

## Characterization of a defensin from the oyster *Crassostrea gigas* - Recombinant production, folding, solution structure, antimicrobial activities, and gene expression

Yannick Gueguen<sup>1</sup>, Amaury Herpin<sup>2,5</sup>, André Aumelas<sup>3</sup>, Julien Garnier<sup>1</sup>, Julie Fievet<sup>1</sup>, Jean-Michel Escoubas<sup>1,&</sup>, Philippe Bulet<sup>4</sup>, Marcelo Gonzalez<sup>1</sup>, Christophe Lelong<sup>2</sup>, Pascal Favrel<sup>2</sup>, Evelyne Bachère<sup>1</sup>

<sup>1</sup> Ifremer-CNRS-Université de Montpellier II ; UMR 5171 – Génome Populations Interactions Adaptation, 2 Place E. Bataillon, CC80, F-34095 Montpellier cedex 5, France ;

<sup>2</sup> Laboratoire de Biologie et Biotechnologies Marines, IBFA, UMR 100 IFREMER-Université de Caen, Physiologie et Ecophysiologie des Mollusques Marins, 14032 Caen cedex, France ;

<sup>3</sup> Centre de Biochimie Structurale, CNRS UMR 5048, INSERM U554, CNRS UMR5048, Université Montpellier I, 29 rue de Navacelles, F-34090 Montpellier Cedex, France ;

<sup>4</sup> Atheris Laboratories, Case Postale 314, CH-1233 Bernex-Geneva, Switzerland ;

<sup>5</sup> Sars International Centre for Marine Molecular Biology, High Technology Centre, 5008 Bergen, Norway.

\*: Corresponding author : [ygueguen@ifremer.fr](mailto:ygueguen@ifremer.fr)

### Abstract:

In invertebrates, defensins were found in arthropods and in the mussels. Here, we report for the first time the identification and characterization of a defensin (*Cg-Def*) from an oyster. *Cg-def* mRNA was isolated from *Crassostrea gigas* mantle using an expressed sequence tag approach. To gain insight into potential roles of *Cg-Def* in oyster immunity, we produced the recombinant peptide in *Escherichia coli*, characterized its antimicrobial activities, determined its solution structure by NMR spectroscopy, and quantified its gene expression *in vivo* following bacterial challenge of oysters. Recombinant *Cg-Def* was active *in vitro* against Gram-positive bacteria but showed no or limited activities against Gram-negative bacteria and fungi. The activity of *Cg-Def* was retained *in vitro* at a salt concentration similar to that of seawater. The *Cg-Def* structure shares the so-called cystine-stabilized  $\alpha$ - $\beta$  motif (CS- $\alpha\beta$ ) with arthropod defensins but is characterized by the presence of an additional disulfide bond, as previously observed in the mussel defensin (MGD-1). Nevertheless, despite a similar global fold, the *Cg-Def* and MGD-1 structures mainly differ by the size of their loops and by the presence of two aspartic residues in *Cg-Def*. Distribution of *Cg-def* mRNA in various oyster tissues revealed that *Cg-def* is mainly expressed in mantle edge where it was detected by mass spectrometry analyses. Furthermore, we observed that the *Cg-def* messenger concentration was unchanged after bacterial challenge. Our results suggest that *Cg-def* gene is continuously expressed in the mantle and would play a key role in oyster by providing a first line of defense against pathogen colonization.

# CHARACTERIZATION OF A DEFENSIN FROM THE OYSTER *CRASSOSTREA GIGAS*: RECOMBINANT PRODUCTION, FOLDING, SOLUTION STRUCTURE, ANTIMICROBIAL ACTIVITIES AND GENE EXPRESSION \*

Yannick Gueguen<sup>1#</sup>, Amaury Herpin<sup>2,5#</sup>, André Aumelas<sup>3</sup>, Julien Garnier<sup>1</sup>, Julie Fievet<sup>1</sup>, Jean-Michel Escoubas<sup>1,&</sup>, Philippe Bulet<sup>4</sup>, Marcelo Gonzalez<sup>1</sup>, Christophe Lelong<sup>2</sup>, Pascal Favrel<sup>2</sup>, Evelyne Bachère<sup>1</sup>

From the <sup>1</sup> Ifremer-CNRS-Université de Montpellier II ; UMR 5171 – Génome Populations Interactions Adaptation, 2 Place E. Bataillon, CC80, F-34095 Montpellier cedex 5, France ; the <sup>2</sup> Laboratoire de Biologie et Biotechnologies Marines, IBFA, UMR 100 IFREMER-Université de Caen, Physiologie et Ecophysiologie des Mollusques Marins, 14032 Caen cedex, France ; the <sup>3</sup> Centre de Biochimie Structurale, CNRS UMR 5048, INSERM U554, CNRS UMR5048, Université Montpellier I, 29 rue de Navacelles, F-34090 Montpellier Cedex, France ; the <sup>4</sup> Atheris Laboratories, Case Postale 314, CH-1233 Bernex-Geneva, Switzerland ; the <sup>5</sup> Sars International Centre for Marine Molecular Biology, High Technology Centre, 5008 Bergen, Norway.

Running title: *Crassostrea gigas* oyster defensin

Address correspondence to Yannick Gueguen, Ifremer-CNRS-Université de Montpellier II ; UMR 5171 – Génome Populations Interactions Adaptation, 2 Place E. Bataillon, CC80, F-34095 Montpellier cedex 5, France. E-Mail : [ygueguen@ifremer.fr](mailto:ygueguen@ifremer.fr)

In invertebrates, defensins were found in arthropods and in the mussels. Here, we report for the first time the identification and characterization of a defensin (*Cg-Def*) from an oyster. *Cg-def* mRNA was isolated from *Crassostrea gigas* mantle using EST approach. To gain insight into potential roles of *Cg-Def* in oyster immunity, we produced the recombinant peptide in *Escherichia coli*, characterized its antimicrobial activities, determined its solution structure by NMR spectroscopy and quantified its gene expression *in vivo* following bacterial challenge of oysters. Recombinant *Cg-Def* was active *in vitro* against Gram-positive bacteria but showed no or limited activities against Gram-negative bacteria and fungi. The activity of *Cg-Def* was retained *in vitro* at a salt concentration similar to that of sea water. The *Cg-Def* structure shares the so-called Cysteine-Stabilized alpha-beta motif (CS- $\alpha\beta$ ) with arthropod defensins, but is characterized by the presence of an additional disulfide bond, as previously observed in the mussel defensin (MGD-1). Nevertheless, despite a similar global fold, the *Cg-Def* and MGD-1 structures mainly differ by the size of their loops and by the presence of two aspartic residues in *Cg-Def*. Distribution of *Cg-def* mRNA in various oyster tissues revealed that *Cg-def* is mainly expressed in mantle edge where it was detected by mass spectrometry analyses. Furthermore, we observed that the *Cg-def*

messenger concentration was unchanged after bacterial challenge. Our results suggest that *Cg-def* gene is continuously expressed in the mantle and would play a key role in oyster by providing a first line of defence against pathogen colonization.

## INTRODUCTION

Antimicrobial peptides (AMPs) are important components of the innate immune system that have been conserved during evolution (1). They constitute a first line of host defence against pathogens in plants and animals (2, 3). We estimate that more than 1,000 antimicrobial peptides have been described at the level of their primary structure (2). They are gathered in the Antimicrobial Sequence Database (<http://www.bbcm.units.it/~tossi/amsdb.html>). AMPs can be classified into three major groups: (i) linear peptides that form amphipathic  $\alpha$ -helices, (ii) cyclic peptides containing cysteine-residue engaged in disulfide bonds and (iii) peptides with an overrepresentation in certain amino acids (proline, arginine, glycine or histidine). Despite their great diversity in terms of size, primary structure, amino acid composition and mode of action, most AMPs are characterized by the preponderance in cationic and hydrophobic amino acids (2). In most of the cases, this amphipathic character is considered as crucial for the interaction of the effective peptide with the membrane of sensitive micro-organisms. This first interaction

seems to be essential whatever the exact mode of action, (i) through disruption of their negatively charged cytoplasmic membranes or (ii) through killing following translocation into the bacteria without membrane lyses and binding to a specific target protein (4). Depending on their tissue distribution, AMPs ensure either a systemic or a local protection of the organism against pathogens.

Among the AMPs, defensins represent an important peptide family. They are abundant and widely distributed in human and animal tissues that are involved in host defence against microbial infection. Defensins are compact cationic peptides, approximately 3-5 kDa in size, containing three or four disulfide bridges, and are active against a wide range of bacteria and fungi (5). The vertebrate defensins can be grouped into three subfamilies, the  $\alpha$ -defensins and  $\beta$ -defensins, which are distinguished on the basis of the connectivity of their six cysteine residues, and more recently the cyclic  $\theta$ -defensins, (6). The  $\alpha$ -defensins are produced constitutively and stored in neutrophils of many animals and in human Paneth cells.  $\beta$ -defensins, that have been identified in many cell types, including epithelial cells and neutrophils, were reported to be inducible or constitutively expressed (3, 7, 8). In mammals, apart from their antimicrobial activities, defensins play an important role in inflammation, wound healing and regulation of specific immunity reactions (9). In contrast with the classification of the vertebrate defensins, based on their secondary structure, the grouping in clear distinct subfamilies of the invertebrate defensins is based on their biological properties, antibacterial *versus* antifungal (2). The invertebrate defensins differ from the vertebrate defensins by their disulfide bridging, (10). Defensins are the most widespread family of invertebrate AMPs and more than 70 different defensins have been isolated in arthropods (insects, ticks, spiders, scorpions) and mollusks (2). Most of the insect defensins were isolated from the hemolymph of experimentally infected animals, whereas in scorpions, termites and mollusks, defensins are present in granular hemocytes of non-infected animals (2). In mollusks, AMPs have only been reported in bivalves, such as in the mussels *Mytilus edulis* (11) and *M. galloprovincialis* (12). Interestingly, defensins from the Mediterranean mussels have been found to display sequence homology to defensins from arthropods, even though they belong to distant phylogenetic groups. Until now, in oyster, antimicrobial activities have been detected in the hemolymph of some species,

however, AMPs have never been fully characterized despite many attempts to purify from hemolymph and other tissues such molecules by RP-HPLC approaches (13).

In the present study, we report the characterization of the first AMP in oyster. The *Crassostrea gigas* defensin (*Cg-Def*) mRNA has been isolated from mantle edge using the Expressed Sequence Tag approach. *Cg-Def* displays sequence homology with members of the arthropod defensin family and shares with the defensins (MGD-1 and -2) from the mussel *M. galloprovincialis*, a fourth pair of cysteine residues. To shed light into the evolutionary conservation of defensins in invertebrates, and the function of this peptide in the defence reactions of the oyster, we expressed *Cg-Def* in *Escherichia coli*, and determined using the refolded peptide, its spectrum of activity against a panel of bacteria and fungi. To draw some structure-function features, the 3D structure of the *Cg-Def* was determined in aqueous solution by  $^1\text{H}$  NMR spectroscopy and molecular modelling, and then compared to MGD-1, its counterpart from mussels. Finally, the *Cg-Def* gene expression was analyzed in response to experimental bacterial challenge of oysters.

