

Insect Antimicrobial Peptides: Structures, Properties and Gene Regulation

Philippe Bulet* and Reto Stöcklin

Atheris Laboratories, Case Postale 314, CH-1233 Bernex, Geneva, Switzerland.

Abstract: Antimicrobial peptides (AMPs) are part of the armament that insects have developed to fight off pathogens. Insect AMPs are typically cationic and often made of less than 100 amino acid residues. Although their structures are diverse, most of the AMPs can be assigned to a limited number of families. The most common structures are represented by peptides assuming a α -helical conformation in organic solutions or disulfide-stabilized β -sheets with or without α -helical domains present. The diverse activity spectrum of these peptides may indicate different modes of action. Genetic analysis in the *Drosophila* model evidenced that multiple signal transduction pathways are activating the genes coding AMPs.

Keywords: antimicrobial peptides, insect immunity, microbial infection, cecropin, defensin, proline-rich, innate defense, cationic peptide.

INTRODUCTION

Nosocomial infections are caused by multiresistant bacterial strains that are often not sensitive to classical antibiotic treatments. The incidence of such infections is exponentially growing in hospitals and clinical environments, reaching several thousands of deaths in the world annually. Approximately 10% of American patients in health care facilities (corresponding to 2 million people every year) acquire clinically significant nosocomial infections. Such infections are predominantly caused by opportunistic microorganisms such as *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas*, *Enterococcus*, *Legionella* and *Aspergillus* spp. Two of the major contributors to these infections are widespread over-prescription and misuse of antibiotic drugs, practices that have promoted the dissemination of a series of particularly harmful bacterial strains that resist to conventional antimicrobial treatments. Another potential main reason for the developed bacterial resistance in humans is the massive use of preventive antibiotics in animal food. Because of a fast growth rate, the frequencies of genetic mutations and selections, and the ability of bacteria to rapidly exchange genes, bacterial resistance to antibiotics seems to take place swiftly in the evolution of bacterial development.

Obviously, if a bacterial pathogen is able to develop or acquire resistance to an antibiotic, this given drug becomes useless in the future treatment of infections caused by the pathogen and alternative treatments are required. Indeed, several new bacterial diseases have been discovered in the past decades. In this context, there is an urgent need for a new generation of antibiotics to complement the panel of drugs that are available to the clinicians and to provide new tools for multitherapy treatment. For decades, one major area of interest for the discovery and study of new antibiotics was

the investigation of AMPs derived from insect immune defense reactions.

With roughly one million characterized species (a total of 30 million of estimated species, 32 orders and more than 600 families worldwide), insects account for the majority of known animals. Important features of their tremendous colonization success and diversity are (i) their short life spans compared to most vertebrates, (ii) their capacity to colonize new niches and to feed on nearly all species of plants and animals and (iii) their ability to mount a harmful immune response. The study of insect immunity has demonstrated alternative antimicrobial strategies, while insect immune defense relies solely on innate immunity (no memory), in vertebrates innate immunity coexists with adaptive immunity (clonal) [1]. One component of the defense weapons developed by insects to rapidly eliminate invading pathogens is the fast and massive production of potent AMPs [2-5]. The large number of AMPs and the various structures of these molecules are not restricted to insects as these features are also present in a diverse range of organisms including microorganisms [6,7], plants [8] and other animals including mammals [9-11]. Approximately 700 AMPs reported in the literature are already listed in the publicly available databases including Swissprot and TrEMBL (<http://www.expasy.org/sprot/sprot-top.html>), AMSDd (<http://www.bbcm.univ.trieste.it/~tossi/pag1.htm>), APD (<http://aps.unmc.edu/AP/main.html>) and ANTIMIC (<http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC/>). Nevertheless, these files are not exhaustive by definition, since they are based almost exclusively on what has been voluntarily submitted by their authors to public databases. Fortunately, several journals are now requesting, before publishing the results, to have a new sequence deposited in any of the specialized database. But this was not the case a few years ago, and even nowadays does by far not apply to all publications. This observation has convinced us to develop our own database going also back to the original manuscripts. It already includes deeply annotated records of more than 1.200 AMPs of animal origin that have been

*Address correspondence to this author at the Atheris Laboratories, Case Postale 314, CH-1233 Bernex, Geneva, Switzerland; E-mail: philippe.bulet@atheris.ch

described in the literature, without taking into account AMPs from bacteria and plants. We expect to reach a total of 1.500 different AMP structures. This work will be described elsewhere (manuscript in preparation).

In insects with complete metamorphosis (holometabolous), AMPs are rapidly and transiently synthesised by the fat body (tissue corresponding to mammalian liver), and by various epithelia [5]. When produced by the fat body, the AMPs are secreted into the hemolymph (blood), from where they can easily diffuse to act throughout the whole animal. Most of the AMPs are produced quite massively, many of them reaching high micromolar concentrations (i.e. mg/L). In contrast, in insects with incomplete metamorphosis (heterometabolous), AMPs are produced by hemocytes (blood cells) in the healthy animal and secreted into the hemolymph upon infection [12]. Insect AMPs are cleaved *in vivo* from a large precursor that contains a signal domain and sometimes a pro-domain upstream or downstream the mature peptide.

Although usually cationic, the primary structures of insect AMPs vary markedly. Members of the most frequent AMP families adopt an α -helical conformation in membrane-mimetic environments [13,14] or their sequence contains pairs of cysteine residues [15]. Another family that is commonly found in insects is represented by peptides with high content in one or two particular amino acids, most frequently proline and/or glycine residues [2-4]. Although this review will focus on the structures and the biological properties of AMPs of insect origin, a short summary of the different transduction signalling pathways controlling AMP gene expression and allowing *Drosophila* to discriminate between a fungal and a bacterial infection will be included. We will also outline the potential development strategies of such peptides for human and veterinary therapeutic use.

