Cloning, post-translational modifications, heterologous expression and ligand-binding of boar salivary lipocalin

Dietrich LOEBEL*[1], Andrea SCALONI†‡[1], Sara PAOLINI§, Carlo FINI§, Lino FERRARA‡, Heinz BREER* and Paolo PELOSI‖[2]

*Institut fur Physiologie, University of Hohenheim, Garbenstrasse 30, 70599 Stuttgart, Germany, †Centro Internazionale Servizi di Spettrometria di Massa, National Research Council, via Pansini 5, 80131 Napoli, Italy, ‡I.A.B.B.A.M., National Research Council, via Argine 1085, 80100 Napoli, Italy, §Dipartimento di Medicina Interna and Sezione INFM, Università di Perugia, via del Giochetto 6, 06126 Perugia, Italy, and ‖Dipartimento di Chimica e Biotecnologie Agrarie, University of Pisa, via S. Michele 4, 56124 Pisa, Italy

Boar submaxillary glands produce the sex-specific salivary lipocalin (SAL), which binds steroidal sex pheromones as endogenous ligands. The cDNA encoding SAL was cloned and sequenced. From a single individual, two protein isoforms, differing in three amino acid residues, were purified and structurally characterized by a combined Edman degradation/MS approach. These experiments ascertained that the mature polypeptide is composed of 168 amino acid residues, that one of the three putative glycosylation sites is post-translationally modified and the structure of the bound glycosidic moieties. Two of the cysteine residues are paired together in a disulphide bridge, whereas the remaining two occur as free thiols. SAL bears sequence similarity to other lipocalins; on this basis, a three-dimensional model of the protein has been built. A SAL isoform was expressed in *Escherichia coli* in good yields. Protein chemistry and CD experiments verified that the recombinant product shows the same redox state at the cysteine residues and that the same conformation is observed as in the natural protein, thus suggesting similar folding. Binding experiments on natural and recombinant SAL were performed with the fluorescent probe 1-aminoanthracene, which was efficiently displaced by the steroidal sex pheromone, as well as by several odorants.

Key words: androstenone, chemical communication, pheromone-binding protein, pheromones, pig.

INTRODUCTION

In several mammals volatile pheromones have been found to be associated with soluble binding proteins [1–3] belonging to the large family of lipocalins. These molecules are single polypeptide chains of approx. 150–170 residues, sharing typical β-barrel folding [4]. Their stable and compact structure encloses a binding site for hydrophobic ligands. The physiological role of several lipocalins has been associated with the transport of hydrophobic compounds in aqueous biological fluids. Sequence similarity studies revealed a high similarity between mammalian pheromone-binding proteins and other lipocalins present in nasal mucosa called odorant-binding proteins (OBPs). It has been proposed that OBPs and pheromone-binding proteins play complementary roles in chemical communication, the first in the perception of pheromones and the second in their delivery [5,6].

Pheromone-binding proteins were first described in the urine of mouse [7] and rat [8], where they are excreted at concentrations of several milligrams per millilitre. The importance of urine in chemical communication in these species soon suggested a related function for such proteins [1]. Furthermore, when excreted in the urine, the mouse urinary proteins are strongly associated with volatile ligands, such as dehydrobrevicomin, 2-sec-butylthiazoline and 4-ethylphenol [9], whose pheromonal activity has been assessed [10]. In the hamster, another lipocalin, called aphrodisin, is secreted in the vaginal discharge and shows pheromonal activity [2,11]. A third biological fluid associated with pheromone transport is the saliva. In the mouse, behavioural experiments have clearly demonstrated its involvement in chemical communication [12]. Lipocalins structurally related to the urinary proteins are known to be secreted in the saliva of this species, although they have not been functionally characterized [13,14]. Finally, components of axillary odour in humans are associated with apolipoprotein-D, another member of the lipocalin family [15,16].

Particularly interesting is the close similarity of these proteins associated with pheromone delivery with OBPs of the nasal area. Thus OBPs, with sequence identity of 90% or more to the urinary and salivary proteins, are expressed in the nasal mucosa of the mouse [17,18], as well as of other species, such as rabbit [19] and porcupine [20]. On the other hand, both subunits of mouse OBP-I share nearly 60% of their residues with hamster aphrodisin [21].

Recent data indicate that pheromone-binding proteins are not only carriers for volatile pheromones, but could themselves be endowed with pheromonal activity. In fact, it has been demonstrated that the urinary proteins of male mice are able to accelerate the onset of puberty in young females, when applied to their nasal cavity. The effect is also produced by the protein devoid of its endogenous ligand and even by a synthetic peptide

Abbreviations used: CAM, carboxamidomethyl; DIG, digoxigenin; ESI–MS, electrospray ionization MS; IPTG, isopropyl β-D-thiogalactoside; MALDI–TOF MS, matrix-assisted laser-desorption ionization–time-of-flight MS; MUP, major urinary protein; OBP, odorant-binding protein; PNGase F, peptide *N*-glycosidase F; poly(A)$^+$, polyadenylated; PTH, phenylthiohydantoin; RACE, rapid amplification of cDNA ends; SAL, salivary lipocalin; TFA, trifluoroacetic acid.

[1] These authors made equal contributions to this study.

[2] To whom correspondence should be addressed (e-mail ppelosi@agr.unipi.it).

The nucleotide sequence data reported will appear in the EMBL, GenBank®, DDBJ and GSDB Nucleotide Sequence Databases under the accession number AJ249974.

reproducing its first six residues [22]. The same proteins activate the production of Ins(1,4,5)P_3, through the action of a G-protein in homogenates of the vomeronasal organ [23]. This effect suggests specific binding of major urinary proteins (MUPs) to class II vomeronasal receptors, recently discovered in the rat [24–26]. These receptors present, in addition to the seven transmembrane domain common to olfactory receptors and class I vomeronasal receptors, a large extracellular domain of approx. 600 amino acids, a possible binding site for proteinaceous ligands.

In the pig the role of male saliva in arousing the female has been known for a long time. Specific sex pheromones, such as 5α-androst-16-en-3-one and 5α-androst-16-en-3-ol, are delivered by the boar through such a vehicle [27]. Binding proteins for these pheromones, called pheromaxeins, have been purified from boar saliva, but have not been further characterized [28,29]. Recently we have purified a very abundant lipocalin from the boar submaxillary gland, which is different from the previously described pheromaxeins [3]. This protein, which we named salivary lipocalin (SAL), reversibly binds the odorant 2-isobutyl-3-methoxypyrazine with a micromolar dissociation constant and contains the boar steroidal pheromones 5α-androst-16-en-3-one and 5α-androst-16-en-3-ol, as endogenous ligands.

In the present study we describe the cloning of boar SAL, its complete structural elucidation (including post-translational modifications) and its expression in a bacterial system. Using volatile pheromones and general odorants, the ligand-binding of both the native and the recombinant proteins was characterized.

MATERIALS AND METHODS

Tissues and reagents

The different tissues were obtained from single male and female mature pigs. After dissection they were rapidly frozen in liquid nitrogen and stored at −70 °C until protein or nucleotide extraction was performed. All reagents were of analytical grade, except those used during Edman degradation analysis (sequencing grade).

Isolation of full-length SAL cDNA

All DNA- and RNA-related methods were performed according to standard procedures [30]. First strand cDNA synthesis for rapid amplification of cDNA ends (RACE) was performed on 3 μg of total submaxillary gland-RNA from a male pig using the Expanded Reverse Transcriptase Kit (Roche). Polyadenylated [poly(A)$^+$] RNA was transcribed with 20 pmol of the lock-docking primer-adapter Q_{total} [5′-TGTAATACGACTCAC-TATAGGGCGAGAATTCAACG$(T)_{18}$vomeronasal-3′] as described previously [31]. For specific amplification of cDNA clones that encode the 3′-end of SAL, degenerated oligonucleotides SAL-1/-2 [5′-AARATHGCNGGNGARTGGTA-(T/C)-3′] were derived from the N-terminal peptide sequence spanning Lys^{17}–Tyr^{23} as described previously [3] and used as sense primers in PCR. Q_{total}-primed cDNA (2 μl) was amplified in a volume of 50 μl with the SAL-1 sense primer and the antisense primer Q_{out} (5′-TGTAATACGACTCACTATAGGG-CG-3′), matching the 5′-end of Q_{total}, with the *Pwo* proof-reading PCR system (Roche). Cycling conditions were carried out using touch down PCR with a first cycle of 2 min at 94 °C, 30 s at 60 °C and 2 min at 72 °C, followed by 20 cycles of 30 s at 94 °C, 30 s at 60 °C (with reduction of 0.5 °C per cycle to 50 °C) and 2 min at 72 °C, followed by an additional 20 cycles of 30 s at 94 °C, 30 s at 50 °C and 2 min at 72 °C. The reaction was completed by a final step at 72 °C for 10 min. Reaction conditions were 10 mM Tris/HCl (pH 8.8), 50 mM KCl, 2 mM $MgSO_4$, 200 μM of each dNTP, 100 pmol of each primer and 5 units of *Pwo* DNA polymerase (Roche). PCR was performed on an MJ Research Peltier Thermocycler PTC 200. For the amplification of the 5′-end of the SAL cDNA, total RNA from submaxillary glands was transcribed using SMART first strand synthesis technology (ClonTech, Palo Alto, CA, U.S.A.) with a SAL-specific antisense cDNA primer and the primer adaptor SMART II (5′-AAGCAGTGGTAACAACGCAGAGTACGCGGG-3′) for the 5′ cDNA ends. Two rounds of nested-PCR were performed on the 5′-SMART cDNA with SMART III (5′-GCAG-TGGTAACAACGCAGTAC-3′) as the sense primer and two SAL-specific antisense primers, SALr1 (5′-ACCATCTCCTA-CTTTGTC-3′) and nested SALr2 (5′-AAAGTCAGTACAC-TCTCCG-3′), as described above. PCR products were separated by electrophoresis on a 1% agarose gel, purified using the QiaexII-gel extraction kit (Qiagen, Chatsworth, CA, U.S.A.) and subsequently ligated into pGEM-5Zf(+) employing the pGEM-T vector system (Promega, Madison, WI, U.S.A.). Sequencing was performed using the ABI prism BigDye ready reaction terminator cycle sequencing kit (Perkin–Elmer, Forster City, CA, U.S.A.), according to the manufacturer's instructions. Samples were run on an ABI Prism 310 Genetic analyser (Perkin–Elmer).

