cDNA Cloning and Sequencing Reveal the Major Horse Allergen Equ c1 to Be a Glycoprotein Member of the Lipocalin Superfamily

Christophe Gregoire, Isabelle Rosinski-Chupin, Jacques Rabillon, Pedro Alzari, Bernard David, Jean-Pierre Dandeu

To cite this version:
Christophe Gregoire, Isabelle Rosinski-Chupin, Jacques Rabillon, Pedro Alzari, Bernard David, et al.. cDNA Cloning and Sequencing Reveal the Major Horse Allergen Equ c1 to Be a Glycoprotein Member of the Lipocalin Superfamily. Journal of Biological Chemistry, American Society for Biochemistry and Molecular Biology, 1996, 271 (51), pp.32951-32959. 10.1074/jbc.271.51.32951. pasteur-03136622

HAL Id: pasteur-03136622
https://hal-pasteur.archives-ouvertes.fr/pasteur-03136622
Submitted on 9 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
The gene encoding the major horse allergen, designated *Equus caballus* allergen 1 (Equ c1), was cloned from total cDNA of sublingual salivary glands by reverse transcription-polymerase chain reaction using synthetic degenerate oligonucleotides deduced from N-terminal and internal peptide sequences of the glycosylated hair dandruff protein. A recombinant form of the protein, with a polyhistidin tail, was expressed in *Escherichia coli* and purified by immobilized metal affinity chromatography. The recombinant protein is able to induce a passive cutaneous anaphylaxis reaction in rat, and it behaves similarly to the native Equ c1 in several immunological tests with allergic patients' IgE antibodies, mouse monoclonal antibodies, or rabbit polyclonal IgG antibodies. Amino acid sequence identity of 49–51% with rodent urinary proteins from mice and rats suggests that Equ c1 is a new member of the lipocalin superfamily of hydrophobic ligand-binding proteins that includes several other major allergens. An RNA blot analysis demonstrates the expression of mRNA Equ c1 in liver and in sublingual and submaxillary salivary glands.

Exposure to animal danders, commonly present in the environment, is known to be a frequent cause of allergy. The inhalation of these potent animal dandruff allergens induces immunoglobulin E antibody (IgE) and subsequent development of asthma in atopic individuals. Among these allergens, a major allergen is defined to be the one that elicits an anaphylactic reaction in a majority of patients, presenting an immediate hypersensitivity response mediated by IgE against the basic raw material (1).

The reasons why a protein is allergenic are not clearly understood to date, although several authors favor the hypothesis of a possible relationship between the structure and the function of proteins and their allergenicity (2). The enzymatic activity of certain proteins has been assumed to have a capacity to enhance the IgE response (2). A family of proteins, the lipocalin superfamily, is known to include several allergens, such as the mouse major urinary protein mMUP (3), the rat α-2-microglobulin (rA2U) (4), the bovine β-lactoglobulin (βlg) (5), the cockroach allergen Bla g4 (6), and the recently described bovine dander allergen Bos d2 (7). Based on this observation, Arruda et al. suggested that lipocalins may contain a common structure that is able to induce the IgE response. Members of this superfamily, which bind or transport small hydrophobic molecules, are generally expressed in the liver and/or secretory glands. This is particularly true for the mMUP and rA2U proteins, which are multigenic families at about 35–40 members in the case of the mMUP family (8) and about 25 for the rA2U (9, 10). These members are differentially expressed in the liver as well as salivary, lacrimal, and other secretory glands (11).

The major horse allergen, Equ c1, is a potent allergen responsible for about 80% of anti-horse IgE antibody response in patients who are chronically exposed to horse allergens. Although much work has been carried out on the isolation and identification of the horse allergenic agents responsible for human hypersensitivity response (12–16), the major horse allergen was only recently purified from hair and dandruff (17).

A previous study by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectric focusing-PAGE showed that Equ c1 appears as a single polypeptide with a relative molecular mass of 21,500 daltons and a pI of 3.9. The purification of Equ c1 allowed the sequencing of the 27 N-terminal amino acids and of internal peptides (18).

To obtain more information on the structural and functional features of Equ c1, we have cloned the corresponding cDNA from the sublingual salivary gland (SLG). Here we report the molecular cloning and sequencing of this cDNA and expression of a recombinant allergen rSLG Equ c1 in a bacterial system. The recombinant protein was compared with natural Equ c1 for its recognition by antibodies raised against the natural Equ c1 in immunoblots and in inhibition/competition enzyme-linked immunosorbent assay (ELISA). We also show that the recombinant protein is able to elicit a rat mast cell degranulation by passive cutaneous anaphylaxis reaction.

Sequence comparisons reveal that Equ c1 is a new member of the lipocalin superfamily.

