Insect Biochemistry and Molecular Biology 39 (2009) 792–800

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

A peptidomics study reveals the impressive antimicrobial peptide arsenal of the wax moth *Galleria mellonella*

Susan E. Brown [a,*], Antoinette Howard [b], Annette B. Kasprzak [b], Karl H. Gordon [b], Peter D. East [b]

[a] CSIRO Entomology, PO Box 225, Dickson ACT 2602, Australia
[b] CSIRO Entomology, Canberra, PO Box 1700, Acton ACT 2601, Australia

ARTICLE INFO

Article history:
Received 24 June 2009
Received in revised form
7 September 2009
Accepted 21 September 2009

Keywords:
Antimicrobial peptide
Galleria mellonella
Peptidomics
Proline-rich peptide
Antifungal
Immune response

ABSTRACT

The complete antimicrobial peptide repertoire of *Galleria mellonella* was investigated for the first time by LC/MS. Combining data from separate trypsin, Glu-C and Asp-N digests of immune hemolymph allowed detection of 18 known or putative *G. mellonella* antimicrobial peptides or proteins, namely lysozyme, moricin-like peptides (5), cecropins (2), gloverin, Gm proline-rich peptide 1, Gm proline-rich peptide 2, Gm anionic peptide 1 (P1-like), Gm anionic peptide 2, galiomicin, gallerimycin, inducible serine protease inhibitor 2, 6tox and heliocin-like peptide. Six of these were previously known only as nucleotide sequences, so this study provides the first evidence for expression of these genes. LC/MS data also provided insight into the expression and processing of the antimicrobial Gm proline-rich peptide 1. The gene for this peptide was isolated and shown to be unique to moths and to have an unusually long precursor region (495 bp). The precursor region contained other proline-rich peptides and LC/MS data suggested that these were being specifically processed and were present in hemolymph at very high levels. This study shows that *G. mellonella* can concurrently release an impressive array of at least 18 known or putative antimicrobial peptides from 10 families to defend itself against invading microbes.

1. Introduction

Antimicrobial peptides and proteins (AMPs) have been identified in plants and animals and shown to play a fundamental role in host defence against bacteria and fungi (Zasloff, 2002). The diversity of AMPs is high, with hundreds of different peptides and proteins identified to date (Wang and Wang, 2004; Brahmachary et al., 2004). Most organisms produce an array of AMPs with different specificities, presumably to maximise their defensive response and adaptability (Zasloff, 2002). AMPs from all organisms can be classified broadly into three groups according to their basic structure. These groups are the α-helical peptides such as the cecropins and moricins, the cysteine containing peptides such as the defensins, and linear peptides with an abundance of particular amino acids such as proline, glycine or tryptophan (Bulet et al., 2003). AMPs are usually derived from a single gene which has an N-terminal precursor region of approximately 60–80 bp including a typical signal sequence, and occasionally there are C-terminal regions that also require processing (Otvos, 2000; Boman et al., 1989). In a few cases, such as the proline-rich apidaecin peptides from bees (Li et al., 2006), the AMPs occur in multiple copies in the gene, and require complex processing to release multiple active peptides from the single precursor. Many AMPs have been shown to destroy membranes, although some such as the proline-rich peptides have recently been shown to have intracellular targets (Shai, 2002; Gennaro et al., 2002).

Insects have been a rich source of AMPs, with a cecropin from the moth *Hyalophora cecropia* (Steiner et al., 1981) being one of the first AMPs isolated. In *Drosophila melanogaster*, genome data has been correlated with earlier biochemical studies to show that the fly produces at least seven different AMPs, most in multiple forms (Boman et al., 1989; Lemaitre and Hoffmann, 2007). Comprehensive study of the array of AMPs produced by other insect species is now possible in a few cases where genomes are known, such as the moth *Bombyx mori*, the mosquito *Anopheles gambiae*, the honeybee *Apis mellifera*, and the beetle *Tribolium castaneum*. For example, in *B. mori*, five different families of AMPs have been identified – the linear α-helical cecropins (including enbocin) and the moricin-like peptides, glycine-rich gloverins and attacins, and the proline-rich lebocins (Xia et al., 2004; Cheng et al., 2006). In other Lepidoptera, cysteine-rich defensin peptides have also been identified, such as gallerimycin and galiomicin in *Galleria mellonella* (Lee et al., 2004; Schuhmann et al., 2003). However, in general, it is challenging to identify short peptides such as AMPs from sequence data alone, so it has been difficult to study the complete immune response of

* Corresponding author. Tel.: +61 2 6276 6960; fax: +61 2 6276 6273.
E-mail address: sue.brown@csiro.au (S.E. Brown).

