Molecular characterization of a male-specific glycosyl hydrolase, Lma-p72, secreted on to the abdominal surface of the Madeira cockroach *Leucophaea maderae* (Blaberidae, Oxyhaloinae)

Richard CORNETTE*†, Jean-Pierre FARINE*¹, Dehbia ABED-VELLARD*, Brigitte QUENNEDEY* and Rémy BROSSUT*

*C.N.R.S., U.M.R. 5548, 'Développement – Communication Chimique', Université de Bourgogne, 6 boulevard Gabriel, 21000 Dijon, France, and †Department of Biology, Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

The epicuticular surface protein Lma-p72 is specific to the abdominal secretions of *Leucophaea maderae* (Madeira cockroach) adult males. Natural Lma-p72 was purified and the complete cDNA sequence determined by reverse-transcription PCR using primers based on Edman degradation fragments. Northern blot and *in situ* hybridization analyses showed that Lma-p72 was expressed in the tergal and sternal glands. Sequence alignment indicates that Lma-p72 is closely related to the family 1 glycosyl hydrolases (EC 3.2.1). Native Lma-p72 was proved to be active in the abdominal secretions and exhibit a β -galactosidase-like activity. However, weak specificity with respect to the C-4 configuration of the substrate was observed. Two main hypotheses were proposed concerning the function of this enzyme: Lma-

p72 could hydrolyse oligosaccharides from the male abdominal secretions, making them more phagostimulatory for the female during the precopulatory behaviour. The protein could also cleave a pheromone–sugar conjugate to release the pheromonal compounds on to the cuticular surface. Such a sugar conjugate could be a transport form. Data from the first *in vivo* inhibition tests indicate that a glycosidase could be directly involved in the production process of some pheromonal compounds in *L. maderae* males.

Key words: cockroach, glycosyl hydrolase, sternal gland, surface protein.

INTRODUCTION

Chemical communication is remarkably developed in insects and involves two main types of compounds. Volatile pheromones play a role mainly in sexual attraction and distance communication, whereas heavier and non-volatile molecules are essential for contact recognition. Until now, cuticular hydrocarbons were considered as the main fraction of a complex mixture of the compounds, commonly called 'cuticular waxes', that cover insect cuticles. The presence of exocrine proteins (surface proteins) in this layer was described in various insects, such as as honeybees (*Apis mellifera*), migratory locusts (*Locusta migratoria*) or stinkbugs (*Pyrhocoris miridae* and *Maladera matrida*), but little is known about their function [1]. Surface proteins have been recently studied in Madeira cockroaches (*Leucophaea maderae*) [2,3] and are found on the abdomen of these insects with different levels of specificity. General surface proteins are secreted on to the cuticular surface of both adult sexes [4,5], whereas other proteins, specific to abdominal glands of the adult male, are involved in the sexual behaviour [6,7].

In the cockroach *Leucophaea maderae*, a volatile sex pheromone produced by the male sternal glands attracts the female from a distance [8]. Once both sexes are in contact, the male raises its wings, displaying its tergal glands, which are covered with proteinaceous secretions. The female climbs on to the male back to feed on these secretions. Then the female is in a proper position for mating and the male grasps its genitalia. Different proteins were recently characterized from these secretions. Two

male-specific tergal proteins, Lma-p18 and Lma-p22, were first characterized as belonging to the calycin superfamily [3,9]. Both of these proteins could bind sexually attractive compounds, similarly to other members of the calycin superfamily [10], or play a pheromonal role by themselves as structurally related lipocalins [11]. Those cockroach calycins were proposed to constitute an orientated cue for the females on to the male tergites (dorsal body plates) [3]. A third exocrine protein, Lma-p54, was described in both sexes [2]. This protein was characterized as belonging to the aspartic-proteinase family and was proposed to be involved in intraspecific contact recognition [5]. These previous studies suggest that surface exocrine proteins may be involved in the chemical communication in cockroaches.

The present paper reports the characterization of Lma-p72, a new male-specific surface protein in *L. maderae*. The complete nucleotide sequence encoding the protein was obtained and the localization of its expression was investigated. The enzymic activity of Lma-p72 was demonstrated, and its involvement in the production of sexually attractive compounds suggested.

EXPERIMENTAL

Insects

All the experimental conditions for rearing the insects and the methods to analyse protein secretion are the same as those previously described [2]. Briefly, insects were immobilized by cooling them 15 min at 4 °C, and the surface abdominal secretions

Abbreviations used: RACE, rapid amplification of cDNA ends; X-Gal, 5-bromo-4-chloroindol-3-yl β -D-galactopyranoside; LPH, lactase-phlorizin hydrolase.

¹ To whom correspondence should be addressed (e-mail Jean-Pierre.Farine@u-bourgogne.fr).

The nucleotide sequence reported in the present paper has been deposited with the EMBL, GenBank®, DDBJ and GSDB Nucleotide Sequence Databases under the accession number AY064214.

wiped with a small paintbrush soaked with water. For SDS/PAGE analysis, proteins were precipitated in 75 % (v/v) acetone. Samples were kept overnight at -20°C , and then centrifuged at 10 000 *g* for 15 min. The pellet was dissolved in water. For enzymic assays and protein purification, the total secretion was directly dissolved in the appropriate buffer.