## EXPERIMENTAL PROCEDURES

**Animals and RT-PCR Amplification of *Cg-def***—Adult oysters, *Crassostrea gigas*, were purchased from a local oyster farm in Normandie (France) or Palavas (Gulf of Lion, France) and kept in sea water at 15°C. A cDNA encoding part of a putative defensin was identified by randomly sequencing clones from a *C. gigas* mantle edge cDNA library constructed in  $\lambda$ -ZAP II (University of Caen). Specific primers were then used to isolate the full-length *Cg-def* cDNA from the same library. Briefly, the 5' missing end of the transcript was identified using the orientated *C. gigas* mantle edge cDNA library as template and using gene specific oligonucleotide (*Cg-Def*R1: 5'-ACCAGAGCGTGGCTGCATCACAG-3') and vector specific oligonucleotide (T3: 5'-AATTAACCCTCACTAAAGG-3') as primers. PCR products were analyzed by electrophoresis on 1.5% (w/v) agarose gel. Fragment of the expected size was extracted from the gel using Qiagen Kit, cloned into the pGEM-T easy using TA cloning kit (Promega, Madison, USA) and sequenced.

### **Screening of *C. gigas* Genomic Library**

A genomic library of *C. gigas* was constructed in  $\lambda$ -DASH<sub>II</sub> (Stratagene, La Jolla, USA). A total of  $1.8 \times 10^6$  independent clones were recovered. After amplification to  $4.5 \times 10^{10}$  pfu/mL,  $2.5 \times 10^5$  recombinant  $\lambda$ -DASH phages were plated at  $5 \times 10^4$  per dish, adsorbed to nitrocellulose membranes and screened at high stringency with digoxigenin-11-dUTP labelled 322 bp encoding the full length *Cg-def* cDNA. Positive clones were purified (Qiagen  $\lambda$  Kit, Courtaboeuf, France) subjected to restriction analysis and Southern blot hybridization using the original probe to confirm that the  $\lambda$ -DASH clones contained *Cg-def* specific sequence. Genomic organization of positive clones was subsequently determined by direct sequencing. Exon/intron boundaries were determined by comparing the genomic sequence to the corresponding cDNA one.

### **Determination of *Cg-def* Expression in Different Tissues**

Total RNA was isolated from adult tissues using Tri-Reagent (Sigma-Aldrich, Lyon, France) according to the manufacturer's instructions. After treatment during 20 min at 37°C with 1U of DNase I (Sigma-Aldrich) to prevent genomic DNA contamination, 1 $\mu$ g of total RNA was reverse transcribed using 1 $\mu$ g of random hexanucleotidic primers (Promega, Madison, USA), 0.5 mM dNTPs and 200U M-MLV Reverse Transcriptase (Promega) at 37°C for 1 h in the appropriate buffer. The reaction was stopped by incubation at 70°C for 10 min. Then, quantitative RT-PCR analysis was performed using the iCycler apparatus (Bio-Rad, Hercules, USA). iQ<sup>TM</sup> SYBR Green supermix PCR kit (Biorad) was used for real time monitoring of amplification (5 ng of template cDNA, 40 cycles: 95°C/15s, 60°C/15s) with the following primers: QsDef 5'-CCACAATCACTGCAAGTCCATTA-3' and QaDef 5'-CTTCCATTACAATCGGTACATG-3' as sense and antisense primers, respectively. Accurate amplification of the target amplicon was checked by performing a melting curve. Using QsGAPDH (5'-TTCTCTTGCCCTCTTGC-3') and QaGAPDH (5'-CGCCAATCCTTGTTGCTT-3') primers, a parallel amplification of oyster GAPDH transcript (EMBL CGI548886) was carried out to normalize the expression data of the *Cg-def* transcript. The relative level of *Cg-def* expression is calculated for 100 copies of the GAPDH housekeeping gene following the formula:  $N=100 \times 2^{(Ct_{GAPDH} - Ct_{Cg-Def})}$ .

### **Recombinant Expression of *Cg-Def***

Recombinant *Cg-Def* was expressed in *E. coli* as N-terminal His<sub>6</sub>-tagged fusion protein using the pET-28a system (Novagen, Madison, USA). By PCR amplification, a Met-coding triodeoxynucleotide was incorporated 5' of the *Cg-def* cDNA and cloned in-frame with the N-terminal His<sub>6</sub> in the *EcoRI/SalI* sites of pET-28a. The *Cg-def* coding cDNA sequence was amplified by PCR using forward primer DefFw1 (5'-GCGCGAATTCATGGGATTTGGGTGTCCGGG) paired with reverse primer DefRv1 (5'-ATATATGTCGACTTACTTCTTTCCATTACAATC GG). The reaction was performed by incubating the reaction mixtures at 94°C for 4 min, followed by successive cycles at 94°C for 1 min, 55°C for 1 min and 72°C for 1 min for 26 cycles using Isis<sup>TM</sup> DNA polymerase (Qbiogene). The underlined codon in the forward primer denotes a Met codon to incorporate a CNBr cleavage site immediately upstream of the N-terminus of the designed peptide according to the method described for recombinant mouse  $\alpha$ -defensin production (14).

Recombinant *Cg-Def* was expressed in *E. coli* Rosetta (DE3) pLysS cells (Novagen) transformed with the pET-28a/*Cg-Def* construct. The cells were grown at 37°C to OD<sub>600</sub> 0.9 in Luria-Bertani (LB) medium (10 g of bacto tryptone, 5 g of bacto yeast extract 10 g of NaCl) supplemented with 50  $\mu$ g/ml of kanamycin. Expression of fusion proteins was induced with 0.5mM isopropyl- $\beta$ -D-1-thiogalactopyranoside. After growth at 37°C for 3 h, bacterial cells were harvested by centrifugation and stored at -20°C. The cells were lysed by resuspending bacteria pellets in 6 M guanidine HCl in 100 mM Tris-HCl, pH 8.1, followed by sonication at 40% amplitude for 2 min using a Vibra cell Sonifier 450 (Branson Ultrasonics, Annemasse, France). The lysate was clarified by centrifugation in a Sorvall SA-600 rotor at 10,000g for 30 min at 4°C prior to protein purification.

### **Purification and Folding of Recombinant *Cg-Def***

His-tagged *Cg-Def* fusion protein was purified by affinity chromatography by incubating cell lysates with nickel-nitrilotriacetic acid resin (Novagen) at a ratio of 25:1 (v/v) in 6 M guanidine HCl, 20 mM Tris-HCl (pH 8.1) for 4 h at 4 °C. Fusion proteins were eluted with two column volumes of 6 M guanidine HCl, 1 M imidazole, 20 mM Tris-HCl (pH 6.4) dialyzed against 5% acetic acid (HOAc) in SpectraPor dialysis membranes (Spectrum Laboratories Inc., Rancho Dominguez, USA) and

lyophilized. The methionine residue introduced at the Cg-Def N-terminus was subjected to CNBr cleavage by dissolving the lyophilized His<sub>6</sub> fusion proteins in 50% formic acid, addition of solid CNBr to 10 mg/ml (final concentration), and incubation of the mixtures for 8 h in the darkness at 25°C. The cleavage reaction was terminated by adding 10 volumes of water, followed by freezing and lyophilization. Then, the cleaved fusion peptide mixture was reduced in 100 mM Tris-HCl buffer at pH 8 in the presence of 100 mM DTT. The reaction mixture was purified to homogeneity using a C<sub>8</sub> reverse-phase high performance liquid chromatography (RP-HPLC) and the fractions of interest lyophilized. The peptide was resolved with a 0-55% acetonitrile gradient developed over 30 min at a flow rate of 5 ml/min on a UP10C8 column (Interchrom modulocart uptisphere 10 C<sub>8</sub>, 250 x 10 mm). The refolding of the Cg-Def was performed at room temperature in 100 mM Tris-HCl buffer at pH 8 during 48 h. Then, the refolded Cg-Def was purified to homogeneity using an additional RP-HPLC step using the same column and eluting conditions as mentioned above. Alternatively, the cleaved fusion peptide mixture was directly folded at pH 8.1 in a buffer solution containing 0.1 M NaHCO<sub>3</sub> and 3 mM reduced and 0.3 mM oxidized glutathione in the presence of 2 M urea and 25% N,N-dimethylformamide (DMF) (15). Then, the folded Cg-Def was purified to homogeneity by RP-HPLC using a C<sub>8</sub> column, as described above. Peptide purity was controlled by acid-urea PAGE and peptide concentration estimated by amino acid analysis and UV absorption at 280 nm based on the extinction coefficients of the molecule. Molecular masses of the purified peptides were determined using matrix-assisted laser desorption ionization mode mass spectrometry (MALDI-TOF-MS) and electrospray ionization mass spectrometry (ESI-MS).

**Antimicrobial Assays**—Antimicrobial activity of recombinant Cg-Def was assayed against several bacteria including the Gram-positive *Micrococcus lysodeikticus*, *Staphylococcus aureus*, *S. haemolyticus*, *Bacillus megaterium*, *Brevibacterium stationi*, *Microbacterium maritopicum*, the Gram-negative *E. coli* 363, *Vibrio alginolyticus* and *Salmonella typhimurium*. The activity of the peptide was also investigated against the following filamentous fungi *Fusarium oxysporum*, *Botrytis cinerea* and *Penicillium crustosum*. Minimum inhibitory concentrations (MICs) were determined in duplicate

by the liquid growth inhibition assay based on the procedure described by Hetru and Bulet (1997) (16). Poor broth (PB: 1% bactotryptone, 0.5% NaCl w/v, pH 7.5) nutrient medium was used for standard bacteria, and saline peptone water (1.5% peptone, 1.5% NaCl, pH 7.2) was used for marine bacteria. Antifungal assay was performed in potato dextrose broth (Difco, Sparks, USA) at half strength supplemented with tetracycline (10 µg/ml final concentration). Growth was monitored spectrophotometrically at 620 nm on a Multiscan microplate reader colorimeter (Dynatech).

Bactericidal or bacteriostatic effect was measured by Colony Forming Unit (CFU) counting following a 15 h incubation at 30°C. A bactericidal kinetic was assayed with *M. lysodeikticus*. Ten microliters of purified peptide, at a concentration 10 times over the MIC value, was mixed with 90 µl of an exponential phase culture of *M. lysodeikticus* at a starting OD<sub>600</sub>=0.01 in PB nutrient medium and incubated at 30°C. Aliquots of 10 µl were plated after 0 min, 2 min, 10 min, 90 min, and 15 h of incubation on nutrient agar plates. The number of CFUs was established after a 15 h incubation at 30°C. As a control, the bacterial culture was incubated with 10 µl of sterile water.