LINEAR AND AMPHIPATHIC α -HELICAL AMPs

The most abundant family of linear AMPs of insect origin are the cecropins including sarcotoxins, hyphancin, enbocin and spodopsin, as well as cecropin-like molecules totalling more than 60 described sequences. Insect α -helical AMPs are 29-42 residues in size, linear and devoid of cysteine residues. Cecropins were the first animal inducible AMPs to be isolated and fully characterized. The first insect cecropin was isolated from the blood of experimentally infected diapausing pupae of the moth *Hyalophora cecropia* (Lepidoptera) [16]. Since this first report, expression of cecropin-like peptides have been documented in several other insect species, which all belong to phylogenetically higher insect orders of Diptera and Lepidoptera [4,5]. A comparison of the sequences revealed that dipteran cecropins form a particularly homologous group with more than 70% identity in their amino acid composition [4] and can reach as high as 100% identity as observed between cecropin IA (also named sarcotoxin IA) from the flesh fly *Sarcophaga peregrina* [17] and cecropin A from the fruit fly *Drosophila melanogaster* [18]. In contrast to dipteran cecropins, cecropins from Lepidoptera present more variation in their primary structure [4]. A characteristic feature of most insect cecropins is the presence of a tryptophan residue at position 1 or 2, and an amidated C-terminus. However, none of these

two canonical features have been found in the mosquito cecropins isolated from an *Aedes albopictus* cell line [19,20] and subsequently from the blood of experimentally infected adults of *Aedes aegypti* [21]. Contrasting to the mosquito cecropins, the *Anopheles gambiae* cecropin, which is also lacking the tryptophan residue, is C-terminally amidated [22]. This absence of tryptophan residue within the N-terminal domain was also reported for a lepidopteran cecropin, namely cecropin D from the silkworm *Bombyx mori* [23]. Interestingly, it has been noted that the *Anopheles* cecropin without the tryptophan is more efficient against yeast and Gram-positive bacteria than *Drosophila* cecropin A, which has a tryptophan residue in position 2 [22]. The combination of increased number of positive charges and no tryptophans may be the reason for the differences in the observed antimicrobial activity. Comparison of amidated *versus* non-amidated cecropins evidenced a higher efficacy for the amidated version [24,25]. A group of cecropin-related peptides, ponericsins (sub-classified in three groups G, W and L), have been isolated from venom glands of the predatory ant *Pachycondylas goeldii* [26]. Sequence similarities exist between cecropins and ponericsins G (60%), and between ponericsins W and melittin (70%), an α -helical toxin isolated from the venom of the honeybee *Apis mellifera* (Hymenoptera). Melittin, the signature peptide for haemolytic activity to mammalian cells, also exhibits some antimicrobial properties [27]. Additional α -helical AMPs have been isolated including the 29 residue ceratotoxins from the medfly *Ceratitis capitata*, stomoxyn, the longest cecropin-like peptide from the stable fly *Stomoxys calcitrans* and spinigerin from the fungus-growing termite *Pseudacanthotermes spiniger* (Isoptera) [28]. As opposed to most AMPs that are secreted into the blood of the insect, the production of ceratotoxins, stomoxyn and spinigerin is not induced by microbial infection. Ceratotoxins, stomoxyn and spinigerin are constitutively present in the secretion of female accessory reproductive glands, in the anterior midgut and in the hemocytes, respectively. Details on the primary structure of the peptides discussed in this section and the insect origin of the molecules are listed in Figure 1.

The α -helical linear AMPs of insect origin are mostly active against bacteria (as opposed to fungi), with a higher efficacy on Gram-negative than on Gram-positive strains. They have no hemolytic activity when tested in the concentration range used for antibacterial activity screening. Interestingly, recent reports established that cecropins and cecropin-like peptides might also affect the growth of filamentous fungi including human and plant pathogens (i.e. *Aspergillus* spp. and *Fusarium* spp.), and yeasts [5]. Stomoxyn, which has a broad activity spectrum, also exhibits trypanolytic activity [29,30].

The solution conformation of a cecropin was first demonstrated on *Hyalophora* cecropin A by circular dichroism (CD) spectroscopy [31]. CD studies revealed that cecropins are unordered in aqueous solution, while in a hydrophobic environment, they adopt a rather stable α -helical conformation. The structure of cecropin A has been further characterized by NMR spectroscopy in water/hexafluoropropanol mixtures, a secondary structure promoting solvent combination [32]. This peptide has a long N-terminal, basic, amphipathic α -helix (residues 5 to 21) and

Organism of origin	Name	Primary structure
<i>Hyalophora cecropia</i>	Cecropin A	KWKLFKKIEKVGQNIRDGIIKAGPAVAVVGQATQIAK*
<i>Sarcophaga peregrina</i>	Sarcotoxin IA	GWLKKIGKKIERVVGQHTRDATIQGLGIAQQAAANVAATAR*
<i>Drosophila melanogaster</i>	Cecropin A	GWLKKIGKKIERVVGQHTRDATIQGLGIAQQAAANVAATAR*
<i>Aedes aegypti</i>	Cecropin A	GGLKKLGKKLEGAGKRVFNAAEKALPVVAGAKALRK
<i>Aedes albopictus</i>	Cecropin A1	GGLKKLGKKLEGVGRVFKASEKALPVAVGIKALGK
<i>Anopheles gambiae</i>	Cecropin A	GRLKKLGKKIEGAGKRVFKAEEKALPVVAGVKAL*
<i>Bombyx mori</i>	Cecropin D	GNFFKDLEKMGQRVRDAVISAAPAVDTLAKAKALGQ*
<i>Pachycondylas goeldii</i>	Ponericin G2	GWKDWLKKGKEWLKAKGPGIVKAAALQAATQ
<i>Cerratitis capitata</i>	Ceratotoxin	SIGSAFKKALPVAKKIGKAALPIAKAALP
<i>Stomoxys calcitrans</i>	Stomoxyn	RGRFKHFNLVKKVKHTISETAHVAKDTAVIAGSGAAVVAAT*
<i>Pseudacanthothermes spiniger</i>	Spinigerin	HVDKKVADKVLKQLRIMRLRL

Figure 1. Primary structures of selected α -helical AMPs. The * marks C-terminal amidation. All sequences reported in the figure are from the Swiss-Prot/TrEMBL database.

a shorter and more hydrophobic C-terminal helix (residues 24 to 37), linked by a Gly-Pro hinge region (Figure 2A).

Structure-activity relationship (SAR) studies outline the importance of a series of different parameters that may influence the activity of the α -helical AMPs: charge, helicity, size, hydrophobic moment and hydrophobicity [14]. SAR studies involve either the modification of the natural peptides or *in silico* design of peptides predicted to adopt α -helical conformations, with the purpose of increasing the

efficacy against the pathogens without generating toxicity to the host. Although SAR studies on α -helical AMPs are dominated by data on cecropins, similar experiments were conducted on magainins, AMPs isolated from frog skin secretions [33]. Efforts were devoted to improve the therapeutic potential of cecropins by combining fragments of different AMPs forming α -helices (cecropins and magainins) or a fragment of *Hyalophora* cecropin A (often the N-terminal amphipathic α -helical domain) with melittin.