Gel electrophoresis

Samples containing proteins were mixed with sample buffer (1:1, v/v) containing 0.1 M Tris/HCl, pH 6.8, 2.25 % (w/v) SDS, 0.36 mM β -mercaptoethanol, 3 % (v/v) glycerol, 2 mM EDTA, 0.05 % (w/v) Bromophenol Blue and heated for 3 min at 95°C before being loaded on to the gels. Proteins were electrophoresed under denaturing conditions on SDS/10 %-(w/v)-PAGE gels [12] on a Bio-Rad mini-gel apparatus. The gels were fixed, then stained for 1 h in a solution of 50 % (v/v) methanol containing 10 % (v/v) acetic acid and 0.25 % (w/v) Coomassie Blue R250 (Merck) and destained in 30 % (v/v) methanol with 10 % (v/v) acetic acid.

Protein purification

The sternal proteinaceous secretion of 100 males was dissolved in 60 mM potassium phosphate, pH 6.0, and Lma-p72 was purified by either gel filtration or anion-exchange chromatography. Exclusion-diffusion chromatography was performed through a Superdex 75 HR 10/30TM column (Amersham Pharmacia Biotech). The gel was equilibrated with 60 mM potassium phosphate, pH 6.0, and the elution carried out at 0.5 ml/min. Anion-exchange chromatography was performed using a Mini Q PE 4.6/50TM column (Amersham Pharmacia Biotech). After loading the sample, the column was washed with 20 mM BisTris propane buffer, pH 7.2, and the elution was achieved using a linear gradient with the same buffer from 0 to 0.4 M NaCl in 33 min, then from 0.4 to 0.8 M NaCl in 12 min. The flow rate was 0.5 ml/min. In both cases, the absorbance was recorded at 280 nm; the collected fractions were submitted to SDS/PAGE analysis and to enzymic assay.

Protein microsequencing

The sternite (ventral-body-plate) surface proteins were separated on SDS/10 %-(w/v)-PAGE gels [12] and electroeluted to a PVDF membrane [13]. Lma-P72-protein-containing bands were cut out, digested *in situ* with trypsin endopeptidase and the resulting peptides were separated by reverse-phase HPLC on a C₁₈ column (LiChrospher[®], Merck). Fractions containing each peptide were sequenced on an automatic protein sequencer (Applied Biosystems), and the sequences of two peptides were thus determined.

cDNA cloning

Polyadenylated RNAs (6 μg) from sternites were reverse-transcribed using Superscript preamplification system kit (Life Technologies).

Lma-P72 cDNA was amplified by PCR using two degenerate oligodeoxynucleotides [TAYCAYATGTAYGAYGARGA, GCRTTRAANGCRAARTARTC (N = A/T/C/G, Y = C/T and R = A/G)] as forward and reverse primers respectively. These oligonucleotides were deduced from the peptidic sequences YHMYDEE and DYFAFNA respectively.

Each cycle comprised 1 min at 94°C , 1 min at 45°C and 2 min at 72°C ; the last cycle was a 10 min extension step at 72°C . Reaction was performed in 50 μl containing 50 pmol of each primer, 0.2 mM deoxynucleotides and 1 unit of *Taq* polymerase (Life Technologies). The 300 bp fragment was subcloned in a T/A cloning plasmid (pGEM-T easy; Promega) and was sequenced on both strands by the dideoxyribonucleotide method [14]. On the basis of the sequence of this fragment, three specific oligonucleotide primers were synthesized and used for rapid amplification of 5' and 3' rapid amplification of cDNA ends (RACE) [15].

In order to obtain the 5' end, the following protocol (5' RACE system; Life Technologies) was used. The first-strand cDNA was poly(C)-tailed using terminal transferase and was used as template for PCR using the oligo(dG/dI) adaptor primer (GGCCACGCGTCTCGACTAGTACGGGGIIGGGIIGGGIIG) and an internal 5' specific primer (TAGAACAGCAAGTAACGCGC). This product was then used as template for a second nested PCR using an Abridged Universal Amplification Primer (GGCCACGCGTCTCGACTAGTAC) and a second 5' specific primer (CCTTTTGTACTCTCACGAAAC).

The 3' end of the transcript was amplified from first-strand cDNA using an oligo(dT)-adaptor primer [GAATTC(T)₁₈(V)₂ (V = A, C or G)] and a 3' specific primer [TACTGCGGACTAC TTCGCC (N = A/T/C/G, Y = C/T and R = A/G)] as reverse and forward primer respectively in the presence of Superscript.

The 5' and 3' fragments were subcloned in a T/A cloning plasmid and sequenced on both strands.

Northern blotting

Total RNA was extracted from male tergal and sternal glands and from whole tergal glands of females and last nymphal stage males. Total RNA (15 μg) was run on a denaturing formaldehyde-agarose gel and transferred to nylon membrane (Hybond N+; Amersham) then hybridized with a ³²P-labelled cDNA probe in hybridization buffer [6 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate)/50 % (v/v) formamide/0.5 % SDS/100 $\mu\text{g}/\text{ml}$ salmon sperm DNA] at 42°C overnight, washed in 0.2 $\times$ SSC/0.1 % SDS at 42°C and autoradiographed.