**EXPERIMENTAL PROCEDURES**

**Materials**—The horse salivary glands were obtained from a slaughterhouse and rapidly frozen in liquid nitrogen after dissection. They were stored at –80 °C until protein and nucleic acid extractions were performed.

**Protein Purification and N-terminal Sequencing**—Equ c1 was purifi-
fied from salivary glands and dander extracts by a combination of size exclusion chromatography in fast protein liquid chromatography (FPLC) and hydrophobic interaction chromatography as described previously (17).

An Equ c 1 tryptic proteolysis was performed for 15 min at 37 °C in a buffer containing 50 mM Tris-HCl, 1 mM CaCl₂, pH 7.0, with an enzyme ratio of 1:1000 (w/w). The sequencing was processed, using the method described by Baw et al. (19), in the microsequencing laboratory of the Pasteur Institute. Protein assays were performed with the colorimetric method using Micro BCA protein assay reagent from Pierce, according to Smith et al. (20).

Preparation of RNA—Total RNA was isolated from sublingual (SLG) and submaxillary (SMG) salivary glands and from liver according to Chirgwin’s protocol (21), modified as described previously (22, 23).

**Equ c 1 cDNA Cloning—** cDNA first strand synthesis was performed on 5 μg of horse SLG total RNA for 1 h at 37 °C in a total volume of 50 μl with 20 pmol of the primer adapter oligo(dT): 5'-AACCCGGCTCGAGGACCAGGCTGCTGCTGCTTTTTTCAAAGGTTGTAAGAT-3', 800 units of Moloney murine leukemia virus reverse transcriptase (Life Technologies, Inc.) in the manufacturer’s buffer. The cDNAs so obtained were amplified by polymerase chain reaction (PCR) with the Opti Prime PCR optimization kit (Stratagene), with the oligomer 5'-GGG GAG TGG TAY AGY ATY TT-3' as primer 1 and the oligomer 5'-GGG GAG TGG TAY ATY ATY TT-3' as primer 2 derived from the Gly³⁵-Sep²⁰⁹ sequence and the 5'-GTS AGP ACR TPR ATR TTY TTY-3' as primer 3 derived from the Glu³⁵-Leu⁷⁹ sequence. The letter Y represents a 50% mixture (w/w) of nucleotides T and C, S a mixture of G and C, and R a mixture of A and G. After a first denaturation cycle at 98 °C for 2 min, 30 cycles of PCR consisting of a 30-s denaturation step at 94 °C followed by annealing at 50 °C for 35 s and elongation at 72 °C for 30 s were carried out in a thermocycler Hybaid (Ceralabo, Aubervilliers, France). Each reaction contained 1 μl of cDNA reaction product, 0.2 mM dNTP, 2.4 units of Taq DNA polymerase, 68.8 pmol of the primer 3, and 34.4 pmol of each other primer. The variable parameters of buffers are pH, MgCl₂, and KCl concentrations. The best amplification was obtained with buffer 6 (10 mM Tris-HCl, pH 8.8, 1.5 mM MgCl₂, and 75 mM KCl) and buffer 12 (10 mM Tris-HCl, pH 9.2, 3.5 mM MgCl₂, and 75 mM KCl). After separation by electrophoresis in a 1.2% agarose gel and purification, the products from the PCR reactions were inserted in pMOS Blue T vector (Amerham LifeSciences). Sequencing was performed after alkaline denaturation by the dideoxy chain termination method (24) using Sequenase version 2.0 (U.S. Biochemical Corp.) and α-³⁵S-labeled dATP.

**Amplification of the cDNA Ends—** The rapid amplification of cDNA ends (RACE) strategy was applied to clone 3' and 5' cDNA extremities. For 5' RACE, 12.5 μl of the first single strand cDNA (as described above) were directly used for dC tailing, for 5 min at 37 °C, in 10 mM Tris-HCl, pH 8.4, 25 mM KCl, 1.25 mM MgCl₂, 50 μM BSA, and 10 units of terminal transferase. Reactions were stopped by increasing the temperature to 65 °C for 10 min. The cDNA amplification was performed in the presence of 5 pmol of the oligomer 5'-GCC CGA GAT GTG AGT-3' corresponding to the nucleotide sequence G³⁵-C³⁸ and the primer adapter oligo(T). All amplified products were cloned in pMOS Blue vector and sequenced as described above.