Analysis of RNA expression

Total RNA from the different tissues was extracted by TRIzol separation as recommended by the supplier (Life Technologies). RNA concentrations were detected at A_{260}. Each RNA sample (10 μg) was heated at 65 °C for 5 min and loaded on to a 25 mM Mops (pH 7.4)/0.67 M formaldehyde/1.2% (w/v) agarose denaturing gel. After electrophoresis, RNA was transferred on to Hybond N$^+$ nylon membrane (Amersham, Little Chalfont, Bucks., U.K.) with 20 × SSC (where 1 × SSC corresponds to 0.15 M NaCl/0.015 M sodium citrate), and was covalently linked to the membrane by baking at 80 °C for 2 h. The full length SAL probe was labelled using the digoxigenin (DIG) DNA labelling kit (Roche) by the random priming method. Hybridization was carried out in DIG easyHyb hybridization buffer (Roche) under high stringency conditions. The membranes were washed in 0.1 × SSC/0.1% SDS at 68 °C and detected according to the manufacturer's protocol (blocking reagent and anti-DIG–alkaline phosphatase conjugated F_{ab} fragments were obtained from Roche). Subsequently, the membranes were exposed to a Fuji RX film for 5–10 h.

Southern blot analysis

High molecular-mass genomic DNA was prepared from muscle tissue of an individual animal according to Ausubel et al. [32]. The genomic DNA was digested with restriction endonucleases, size fractionated on a 0.8% agarose gel and transferred to Hybond N$^+$ nylon membrane using standard protocols [30]. The blot was hybridized under high stringency conditions to the DIG-labelled DNA probe, as described previously [33]. The blot was developed as described above.

Construction of the expression plasmid and protein expression

The coding sequence of SAL was amplified by PCR using the full-length cDNA clone as a template. Specific primers, matching the SAL sequence, were designed with 5′-extended *Bam*H1 and *Hin*dIII sites. The sense primer containing the *Bam*H1 site (underlined) was 5′-AATACGGATCCACACAAGGAAGCA-GGCCAAG-3′ and the antisense primer containing the *Hin*dIII site (underlined) was 5′-CATCGCAAGCTTCACTCAGCAC-

TGGACTCCTGG-3′. The PCR product was double-digested with *Bam*H1 and *Hin*dIII and subsequently cloned into the corresponding site of the expression vector pQE31 (Qiagen) under control of the T5/*lac*Z hybrid promoter system. The derived plasmid construct was confirmed by DNA sequencing and transformed into *Escherichia coli* BL21.

The expression vector contains the ampicillin resistance gene and generates a recombinant fusion product with a Met-Arg-Gly-Ser-$(His)_6$-Thr-Asp-Pro tag at the N-terminus of the protein. The expression was regulated by co-transformation of the pREP4 plasmid, containing a kanamycin resistance gene and multiple copies of the *lac*I repressor gene. Induction was performed when the bacterial culture had reached an A_{600} of 0.5–0.7, by adding 0.2 mM isopropyl β-D-thiogalactoside (IPTG) to the medium for 3 h at 37 °C. Cells were harvested by centrifugation at 5000 ***g*** for 15 min and resuspended in 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 20 mM imidazole, 0.25 mg/ml lysozyme and 1 mM PMSF, and stored on ice for 1 h.

Purification of the native and recombinant SAL

Native protein was purified from a single male pig submaxillary gland by conventional chromatographic methods, as previously reported [3]. Recombinant SAL was purified as described previously [34]. In both cases individual components were isolated in a homogeneous form by a final chromatographic step on a Vydac C_4 column 214TP54 (250 mm × 4.6 mm; 5 μm diameter beads; 300 Å pore size; The Separations Group, Hesperia, CA, U.S.A.). Proteins were dissolved in 0.1 % trifluoroacetic acid (TFA), loaded on to the column and eluted by a linear gradient from 10 % to 60 % acetonitrile in 0.1 % TFA over 30 min, at a flow rate of 1 ml/min.

Western blot analysis

Protein samples and extracts were separated by electrophoresis under denaturing conditions [SDS/PAGE (15 % polyacrylamide)] and then electroblotted on to a nitrocellulose membrane, according to Kyhse-Andersen [35]. After treatment with 0.2 % gelatin from porcine skin (Sigma)/0.05 % Tween 20 in PBS for 2 h, the membrane was incubated with the crude antiserum against native SAL at a dilution of 1:10000 and then with goat anti-rabbit IgG–horseradish peroxidase conjugate (dilution 1:10000). Immunoreacting bands were detected by treatment with 4-chloro-1-naphthol.

Preparation of polyclonal antibodies

Antisera were obtained by injecting an adult rabbit subcutaneously and intramuscularly with 400 μg of protein purified from a boar submaxillary gland, followed by two additional injections of 250 μg after 18 and 30 days. The protein was emulsified with an equal volume of Freund's complete adjuvant for the first injection and with an equal volume of incomplete adjuvant for further injections. Animals were bled 10 days after the last injection, and the serum was partially purified by precipitation in 45 % ammonium sulphate.

Lectin assays

To identify N-glycosylated or O-glycosylated sites, samples of the native and recombinant proteins were incubated with *N*-glycosidase or *O*-glycosidase and analysed by SDS/PAGE, together with samples of untreated protein. To assay the nature of linked oligosaccarides, 2 μg of SAL samples, carboxypeptidase Y, transferrin, asialofetuin and fetuin (as control standard glycoproteins; glycan-differentiation kit; Boehringer Mannheim, Mannheim, Germany), and acylpeptide hydrolase (as control blank) were spotted directly on to PVDF membranes. After incubation in 50 mM Tris/HCl (pH 7.0)/50 mM NaCl with DIG-labelled lectins, from *Galanthus nivalis* (specific for terminal mannose), *Sambucus nigra* [specific for sialic acid α(2–6)galactose], M*aackia amurensis* [specific for sialic acid α(2–3)galactose], peanut [specific for galactose β(1–3)*N*-acetylgalactosamine] and *Datura stramonium* [specific for galactose β(1–4)*N*-acetylglucosamine], membranes were exhaustively washed. The immunological detection was performed by a colorimetric reaction using anti-DIG–alkaline phosphatase conjugates. The experiments were performed in duplicate.

Chemical and enzymic reactions, and peptide separation

Before carrying out sequence analysis, protein samples were reduced and alkylated under the conditions described previously [36]. Parallel alkylation experiments were performed under denaturing conditions without treatment with dithiothreitol. Proteins were freed from excess reagents by an HPLC desalting step on a Vydac C_4 column as described above.

Native and carboxyamidomethylated proteins were digested with trypsin or endoprotease Glu-C (Boehringer Mannheim) in 0.4 % ammonium bicarbonate (pH 7.5), at 37 °C for 18 h, using an enzyme/substrate ratio of 1:50 (w/w). Peptide mixtures were deglycosylated by incubating with 0.15 units of peptide *N*-glycosidase F (PNGase F) overnight in 0.4 % ammonium bicarbonate (pH 8.0) at 37 °C. Peptide mixtures were fractionated by reverse-phase HPLC on a Vydac C_{18} column 214TP52 (250 mm × 2.1 mm; 5 μm diameter beads; 300 Å pore size; The Separations Group) using a linear gradient from 5 % to 60 % acetonitrile in 0.1 % TFA over 60 min, at a flow rate of 0.2 ml/min. Individual components were manually collected and dried down in a Speed-vac centrifuge (Savant).

Mass spectrometry and amino acid sequence analysis

Intact proteins were submitted to electrospray ionization MS (ESI–MS) analysis, which was performed using a BIO-Q triple quadrupole mass spectrometer (Micromass, Wythenshaw, Manchester, U.K.). Samples were dissolved in 1 % (v/v) acetic acid and 2–10 μl was injected into the mass spectrometer at a flow rate of 10 μl/min. The quadrupole was scanned from m/z 500–1800 at 10 s/scan and the spectra were acquired and elaborated using the MASSLYNX software. Calibration was performed by the multiply charged ions from a separate injection of myoglobin (molecular mass of 16951.5 Da). All mass values are reported as average masses.

