Mar Biotechnol (2011) 13:857–867
DOI 10.1007/s10126-010-9345-4

ORIGINAL ARTICLE

Diversity of Coding Sequences and Gene Structures of the Antifungal Peptide Mytimycin (MytM) from the Mediterranean Mussel, *Mytilus galloprovincialis*

Molruedee Sonthi · Mylène Toubiana · Alberto Pallavicini · Paola Venier · Philippe Roch

Received: 6 July 2010 / Accepted: 30 November 2010 / Published online: 19 January 2011

Abstract Knowledge on antifungal biomolecules is limited compared to antibacterial peptides. A strictly antifungal peptide from the blue mussel, *Mytilus edulis* named mytimycin (MytM) was reported in 1996 as partial NH_2 33 amino acid sequence. Using back-translations of the previous sequence, MytM-related nucleotide sequences were identified from a normalized *Mytilus galloprovincialis* expressed sequence tag library. Primers designed from a consensus sequence have been used to obtain a fragment of 560 nucleotides, including the complete coding sequence of 456 nucleotides. Precursor is constituted by a signal peptide of 23 amino acids, followed by MytM of 54 amino acids (6.2–6.3 kDa, 12 cysteines) and C-terminal extension of 75 amino acids. Only two major amino acid precursor sequences emerged, one shared by *M. galloprovincialis* from Venice and Vigo, the other belonging to *M. galloprovincialis* from Palavas, with nine amino acid differences between the two MytM. Predicted disulfide bonds suggested the presence of two constrained domains joined by amino acidic NIFG track. Intriguing was the presence of conserved canonical EF hand-motif located in the C-terminus extension of the precursor. The *MytM* gene was found interrupted by two introns. Intron 2 existed in two forms, a long (1,112 nucleotides) and a short (716 nucleotides) one resulting from the removal of the central part of the long one. Both the short (GenBank FJ804479) and the long (GenBank FJ804478) genes are simultaneously present in the mussel genome.

Keywords Antimicrobial peptide · Mytimycin · Polymorphism · Molecular diversity · Gene structure · Innate immunity · Bivalve · Mussels · Mollusks · Mytilus

Abbreviations

EST	Expressed sequence tag
gDNA	Genomic DNA
MytM	Mature moiety of mytimycin precursor
MytM-P	Mature moiety of mytimycin from Palavas
MytM-V	Mature moiety of mytimycin from Vigo and Venice
UTR	Untranslated region

Introduction

The development of large-scale aquaculture resulted in increasing catastrophic losses due to various diseases, both in fish, crustacean, and molluscs. Although utilization of antibiotics partially solved the problem in the past, their abuse accelerated the selection of drug-resistant micro-organisms which ultimately might display negative effects on human health, and was therefore banned by international

Electronic supplementary material The online version of this article (doi:10.1007/s10126-010-9345-4) contains supplementary material, which is available to authorized users.

M. Sonthi · M. Toubiana · P. Roch (✉)
Ecosystèmes Lagunaires, CNRS-Université Montpellier 2,
cc093, place E. Bataillon,
34095 Montpellier, Cedex 05, France
e-mail: Philippe.Roch@univ-montp2.fr

A. Pallavicini
Department of Life Science, University of Trieste,
P.le Valmaura, 9,
34148 Trieste, Italy

P. Venier
Department of Biology, University of Padova,
Via Bassi 58/B,
35131 Padova, Italy

regulations. In absence of possible vaccination, one alternative is to stimulate the natural anti infectious mechanisms from the animals. To achieve such goal, we need better understanding of the innate immunity. Since decades, precious information has been collected regarding antibacterial defences and numerous sequences were included in specific databases. For instance, the antimicrobial peptide databse 2 (http://aps.unmc.edu/AP/main.php) contained nearly 2,000 entries with 1,242 antibacterial and 483 antifungal proteins (Wang et al. 2009). Regarding antifungal proteins, 13 classes have been described considering only the proteins of molecular mass greater than 5 kDa (Selitrennikoff 2001). Most of these classes referred to plant proteins and only a restricted number of antifungal peptides have been isolated from invertebrates. Best known are the ones from arthropods, such are the antifungal protein (AFP) from the flesh fly *Sarcophaga peregrine* larvae (Iijima et al. 1993), drosomycin (Fehlbaum et al. 1994), and metchnikovin (Levashina et al. 1995) from the fruit fly *Drosophila melanogaster*, thanatin from the hemipteran insect *Podisus maculiventris* (Fehlbaum et al. 1996), penaeidins from the Pacific white shrimp *Litopenaeus vannamei* (Destoumieux et al. 1997) and from other *Penaeus* species (Gueguen et al. 2006a), heliomicin from the lepidopteran *Heliothis virescens* (Lamberty et al. 1999), termicin and spinigerin from the termite *Pseudacanthotermes spiniger* (Lamberty et al. 2001), scarabaecin from the coconut rhinoceros beetle *Oryctes rhinoceros* (Tomie et al. 2003), antilipopolysaccharide factor from the black tiger shrimp *Penaeus monodon* (Somboonwiwat et al. 2005) and from the Chinese silkworm *Bombyx mori* (Zhang et al. 2008). Even cecropin, the first antimicrobial peptide isolated in 1981 from the giant silk moth *Hyalophora cecropia*, was later recognized as possessing also antifungal properties (Ekengren and Hultmark 1999).