0965-1748/$ – see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.ibmb.2009.09.004

a single insect species in order to identify the array of AMPs produced and their interactions.

Peptidomic studies have the potential to provide insight into the systemic response of an organism to stimuli such as a bacterial infection. Such direct functional studies are particularly important in an era when function is often inferred only by similarity to a known gene sequence. However, to date there have been only a few peptidomic studies on the immune response of insects, and these were all in *Drosophila* (Schoofs and Baggerman, 2003; Levy et al., 2004; Verleyen et al., 2006). Furthermore, current proteomics methods are often not suitable for studying the peptides of the immune response. This is because the AMPs are typically 2–5 kDa, which is too large for *de novo* sequencing of the entire peptide and too small for easy analysis by 2D gels (Baggerman et al., 2004). Furthermore, they are often rich in arginine and lysine residues which decreases the usefulness of the standard trypsin digests. This study sought to overcome these limitations in order to investigate the total AMP repertoire of *G. mellonella* in response to a bacterial infection. This paper thus describes a peptidomics study of hemolymph from immune-stimulated *G. mellonella*, including evidence for the translation of six *G. mellonella* AMPs previously known only from cDNA sequences, and new knowledge about an unusual proline-rich peptide. It also gives a picture of the total *G. mellonella* peptide response to an infection, demonstrating the simultaneous release of a large repertoire of at least 18 AMPs from at least 10 families.

2. Materials and methods

2.1. Peptidomics of G. mellonella hemolymph

Control and immune-stimulated hemolymph samples of *G. mellonella* were prepared and purified as described previously (Brown et al., 2008). In summary, larvae were injected with buffer (control hemolymph) or *Micrococcus luteus* and *Escherichia coli* bacteria (immune hemolymph), and the larval hemolymph isolated after 48 h and concentrated. Both hemolymph samples were partially purified by C_{18} solid phase extraction (unfractionated samples). The immune hemolymph was then purified further on a C_{18} semi-preparative HPLC column and 11 active fractions chosen based on their antimicrobial activity in an inhibition zone plate assay (fractionated sample) (Fig. 1).

For the LC/MS peptidomics analysis, three hemolymph samples were analysed – the unfractionated control, unfractionated immune and fractionated immune hemolymph. All three samples (2 μl) were digested with trypsin or Glu-C(DE) (Roche Applied Science), and the immune hemolymph fractions were also digested with Asp-N (Roche Applied Science). Prior to protease digest all samples were treated with DTT and iodoacetamide to modify cysteine residues. DTT (approximately 15 mM) was added to each sample before heating at 90 °C for 15 min followed by cooling to room temperature. Each sample was then diluted approximately two-fold by adding an equal amount of the appropriate 2x protease digest buffer (see below) plus fresh iodoacetamide (approximately 7.5 mM), and left in the dark for 30 min. Trypsin (7 ng/μl) digests were carried out in 100 mM TRIS buffer at pH 8.5 at 37 °C for approximately 3 h or overnight. Glu-C (DE) (0.25 ng/μl) digests were carried out in 100 mM sodium phosphate buffer at pH 7.8 at room temperature for approximately 3 h. The Asp-N (0.2 ng/μl) digest was carried out in 50 mM sodium phosphate buffer at pH 8 overnight. All samples were acidified to approximately 0.1% formic acid, centrifuged briefly and transferred to a 96 well plate. Digested peptide samples were loaded onto a Zorbax SB-C_{18} 5 μm 150 × 0.5 mm column (Agilent) running on an 1100 capillary liquid chromatography system (Agilent). Bound peptides were eluted with a gradient of 20–50% acetonitrile plus 0.1% formic acid at 5 μl/min over 28 min.

