Available online at www.sciencedirect.com
SCIENCE DIRECT®
Developmental and Comparative Immunology 30 (2006) 283–288
Developmental & Comparative Immunology
www.elsevier.com/locate/devcompimm

PenBase, the shrimp antimicrobial peptide penaeidin database: Sequence-based classification and recommended nomenclature

Yannick Gueguen[a,*], Julien Garnier[a], Lorenne Robert[a], Marie-Paule Lefranc[b], Isabelle Mougenot[c], Julien de Lorgeril[a], Michael Janech[a], Paul S. Gross[d], Gregory W. Warr[d], Brandon Cuthbertson[d], Margherita A. Barracco[e], Philippe Bulet[f], André Aumelas[g], Yinshan Yang[g], Dong Bo[h], Jianhai Xiang[h], Anchalee Tassanakajon[i], David Piquemal[j], Evelyne Bachère[a]

[a]*Ifremer-CNRS-University of Montpellier 2, UMR 5171 Génome Population Interaction Adaptation, 2 Place E. Bataillon, CC80, F-34095 Montpellier cedex 5, France*
[b]*Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier II, Institut Universitaire de France, UPR CNRS 1142, IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France*
[c]*Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier, LIRMM, UMR 5506 CNRS-UM2, 161 rue Ada 34392 Montpellier cedex 5, France*
[d]*Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA*
[e]*Departamento de Biologia Celular, Embriologia e Genética, Universidade Federal de Santa Catarina, Florianópolis, Brazil*
[f]*Atheris Laboratories, Case postale 314, CH-1233 Bernex-Geneva, Switzerland*
[g]*Centre de Biochimie Structurale, Université Montpellier I, 15 avenue Charles Flahault, F-34060 Montpellier cedex 2, France*
[h]*Institute of Oceanology, Chinese Academy of Science, 7 Nanhai Road, Qingdao 266071, People's Republic of China*
[i]*Faculty of Science, Shrimp Molecular Biology and Genomics Laboratory, Department of Biochemistry, Chulalongkorn University, Bangkok 10330, Thailand*
[j]*Skuld-Tech Company, Université Montpellier II, Place E. Bataillon, CC091, F-34095 Montpellier cedex 5, France*

Received 8 March 2005; revised 1 April 2005; accepted 7 April 2005
Available online 19 May 2005

Abstract

Antimicrobial peptides play a major role in innate immunity. The penaeidins, initially characterized from the shrimp *Litopenaeus vannamei*, are a family of antimicrobial peptides that appear to be expressed in all penaeid shrimps. As of recent, a large number of penaeid nucleotide sequences have been identified from a variety of penaeid shrimp species and these sequences currently reside in several databases under unique identifiers with no nomenclatural continuity. To facilitate research in this field and avoid potential confusion due to a diverse number of nomenclatural designations, we have made a systematic effort to collect, analyse, and classify all the penaeidin sequences available in every database. We have identified a common penaeidin signature and subsequently established a classification based on amino acid sequences. In order to clarify the naming

* Corresponding author. Tel.: +33 4 67 14 47 07; fax: +33 4 67 14 46 22.
E-mail address: ygueguen@ifremer.fr (Y. Gueguen).

0145-305X/$ - see front matter
doi:10.1016/j.dci.2005.04.003

process, we have introduced a 'penaeidin nomenclature' that can be applied to all extant and future penaeidins. A specialized database, PenBase, which is freely available at http://www.penbase.immunaqua.com, has been developed for the penaeidin family of antimicrobial peptides, to provide comprehensive information about their properties, diversity and nomenclature.