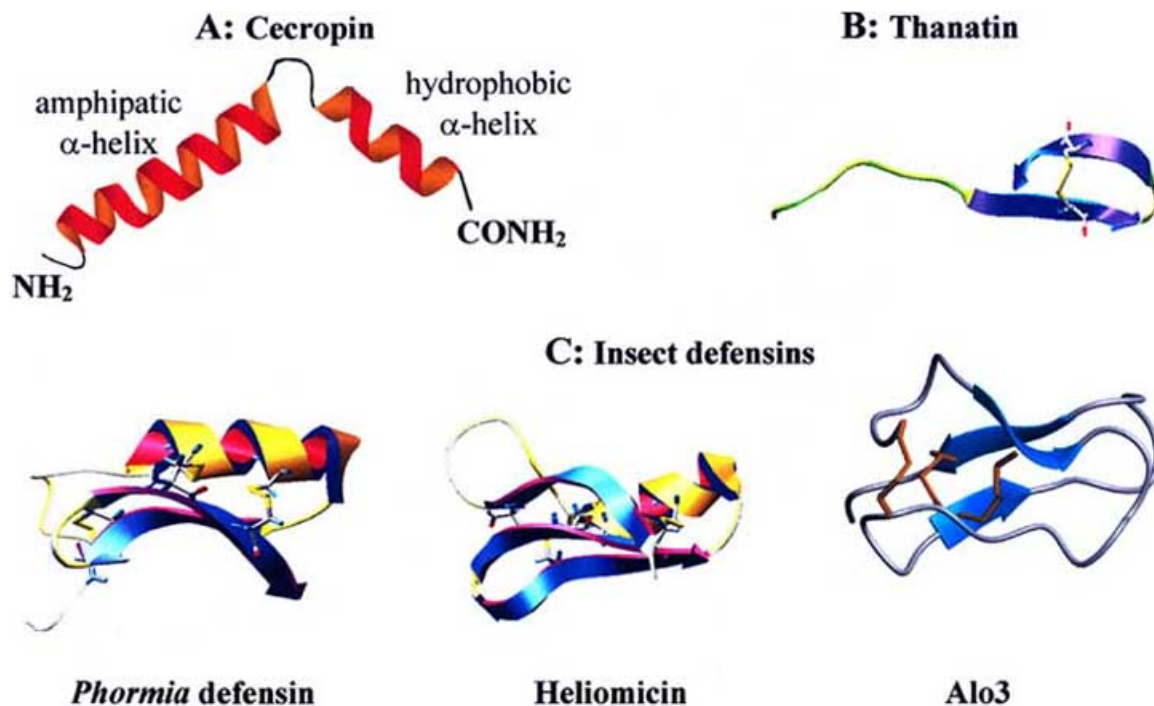


Figure 2. Global fold of five selected insect AMPs. *Hyalophora* cecropin A, (A), thanatin from *Podisus maculiventris* (B), the antibacterial *Phormia* defensin, the antifungal heliomicin from *Heliothis virescens* and Alo3 from *Acrocisus longimanus* (C). NH₂ stands for the N-terminus and CONH₂ for the amidated C-terminus.

CYSTEINE-STABILIZED PEPTIDES, MIXED α -HELICAL/ β -SHEET, TRIPLE-STRANDED β -SHEET OR HAIRPIN-LIKE STRUCTURES

AMPs with an even number of cysteine residues are widely distributed in insects. Based on their secondary structure in aqueous solutions or by sequence homology, they can be classified to three main categories: (i) peptides belonging to the family of defensins with a α -helix/ β -sheet (,) mixed structure [2,5,15], (ii) peptides with a triple-stranded antiparallel β -sheets [34] and (iii) peptides forming a hairpin-like β -sheet structure [5,35].

INSECT DEFENSINS

Defensins were found in every insect species investigated to date. This family of small, cationic peptides includes the phormicins, sapecins, royalisin and spodopterin [36]. Defensins, were first reported from cell cultures of the flesh fly *Sarcophaga peregrina* [37] and from experimentally injured larvae of the black blowfly (*Phormia terranova*) [38]. More than 60 defensins have been isolated from insects belonging to phylogenetically recent orders (Diptera, Lepidoptera, Coleoptera, Hymenoptera) and to the ancient order of Odonata (dragonfly). Surprisingly, the primary structure of the dragonfly defensin, with over 75% identity, is closely related to the mollusk and scorpion defensins [2]. Comparison of the peptide sequences revealed 70 to almost 100% homology among the group of dipteran defensins and indicated that these defensins have more than 35% similarity to the defensin from the dragonfly. This is particularly interesting as dragonflies evolved about 100 million years before the emergence of the dipteran insects. Defensins are larger than cecropins with a size of 33-46 residues, and three to four internal disulfide bridges stabilize their structure. A further N-terminally extended defensin has been reported in the gut of the stable fly [39], while C-terminally extended defensins/royalisin have been reported in the bumblebee *Bombus pascuorum* (Hymenoptera) and in the royal jelly of the honeybee. In contrast to cecropins that are frequently C-terminally amidated, only two defensins (royalisin and bumblebee defensin) were found to be amidated at their C-terminus. In respect to their *in vitro* activity against bacteria or filamentous fungi, defensins can be classified in two sub-families: antibacterial defensins that stop preferentially the growth of bacteria and antifungal defensins that are predominantly effective against filamentous fungi. While defensins with antibacterial activities are abundantly discussed in the literature, defensins with predominantly or strictly antifungal properties are poorly documented. Indeed only few antifungal defensins have been reported [5]: drosomycin from *Drosophila*; heliomicin from the tobacco budworm *Heliothis virescens* (Lepidoptera), gallerimycin from the greater wax moth larvae *Galleria mellonella* (Lepidoptera) [40], termicin from the isopteran *P. spiniger* and Alo13 from the harlequin beetle *Acrocinus longimanus* (Coleoptera) [34]. The relatively low number of strictly antifungal defensins as reported in the literature is certainly not a reflection of the lack of insect antifungal peptides but rather a lack of deliberate search for such molecules.