Related to molluscs, first report on antibacterial activity was from mucus of the giant snail, *Achatina fulica*; achacin of 150 kDa composed of two subunits (Kubota et al. 1985). Also from Gastropod was aplysianin P, a 60-kDa bacteriostatic glycoprotein from the purple fluid of the sea hare, *Aplysia kurodai* (Yamazaki et al. 1990). It was only in 1996 that true antimicrobial peptides (i.e., cationic, 15–45 amino acid residues) have been isolated simultaneously from the Mediterranean mussel, *Mytilus galloprovincialis* (Hubert et al. 1996) and the blue mussel, *Mytilus edulis* (Charlet et al. 1996) with in that last species, a 6,233 Da strictly antifungal peptide named mytimycin (MytM) containing 12 cysteines and reported as NH_2-terminal sequence of 33 amino acids (Charlet et al. 1996). Since that time, numerous antimicrobial peptides have been isolated from marine bivalves, only some of them sharing antibacterial and antifungal activities, such are myticin B (Mitta et al. 1999a), defensin MGD1 (Mitta et al. 1999b) and mytilin B and D (Mitta et al. 2000b) from *M. galloprovincialis*, and defensin from the oyster *Crassostrea gigas* (Gueguen et al. 2006b). Extensive efforts performed to isolate antimicrobial peptides or to characterize their mRNAs in *Mytilus* identified several families and multiple isoforms of antimicrobial peptides, nevertheless, MytM was never reevidenced.

In the present report, (1) we identified a number of putative MytM sequences from a normalized *M. galloprovincialis* expressed sequence tag (EST) library, (2) established the full coding sequences from hemocytes of *M. galloprovincialis*, (3) evaluated their diversity in mussel populations from 3 different geographic origins, (4) looked for phylogenetic relationships, and (5) reported the structures of the two MytM genes

Material and Methods

Mussels

Adult mussels, *M. galloprovincialis* (6–7 cm shell length), were collected in three different locations during June–July 2005–2007. Mussels from Ria de Vigo (Vigo, Spain) were purchased from the shellfish farm Mariscos Ria de Vigo SL. Mussels from the French Mediterranean coast were purchased from the marine farm Les Compagnons de Maguelone (Palavas, France). Mussels from the Adriatic Sea were purchased from the marine farm Mitilpesca (Venice, Italy). Mussels were acclimated for 24 h in the laboratory in a flow-through system of oxygenated sea water before sampling.

Hemocyte Collection and RNA Extraction

Hemolymph was collected from the posterior adductor muscle with a 1 ml syringe containing 0.2 ml of the anticoagulant modified Alsever's solution buffer. Hemolymphs of 10 mussels per sampling site were pooled and hemocytes were pelleted by 15 min centrifugation at 800×*g*, 4°C. Total RNA was extracted with Trizol Reagent protocol (Invitrogen) and the pellets were resuspended in 20 µl of sterile water. For genomic analyses, hemocytes from individual mussels were kept separate.

RT, PCR, and cDNA Cloning

First-strand cDNAs were synthesized using random hexaprimers (Invitrogen Life Technologies) and murine leukemia virus reverse transcriptase (Promega), and purified through QIAquick Column (Qiagen) then kept in sterile water at −20°C until use.

Forward F1 and reverse R1 PCR primers were hand-designed according to the consensus sequence from the

EST library MgNOR01 (Venier et al. 2009). Following denaturizing of 35 ng cDNA template for 2 min at 94°C, PCR was performed by 30 cycles of denaturing at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1 min, with final extension at 72°C for 5 min. The unique amplicon of 437 nucleotides was cloned in the plasmid pCR 2.1 TOPO from the TOPO TA Cloning kit (Invitrogen). Extraction and purification of the insert were done using Wizard Plus SV Miniprep kit (Promega) and specificity has been confirmed with several sequencings (Millegen, Labèges-France).

The 5′-extension was performed using reverse primer R3, and nested PCR was with R4 then R5 using 5′/3′ rapid amplification of cDNA ends (RACE) kit according to manufacturer's protocol (Roche).

Full coding sequence was amplified using the forward primer F3 designed from extended sequence obtained from 5′-RACE and reverse primer R7, according to the following program: 1 μg of cDNA template, initial denaturing at 94°C for 2 min followed by 35 cycles of denaturing at 94°C for 1 min, annealing at 55°C for 1 min and elongation at 72°C for 1 min, and a final extension at 72°C for 7 min. The unique amplicon of 560 nucleotides was cloned as previously. White *Escherichia coli* colonies were individually transferred to deep agar containing Luria-Bertoni medium and 50 μg/ml kanamycin, distributed into 96-well microtiter plates and sent to Agowa GmbH (Berlin, Germany) for sequencing based on M13 universal primers. Each clone was bidirectional sequenced and the sequences corrected accordingly.

cDNA Sequence Analysis

Several adjustments and comparisons have been made: (1) the nucleotides from upstream and downstream the primers, including the primer sequences, were removed, (2) all the sequences were aligned using Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html) and clustered according to nucleotide sequences, (3) untranslated regions (UTRs) were removed and coding sequences compared using Multalin. The different coding sequences were translated into pro-peptides (http://www.expasy.ch/tools/dna.html). Structural analysis identified signal peptide, mature region and C-term extension, and the resulting amino acid sequences have been compared using Multalin. Isoelectric point and instability index were calculated by http://au.expasy.org/cgi-bin/protparam.html. Predicted disulfide bonds were established by http://scratch.proteomics.ics.uci.edu. Predictable structure of MytM has been established using Phyre (http://www.sbg.bio.ic.ac.uk/phyre/). Relationship with known proteins was investigated using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Presence of conserved domains was analyzed using http://www.expasy.org/scanprosite/. Evolutionary relationships of antifungal moieties were inferred using the neighbor-joining algorithm MEGA-4 (http://www.megasoftware.net). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogeny tree. The evolutionary distances were computed using the maximum composite likelihood method.

Genome Sequence Analysis

Genomic DNA (gDNA) from individual *M. galloprovincialis* was extracted from hemocytes with DNAzol (Invitrogen). PCR amplification sibling the 3′end was done with primers F1 and R7. For 5′end, PCR amplification was done with primers F3 and R3. PCR programs were identical to the one used with F1-R1 (see "RT, PCR, and cDNA Cloning" section). Sequence analysis consisted in the removal of nucleotides from upstream and downstream the primers, including the primer sequences, and alignments using Multalin.