**NMR Spectroscopy**—NMR samples of Cg-Def were prepared either in a 95:5 H<sub>2</sub>O:D<sub>2</sub>O mixture (v/v) or in 99.98% D<sub>2</sub>O to yield a 0.6-0.8 mM solutions. The pH was adjusted to the desired values by addition of DCl or NaOD and measured at room temperature with a 3-mm electrode. They are given uncorrected for the deuterium isotopic effect. Proton chemical shifts were referenced with respect to sodium 4,4-dimethyl-4-silapentane-1-sulfonate according to the IUPAC recommendations. <sup>1</sup>H NMR experiments were performed on a Bruker Avance 600 spectrometer equipped with a triple resonance cryoprobe and spectra were recorded in the temperature range of 10 to 30°C. In all experiments, the carrier frequency was set at the center of the spectrum at the water frequency. Double-quantum filtered-correlated spectroscopy (DQF-COSY) (17, 18), z-filtered total-correlated spectroscopy (z-TOCSY) (19, 20) and nuclear Overhauser effect spectroscopy (NOESY) (21) spectra were acquired in the phase-sensitive mode using the States-TPPI method (22). For spectra recorded in H<sub>2</sub>O, the water resonance was suppressed by the WATERGATE method (23), except for the DQF-COSY spectra where a low-power irradiation was used. The z-TOCSY spectra were obtained with a

mixing time of 90 ms and NOESY spectra with mixing times of 100, 150 and 300 ms. Data were processed by using the XWINNMR software. The full sequential assignment was achieved using the general strategy described by Wüthrich (24). To identify the slow exchanging amide protons, the sample was dissolved in D<sub>2</sub>O at 25 °C and first a TOCSY (80 mn) and then a NOESY (180 mn) experiments were recorded for their identification.

**Structure Calculation and Analysis**—The NOESY cross-peaks were measured from two NOESY spectra (pH 3.25, 20 and 25°C) with a mixing time of 150 ms, and subsequently divided into five classes, according to their intensities. Very strong, strong, medium, weak, and very weak NOEs were then converted into 2.5, 3.0, 3.5, 4.0, and 5.0 Å distance constraints, respectively. The  $\phi$  angle restraints were derived from the 3JNH-C $\alpha$ H coupling constants, and the  $\chi_1$  angle restraints were derived from the combined analysis of the 3JH $\alpha$ -H $\beta$ , $\beta'$  coupling constants and intra-residues NOEs, respectively. To calculate the 3D structures, distance and dihedral angle restraints were used as input in the DYANA program that uses simulated annealing combined with molecular dynamics in torsion angle space (25). In the first stage of the calculation, an initial ensemble of 20 structures was generated from a template structure with randomized  $\phi$ ,  $\psi$  dihedral angles and extended side chains. In preliminary calculations, neither hydrogen bond nor disulfide bond was used as restraint. Hydrogen bonds were considered as present if the distance between heavy atoms was less than 3.5 Å and the donor-hydrogen-acceptor angle was greater than 120°. Finally, to calculate the Cg-Def 3D structure, 456 NOE-derived distances and 13 dihedral constraints including the cis conformation of the Cys4-Pro5 amide bond ( $\omega = 0^\circ$ ) and the disulfide bonds were used as input. A calculation of 60 conformers was carried out, and the resulting 10 structures with a minimum of restrained violations (no violation > 0.3 Å) were analyzed with INSIGHT 97 (Molecular Simulation Inc., San Diego). The Ramachandran analysis was performed with PROCHECK (26) and the limits of the secondary structure elements and the van der Waals surfaces were determined with STRIDE (27). The chemical shifts and coordinates of Cg-Def are deposited in the BioMagResBank (BMRB entry: 6849) and the Protein Data Bank (PDB entry: 2B68), respectively.

**Analyses of *C. gigas* Mantle by RP-HPLC and MALDI-TOF MS**—Mantle tissue from 200 bacteria challenged *C. gigas* oysters was collected, rapidly frozen with liquid nitrogen and ground to fine powder. Then, the mantle sample was diluted in 10% acetic acid, homogenized using an ultra-turax and left at 4°C under gentle stirring for 15 h. Extracts were centrifuged at 8000 g for 20 min at 4°C and the supernatant prepurified by solid-phase extraction on Sep-Pak C<sub>18</sub> cartridges (Waters Associates) equilibrated with acidified water (0.05% trifluoroacetic acid). After washing with acidified water, peptides were eluted with 100% acetonitrile containing 0.05% trifluoroacetic acid. The Sep-Pak fraction was then lyophilized, reconstituted in ultrapure water and loaded on a C<sub>8</sub>-reversed-phase high performance liquid chromatography (RP-HPLC) on an UP10C<sub>8</sub> column (Interchrom modulocart uptisphere 10 C<sub>8</sub> 250 X 10 mm) and elution was performed using a 0-55% acetonitrile gradient developed over 30 min at a flow rate of 5 ml/min. Fractions were hand collected, lyophilized, resuspended in 100  $\mu$ l of ultrapure water and then assayed for antimicrobial activity. Finally, the active fractions were purified in a third step by C<sub>18</sub> RP-HPLC, on a Symetry Shield RP<sub>18</sub> column (Waters, 5 $\mu$ m, 4.6x250) using a 0-45% acetonitrile gradient developed over 45 min at a flow rate of 1 ml/min. The corresponding fractions were then analyzed by MALDI-TOF-MS.

**Bacterial Challenge and Quantification of Cg-def Gene Expression**—Adult *C. gigas* oysters were stimulated by bath with heat-killed bacteria (5.10<sup>8</sup> bacteria/liter). The stimulation was performed with three bacterial species, *M. lysodeikticus*, *Vibrio splendidus* and *V. anguillarum*. The bacterial strains were grown separately overnight, in saline peptone water (SPW) at 25°C for *Vibrio* strains and in Luria-Bertani (LB) medium for *M. lysodeikticus* (30°C). Bacteria were collected by centrifugation (10000 g, 5 min) and suspended in fresh growth medium. Bacteria concentration was calculated from the optical density at 550 nm (1 unit OD<sub>550</sub> corresponds to 5x10<sup>8</sup> bacteria/ml). Mantle tissue samples (100 to 300 mg) were then collected at two times post-stimulation (24 and 48 h) and washed in sterile sea water, cut into small pieces and incubated overnight at 4°C in Trizol reagent (1 ml/100 mg of tissue). Total RNAs were extracted following manufacturer instructions (Invitrogen™) and treated with DNase Turbo

(Ambion). The experiments were done in triplicate and, to minimize individual variability, at least ten oysters were used in each experimental condition.

Following heat denaturation (70°C for 5 min), reverse transcription was performed using 1 µg of total RNA prepared with 50 ng/µl oligo-(dT)<sub>12-18</sub> in a 50 µl reaction volume containing 1mM dNTPs, 1unit/µl of RnaseOUT™ (Invitrogen) and 200 units/µl M-MLV reverse transcriptase in reverse transcriptase buffer. PCR amplifications were performed with the LightCycler™ (Roche) in the presence of SYBR-Green™ (Master SYBR Green™) with the following primers: Defm1 5'- GATGGTTTCTGCAGACATGG -3' and Defm2 5'- CACAGTAGCCCGCTCTACAAC -3' as sense and antisense primers, respectively. The gene encoding the elongation factor (EF, GenBank AB122066) was used as internal control. For EF the forward and reverse primers were EF (5'- ATGCACCAAGGCTGCACAGAAAG-3') and EFR (5'-TCCGACGTATTTCTTTGCGATGT-3'), respectively. Samples were run under the following conditions: 95°C (10 min); and 39 cycles of 95°C (10 s), 62°C (5 s) and 72°C (15 s). For further expression level analysis, the crossing points (CP) were determined for each transcript using the LightCycler software. Specificity of RT-PCR product was analyzed on agarose gel and melting curve analysis. The copy ratio of each analyzed cDNA was determined as the mean of three replicates. The relative expression ratio of *Cg-def* was calculated based on the CP deviation of each RT-PCR product of RNA extracted from stimulated oyster versus the appropriate control sample, and expressed in comparison to the reference gene *EF*. The relative expression ratio of *Cg-Def* was calculated based on the delta-delta method for comparing relative expression results (28)

## RESULTS

### *Cg-def* cDNA Cloning and Primary Structure

**Analyses**—The complete cDNA of *C. gigas* defensin (*Cg-Def*) was obtained by 5' amplification of cDNA ends-PCR on a *C. gigas* mantle edge cDNA library, as described in Materials and Methods (GenBank AJ565499). The oyster *Cg-def* cDNA contained 323 bp, comprising a coding region of 195-bp, a 110-bp 3' untranslated region containing a polyadenylation consensus sequence (AATAAA) at position 66 to 72 of the 3'UTR and poly(A) tail. The 195-bp coding region encoded a 65-amino-acid propeptide (GenBank CAJ19280). The deduced amino acid sequence starts

with a 22-residue signal peptide and the cleavage site for signal peptidase is most likely located after the alanine residue preceding the glycine in position 23 as predicted by SignalP 3.0 software (data not shown). *Cg-Def* is not synthesized as a precursor with a C-terminal extension or a N-terminal pro-region as observed with the Mediterranean mussel or the *Drosophila* defensins, respectively. The amino acid sequence of the mature peptide was aligned with other defensins from the “arthropod defensin family” available in GenBank and Expasy, that contains defensins from arthropods, mollusks and fungi (Fig. 1A and 1B). As already observed for the mussel defensins MGD-1 and MGD-2, *Cg-Def* is an original member of this family due to the presence of two extra-cysteines residues. *Cg-Def* shares 50% of identity with MGD-1 from *Mytilus galloprovincialis*, but has a unique doublet of lysine residue at the C-terminal part of the molecule (Fig. 1A). Less identity was shared with defensin from the scorpion *Androctonus australis* (46%) and the tick *Ixodes scapularis* (44%). The selected members of the “arthropod defensin family” belonging to the so-called primitive defensin family show a high degree of conservation with both *Cg-Def* and the different mussel defensins (Fig. 1A) (29). Interestingly, very recently, a 40-amino acid AMP named plectasin (GenBank number CAI83768) with marked homologies to the arthropod and mollusk defensins has been isolated from the saprophytic fungus *Pseudoplectania nigrella*.

**Genomic Organization and Analysis of the Promoter Region**—*Cg-def* gene (EMBL AM050547) harbours a unique intron whose splice junction follows the AG-GT rule (Fig. 1C). The genomic organization of *Cg-def* is similar to that of the mussel and scorpion defensin genes (29). It also displays the features of the other arthropod and mollusk defensin genes studied so far that demonstrate a genetic relatedness: (i) the exon encoding the mature defensin is never split by any intron, (ii) the intron flanking the exon encoding the mature defensin shows for all genes a strict level of phase conservation (phase 1) (29). Analysis of the 3.6 kb genomic DNA sequence upstream from the putative ATG start codon identified several consensus sequences for transcription factors commonly observed in promoters of cellular house keeping genes (Sp1), genes expressed during early embryogenesis (AP2, Gata 1, Gata 2), development (Zeste), cell cycle regulation (H1TF2, H4TF1 and 2), cell differentiation

(Gata 2, Ik1, Ik2, Ik3, IRF2, Oct1, 2, 4, 6 and 9) and genes expressed in a tissue specific manner (Pit-1, SRF).