Keywords: Penaeidin; Penaeid shrimp; Antimicrobial peptide; Crustacea; Sequence database; Nomenclature

1. Introduction

Antimicrobial peptides (AMP) play a major role in innate immunity, conserved in evolution, and present in all phyla of the living kingdom. They are mostly cationic and amphipathic molecules although they present a great diversity in terms of structural features as well as biological properties and functions. More than 850 antimicrobial peptides have now been discovered in plants, vertebrates, and invertebrates (Antimicrobial sequences database, AMSDb http://www.bbcm.units.it/~tossi/amsdb.html [1]) and very recently, two databases dedicated to antimicrobial peptides, the Antimicrobial peptide database (APD) [2] and ANTIMIC [3], were created. To date, in penaeid shrimps, two kinds of antimicrobial peptides have been fully characterized, namely the penaeidins from hemocytes [4] and anionic hemocyanin-derived peptides isolated from shrimp plasma [5]. In addition, recent studies utilizing a genomic approach led to the characterization of other antimicrobial effectors in shrimp, i.e. anti-LPS factor (ALF) and crustin [6–8].

Penaeidins were first characterized from *Litopenaeus vannamei* using a biochemical approach and molecular cloning techniques. Three peptides (initially named penaeidin 1, 2, and 3) were isolated in their active and mature forms (5.48–6.62 kDa) from the hemocytes of shrimp [4]. However, subsequent phylogenetic analysis indicated that penaeidin 1 and -2 could be classified in the same class [9]. Penaeidins are composed of an N-terminal proline-rich domain, followed by a C-terminal domain containing six cysteine residues organized in two doublets. This overall structure is quite unique among the AMP families [4]. The antimicrobial activity spectrum of penaeidin 2 and 3 from *L. vannamei* has been established through the production and analysis of recombinant peptides [10]. More recently, a new subgroup of penaeidins, named penaeidin 4 by their authors, has been identified in *L. vannamei* using a genomic approach [9], and a synthetic chemical peptide was produced to investigate and characterize the functional properties and spectrum of activity [11]. Penaeidins possess antibacterial activity predominantly directed against Gram-positive bacteria and antifungal activity against filamentous fungi. Very recently, the solution structure of the recombinant penaeidin 3 from *L. vannamei* and of the synthetic penaeidin 4 from *L. setiferus* have been determined, revealing the overall organization of the two domains and the arrangement of the disulfide bonds [12,13].

Recent studies have revealed the presence of penaeidin mRNAs in different penaeid shrimp species [6,8,14,15]. Moreover, in both *L. vannamei* and *L. setiferus*, the penaeidin subgroups (penaeidin 2, 3, and 4) were shown to be expressed in a single individual [6]. Expressed sequence tag (EST) analysis from hemocyte cDNA libraries has shown that penaeidin transcripts are very abundant. In fact, penaeidins appear to represent 10.7 and 20% of all the sequences isolated from hemocytes of *L. vannamei* and *L. setiferus*, respectively [6].

Most groups working on shrimp from different parts of the world have now isolated numerous penaeidin sequences mostly by genomic approaches, and this family appears to be ubiquitous among penaeid shrimp species. At the moment more than 200 sequences can be found in the EMB/GenBank/DDBJ databases, some of which have yet to be published. The current nomenclature for penaeidins is very confusing, with several different names given to the same gene product and distinct names (i.e. penaeidin number) given to variants of the same molecule. Therefore, we suggest that the 'penaeidin field' adopts a common nomenclature based on amino acid sequence similarity. Herein, we introduce a new database, PenBase, to provide comprehensive information about penaeidin properties, diversity and nomenclature. PenBase has been implemented according to the IMGT Scientific chart rules that are based on the IMGT-ONTOLOGY concepts [16].

The description labels, the subgroup and penaeidin nomenclature, the amino acid unique numbering, and the penaeidin origin are based on the DESCRIPTION, CLASSIFICATION, NUMEROTATION and OBTENTION concepts, respectively [17–19]. Owing to that standardized approach, PenBase provides for the first time a coherent management of data related to penaeidins.