In situ hybridization

The 5'-TAGAACAGCAAGTAACGCGC-3' oligonucleotide was end-labelled by T₄ polynucleotide kinase and used as a probe for *in situ* hybridization. Cryosections were deposited on poly-(L-lysine)-coated slides. After three washes in PBS (130 mM NaCl/7 mM NaPO₄/3 mM NaH₂PO₄, pH 7), sections were incubated in 0.2 M HCl for 10 min and acetylated using acetic anhydride, and then dehydrated in increasing concentrations of ethanol [50, 75, 95 and 100 % (v/v)]. Desiccated sections were incubated for 1 h at 25°C in prehybridization buffer [4 $\times$ SSC/1 $\times$ Denhardt's solution (0.02 % Ficoll 400/0.02 % polyvinylpyrrolidone/0.02 % BSA)]. The probe was dissolved in hybridization buffer [40 % (v/v) formamide/10 % (w/v) dextran sulphate/1 $\times$ Denhardt's solution/4 $\times$ SSC/0.5 mg/ml salmon sperm DNA] and deposited on each section for 16 h at 25°C (15 $^{\circ}\text{C}$ under the calculated melting temperature). After six washes (20 min each) in 4 $\times$ SSC, and one wash (45 min) in 2 $\times$ SSC, slides were dried in a desiccator, coated with Ilford K5 emulsion, and exposed for about 1 week at 4°C before observation.

Enzymic assays

Total tergal proteinaceous secretion from males or females (10 individuals) was diluted in water containing 1.8 mM 5-bromo-4-chloroindol-3-yl β -D-galactopyranoside (X-Gal). Proteinase KTM (0.5 mg; 30 units/mg; Sigma) or 10 mM D-gluconolactone were added in some samples. The final reaction mix was incubated for 16 h at room temperature and the absorbance was measured at 420 nm. For enzymic assays with the Lma-p72 fractions purified by chromatography, the reaction was performed in a buffer containing 60 mM potassium phosphate, pH 6.0, and 1.8 mM X-Gal.

In vivo inhibition tests

In the first experiment, 15 day-old-males ($n = 10$) isolated at the imaginal moult were injected with an aqueous solution of 2 M D-gluconolactone (2 μ l). The second experiment involved applying on to the sternite surface, daily for 3 days, a solution of 4 M D-gluconolactone. On the fourth day, volatile compounds were extracted for 2 h in methylene chloride after dissection of the sternites. Other treated males ($n = 10$) were used for quantifying their cuticular hydrocarbons (extraction was for 5 min in pentane). 2-Phenylethanol or tetracosane were added to the extracts as internal standards. After concentration under a stream of nitrogen, aliquots were analysed by GLC using a Chrompack (Middleburg, The Netherlands) CP 437 A gas chromatograph fitted with a split-splitless injector (30 s split, 20 ml/min) and a flame-ionization detector. Injector and detector temperatures were respectively 260 and 280 °C. For both experiments, the column used was a DB 1701TM fused silica capillary column (30 m long $\times$ 0.32 mm internal diameter; Chrompack). Hydrogen was used as carrier gas at a velocity of 25 ml/min at room temperature. The oven temperature was programmed at 40 °C for 2 min, 20 °C/min for 1 min, then 2 °C/min to 240 °C. The signal was recorded and integrated on a computer fitted with Maestro software (Chrompack). Results were compared two-by-two with a Mann–Whitney *U*-test.

RESULTS

Analysis of the secretions

SDS/PAGE analysis showed that abdominal proteinaceous secretions of *L. maderae* contained four major proteins (Figure 1). As previously described, a 54 kDa protein, named Lma-p54, was present on the abdominal surface of both sexes, whereas two other proteins with low apparent molecular masses, respectively Lma-p18 and Lma-p22, were specific to the male tergites [2]. A major protein band with an apparent molecular mass of 72 kDa, named Lma-p72, was also male-specific, but was present in larger amount on sternites than on tergites. None of these four major proteins were observed in nymphs.

Isolation and characterization of Lma-P72 cDNA

A PCR approach coupled to 3' and 5' RACE was developed to isolate the full-length cDNA coding the Lma-P72 surface protein of the male tergites and sternites. Electrophoretic analysis of the selected amplified cDNA on 1%-agarose gel showed a single PCR band of 314 bp.