Antibacterial defensins are particularly efficacious against Gram-positive bacteria, including human pathogens, with minimal inhibitory concentrations (MIC) often below the μ M range. However, these peptides kill bacteria only in test media low of ionic strength. Increasing the salt concentration of the culture medium present in physiological conditions rapidly decreases the *in vitro* efficacy of antibacterial defensins. Gram-negative bacteria, yeast and filamentous fungi are less sensitive to these defensins. As a positive feature, no lytic activity has been recorded on red blood cells even at concentrations higher than several hundred μ M. Most antibacterial defensins kill bacteria in less than a minute with the exception of the dragonfly defensin that requires at least 30 minutes to exert noticeable activity. *Phormia* defensin disrupts the permeability barrier of the cytoplasmic membrane of the Gram-positive strain *Micrococcus luteus*, resulting in a loss of cytoplasmic potassium, a partial depolarization of the inner membrane, a decrease in cytoplasmic ATP, and an inhibition of respiration [41]. It remains to be established whether the mode of action of the dragonfly defensin and comparable molecules from mollusks and scorpions is similar to the one exhibited by the *Phormia* defensin. In contrast to the antibacterial defensins that have a rather broad spectrum of activity, the activity of most of the antifungal defensins is restricted to filamentous fungi and yeast cells. From these, only termicin has the potency to affect the growth of a limited number of Gram-positive bacteria. Strictly antifungal defensins act through the inhibition of spore germination or hyphal perforation. Recently, Thevissen and co-workers established that *RsAFP2*, a plant antifungal peptide from radish seed (*Raphanus sativus*) with sequence similarities to heliomicin, interacts with fungal glucosylceramides in an initial step of the process of fungal growth inhibition [42]. While this interaction is not by itself sufficient, it is nevertheless necessary to induce fungal growth arrest. As far as the mode of action of antifungal defensins is concerned, breakthrough studies are yet to surface. Because the size of insect defensins is at the upper limit of conventional solid-phase chemical peptide synthesis, for SAR studies, most of these molecules have to be produced using recombinant strategies. Figure 3 details the primary structure of insect defensins and the organisms where these peptides were isolated.

The 3D solution structures of *Phormia* and *Sarcophaga* defensins and of four antifungal defensins (drosomycin, heliomicin, termicin and Alo3) have been solved on recombinant molecules by $^1\text{H-NMR}$ spectroscopy [5,42]. The structure of the antibacterial defensins and termicin can be described as an α -helical domain and two antiparallel β -strands stabilized by two disulfide bridges (). For heliomicin and drosomycin an additional scaffold is presented by another short N-terminal β -strand (). The typical and unique triple-stranded β -sheet scaffold of insect defensins () has been observed for Alo3 isolated from the harlequin beetle. A second interesting finding is that Alo3 has all the structural characteristics of bacterial growth inhibitors of the cystine-knot family [42], opening the perspectives of a new mode of action for Alo3-type antifungal molecules (see Figure 2C).

THANATIN, A HAIRPIN-LIKE β -SHEET PEPTIDE

Among the group of insect AMPs containing an even number of cysteine residues, thanatin is the smallest one containing only 21 amino acids including two cysteine residues. Thanatin, with noticeable sequence similarities to frog AMPs of the brevinin family [2], has been isolated from the spined soldier bug *Podisus maculiventris* [43]. Thanatin has a broad spectrum of activity, killing (at μ M concentrations) both Gram-positive and Gram-negative bacteria, filamentous fungi and yeasts without measurable toxicity for red blood cells. A recent study by Pagès and co-workers on the activity of thanatin against multidrug resistant bacteria isolated from hospitalized patients (*Enterobacter aerogenes* and *Klebsiella pneumonia*, [44]) evidenced that the activity of this peptide depends on the size of the lipopolysaccharide/endotoxin (LPS), a component of the outer membrane of Gram-negative bacterial cell wall. The authors suggested that the accessibility of thanatin to an internal target is improved when the size of LPS is decreased. SAR studies established that an all-D enantiomer is ineffective against Gram-negative bacteria, but exhibits the same level of activity as the natural L peptide on fungi. The *in vitro* efficacy of all-D thanatin against Gram-positive bacteria is strain dependent [43]. This suggests that for killing different types of microorganisms, thanatin uses different mechanisms of action, involving a stereospecific interaction with a bacterial target. Additional SAR studies on thanatin included activity measurements of N- and/or C-terminally truncated forms [43], or analogues containing amino acid substitutions within the disulfide loop [45], and functional mapping by random mutagenesis in *E. coli* combined with the determination of the solution structure by NMR spectroscopy and molecular dynamics calculations [46]. It appears that (i) elimination of the C-terminal

methionine residue diminishes the efficacy against Gram-negative bacteria, and removal of the last two residues completely abolishes the activity; (ii) the deletion of the five N-terminal residues retains activity against some Gram-negative bacteria, while the removal of two additional residues is deleterious for the activity; (iii) the deletion of amino acids within the disulfide loop increases the efficacy against Gram-positive strains, but an insertion of an alanine residue to the disulfide loop decreases the activity against all types of bacteria; and (iv) maintenance a hydrogen bonding network within the C-terminal loop linking the two antiparallel β -strands is needed for the activity.

The 3D structure of thanatin has been elucidated by 1 H 2D NMR spectroscopy and molecular modelling [46,47]. The peptide adopts a well-defined, two-stranded β -sheet structure, stabilized by the intramolecular disulfide bridge, delineating a box extended by a rather flexible N-terminal tail. While the central part of the molecule is hydrophobic, the outside sites (loop and N-terminal tail) are rich in hydrophilic residues (see Figure 2B).

Thanatin appears to be a good candidate for the development of novel therapeutic agents against multidrug resistant bacteria because this peptide (i) is devoid of toxicity toward eukaryotic cells, (ii) kills bacteria including those featuring multidrug resistant phenotypes, (iii) acts on bacteria through different modes of action, and (iv) is short enough to facilitate SAR studies.

LINEAR PROLINE-RICH AMPs

The proline-rich peptides are linear molecules and are made up from 14-39 amino acids (the two extremes are 14 residues for the apidaecin from the baldfaced hornet and 39 residues for abaecin from a bumblebee), [2,48]. They are

