoglycan and glucan. *J Immunol (Baltimore Md 1950)* (2007) 179:8425–34. doi: 10.4049/jimmunol.179.12.8425
38. Wang L, Song X, Song L. The oyster immunity. *Dev Comp Immunol* (2018) 80:99–118. doi: 10.1016/j.dci.2017.05.025
39. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat Commun* (2017) 8:15451. doi: 10.1038/ncomms15451
40. Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* (2006) 313:1126–30. doi: 10.1126/science.1127119
41. Kieser KJ, Kagan JC. Multi-receptor detection of individual bacterial products by the innate immune system. *Nat Rev Immunol* (2017) 17:376–90. doi: 10.1038/nri.2017.25
42. Grimes CL, Ariyananda LDZ, Melnyk JE, O'Shea EK. The innate immune protein Nod2 binds directly to MDP, a bacterial cell wall fragment. *J Am Chem Soc* (2012) 134:13535–7. doi: 10.1021/ja303883c
43. Hsu LC, Ali SR, McGillivray S, Tseng PH, Mariathasan S, Humke EW, et al. A NOD2-NALP1 complex mediates caspase-1-dependent IL-1beta secretion in response to bacillus anthracis infection and muramyl dipeptide. *Proc Natl Acad Sci USA* (2008) 105:7803–8. doi: 10.1073/pnas.0802726105
44. Girardin SE, Boneca IG, Viala J, Chamaillard M, Labigne A, Thomas G, et al. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* (2003) 278:8869–72. doi: 10.1074/jbc.C200651200
45. Drickamer K, Taylor ME. Recent insights into structures and functions of c-type lectins in the immune system. *Curr Opin Struct Biol* (2015) 34:26–34. doi: 10.1016/j.sbi.2015.06.003
46. Williams AF, Barclay AN. The immunoglobulin superfamily—domains for cell surface recognition. *Annu Rev Immunol* (1988) 6:381–405. doi: 10.1146/annurev.iy.06.040188.002121
47. Huang S, Wang X, Yan Q, Guo L, Yuan S, Huang G, et al. The evolution and regulation of the mucosal immune complexity in the basal chordate amphioxus. *J Immunol* (2011) 186:2042–55. doi: 10.4049/jimmunol.1001824