**Production, Folding and Purification of Recombinant Cg-Def**—In order to produce sufficient amounts of Cg-Def for detailed studies (5 mg scale) of its biological activity and its 3D structure, Cg-Def was expressed in *E. coli* Rosetta (DE3) pLysS cells transformed with the pET-28a/Cg-Def construct. Recombinant His<sub>6</sub>-tagged fusion protein was purified by affinity chromatography from the bacteria cell lysates. After chemical cleavage with CNBr and refolding, Cg-Def was purified by RP-HPLC, and its purity was judged by analytical RP-HPLC, MALDI-TOF and ESI-MS mass spectrometry analyses (Fig. 2, panels A and B). The oxidative folding of Cg-Def was assessed under two different conditions. First, after complete reduction with DTT, purified denatured Cg-Def was refolded during 48 h at room temperature in the presence of a Tris-HCl 100 mM pH 8 buffer. The refolding process was followed by RP-HPLC revealing at 48 h an additional more hydrophobic peak eluting around 35 min (Fig. 2, panel A, asterisk). By ESI-MS analyses, this hydrophobic fraction was found to contain a molecule with a molecular mass of 4,634.10 Da (see Fig. 2 panel B1). The value measured by ESI-MS is in perfect agreement with the calculated average molecular mass of 4,634.34 Da. This suggests that the recombinant Cg-Def has its eight cysteine residues paired. Alternatively, as described by Wu et al. (15), the use of DMF and urea was shown to enhance 5-times the folding efficiency of Cg-Def (data not shown). Using this protocol, the purified recombinant Cg-Def was also found to have a molecular mass measured by ESI-MS that is in perfect agreement to the calculated average molecular mass (see Fig.2, panel B1). The purified refolded peptide was then submitted to antimicrobial assays and NMR spectroscopy studies.

**Antimicrobial Activity of Cg-Def**—The antimicrobial activity of the recombinant Cg-Def was determined against a panel of micro-organisms including Gram-positive, Gram-negative bacteria and filamentous fungi. The MIC values obtained are reported in Table 1. The peptide was active at very low concentration against most of the Gram-positive bacteria tested including *Micrococcus lysodeikticus*, *Staphylococcus aureus* and the marine bacteria *Brevibacterium stationis* and *Microbacterium*

*maritypicum*. However, Cg-Def was not active at 20  $\mu$ M against *Bacillus megaterium*. Cg-Def showed no activity at 20  $\mu$ M against the Gram-negative bacteria *Vibrio alginolyticus* and *Salmonella typhimurium*. At higher concentration (35  $\mu$ M), the peptide was active against *E. coli*. Cg-Def displayed anti-fungal activity against *Fusarium oxysporum* at relatively high concentrations (9  $\mu$ M; Table 1). Experiments were conducted to examine the bactericidal effects of Cg-Def. The bactericidal activity of the peptide was assessed by plating cultures and the number of CFUs counted after overnight incubation at 30°C. Cg-Def exerted bactericidal effects against all the Gram-positive bacteria tested, excepted *M. maritypicum* (Table 1). When *M. lysodeikticus* was incubated with Cg-Def at concentration 10 times higher than the MIC value, all bacteria were killed in few minutes (Table 2).

In order to determine if the high salinity of the sea water might modify the efficacy of Cg-Def *in vivo*, the effect of the peptide on bacterial growth was tested *in vitro* at NaCl concentrations ranging from 85 mM to 1 M. Cg-Def is highly active against *M. lysodeikticus* and *B. stationis* even at 600 mM NaCl, a concentration closed to the value measured in sea water. Interestingly, the MIC value is unchanged throughout the NaCl concentration range tested [85 mM - 600 mM] (0.01  $\mu$ M for *M. lysodeikticus* and 0.2  $\mu$ M for *B. stationis*; Table 3). At 1 M NaCl, *M. lysodeikticus* did not grow in the control and a low bacterial growth can be observed for the marine bacteria *B. stationis*. In contrast to most AMPs, Cg-Def seems to conserve its antibacterial activity at such a high salt concentration (30, 31).

**NMR Structural Study**—The 2D NMR spectra (TOCSY, DQF-COSY and NOESY) of Cg-Def were recorded at different temperatures, ranging from 17 to 32°C. The identification of all the spin systems of Cg-Def was obtained by analysis and comparison of DQF-COSY, TOCSY, and NOESY spectra according to the strategy described by Wüthrich (24). Nevertheless, we noticed that the 2D spectra showed, for each residue of the D38CNGK42 sequence, two spin systems approximately equally populated. This suggests a heterogeneity for this C-terminal part inferred to result from the partial deamidation to yield the Asp40 and isoAsp40 mixture. The AsnGly sequence (Asn40Gly41) has been shown to experience the highest propensity for the deamidation process through the succinimide intermediate (32). Indeed, such a

deamination process is known to occur and depend upon primary sequence, 3D structure, and solution properties such as pH, temperature, ionic strength, and buffer ions (33). The two spin systems were assigned to the initial molecule (Asn40) and to the Asp40 analogue. These two similar spin systems were identified by recording TOCSY spectra at several pH values in the 2-5 pH range to monitor the ionization state change of the carboxyl group. The spin system sensitive to the ionization state change was assigned to Asp40. The other one, unaffected in this pH range, was assigned to Asn40. Only a small amount of the third expected compound (isoAsp40) was formed. It was detected in the TOCSY experiment but not fully characterized in the NOESY.

Chemical shifts of the recombinant *Cg-Def* are reported in the Table S1 provided as “Supplementary Materials”. Both amide and alpha protons resonances were spread in a large range of chemical shifts and some atypical values were measured. When compared to statistical values, among the alpha proton resonances, those of Ala27 (1.87 ppm) and of Trp31 (3.17 ppm) were upfield shifted by around 2.5 and 1.5 ppm, respectively. The large spreading of amide signals is illustrated by the two extreme values measured for Ala27 (9.84 ppm) and for Ile18 (6.91 ppm). Altogether, the large spreadings of chemical shifts suggest a highly constrained structure, in agreement with the presence of the four disulfide bonds. This is also supported by several large non-equivalencies such as for the Gly3 and Gly23 alpha protons resonances for which non-equivalencies of 1.24 ppm and 1.29 ppm were measured, respectively. Similarly, non-equivalencies of 1.21 ppm and 1.02 ppm were also observed for  $\beta\beta'$  signals of Cys4 and Cys11, respectively. The large difference of chemical shifts measured between the alpha (at 4.14 ppm) and the beta (at 3.08 ppm) protons of Thr35 residue and in a lesser extent for Thr37 (at 4.61 and 4.05 ppm) also suggests that they belong to highly constrained parts of the molecule. By comparison, chemical shifts of Thr29 (at 4.45 and 4.52 ppm) were very close to that usually observed for a less constrained sequence. The NOESY data showed several  $d_{NN}$  and  $d_{\alpha\alpha}$  NOEs suggesting that the molecule both contains a helical and a  $\beta$ -stranded parts. Moreover, the strong  $d_{\alpha}Cys^4-\alpha Pro^5$  NOE and the absence of the  $d_{\alpha}C^4-\delta\delta-P^5$  NOE unambiguously indicated a cis conformation for the Cys4-Pro5 amide bond. Among the amide protons found in slow exchange, eight of them were successive and span the 13-21 sequence. The amide protons of Ala22, Thr35,

Cys36 and Thr37 residues were also found in slow exchange.

**Solution Structure of *Cg-Def***—A stereoview of a family of 10 *Cg-Def* conformers is reported in Fig. 3. The global structure of *Cg-Def* was well defined and displays the Cystine Stabilized- $\alpha\beta$  motif (CS- $\alpha\beta$ ) that consists of an helical structure and two  $\beta$  strands cross linked by three disulfide bonds (34) and sometimes by a fourth disulfide bond (35-37). The backbone atoms of the ten conformers were overlaid for the well-defined regions spanning residues 4-39 resulting in a pairwise average rmsd of  $0.43 \pm 0.09$  and  $0.98 \pm 0.16$  Å for the backbone and heavy atoms, respectively. The Ramachandran plot of all residues (except for the glycine and proline residues) of the ten best conformers indicated that 89.7% were located in the most favoured and the additional allowed regions, and 5.8% in the generously allowed regions. A total of 4.4% of the residues were found in the disallowed regions. Limits of secondary structure elements indicated that the *Cg-Def* structure mainly consisted of an helical part (8-18 residues), two  $\beta$ -strands (residues 22-26 and 33-37), and three loops (residues 1-7, 19-21 and 27-32 for loops L1 L2 and L3, respectively) (Fig. 1B and 3). Three successive type IV turns were identified in L1 (1-4, 3-6, and 4-7) and L3 (27-30, 28-31 and, 29-32) loops, and a type I turn was found for the 26-29 sequence. It is interesting to note that the Cys4-Pro5 amide bond adopted the unusual cis conformation. Finally, the global fold is highly constrained by the 4-25, 11-34, 15-36, and 20-39 disulfide bonds. The *Cg-Def* helix exhibits an amphiphilic character with a hydrophobic side contributing to the hydrophobic core of the molecule and a hydrophilic side accessible to the solvent including mainly, Lys10, Asn13, His14, Lys16 and, Ser17 residues. Nevertheless, the Leu9 and Ile18 hydrophobic residues are solvent exposed. The hydrophobic core mainly includes the four disulfide bonds, the Ala22 side chain and is extended to the L1 (Pro5) and L3 hydrophobic loops (Ala27, Ala28, Leu30, Trp31, Leu32). As expected, the positively (Lys10, Lys16, Arg21, Arg33, Lys42, Lys43) and negatively charged (Asp26, Asp38) residues are solvent exposed.

The *Cg-Def* structure was compared to that of MGD-1 (PDB entry: 1FJN) whose 3D structure was previously determined (37). Their superimposition showed that the helical and  $\beta$ -strands elements were very conserved (Fig. 4).

### *Cg-def* mRNA and Peptide Tissue

**Localization**—To investigate the tissue distribution of *Cg-def* expression, we analyzed by real-time RT-PCR mRNA content in various oyster tissues. Low or no *Cg-def* mRNA amount was measured in most of the organs analyzed including hemocytes, heart, digestive gland and gills whereas a high mRNA level, around 20 to 60 fold the level measured in the remaining animal tissues, was detected in mantle (Fig. 5). In non-stimulated *C. gigas* oysters, the mantle appears as the main tissue expressing *Cg-def* transcripts.

In order to demonstrate the presence of native *Cg-Def* in mantle, an acidic extract of this tissue was prepared, subjected first to RP-HPLC and then to antibacterial activity screening. Following solid-phase extraction of the acidified extracts from oyster mantle, RP-HPLC revealed the presence of two fractions with antimicrobial activity. These two fractions were subjected to molecular mass fingerprint analysis by MALDI-TOF-MS. In order to ascertain the molecular mass measured, the well-folded recombinant *Cg-Def* was used as external calibrant (see Fig. 2 panel B2). A difference of 3 Da was observed between the molecular mass measured by ESI-MS (4,634.10 Da, panel B1) and the one measured by MALDI-TOF-MS (4,637.96 Da, arrow in panel B2). Following solid-phase extraction of the acidified extracts from oyster mantle, RP-HPLC revealed the presence of two fractions with antimicrobial activity. The less hydrophobic fraction (see Fig. 2, panel C) contains a component with a molecular mass at 4637.96 corresponding to that measured by MALDI-TOF-MS for the recombinant folded *Cg-Def*. This peptide with antimicrobial activity against *M. lysodeycticus* might likely correspond to the native form of *Cg-Def*. Interestingly, the second active fraction (more hydrophobic), that has not yet been identified, showed a close but different molecular mass of 4631.45 Da (see Fig. 2, arrow in panel D).