Organism of origin	Name	Primary structure
Antimicrobial peptide with a single disulfide bridge		
<i>Podisus maculiventris</i>	Thanatin	GSKPKVPIIY <u>C</u> NRRTGK <u>C</u> QRM
Antimicrobial peptides with three disulfide bridges		
<i>Sarcophaga peregrina</i>	Sapecin	AT <u>C</u> DLLSGTGINHSACAAH <u>C</u> LLRGNRGGY <u>C</u> NGKAV <u>C</u> VCRN
<i>Phormia terranovae</i>	Defensin A	AT <u>C</u> DLLSGTGINHSACAAH <u>C</u> LLRGNRGGY <u>C</u> NGKGV <u>C</u> VCRN
<i>Stomoxys calcitrans</i>	SmD1	AAKPMGIT <u>C</u> DLLSLWKVGHAA <u>C</u> AAH <u>C</u> LVLDGVDGGY <u>C</u> TKEGL <u>C</u> V <u>C</u> KE
<i>Galleria mellonella</i>	Gallerimycin	GVTITVKPPFP <u>G</u> VFYE <u>C</u> IANC <u>R</u> SRGYKNGGY <u>C</u> TING <u>C</u> Q <u>C</u> LR
<i>Heliiothis virescens</i>	Heliomicin	DKLIGS <u>C</u> VWGAVNYS <u>D</u> CNGE <u>C</u> KRRGYKGGH <u>C</u> GSFANVN <u>C</u> W <u>C</u> ET
<i>Pseudacanthohermes spiniger</i>	Termicin	A <u>C</u> NFQS <u>C</u> WAT <u>C</u> QAQHSIYFRRAF <u>C</u> DRS <u>C</u> K <u>C</u> VFVRG*
<i>Apis mellifera</i>	Royalisin	VT <u>C</u> DLLSFKGQVND <u>S</u> ACAA <u>N</u> CLSLGKAGGH <u>C</u> EKVG <u>C</u> I <u>C</u> RKTSFKDLWDKRF*
<i>Aeschna cyanea</i>	Defensin	GFG <u>C</u> PLDQM <u>Q</u> CHR <u>H</u> CQTITGRSGGY <u>C</u> SGPLKLT <u>C</u> T <u>C</u> YR
Antimicrobial peptide with four disulfide bridges		
<i>Drosophila melanogaster</i>	Drosomycin	D <u>C</u> LSGRYKGP <u>C</u> AVWDNET <u>C</u> RRV <u>C</u> KEEGRSSGH <u>C</u> SPSLK <u>C</u> W <u>C</u> EG <u>C</u>

Figure 3. Primary structures of selected insect defensins and of thanatin. The * marks C-terminal amidation. Cysteine residues are underlined and printed in bold. All sequences reported in the figure are from the Swiss-Prot/TrEMBL database. The exact N-terminus of gallerimycin has not been experimentally defined. The presence of a dibasic cleavage site in the precursor molecule supports the hypothetical N-terminus proposed here.

classified in two sub-families: short-chain (less than 20 residues) and long-chain peptides (more than 20 amino acids). They have been isolated from bees and wasps (abaecins, apidaecins), from ants (formaecins), from the *D. melanogaster* (drosocins and metchnikowins), from the silkworm *Bombyx mori*, the cabbage looper moth *Trichoplusia ni* (lebocins) and from bugs (pyrrhocoricin, metalnikowins).

In addition to be over-represented in the sequences (with more than 25% of the total amino acid content), proline residues are frequently associated in doublets and triplets with basic residues (arginine or histidine). The PRP/PHP motifs are either evenly distributed along the entire peptide sequence or are concentrated in certain subdomains. Of particular interest is the presence, on some of the proline-rich AMPs, of O-linked carbohydrates attached to threonine and serine residues. The first member of naturally O-glycosylated AMPs to be isolated was drosocin as early as in 1993 [49]. Since this first report, AMPs bearing such post-translational modifications have been isolated from the European sap-sucking bug *Pyrrhocoris apterus* [50], the silkworm [51] and from the bulldog ant *Myrmecia gulosa* [52]. The biological role of the sugars remains contradictory as attachment of mono- and disaccharide side-chains improves the *in vitro* efficacy of drosocin and formaecin, but synthetic glycan-free pyrrhocoricin is more active than its natural O-glycosylated parent analogues. To make the situation more complicated, elongation of the glycan chain on lebocins tends to reduce the antimicrobial activity. Lebocins exhibit 40% structural similarity to abaecins and drosocin has 47% and 53% sequence similarity with apidaecins and pyrrhocoricin, respectively. The primary structure of the proline-rich peptides is found in Figure 4, together with the organisms of origin of these molecules.

Conformational analyses by NMR and CD spectroscopy in aqueous solution reveal that drosocin and its non-glycosylated forms essentially assume random coil structures. Adding 50% trifluoroethanol (v/v) induces the appearance of a small population of ordered conformations, most probably turns, and glycosylation results in a subtle shift toward a more extended conformation around the O-glycosylated threonine residue [53].

Remarkably, most of the short-chain proline-rich AMPs have a high selectivity toward Gram-negative bacteria belonging to the *Enterobacteriaceae* family, while Gram-positive strains remain mostly non-susceptible [48]. Interestingly, the long-chain abaecins and lebocins are active against both Gram-negative and Gram-positive bacteria, while metchnikowin exclusively kills filamentous fungi [2]. For lebocins, the activity is strongly salt dependent as it is significantly reduced even at physiological salt concentrations. Lebocins show synergistic activity with cecropin and when mixed together may improve the efficacy of unrelated AMPs by decreasing their MIC.

SAR studies on short-chain proline-rich AMPs, highlight some interesting biological properties. Unlike the α -helical and most of the cysteine stabilized AMPs that kill bacteria within minutes through a non-stereospecific fashion [54], the short-chain proline rich AMPs need hours to kill bacteria. In addition, all-D enantiomers of apidaecins, drosocins and pyrrhocoricins are totally ineffective suggesting that these peptides are bactericidal through a mechanism that includes stereoselective elements without any pore-forming process. Castle and co-workers [55] proposed a model of permease/transporter-mediated mechanism for the mode of action of apidaecins on the Gram-negative bacterium *E. coli*. Indeed, apidaecin uptake in *E. coli* is energy-driven, irreversible and can be partially competed by proline in a stereospecific manner. Complementary information on the mode of action of the short-chain proline-rich AMPs has been obtained following SAR studies on the drosocin/pyrrhocoricin model [56]. For example, Otvos and co-workers observed that shortening these peptides from either termini or cyclizing the molecule in a head-to-tail fashion destroys their activity. Using biotin- and fluorescein-labeled peptides, the same group evidenced that apidaecin, drosocin and pyrrhocoricin interact with the 70-kDa *E. coli* heat shock protein DnaK [57]. It was proposed that pyrrhocoricin (and perhaps drosocin) bind to the multihelical lid above the substrate-binding pocket of *E. coli* DnaK resulting in the inhibition of chaperone-assisted protein folding [58]. SAR studies and *in vivo* trials on mice, performed on pyrrhocoricin analogues including protected dimeric forms, evidenced a statistically significant reduction