Matrix-assisted laser-desorption ionization–time-of-flight (MALDI–TOF) mass spectra were recorded using a Voyager DE MALDI–TOF mass spectrometer (Perkin–Elmer); a mixture of analyte solution, α-cyano-4-hydroxycinnamic acid and bovine insulin was applied to the sample plate and dried *in vacuo*. Mass calibration was performed using the molecular ions from the bovine insulin at 5734.54 Da and the matrix at 379.06 Da as internal standards. Raw data were analysed using computer software provided by the manufacturer and are reported as average masses. Assignments of the recorded mass values to individual peptides were performed on the basis of their molecular mass.

Amino acid sequence was determined by direct Edman degradation using a Perkin–Elmer-Applied Biosystems 477A pulsed-liquid protein sequencer equipped with a Perkin–Elmer-Applied

Biosystems 120A HPLC apparatus for phenylthiohydantoin (PTH)-amino acid identification.

CD measurements

CD spectra were recorded on a Jasco J-900 spectropolarimeter at 25 °C between 190 nm and 300 nm. The spectra were accumulated 10 times with a bandwidth of 1.0 nm and a resolution of 1 nm at a scan speed of 200 nm/min. The purified protein samples were diluted into 10 mM potassium phosphate buffer (pH 7.5) at a final concentration of 1 mg/ml. Protein concentrations were determined at A_{280} assuming $\epsilon = 18996\ M^{-1}cm^{-1}$ and a molecular mass of 21000 Da. Data were expressed as molar ellipticity based on the protein concentration.

Intrinsic fluorescence measurements and fluorescence binding assays

The tryptophan fluorescence emission spectra of native and recombinant SAL were recorded on a Jasco FP-750 spectrofluorimeter. Protein concentration was 1 μM in 50 mM Tris/HCl (pH 7.5). The temperature was maintained at 25 ± 0.1 °C by a temperature controller, Jasco ETC-272T, utilizing the Peltier effect, and spectra were recorded at 1 nm intervals, at a scan speed of 60 nm/min with five accumulations. The excitation and emission band widths were 5 nm and 10 nm respectively.

The binding curve to the fluorescent probe 1-aminoanthracene was obtained by exciting mixtures of the protein (1 μM) and the ligand (0.1 to 10 μM) in 50 mM Tris/HCl (pH 7.4) at 256 nm and recording emission spectra between 300 and 600 nm. Experimental values represented means of three determinations and they were fitted to a hyperbolic curve using the computer program ORIGIN 5.0 (Microcal Software, Inc., Northampton, MA, U.S.A.). Goodness-of-fit was estimated using Chi-square tests. The concentration of the fluorescent probe bound to the protein was evaluated from the intensity of the emission spectrum at 488 nm, assuming a stoichiometry of protein/ligand of 1:1. The dissociation constant of the complex SAL–1-aminoanthracene was obtained from the binding-curve using the computer program ENZPACK3 (Biosoft, Cambridge, U.K.). Competitive binding experiments were performed in similar conditions as described above, using the protein and the fluorescent ligand 1-aminoanthracene at concentrations of 1 and 10 μM respectively. Competing odorants were used at final concentrations of 1, 3, 10, 30 and 100 μM.

Molecular modelling

Sequence analysis was performed using the HUSAR 3.0 software package (EMBL, Heidelberg, Germany) based on the sequencing analysis package GCG 7.2. Pairwise and secondary structure-driven sequence alignments were obtained by using the CLUSTAL W multiple alignment and PHD programs [37].

Computer modelling was performed on a Silicon Graphics workstation. Three-dimensional models of SAL isoforms were constructed on the basis of the crystallographic structures of mouse MUP [38], pig OBP-I [39] and bovine OBP [40,41], taken from the Brookhaven Protein Data Bank, using the PROMOD II program version 2.01 [42]. The N-terminal hexapeptide and C-terminal octapeptide of the protein were not included in the model construction because these residues are absent from the other sequences.

RESULTS

Molecular cloning of the SAL cDNA

The submaxillary glands of the male pig were used for RNA extraction and reverse-transcription into total cDNA, based on the information that they contain large amounts of SAL [3]. Using a RACE-PCR technique, with degenerated forward primers, designed on the N-terminal amino acid sequence spanning Lys^{17}–Tyr^{23}, and Q_{out} as a specific reverse primer matching the cDNA 3′-ends, we have amplified a specific DNA fragment of 800 bp. After cloning of this amplification product into a pGEM-T vector, positive clones were analysed by DNA sequencing. To obtain full-length cDNA information on SAL a 5′-RACE strategy was used. The forward primer template for PCR amplification was linked on to the cDNA 5′-end by a short oligonucleotide, whereas the reverse primer sequence was deduced from the sequence information on the 3′-RACE clone. The amplification product of 300 bp was cloned and analysed by DNA sequencing.

Sequence of SAL cDNA

The nucleotide sequence for the SAL gene, together with the deduced amino acid sequence, shown in Figure 1, was verified from the analysis of 16 clones obtained from four independent PCR experiments. The SAL-encoding cDNA is 908 bp long and contains an open reading frame of 573 bp, flanked by 58 bp at its 5′-end and 259 bp at its 3′-end, including a consensus polyadenylation signal AATAAA at positions A^{891}–A^{896}. The open reading frame encodes a protein of 191 amino acids. Analysis of the deduced amino acid sequence revealed that the 5′-end of the coding region contains a typical signal sequence (Figure 1). A putative signal peptidase cleavage site can be assigned after the deduced first 16 amino acids (…Gly-Leu-Thr-Leu-Ala-Ser-Ser-↓-His-Lys-Glu-Ala-Gly…), thus generating a protein with an N-terminus identical to that determined for the purified protein experimentally [3]. Three putative N-glycosylation sites were present at positions Asn^{38}, Asn^{53} and Asn^{99}.

Northern blot analysis

To determine whether the expression of the SAL gene was restricted to the submaxillary glands, different tissues of adult male pig were analysed by Northern blotting (Figure 2A). The labelled probe detected two diffuse bands of 1200 bp and 950 bp in the submaxillary glands of male pig, indicating strong tissue-specific expression of the SAL gene. Whether the two bands were products of two genes, different splicing variants of a single gene or due to the presence of poly(A)$^+$ tails of different sizes was not investigated. The same analysis also revealed that the expression level of SAL mRNA was much lower in submaxillary glands of castrated male pigs than in sexually mature individuals, while it was not detectable in female submaxillary glands.

Southern blot analysis

To assess the number of SAL genes, Southern blot analysis was performed on male pig genomic DNA. The DNA was digested with the restriction enzymes *Kpn*I, *Pst*I, *Sac*I, *Hin*dIII and *Eco*RI, which do not cut the cDNA probe. The SAL probe stained two bands in each lane (Figure 2B), indicating the presence of two homologous genes. This result is in line with identification of a second SAL isoform by the combined Edman degradation/MS approach.

```
1     AGAGGGGAGAGATTCAGTCCCAGTGACTACAGAGGAGACAGTGCTGTCCGCTGCCAAG ATG AAG CTG CTG      70
1                                                                 M   K   L   L       4

71    CTC TTG CTG TGT CTG GGG CTG ACT CTA GCC TCT TCC CAC AAG GAA GCA GGC CAA GAT     127
5     L   L   L   C   L   G   L   T   L   A   S   S   H   K   E   A   G   Q   D      23

128   GTT GTG ACA AGC AAC TTC GAT GCT TCA AAG ATT GCC GGC GAG TGG TAT TCC ATT CTT     184
24    V   V   T   S   N   F   D   A   S   K   I   A   G   E   W   Y   S   I   L      42
                                          *
185   TTG GCC TCA GAC GCC AAG GAA AAT ATA GAA GAA AAT GGT AGC ATG AGA GTT TTT GTG     241
43    L   A   S   D   A   K   E   N   I   E   E   N   G   S   M   R   V   F   V      61
                          *
242   GAG CAT ATC CGT GTC TTG GAC AAC TCT TCC CTA GCC TTT AAA TTT CAG AGA AAG GTA     298
62    E   H   I   R   V   L   D   N   S   S   L   A   F   K   F   Q   R   K   V      80

299   AAC GGA GAG TGT ACT GAC TTT TAT GCG GTT TGT GAC AAA GTA GGA GAT GGT GTG TAT     355
81    N   G   E   C   T   D   F   Y   A   V   C   D   K   V   G   D   G   V   Y      99
                                                            *
356   ACA GTT GCC TAC TAT GGA GAG AAC AAA TTT CGC CTA CTT GAA GTG AAC TAT TCT GAC     412
100   T   V   A   Y   Y   G   E   N   K   F   R   L   L   E   V   N   Y   S   D     118

413   TAT GTC ATT TTA CAC CTT GTG AAT GTC AAT GGC GAC AAA ACA TTC CAG CTG ATG GAG     469
119   Y   V   I   L   H   L   V   N   V   N   G   D   K   T   F   Q   L   M   E     137

470   TTC TAC GGC CGA AAA CCA GAT GTG GAG CCA AAA CTC AAG GAC AAG TTT GTG GAG ATT     526
138   F   Y   G   R   K   P   D   V   E   P   K   L   K   D   K   F   V   E   I     156

527   TGC CAA CAA TAT GGC ATC ATC AAG GAA AAC ATA ATC GAT CTG ACC AAA ATT GAT CGC     583
157   C   Q   Q   Y   G   I   I   K   E   N   I   I   D   L   T   K   I   D   R     175

584   TGC TTC CAA CTC CGA GGG AGT GGA GGA GTC CAG GAG TCC AGT GCT GAG TGA AGGCCTC     641
176   C   F   Q   L   R   G   S   G   G   V   Q   E   S   S   A   E   *             191

642   CTCATGTGGGCTCCAGGATCTTCTCCTCCTTGGTCCCCCCATCATCTGGTGACAAGTCCTGTGATCCAGTTTCCA    716
717   TCCCGAGCCCACAGAAAGCCTGAAAGGCATTATCTCTGCCTCTTCAGGATCTTCCCTAATTATCTAGGAAGATTC    791
792   ATCAGCTCACCAAGAATCAAAAGTGTTCTCCAAATTCCTGACTCCCTCCGCCAGACCCAGGACAGGGCCACCATG    866
867   AGGAGAACCTCCTCTACCAGGTCAATAAATGATTCTCCTTAC                                    908
```

Figure 1 Nucleotide sequence of porcine SAL and the deduced amino acid sequence

The start and stop codons are underlined and the polyadenylation signal is shown in bold. The three putative N-glycosylation sites are marked by asterisks. The amino acids in bold correspond to the signal peptide and the amino acid sequence used to deduce the sequence for degenerate primers is double underlined. The length of the sequence is 908 bp, and the open reading frame is 573 bp, encoding a protein of 191 amino acids.