**Real-Time PCR Analyses of *Cg-def* transcript levels after oyster bacterial challenges**—In order to determine the expression pattern of *Cg-def* during bacterial challenge, two batches of oysters were selected. In the first one, oysters were stimulated by bath with killed bacteria (see Material and Methods) and in the second one, non-stimulated oysters were used as control. Then, qRT-PCR analyses were performed with total RNA extract from mantle collected at two times post challenge (24 and 48 h). No

striking discrepancies were observed for *Cg-def* expression between naive and bacteria challenged oysters at the different times analyzed (Fig. 6). These preliminary analyses revealed that *Cg-def* mRNA are present in naive *C. gigas* oyster and that, the level of transcript was unaffected by the bacterial-challenge performed.

## DISCUSSION

Efficient host defence mechanisms are needed to neutralize microbial invasions to which living organisms are exposed. In higher vertebrates, innate and adaptive immunity are present. In contrast, invertebrate and plant defence against pathogens takes place exclusively through mechanisms that are part of the innate immunity. The involvement of AMPs in natural resistance to infection is sustained by their strategic location in phagocytes, in body fluids and at the epithelial level, *ie* at interfaces between the organisms and its environment. In oyster, even if antimicrobial activities have been detected in the hemolymph of some species (38), no AMP has been fully characterized despite many attempts to purify such molecules by biochemical approaches from hemolymph and other tissues (13). In this report, and for the first time in an oyster, we describe the characterization of an AMP isolated from mantle tissue. Using molecular approaches, a new member of the widespread defensin family, which is present in animal and plant kingdoms, was identified. Usually, three disulfide bonds characterize this defensin family (Fig. 1A). The amino acid sequence of the mature *C. gigas* defensin displays interesting homologies with defensins from the “arthropod family” including the mussel’s defensins. Moreover, as observed with the defensins MGD-1, MGD-2 and MGD-2b from *M. galloprovincialis*, *Cg-Def* has four disulfide bonds. This additional bridge was proposed to render the peptide more stable in high osmolarity media such as in sea water (37). However, this fact is not common to all marine mollusk defensins because the mussel *M. edulis* defensins do not bear this additional disulfide bond (Fig. 1A) (11). Interestingly, *Cg-Def* possesses a N-terminal pre-segment presumed to be a signal sequence for translocation to the lumen of the rough endoplasmic reticulum, but has no C-terminal extension nor a N-terminal pro-region as observed in the mussel or the *Drosophila* defensins, respectively (39). This property would suggest a different way of processing for the oyster defensin.

To investigate its biological properties, *Cg-Def* was produced *in E. coli* and refolded *in vitro*. This recombinant approach was efficient, yielding enough material (approx. 1 mg of well-folded pure peptide per liter of culture) for functional assays and structural characterization. A challenging aspect for the production of *Cg-Def* in *E. coli* was the presence of eight cysteine residues in the mature peptide. Indeed, the correct refolding of proteins with several disulfide bonds is a general problem. As discussed by Harder *et al.*, only few reports describe the production of AMP in bacteria, a fact that reflects the difficulty of expressing and folding such cysteine rich molecules in a bacterial host system (6). Previous reports described the production of human  $\alpha$ - and  $\beta$ -defensins in *E. coli* (6, 40). We report here for the first time the efficient production in a bacterial-host of an AMP with four cysteine bridges that can be properly refolded *in vitro*. The activity spectrum of the recombinant *Cg-Def* was evaluated against a panel of bacteria and fungi. Consistent with studies on invertebrate defensins (2), *Cg-Def* was active *in vitro* against Gram-positive bacteria but showed no or limited activities against Gram-negative bacteria and fungi. Most interestingly, *Cg-Def* kept its antibacterial properties in the presence of high NaCl concentrations (up to 1M). This favours the hypothesis that *Cg-Def* retains its activity in *in vivo* conditions (sea water) and would play a role in the antimicrobial defence of the oyster *C. gigas*. For these reasons, *Cg-Def* represents a model of choice for structure-activity relationship studies to design a lead antibacterial drug to treat infections of bacteria in a salt-rich environment (41). This is particularly interesting for cystic fibrosis, for which it has been suggested that the primary genetic defect increases the salt content of fluid lining the airway, which reversibly inactivates antimicrobial molecules (42, 43).

As expected from the disulfide bridges array, the global fold of *Cg-Def* includes the CS- $\alpha\beta$  motif observed in the defensins isolated from mussels (37), insects (34) and plants (44). The *Cg-Def* and MGD-1 sequences were aligned and their structures compared (Fig. 1B and 4). This alignment revealed 50 % of identity (23 residues with three gaps) including eight cysteines (45). *Cg-Def* and MGD-1 share the fourth disulfide bond that contributes to the stability and to the rigidity of their 3D structure. The identical residues were mainly gathered in four segments (residues 1-5, 14-18, 23-25 and 32-36). Beyond their sequence comparison the *Cg-Def* structure was compared to that of MGD-1 (PDB entry: 1FJN) (37). Their

superimposition showed that the helical and  $\beta$ -strand elements were rather conserved and the best fit was obtained for the backbone superimposition of the residues 1-4, 7-17, 21-25, 34-36 of *Cg-Def* with residues 1-4, 6-16, 21-25, 33-35 of MGD-1 (23 residues). A rmsd of 1.04 Å was measured for the backbone. Moreover, the two structures share the Cys4-Pro5 cis-amide bond and a similar disulfide bond pattern. Nevertheless, some differences were observed for the Cys20-Cys39 disulfide bond and for the length of the three loops. While the three equivalent disulfide bonds (4-25, 11-34, 15-36 for *Cg-Def* and 4-25, 10-33, 14-35 for MGD-1) well overlaid, the fourth one (Cys20-Cys39 for *Cg-Def* and Cys21-Cys38 for MGD-1) is significantly different mainly due to the translation of Cys20 by one residue towards the center of the molecule. This is certainly due to the shorter L2 loop (one residue less) joining the helix and the first  $\beta$ -strand. The consequence is a shift of this disulfide bridge that locates on the opposite side of the  $\beta$ -sheet ("above" in *Cg-Def* and "below" in MGD-1 as displayed in Fig. 4). Conversely, the L1 and L3 loops of *Cg-Def* display one additional residue. Whereas the sequences of the L1 loops including the Cys4-Pro5 cis amide bond are well conserved (Fig. 4), the sequences of the L3 loops are totally different. This latter is much more hydrophobic (A27ATLWL32 for *Cg-Def*) than that of MGD-1 (G27WHRL31) that contains two positively charged residues. Nevertheless, due to the positions in the sequence (at the end for *Cg-Def* and at the beginning for MGD-1), it appeared that, they have opposite location with regard to the loop. In contrast, the Leu32(*Cg-Def*)/Leu31(MGD-1) side chains were similarly located (Fig. 4B). Since it has been shown that the L3 loop of MGD-1 is responsible for a large part of the antibacterial activity (46), this significant structural difference is worth being noted.

It is commonly admitted that the surface distribution of hydrophobic and hydrophilic side chains in AMPs is essential for their antimicrobial activity. The alignment showed that several of them (Phe2, Pro5, Ile18, Tyr24, Leu32 and His14, Lys16, Arg33) were conserved (Fig. 1 and see below). However, the two antimicrobial sequences mainly differ by the presence of two aspartic residues (Asp26 and Asp38) and by the addition of a dibasic peptide (Lys42 and Lys43) at the C-terminus. Although, the *Cg-Def* and MGD-1 surfaces display significant differences mainly due to the two aspartic acids, the hydrophobic L3 loop and, by the "above"/"below" locations of the Cys20-Cys39/Cys21-Cys38 disulfide

bonds, several hydrophobic (Phe2/2, Pro5/5, Leu9/Tyr8, Ile18/17, Tyr24/24, Leu32/31 and charged residues (Lys10/Arg12, His14/13, Arg21/Lys15, Arg33/32, Lys42/Arg20, Lys43/Arg37) are similarly located at the surface giving rise to a comparable location of hydrophobic and hydrophilic clusters (Fig. 4B).

Our results showed that in non-stimulated animals, the mantle appears as the main tissue expressing *Cg-Def*. Although, we were not able to isolate by HPLC significant amount of pure defensin from the oyster, we clearly detected its presence by mass spectrometry analyses of mantle tissue (Fig. 2, panel C). The results presented here are in contrast with those reported for the mollusk, *M. galloprovincialis*. In the mussel, AMPs are only produced in hemocytes where they are stored and released following bacterial challenges (45). In the well-studied insect model, *Drosophila melanogaster*, AMPs are predominantly produced in the fat body (the functional equivalent of mammalian liver) and secreted into the blood in response to a microbial challenge, which characterizes the systemic response in insects. Otherwise, the expression of these peptides is also locally regulated in the surface epithelia of several *Drosophila* tissues (47), as also observed for the expression of peptides in mammalian epithelia. For example, *Drosophila* defensin is expressed in the

labellar glands and in the seminal receptacle and spermatheca (for a review see Imler and Bulet 2005 (48)). Additionally, in mammals, AMPs are constitutively produced by blood cells (49). Apart from insects, most of the AMPs reported in the different groups of invertebrates have been isolated from hemocytes (13). Our preliminary results would suggest that *Cg-Def* is continuously expressed in the oyster mantle, *ie* at epithelial surface, indicating that this AMP has an important function in host protection against environmental microbes. Indeed, in oyster, the mantle is a site of intense exposure to external environment that represents a dynamic ecosystem for numerous bacterial species, including commensals and potential pathogens. A more detailed study of *Cg-def* gene expression in response to various microbial challenges will be performed to gain insight into the role of *Cg-Def* in oyster immunity. In addition, AMPs isolation and characterization must be investigated from other oyster tissues. These data on AMPs are of great interest to understand how the oyster immune system interacts with the commensal flora and how it responds to environmental pathogens.

#### ACKNOWLEDGEMENTS

We are very grateful to Dr. A. J. Ouellette for excellent advices during defensin recombinant expression. We also thank B. Romestand, J. de Lorgeril, G. Desserre and M. Leroy for helpful discussion and assistance.