Results

Identification of Mytimycin-Related Sequences

The screening of the *M. galloprovincialis* EST integrated database called MytiBase (available at http://mussel.cribi.unip.it) with the 33 amino acid N-terminal sequence reported from the blue mussel, *M. edulis* (accession number Swissprot P81614) (Charlet et al. 1996) resulted in the identification of one sequence, annotated mytimycin-like0. Second blast performed on the same MytiBase with the mytimycin-like0 sequence identified 16 sequences, originally not annotated, but which can be organized in 5 mytimycin-like clusters (Fig. 1a). Differences in amino acid sequences appeared between them and with the original *M. edulis* sequence. One consensus sequence (Fig. 1b) was constructed and PCR primers (F1 and R1) have been designed to look for mytimycin transcripts in hemocytes from the Palavas-France *M. galloprovincialis* mussels. Performing later searches at NCBI with nucleotide (BlastN) and translated nucleotide (BlastX) queries revealed strong similarities (*E* value $5E^{-70}$) between the five mytimycin-like sequences and segments of the two genomic mytimycin sequences we deposited in GenBank (Fig. 2), confirming the mytimycin origin of the five EST sequences. In addition, weak similarity (*E* value $5E^{-08}$) was observed with the original *M. edulis* partial sequence.

mRNA Sequences and Structures

Alignment of F1-R1 amplicon with the consensus sequence revealed that the sequence of 437 nucleotides

Fig. 1 Identification of mytimycin-like from *M. galloprovincialis*. **a** Alignment of the amino acid sequence P81614 reported from *M. edulis* with the amino acid translation of the five consensus nucleotide sequences identified from a normalized *M. galloprovincialis* EST library. Cysteines are *gray boxed*. **b** Inferred consensus sequence and primer F1 and R1 used to screen *M. galloprovincialis* hemocyte mRNA from Palavas

obtained corresponded to mytimycin, with few differences (not illustrated). The mussel EST library from which the mytimycin-like sequences came from was constructed from mixed samples of RNA coming from variously challenged *M. galloprovincialis* from Spain and Italy. Consequently, the observed differences cannot be specifically related to any stimulation or geographic origin.

The consensus sequence included the polyA tail, but was incomplete on the 5′ end (Fig. 1b). In order to design a forward primer within the 5′-UTR, allowing amplification of the full coding sequence, 5′-RACE was performed and resulted in 28 additional nucleotides. A new primer F3 was designed within the 5′-UTR and PCR with F3 and R7 gave a 560 nucleotide amplicon, reduced to 456 nucleotides when considering only the full coding sequence (Fig. 2). Alignments with original partial amino acid sequence located the first amino acid of MytM as D or A, followed by a duplet of C (Fig. 1a). The presence of an upstream signal peptide of 23 amino acids was deduced from translation of the mRNA 5′-end 120 nucleotide sequence, and from the presence of a methionine downstream a stop codon.

Taking into account the MM of 6,233.5 Da measured by mass spectrometry for *M. edulis* MytM (Charlet et al. 1996) and considering 12 cysteines engaged in six intrachain disulfide bonds, the theoretical MM of linear *M. edulis* MytM would be 6,245.5 Da. The MytM sequence from *M. galloprovincialis*, including 54 amino acids and 12 cysteines, would display the theoretical MM of 6,247.1 Da, consistent enough to consider *M. galloprovincialis* MytM constituted of 54 amino acids. It must be noted that the N-terminal amino acid sequence translated from the *M. galloprovincialis* consensus sequence was not identical to the one from *M. edulis* and we hypothesized that the rest of the molecule will not be identical, explaining the discrepancy in MM. The 75 amino acid sequence downstream corresponded to a C-terminal extension ended by the stop codon TAG. Separate BLAST analysis of amino acid sequences from signal, MytM and C-terminal extension, were not revealing any similar already recorded sequences. Consequently, mytimycin precursor appeared original.

mRNA Polymorphism and Codon Changes

To exclude errors related to the sequencing process, the 96 clones contained in one full microtiter plate have been resequenced. Only few single nucleotide changes were observed in randomly distributed locations. Such changes were considered as resulting from errors introduced during sequencing process, and only nucleotide changes located at the same position in both forward and reverse sequences were noted. In addition, we eliminated the few nucleotide changes found in only one clone, as resulting probably from errors introduced by the first PCR before cloning. Consequently, the polymorphism analyzed herein was based on nucleotide changes found at the same position in at least two clones.

Within a Pool of 10 Hemolymph mRNAs from M. galloprovincialis of Palavas, France A total of 181 clones were randomly selected, sequenced and virtually translated in amino acids. Four consistent nucleotide changes located at −4, +6, +48, and +300 (Table 1 Supplementary data) have been identified, defining nine different coding sequences. Frequencies of occurrence were extremely different, with two coding sequences accounting for 88% of the 181 clones and five coding sequences present in only