Analysis of SAL protein

On the basis of the tissue- and sex-specific expression of SAL mRNA (Figure 2), we probed the expression of the protein by Western blot analysis using a polyclonal antiserum, raised against a sample of SAL purified from boar submaxillary glands (results not shown). The results clearly showed that the crude extract of sow submaxillary gland did not cross-react, while the extract of the boar gland was heavily stained, confirming the sex-specificity of SAL. Among the nasal tissues, no reacting bands were visible in the vomeronasal or in the olfactory extracts, while a protein of the respiratory extract, migrating with higher apparent molecular mass than SAL, was recognized by the antiserum. This protein, which we call OBP-IV, has been purified by immunoaffinity chromatography and is currently being characterized.

Heterologous expression of SAL is an essential prerequisite for detailed functional investigation. The recombinant polypeptide was produced in a bacterial system, after cloning the corresponding cDNA sequence into a pQE31 plasmid. This expression vector generates a fusion protein containing 13 additional amino acids at its N-terminus, including a His_6-tag. Using *E. coli* BL21(DE3), transfected with this vector construct, several clones were tested for SAL expression. Optimal production was obtained 3 h after induction with IPTG. The bacterial expression and purification of recombinant SAL was analysed by SDS/PAGE (15% polyacrylamide) (Figure 3A). The protein migrates as a single band of 21 kDa, which is absent in the non-induced *E. coli* culture. The recombinant SAL was overexpressed in high quantities as a soluble protein and purified by Ni^{2+}-nitriloacetate–agarose affinity chromatography. This protocol afforded approx. 15 mg of protein at a purity of > 98% from 2 litres of *E. coli* culture. The recombinant protein, when analysed by isoelectric focusing in a pH 4–6 gradient of Ampholines, showed a pI of 5.6. The calculated value for this polypeptide is 5.58. The native SAL gave two bands at pI 4.5 and 4.7 [3].

Isoform characterization

SAL, purified from crude extracts of one pig submaxillary gland, as previously described [3], was subjected to analytical reverse-phase HPLC and eluted in a single chromatographic peak. Electrophoretic analysis of this fraction resulted in a broad band probably due to the presence of linked oligosaccharides, as demonstrated by concanavalin A binding after electroblotting on to nitrocellulose membranes. Strong cross-reaction was observed with native SAL, while no staining was detectable with the recombinant polypeptide, as expected. To verify the presence of N-glycosylated or O-glycosylated sites, samples of native and recombinant proteins were treated with *N*-glycosidase and *O*-glycosidase. Figure 3(B) shows the SDS/PAGE analysis of the products obtained; native SAL (lane 1) migrated as two bands corresponding to apparent masses of 24 and 21.5 kDa. After treatment with *N*-glycosidase, these two bands were moved to apparent masses of 22 and 20 kDa respectively (lane 2). In contrast, the apparent molecular mass of the recombinant SAL was not affected by the *N*-glycosidase treatment. Both native and recombinant SAL were refractive to the action of *O*-glysosidase. The nature of the N-linked oligosaccharides was assayed by lectin-binding assays. The native protein was recognized specifically by *G. nivalis* agglutinin, which binds to terminal

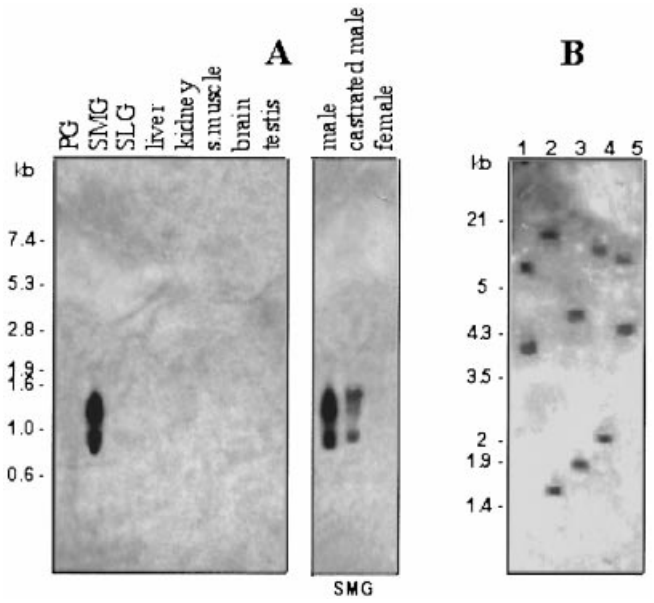

Figure 2 Northern blot (A) and Southern blot (B) analysis

(**A**) Tissue distribution of SAL mRNA in adult male pig. Total RNA samples (10 μg) were analysed as described in the Materials and methods section. The blots were hybridized with a DIG-labelled SAL cDNA probe. A diffuse double band corresponding to 1200 bp and 950 bp is visible only in submaxillary glands. Sex-specific expression of SAL in pig was also compared between adult male, castrated male and female. The position of the RNA marker is indicated. (**B**) Boar genomic DNA was digested with *Kpn*I (lane 1), *Pst*I (lane 2), *Sac*I (lane 3), *Hin*dIII (lane 4) and *Eco*RI (lane 5) and analysed by Southern blotting under high stringency conditions employing a DIG-labelled SAL cDNA probe. The positions of *Hin*dIII and *Hin*dIII/*Eco*RI digested λ-DNA size markers are indicated. PG, parotid glands; SLG, sublingual glands; SMG, submaxillary glands.

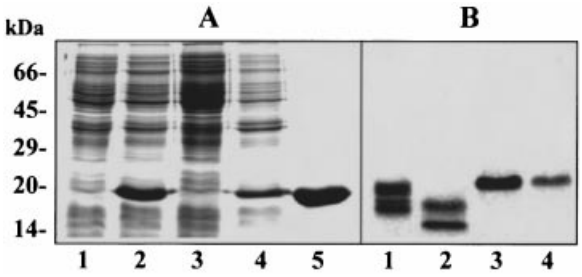

Figure 3 SDS/PAGE analysis of native and recombinant SAL

(**A**) Prokaryotic expression and purification of recombinant SAL. The expression and purification were analysed by SDS/PAGE (15% polyacrylamide) followed by Coomassie Blue staining. Lane 1, non-induced bacteria; lane 2, IPTG-induced bacteria; lane 3, inclusion bodies; lane 4, bacterial cytosol; lane 5, Ni^{2+}-nitriloacetic acid purified fusion protein. A polypeptide band of 21 kDa is only visible in lanes 2, 4 and 5. (**B**) SDS/PAGE of native and recombinant SAL. Lane 1, native SAL; lane 2, native SAL digested with *N*-glycosidase; lane 3, recombinant SAL; lane 4, recombinant SAL digested with *N*-glycosidase.

mannose residues, and *D. stramonium* agglutinin, specific for galactose β(1–4)*N*-acetylglucosamine.

Aliquots of the native protein were directly submitted to ESI–MS analysis yielding a transformed mass spectrum where two distinct series of components differing by 203 Da were identified in the spectrum. The measured molecular mass was slightly different from that previously reported [3]. In particular, species at 20207.5 $\pm$ 2.8 Da, 20410.8 $\pm$ 2.7 Da, 20613.4 $\pm$ 2.6 Da, 20816.4 $\pm$ 2.3 Da, 21019.9 $\pm$ 3.5 Da, and species with molecular masses of 20221.9 $\pm$ 3.0 Da, 20424.9 $\pm$ 2.5 Da, 20627.8 $\pm$ 2.0 Da, 20830.8 $\pm$ 3.1 Da, 21034.0 $\pm$ 3.2 Da were observed in the spectrum. These results suggested a high heterogeneity of the protein molecule, probably due to the presence of oligosaccharide chains.