## REFERENCES

1. Yang, D., Biragyn, A., Kwak, L. W., and Oppenheim, J. J. (2002) *Trends Immunol* **23**, 291-296
2. Bulet, P., Stocklin, R., and Menin, L. (2004) *Immunol Rev* **198**, 169-184
3. Boman, H. G. (2003) *J Intern Med* **254**, 197-215
4. Brogden, K. A. (2005) *Nat Rev Microbiol* **3**, 238-250
5. Boulanger, N., Lowenberger, C., Volf, P., Ursic, R., Sigutova, L., Sabatier, L., Svobodova, M., Beverley, S. M., Spath, G., Brun, R., Pesson, B., and Bulet, P. (2004) *Infect Immun* **72**, 7140-7146
6. Harder, J., Bartels, J., Christophers, E., and Schroder, J. M. (2001) *J Biol Chem* **276**, 5707-5713
7. Izadpanah, A., and Gallo, R. L. (2005) *J Am Acad Dermatol* **52**, 381-390; quiz 391-382
8. Raj, P. A., and Dentino, A. R. (2002) *FEMS Microbiol Lett* **206**, 9-18
9. Sima, P., Trebichavsky, I., and Sigler, K. (2003) *Folia Microbiol (Praha)* **48**, 123-137
10. Dimarcq, J.-L., Bulet, P., Hétru, C., and Fofmann, J. A. (1998) *Biopolymers* **47**, 465-477
11. Charlet, M., Chernysh, S., Philippe, H., Hétru, C., Hoffmann, J. A., and Bulet, P. (1996) *J. Biol. Chem.* **271**, 21808-21813
12. Mitta, G., Vandenbulcke, F., and Roch, P. (2000) *FEBS Letters* **486**, 185-190
13. Bachère, E., Gueguen, Y., Gonzalez, M., de Lorgeril, J., Garnier, J., and Romestand, B. (2004) *Immunol. Rev.* **198**, 149-168
14. Satchell, D. P., Sheynis, T., Kolusheva, S., Cummings, J., Vanderlick, T. K., Jelinek, R., Selsted, M. E., and Ouellette, A. J. (2003) *Peptides* **24**, 1795-1805
15. Wu, Z., Powell, R., and Lu, W. (2003) *J Am Chem Soc* **125**, 2402-2403
16. Hetru, C., and Bulet, P. (1997) in *Methods in Molecular Biology* (Shafer, W. M., ed) Vol. 78, pp. 35-49, Humana Press Inc., Totowa, NJ
17. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117**, 479-485.
18. Derome, A. E., and Williamson, M. P. (1990) *Journal of Magnetic Resonance* **88**, 177-185
19. Bax, A., and Davis, G. D. (1985) *Journal of Magnetic Resonance* **65**, 355-360
20. Rance, M. (1987) *J. Magn. Reson.* **74**, 557-564
21. Macura, S., Huang, Y., Sutter, D., and Ernst, R. R. (1981) *J. Magn. Reson.* **43**, 259-281
22. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson.* **85**, 393-399
23. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* **2**, 661-665.
24. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & sons, New York
25. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J Mol Biol* **273**, 283-298
26. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8**, 477-486.
27. Frishman, D., and Argos, P. (1995) *Proteins* **23**, 566-579
28. Livak, K. J., and Schmittgen, T. D. (2001) *Methods* **25**, 402-408
29. Froy, O., and Gurevitz, M. (2003) *Trends Genet* **19**, 684-687
30. Landon, C., Thouzeau, C., Labbe, H., Bulet, P., and Vovelle, F. (2004) *J Biol Chem* **279**, 30433-30439
31. Zaballo, A., Villares, R., Albar, J. P., Martinez, A. C., and Marquez, G. (2004) *J Biol Chem* **279**, 12421-12426
32. Li, B., Gorman, E. M., Moore, K. D., Williams, T., Schowen, R. L., Topp, E. M., and Borchardt, R. T. (2005) *J Pharm Sci* **94**, 666-675
33. Robinson, N. E. (2002) *Proc Natl Acad Sci U S A* **99**, 5283-5288
34. Cornet, B., Bonmatin, J. M., Hetru, C., Hoffmann, J. A., Ptak, M., and Vovelle, F. (1995) *Structure* **3**, 435-448
35. Savarin, P., Romi-Lebrun, R., Zinn-Justin, S., Lebrun, B., Nakajima, T., Gilquin, B., and Menez, A. (1999) *Protein Sci* **8**, 2672-2685

36. Guijarro, J. I., M'Barek, S., Gomez-Lagunas, F., Garnier, D., Rochat, H., Sabatier, J. M., Possani, L., and Delepierre, M. (2003) *Protein Sci* **12**, 1844-1854
37. Yang, Y. S., Mitta, G., Chavanieu, A., Calas, B., Sanchez, J. F., Roch, P., and Aumelas, A. (2000) *Biochemistry* **39**, 14436 - 14447
38. Anderson, R. S., and Beaven, A. E. (2001) *Aquat Living Resour* **14**, 343-349
39. Mitta, G., Hubert, F., Dyrynda, E. A., Boudry, P., and Roch, P. (2000) *Develop. Comp. Immunol.* **24**, 381-393
40. Ouellette, A. J., Satchell, D. P., Hsieh, M. M., Hagen, S. J., and Selsted, M. E. (2000) *J Biol Chem* **275**, 33969-33973
41. Friedrich, C., Scott, M. G., Karunaratne, N., Yan, H., and Hancock, R. E. (1999) *Antimicrob Agents Chemother* **43**, 1542-1548
42. Goldman, M. J., Anderson, G. M., Stolzenberg, E. D., Kari, U. P., Zasloff, M., and Wilson, J. M. (1997) *Cell* **88**, 553-560
43. Aarbiou, J., Rabe, K. F., and Hiemstra, P. S. (2002) *Ann Med* **34**, 96-101
44. Da Silva, P., Jouvensal, L., Lamberty, M., Bulet, P., Caille, A., and Vovelle, F. (2003) *Protein Sci* **12**, 438-446
45. Mitta, G., Vandenbulcke, F., Hubert, F., and Roch, P. (1999) *J. Cell Sc.* **112**, 4233-4242
46. Romestand, B., Molina, F., Richard, V., Roch, P., and Granier, C. (2003) *Eur J Biochem* **270**, 2805-2813
47. Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J. M., Lemaitre, B., Hoffmann, J. A., and Imler, J. L. (2000) *Immunity* **13**, 737 - 748
48. Imler, J. L., and Bulet, P. (2005) *Chem Immunol Allergy* **86**, 1-21
49. Lehrer, R. I., and Ganz, T. (1999) *Curr. Opin. Immunol.* **11**, 23-27
50. Casteels, P., Ampe, C., Jacobs, F., and Tempst, P. (1993) *J. Biol. Chem.* **268**, 7044-7054
51. Mitta, G., Hubert, F., Noël, T., and Roch, P. (1999) *Eur. J. Biochem.* **264**, 1-9

## FOOTNOTES

# These authors contributed equally to this work.

& Present address: Ecologie microbienne des insectes et interactions hôte-pathogène, UMR INRA Université de Montpellier II, 2 Place E. Bataillon, CC80, F-34095 Montpellier cedex 5, France.

\* Works in the author' laboratories are supported by the Ifremer, the CNRS, The INSERM and the University of Montpellier II and Caen. This study was part of a collaborative project supported by the European Commission, DG XII, in the program International Co-operation with Developing Countries, INCO-DC, Contract n° ICA4-CT-2001-10023 (IMMUNAQUA). This study was also part of a French program MOREST funded by Ifremer, the Regions "Basse-Normandie, Bretagne, Pays de la Loire and Poitou-Charentes and by the "Conseil Général du Calvados". Atheris Laboratories was supported by the Swiss Federal Office for Education and Science contract n° 02.0360. A. Herpin, C. Lelong and P. Favrel were financially supported by the Conseil Régional de Basse-Normandie, Agence de l'Eau "Seine Normandie" and FEDER N°4474 (program PROMESSE).

The abbreviations used are: *Cg-Def*, *Crassostrea gigas* defensin; DG, distance geometry; DQF-COSY, 2D double-quantum filter correlation spectroscopy; MALDI-TOF-MS; matrix-assisted laser desorption/ionization time of flight mass spectrometry; ESI-MS, electrospray ionization mass spectrometry; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D nuclear Overhauser effect spectroscopy; RP-HPLC, reverse-phase high performance liquid chromatography; rmsd, root mean square deviation; TOCSY, total correlation spectroscopy; TPPI, time proportional phase incrementation; PAGE, polyacrylamide gel electrophoresis; qRT-PCR, quantitative reverse transcriptase-polymerase chain reaction, CFU, colony forming unit.

**Table 1. Antimicrobial activity of recombinant Cg-Def.**

MIC values are expressed as the interval of concentration [a]-[b], where [a] is the highest concentration tested at which microbial growth can be observed and [b] is the lowest concentration that causes 100% growth inhibition (50).

Microorganisms	MIC ( $\mu\text{M}$ )	Bactericidal effect of Cg-Def
<b>Bacteria</b>		
Gram-positive		
<i>Micrococcus lysodeikticus</i>	0.005-0.01	Yes
<i>Staphylococcus aureus</i>	1.25-2.5	Yes
<i>Brevibacterium stationis</i>	0.1-0.2	Yes
<i>Microbacterium maritopicum</i>	0.5-1	No
<i>Bacillus megaterium</i>	>20	
Gram-negative		
<i>Escherichia coli</i> 363	35-20	nd <sup>a</sup>
<i>Vibrio alginolyticus</i>	>20	
<i>Salmonella typhimurium</i>	>20	
<b>Filamentous fungi</b>		
<i>Fusarium oxysporum</i>	9-4.5	
<i>Botrytis cinerea</i>	>20	
<i>Penicillium crustosum</i>	>20	

<sup>a</sup> not determined

**Table 2: Kinetic of killing of Cg-Def on *M. lysodeikticus*.**

Cg-Def at a final concentration (0.1 $\mu$ M) 10 times higher than the MIC value or water (control) was added to an exponential growth phase culture of *M. lysodeikticus* in PBS medium. Aliquots were removed at various times, and the number of colony-forming units/mL ( $10^{-4}$  CFU.mL<sup>-1</sup>) was determined after incubation of the LB agar plates during overnight at 37 C.

<b>Time of incubation</b>	<b>Control</b>	<b>Cg-Def</b>
0 min	40.7	40.7
2 min	41.6	0
10 min	49.2	0
90 min	66.4	0
15 h	nd <sup>a</sup>	0

<sup>a</sup> not determined

**Table 3. Effect of sodium chloride on Cg-Def antimicrobial activity.**

MIC values were determined on the bacterial strains *Micrococcus lysodeikticus* and *Brevibacterium stationis* using liquid growth assay. In this table, only the lowest concentration of Cg-Def that causes 100% growth inhibition is indicated. Final OD<sub>600</sub> of control *M. lysodeikticus* and *B. stationis* cultures without Cg-Def in the same corresponding medium is indicated in parentheses.

Final NaCl concentration in poor broth	MIC in $\mu\text{M}$ (OD <sub>600</sub> after 24 h in the control experiment without peptide)	
	<i>M. lysodeikticus</i>	<i>B. stationis</i>
85 mM <sup>a</sup>	0,01 (0,6)	0,2 (0,55)
300 mM	0,01 (0,4)	0,2 (0,5)
600 mM <sup>b</sup>	0,01 (0,15)	0,2 (0,35)
1000 mM	- <sup>c</sup>	0,06 (0,1)

<sup>a</sup> Control poor broth.

<sup>b</sup> NaCl concentration close to the value measured in sea water.