**Bacterial Expression of Recombinant Equ c 1—** A cDNA corresponding to the nearly complete Equ c 1 sequence was amplified by PCR and cloned in a pET vector. Primers for PCR were designed to specifically hybridize with Equ c 1 cDNA and contained EcoRI and XhoI sites. The primers used were 5'-CTT GAA TTC ATC GAG GGA AGA GAA GAC AGAT GGT GCG GCG-3' (5' end primer) and 5'-CCA CTC GAG GAA GTA TCC ACT GTC-3' (3' end primer). In addition, the 5' primer provides the recombinant protein with a new proteolytic cleavage site for the factor Xa. PCR products were cloned into the EcoRI/XhoI sites of the plasmid pET 28 (a) under control of the T7 lac promoter (Fig. 1). This expression vector contains the kamcyanin resistance gene and a His⁸ tag at the N terminus of the recombinant protein. Competent Escherichia coli XL1 cells were transformed, and supercoiled plasmid was sequenced and transfected in E. coli BL 21 (DE3). Induction was performed by adding isopropyl β-D-thiogalactopyranoside to the medium at a final concentration of 1 mM for 180 min at 37 °C. Induction was controlled by taking aliquots every 30 min. Cells were then harvested by centrifugation and resuspended in 50 mM Tris-HCl, pH 7.0, containing 1% (v/v) Triton X-100 and 100 μg/ml lysozyme. The cells were incubated for 15 min at 30 °C, and the DNA was disrupted by sonication. The supernatant obtained after centrifugation was filtered on a 0.2-μm membrane and dialyzed against phosphate-buffered saline (PBS) with 0.5 mM NaCl. The resulting product was used for chromatographic purification.

**Purification of the Recombinant Equ c 1—** An HR 5/5 column was packed with chelating Sepharose fast flow (Pharmacia Biotech, Inc.), washed according to the manufacturer’s suggestions, and charged until saturated with metal ions from a 0.5% (w/v) copper(II) chloride solution. After thorough rinsing with water, the column was presaturated with buffer (PBS/0.5 mM NaCl) containing 10 mM imidazole (25). After equilibration of the column with the starting buffer (PBS/0.5 mM NaCl), 6 column volumes of supernatant were loaded, and the unbound material was collected. Competitive elution was carried out using imidazole at 40 and 120 mM (PBS/0.5 mM NaCl), pH 7.0, collecting 6 column volumes at each step (26). The whole process was controlled by an FPLC apparatus (Pharmacia). The fractions were concentrated using stirred cell ultrafiltration with a PM 10 membrane (Amicon) and dialyzed against the proteolytic buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM CaCl₂). Digestion with the factor Xa was performed overnight at 30 °C. After proteolysis, the digest was dialyzed to remove the small digest peptides and lyophilized.

**SDS-PAGE and Western Blot—** All analysis of the different fractions was performed with the Adjustable Stab Gel kit ASC 400 (Phalan) using 18% acrylamide/bisacrylamide (29:1) gels (27). Proteins were visualized with Coomassie Blue and/or silver nitrate staining. Electroblotting experiments were performed using nitrocellulose membrane (Schleicher & Schull). For immunological detection, polyclonal antibodies from human and rabbit sera and mouse monoclonal antibody directed against Equ c 1 were used.

The rabbit immunization was performed by intradermal injection of 100 μg of pure allergen. Sixteen patients with established allergy to natural Equ c 1 were selected, and a pool from three nonallergic healthy donors was used as negative control. Bound IgE were detected using peroxidase conjugated to rabbit anti-human IgE. When mouse mAb anti-Equ c 1 or the polyclonal rabbit IgG was used, the detection was...
performed with peroxidase conjugated to rabbit anti-mouse IgG or peroxidase conjugated to goat anti-rabbit IgG, respectively, using the diaminobenzidine tetrachloride as a specific reagent.

The mouse anti-HD Equ c1 mAb were prepared in the Hybrid lob of the Pasteur Institute according to the methods described by Kohler and Milet in (28).

Passive Cutaneous Anaphylaxis—Each mouse was immunized subcutaneously at day 0 and boosted at days 21 and 35 with 5 μg of antigen (purified HD Equ c1 protein extract from horse hair dandruff, horse serum albumin, or ovalbumin) in the presence of 4% (w/w) Al(OH)3 in a physiological solution.

Each mouse was bled after being anesthetized, at day 42 by retro-orbital puncture in order to study IgG immune response. The IgG antibody titers were determined by the passive cutaneous anaphylaxis reaction in rats (29).

Samples were diluted in a physiological solution and 100-μl aliquots inoculated intradermally on the shaved back of Lewis rats. Twenty-four hours later, each rat was challenged by intravenous inoculation in the tail of 1 ml of a physiological solution containing 50 μg of antigen and 0.5% Evans blue. Thirty minutes later, rats were killed, and skin was excised for examination. The reciprocal of the highest dilution giving a bluing reaction of 10-mm diameter was taken as the passive cutaneous anaphylaxis titer.