Organism of origin	Name	Primary structure
<i>Drosophila melanogaster</i>	Drosocin	GKPRPY <u>S</u> PRP <u>T</u> SHRPPIRV
	Metchnikowin	HRHQGPIFDTRPSPFNPNQPRPGPIY
<i>Apis mellifera</i>	Apidaecin IA	GNNRPVYIPQPRPPHPI
<i>Bombus pascuorum</i>	Abaecin	FVPYNPPRPGQSKPFPSFPGHGFNPKIQWPYPLPNPGH
<i>Myrmecia gulosa</i>	Formaecin 1	GRPNPVNNKP <u>T</u> PHPRL
<i>Bombyx mori</i>	Lebocin 3	DLRFLYPRGKLPVPT <u>L</u> PPFNPKPIYIDMGNRY
<i>Trichoplusia ni</i>	Lebocin	SLPSLRLPGRNFP <u>I</u> PP <u>T</u> PPFVVKPRRFPIYV
<i>Pyrrhocoris apterus</i>	Pyrrhocoricin	VDKGSYLPRP <u>T</u> PPRPPIYNRN
<i>Palomena prasina</i>	Metalnikowin I	VDKPDYRPRRPPNM

Figure 4. Primary structures of selected insect proline-rich AMPs. All sequences listed in the figure are from the Swiss-Prot/TrEMBL database. The threonine or serine residues for O-glycosylation are underlined and printed in bold letters.

and elimination of the pathogenic bacteria in animals treated with the designed peptides [59,60]. As prominent sequence variations exist in the pyrrolic-binding region of eucaryotic and procaryotic DnaK molecules, short-chain proline-rich AMPs may represent promising starting points to design strain-specific antibiotics without activity on mammalian cells.

SENSING OF MICROORGANISMS BY THE FRUIT FLY *DROSOPHILA MELANOGASTER*: ACTIVATION OF TWO SIGNALLING PATHWAYS

Being devoid of adaptive immune reactions, *D. melanogaster* with its well-known genetic composition is recognized as an ideal model system for studying the innate immune defenses of insects. In *Drosophila*, seven distinct inducible AMPs have been identified [5,61]. Their activity is directed either against Gram-positive (defensin) and Gram-negative (diptericin, drosocin, attacins, cecropins) bacteria, as well as against filamentous fungi (drosomycin, metchnikowin). This arsenal of AMPs is released into the blood stream of the insect following synthesis by the fat body tissue of experimentally infected flies. The circulating concentrations of the AMPs reach overall values of 0.5 mM, which is far higher than the concentration required to kill most of the microorganisms *in vitro*.

The observation that a natural infection with an entomopathogenic filamentous fungus causes only the activation of the gene coding the antifungal peptides and not those coding the antibacterial ones suggested that *Drosophila* might discriminate between a bacterial and a fungal infection [62]. During the past few years, significant progress has been made towards the understanding of pathogen-sensing by *Drosophila* [63,64]. Two different signal transduction pathways, Toll and Immune-deficiency (Imd), emerged [64,65], and ensuing genetic and molecular studies led to a detailed characterization of these pathways. Most recently, it has been shown that microbial recognition, which occurs upstream of the two pathways, is modulated for Toll activation by at least peptidoglycan-recognition proteins (PGRPs) and Gram-negative binding proteins (GNBPs), the gene products *semmelweis* and *osiris*, respectively [64]. The Toll pathway predominantly activates the expression of the gene encoding the antifungal peptide drosomycin through a cascade of events. Toll-activation can be observed in response to either fungal infection or infection with Gram-positive bacteria. In contrast, the Imd pathway was found to be increasingly involved in the resistance of *Drosophila* to Gram-negative infections. Since excellent recent reviews can be consulted on fine details of these two pathways [64,65], only the main features of the pathways will be outlined here, with emphasis on their activation process.

THE TOLL PATHWAY

The activation of the Toll signalling pathway is dependent on a cleaved form of the cystine-knot cytokine-like protein Spaetzle. A dimeric form of the mature Spaetzle binds to the ectodomain of two Toll receptors. The intracytoplasmic domain of Toll then interacts with three proteins including Pelle, which contains a serine-threonine

kinase domain. This process results in the activation of either a dorsal-related immunity factor (DIF) in adults or Dorsal and/or DIF in larvae, both factors being closely related to NF- κ B proteins. The final result of this signalling is the dissociation of these proteins from the phosphorylated Cactus protein. Dif and/or Dorsal then translocate to the nucleus of immune cells where they activate the target immune genes such as *drosomycin*.

THE IMD PATHWAY

This pathway is predominantly implicated in the regulation of the genes encoding anti Gram-negatives AMPs, including the cecropins, drosocin, diptericin (a glycine-rich AMP) and attacins. The activation of the genes encoding these AMPs involves a NF- κ B protein (Relish) that is cleaved into two parts, together with the nuclear translocation of the N-terminal cleaved-domain. Ultimately promoters of the above-mentioned antimicrobial genes are activated in the nucleus. In cases when the sequence of events downstream of Imd was not fully dissected, genetic data suggested the involvement of a series of activation partners including a MAP kinase.

PROCESS OF ACTIVATION OF THE SIGNALLING PATHWAYS

There are several data available that support the existence of pattern-recognition molecules in *Drosophila* as well as in other insects. For example, the *Drosophila* genome contains 13 genes belonging to the PGRP family and three to the GNBPs. Mutagenesis studies revealed that the two genes *semmelweis* and *osiris*, which encode a PGRP (PGRP-SA) and a GBP, respectively, are involved in the survival of the flies to Gram-positive infections. Activation of the Toll pathway by fungal infection is not driven by the *semmelweis* and *osiris* genes, but occurs through *persephone*, a gene that encodes a blood trypsin-like protease, a mediator of Spaetzle and the Toll activation maturation. Recently, a member of the PGRP family (PGRP-LC) has been found to have a crucial role in the activation of the Imd pathway. PGRP-LC triggers the activation of this pathway upon encountering either a peptidoglycan or a bacterial lipopolysaccharide. In conclusion, recent studies on PGRPs and GNBPs provide correlative evidences for the notion that *Drosophila* has the capability to discriminate among pathogens and is able to mount the most appropriate immune response to fight off any given infection.

CONCLUDING REMARKS

In this review, we highlighted the structural and biological properties of insect AMPs and when available, provided results on SAR studies. What clearly emerging from the long list of structures available is that most of the AMPs from insects belong to a limited number of structural families. This reflects conserved structures not only within insects, but also within the animal and plant kingdoms. For example, insect defensins (antibacterial or antifungal) have their counterparts in arachnids, mollusks, plants and also to some extent in vertebrates. This suggests that such antimicrobial molecules may play an important role in the defense reactions developed by the host to fight off

infections. The next step is the validation of the *in vitro* results in animal models of infections or on naturally infected animals *in vivo*. The secretion of the AMPs in the bloodstream of insects where the antimicrobial peptide concentrations can reach 0.5 mM without any toxicity to the host is doubtless an element in favour of the potential development of those peptides as new systemic therapeutic drugs. SAR studies on the small-size proline-rich peptides, on thanatin and on the antifungal defensins have generated interesting data that are positioning these peptides as potential probes for R&D. However, one of the main concerns regarding the use of insect AMPs as therapeutic drugs is their production cost, which remains high compared to conventional antibiotics. On a positive note, peptide-manufacturing prices clearly tend to decrease with technological improvements brought upon the emergence of several novel peptide drugs on the pharmaceutical market. Another key question will be to elucidate whether bacteria or fungi are able to develop resistance to AMPs similar to those they have developed against conventional antibiotics. Finally, we have to keep in mind that nature has generated its own mutational program through evolution and biodiversity and that we still have a lot to learn from it.