<sup>c</sup> No bacterial growth in the control experiment

## Figure legends

**Figure 1: Sequence alignment and genomic organization of representative of the defensin family AMP.** (A) Amino acid sequence alignment of defensins from arthropods, mollusks and a fungus. Oyster *Crassostrea gigas* (Cg-Def); mussels *Mytilus edulis* (DEFA-MYTED, [P81610], DEFB-MYTED, [P81611]) and *M. galloprovincialis* (MGD-1, [P80571]; MGD-2, [AAD52660]); tick *Dermacentor variabilis* (varisin, VSNA1, [AAO24323]), *Ornithodoros moubata* (Om-DEF-C, [AB041814]), *Haemaphysalis longicornis* (HIDfs, [BAD93183]), *Boophilus microplus* (BOMICR, [AAO48943]) and *Ixodes scapularis* (IXSCA, scapularisin, [AAV74387]); dragonfly *Aeschna cyanea* (DEFI-AESCY, [P80154]); scorpion *Androctonus australis* (DEF4-ANDAU, [P56686]) and *Leiurus quinquestriatus* (DEF4-LEIQH, [P41965]) and the saprophytic fungus *Pseudoplectania nigrella* (plectasin, [CAI83768]). The nomenclature derived from the SwissProt database and the accession numbers are in brackets. Identical residues are shaded. The two additional cysteine residues found in the Cg-Def, MGD-1 and MGD-2 are in bold. (B) Cg-Def and MGD-1 sequence alignment manually modified according to the structure comparison (this work). The similar elements of structure are grey shaded and reported above. Conserved residues are in bold. Notice the shift for one cysteine (C20/C21) out of the eight. (C) Organization of *Cg-def* gene and the related genes coding MGD-2 and myticin A from the mussel *Mytilus galloprovincialis*. Boxes represent the coding region; the signal sequence appears in black and the mature peptide in white. Grey lines represent the untranslated regions of the exons. Potential sites of cleavage (mono [R] or dibasic [RR] and [KK]) involved in the splitting of the mature peptide (white) from the C-terminal peptide spotted) are indicated. Accession numbers: Cg-Def: CAJ19280, Cg-Def genomic sequence AJ582630. Gene organization of MGD-2b is from Mitta *et al.* (39) and incomplete structure of Myticin A (dotted box) is deduced from Mitta *et al.* (51)

**Figure 2: Quality control of recombinant *Crassostrea gigas* defensin (Cg-Def) and detection of the natural form in acidic extracts of mantle tissue.** (A) Following reduction, refolding of the recombinant peptide was performed at room temperature in the presence of Tris-HCl buffer. The refolding was monitored by RP-HPLC using an acetonitrile gradient in acidified water (for details see Materials and Methods). The grey and black lines correspond to the linear and to the refolded (see \*) recombinant peptides, respectively. The molecular mass of the refolded peptide was measured by ESI-MS (B1) and MALDI-TOF-MS (B2). The molecular mass (average value) measured at 4,634.10 Da is in perfect agreement with the calculated average molecular mass at 4,634.34 Da. A difference of 3 Da was observed between the molecular masses measured using MALDI-TOF-MS ( $m/z$  4638.46, see B2) versus the one detected in ESI-MS at 4,634.10 Da ( $m/z$  4,635.10). (C, D) Following purification by RP-HPLC of an acidic extract of mantle tissue, the two fractions active against the tested microorganisms were analyzed by MALDI-TOF-MS. The first fraction (see the arrow head in panel C) eluting at the same retention time as the recombinant-refolded Cg-Def contains ions of  $m/z$  4,638.96 corresponding to the one measured by MALDI-MS for the recombinant peptide. In the second bioactive fraction (see the arrow head in panel D), ions of  $m/z$  4,631.45 were detected.

**Figure 3: Structure of Cg-Def.** Stereo-view of the ten best conformers of Cg-Def. The 4-39 heavy atoms of the backbone were used for the superimposition. The mean pairwise rmsd is of  $0.43 \pm 0.09$  Å. The four disulfide bridges are labeled and displayed as dashed lines. Hydrophobic, positively and negatively charged side chains are colored in green, blue and red, respectively.

**Figure 4: Structure comparison of Cg-Def and MGD-1.** (A) Global fold comparison of the Cg-Def (left) and MGD-1 (right) structures. The structures prepared using MOLSCRIPT show the arrangement of the four disulfide bonds (in yellow). (B) - With a similar orientation, comparison of their Van der Waals surface showing the location of the hydrophobic (green) and positively charged (blue) residues. For this orientation, the W31 side chain of Cg-Def is hidden. Most of the labeled residues are conserved in the two peptides. Aspartic and cysteine residues are in red and yellow, respectively.

**Figure 5: Tissue expression of *Cg-def* mRNA analyzed by real time quantitative RT-PCR.** Each value is the mean of four pools of four animals. Hc : Hemocytes ; M : Mantle ; PAM : Posterior Adductor Muscle ; DG : Digestive Gland ; G : Gills ; H : Heart ; LP : Labial Palps. ND: Not detected. Bars represent the relative *Cg-def* transcript levels normalized to GAPDH transcript levels.

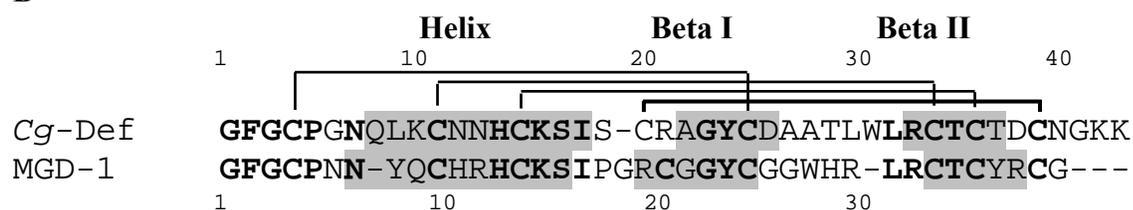
**Figure 6: *Cg-def* mRNA expression in oyster mantle following a bacterial challenge.** Results are means  $\pm$  SE of three independent experiments realized on a pool of ten oysters from non-stimulated (white) and stimulated (black) oysters, at 24 and 48 h post-infection. Bars represent the relative *Cg-def* transcript levels normalized to elongation factor transcript levels.

Figure 1

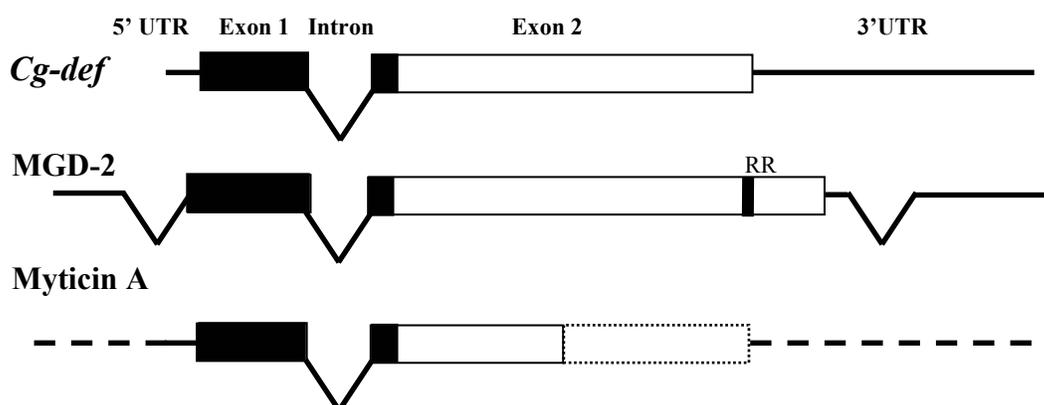
A

<i>Cg</i> -DEF	GFGC--PGNQLK--CNNHCKSIS--CRAGYCDAA TLWLRCTCTDCNGKK
MGD-1	GFGC--PNN-YQ--CHRHCKSIPGR <b>C</b> -GGYCGGWHR-LRCTCYR <b>CG</b>
MGD-2	GFGC--PNN-YA--CHQHCKSIRGY <b>C</b> -GGY CAGWFR-LRCTCYR <b>CG</b>
DEFA-MYTED	GFGC--PND-YP--CHRHCKSIPGRX-GGYCGGXHR-LRCTCYR
DEFB-MYTED	GFGC--PND-YP--CHRHCKSIPGRY-GGYCGGXHR-LRCTC
DEFI-AESCY	GFGC--PLDQMQ--CHRHCQTITGRS-GGYCSGPLK-LTCTCYR
VSNA1	GFGC--PLNQGA--CHNHCRSIRRR--GGYCSGI IK-QTCTCYRN
DEF4-ANDAU	GFGC--PFNQGA--CHRHCRSIRRR--GGY CAGLFK-QTCTCYR
DEF4-LEIQH	GFGC--PLNQGA--CHRHCRSIRRR--GGY CAGFFK-QTCTCYRN
<i>Om</i> -DEF-C	GYGC--PFNQYQ--CHSHCSGIRGYK-GGYCKGLFK-QTCNCY
HIDfs	GFGC--PLNQGA--CHNHCRSIGRR--GGY CAGI IK-QTCTCYRK
BOMICR	GFGC--PFNQGA--CHRHCRSIRRR--GGY CAGLIK-QTCTCYRN
Scapularisin	GFGC--PFDQGA--CHRHCQSIGRR--GGY CAGFIK-QTCTCYHN
Plectasin	GFGCNGPWEDDMQCHNHCKSIKGYK-GGYCAKGGFV--CKCY