2. Penaeidin signature determination

We have made a systematic effort to collect and analyse all publicly accessible penaeidin sequence data. Penaeidin sequences were analysed from eight shrimp species: *L. vannamei* (Pacific White Shrimp or White leg Shrimp), *L. setiferus* (Atlantic White Shrimp or Northern White Shrimp), *L. stylirostris* (Blue Shrimp), *Farfantepenaeus paulensis* (Sao Paulo Shrimp), *L. schmitti* (Southern white shrimp), *Penaeus semisulcatus* (Green tiger Prawn), *Fenneropenaeus chinensis* (Fleshy Prawn) and *Penaeus monodon* (Giant or Black tiger prawn). Three different categories of data were collected: cDNA sequences, peptidic sequences with no corresponding nucleotide sequence and nucleotide sequences originating from EST projects. The determination of the family signature is based on common amino acid sequences shared by penaeidins, and that include two regions, the signal peptide (L-Region) and the mature peptide (Peptide-Region). In our analysis, 104 sequences coming from EST project were not used because they contained uncertain nucleotides sequences in the Peptide-Region. Multiple alignments using ClustalX [20] were constructed with 110 sequences and a penaeidin family signature was designated based on the coding amino acid sequences. Many sequences appear 100% identical at the amino acid level while only 34 penaeidin sequences differ at the level of at least one amino acid. The Pepide-Region from the penaeidin family is characterized by the presence of two domains, a proline-rich N-terminal domain and a cysteine-rich C-terminal domain presenting two doublets of cysteine residues. NMR solution structure studies revealed that the N-terminal domain rich in proline residues is unconstrained. In contrast, the structure of the C-terminal domain stabilized by three conserved disulfide bonds in the 1–3, 2–5 and 4–6 patterns is well defined [12]. This domain consists in an amphipathic helix linked to upstream and downstream coils by the 1–3 and 2–5 disulfide bonds and these two coils are in turn linked together by the 4–6 disulfide bond. This disulphide arrangement would be characteristic of the penaeidin family according to recent studies on solution structure of penaeidin 4 for which a similar pattern was determined [13]. From the alignment and structural studies, the signature of the penaeidin family has been defined as shown in Fig. 1 by 10 conserved amino acids in the signal peptide sequence; a threonine and two proline residues conserved in the N-terminal domain; and the conserved cysteine array of the C-terminal structured domain.

3. Penaeidin nomenclature and subgroup determination

A detailed analysis of the sequence alignments led to the classification of all the existing penaeidin amino acid sequences in three distinct subgroups: Penaeidin 2 (PEN2), Penaeidin 3 (PEN3), Penaeidin 4 (PEN4). Based on these analyses, we introduced the following nomenclature that will be now used for the penaeidins. The name of a penaeidin is composed of (i) six letters in italics (three for the genus and three for the species)

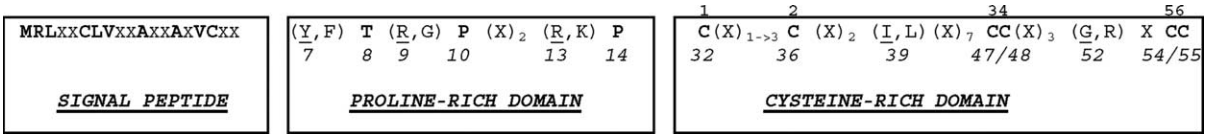

Fig. 1. Penaeidin signature. In bold are represented the amino acids conserved in all the penaeidins. X is for any residue. Numbers 1–6 above the sequence are related to the cysteine rank in the mature sequence and used in the text to describe the disulfide bond pattern. Numbers below the sequence correspond to the amino acid position in the mature peptide.

followed by a space, (ii) the penaeidin subgroup followed by a hyphen and (iii) a number for the identification inside the subgroup (Example: *Litvan* PEN3-1: sequence from *L. vannamei*, PEN3 subgroup, first sequence identified. The determination of the penaeidin subgroups (PEN2, PEN3 and PEN4) is based on the amino acid sequences of the mature peptide (Peptide-Region). Following amino acid alignment analyses, conserved key residues appear to be characteristic for each of the three subgroups (Fig. 2). These key residues are highlighted in grey in Fig. 2. For example, all penaeidins of the 'subgroup PEN3' possess the following signature: Gln^{1}, Gly^{5}, Arg^{13}, Val^{18} or Gly^{18}, Ser^{35}, Arg^{37} or His^{37} or Pro^{37}, Gln^{43}, Ser^{46}or Ala^{46}.