ACKNOWLEDGEMENTS

We would like to thank Drs Françoise Vovelle and Céline Landon from the Centre de Biophysique Moléculaire UPR 4301-CNRS (Orléans, France) for providing us the global folds of the peptides reported in figure 2. We are grateful to Dr Laszlo Otvos from Wistar Institute (Philadelphia, USA), Dr Laure Menin and to Céline Duhot from Atheris Laboratories, and to OFES for financial support (OFES N°. 02.0360).

REFERENCES

- [1] Hoffmann, J.A., Kafatos, F.C., Janeway, C.A. and Ezekowitz, R.A. (1999) *Science*, 284, 1313-1318.
- [2] Bulet, P., Hetru, C., Dimarcq, J.-L. and Hoffmann, D. (1999) *Dev. Comp. Immunol.*, 23, 329-344.
- [3] Otvos, L. Jr. (2000) *J. Pept. Sci.*, 6, 497-511.
- [4] Hetru, C., Hoffmann, J.A. and Hancock, R.E.W. (2002) In *Peptide Antibiotics*, (Dutton, C.J., Haxell, M.A., McArthur, H.I.I. and Wax, R.G., Eds.) pp. 117-144. Marcel Dekker Inc., New York and Basel.
- [5] Bulet, P., Charlet, M. and Hetru, C. (2003) In *Innate Immunity*, (Ezekowitz, R.A.B. and Hoffmann, J.A., Eds.) pp. 89-107. Humana Press, Totowa, NJ.
- [6] Pag, U. and Sahl, H.G. (2002) In *Peptide Antibiotics*, (Dutton, C.J. Haxell, M.A. McArthur, H.I.I. and Wax, R.G., Eds.) pp. 47-80. Marcel Dekker Inc., New York and Basel.
- [7] Nes, I.F., Holo, H., Fimland, G., Hauge, H.H. and Nissen-Meyer, J. (2002) In *Peptide Antibiotics*, (Dutton, C.J., Haxell, M.A., McArthur, H.I.I. and Wax, R.G., Eds.) pp. 81-116. Marcel Dekker Inc., New York and Basel.
- [8] Garcia-Olmedo, F., Molina, A., Alamillo, J.M. and Rodriguez-Palenzuela, P. (1998) *Biopolymers*, 47, 479-491.
- [9] Zasloff, M. (2002) *Nature*, 415, 389-395.
- [10] Brogden, K.A. and Ackermann, M. McCray, P.B. Jr. and Tack, B.F. (2003) *Int. J. Antimicrob. Agents*, 22, 465-478.
- [11] Ganz, T. and Lehrer, R.I. (2003) In *Innate Immunity*, (Ezekowitz, R.A.B. and Hoffmann, J.A., Eds.) pp. 287-303. Humana Press, Totowa, NJ.
- [12] Lamberty, M., Zachary, D., Lanot, R., Bordereau, C., Robert, A., Hoffmann, J.A. and Bulet, P. (2001) *J. Biol. Chem.*, 276, 4085-4092.
- [13] Dathe, M. and Wieprecht, T. (1999) *Biochim. Biophys. Acta*, 1462, 71-87.
- [14] Tossi, A., Sandri, L. and Giangaspero, A. (2000) *Biopolymers*, 55, 4-30.
- [15] Dimarcq, J.L., Bulet, P., Hetru, C. and Hoffmann, J. (1998) *Biopolymers*, 47, 465-477.
- [16] Steiner, H., Hultmark, D., Engström, A., Bennich, H. and Boman, H.G. (1981) *Nature*, 292, 246-248.
- [17] Okada, M. and Natori, S. (1985) *J. Biol. Chem.*, 260, 7174-7177.
- [18] Kylsten, P., Samakovlis, C. and Hultmark, D. (1990) *EMBO J.*, 9, 217-224.
- [19] Sun, D., Eccleston, E.D. and Fallon, A.M. (1998) *Biochem. Biophys. Res. Commun.*, 249, 410-415.
- [20] Sun, D., Eccleston, E.D. and Fallon, A.M. (1999) *FEBS Lett.*, 454, 147-151.
- [21] Lowenberger, C., Charlet, M., Vizioli, J., Kamal, S., Richman, A., Christensen, B.M. and Bulet, P. (1999) *J. Biol. Chem.*, 274, 20092-20097.
- [22] Vizioli, J., Bulet, P., Charlet, M., Lowenberger, C., Blass, C., Muller, H.M., Dimopoulos, G., Hoffmann, J., Kafatos, F.C. and Richman, A. (2000) *Insect Mol. Biol.*, 9, 75-84.
- [23] Hara, S., Taniai, K., Kato, Y. and Yamakawa, M. (1994) *Comp. Biochem. Physiol.*, 108B, 303-308.
- [24] Li, Z.Q., Merrifield, R.B., Boman, I.A. and Boman, H.G. (1988) *FEBS Lett.*, 31, 299-302.
- [25] Andreu, D., Merrifield, R.B., Steiner, H. and Boman, H.G. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 6475-6479.
- [26] Orivel, J., Redeker, V., Le Caer, J.P., Krier, F., Revol-Junelles, A.M., Longeon, A., Chaffotte, A., Dejean, A. and Rossier, J. (2001) *J. Biol. Chem.*, 276, 17823-17829.
- [27] Boman, H.G., Wade, D., Boman, I.A., Wahlin, B. and Merrifield, R.B. (1989) *FEBS Lett.*, 259, 103-106.
- [28] Bulet, P., Stöcklin, R. and Menin, L. (2004) In *Immunol. Reviews*, Blackwell Publishing Ltd., in press.
- [29] Boulanger, N., Munks, R.J., Hamilton, J.V., Vovelle, F., Brun, R., Lehane, M.J. and Bulet, P. (2002) *J. Biol. Chem.*, 277, 49921-49926.
- [30] Haines, L.R., Hancock, R.E. and Pearson, T.W. (2003) *Vector Borne Zoonotic Dis.*, 3, 175-186.
- [31] Steiner, H. (1982) *FEBS Lett.*, 137, 283-287.
- [32] Holak, T.A., Engstrom, A., Kraulis, P.J., Lindeberg, G., Bennich, H., Jones, T.A., Gronenborn, A.M. and Clore, G.M. (1988) *Biochemistry*, 27, 7620-7629.
- [33] Zasloff, M. (1987) *Proc. Natl. Acad. Sci. USA*, 84, 5449-5453.
- [34] Barbault, F., Landon, C., Guenneugues, M., Meyer, J.P., Schott, V., Dimarcq, J.L. and Vovelle, F. (2003) *Biochemistry*, 42, 14434-14442.
- [35] Mandard, N., Bulet, P., Hetru, C., Landon, C. and Vovelle, F. (2002) In *Membrane Interacting Peptides and Proteins*, (Heitz, F., Ed) pp.155-171. Research Signpost, Trivandrum, India.
- [36] Volkoff, A.N., Rocher, J., d'Alençon, E., Bouton, M., Landais, I., Quesada-Moraga, E., Vey, A., Fournier, P., Mita, K. and Devauchelle, G. (2003) *Gene*, 319, 43-53.
- [37] Matsuyama, K. and Natori, S. (1988) *J. Biol. Chem.*, 263, 17112-17116.
- [38] Lambert, J., Keppi, E., Dimarcq, J.L., Wicker, C., Reichhart, J.M., Dunbar, B., Lepage, P., Van Dorsseleer, A., Hoffmann, J.A., Fothergill, J. and Hoffmann, D. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 262-266.
- [39] Lehane, M.J., Wu, D. and Lehane, S.M. (1997) *Proc. Natl. Acad. Sci. USA*, 94, 11502-11507.
- [40] Schuhmann, B., Seitz, V., Vilcinskis, A. and Podsiadlowski, L. (2003) *Arch. Insect Biochem. Physiol.*, 53, 125-133.
- [41] Cociancich, S., Ghazi, A., Hetru, C., Hoffmann, J.A. and Letellier, L. (1993) *J. Biol. Chem.*, 268, 19239-19245.
- [42] Thevisen, K., Warnecke, D.C., Francois, I.E., Leipelt, M., Heinz, E., Ott, C., Zahringer, U., Thomma, B.P., Ferket, K.K. and Cammue, B.P. (2004) *J. Biol. Chem.*, 279, 3900-3905.
- [43] Fehlbaum, P., Bulet, P., Chernysh, S., Briand, J.P., Roussel, J.P., Letellier, L., Hetru, C. and Hoffmann, J.A. (1996) *Proc. Natl. Acad. Sci. USA*, 93, 1221-1225.
- [44] Pagès, J.M., Dimarcq, J.L., Quenin, S. and Hetru, C. (2003) *Inter. J. Antimicrob. Agents*, 22, 265-269.
- [45] Lee, M.K., Cha, L., Lee, S.H. and Hahn, K.-S. (2003) *J. Biochem. Mol. Biol.*, 35, 291-296.