Table 1 MALDI–TOF MS analysis of tryptic and endoprotease Glu-C digests of reduced and carboxamidomethylated samples of native SAL following treatment with PNGase F

MH^+	Peptide	Modification
Trypsin		
1833.6	(1–17)	
1568.4	(3–17)	
1637.7	(18–32)	
1178.8	(33–42)	
899.7	(43–49)	
857.9	(43–49)	$Val^{45} \rightarrow Ala$, $Ile^{48} \rightarrow Val$
4080.9	(43–77)	2CAM
4066.6	(43–77)	2CAM, $Val^{45} \rightarrow Ala$, $Ile^{48} \rightarrow Val$, $Ala^{73} \rightarrow Val$
1095.7	(50–59)	
450.5	(60–62)	
1806.5	(63–77)	2CAM
1834.9	(63–77)	2CAM, $Ala^{73} \rightarrow Val$
1678.9	(64–77)	2CAM
1706.7	(64–77)	2CAM, $Ala^{73} \rightarrow Val$
3598.9	(64–94)	2CAM
1635.8	(78–92)	
4338.5	(78–115)	
2418.5	(95–115)	
4486.1	(95–132)	
1292.3	(116–125)	
1741.8	(135–148)	CAM
4016.3	(135–168)	2CAM
1498.5	(137–148)	CAM
945.9	(149–156)	
1330.1	(149–159)	
403.4	(157–159)	
1108.2	(157–164)	CAM
723.7	(160–164)	CAM
981.9	(160–168)	CAM
Endoprotease Glu-C		
1809.4	(4–21)	
3672.4	(4–33)	
1395.9	(22–33)	
1749.5	(22–36)	
1881.7	(22–37)	
1039.4	(38–46)	
1011.4	(38–46)	$Val^{45} \rightarrow Ala$
5095.6	(47–90)	2CAM
5109.3	(47–90)	2CAM, $Ile^{48} \rightarrow Val$, $Ala^{73} \rightarrow Val$
2653.8	(68–90)	2CAM
2681.6	(68–90)	2CAM, $Ala^{73} \rightarrow Val$
1520.8	(77–90)	
920.1	(91–97)	
2813.2	(98–121)	
2196.1	(122–139)	
5573.3	(122–168)	
1252.2	(140–149)	CAM
2163.9	(150–168)	

In order to verify the protein primary structure deduced from the nucleotide sequence, reduced and carboxamidomethylated samples of natural SAL were analysed by MS. A portion of the SAL tryptic peptide mixture, obtained after PNGase F digestion, was submitted directly to MALDI–TOF MS analysis. The results obtained are shown in Table 1. Most of the signals recorded in the spectrum were associated with the corresponding peptide along the amino acid sequence on the basis of its mass value and

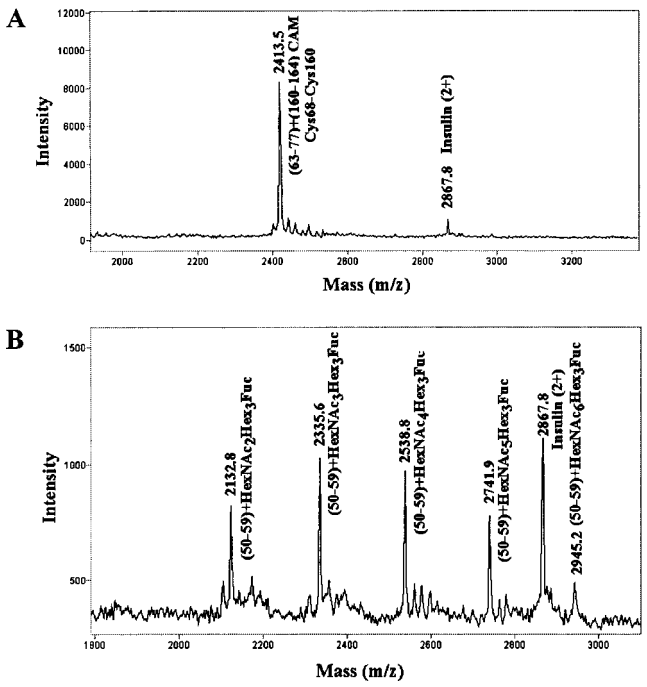

Figure 4 Post-translational modification analysis of porcine SAL

(**A**) MALDI–TOF mass spectrum of a disulphide bridge-containing peptide collected from narrow-bore HPLC analysis of SAL, carboxamidomethylated under native conditions following digestion with trypsin. The two cysteine residues involved in the disulphide bond, as identified by automated Edman degradation, are also reported. (**B**) MALDI–TOF mass spectrum of the glycopeptide collected from narrow-bore HPLC analysis of carboxamidomethylated SAL following digestion with trypsin. The signal recorded in the spectrum was assigned to a peptide containing a putative N-glycosylation site at one of the three putative N-glycosylation sites present within the protein sequence. The sugar composition of the N-linked glycopeptide is also reported. Hex, hexose; HexNAc, *N*-acetylhexosamine; Fuc, fucose.

trypsin specificity. However, the spectrum showed the presence of four clear unexpected MH$^+$ signals at m/z 981.9, 857.9, 1706.7 and 1834.9 which did not correspond with any of the theoretical values for the tryptic peptides. Therefore the digest was separated by reverse-phase HPLC, and the fractions showing the indicated mass values were subjected to Edman degradation. On the basis of the results obtained, the first signal was associated with the C-terminal peptide carboxamidomethylated peptide 160–168 (CAM), which demonstrated the existance of a C-terminal processing phenomenon, occurring either *in vivo* and/or during the purification procedure. The remaining MH$^+$ signals were attributed to peptides (43–49), (64–77)CAM$_2$ and (63–77)CAM$_2$, respectively, where the amino acid replacements Val^{45} → Ala, Ile^{48} → Val and Ala^{73} → Val occurred. The co-existence in this mixture of the corresponding peptides not presenting these substitutions was verified by MALDI–TOF MS and amino acid sequencing. Furthermore, mass analysis excluded the presence of other amino acid replacements and highlighted the occurrence of partial hydrolysis products. Among these, two distinct forms of the peptide (43–77)CAM$_2$ presenting MH$^+$ signals at m/z 4080.9 and 4066.6 were identified; the first corresponded to the primary structure deduced from the nucleotide sequence and the second bore the amino acid replacements Val^{45} → Ala, Ile^{48} → Val and Ala^{73} → Val. Complementary mass-mapping experiments carried out using endoprotease Glu-C as the proteolytic agent confirmed earlier results and allowed us to cover the entire protein sequence (1–168) (Table 1). Therefore these results clearly demonstrated that SAL is constituted of two distinct species: isoform A, representing the polypeptide chain reported in Figure 1, and isoform B, differing from the previous isoform by the presence of the three amino acid substitutions reported above, as suggested from results reported in Figure 2(B).

Similarly, ESI–MS analysis of the recombinant product yielded a mass value of 21499.2 ± 1.9 Da, in good agreement with the theoretical one (21498.0 Da) calculated on the basis of the entire amino acid sequence reported in Figure 1, bearing the His$_6$ tag and assuming that two of the cysteine residues were involved in a disulphide pair. In addition, the correct nature of the amino acid sequence was verified by peptide mapping experiments.

Post-translational modifications

Experiments on the redox state of the cysteine residues present in native and recombinant SAL demonstrated that two of the four -SH groups were titratable by Ellman's reagent. Therefore samples of protein were alkylated with iodoacetamide under denaturing conditions without treatment with dithiothreitol; the CAM groups introduced blocked the reactivity of the -SH groups, thus avoiding scrambling phenomena during enzymic digestion. In order to ascertain the nature of the disulphides and other post-translational modifications present in native and recombinant SAL, protein samples were desalted and digested with trypsin as described in the Materials and methods section. The resulting peptide mixtures were resolved by reverse-phase HPLC and the whole fractions were analysed by MALDI–TOF MS.

In the case of the native protein, most of the peptides obtained showed MH$^+$ in agreement with the mass values reported in Table 1. On the contrary, four fractions in the range 35–40 min showed the presence of different components with MH$^+$ at m/z 2285.5, 2413.5, 2313.9 and 2441.7, which were associated with peptides (64–77)CAM + (160–164) and (63–77)CAM + (160–164) of isoforms A and B, respectively, connected by disulphide bridges. Figure 4(A) reports the spectrum obtained from the peak eluting at 37.9 min. These findings were confirmed by reduction of an aliquot of these species with dithiothreitol followed by re-examination by MALDI–TOF MS, which showed the presence of the reduced fragments. Furthermore, the remaining material was subjected to automated Edman degradation leading to the amino acid sequences expected; no PTH-Cys was observed in the cycles corresponding to Cys^{160}, PTH-Cys was detected at the cycles corresponding to Cys^{68} while PTH-CysCAM was present at the cycles corresponding to Cys^{75}. These results, together with the observation that Cys^{141} was always contained in the peptides in carboxamidomethylated form, demonstrated that SAL is characterized by the presence of the disulphide bridge Cys^{68}–Cys^{160}, while Cys^{75} and Cys^{141} are both present in reduced form.

Similarly, the presence in the MALDI–TOF spectra of adjacent peaks differing by 203 mass units made the fraction containing glycopeptides immediately recognizable. In fact, the peak eluting at 28.3 min showed the spectrum reported in Figure 4(B). The signals at m/z 2132.8, 2335.6, 2538.8, 2741.9 and 2945.2 were associated with glycosylated forms of the peptide (50–59). On the basis of the known biosynthetic pathway of N-linked oligosaccharides and the molecular mass of the peptide moiety, these signals were interpreted as the fragment (50–59) bearing complex-type N-linked glycans. These were shown to contain up to four *N*-acetylglucosamine residues bound to the fucosylated pentasaccharide core linked to Asn^{53}. No signal corresponding to the unglycosylated peptide was observed in all of the fractions

analysed from the peptide mixture. In contrast, peaks at 10.9 and 46.2 min presented intense signals at m/z 1178.8 and 2418.5, that were attributed to fragments (33–42) and (95–115) respectively. No peaks related to glycosylated forms of these peptides were observed. These findings demonstrated that Asn^{38} and Asn^{99} do not exhibit any post-translational modifications.