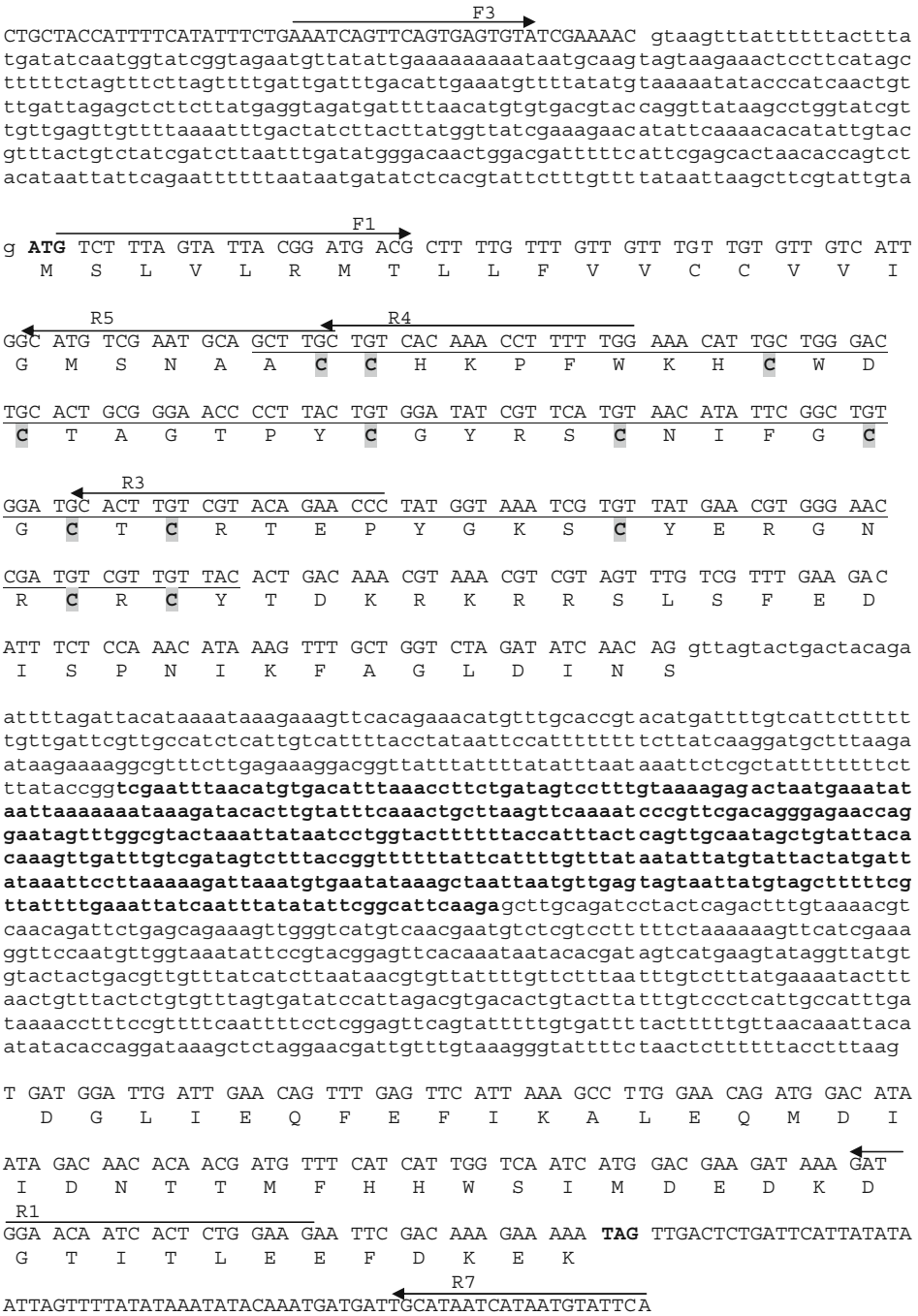

Fig. 2 Nucleotide sequence of *mytimycin* genes from *M. galloprovincialis* from Palavas, MMg1 (GenBank FJ804479) and MMg2 (GenBank FJ804478), and predicted amino acid sequences of the unique precursor. The translation start (ATG) and stop (TAG) codons are in *bold*. The coding region corresponding to the 54 amino acids of MytM is underlined, with the 12 cysteines *gray boxed*. The 23 amino acid sequence upstream MytM corresponds to signal peptide and the 75 amino acid sequence downstream MytM corresponds to C-terminal extension. The two introns are in lower cases with additional part within intron 2, specific to MMg2 gene, in *bold*. Primer locations are indicated by *long arrows*

one or two clones. Interestingly, only the nucleotide change located within the signal peptide (at −4) resulted in amino acid substitution, asparagine found in 126 clones, being replaced by lysine found in 55 clones. The three other nucleotide changes were silent and no mutation was noticed within partial UTRs (not illustrated). Consequently, mytimycin precursor from Palavas existed as two amino acid sequences, MytM being unique.

Within a Pool of 10 Hemolymph mRNAs from M. galloprovincialis of Venice, Italy Ninety-three clones were randomly selected, sequenced and virtually translated in amino acids. Six consistent nucleotide changes were located at −60, −59, −33, +216, +305, and +355 (Table 1 Supplementary data), defining six different coding sequences. As for mussels from Palavas, France, frequencies of occurrence were extremely different, with one coding sequence accounting for 94% of the clones. Only nucleotide change at +216 was silent. Nucleotide changes at −60 and −59 resulted in proline or leucine in one clone each compared to serine in 91 clones. Nucleotide change at +305 resulted in serine in two clones compared to asparagine in 91 clones.

Table 1 Amino acid sequences of mytimycin precursors and characteristics

Amino acid sequences of major mytimycin precursors, with predicted disulfide bonds	
Mytimycin-P	
Signal peptide	MSLVLRMTLLFVVCCVVIGMSNA
MytM-P	2 3 11 14 21 26 31 33 35 44 51 53 ACCHKPFWKHCWDCTAGTPYCGYRSCNIFGCGCTCRTEPYGKSCYERGNRCRCY
C-terminal extension	TDKRKRRSLSFEDISPNIKFAGLDINSDGLIEQFEFIKALEQMDIIDNTTMFHHWSIMDEDKDGTITLEEFDKEK
Mytimycin-V	
Signal peptide	MSLSLRMTLLFVICCVVIGMANA
MytM-V	2 3 11 14 21 26 31 33 35 44 51 53 DCCHRPYYYHCWDCTAATPYCGYRPCNIFGCGCTCRTEPHGKSCYERGDRCRCY
C-terminal extension	SDKRRRRSLSFEDMSANIKFAGLDINSDGLIEQFEFIKALEQMDITDNTTMFHHWSIMDEDKDGAITLEEFDKEN

Deduced amino acid sequences of the two mytimycin precursors found in majority among *M. galloprovincialis* from Palavas (Mytimycin-P) and from Venice and Vigo (Mytimycin-V). Different amino acids are *gray boxed*. Note the identical predicted cysteine patterns identifying two constrained domains joined by the common NIFG sequence (*double underlined*)

Finally, nucleotide change at +355 resulted in serine in two clones compared to alanine in 91 clones. None of the detected changes mapped neither within the open reading frame corresponding to MytM nor within partial UTRs (not illustrated). Consequently, precursor from Venice encompassed five amino acid sequences, MytM being unique.