Specific primers taken from the corresponding sequence were then used for the amplification of the 5' and 3' ends of the corresponding cDNA. The cDNA sequence obtained presents an open reading frame of 1602 nucleotides, followed by a 3'

Figure 1 SDS/PAGE separation of the surface proteins from *L. maderae* abdominal secretions

The protein amount applied to each lane corresponded to three individual equivalents (that is, the quantity of secretion corresponding to three males, three females or three nymphs). Surface proteins were wiped from male tergites (lane A) and sternites (lane B), female tergites (lane C) and sternites (lane D), nymphal tergites (lane E) and sternites (lane F). MM, molecular mass standards.

untranslated sequence of 138 nucleotides. Only five nucleotides from the 5' untranslated sequence were obtained. The deduced protein is composed of 534 amino acids with a calculated molecular mass of 61.3 kDa and a pI of 5.61. The difference between the molecular mass observed on SDS/PAGE and deduced from the amino acid sequence of Lma-p72 may be due to post-translational modifications of the protein; Lma-p72 may also be glycosylated or phosphorylated. The sequences obtained after Edman degradation of the protein extracted from the male sternal secretion were found from positions 93 to 95, from 240 to 249 and from 338 to 346. The hydrophobicity profile [16] revealed that the first 18 amino acids are hydrophobic and that they probably correspond to a signal peptide with a putative cleavage site [17] between residues 18 (cysteine) and 19 (leucine). This signal-peptide prediction confirms that the first methionine residue of the peptide sequence should correspond to the initiation methionine residue.

When submitted to the BLAST program [18], Lma-p72 protein showed significant sequence similarity with family 1 glycosyl hydrolases (EC 3.2.1.-). The percentages of identity comprised between 39.5 and 44.6 % with a β -glycosidase from the midgut of the coleopteran *Tenebrio molitor* (the yellow mealworm) [19], with intestinal lactase-phlorizin hydrolases (LPHs) from rat [20] and from human [21] and with a human cytosolic β -glucosidase (GenBank[®] accession number AF317840). The patterns defining the family 1 of glycosyl hydrolases in the N-terminal region and in the C-terminal region (around the active-site glutamate residue) are present in the Lma-p72 sequence. Secondary-structure prediction shows that Lma-p72 would contain 13 β -strands and 13 α -helices [22]. Tertiary-structure prediction with the program 3D-PSSM [23] concludes that Lma-p72 presents a ($\beta\alpha$)₈ or TIM barrel folding, which is characteristic of the family 1 glycosyl hydrolases.

RNA expression profile

The first obtained 314 bp cDNA was used as a probe to check the male-specificity by Northern-blot analysis. As shown in Figure 2, a single mRNA of approx. 1900 nucleotides was detected in adult male tergite and sternite extracts, while no signal was observed

Figure 2 Northern hybridization analysis of *L. maderae* RNA with Lma-p18 cDNA

A Northern blot was performed with total RNA (15 µg) from adult male tergites (a), sternites (b) and midguts (c), nymphal male tergites (d), sternites (e) and midguts (f), adult female tergites (g), sternites (h) and midguts (i) and female nymph tergites (j), sternites (k) and midguts (l). Ethidium bromide-stained rRNAs indicate that approximately equal amounts of RNA were loaded.

in midguts. Lma-p72 transcript was not expressed in nymphs and females, as expected from SDS/PAGE analysis. Compared with SDS/PAGE analysis, the expression of Lma-p72 by Northern blotting seems to be higher in sternites than in tergites. This may be due to differing determination of the sample concentration, which was made as individual equivalents for SDS/PAGE and in

mg of total RNA for Northern blotting. So the concentration of the Lma-p72 transcript among total RNA would seem to be the same in both tergites and sternites, while the general amount of expression would be higher in sternites.

Histological localization

In situ hybridization analyses were undertaken on tergite, sternite and midgut sections. Lma-p72 transcript was expressed faintly in the whole length of the glandular epithelium of the male second tergite (Figure 3A). However, in sagittal sections, labelling was more accentuated in the glandular region underlying the knob-like structure which is nibbled by females during mating behaviour (Figure 3B). Labelling of the well-developed sternal glandular epithelium was uniform in depth (Figure 3C), whereas the whole gland was strongly labelled (Figure 3D). Lma-p72 is specific to adult males, and the transcript was not expressed in the non-glandular epidermal epithelium from nymphal tergites (Figure 3E). As expected from Northern-blot results, no labelling was observed in the intestinal epithelium at the level of midgut (Figure 3F).

Figure 3 *In situ* hybridization analysis with Lma-p72 radioactive probe

Cryosections from various tissues of *L. maderae* hybridized with the γ - 32 P-labelled Lma-p72 oligonucleotide probe. (A) Sagittal section of the second tergite of adult male. The glandular area ('gl') is labelled (arrows) especially under the cuticular knob-like structure ('kn'). (B) Enlargement of the region under the knob-like structure of the same tergite. (C) Enlargement of the anterior region of an adult male sternite showing a stronger labelling of the gland ('gl'). The canaliculi, from which the secretion flows outside, are seen in the cuticle ('cu'). (D) Sagittal section of an adult male sternite. The glandular epithelium ('gl') is entirely and uniformly labelled. (E) Sagittal section of a nymph second tergite. The non-glandular epidermis ('ep') is not labelled. (F) Cross-section of male midgut. There is no expression of Lma-p72 transcript in the intestinal epithelium ('ie').