Inhibition/Competition Experiments—These experiments were performed using ELISA as follows. Each well of the assay plate (Maxisorb, Nunc, Roskild, Denmark) was coated with 100 μl of a highly purified HD Equ c1 or rSLG Equ c1, 10 μg/ml in 0.1 M carbonate/bicarbonate buffer, pH 9.6. After saturation of the unoccupied sites with 0.5% BSA in PBS and appropriate washing, mAbs, after being previously preincubated 1 h at 37 °C with different dilutions of competitor, were added in duplicate to the sample-coated wells and incubated for 1 h at 37 °C. Bound mAb and rabbit antibodies were detected with peroxidase-conjugated rabbit anti-mouse IgG (Sigma) and peroxidase-conjugated goat anti-rabbit IgG, respectively, and revealed with o-phenylenediamine according to the manufacturer’s recommendations.

Determination of Sugar Content—A study was done to perform de-glycosylation on Equ c1, using anhydro trifluoromethanesulfonic acid, as described by Sojar and Bahl (30). Each dry sample was acid-treated with a mixture of trifluoromethane sulfonic acid and toluene for 4 h at 20 °C. Then trifluoromethane sulfonic acid was neutralized by adding to the reaction mixture pyridine and ammonium bicarbonate, and dialyzed against 50 mM Tris/HCl, pH 7.5, 100 mM NaCl. Each sample was submitted to electrophoresis in SDS-PAGE. Gels were stained with silver nitrate. Analysis of the saccharide composition of the HD Equ c1 and Saliva Equ c1 was done using gas phase chromatography after acidic treatment, as described by Kamerling et al. (31).

RNA Analysis—Total mRNA was electrophoresed in an agarose/formaldehyde gel (32) transferred to a nylon membrane, and hybridized with the Equ c1 cDNA probe. The probe was the full-length cDNA insert labeled by the random priming method (33).

The search for homologies between the deduced amino acid sequence of Equ c1 and the proteins of the Swiss-Prot data base or the Equ c1 cDNA and the GenBank™ nucleotide sequence database were done, respectively, with the FASTP and FASTN program according to Altschul et al. (34).

RESULTS

Molecular Cloning of the Equ c1 cDNA—Tryptic fragments were generated from HD Equ c1 isolated and purified from horse hair dandruff extract by a combination of size exclusion chromatography and hydrophobic interaction chromatography. These fragments were microsequenced, and two of them (shown in boldface type on Fig. 2) were used to design three degenerate primers. The design of the primers took into consideration the codon usage in horse.

It was previously demonstrated by Dandeau et al. (17) that Equ c1 from different sources, i.e. saliva, urine, and hair dandruff extracts, are similarly recognized by antibodies. Salivary secretions contain the highest amount of Equ c1 protein; therefore, the salivary glands were chosen to clone Equ c1 cDNA. Among the tested salivary glands, the sublingual glands had the highest level of Equ c1 immunoreactivity and were selected to prepare mRNA.

The mRNAs so obtained were reverse-transcribed, and the Equ c1 cDNA was amplified by PCR using a mixture of the three primers. This reverse transcription-PCR resulted in a DNA fragment of about 400 base pairs in length that was cloned in pMOS Blue; several positive clones were sequenced. In a second step, 5’ and 3’ ends of the SLG Equ c1 cDNA were obtained using a 5’ and 3’ RACE strategy. The two amplification products of 250 and 450 base pairs for the 5’ and 3’ RACE, respectively, were cloned and sequenced.

Sequence of the Equ c1 cDNA—The full-length sequence of Equ c1 cDNA and the deduced amino acid sequence are shown in Fig. 2. The SLG Equ c1 cDNA is 923 nucleotides long with an open reading frame of 560 nucleotides (excluding the stop codon), coding for a 187-amino acid protein. All peptides from HD Equ c1 can be localized in the SLG Equ c1 sequence and start after an arginine or a lysine residue, according to the tryptic proteinolysis consensus sites. However, some differences in the amine acid sequence can be observed between rSLG Equ c1 from sublingual salivary gland and the tryptic peptides obtained from HD Equ c1. These differences are not PCR artifacts, because our nucleotide sequence results from the analysis of 12 clones from four independent PCR experiments. These differences are in the internal peptides, at positions 62 (Ala/Leu), 90 (Phe/Ala), 136 (Phe/Leu), 146 (Ser/Asp), 172 (Lys/Gln), and 173 (Ile/Thr). All analyzed clones contained a 3’ noncoding region of 298 nucleotides and a poly(A) tail 23 base pairs, downstream from a consensus polyadenylation signal AATAAA at position A666/A693. All clones sequenced have identical 5’ ends with a noncoding region of 63 nucleotides and with an open reading frame beginning at A34.

