

cDNA cloning of an adult male putative lipocalin specific to tergal gland aphrodisiac secretion in an insect (*Leucophaea maderae*)

Ahmed Korchi¹, Rémy Brossut, Hervé Bouhin, Jean Delachambre*

Développement-Communication chimique, Unité Mixte de Recherche 5548 du Centre National de la Recherche Scientifique, Université de Bourgogne, 6 Bd Gabriel, 21 000 Dijon, France

Received 13 February 1999; received in revised form 15 March 1999

Abstract Lma-P22 is a cuticular surface protein specific to the tergal gland secretion of *Leucophaea maderae* adult males which is ingested by females just before copulation. The complete Lma-P22 cDNA sequence was determined by RT-PCR using primers based on Edman degradation fragments. The recombinant protein expressed in *Escherichia coli* was recognized by an anti-Lma-P22 antibody. Northern blot analysis indicates that the corresponding mRNA is transcribed only in the epidermis of male tergites. Sequence analysis indicated that Lma-P22 deduced protein belongs to the lipocalin family. Lipocalins are extracellular proteins which carry hydrophobic compounds and some of them can bind sexual pheromone in vertebrates. Lma-P22 is the first example of a lipocalin-like protein involved in insect sexual behavior.

© 1999 Federation of European Biochemical Societies.

Key words: Cockroach; Cuticle surface protein; Lipocalin; Sex behavior; Tergal gland

1. Introduction

Insects are remarkable chemists and present numerous epidermal glands which synthesize a large array of molecules involved in chemical communication. Most of them are volatile and play a role in the attraction of sexual partners, because they can be perceived relatively far away. Many studies have been devoted to these compounds because of the important role that they play in chemical communication. Heavier products, such as proteins, can also be secreted onto the cuticular surface, either on the whole body [1–3] or at very localized spots corresponding to specific glandular regions [3–5]. However, little attention has been paid to their possible role as contact signals and none of these proteins is as yet known to play a pheromonal role.

In cockroaches, the tergal protein secretion carried by adult males has been described as playing an aphrodisiac influence on the female [6]. During courtship, once the two sexual partners are close to each other, the male exposes his tergal glands and the female mounts on his back to ingest his aphrodisiac secretion [7]. This ‘feeding’ behavior is a ritual step which makes it possible to immobilize the female in a favorable position for copulation. In a previous work [3], the analysis of the *Leucophaea maderae* tergal gland secretion revealed that one protein (Lma-P22) was specific to the male tergite secretion eaten by the female. Such sexual specificity suggested

that Lma-P22 could be involved in attracting the female. Therefore, its molecular characterization was undertaken in order to find a relation between its structure and its putative function during cockroach courtship ritual.

This paper reports the complete nucleotide sequence of the cDNA encoding Lma-P22. The deduced amino acid sequence strongly suggests that Lma-P22 is a putative member of the lipocalin family. The fact that several vertebrate lipocalins, such as mouse urinary proteins and hamster aphrodisin, can bind odorous ligands in order to act as sex attractants suggests that Lma-P22 plays an analogous function during *L. maderae* courtship behavior.

2. Materials and methods

2.1. Animals

The conditions for rearing insects and the methods to analyze protein secretion have been previously described [3].

2.2. Protein microsequencing

Surface proteins from the second tergites were separated on 12% SDS-PAGE [8] and electroeluted to a PVDF membrane [9]. Lma-P22 protein containing bands were cut out and digested in situ by trypsin endopeptidase. The resulting peptides were separated by reverse phase high-performance liquid chromatography on a C18 column (Lichrospher, Merck). The fractions containing each peptide were lyophilized, sequenced by an automatic protein sequencer (Applied Biosystems) and the sequence of five peptides determined.

2.3. cDNA cloning

Poly(A)⁺ RNA, extracted from the second tergite of 5 day old adults, was reverse transcribed using the Superscript preamplification system (Life Technologies). The obtained cDNA was amplified by PCR using two degenerate oligodeoxynucleotides (GCRTART-GYTCNAGRTTCTC, CTCCAYTCNGGYAAATG, N = A/T/C/G, Y = C/T and R = A/G) deduced from the peptide sequences as forward and reverse primers, respectively. These oligonucleotides correspond to the peptide sequences ENLEHYA and QLEPWS. Each cycle comprised 1 min at 94°C, 1 min at 46°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 200 bp fragment was subcloned in a T/A cloning plasmid (pMOSBlue, Amersham) and was sequenced on both strands [10] by double stranded sequencing using 2.0 Sequenase (USB). Based on the sequence of this fragment, specific oligonucleotide primers were synthesized and used for rapid amplification of 3' and 5' cDNA ends (RACE) [11].