Within a Pool of 10 Hemolymph mRNAs from M. galloprovincialis of Vigo, Spain Ninety-two clones were randomly selected, sequenced and virtually translated in amino acids. A total of 35 nucleotide changes were located, defining 10 different coding sequences (Table 1 Supplementary data). One particular coding sequence was present in 69 clones, and another present in 10 clones. Other coding sequences have been found in 1–3 clones only. Eight mutations were silent. Amino acid replacements affected the signal peptide (four cases) as well as the C-terminal extension (eight cases). Unusual was the presence of nine amino acid replacements within MytM coding sequence. Meanwhile, one of these sequences was found in 82 clones, another in seven clones, and the three other amino acid sequences were found in only one clone each. One nucleotide change was observed in the 5′UTR and five in the 3′UTR (not illustrated). Consequently, precursor from Vigo existed as seven amino acid sequences, with five MytM amino acid sequences.

Diversity of Mytimycin Precursors

Alignments of the virtually translated precursor sequences from Palavas, Venice, and Vigo have been performed. Among the 21 different amino acid sequences, the majority from Venice (88 clones out 93) and the majority from Vigo (72 clones out of 92) were identical, and named mytimycin-V. Such sequence was different from the majority one from Palavas (126 clones out of 181), named mytimycin-P. Consequently, among the pooled *M. galloprovincialis* originated from three distant geographic areas, only two mytimycin precursors were dominant (Table 1), a phenomenon confirmed by preliminary comparisons performed on new FLX 454 sequence MytM data.

Alignments have been also performed addressing separately the signal region, MytM, and C-terminal extension (Table 2). The two signal amino acid sequences from Palavas differed by only one amino acid. Such sequences

Table 2 Diversity of signal peptide, MytM and C-terminal extension amino acid sequences from *M. galloprovincialis* from Palavas, Venice and Vigo. The original MytM partial NH_2 amino acid sequence reported in 1996 from *M. edulis* was added for comparison; *gray boxed* are the amino acids different from Palavas and from Venice sequences. Note a unique sequence of MytM expressed by the mussels from Palavas and from Venice

Signal peptide

Amino acid sequences	Number of clones		
	Palavas	Venice	Vigo
MSLVLRMTLLFVVCCVVIGMSNA	126	-	5
.....K.	55	-	4
...S.....A..	-	1	-
...S.....I.....A..	-	90	83
...S.....I.....D.A..	-	-	-
...P.....I.....A..	-	1	-
...L.....I.....A..	-	1	-

MytM

Amino acid sequences	Number of clones		
	Palavas	Venice	Vigo
DCCRKPFRKHCWDCTAGTPYYGYSTRNIFGCTC--- from *M. edulis*			
ACCHKPFWKHCWDCTAGTPYCGYRSCNIFGCGCTCRTEPYGKSCYERGNRCRCY	181	-	5
.....P.....	-	-	3
.....H.....D.....	-	-	1
D...R.YYY.....A.....H.....D.....	-	-	1
D...R.YYY.....A.....P.....H.....D.....	-	93	82

C-terminal extension

Amino acid sequences	Number of clones		
	Palavas	Venice	Vigo
TDKRKRRSLSFEDISPNIKFAGLDINSDGLIEQFEFIKALEQMDIIDNTTMFHHWSIMDEDKDGTITLEEFDKEK	181	-	7
.....T.....A.....N	-	-	1
S...R.....M.A.....T.....A.....N	-	91	74
S...R.....M.A.....T.S.....S.....N	-	2	-
S...R.....M.A.....T.....G....A.....N	-	-	10

were absent from Venice and only marginally present in Vigo. On the opposite, the majority signal amino acid sequence from Venice was identical to the majority one from Vigo, and was absent from Palavas. Similarly, the unique MytM amino acid sequence from Palavas was absent from Venice and marginally present in Vigo, whereas Venice and Vigo shared their majority MytM amino acid sequence, not present in Palavas. Identical observation was made related to C-terminal extension: unique amino acid sequence from Palavas absent from Venice and marginally present in Vigo; Venice and Vigo sharing their majority amino acid sequence, absent in Palavas.

Patterns of Variation and Structure Prediction

Differences between the two majority precursors, mytimycin-P and mytimycin-V, were within the signal peptide (three amino acid different), MytM (nine amino acid different), and C-terminal extension (seven amino acid different; Table 1). Folded MytM possessed a calculated molecular mass from 6,235.1 (MytM-P) to 6,334.1 Da (MytM-V). They belong both to cationic peptides with similar estimated pI above 8. Their elevated calculated instability indexes (68.52 and 53.65) revealed possible rapid degradation of the peptides (Table 3). Disulfide bonds were predicted identical for both MytM-P and MytM-V as 2–11, 3–21, 14–26, 31–44, 33–51, and 35–53, suggesting the existence of two distinct domains joined by the common NIFG sequence (Table 1). Remarkable was the fact that the COOH terminus domain of both MytM, including 24 amino acids and three disulfide bonds, appeared very stable with an instability index of 35.51 and 35.12, compared to 90.72 and 60.18 for the NH_2-terminal domains. Three-dimensional structure failed to be predicted by SWISS MODEL (http://swissmodel.expasy.org/workspace) considering either MytM sequences, COOH- or NH_2-terminus domains. Submission of the five clustered amino acid sequences obtained from translations of the nucleotide sequences identified in the normalized *M. galloprovincialis* EST library (see Fig. 1a) to ScanProsite resulted in a constant signature related to canonical calcium-binding EF hand domain (not illustrated). Concerned amino acids were from 98 to 110 and from 133 to 145 although some shifting of few amino acids was observed according to the sequence.

Phylogenetic Relationships

Fourteen amino acid sequences of antifungal peptides have been aligned with MytM-P and MytM-V revealing two distinct clusters (Fig. 3). The two MytM sequences segregated with the cysteine-stabilized peptides (scarabaecin, thanatin, termicin, drosomycin, heliomicin, penaeidin-2 and 3, defensin MGD-1, mytilin B) but not with the linear peptides (metchnikovin, cecropin, spinigerin-1 and 2). Closest relationships were with termicin and mytilin B.