Analysis of the deduced amino acid sequence revealed that the 5’ end of the coding region contains a typical signal sequence (35) (Fig. 3A). According to the Von Heijne weight matrix method (36), a favored putative signal peptide cleavage site can be assigned between the Ala137 and Gln138 residues, generating a protein beginning with QQEENSDVAI. In contrast, the N-terminal end of the protein initially purified from hair dandruff (SDVAI) would result from a cleavage between Asn20 and Ser21, which is not predicted by Von Heijne’s rules. Equ c1 was purified from saliva, and the microsequencing of its N-terminal peptide revealed a mixture of three sequences, one of them beginning at the predicted Gln16, but the others at Gln22 and Ser21, respectively (Fig. 3B). Whether these N-terminal ends are due to cleavage by signal peptidase at different sites or are generated by proteolytic processing of the secreted protein is not known. Such heterogeneous N-terminal ends were also reported for human tear albumin (37), another member of the lipocalin superfamily.

Excluding the putative signal peptide, the protein contains two cysteine residues at positions 83 and 176. In a previous study, we observed an increase in the apparent molecular mass of Equ c1 from 21,500 to 25,000 daltons in SDS-PAGE gels under reducing conditions, indicating that these two cysteines could form a disulfide bridge. Equ c1 is highly rich in charged residues and aromatic residues. The calculated pI is 4.57, a value close to that determined by Dandeau et al. Glycosylation of Equ c1—Two putative N-glycosylation sites are present at positions Asn63 and Asn68. A Glycosylation of HD and SLG Equ c1 was confirmed by gas phase chromatography, which revealed the presence of approximately 8.6% (w/w) of carbohydrates, representing 1,850 daltons. These results could explain the decrease in apparent molecular weight of Equ c1 in SDS-PAGE (Fig. 4) and the modification of the pI after deglycosylation.

Analysis of the sugar residue composition in Table I shows the presence of GaINAc, Gal, NeuAc, GlicNAc, and Man. Carbohydrates attached to proteins can be classified into two groups, N-glycans and O-glycans. All N-glycans contain a common structure, Manα1→6(Manα1→3)Manβ1→4GlicNAcβ1→
4GlcNAc→Asn, called the trimannosyl core. Molecular ratio results (second column in Table I) indicate unambiguously the presence of this core and, therefore, the presence in the glucidic part of Equ c1 of one N-glycan member of the biantennary complex type that contains three mannose residues. One N-acetyl-lactosamine (Gal\(\beta_1\)4GlcNAc) is attached to the outer two α-mannose residues, followed by sialic acid residues (for Equ c1) or additional N-acetyllactosamines (38).

The presence of GalNAc only found in the O-glycan components, except for several hormones (38), suggests that the protein is also O-glycosylated.

Expression of Equ c1 as a Recombinant Protein—A recombinant protein, starting at Glu19, was produced in a bacterial system, after cloning of the corresponding cDNA sequence in a pET 28 plasmid. This plasmid allows bacterial expression of a recombinant protein with a 40-amino acid polypeptide tail containing a polyhistidine tag to its N-terminal end (Fig. 1). To allow the production of a recombinant protein without any added amino acid, a factor Xa proteolytic site (LEFIEGR2ENSDVA) was inserted between the tail and the recombinant protein.

Two recombinant clones were tested for rSLG Equ c1 expression. Optimal production was obtained after a 150-min induction by isopropyl β-D-thiogalactopyranoside. A protein determination assay showed that rSLG Equ c1 represents about 30% of the total bacterial protein. This protein was essentially present in the supernatant of the bacterial extracts. A single purification step by immobilized metal affinity chromatography was sufficient to obtain pure rSLG Equ c1, which migrates as a single band of 19.5–20 kDa in an 18% SDS-PAGE gel (Fig. 4, lane A) after cleavage by factor Xa. This molecular mass is compatible with the calculated mass of 19,469 daltons and rather similar to that of deglycosylated natural Equ c1.

Antigenicity of the Recombinant Protein—The recombinant protein was tested for its antigenic recognition by different antibodies raised against HD Equ c1, i.e. three mouse monoclonal antibodies (mAbs 118 and 197, which recognize two different linear epitopes, and mAb 220, which recognizes a conformational epitope), 2 mouse and rabbit polyclonal antibodies (IgG), and human IgE from the sera of 16 patients suffering from horse allergic reactions (characterized in Ref. 17).