On the basis of these results, the mass values reported for the native protein in Figure 4(B) were associated with two truncated (1–168) isoforms [the arithmetic difference calculated between the two mass values (ΔM) = 14 Da], having a single disulphide bridge between Cys^{68} and Cys^{160} and a fucosylated pentasaccharide core linked to Asn^{53} and bearing between zero and four *N*-acetylglucosamine residues. Therefore the theoretical values calculated on the basis of the structures determined in the present study were in perfect agreement with the mass values experimentally measured by ESI–MS (isoform A: $FucHexNAc_2Hex_3$, 20222.1 Da; $FucHexNAc_3Hex_3$, 20425.2 Da; $FucHexNAc_4Hex_3$, 20628.3 Da; $FucHexNAc_5Hex_3$, 20831.4 Da; $FucHexNAc_6Hex_3$, 21034.5 Da. Isoform B: $FucHexNAc_2Hex_3$, 20208.1 Da; $FucHexNAc_3Hex_3$, 20411.2 Da; $FucHexNAc_4Hex_3$, 20614.3 Da; $FucHexNAc_5Hex_3$, 20817.4 Da; $FucHexNAc_6Hex_3$, 21020.5 Da. Where HexNAc corresponds to *N*-acetylhexosamine, Hex corresponds to hexose and Fuc corresponds to fructose).

Parallel MS/Edman degradation experiments on recombinant SAL detected in the tryptic digest the presence of fragments (64–77)CAM + (160–164) and (63–77)CAM + (160–164), connected by the disulphide bridges Cys^{68}–Cys^{160}, as well as peptides (135–148)CAM and (137–148)CAM. These results indicate that the cysteine residues are present in the recombinant protein in the same redox state as in the native SAL.

Secondary structure analysis

To explore the correct folding of the recombinant protein in comparison with native protein the secondary structure of both proteins was analysed by CD spectroscopy. The CD spectra of purified native SAL and His_6-tagged recombinant SAL were similar, both displaying one minimum at 218 nm, typical for β-sheet structures (results not shown). The minor differences between the two spectra, between 190 nm and 200 nm, were attributed to the presence of the N-terminal fusion tail in the recombinant protein. Secondary structure prediction, based on the analysis of the spectra between 200 nm and 240 nm with the programme k2d [43,44] returned 7% α-helix, 47% β-sheet and 45% random coil content for the native SAL and 6% α-helix, 48% β-sheet and 45% random coil content for the recombinant protein. This is consistent with the structural elements reported for other lipocalins [4].

The boar SAL contains a single tryptophan residue at position 22, part of the motif conserved in all lipocalins. Excitation of the native protein at 295 nm produced a fluorescence spectrum with maximal emission at 331 nm (results not shown). This indicated a highly hydrophobic environment for the tryptophan residue, as observed for porcine OBP-I [39,45]. In the recombinant protein the emission peak was found at 335 nm, suggesting a slightly less hydrophobic environment (results not shown). This was tentatively ascribed to minor differences in the folding of the two polypeptides, related to the absence of the glycan moiety or to the presence of the His_6 tag in the recombinant protein.

Ligand-binding

The ligand-binding properties of both the native and the recombinant SAL were assayed with the use of the fluorescent probe 1-aminoanthracene, already successfully employed with porcine OBP-I [45]. In this case, competitive binding experiments showed that 1-aminoanthracene bound at the same binding site as 2-isobutyl-3-methoxypyrazine and other odorants. When 1-aminoanthracene was added to a solution of native SAL, we did not observe any change in the fluorescent properties of the ligand, that could indicate specific interactions with the protein. On the hypothesis that binding of 1-aminoanthracene could be prevented by endogenous ligands present in the purified protein [3], native SAL was subjected to extensive dialysis before being used in binding assays. The dialysed sample showed binding activity to 1-aminoanthracene, that was measured by the intensity of the emission peak. In fact, bound 1-aminoanthracene emitted at the wavelength of 488 nm, whereas the emission λ_{max} of the free ligand occurred at 537 nm, with an increase in fluorescence of approx. 10-fold. These data, when compared with those obtained for OBP-I (fluorescence maximum at 481 nm and 80-fold intensity increase [45]), indicated a similar, but less hydrophobic binding pocket in the salivary protein. A binding curve was plotted by measuring the fluorescence intensity at 488 nm as a function of ligand concentration (Figure 5). We measured a binding constant of 12 μM, approx. one order of magnitude higher with respect to OBP-I (1.3 μM [45]). Parallel experiments, performed with a sample of the recombinant protein, yielded similar results, with a value for the dissociation constant of 9 μM (Figure 5).

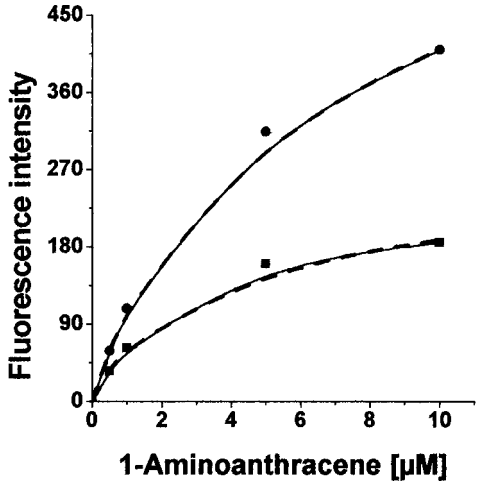

Figure 5 Binding curves of 1-aminoanthracene to native (●) and recombinant (■) SAL

The hyperbolic function to which data were fitted is represented by dashed lines. Dissociation constants were 12 μM and 9 μM respectively.

The binding of selected odorants was also probed both for the native and the recombinant protein, by measuring the displacement of 1-aminoanthracene from the complex caused by their addition. The ligands included odorants, such as 2-isobutyl-3-methoxypyrazine, 3,7-dimethyl-1-octanol and 2-phenylethanol, previously used with OBPs, as well as the natural pheromone 5α-androst-16-en-3-one. The competitive binding experiments indicated that all of the ligands tested did, to different extents, displace the fluorescent probe from the complex. However, the natural pheromone was by far the best ligand, with an affinity approx. one order of magnitude higher than odorants, such as 2-isobutyl-3-methoxypyrazine and 3,7-dimethyl-1-octanol, known to be good ligands for OBPs [46]. On the other hand, 2-phenylethanol, a poor ligand for OBPs, also showed

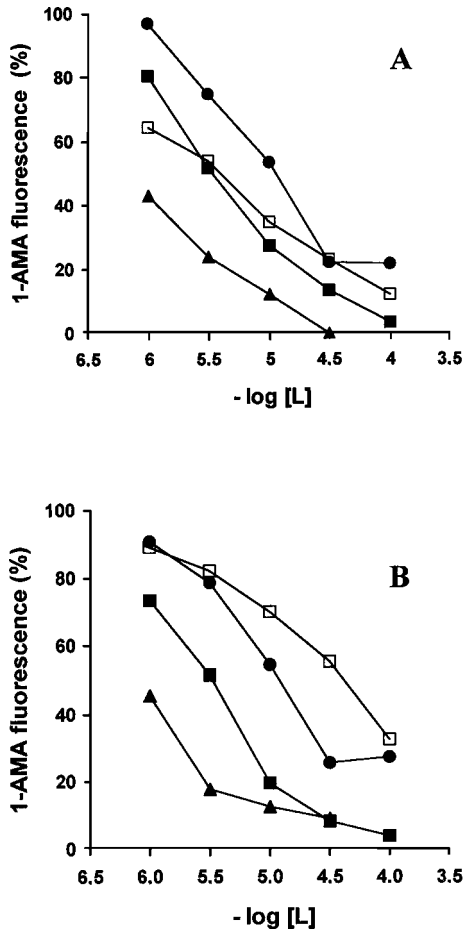

Figure 6 Competitive binding assays of native (A) and recombinant (B) SAL with various odorants

1-Aminoanthracene was used as the ligand and the decrease in fluorescence intensity was plotted as a function of competitive odorant concentration. 5α-Androst-16-en-3-one, filled triangles; 3,7-dimethyl-1-octanol, solid squares; 2-phenylethanol, solid circles; 2-isobutyl-3-methoxypyrazine, open squares.

only a weak affinity for SAL. In parallel experiments, the recombinant protein yielded very similar results (Figure 6).