One penaeidin peptide name (example *Litvan* PEN3-1) can correspond to several related sequences which may differ at the nucleotide level without amino acid changes in the PEPIDE-REGION. The different cDNA sequences that correspond to nucleotide variations are identified by an asterisk followed by a number, *01 being assigned to the first cDNA

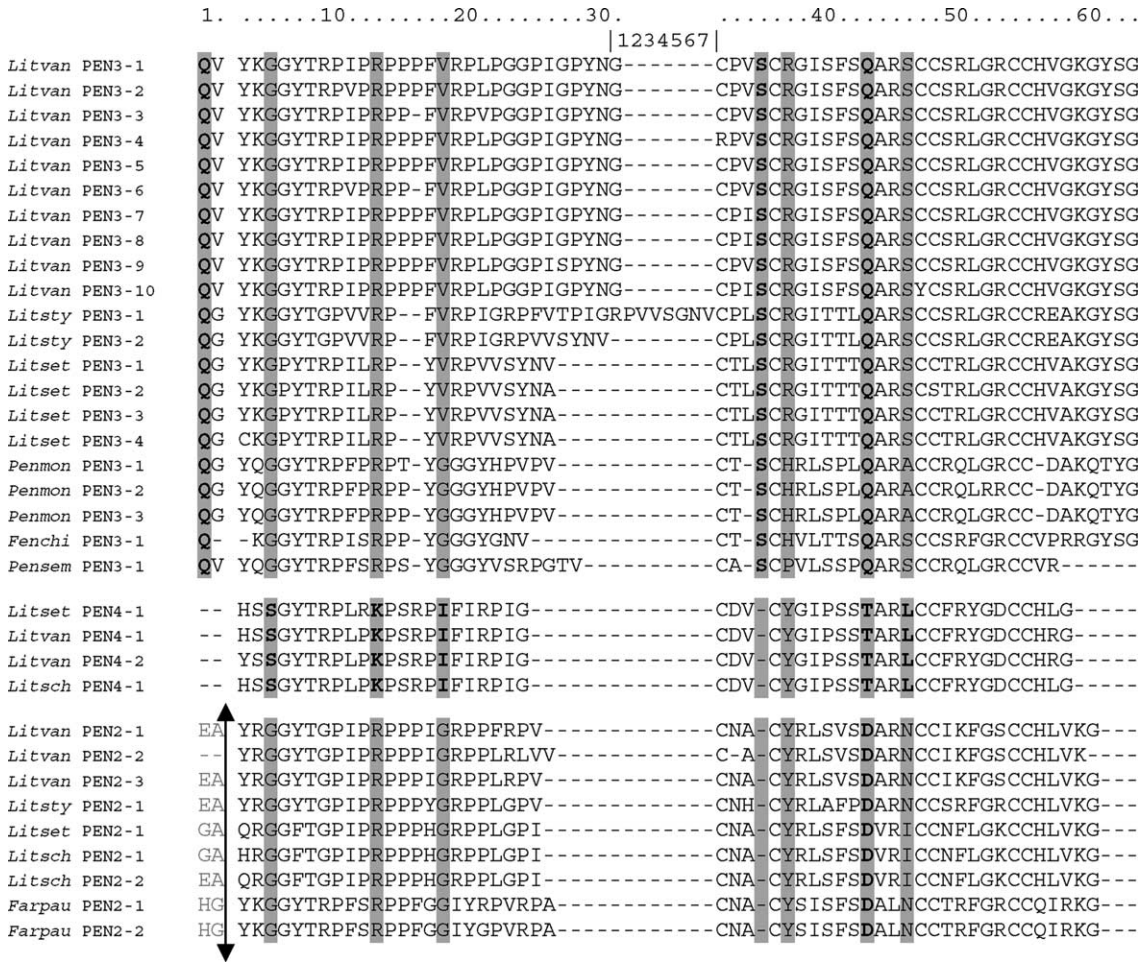

Fig. 2. Determination of the penaeidin subgroup signatures. In grey are represented the key amino acids that are used for the subgroup signatures. In bold, amino acids, which are specific of only one subgroup. The double arrow indicates the peptide signal cleavage site. *Litopenaeus vannamei*: *Litvan*, *Litopenaeus setiferus*: *Litset*, *Litopenaeus stylirostris*: *Litsty*, *Litopenaeus schmitti*: *Litsch*, *Penaeus semisulcatus*: *Pensem*, *Fenneropenaeus chinensis*: *Fenchi*, *Penaeus monodon*: *Penmon*, *Farfantepenaeus paulensis*: *Farpau*.

sequence (example *Litvan* PEN3-1*01). In addition, one penaeidin peptide name and sequence (example *Litvan* PEN3-1), can also correspond to one or several EST sequences. If an EST sequence is identical to an already known cDNA Pepide-Region nucleotide sequence, the EST is classified under the same name (example *Litvan* PEN3-1*01) in the EST sequence section. If not, the EST is classified as an EST sequence (example *Litvan* PEN3-1*EST1) until the sequence is confirmed by sequencing and classified as a cDNA.

4. Organization of the database

PenBase was built on the Debian Sarge GNU/Linux operating system (http://www.debian.org/) using the open-source softwares Apache 2 web server (http://www.apache.org/), PHP script language (http://www.php.net/) and postgreSQL relational database system (http://www.postgresql.org/). The PenBase home page contains an introduction to the penaeidin family and lists the accessible sections in the database:

1. Properties
2. Signatures and subgroups
3. Nomenclature
4. Penaeidin database
5. Primer library
6. Bibliography
7. Forum
8. Penaeidin links