Immunoblot analysis after SDS-PAGE (Fig. 5), performed on the total bacterial extract, shows that the three mAbs bind a...
cDNA Cloning and Sequencing of the Major Horse Allergen Equ c1

The sugar content was determined by gas phase chromatography on pure HD/SLG Equ c1 (31). The relative weight ratio was given for each monosaccharide. The molecular ratio (column 2) was compared with the theoretical ratio for one N-glycan biantennary complex type given in parenthesis.

| Monosaccharide | Weight | Molecular ratio |
|----------------|--------|----------------|
| Man            | 1.8    | 3.0 (3)        |
| Gal            | 1.5    | 2.5 (2)        |
| GalNAc         | 0.53   | 0.7 (0)        |
| GlcNAc         | 2.6    | 3.5 (4)        |
| NeuAc          | 2.2    | 2.5 (2)        |

The biological activity of rSLG Equ c1 was also tested by passive cutaneous anaphylaxis on several rats as described under “Experimental Procedures.” The mouse sera were harvested after animal immunization with HD Equ c1, hair dandruff extract, or control proteins (horse serum albumin or ovalbumin). The results in Table II show that rSLG Equ c1 and HD Equ c1 are used as competitors; 100 and 50% inhibition are obtained with 20 µg/ml and 100 µg/ml, respectively, of either of them. This result suggests that the majority of the HD Equ c1 epitopes are present on the recombinant protein structure.
Homologies of Equ c1 with Proteins of the Lipocalin Superfamily—Homology searches in the sequence data bases show that Equ c1 has sequence similarities with other members of the lipocalin superfamily (Fig. 7). The best score was obtained with the mouse major urinary proteins clac1 MUP4, the cSmx1 MUP5 (cloned from lachrimal and submaxillar glands, respectively), and rA2U with homology ranging from 49 to 51% of identity and 76% of conservative mutations.

Sequence alignment shows that the two cysteines, Cys83 and Cys176, that form a disulfide bond in mMUP and rA2U, as well as the majority of other lipocains, are conserved (39). Only one potential N-glycosylation site, corresponding to position Asn253, is present in rA2U and is absent from the mMUP. The other site, at position Asn468, is specific to Eq c1 and is due to the insertion of a serine residue at position 69.

Three motifs, relatively well conserved among lipocains, have been described by Flower et al. (40). Two of these motifs are found in Eq c1 (Fig. 7). The most highly conserved amino acid sequences with the lipocain superfamily are Lys32, Gly35, Xaa36-Trp37-Tyr38, Ile40, Leu42-Ala43-Ser44-Asp45 in motif 1 and Arg141-Glu142-Pro143-Asp144, Ile149-Lys150-Glu151, Phe153 in motif 3 (41). The other conserved motif is TDY (structurally conserved region 2), while Phe109, Ile111, and Asp117 seem to be less conserved in the Eq c1 sequence. However, this motif is also absent from a number of true lipocain members, such as the human tear albumin (37), von Ebner’s gland protein (42), and hamster aphrodisin (43), and is less conserved in the bilin-binding protein (44), the α1-microglobulin (45), and rat odorant protein (46).

Tissue Expression of Eq c1 mRNA—To study the distribution of Eq c1 in the horse, total RNA was prepared from SLG and SMG salivary glands as well as from the liver, and it was analyzed by RNA blot hybridization (Fig. 8). Eq c1 mRNA was detected in each twice; however, the level in the SMG and liver is about 100 times lower than in the SLG. In addition, Eq c1 mRNA in liver seems to be slightly longer. Whether this is due to a true difference of size or to the presence of a longer poly(A) tail in liver Eq c1 mRNA was not investigated.

DISCUSSION

This paper reports the cloning, characterization, and expression in a bacterial system of the cDNA corresponding to a major horse allergen, Eq c1. This cDNA was cloned from the SLGs and some differences were noted between its deduced amino acid sequence and peptides generated from a protein purified from horse hair dandruff extract (HD Eq c1). Indeed, 6 amino acids out of 79 are different between the two sequences. Some of these changes are conservative. One likely explanation of these differences is that HD Eq c1 and SLG Eq c1 belong to the same multigenic family, whose members are tissue-specifically expressed, as was reported for rodent urinary proteins from mouse and rat (47). During the cloning of SLG Eq c1, we obtained no evidence for another member of this family being expressed in salivary sublingual glands; however, we cannot exclude the possibility that the choice of primers for reverse transcription-PCR might have favored the cloning of one cDNA only. An RNA blot study revealed the presence of mRNAs hybridizing with SLG Eq c1 cDNA in submaxillary glands and in liver too. Synthesis in the liver could explain the presence of Eq c1 in the horse’s urine (18), since it was reported for proteins of the MUP family in rat and mouse (48).

Despite the slight differences in their amino acid sequences and the absence of glycosylation in rSLG Eq c1, rSLG Eq c1 and HD Eq c1 are similarly recognized in our immunoblotting studies and inhibition/competition ELISA experiments. Moreover, the results obtained in inhibition/competition ELISA with three mAbs and with rabbit antibodies raised against HD Eq c1 suggest that all IgG epitopes of HD Eq c1 are also present in rSLG Eq c1, and thus in SLG Eq c1. In addition, at least
some of the IgE epitopes are also present in rSLG Equ c1, since rSLG Equ c1 is recognized by IgE from allergic patients in immunoblot experiments and binds to mouse IgE in passive cutaneous anaphylaxis experiments, resulting in the induction of a specific immediate hypersensitivity response in rats pre-sensitized with HD Equ c1. Together, these results suggest that neither the differences in amino acids nor the absence of glycosylation in the bacterially expressed protein affects the global conformation of the protein.