- [46] Taguchi, S., Kuwasako, K., Suenaga, A., Okada, M. and Momose, H. (2000) *J. Biochem.*, 128, 745-754.
- [47] Mandard, N., Sodano, P., Labbe, H., Bonmatin, J.M., Bulet, P., Hetru, C., Ptak, M. and Vovelle, F. (1998) *Eur. J. Biochem.*, 256, 404-410.
- [48] Otvos, L. Jr. (2002) *Cell. Mol. Life Sci.*, 59, 1138-1150.
- [49] Bulet, P., Dimarcq, J.L., Hetru, C., Lagueux, M., Charlet, M., Hegy, G., Van Dorsselaer, A. and Hoffmann, J.A. (1993) *J. Biol. Chem.*, 268, 14893-14897.
- [50] Cociancich, S., Dupont, A., Hegy, G., Lanot, R., Holder, F., Hetru, C., Hoffmann, J.A. and Bulet, P. (1994) *Biochem. J.*, 300, 567-575.
- [51] Hara, S. and Yamakawa, M. (1995) *Biochem. J.*, 310, 651-656.
- [52] Mackintosh, J.A., Veal, D.A., Beattie, A.J. and Gooley, A.A. (1995) *J. Biol. Chem.*, 273, 6139-6143.
- [53] McManus, A.M., Otvos, L., Jr. Hoffmann, R. and Craik, D.J. (1999) *Biochemistry*, 38, 705-714.
- [54] Oren, Z. and Shai, Y. (1998) *Biopolymers*, 47, 451-463.
- [55] Castle, M., Nazarian, A., Yi, S.S. and Tempst, P. (1999) *J. Biol. Chem.*, 274, 32555-32564.
- [56] Hoffmann, R., Bulet, P., Urge, L., Otvos, L. Jr. (1999) *Biochim. Biophys. Acta*, 1426, 459-467.
- [57] Otvos, L. Jr. O., I. Rogers, M.E., Consolvo, P.J., Condie, B.A., Lovas, S., Bulet, P. and Blaszczyk-Thurin, M. (2000) *Biochemistry*, 39, 14150-14159.
- [58] Kragol, G., Lovas, S., Varadi, G., Condie, B.A., Hoffmann, R. and Otvos, L. Jr. (2001) *Biochemistry*, 40, 3016-3026.
- [59] Cudic, M., Condie, B.A., Weiner, D.J., Lysenko, E.S., Xiang, Z.Q., Insug, O., Bulet, P. and Otvos, L. Jr. (2002) *Peptides*, 23, 2071-2083.
- [60] Cudic, M., Lockatell, C.V., Johnson, D.E. and Otvos, L. Jr. (2003) *Peptides*, 24, 807-820.
- [61] Meister, M., Hetru, C. and Hoffmann, J.A. (2000) In *Current topics in microbiology and immunology*, (Du Pasquier, L. and Litman, G.W., Ed), pp. 17-36. Springer-Verlag, New York.
- [62] Lemaitre, B., Reichhart, J.M., Hoffmann, J.A. (1997) *Proc. Natl. Acad. Sci. USA*, 94, 14614-14619.
- [63] Hultmark, D. (2003) *Curr. Opin. Immunol.*, 15, 12-19.
- [64] Hoffmann, J.A. (2003) *Nature*, 426, 33-38.
- [65] Royet, J., Meister, M. and Ferrandon, D. (2003) In *Innate Immunity*, (Ezekowitz, R.A.B. and Hoffmann, J.A., Eds.) pp. 89-107. Humana Press, Totowa, NJ.

Received on February 12, 2004, accepted on March 22, 2004.