The Database section lists all known penaedins sorted by shrimp species or subgroups. Individual penaeidin records contain detailed information about each penaeidin including peptide name, nucleotide and peptide sequence data, biochemical properties (including p*I*, molecular mass, signal peptide length, proline-rich and cysteine-rich domain lengths), contact information, references (including authors, article titles, journal, page, volume and year) and links to PubMed and EMBL/GenBank. In this section, Blast and alignment tools are available. The 'Signature and subgroup' section describes the criteria for penaeidin identification and classification. The forum section is a place where anyone can exchange information or ask questions regarding penaeidins. References throughout the database are linked to the Bibliography section. The bibliography includes a list of all the papers related to penaeidins.

5. Availability and future developments

PenBase can be accessed directly at http://www.penbase.immunaqua.com. Users of PenBase are requested to cite this article when referencing the database. Researchers in this field are invited to make suggestions and comments by using the forum section. PenBase contains more than 200 entries of penaeidins and is expected to quickly grow up with the rapid development of genomic and proteomic projects. As more information about penaedins becomes available, the database will be expanded and improved accordingly.

Acknowledgements

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 no ICA4-CT-2001-10023 (IMMUNAQUA, http://www.immunaqua.com). It is also supported by IFREMER, CNRS and the University of Montpellier II. M. Janech was supported by the National Science Foundation, International Research Fellowship Program. M.A. Barracco was supported by a post-doctoral scholarship given by CAPES (Brazil) and J. de Lorgeril was supported by a doctoral funding from the New Caledonian Territories and the IFREMER. P. Bulet was supported by the Swiss Federal Office for Education and Science contract no 02.0360. This work was partially supported by the National Science Foundation (NSF grant #MCB0110576 to PS Gross & MCB0315393 to GW Warr), the National Marine Fisheries Service (NA03NMF4720362 to PS Gross & GW Warr), the South Carolina Sea Grant Consortium (R/MT-6 to PS Gross), United States Department of Agriculture (USDA NRICGP-CSREES/AREA Grant# 2002-35201-11620 to PS Gross) and the South Carolina Department of Natural Resources. Any opinions, findings, and conclusions or recommendations

expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting bodies mentioned herein.