Molecular modelling

The sequence identity in pairwise comparisons between SAL and members of the lipocalin family whose three-dimensional structure is known was approx. 40% or less. Nevertheless, on the basis of the similarity observed and secondary structure predictions, we hypothesize a lipocalin-fold for SAL. The schematic structural model obtained for isoform A consisted of the typical nine-stranded β-barrel containing two helices observed in other members of the lipocalin family (results not shown). Structural analysis in the region (120–135) seemed to reveal Lys^{126} as the major element responsible for the narrow turn present in SAL between sheet 8 and helix 2. This polypeptide segment was reported as crucial in the domain swapping observed in bovine OBP and is responsible for protein dimerization [40,41]. Primary structure alignment revealed high sequence identities in this region between SAL, MUP and porcine OBP-I, while in bovine OBP a deletion was present at position 126 (Figure 7). Therefore the presence of a longer and more flexible segment may allow a narrower turn in the molecular structure, thus generating the monomeric form observed in the case of native SAL. Furthermore, the presence in the protein of the disulphide bridge connecting Cys^{68} and Cys^{160}, conserved in MUP and porcine OBP-I, prevents any possible molecular rearrangment associated with domain swapping phenomena.

DISCUSSION

The complete amino acid sequence of SAL, obtained through the cloning of the corresponding cDNA, clearly assigns the protein to the lipocalin family. Members of this family display rather low levels of sequence conservation; therefore, several subclasses can be identified. A database similarity search using Basic Alignment Tool programs of the EMBL Heidelberg unix sequence analysis resources found best similarity values for SAL with the major horse allergen Equ c1 [47], the mouse MUP [19] and the rat α_2-urinary globulin [48], with amino acid identities ranging between 64% and 57% (Figure 7). Urinary proteins are pheromone

```
                1                              30                *  *           60          •     *
pig SAL         HKEAGQDVVTSNFDASKIAGEWYSILLASDAKENIEENGSMRVFVEHIRVLDNSSLAFKF-QRKVNGECTDFYAVCDKVG-DGVYTVAY
mus MUP-6          EEASSTGRNFNVEKINGEWHTIILASDKREKIEDNGNFRLFLEQIHVLEN-SLVLKF-HTVRDEECSELSMVADKTEKAGEYSVTY
rat α2-uG          EEASSTRGNLDVAKLNGDWFSIVVASNKREKIEENGSMRVFMQHIDVLEN-SLGFKF-RIKENGECRELYLVAYKTPEDGEYFVEY
pig OBP-I             QEPQPEQDPFELSGKWITSYIGSSDLEKIGENAPFQVFMRSIEFDDKESKVYLNFFSKENGICEEFSLIGTKQ-EGNTYDVNY
bov OBP             AQEEEAEQNLSELSGPWRTVYIGSTNPEKIQENGPFRTYFRELVFDDEKGTVDFYFSVKRDGKWKNVHVKATKQD-DGTYVADY
equ C-1         QQEENSDVAIRNFDISKISGEWYSIFLASDVKEKIEENGSMRVFVDVIRALDNSSLYAEY-QTKVNGECTEFPMVFDKTEEDGVYSLNY

                                  ----β1-----  --α1-        --- β2---    ----- β3---   ----- β4-----     -- β5-

                 90                              120                             150         •
pig SAL         YGENKFRLLEVNYSDYVILHLVNVNGDKTFQLMEFYGRKPDVEPKLKDKFVEICQQYGIIKENIIDLTKIDRCFQLRGSGGVQESSAE
mus MUP-6       DGFNTFTIPKTDYDNFLMAHLINEKDGETFQLMGLYGREPDLMSDIKERFAQLCEEHGILRENIIDLSNANRCLQARE
rat α2-uG       DGGNTFTILKTDYDRYVMFHLINFKNGETFQLMVLYGRTKDLSSDIKEKFAKLCEAHGITRDNIIDLTKTDRCLQARG
pig OBP-I       AGNNKFVVSYASETALIISNINVDEEGDKTIMTGLLGKGTDIEDQDLEKFKEVTRENGIPEENIVNIIERDDCPA
bov OBP         EGQNVFKIVSLSRTHLVAHNINVDKHGQTTELTELFVK-LNVEDEDLEKFWKLTEDKGIDKKNVVNFLENEDHPHPE
equ C-1         DGYNVFRISEFENDEHIILYLVNFDKDRPFQLFEFYAREPDVSPEIKEEFVKIVQKRGIVKENIIDLTKIDRCFQLRGNGVAQA

                 -----β6----   ---- β7----   ---- β8-----      ------α2-----       --β9--
```

Figure 7 Sequence alignment of porcine SAL isoform A with structurally related lipocalins

Amino acids identical in all aligned sequences are in bold. Cysteine residues involved in disulphide bridges are highlighted with circles. The three amino acids that differ between the SAL isoforms are marked by asterisks. Predicted secondary structure elements are indicated. *mus*, mouse; *bov*, bovine; *equ*, horse.

carriers, as reported in the Introduction section, and are secreted, at least in mouse [9], with the specific pheromone molecules tightly bound. These properties have previously been described for SAL [3] and further support the notion for analogous roles of these proteins in mouse and pig.

Additional structural similarities between SAL and urinary proteins concern the post-translational modifications. A glycan moiety is present in the proteins of both species. In SAL, a single oligosaccharide complex is linked to Asn^{53}, one of the three putative glycosylation sites within the polypeptide chain. It is difficult to speculate on the function of the carbohydrate chain, although it is clear that it does not affect ligand-binding properties, as shown by the similar binding data obtained for the recombinant protein, which lacks the glycan moiety. As in mouse MUP and rat α_2-urinary globulin, Cys^{68} and Cys^{160} are connected by a single disulphide bridge, while those at positions 75 and 141 are present in their reduced forms. Cys^{75} is not conserved in the other sequences and Cys^{141} is replaced by valine in horse Equ c1 (Figure 7).

It has been demonstrated for hamster aphrodisin, another member of the lipocalin family, that the non-volatile protein itself is essential for pheromonal function [2]. Also the MUPs, considered as pheromone carriers, exhibit pheromonal activity, even when devoid of their natural ligands [22]. That the protein itself may act as a pheromone has not been reported for pig SAL, but can be speculated upon, on the basis of structural similarity with these proteins. In the mouse the same physiological effects observed with MUP were also produced by a small peptide corresponding to the first six residues of the protein. The marked differences of MUP and SAL in their N-termini, as opposed to their general structural similarity, may suggest that species recognition is probably encoded in the first segment of these proteins.

The idea that the two SAL isoforms identified might be specific for the two pheromone components found to be associated with them is quite appealing and is supported by the observation that, in the model of this protein (see below) all three variable residues are located inside the binding pocket.

The expression of SAL in *E. coli* has paved the way to a detailed structural analysis of this protein and its mode of binding to the pheromone, through X-ray crystallography and NMR spectrometry. In fact, the absence of the glycan moiety and the His_6-tag at the N-terminus do not appreciably affect the properties of recombinat SAL as compared with the native protein. The similarity of the two proteins is documented by several observations: (1) the CD spectra of the two polypeptides are similar, indicating the same secondary structure; (2) the intrinsic fluorescence of the single tryptophan present in the protein indicates in both cases that this residue is located in a hydrophobic environment, i.e. inside the protein core, although not necessarily in the binding site; (3) binding curves for the fluorescent probe 1-aminoanthracene afford similar values for the dissociation constants; (4) ligand-binding specificity, measured with various ligands, is about the same in the two protein samples; in both cases the best ligand was the specific pheromone 5α-androst-16-en-3-one.

Binding properties of both native and recombinant SAL were determined spectroscopically using the fluorescent ligand 1-aminoanthracene. In the hydrophobic binding pocket of lipocalins, the emission peak of 1-aminoanthracene is shifted to a lower wavelength and its intensity increases by one or two orders of magnitude. Similar results were previously reported for porcine OBP-I [45]. The native protein, which contains steroidal compounds as endogenous ligands, only showed binding activity after these compounds were removed by displacement with high concentrations of 3,7-dimethyloctanol, a ligand of medium affinity, followed by extensive dialysis. The ligand 1-aminoanthracene binds to native SAL with a dissociation constant of 12 μM, sufficient to employ this fluorescent probe in binding experiments. The recombinant SAL binds 1-aminoanthracene with similar strength ($K_d = 9\ \mu M$), although the maximum fluorescence intensity was approx. half of the value observed with the native sample, which could be due to a quenching effect of the His_6-tag.

A three-dimensional model of the protein isoforms has been generated based on the crystallographic structure of other odorant- and pheromone-binding proteins. A single flat hydrophobic cavity was observed inside the β-barrel structure that should accomodate the polycyclic molecules often observed as natural and synthetic ligands. Among the three amino acids that vary between the two isoforms, two are observed inside the β-barrel ($Val^{45} \rightarrow Ala$ and $Ala^{73} \rightarrow Val$) (results not shown); the third residue ($Ile^{48} \rightarrow Val$) affects a position adjacent to the β-strands. The notion that the minor structural differences may lead to slight different ligand specificities is in line with the previous finding that the isolated native protein – probably both isoforms – contains two natural ligands [3].

This study was part of our continuing efforts to elucidate the structure–function relationships in classes of proteins involved in chemical communication. The functional role of the three amino acid replacements observed between SAL isoforms will be definitively understood after the resolution of the three-dimensional structure of these proteins and the mapping of their binding sites. Crystallographic studies on recombinant SAL, now in progress, will be facilitated by the data reported in the present study.

This work was supported by EEC grant no. BIO4-CT98-0420 to H.B. and P.P., by the Deutsche Forsuchgemeinschaft and the BMBF to H.B., by Italian National Research Council grants to L.F. and A.S., by MURST Ricerca Scientifica 40% grant to P.P. and by EEC Training and Mobility Program 'Access to Large-Scale Facilities' grant no. ERB FMGECT95 0061 to A.S.