Gene Structures

PCR has been performed on gDNA isolated from one mussel using primers F1 and R7. Two amplicons of 1,246 and 1,642 nucleotides have been obtained, compared to the 437 nucleotide fragment obtained from mRNA, suggesting the existence of intron(s). Several sequencing of both amplicons revealed the presence of one intron interrupting the C-terminal extension with typical gt and ag flanking nucleotides (Fig. 2). Intriguingly, such intron existed in two different forms, a long of 1,112 nucleotides and a short of 716 nucleotides which was in fact the long intron with 396 nucleotides missing to the central part. Similar analysis has been done sibling the 5′-end with primers F3 and R3. A single amplicon of 674 nucleotides has been obtained whereas a fragment of 214 nucleotides was expected according to cDNA. The presence of a suspected intron of 460 nucleotides has been confirmed by sequencing, revealing its location just upstream the ATG of the first methionine. Moreover, the open reading frame corresponding to signal and MytM was not interrupted by any intron. The two gene sequences have been deposited in GenBank with the accession numbers FJ804478 for MMg2, including the long intron 2, and FJ804479 for MMg1, including the short intron 2.

gDNA from 16 mussels have been individually extracted and submitted to PCR with F3-R3 primers and with F1-R7

Table 3 Some characteristics of the two MytM

	MytM-P		MytM-V	
Molecular mass (as folded peptide)	6,235.1		6,334.1	
Estimated pI	8.66		8.03	
Estimated charge at pH 7.0	5.0		2.3	
Instability indexes	68.52		53.65	
	NH_2 domain	COOH domain	NH_2 domain	COOH domain
	90.72	35.51	60.18	35.12

Note the cationic nature and the relative instability of the two molecules. Remarkable is the difference in stability between the two domains, with highly instable NH_2 domains versus stable COOH domains

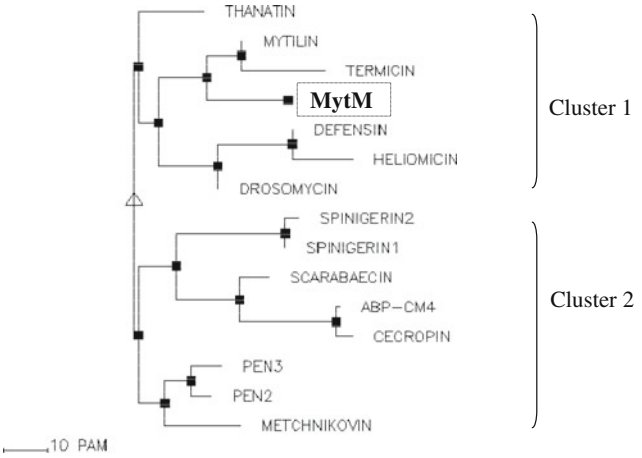

Fig. 3 Relationships between amino acid sequences of MytM and some other antifungal peptides presented as a phylogeny reconstruction using the method of minimum evolution considering 16 amino acid sequences from active moieties (see "Introduction" for references). The two amino acid sequences of MytM-V and MytM-P are too similar to be evolutionary separated by the software. Note the existence of two distinct clusters, cluster 1 containing only cysteine-stabilized sequences

primers. In all the 16 mussels, the presence of two bands of expected size demonstrated the simultaneous presence of MMg1 and MMg2. Consequently, each genome of *M. galloprovincialis* contained at least one copy of the 2 different *mytimycin* genes, the long and the short one, either located tandemly or as alleles.

Discussion

Number of expressed sequence tags (EST) dramatically increased in the last years, among which about 50% are unknown. Mining the libraries gave rise to the discovery of new molecules, such are the *C. gigas* defensin (Gueguen et al. 2006b), *Ruditapes decussatus* mytilin and myticins (Gestal et al. 2007) or C1q-domain (Prado-Alvarez et al. 2009), *Chlamys farreri* thioester-containing protein (Zhang et al. 2007), *Haliotis discus* C-type lectin (Wang et al. 2008), and the phenomenon of high-sequence variability of *M. galloprovincialis* myticin C (Pallavicini et al. 2008), for instance. Data mining of a normalized *M. galloprovincialis* EST library, now publicly available, made evident several mytimycin-related sequences, recently confirmed by 454 pyrosequencing using primers designed on the EST mytimycin-like transcripts (not illustrated).

Referring to mRNAs isolated from hemocytes, the mytimycin coding sequence differed in single nucleotide positions according to the geographical origin of the mussels and to the species: (1) few nucleotide replacements within Palavas and Venice mussels, and no one in common and (2) numerous nucleotide changes in Vigo's mussels, including all the ones from Palavas or Venice. Vigo is also the only population showing nucleotide changes in the UTRs: one out of 8 nucleotides in 5′ and five out of 56 nucleotides in 3′. The detected levels of transcript variation suggest that, in the three geographical areas, differences in pathogen load and possibly other factors, may have forced the *mytimycin* gene to diversify, as reported for mosquito defensin (Dassanayake et al. 2007). As expected, translation into amino acids also revealed specific patterns of variation: two precursors amino acid sequences from Palavas mussels, five from Venice, and seven from Vigo, but with one major sequence per group. In contrast, diversity of MytM was restricted to five amino acid sequences from Vigo and to a unique amino acid sequence from Palavas and Venice. Large coding sequence diversity resulting in restricted number of precursors and a unique mature peptide was also reported for mytilin B (Parisi et al. 2009) whereas, myticin C displays high sequence diversity (Pallavicini et al. 2008) even within individual mussels (Costa et al. 2009).