References

[1] Tossi A, Sandri L. Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. Curr Pharm Des 2002;8: 743–61.

[2] Wang Z, Wang G. APD: the antimicrobial peptide database. Nucleic Acids Res 2004;32:590–2.

[3] Brahmachary M, Krishnan S, Koh J, Khan A, Seah S, Tan T, et al. ANTIMIC: a database of antimicrobial sequences. Nucleic Acids Res 2004;32:586–9.

[4] Destoumieux D, Bulet P, Loew D, Van Dorsselaer A, Rodriguez J, Bachère E. Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). J Biol Chem 1997;272:28398–406.

[5] Destoumieux-Garzon D, Saulnier D, Garnier J, Jouffrey C, Bulet P, Bachère E. Crustacean immunity: antifungal peptides are generated from the C-terminus of shrimp hemocyanin in response to microbial challenge. J Biol Chem 2001;276: 47070–7.

[6] Gross PS, Bartlett TC, Browdy CL, Chapman RW, Warr GW. Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. Dev Comp Immunol 2001;25:565–77.

[7] Bartlett TC, Cuthbertson BJ, Shepard EF, Chapman RW, Gross PS, Warr GW. Crustins, homologues of an 11.5 kDa antimicrobial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *L. setiferus*. Mar Biotechnol 2002; 4(3):278–93.

[8] Supungul P, Klinbunga S, Pichyangkura R, Jitrapakdee S, Hirono I, Aoki T, et al. Identification of immune-related genes in hemocytes of Black Tiger Shrimp (*Penaeus monodon*). Mar Biotechnol 2002;4:487–94.

[9] Cuthbertson BJ, Shepard EF, Chapman RW, Gross PS. Diversity of the penaeidin antimicrobial peptides in two shrimp species. Immunogenetics 2002;54:442–5.

[10] Destoumieux D, Bulet P, Strub JM, Van Dorsselaer A, Bachère E. Recombinant expression and range of activity of penaeidins, antimicrobial peptides from penaeid shrimp. Eur J Biochem 1999;266:335–46.

[11] Cuthbertson BJ, Bullesbach EE, Fievet J, Bachère E, Gross PS. A new class (penaeidin class 4) of antimicrobial peptides from the Atlantic white shrimp (*Litopenaeus setiferus*) exhibits target specificity and an independent proline-rich-domain function. Biochem J 2004;381:79–86.

[12] Yang Y, Poncet J, Garnier J, Zatylny C, Bachère E, Aumelas A. Solution structure of the recombinant penaeidin-3, a shrimp antimicrobial peptide. J Biol Chem 2003;278: 36859–67.

[13] Cuthbertson BJ, Yang Y, Bachère E, Gross PS, Aumelas A. Synthesis and solution structure of synthetic penaeidin 4 with structural and functional comparisons to penaeidin 3. J Biol Chem 2005;280:16009–18.

[14] Rojtinnakorn J, Hirono I, Itami T, Takahashi Y, Aoki T. Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. Fish Shellfish Immunol 2002;13:69–83.

[15] Kang CJ, Wang JX, Zhao XF, Yang XM, Shao HL, Xiang JH. Molecular cloning and expression analysis of the Ch-penaeidin, an antimicrobial peptide from Chinese shrimp, *Fenneropenaeus chinensis*. Fish Shellfish Immunol 2004;16: 513–25.

[16] Giudicelli V, Lefranc MP. Ontology for immunogenetics: IMGT-ONTOLOGY. Bioinformatics 1999;15:1047–54.

[17] Lefranc MP. IMGT, the international ImMunoGeneTics database: a high-quality information system for comparative immunogenetics and immunology. Dev Comp Immunol 2002; 26:697–705.

[18] Warr GW, Clem LW, Söderhall K. Editorial—The international ImMunoGeneTics database IMGT. Dev Comp Immunol 2003;27:1–2.

[19] Lefranc MP, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, et al. IMGT, the international ImMunoGeneTics information system®. Nucleic Acids Res 2005;33:593–7.

[20] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–82.