Figure 4 β -Galactosidase activity in *L. maderae* abdominal surface secretions

Hydrolysis of X-Gal was tested in male tergal secretion alone (M), with addition of 15 units of Proteinase K (+K) or 10 mM D-gluconolactone (+G) and in female tergal secretion (F). A control contained only male tergal secretion (C). Experiments were repeated ten times and the average absorbances at 420 nm (means $\pm$ S.E.M.) are shown.

β -Galactosidase activity of Lma-p72

In order to verify whether Lma-p72 was an active glycosyl hydrolase, we first used X-Gal as a chromogenic substrate to verify the presence of a β -galactosidase activity in the whole proteinaceous tergal secretions. Our data showed that tergal and sternal secretions from males cleaved X-Gal efficiently (Figure 4). The β -galactosidase activity was fully inhibited by the addition of Proteinase K in the reaction mixture, and the measured absorbance at 420 nm diminished to a level similar to that of the control containing no X-Gal. These observations indicate that the β -galactosidase activity is due to an enzyme. The β -galactosidase activity in male tergal secretions was also fully inhibited by the addition of D-gluconolactone, which is a general inhibitor of glucosidases. Regarding these results, we can conclude that the enzyme responsible for the β -galactosidase activity is also able to bind β -galactosides and β -glucosides. The lack of activity in female tergal secretions is congruent with the general expression pattern of Lma-p72. Consequently, Lma-p72 has been purified from sternal secretions by chromatography in order to verify whether Lma-p72 was actually responsible for the β -galactosidase activity in male abdominal secretions. Anion-exchange chromatography allowed a good separation of the two major proteins of the sternal secretions: Lma-p54 and Lma-p72. The analysis of the fractions obtained revealed that the β -galactosidase activity was eluted along with the Lma-p72 protein.

These results confirm that Lma-p72 is an active family 1 glycosyl hydrolase that exhibits a β -galactosidase-type activity. Nevertheless, Lma-p72 is fully inhibited by the D-gluconolactone and should present a low specificity with respect to the C-4 configuration of the substrate.

In vivo inhibition of Lma-p72

Adult males of *L. maderae* were treated with D-gluconolactone in order to evaluate the potential effect of Lma-p72 inhibition on the production of pheromonal or cuticular hydrocarbon components. Two kinds of treatment were used: D-gluconolactone was either injected into the abdomen or topically applied on to the surface of the sternites.

Figure 5 In vivo effect of D-gluconolactone on the production of pheromonal compounds (A) and cuticular hydrocarbons (B) in male *L. maderae*

The amount of each compound was quantified by GLC in control (black bars) and treated insects after injections (white bars) or topical application (grey bars) of D-gluconolactone. Abbreviations: 3H2B, 3-hydroxybutan-2-one; R,R BUT, (2R,3R)-butanediol; SENEC. Ac., senecioic acid; (E)-2-OCT. Ac., (E)-oct-2-enoic acid; H1, n-heptacosane; H2, 11- and 13-methylheptacosane; H3, 3-methylheptacosane. Experiments were repeated ten times and the average amount (ng) of each compound (mean $\pm$ S.E.M.) is represented. * P < 0.05; ** P < 0.01.

The male sexual pheromone of *L. maderae* contained four main volatile compounds: 3-hydroxybutan-2-one, (2R,3R)-butanediol, senecioic acid ($\beta\beta$ -dimethylacrylic acid) and (E)-oct-2-enoic acid (J.-P. Farine, unpublished work). After both D-gluconolactone treatments, the production of each compound was appreciably inhibited (Figure 5A). The decrease in production only of 3-hydroxybutan-2-one was not significant; the high inter-individual variability observed for this compound could explain these results. The inhibition of the synthesis of senecioic acid was significant only after topical applications. Consequently, D-gluconolactone seems to inhibit the production of at least three of the four major pheromonal compounds. Nevertheless, D-gluconolactone is known to be a general inhibitor of glucosidases and, thus, may inhibit not only Lma-p72, but also other glucosidases, which are important for the metabolism of the insects.

Four main cuticular hydrocarbons are found in *L. maderae*: n-heptacosane, a blend of 11- and 13-methylheptacosane, and 3-methylheptacosane (J.-P. Farine, unpublished work). No significant difference was observed for any of these hydrocarbons between the control and the treated insects after D-gluconolactone treatments (Figure 5B). So, if there is any physiological perturbation due to D-gluconolactone, this did not affect the production of hydrophobic compounds on the abdominal surface.

Consequently, a glycosidase should be directly involved in the production pathway of the pheromonal compounds.

DISCUSSION

We report here the molecular characterization of Lma-p72, a new male-specific cockroach surface protein. The sequence analysis revealed that Lma-p72 shares similarity to the family 1 glycosyl hydrolases. Within the large family of glycosyl hydrolases [24], the family 1 glycosyl hydrolases is part of a wider group of glycosidases named 'clan GH-A', which comprises more than 250 members with 18 different substrate specificities [25]. This family contain the β -glucosidases and the β -galactosidases [26–30]. Glycosidases are known to cleave glycosidic bonds, releasing the terminal aglycone moiety of various disaccharides or oligosaccharides. Family 1 glycosyl hydrolases generally present a low substrate specificity, especially in regard to the C-4 configuration, which means that they are active on both β -glucosides and β -galactosides [31]. Our results reveal that it is certainly the case that Lma-p72 exhibits a β -galactosidase activity inhibited by a glucosidase inhibitor (D-gluconolactone). Glycosidic-bond hydrolysis by β -glycosidases is mediated by the two glutamic acid residues that confer the active site. Those two glutamic acid residues were identified as nucleophile and general acid–base respectively [27,32,33] and are located in two conserved sequence patterns: NEP and ITENG (one-letter amino acid code) [34]. In the Lma-p72 peptide sequence, these glutamic acid residues were found at the positions 202 and 414, along with the conserved patterns.