The search in the sequence data base revealed homology with members of the lipocalin superfamily, in particular with cLac1 MUP4 and cSmx1 MUP5. Members of this family share a common structure as was shown by the x-ray crystal structures of retinol-binding protein (49), b-lactoglobulin (50), and MUP (51). The folding architecture of lipocalins consists of an eight-stranded b-barrel followed by a single a-helix and a short C-terminal b-strand (Fig. 9). The eight anti-parallel strands are arranged in two orthogonal b-sheets that leave a small hydrophobic cavity within the barrel (52). This pocket is in a highly apolar environment, appropriate for binding and transport of small hydrophobic molecules through a hydrophilic medium. The binding pocket is entirely formed by aliphatic and aromatic side chains from the inner faces of the two b-sheets (these positions are indicated by arrows in the alignment shown in Fig. 7).

A structural model of Equ c1 (Fig. 9) was constructed from the x-ray coordinates of the mouse MUP1 model by Bocskei et al. (51) using the program QUANTA (MSI). This modeling was facilitated by the absence of amino acid insertions and deletions between the two proteins, with two exceptions: the insertion of Asp22 at the N terminus and Ser 69 in the b-hairpin loop between the second and the third strands of Equ c1. At positions where the two proteins differed, the amino acid sequence was substituted, and the side chains were rebuilt using stereoechemical criteria. The model was finally submitted to an overall energy minimization. As can be seen in Fig. 9, many of the amino acids of the presumed binding pocket (Ile63, Leu71, Phe109, Ile111, Leu124, Leu135, and Tyr139) are strictly conserved or have conservative amino acid substitutions in SLG Equ c1 when compared with rA2U/mMUP. The most noticeable differences are the substitution of Ala75 in Equ c1 by Leu/Phe in rA2U/mMUP and the substitution of Phe90 in the adjacent b-strand of Equ c1 by alanine. Although the hydrophobic character of the binding pocket is maintained, these
changes might modulate its shape and specificity.

In addition, the two possible N-glycosylation sites, which are not present in MUP1, are found in Equ c1 in exposed protein loops accessible to the solvent (Fig. 9), suggesting that the presence of an N-glycan does not interfere with the structure of the binding pocket. Moreover, the two cysteine residues that form a disulfide bridge linking the C-terminal part of the protein to the β-barrel (Fig. 9) in r2A2U/mMUP (Fig. 7) and in the majority of other lipocalins are also conserved in Equ c1 (positions 83 and 176).

This structural model, therefore, suggests that Equ c1 could adopt the same tertiary structure as that described for other lipocalins. The exact physiological role of Equ c1 has not been established yet. Its presence in the urine of adult mares and stallions and its absence in the urine of yearlings (18) suggests that Equ c1 is only synthesized at sexual maturity. Thus, its physiological role could be similar to that of rodent urinary protein of mice and rats (pheromone-binding protein) but not completely identical, since these two proteins are essentially produced in males.

Our results allow us to add Equ c1 to the list of lipocalins able to induce an IgE response, thus enhancing the hypothesis of Arruda (6) that lipocalins could have an intrinsic property to stimulate IgE production, thus enhancing the hypothesis of a cross-reactivity remains to be clearly established.

In this context, where some members of the lipocalin superfamily may have an intrinsic property to stimulate IgE production, the obtainment of a recombinant wild-type protein and of suitable mutants that can induce a biological activity will be an important tool to study the determinants involved in allergic reactions. Moreover, rSLGEquc1 may also help in the diagnosis of the allergic reaction to horses.

Acknowledgments—We thank Prof. F. Rougeon for continuous support and useful discussions and Dr. T. Fontaine for the determination of the monosaccharide composition. We also thank Dr. Bernadac for collection of horse saliva and hair dandruff extracts, and Dr. B. Laode for a critical reading of this manuscript.