In order to obtain the 5' end, the following protocol was used. Poly(A)⁺ RNAs (2 µg) from the second tergite were reverse transcribed using a specific oligonucleotide (as a reverse primer) in the presence of Superscript reverse transcriptase (Life Technologies). The first strand cDNA was then poly(C) tailed using terminal transferase and the double strand was generated by extension of an oligo(dG/dI) adapter primer (GGCCACGCGTCGACTAGTACGGGIIIGGGIIG-GGIIG). This product was used as a template for PCR using the adapter primer (GGCCACGCGTCGACTAGTA) and a 5' specific primer (CTTACTGAACATGACGACCG).

The 3' end of the transcript was amplified by reverse transcription poly(A)⁺ RNA (2 µg) using a oligo(dT) adapter primer (GAATT-

*Corresponding author. Fax: (33) 3 80 39 62 89.
E-mail: jdelach@u-bourgogne.fr

¹ Present address: Institut für Physiologie, Universität Hohenheim (230), 70593 Stuttgart, Germany.

C(T)18(V)2, V = A, C or G) as the reverse primer, in the presence of Superscript. This product was then used as template for PCR amplification using the adapter primer (CCTCTGAAGGTTCCAGAATC-GATAG) and a specific primer (CGGTCGTCATGTTTCAGTAA-GAAATGTCC) as the forward and reverse primers, respectively. A second nested PCR was performed using the adapter primer and a second specific primer (GGTGAGCAAGAGCAGCCC). The 5' and 3' fragments were subcloned in a T/A cloning plasmid and sequenced on both strands. The CLUSTAL program [12] was used for multiple sequence alignment.

2.4. Northern blotting

Total RNA was extracted from male tergal glands, and from tergal glands of females and of last stage larvae of both sexes. Corresponding poly(A)⁺ RNA (1 µg) were allowed to run on a denaturing formaldehyde agarose gel, then transferred to nylon membrane (Hybond N⁺, Amersham) and hybridized overnight at 42°C with a ³²P-labelled cDNA probe in the hybridization buffer (SSC 6×, formamide 40%, Denhardt 5×, dextran sulfate 10% and salmon sperm DNA 100 µg/ml) before being washed in 0.2×SSC, 0.1% SDS at 65°C and autoradiographed.

2.5. Western blotting

Western blot analysis of recombinant Lma-P22 protein was performed as previously described [3].

2.6. Expression of Lma-P22 cDNA in *Escherichia coli*

The coding sequence (without signal peptide) was obtained by a series of three PCRs using oligonucleotides specific to 5' and 3' ends. This construct was made in pQE-30 expression vector (Qiagen) which contained a histidine tagged region upstream of the cloning site. The His-tagged protein was expressed in M15 *E. coli* and purified on Ni-NTA columns according to the manufacturer's instructions.

3. Results

3.1. Isolation and characterization of the Lma-P22 cDNA

A PCR approach coupled to 5' and 3' RACE was used to isolate the full-length cDNA encoding the Lma-P22 surface protein present on the second tergite of the male [3]. Electrophoretic analysis of the selected amplified cDNA on 1% agarose gel showed a single PCR band of 212 bp. This product was used as a probe to check the male specificity by Northern blot analysis. As shown in Fig. 1, a single mRNA of approximately 800 nucleotides was detected only in adult male tergite extracts, while no signal was observed in control tergites from females and from larvae of both sexes.

Specific primers deduced from the 212 bp sequence were

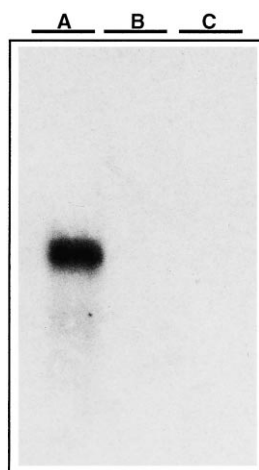


Fig. 1. Northern blot analysis of epidermal gland mRNAs from male second tergites (A), female (B), and larval (C) tergites (1–7) hybridized with the 212 bp PCR fragment.