Overall, the primary structures of antimicrobial peptides are highly diverse. Tested mainly against bacteria, some of them were reported as also capable of antifungal activity. They encompassed sequences without cysteine, such are metchnikovin (3 kDa), spinigerin (3 kDa), cecropins (4.6 kDa; Boman et al. 1974) and AFP (7.1 kDa), and sequences stabilized by intrachain cysteine bonds, such are thanatin (2.5 kDa, two cysteines), scarabaecin (4 kDa, two cysteines), antilipopolysaccharide factor (11.0 kDa, two cysteines), termicin (4.2 kDa, six cysteines), heliomicin (4.7 kDa, six cysteines), penaeidins (5.7–6.1 kDa, six cysteines), mytilin B (3.9 kDa, eight cysteines), defensin MGD-1 (4.4 kDa, eight cysteines), and drosomycin (4.9 kDa, eight cysteines; see "Introduction" section for references). Our study on MytM confirmed the unusual structure of such strictly antifungal peptide of 6.2–6.3 kDa, containing 12 cysteines engaged in six intra molecular disulfide bonds defining two domains. Combining the analysis of cDNAs derived from *M. galloprovincialis* of three different geographical origins, only two major mytimycin precursors have been found with the same sequence shared between mussels from Vigo and Venice. One can hypothesize that this situation resulted from selection due to ancient coevolution between microbes and the immune system, as suggested for myticin C (Pallavicini et al. 2008) and defensin MGD2 but not demonstrated for mytilin B (Boon et al. 2009; Parisi et al. 2009).

Three-dimensional structures of some antifungal peptides are available. In particular, termicin, drosomycin, heliomicin, mussel defensin MGD-1, and mussel mytilin B have similar CSαβ motifs (Da Silva et al. 2003). Predicted disulfide bond pattern of MytM suggested the

existence of two distinct domains linked by the four amino acid NIFG. Elevated instability indexes of MytM are similar to the ones of mytilin C and myticin A, suggesting rapid degradation of secreted MytM although the peptide is stabilized by six disulfide bonds. On the contrary, other antimicrobial peptides such are mytilin A and myticin B, which included only four disulfide bonds, appeared very stable. In fact, instability of MytM seemed to be due to the NH_2-terminus domain. Analyzed separately, the sequence of 24 amino acids from COOH terminus domain appeared very stable, suggesting such domain is responsible for the antifungal activity.

Computer search for conserved motifs identified the canonical calcium-binding EF hand helix-loop-helix structural domain. Such motif consists in two α helices positioned roughly perpendicular and linked by a short loop region (Lewit-Bentley and Rety 2000). Upon binding to Ca^{2+}, the motif undergoes conformational changes that enable Ca^{2+} to regulate various functions. Its structure includes usually paired motifs of 29 amino acids but also exists as single, with different secondary structure, and even variations in length. Identification of such motif in mytimycin precursor amino acid sequence suggested calcium ions are important. On the other hand, EF hand-like domain is contained in the C-terminus extension (from amino acid 98 to 145), and therefore does not participate to the antifungal activity and its role in the C-terminus region remains to be investigated.

The gene of *M. galloprovincialis* mytimycin is interrupted by two introns. Remarkable is intron 2 which included two forms: a long one of 1,112 nucleotides and a short one of 716 nucleotides resulting from the removal of 396 nucleotides from the central part of the long one. The presence of two introns with great variability was also observed for the gene of myticin C (Pallavicini et al. 2008), whereas different are the structures of defensin MGD2, mytilin B (Mitta et al. 2000a) and myticin B (GenBank EU088427) genes with three introns each. Common feature is that none of the introns interrupted the open reading frame corresponding to mature peptides. The peculiar location of mytimycin intron 1, just upstream the first ATG was identical to the one of mytilin B gene (Mitta et al. 2000a). The two *mytimycin* genes, defined by the presence of a long or a short intron 2, are simultaneously present in all the *M. galloprovincialis* analyzed so far. In contrast, defensin MGD2 (Mitta et al. 2000a) and myticin C genes (Costa et al. 2008) are probably present as a single copy in the *M. galloprovincialis* genome.

In conclusion, we revealed the existence of the antifungal peptide MytM in *M. galloprovincialis*, previously only partially evidenced in *M. edulis*. We established the complete coding sequence sequences and studied the diversity of mytimycin precursors within and between three geographical mussel populations. In addition, we established the primary structures of *mytimycin* genes. The molecular informations reported here completed the panel of anti-infectious molecules used by mussels. Next step must address the antifungal activity, previously tested only against the red bread mold *Neurospora crassa* (Charlet et al. 1996), and gene regulation to provide new insights into the control of fungal diseases in mussels which might be of more general interest to prevent dramatic infestations by fungi.

Acknowledgments This collaborative work was partly funded by the EU program IMAQUANIM (FOOD-CT-2005-007103). MT was fully supported by IMAQUANIM. Thanks to Umberto Rosani for the preliminary comparison performed on new FLX 454 sequence MytM data.

References

Boman HG, Nilsson-Faye I, Paul K, Rasmuson T Jr (1974) Insect immunity. I. Characteristics of an inducible cell-free antibacterial reaction in hemolymph of *Samia cynthia* pupae. Infect Immun 10:136–145

Boon E, Faure MF, Bierne N (2009) The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. J Mol Evol 68:461–474

Charlet M, Chernysh S, Philippe H, Hetru C, Hoffmann JA, Bulet P (1996) Innate immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. J Biol Chem 271:21808–21813

Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, Figueras A (2008) Myticin C variability in mussel: ancient defence mechanism or self/nonself-discrimination? Dev Comp Immunol 32:213–226

Costa MM, Dios S, Alonso-Gutierrez J, Romero A, Novoa B, Figueras A (2009) Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). Dev Comp Immunol 33:162–170

Da Silva P, Jouvensal L, Lamberty M, Bulet P, Caille A, Vovelle F (2003) Solution structure of termicin, an antimicrobial peptide from the termite *Pseudacanthotermes spiniger*. Protein Sci 12:438–446

Dassanayake RS, Silva Gunawardene YI, Tobe SS (2007) Evolutionary selective trends of insect/mosquito antimicrobial defensin peptides containing cysteine-stabilized alpha/beta motifs. Peptides 28:62–75

Destoumieux D, Bulet P, Loew D, Van Dorsselaer A, Rodriguez J, Bachere E (1997) Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). J Biol Chem 272:28398–28406

Ekengren S, Hultmark D (1999) Drosophila cecropin as an antifungal agent. Insect Biochem Mol Biol 29:965–972