B



C



**Figure 2**  
(black and white figure)

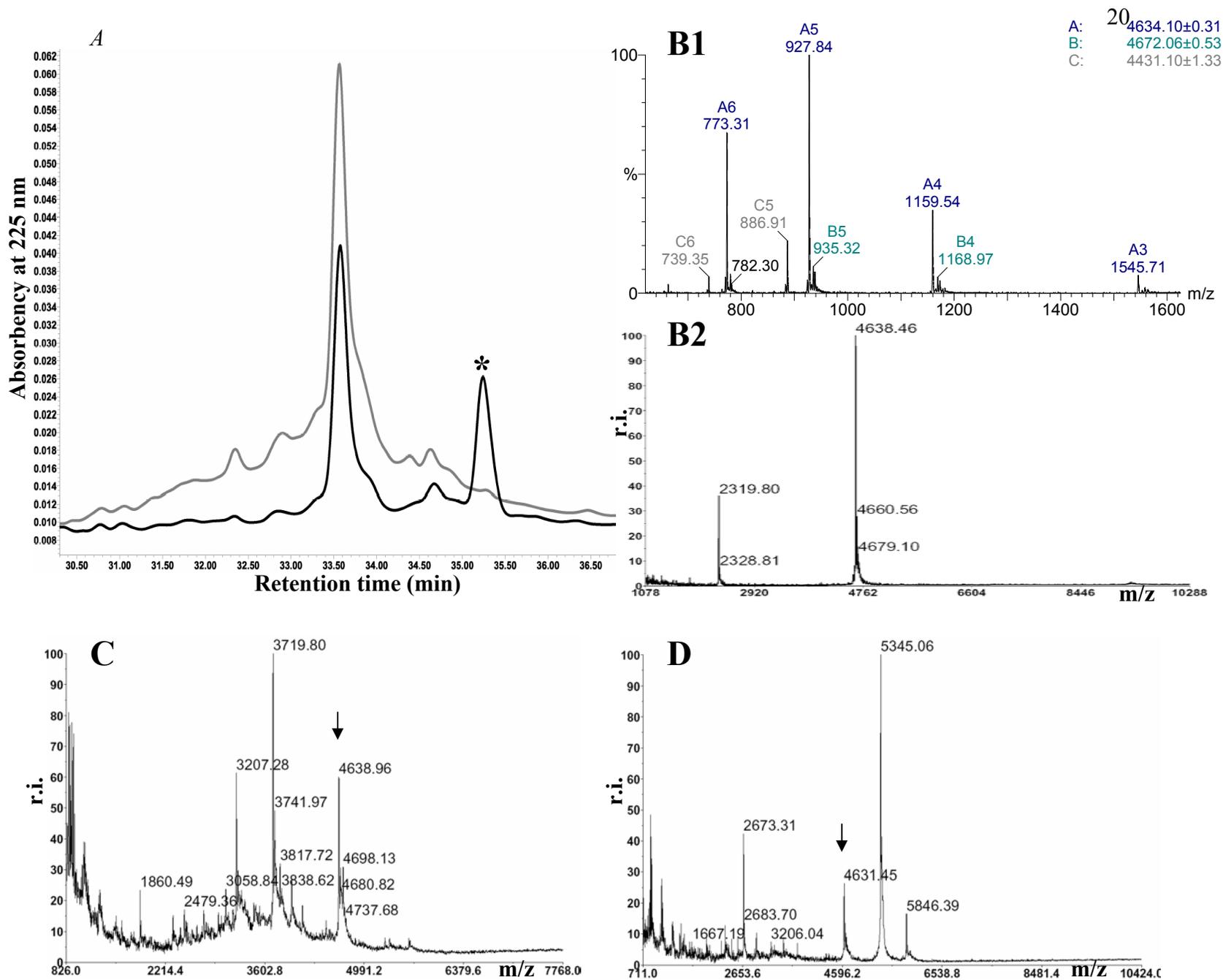


Figure 3

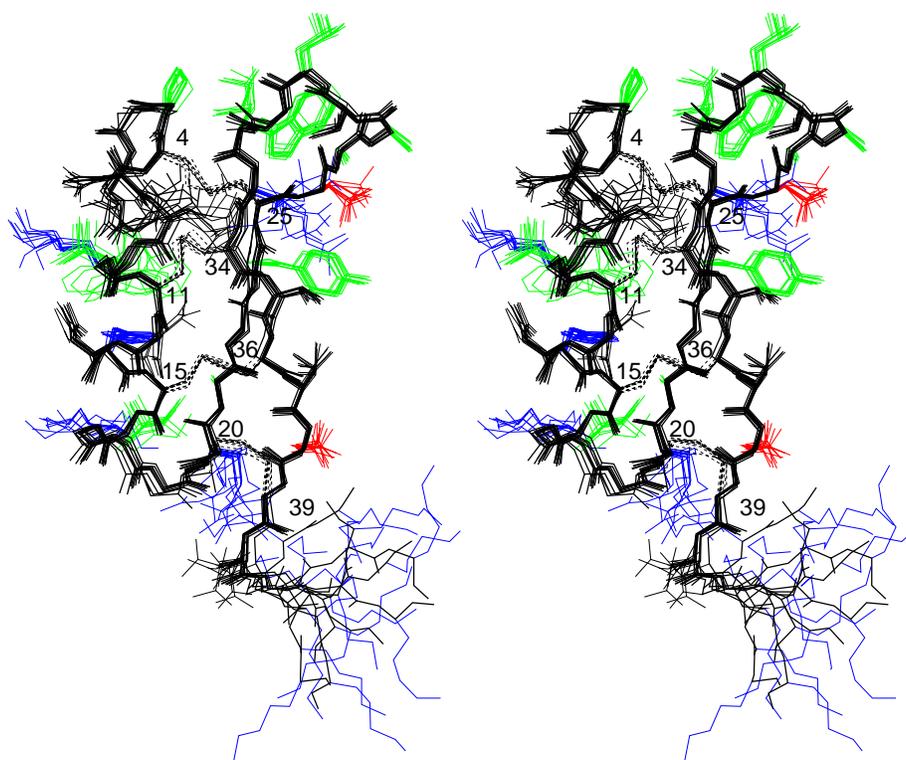
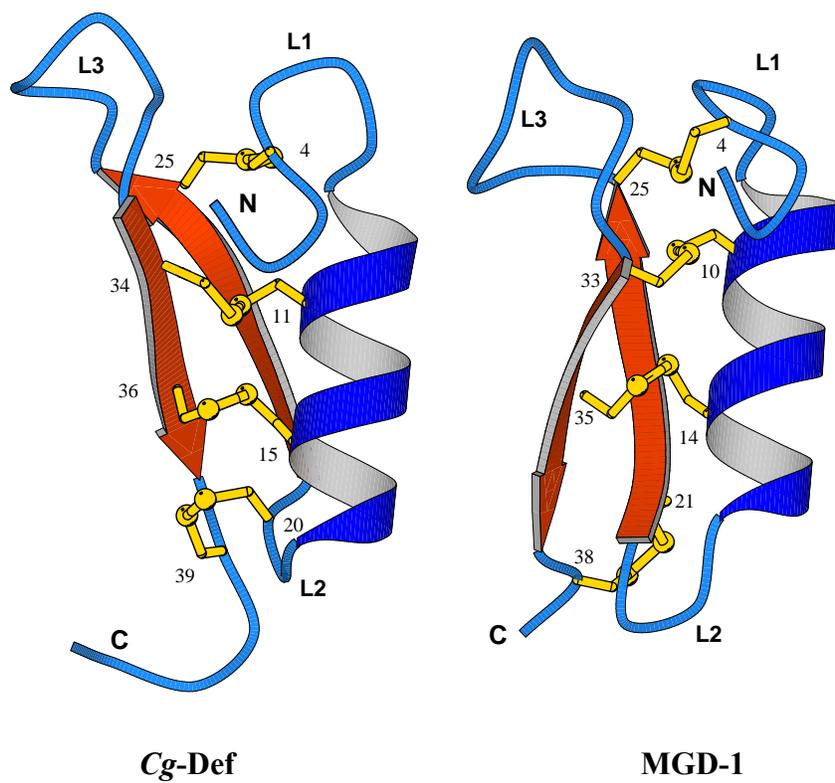


Figure 4

A



B

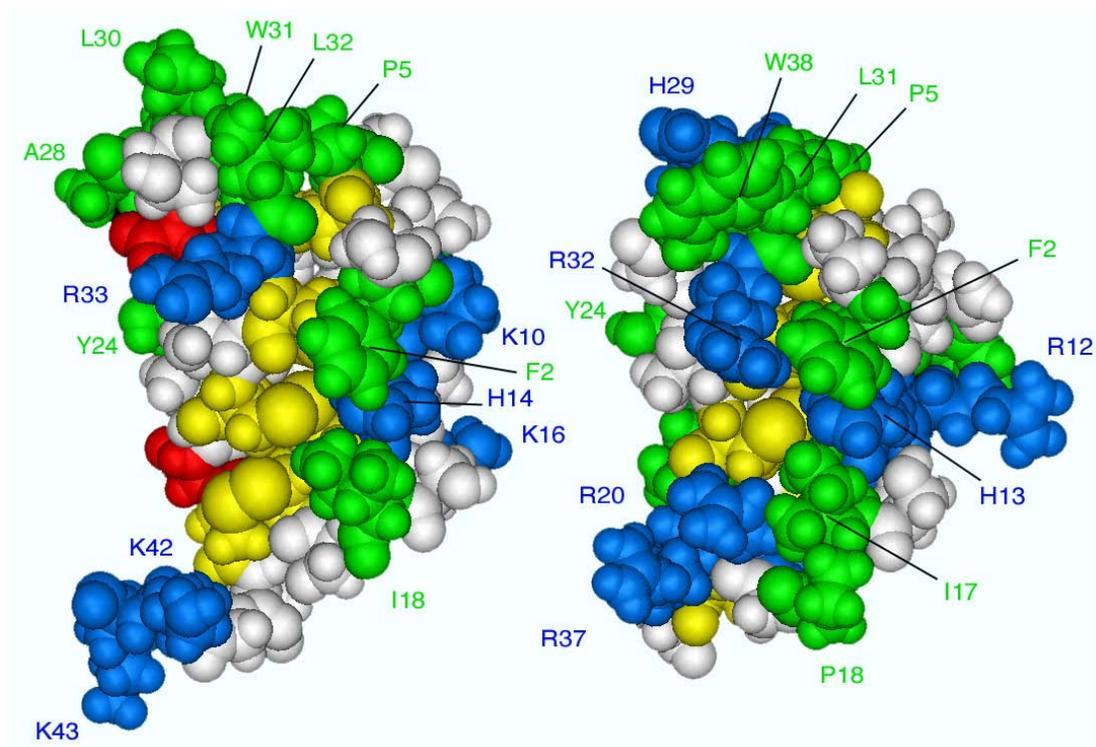


Figure 5

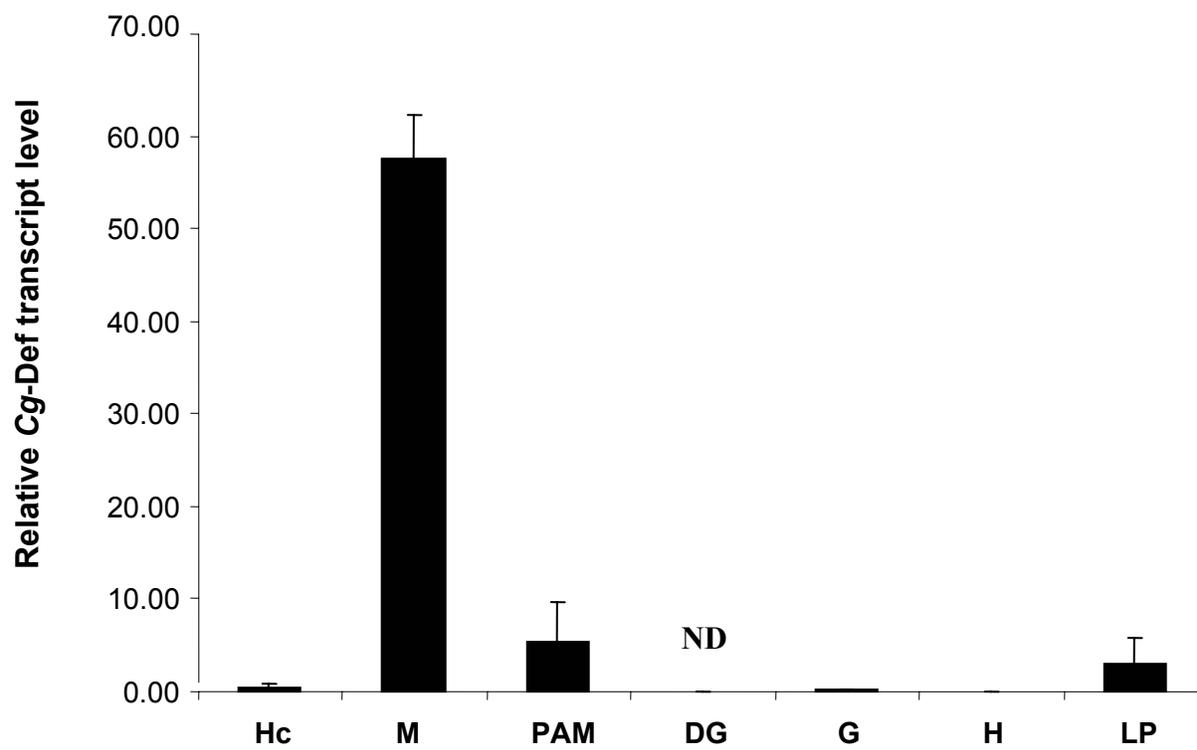


Figure 6