REFERENCES

1 Cavaggioni, A., Findlay, J. B. C. and Tirindelli, R. (1990) Ligand binding characteristics of homologous rat and mouse urinary proteins and pyrazine binding protein of the calf. Comp. Biochem. Physiol. B **96**, 513–520
2 Singer, A. G., Macrides, F., Clancy, A. N. and Agosta, W. C. (1986) Purification and analysis of a proteinaceous aphrodisiac pheromone from hamster vaginal discharge. J. Biol. Chem. **261**, 13323–13326
3 Marchese, S., Pes, D., Scaloni, A., Carbone, V. and Pelosi, P. (1998) Lipocalins of boar salivary glands, binding odours and pheromones. Eur. J. Biochem. **252**, 563–568
4 Flower, D. R. (1996) The lipocalin protein family: structure and function. Biochem. J. **318**, 1–14
5 Pelosi, P. (1994) Odorant-binding proteins. Crit. Rev. Biochem. Mol. Biol. **29**, 199–228
6 Pelosi, P. (1996) Perireceptor events in olfaction. J. Neurobiol. **30**, 3–19
7 Finlayson, J. S., Asofsky, R., Potter, M. and Runner, C. C. (1965) Major urinary protein complex of normal mice: origin. Science **149**, 981–982
8 Dinh, B. L., Tremblay, A. and Dufour, D. (1965) Immunochemical study of rat urinary proteins: their relation to serum and kidney proteins. J. Immunol. **95**, 574–582
9 Bacchini, A., Gaetani, E. and Cavaggioni, A. (1992) Pheromone-binding proteins in the mouse mus musculus. Experientia **48**, 419–421
10 Jemiolo, B., Xie, T. M. and Novotny, M. (1992) Urine marking in male mice: responses to natural and synthetic chemosignals. Physiol. Behav. **52**, 521–526
11 Henzel, W. J., Rodriguez, H., Singer, A. G., Stults, J. T., Macrides, F., Agosta, W. C. and Niall, H. (1988) The primary structure of aphrodisin. J. Biol. Chem. **263**, 16682–16687
12 Marchlewska-Koj, A., Pochron, E. and Sliwowska, A. (1990) Salivary glands and preputial glands of males as source of estrus-stimulating pheromone in female mice. J. Chem. Ecol. **16**, 2817–2822

13 Shaw, P. H., Held, W. A. and Hastie, N. D. (1983) The gene family for major urinary proteins: expression in several secretory tissues of the mouse. Cell **32**, 755–761
14 Shahan, K. M., Denaro, M., Gilmartin, M., Shi, Y. and Derman, E. (1987) Expression of six mouse major urinary protein genes in the mammary parotid sublingual submaxillary and lachrymal glands and in the liver. Mol. Cell. Biol. **7**, 1947–1954
15 Spielman, A. I., Zeng, X. N., Leyden, J. J. and Preti, G. (1995) Proteinaceous precursors of human axillary odor: isolation of two novel odor-binding proteins. Experientia **51**, 40–47
16 Zeng, C., Spielman, A. I., Vowels, B. R., Leyden, J. J., Biemann, K. and Preti, G. (1996) A human axillary odorant is carried by apolipoprotein D. Proc. Natl. Acad. Sci. U.S.A. **93**, 6626–6630
17 Pes, D. and Pelosi, P. (1995) Odorant-binding proteins of the mouse. Comp. Biochem. Physiol., B: Comp. Biochem. **112**, 471–479
18 Utsumi, M., Ohno, K., Kawasaki, Y., Tamura, M., Kubo, T. and Tohyama, M. (1999) Expression of major urinary protein genes in the nasal glands associated with general olfaction. J. Neurobiol. **39**, 227–236
19 Garibotti, M., Navarrini, A., Pisanelli, A. M. and Pelosi, P. (1997) Three odorant-binding proteins from rabbit nasal mucosa. Chem. Senses **22**, 383–390
20 Ganni, M., Garibotti, M., Scaloni, A., Pucci, P. and Pelosi, P. (1997) Microheterogeneity of odorant-binding proteins in the porcupine revealed by N-terminal sequencing and mass spectrometry. Comp. Biochem. Physiol., B: Comp. Biochem. **117**, 287–291
21 Pes, D., Mameli, M., Andreini, I., Krieger, J., Weber, M., Breer, H. and Pelosi, P. (1998) Cloning and expression of odorant-binding proteins Ia and Ib from mouse nasal tissue. Gene **212**, 49–55
22 Mucignat-Caretta, C., Caretta, A. and Cavaggioni, A. (1995) Acceleration of puberty onset in female mice by male urinary proteins. J. Physiol. (London) **486**, 517–522
23 Krieger, J., Schmitt, A., Lobel, D., Gudermann, T., Schultz, G., Breer, H. and Boekhoff, I. (1999) Selective activation of G protein subtypes in the vomeronasal organ upon stimulation with urine-derived compounds. J. Biol. Chem. **274**, 4655–4662
24 Ryba, N. J. and Tirindelli, R. (1997) A new multigene family of putative pheromone receptors. Neuron **19**, 371–379
25 Herrada, G. and Dulac, C. (1997) A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. Cell **90**, 763–773
26 Matsunami, H. and Buck, L. B. (1997) A multigene family encoding a diverse array of putative pheromone receptors in mammals. Cell **90**, 775–784
27 Katkov, T., Booth, W. D. and Gower, D. B. (1972) The metabolism of 16-androstenes in boar salivary glands. Biochim. Biophys. Acta **270**, 546–556
28 Booth, W. D. and White, C. A. (1988) The isolation, purification and some properties of pheromaxein, the pheromonal steroid-binding protein, in porcine submaxillary glands and saliva. J. Endocrinol. **118**, 47–57
29 Booth, W. D. and von Glos, K. I. J. (1991) Pheromaxein, the pheromonal steroid-binding protein, is a major protein synthesized in porcine submaxillary salivary glands. J. Endocrinol. **128**, 205–212
30 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor
31 Freitag, J., Beck, A., Ludwig, G., von Buchholtz, L. and Breer, H. (1999) On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). Gene **226**, 165–174

Received 24 February 2000/18 May 2000; accepted 26 June 2000

32 Ausubel, F. M., Brent, R., Kingston, R. F., Moore, D. D., Seidmann, J. G., Smith, J. A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene and Wiley, New York
33 Engler-Blum, G., Meier, M., Frank, J. and Müller, G. A. (1993) Reduction of background problems in nonradioactive Northern and Southern blot analyses enables higher sensitivity than ^{32}P-based hybridizations. Anal. Biochem. **210**, 235–244
34 Loebel, D., Marchese, S., Krieger, J., Pelosi, P. and Breer, H. (1998) Subtypes of odorant binding proteins: heterologous expression and assessment of ligand binding. Eur. J. Biochem. **254**, 318–324
35 Kyhse-Andersen, J. (1984) Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. J. Biochem. Biophys. Methods **10**, 203–209
36 Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R. and Pelosi, P. (1999) Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. Eur. J. Biochem. **262**, 745–754
37 Thompson, J. D., Higgins, G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680
38 Böcskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, E. V., Cavaggioni, A., Findlay, J. B. C. and North, A. C. T. (1992) Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. Nature (London) **360**, 186–188
39 Spinelli, S., Ramoni, R., Grolli, S., Bonicel, J., Cambillau, C. and Tegoni, M. (1998) The structure of the monomeric porcine odorant binding protein sheds light on the domain swapping mechanism. Biochemistry **37**, 7913–7918
40 Bianchet, M. A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S. H., Monaco, H. L. and Amzel, L. M. (1996) The three dimensional structure of bovine odorant-binding protein and its mechanism of odor recognition. Nat. Struct. Biol. **3**, 934–939
41 Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. and Cambillau, C. (1996) Domain swapping creates a third putative combining site in bovine odorant binding protein dimer. Nat. Struct. Biol. **3**, 863–867
42 Peitsch, M. C. (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochem. Soc. Trans. **24**, 274–279
43 Andrade, M. A., Chacon, P., Merelo, J. J. and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng. **6**, 383–390
44 Merelo, J. J., Andrade, M. A., Prieto, A. and Moran, F. (1994) Proteinotopic feature maps. Neurocomputing **6**, 1–12
45 Paolini, S., Tanfani, F., Fini, C., Bertoli, E. and Pelosi, P. (1999) Porcine odorant-binding protein: structural stability and ligand affinities measured by fourier-transform infrared spectroscopy and fluorescence spectroscopy. Biochim. Biophys. Acta **1431**, 179–188
46 Dal Monte, M., Centini, M., Anselmi, C. and Pelosi, P. (1993) Binding of selected odorants to bovine and porcine odorant-binding proteins. Chem. Senses **18**, 713–721
47 Gregoire, C., Rosinski-Chupin, I., Rabillon, J., Alzari, P. M., David, B. and Dandeu, J. P. (1996) cDNA cloning and sequencing reveal the major horse allergen Equ c1 to be a glycoprotein member of the lipocalin superfamily. J. Biol. Chem. **271**, 32951–32959
48 Gao, F., Endo, H. and Yamamoto, M. (1989) Length heterogeneity in rat salivary gland alpha 2 mu globulin mRNAs: multiple splice-acceptors and polyadenylation sites. Nucleic Acids Res. **17**, 4629–4636