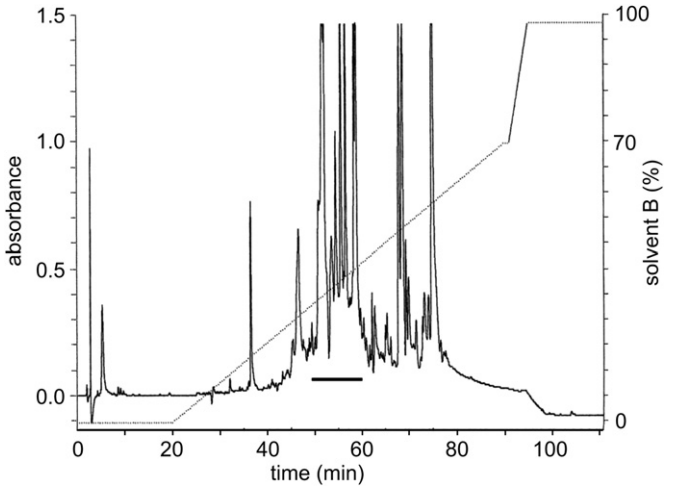

Fig. 1. HPLC trace from a semi-preparative C_{18} column showing the purification of AMPs from immune-stimulated *G. mellonella* hemolymph. The approximate locations of fractions which showed antimicrobial activity are marked with a solid line. The dotted line indicates the gradient used for peptide elution, corresponding to 0–70–100% solvent B (equivalent to 2–67–95% acetonitrile), where solvent A is 2% acetonitrile/0.065% TFA and solvent B is 95% acetonitrile/0.05% TFA.

Eluate from the column was introduced to an XCT ion trap mass spectrometer (Agilent) through a micronebuliser electrospray ion source. As peptides were eluting from the column, the ion trap collected full spectrum positive ion scans (100–2200 *m*/*z*) followed by four MS/MS scans of ions observed in the full spectrum according to the instrument's 'SmartFrag' and 'Peptide Scan' settings. Once two fragmentation spectra were collected for any particular *m*/*z* value it was excluded from selection for analysis for a further 30 s to avoid collecting redundant data.

LC/MS data were analysed using Spectrum Mill MS Proteomics Workbench (Rev A.03.03.078, Agilent Technologies). Data were extracted for each LC/MS run using default parameters, except that the "Merge scans with same precursor" parameter was set to 30 s and Cys residues were assumed to have a fixed carbamidomethylation modification. The MS data was analysed in two basic ways – first to identify digest peptides from known proteins and peptides, and second to identify novel sequences based on *de novo* sequencing. For the former, a non-redundant database of all Lepidoptera proteins was extracted from NCBI (30/4/08) and augmented with any *G. mellonella* AMP sequences known but not deposited in the database. A table of the 24 known and putative AMPs from *G. mellonella*, including the partial sequences of gloverin and Inducible Serine Protease Inhibitors (ISPI) 1–3, is shown in Supplementary Table 1. The known *G. mellonella* AMPs belong to 10 different families. Note that early analysis using a non-redundant all species database extracted from NCBI did not discover any extra matches than using the smaller lepidopteran database, so the latter was used in all subsequent processing. MS/MS identity searches were conducted with variable carbamidomethylation and methionine oxidation allowed, and false positives reduced using reversed database scoring, proton mobility scoring and dynamic peak thresholding (weightings favouring more probable ionisation and fragmentation patterns). The searches were conducted first with higher stringency (digest enzyme specified and two missed cleavages allowed) and then with lower stringency (no enzyme specified and two missed cleavages allowed). The latter conditions were necessary for confident identification of AMPs of this size, which were too long to be sequenced completely without digestion but which only produced a few peptides from protease digests. The spectra were then autovalidated using the default settings except that the protein filtering score was reduced to 12. The scoring in Spectrum Mill is such that a protein match with multiple