1	TAAAGAACTCAGATATAATAA
23	CTGTATTACTATTCATATTTTACGGCGCAGTGTGATAAGTTTCAACGCCAGAAAAA
82	ATG AAT TCA CTG GCT GGC TCA CTG GCA TGT CTT GCT GTT GGC ATT
81	Met Asn Ser Leu Ala Gly Ser Leu Ala Cys Leu Ala Val Ala Ile
127	TTA AAT GTC AAC GGC GAC TGC GAC TTT GGA CCT ATA TTT CGA TTT
126	Leu Asn Val Asn Gly Asp Cys Asp Phe Gly Pro Ile Phe Arg Phe
172	CAT CCA AAT TGG ATG GAA AAT ACA TGG TAT GTT GTC TAT TCA TCC
31	His Pro Asn Trp Met Glu Asn Thr Trp Tyr Val Val Tyr Ser Ser
217	CCT TCT GCC TTC GAT GAG GCT AAC AAT GTT AGC TTT TCA TAT GAG
46	Pro Ser Ala Phe Asp Glu Ala Asn Asn Val Ser Phe Ser Tyr Glu
262	CTT CAA GGA CAC TCG CAC TAT GTG GCA CAC GTT GCA GTT TTT
61	Leu Gln Gly His Ser His Tyr Val Ala His Val Gly Ile Thr Phe
307	GCA GAC ATA GGC ACA GCT AAA GAA ATT AAT GGT ACA GTG ACT GCA
76	Ala Asp Ile Gly Thr Ala Lys Glu Ile Asn Gly Thr Val Thr Ala
352	CTA GAC TTT GGC ACC AAG TTC AAC GCA TGT CTT CCT GAA TGG AGT
91	Leu Asp Phe Gly Thr Lys Phe Asn Val Gln Leu Pro Glu Trp Ser
397	AAA TAC AGT GGA ACC TAC CGA GTA ACA GCA CTA GAG TAT GGG AAT
106	Lys Tyr Ser Gly Thr Tyr Arg Val Thr Ala Leu Glu Tyr Gly Asn
442	TAT TTA ATT GCA AAA GGT TGC CCT GAA CAA CTT ACT GTT AAA TCT
121	Tyr Leu Ile Ala Lys Gly Cys Pro Glu Lys Ser Thr Val Lys Ser
487	TTT ACG GTC GTC ATG TTC AGT AAG AAA TGT CCT GAT GAA GCG TCG
136	Phe Thr Val Val Met Phe Ser Lys Lys Cys Pro Asp Glu Ala Ser
532	GTA GGT GCA GCA AGA GCA GCC CTT AAG AAA TAC CTG AAT GAG AAT
151	Val Gly Ala Ala Arg Ala Ala Leu Lys Lys Tyr Leu Asn Glu Asn
577	CTT GAA CAT TAT GCC AAG GAT ACT TTT CTG AAT TGT CCT TAA AAT
166	Leu Glu His Tyr Ala Lys Asp Thr Phe Leu Asn Cys Pro
622	TGAGTGAAATATGAATGTATATCAAGATTATACACGACGCCGATATATACACT
681	ATTAACAGATAAATACACACACACACACGACATATATATAATTATATATATG
740	AATCTTGAATATGTATGTGAACATGAAAAAGTAAATAATGTATAACATTGTCTAAC
799	AGTAAAAA

Fig. 2. Full-length nucleotide and deduced amino acid sequences of Lma-P22. The arrow corresponds to the putative cleavage site of the signal peptide. The putative polyadenylation site is in bold. Amino acid sequences obtained after Edman degradation are underlined. (*, putative N-glycosylation sites).

then used to amplify the 5' and 3' ends of the corresponding cDNA. Its complete sequence (EMBL accession number: AJ223986) presents an open reading frame of 534 nucleotides, comprised between 5' and 3' untranslated sequences of 81 and 202 nucleotides (Fig. 2). The deduced protein is composed of 178 amino acids with a calculated molecular weight of 19.7 kDa and an isoelectric point (pI) of 5.93. The sequences obtained after Edman degradation of the protein extracted from the male tergal secretion were found at positions 83–91 and 97–119. The hydrophobicity profile [13] revealed that the first 20 amino acids are hydrophobic and that they probably correspond to a signal peptide with a putative cleavage site [14] between residues 20 (glycine) and 21 (aspartic acid).