On the basis of their stereochemical behaviour and catalytic mechanism, glycosidases were separated in 'retaining glycosidases' and 'inverting glycosidases' [35]. In both types, the active site contains a carboxy group (corresponding to Glu²⁰² in the Lma-p72 numbering) and a carboxylate group (corresponding to Glu⁴¹⁴ in the Lma-p72 numbering), which are always conserved in each glycosidase family. In the inverting glycosidases, these residues, which show an average distance of 9.3 Å (0.93 nm), function as an acid or a basic catalyst respectively and operate with a single displacement of the leaving group. By contrast, retaining glycosidases, in which the two residues are closer, follow a two-step mechanism with formation of a covalent glycosyl-enzyme intermediate [36]. The carboxy group functions as a general acid–base catalyst and the carboxylate as the nucleophile of the reaction [37,38]. All the family 1 glycosyl hydrolases are retaining glycosidases and the two glutamic acids involved in the catalytic process are located just after two β -sheets [39].

The active site of β -glycosidase may be made of several monosaccharide residue-binding subsites. The glycosidic linkage to be cleaved corresponds to the linkage between –1 (glycone) and +1 (aglycone) subsite [19,40]. The binding of the monosaccharide constituting the glycone moiety with the –1 subsite is realized by interactions with strongly conserved amino acid residues. Conserved His¹⁵⁷ and Asn²⁰¹ residues (Lma-p72 numbering) were suggested to interact with the C-2 and C-3 hydroxy groups, whereas Gln⁵⁴ and Glu⁴⁶³ would form hydrogen bonds with the C-4 hydroxy group. Additionally, Trp⁴⁵⁶ would form a hydrophobic platform interacting with the C-6 hydroxymethyl group. As this hydrophobic platform should interact similarly with different monosaccharides, the specificity of the –1 subsite must rely on the spatial distribution of the polar groups (His¹⁵⁷, Asn²⁰¹, Gln⁵⁴ and Glu⁴⁶³) inside that subsite [39]. Subsites +1 and following would form a cleft containing mainly either hydroxy groups or hydrophobic groups, and would consequently bind preferentially

either a hydrophilic oligosaccharide or a hydrophobic aglycone portion respectively [19,39].

In *L. maderae*, Lma-p72 is specific to adult males and the expression of its transcript was confirmed by Northern blotting and *in situ* hybridization in tergal and sternal glands. The lack of expression of the protein in the midgut was quite surprising compared with the most similar intestinal enzymes (*T. molitor* β -glycosidase and mammalian LPHs [19–21]). LPHs are major enzymes in the small intestine of young suckling mammals and are involved in the digestion of lactose and other glycosidic substrates from the milk [41–48]. The Lma-p72 active site containing the nucleophile glutamic acid shows a high identity not only with the active site of the domain IV in rat LPH, but also with the active site of the domain III in human LPH. Consequently, at this stage of knowledge, and only on the basis of sequence similarity, a preferential substrate for Lma-p72 cannot be proposed definitely (e.g. lactase-like activity or phlorizin hydrolase-like activity). Furthermore, the digestive β -glycosidase from *T. molitor*, which is the most similar to Lma-p72, shows weak substrate specificity and is able to hydrolyse disaccharides, the polymer laminarin and alkyl glucosides as well [19]. Such weak substrate specificity is common in digestive enzymes. By contrast, Lma-p72, which is specific to males and expressed only in the abdominal glands, should have a more specific substrate specificity.

In comparison with the PH site of the LPH [41,50,51], Lma-p72 could cleave male-specific molecules associated with sugars. Such sugars were recently identified in the tergal secretions of the German cockroach, *Blattella germanica* [49]. Compounds from the sexual pheromone or cuticular hydrocarbons may be such aglycones, the sugar-associated component constituting a transport form or a precursor. According to this hypothesis, our data showed that Lma-p72 occurs in larger amounts on the sternites, and *in situ* hybridization showed that Lma-p72 transcript is expressed into the glandular epithelium, which is also responsible for the production of the sex pheromone. Furthermore, our *in vivo* tests demonstrated a significant inhibition for three of the pheromonal compounds. These results are congruent with the hypothesis of the existence of a glycosidase involved in the production pathways of pheromone components in *L. maderae*, and Lma-p72 could be a good candidate for this function. Further experiments using more specific inhibitors will be needed to determine the exact role of this protein.