Fehlbaum P, Bulet P, Chernysh S, Briand JP, Roussel JP, Letellier L, Hetru C, Hoffmann JA (1996) Structure–activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides. Proc Natl Acad Sci U S A 93:1221–1225

Fehlbaum P, Bulet P, Michaut L, Lagueux M, Broekaert WF, Hetru C, Hoffmann JA (1994) Insect immunity. Septic injury of *Drosophila* induces the synthesis of a potent antifungal peptide with sequence homology to plant antifungal peptides. J Biol Chem 269:33159–33163

Gestal C, Costa M, Figueras A, Novoa B (2007) Analysis of differentially expressed genes in response to bacterial stimulation in hemocytes of the carpet-shell clam *Ruditapes decussatus*: identification of new antimicrobial peptides. Gene 406:134–143

Gueguen Y, Garnier J, Robert L, Lefranc MP, Mougenot I, de Lorgeril J, Janech M, Gross PS, Warr GW, Cuthbertson B, Barracco MA, Bulet P, Aumelas A, Yang Y, Bo D, Xiang J, Tassanakajon A, Piquemal D, Bachere E (2006a) PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. Dev Comp Immunol 30:283–288

Gueguen Y, Herpin A, Aumelas A, Garnier J, Fievet J, Escoubas JM, Bulet P, Gonzalez M, Lelong C, Favrel P, Bachere E (2006b) Characterization of a defensin from the oyster *Crassostrea gigas*. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. J Biol Chem 281:313–323

Hubert F, Noel T, Roch P (1996) A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis*). Eur J Biochem 240:302–306

Iijima R, Kurata S, Natori S (1993) Purification, characterization, and cDNA cloning of an antifungal protein from the hemolymph of *Sarcophaga peregrina* (flesh fly) larvae. J Biol Chem 268:12055–12061

Kubota Y, Watanabe Y, Otsuka H, Tamiya T, Tsuchiya T, Matsumoto JJ (1985) Purification and characterization of an antibacterial factor from snail mucus. Comp Biochem Physiol C 82:345–348

Lamberty M, Ades S, Uttenweiler-Joseph S, Brookhart G, Bushey D, Hoffmann JA, Bulet P (1999) Insect immunity. Isolation from the lepidopteran *Heliothis virescens* of a novel insect defensin with potent antifungal activity. J Biol Chem 274:9320–9326

Lamberty M, Zachary D, Lanot R, Bordereau C, Robert A, Hoffmann JA, Bulet P (2001) Insect immunity. Constitutive expression of a cysteine-rich antifungal and a linear antibacterial peptide in a termite insect. J Biol Chem 276:4085–4092

Levashina EA, Ohresser S, Bulet P, Reichhart JM, Hetru C, Hoffmann JA (1995) Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with antibacterial and antifungal properties. Eur J Biochem 233:694–700

Lewit-Bentley A, Rety S (2000) EF hand calcium-binding proteins. Curr Opin Struct Biol 10:637–643

Mitta G, Hubert F, Dyrynda EA, Boudry P, Roch P (2000a) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. Dev Comp Immunol 24:381–393

Mitta G, Hubert F, Noel T, Roch P (1999a) Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. Eur J Biochem 265:71–78

Mitta G, Vandenbulcke F, Hubert F, Roch P (1999b) Mussel defensins are synthesised and processed in granulocytes then released into the plasma after bacterial challenge. J Cell Sci 112:4233–4242

Mitta G, Vandenbulcke F, Hubert F, Salzet M, Roch P (2000b) Involvement of mytilins in mussel antimicrobial defense. J Biol Chem 275:12954–12962

Pallavicini A, Costa MM, Gestal C, Dreos R, Figueras A, Venier P, Novoa B (2008) High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host–pathogen interactions. Dev Comp Immunol 32:213–226

Parisi MG, Li H, Toubiana M, Parrinello N, Cammarata M, Roch P (2009) Polymorphism of mytilin B mRNA is not translated into mature peptide. Mol Immunol 46:384–392

Prado-Alvarez M, Gestal C, Novoa B, Figueras A (2009) Differentially expressed genes of the carpet shell clam *Ruditapes decussatus* against *Perkinsus olseni*. Fish Shellfish Immunol 26:72–83

Selitrennikoff CP (2001) Antifungal proteins. Appl Environ Microbiol 67:2883–2894

Somboonwiwat K, Marcos M, Tassanakajon A, Klinbunga S, Aumelas A, Romestand B, Gueguen Y, Boze H, Moulin G, Bachere E (2005) Recombinant expression and anti-microbial activity of anti-lipopolysaccharide factor (ALF) from the black tiger shrimp *Penaeus monodon*. Dev Comp Immunol 29:841–851

Tomie T, Ishibashi J, Furukawa S, Kobayashi S, Sawahata R, Asaoka A, Tagawa M, Yamakawa M (2003) Scarabaecin, a novel cysteine-containing antifungal peptide from the rhinoceros beetle, *Oryctes rhinoceros*. Biochem Biophys Res Commun 307:261–266

Venier P, De Pitta C, Bernante F, Varotto L, De Nardi B, Bovo G, Roch P, Novoa B, Figueras A, Pallavicini A, Lanfranchi G (2009) MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. BMC Genomics 10:72

Wang G, Li X, Wang Z (2009) ADP2: the updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Research 37:933–937

Wang N, Whang I, Lee J (2008) A novel C-type lectin from abalone, *Haliotis discus* discus, agglutinates *Vibrio alginolyticus*. Dev Comp Immunol 32:1034–1040

Yamazaki M, Ohye H, Kisugi J, Kamiya H (1990) Bacteriostatic and cytolytic activity of purple fluid from the sea hare. Dev Comp Immunol 14:379–383

Zhang H, Song L, Li C, Zhao J, Wang H, Gao Q, Xu W (2007) Molecular cloning and characterization of a thioester-containing protein from Zhikong scallop Chlamys farreri. Mol Immunol 44:3492–3500

Zhang J, Wu X, Zhang SQ (2008) Antifungal mechanism of antibacterial peptide, ABP-CM4, from *Bombyx mori* against Aspergillus niger. Biotechnol Lett 30:2157–2163