3.2. Expression of recombinant Lma-P22 in bacteria

In order to confirm that this sequenced cDNA corresponds to the Lma-P22 protein detected by antibodies raised against the natural protein, the expression of this cDNA in bacteria and the identification of recombinant protein by immunolabelling were carried out. The cDNA coding region without the signal peptide was amplified using three successive PCRs. The products of each PCR were used as matrix for following amplification. The 555 bp cDNA was cloned in pQE-30 expression plasmid and expressed in *E. coli* M15 after IPTG induction. The recombinant colonies grown in the absence of IPTG (Fig. 3A) were used as control. After a 5 h incubation, the expressed products were subjected to SDS-PAGE, before or after purification on Ni-NTA columns, then analyzed by Western blotting. The electrophoretic profiles showed a main 20 kDa band specific to IPTG-induced culture extracts (Fig. 3). The same band was retained by Ni-NTA columns and recognized by the anti-Lma-P22 antibody (Fig. 3B).

These results confirm that the cloned cDNA encodes the Lma-P22 protein.

3.3. Expression of *Lma-P22* transcripts

The fact that a large amount of the Lma-P22 protein was present in the secretion from the male second tergite [3] led us to check whether the corresponding mRNA was also highly produced in the same organ. As shown in Fig. 3C, the *Lma-P22* transcript was only expressed in the epidermis of tergites 2, 3 and 4. No *Lma-P22* signal was detected either in the other tergites or in the epidermal extracts from male sternites (data not shown).

4. Discussion

This work is the first report of the molecular cloning of an epidermal cDNA encoding a protein (Lma-P22) secreted onto the cuticular surface of an insect. Northern blot analysis demonstrated that the corresponding mRNA is very specific because it was only found in the epidermis of tergites 2, 3 and 4 in the adult male cockroach. These results partly contradict a previous Western blot analysis which suggested that the corresponding protein was only specific to the second male tergite [3]. However, this immunological analysis was performed on proteins from a pool of tergites 3–8, in which the concentration of Lma-P22 was probably too weak to give a positive signal.

The difference between the calculated MW of the deduced Lma-P22 amino acid sequence (19.7 kDa) and the estimated MW of the corresponding SDS-PAGE band (22 kDa) could be due to glycosylations as previously suggested [3]. Moreover, two putative *N*-glycosylation sites were found at positions Asn-54 and Asn-85 using the SAPS program [14].

As shown in Fig. 4, the best alignment of Lma-P22 was found with *Manduca sexta* insecticyanin A [16], a human allergen (BLG4) cloned in *Blattella germanica* [17], *Galleria mel-*

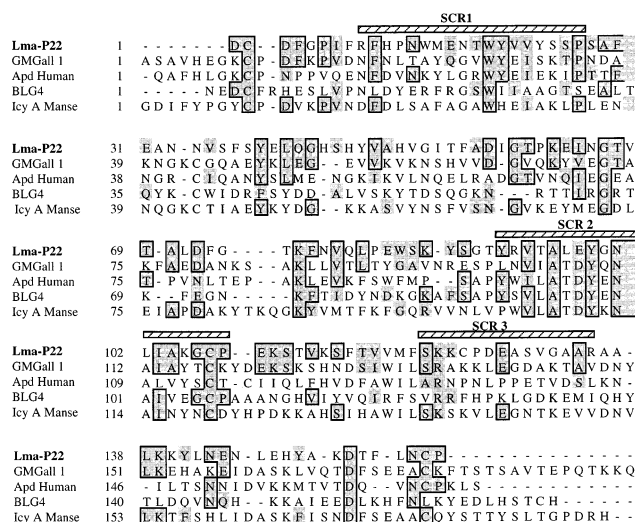


Fig. 4. Alignment of Lma-P22 sequence with gallerin (GMGall 1, EMBL Q24996), human apolipoprotein D (Apo Human [17]), *Blattella* human allergen (BLG4 [16]) and with *Manduca* insectacyanin (Icy A Manse [15]). Identical amino acids are boxed in gray, conserved substitutions are indicated in gray. The structurally conserved regions defining the lipocalin family are indicated (SCR1, 2 and 3). Gaps were introduced to improve the alignment.