This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Université de Bourgogne. We are grateful to H. Bouhin, J. P. Charles and A. Quennedey for helpful discussions. We also thank M. Bentejac (Laboratoire de Biologie Moléculaire et Cellulaire, INRA, Dijon, France) for his help concerning the purification of Lma-p72, and C. Cambillau (CNRS, UMR 6098, Université de Provence, Marseille, France) for his constructive comments.

REFERENCES

- 1 Zupko, K., Sklan, D. and Lensky, Y. (1993) Proteins of the honeybee (*Apis mellifera* L.) body surface and exocrine gland secretions. *J. Insect Physiol.* **39**, 41–46.
- 2 Korchi, A., Farine, J. P. and Brossut, R. (1998) Characterization of two male-specific polypeptides in the tergal glands secretions of the cockroach *Leucophaea maderae* (Dictyoptera, Blaberidae). *Insect Biochem. Mol. Biol.* **28**, 113–120.
- 3 Cornette, R., Farine, J. P., Quennedey, B. and Brossut, R. (2001) Molecular characterization of a new adult male putative calycin specific to tergal aphrodisiac secretion in the cockroach *Leucophaea maderae*. *FEBS Lett.* **507**, 313–317.
- 4 Abed, D., Farine, J. P., François, J. and Brossut, R. (1995) Protein and amino acid composition of the tergal gland secretions of *Blatta orientalis* and *Eurycotis floridana* (Dictyoptera: Blattellidae). *Chemoecology* **1**, 19–25.
- 5 Cornette, R., Farine, J. P., Quennedey, B., Rivière, S., and Brossut, R. (2002) Molecular characterization of Lma-p54, a new epicuticular surface protein in the cockroach *Leucophaea maderae* (Dictyoptera, Oxyhaloinae). *Insect Biochem. Mol. Biol.* **32**, 1635–1642.