lonella gallerin, homologous to *Manduca* insecticyanin (EMBL, Q24996; not published) and human apolipoprotein D [18]. All these extracellular proteins, with a MW close to 20 kDa, belong to the lipocalin family whose members are known to bind and carry small hydrophobic compounds. The degree of similarity of Lma-P22 with these proteins varied from 17 to 26%. These values are commonly found within the members of this family and rarely exceed 20% [19,20]. When we only considered the structurally conserved regions, which are the main characteristics of lipocalins [20], a much higher percentage of identity was found (55–61%; Fig. 4, SCR1–3). The crystal structure of several lipocalins showed eight anti-parallel β -barrels and a C-terminal α -chain [21]. The predicted secondary structure of Lma-P22, obtained using the SSPAL program [22], showed a similar folding. Taken together, these results strongly suggest that Lma-P22 belongs to the lipocalin family.

The resulting tertiary structure of such lipocalins forms a non-polar pocket known to be used as an internal site for hydrophobic ligands such as retinoids, odorous molecules, chromophores, lipids, pheromones or steroids [21,23]. The fact that Lma-P22 is a part of the male tergal gland secretion ingested by the female during the mating courtship raises the possibility that this protein can bind and carry some pheromonal ligand at the cuticle surface, therefore stimulating the female sexual receptivity. The best examples of pheromone-binding lipocalins were found in the hamster where aphrodisin stimulates the copulatory behavior of the male [24], and in the mouse where major urinary proteins (MUPs; secreted by many glands and excreted in large quantities in male urines) stimulates female sexual behavior [25,26]. Three MUP pheromonal ligands including 2*S*-butyl-thiazoline and 2,3-dehydroxobrevicomin were identified [27], but it was shown that the protein alone can accelerate female sexual maturity, thus itself acting as a pheromone [28]. However, when aphrodisin was expressed in bacteria, it was inactive and it could be

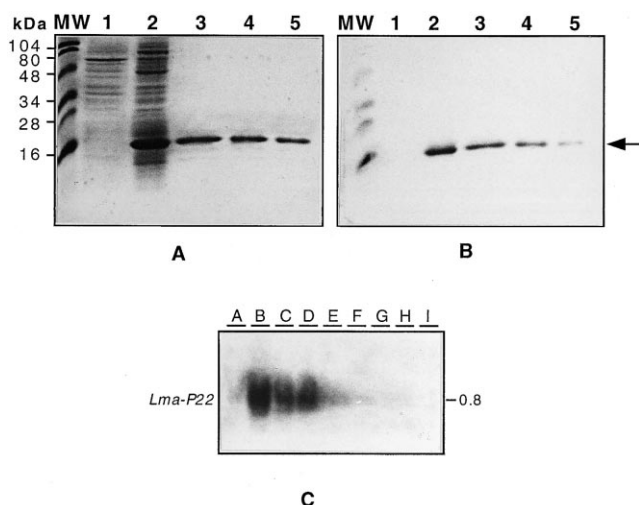


Fig. 3. A: SDS-PAGE of proteins extracted from recombinant colonies of *E. coli* M15 incubated without (1) and with IPTG induction (2–5). SDS-PAGE of the recombinant protein (arrow) after purification on Ni-NTA columns (3–5). B: Western blot analysis showing that the recombinant protein is recognized by the anti-Lma-P22 antibody. C: Northern hybridization analysis of epidermal gland RNAs (15 μ g) extracted from tergites 1–9 of 5 day old adult males, with the *Lma-P22* cDNA probe. RNAs size is indicated in kb.

active only when associated with a ligand found in the organic phase of the vaginal discharge [29].