- 6 Roth, L. M. (1969) The evolution of male tergal glands in the Blattaria. *Ann. Entomol. Soc. Am.* **62**, 176–208
- 7 Brossut, R. and Roth, L. M. (1977). Tergal modifications associated with abdominal glandular cells in the Blattaria. *J. Morphol.* **151**, 259–298
- 8 Sreng, L. (1984) Morphology of the sternal and tergal glands producing the sexual pheromones and the aphrodisiacs among the cockroaches of the subfamily Oxysaloinae. *J. Morphol.* **182**, 279–294
- 9 Korchi, A., Brossut, R., Bouhin, H. and Delachambre, J. (1999) cDNA cloning of an adult male putative lipocalin specific to tergal gland aphrodisiac secretion in an insect (*Leucophaea maderae*). *FEBS Lett.* **449**, 125–128
- 10 Flower, D. R. (1996) The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14
- 11 Cavaggioni, A. and Muscignat-Caretta, C. (2000) Major urinary proteins, α_{2U} -globulins and aphrodisin. *Biochim. Biophys. Acta* **1482**, 218–228
- 12 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685
- 13 Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- 14 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- 15 Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
- 16 Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132
- 17 Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
- 18 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
- 19 Ferreira, A. H., Marana, S. R., Terra, W. R. and Clelia, F. (2001) Purification, molecular cloning, and properties of a β -glycosidase isolated from midgut lumen of *Tenebrio molitor* (Coleoptera) larvae. *Insect Biochem. Mol. Biol.* **31**, 1065–1076
- 20 Duluc, I., Boukamel, R., Mantei, N., Semenza, G., Raul, F. and Freund, J. N. (1991) Sequence of the precursor of intestinal lactase-phlorizin hydrolase from fetal rat. *Gene* **103**, 275–276
- 21 Harvey, C. B., Fox, M. F., Jeggo, P. A., Mantei, N., Povey, S. and Swallow, D. M. (1993) Regional localization of the lactase-phlorizin hydrolase gene, LCT, to chromosome 2q21. *Ann. Hum. Genet.* **57**, 179–185
- 22 Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893
- 23 Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520
- 24 Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788
- 25 Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696
- 26 Gonzalez-Candelas, L., Ramon, D. and Polaina, J. (1990) Sequences and homology analysis of two genes encoding β -glucosidases from *Bacillus polymyxa*. *Gene* **95**, 31–38
- 27 Withers, S. G., Warren, A. J., Street, I. P., Rupitz, K., Kempton, J. B. and Aebersold, R. (1990) Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a glycosidase. *J. Am. Chem. Soc.* **112**, 5887–5889
- 28 Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316
- 29 Henrissat, B. (1991) Sequence homology between a β -galactosidase and some β -glucosidases. *Protein Sequences Data Anal.* **4**, 61–62
- 30 El Hassouni, M., Henrissat, B., Chippaux, M. and Barras, F. (1992) Nucleotide sequences of the *arb* genes, which control β -glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli* bgl operon and evidence for a new β -glycohydrolase family including enzymes from eubacteria, archaeobacteria, and humans. *J. Bacteriol.* **174**, 765–777
- 31 Gräbnitz, F., Seiss, M., Rucknagel, K. P. and Staudenbauer, W. L. (1991) Structure of the β -glucosidase gene bglA of *Clostridium thermocellum*. Sequence analysis reveals a superfamily of cellulases and β -glycosidases including human lactase/phlorizin hydrolase. *Eur. J. Biochem.* **200**, 301–309
- 32 Trimbur, D. E., Warren, R. A. and Withers, S. G. (1992) Region-directed mutagenesis of residues surrounding the active site nucleophile in β -glucosidase from *Agrobacterium faecalis*. *J. Biol. Chem.* **267**, 10248–10251
- 33 Wang, Q., Trimbur, D., Graham, R., Warren, R. A. and Withers, S. G. (1995) Identification of the acid/base catalyst in *Agrobacterium faecalis* β -glucosidase by kinetic analysis of mutants. *Biochemistry* **34**, 14554–14562
- 34 Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausin, C., Querol, E., Planas, A. and Perez-Pons, J. A. (2001) Mechanism of the family 1 β -glucosidase from *Streptomyces* sp: catalytic residues and kinetic studies. *Biochemistry* **40**, 5975–5982
- 35 Hehre, E. J. (2000) A fresh understanding of the stereochemical behavior of glycosylases: structural distinction of "inverting" (2-MCO-TYPE) versus "retaining" (1-MCO-TYPE) enzymes. *Adv. Carbohydr. Chem. Biochem.* **55**, 265–310
- 36 Moracci, M., Trincone, A., Perugini, G., Ciaramella, M. and Rossi, M. (1998) Restoration of the activity of active-site mutants of the hyperthermophilic β -glucosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* **37**, 17262–17270
- 37 Koshland, D. E. J. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* **28**, 416–436
- 38 White, A. and Rose, D. R. (1997) Mechanism of catalysis by retaining β -glycosyl hydrolases. *Curr. Opin. Struct. Biol.* **7**, 645–651
- 39 Marana, S. R., Jacobs-Lorena, M., Terra, W. R. and Ferreira, C. (2001) Amino acid residues involved in substrate binding and catalysis in an insect digestive β -glycosidase. *Biochim. Biophys. Acta* **1545**, 41–52
- 40 Davies, G. J., Ducros, V., Lewis, R. J., Borchert, T. V. and Schulein, M. (1997) Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. *J. Biotechnol.* **57**, 91–100
- 41 Leese, H. J. and Semenza, G. (1973) On the identity between the small intestinal enzymes phlorizin hydrolase and glycosylceramidase. *J. Biol. Chem.* **248**, 8170–8173
- 42 Lorenz-Meyer, H., Blum, A. and Semenza, G. (1972) Lactase phlorizin-hydrolase complex of the human small intestine. *Verh. Dtsch. Ges. Inn. Med.* **78**, 1413–1418
- 43 Colombo, V., Lorenz-Meyer, H. and Semenza, G. (1973) Small intestinal phlorizin hydrolase: the " β -glycosidase complex". *Biochim. Biophys. Acta* **327**, 412–424
- 44 Birkenmeier, E. and Alpers, D. H. (1974) Enzymatic properties of rat lactase-phlorizin hydrolase. *Biochim. Biophys. Acta* **350**, 100–112
- 45 Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., Hunziker, W. and Semenza, G. (1988) Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *EMBO J.* **7**, 2705–2713
- 46 Wacker, H., Keller, P., Falchetto, R., Legler, G. and Semenza, G. (1992) Location of the two catalytic sites in intestinal lactase-phlorizin hydrolase. Comparison with sucrase-isomaltase and with other glycosidases, the membrane anchor of lactase-phlorizin hydrolase. *J. Biol. Chem.* **267**, 18744–18752
- 47 Zecca, L., Mesonero, J. E., Stutz, A., Poiree, J. C., Giudicelli, J., Cursio, R., Gloor, S. M. and Semenza, G. (1998) Intestinal lactase-phlorizin hydrolase (LPH): the two catalytic sites; the role of the pancreas in pro-LPH maturation. *FEBS Lett.* **435**, 225–228
- 48 Arribas, J. C., Herrero, A. G., Martin-Lomas, M., Canada, F. J., He, S. and Withers, S. G. (2000) Differential mechanism-based labeling and unequivocal activity assignment of the two active sites of intestinal lactase/phlorizin hydrolase. *Eur. J. Biochem.* **267**, 6996–7005
- 49 Nojima, S., Nishida, R., Kuwahara, Y. and Sakuma, M. (1999) Nuptial feeding stimulants: a male courtship pheromone of german cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae). *Naturwissenschaften* **86**, 193–196
- 50 Day, A. J., DuPont, M. S., Ridley, S., Rhodes, M., Rhodes, M. J., Morgan, M. R. and Williamson, G. (1998) Deglycosylation of flavonoid and isoflavonoid glycosides by human small intestine and liver β -glucosidase activity. *FEBS Lett.* **436**, 71–75
- 51 Day, A. J., Canada, F. J., Diaz, J. C., Kroon, P. A., McLauchlan, R., Faulds, C. B., Plumb, G. W., Morgan, M. R. and Williamson, G. (2000) Dietary flavonoid and isoflavone glycosides are hydrolysed by the lactase site of lactase phlorizin hydrolase. *FEBS Lett.* **468**, 166–170