It finally seems that the vertebrate lipocalins which are involved in sex behavior may act either alone or with their ligands. Lma-P22 is the first characterized protein which seems to be involved in an insect sex behavior. However, its ligand remains unknown. The protein ingested by the female may play a role by itself (as observed for MUP), but the possibility that it binds an attractive ligand is supported by previous work [30] which provided evidence that in *L. maderae*, the male tergal gland secretion contains an organic phase composed of hydrocarbons and of sex pheromone compounds, which could be good candidates as ligands of Lma-P22. It would thus be interesting to determine whether the 'aphrodisiac' property is caused by the Lma-P22 protein itself or by its association with such pheromonal putative ligands. The possibility that we have of synthesizing large quantities of this protein will allow us to check these various hypotheses after screening the putative ligand of Lma-P22 and carrying out behavioral tests.

Acknowledgements: We are grateful to J.P. Farine, B. Quennedey, and V. Royer for helpful discussions. We thank J.-F. Ferveur for critical reading of the manuscript. This work was supported by Centre National de la Recherche Scientifique and the Université de Bourgogne.

References

- [1] Roth, L.M. (1969) *Ann. Entom. Soc. Am.* 62, 176–208.
- [2] Zupko, K., Sklan, D. and Lensky, Y. (1993) *J. Insect Physiol.* 39, 41–46.
- [3] Korchi, A., Farine, J.P. and Brossut, R. (1998) *Insect Biochem. Mol. Biol.* 28, 113–120.
- [4] Plattner, H., Salpeter, M., Carrel, J.E. and Eisner, T. (1972) *Z. Zellforsch.* 125, 45–87.
- [5] Abed, D., Farine, J.P., François, J. and Brossut, R. (1995) *Chemoecology* 1, 19–25.
- [6] Birch, M.C. (1974) in: *Pheromones* (Birch, M., Ed.), pp. 115–134, North Holland, Amsterdam.
- [7] Sreng, L. (1984) *J. Morphol.* 182, 279–294.
- [8] Laemmli, U.K. (1970) *Nature* 227, 680–685.
- [9] Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.
- [10] Sanger, F.S., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [11] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8998–9002.
- [12] Higgins, D.G. and Sharp, P.M. (1988) *Gene* 73, 237–244.
- [13] Kyte, J. and Doolittle, R. (1982) *J. Mol. Biol.* 157, 105–132.
- [14] Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997) *Protein Eng.* 10, 1–6.
- [15] Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B.E. and Karlin, S. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2002–2006.
- [16] Riley, C.T., Barbeau, B.K., Keim, P.S., Kezdy, F.J., Heinrikson, R.L. and Law, J.H. (1984) *J. Biol. Chem.* 259, 13159–13165.
- [17] Arruda, L.K., Vailes, L.D., Hayden, M.L., Benjamin, D.C. and Chapman, M.D. (1995) *J. Biol. Chem.*, 270, (52) 31196–31201.
- [18] Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K. and Lawn, R. (1986) *J. Biol. Chem.* 261, 16535–16539.
- [19] Flower, D.R. (1995) *J. Mol. Recogn.* 8, 185–195.
- [20] Flower, D.R., North, A.C.T. and Attwood, T.K. (1993) *Protein Sci.* 2, 753–761.
- [21] Flower, D.R. (1996) *Biochem. J.* 318, 1–14.
- [22] Salamov, A.A. and Solov'yev, V.V. (1997) *J. Mol. Biol.* 268, 31–36.
- [23] Patel, D.D., Forder, R.A., Soutar, A.K. and Knight, B.L. (1997) *Biochem. J.* 324, 371–377.
- [24] Henzel, W.J., Rodriguez, H., Singer, A.G., Stults, J.T., Macrides, F., Agosta, W.C. and Niall, H. (1988) *J. Biol. Chem.* 263, 16682–16687.
- [25] Knopf, J.L., Gallagher, J.A. and Held, W.A. (1983) *Mol. Cell Biol.* 3, 2232–2241.
- [26] Shaw, P.H., Held, W.A. and Hastie, M.D. (1983) *Cell* 32, 755–761.
- [27] Bacchini, A., Gaetani, E. and Cavaggioni, A. (1992) *Experientia* 48, 419–421.
- [28] Mucignat-Caretta, C., Caretta, A. and Cavaggioni, A. (1995) *J. Physiol.* 486, 517–522.
- [29] Singer, A.G. and Macrides, F. (1990) *Chem. Senses* 15, 199–203.
- [30] Sirugue, D. (1992) Thesis, Université de Bourgogne.