REFERENCES
1. Lowenstein, H., Markussen, B., and Weeke, B. (1976) *Int. Arch. Allergy Appl. Immunol.* **51**, 38–47
2. Dudler, T., Machado, D. C., Kolbe, L., Annand, R. R., Rhodes, N., Gelb, M. H., Kost, K., Suter, M., and Ajello, R. A. (1995) *J. Immunol.* **155**, 2605–2613
3. Lorusso, J. R., Moffat, S., and Ohman, J. L. (1986) *J. Allergy Clin. Immunol.* **78**, 928–937
4. Wulla, A. F., and Longbottom, J. L. (1985) *J. Allergy Clin. Immunol.* **75**, 242–251
5. Ball, G., Shelton, M. J., Walsh, B. J., Hill, D. J., Hasking, C. S., and Howden, M. E. (1994) *Clin. Exp. Allergy* **24**, 758–764
6. Arruda, L. K., Vailes, L. D., Hayden, M. L., Benjamin, D. C., and Chapman, M. D. (1995) *J. Biol. Chem.* **270**, 31196–31201
7. Mantyjarvi, R. A., Rytkonen, M., Pentikainen, J., Rautiainen, J., Virtanen, T., Santan, H., and Laatikainen, R. (1996) *J. Allergy Clin. Immunol.* **97**, 212 (Abstr. 117)
8. Shaham, K., Gilmaritin, M., and Derman, E. (1987) *Mol. Cell. Biol.* **7**, 1958–1946
9. Unterman, R. D., Lynch, K. R., Nakhaski, H. L., Dolan, K. P., Hamilton, J. W., Cohn, D. V., and Feigelson, P. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3478–3482
10. Dolan, K. P., Unterman, R., McLaughlin, M., Nakhaski, H. L., Lynch, K. H., and Feigelson, P. (1982) *J. Biol. Chem.* **257**, 13527–13534
11. Gao, F., Endo, H., and Yamamoto, M. (1989) *Nucleic Acids Res.* **17**, 4629–4636
12. Storbeck, D. R. (1985) *J. Biochem.* **95**, 1823–1827
13. Ceska, M., and Hulten, E. (1972) *Int. Arch. Allergy Appl. Immunol.* **43**, 427–433
14. Porosius, G., Brandt, R., Hulten, E., and Yman, L. (1973) *Int. Arch. Allergy Appl. Immunol.* **44**, 679–691
15. Lowenstein, H. (1978) *Int. Arch. Allergy Appl. Immunol.* **57**, 349–357
16. Franke, D., Maasch, H. J., Wahl, R., Schultz-Werninghaus, G., and Breetting, H. (1990) *Int. Arch. Allergy Appl. Immunol.* **92**, 309–317
17. Dandeau, J. P., Rabillon, J., Dreyfus, M., Babinet, C., and Rougeon, F. (1975) *J. Chromatogr.* **107**, 23–31
18. Dandeau, J. P., Rabillon, J., Dreyfus, M., Babinet, C., and Rougeon, F. (1975) *EMBO J.* **4**, 983–987
19. Bauw, G., Van Damme, J., Puyve, M., Vandenkerckhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B., and Cels, J. E. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7701–7705
20. Smith, P. K., Krohn, R. I., Hermanson, G. T., Malia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. L. (1985) *Anal. Biochem.* **148**, 155–161
21. Chirgwin, J. M., Prybyla, E. A., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
22. Aufray, C., and Rougeon, F. (1986) *J. Biochem.* **109**, 303–314
23. Trenit, D., Dreyfus, M., Babinet, C., and Rougeon, F. (1987) *EMBO J.* **6**, 983–987
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
25. Sulkowski, E. (1985) *Trends Biochem. Sci.* **3**, 1–7
26. Casey, J. L., Keep, P. A., Chester, K. A., Robson, L., Hawkins, R. E., and Begent, R. H. (1995) *J. Immunol. Methods* **179**, 105–116
27. Laemmli, U. K. (1970) *Nature* **227**, 680–685
28. Kohler, G., and Micle, J. (1975) *Nature* **256**, 495–497
29. Ovary, Z., Caienza, S. B., and Kojima, S. (1975) *Int. Arch. Allergy Appl. Immunol.* **48**, 16–21
30. Sijar, H. T., and Bahl, O. P. (1987) *Methods Enzymol.* **138**, 341–350
31. Kamerling, J. P., Gerwig, G. J., Vliegenthart, J. F., and Clamp, J. R. (1975) *Biochem. J.* **151**, 491–495
32. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 5201–5205
33. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
34. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
35. Von Heijne, G. (1986) *J. Mol. Biol.* **184**, 99–105
36. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
37. Redl, B., Holzfeind, P., and Lottspeich, F. (1995) *J. Biol. Chem.* **270**, 20282–20287
38. Baenziger, J. U., and Green, E. D. (1988) *Biochim. Biophys. Acta* **947**, 287–306
39. Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990) *Proteins* **8**, 44–61
40. Flower, D. R., North, A. C., and Attwood, T. K. (1993) *Protein Sci.* **2**, 753–761
41. Nagata, A., Suzuki, Y., Igarashi, M., Eguchi, N., Toh, H., Urade, Y., and Hayashi, O. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4020–4024
cDNA Cloning and Sequencing of the Major Horse Allergen Equ c1

42. Schmale, H., Holtgreve-Grez, H., and Christiansen, H. (1990) Nature 343, 366–369
43. 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
44. Suter, F., Kayser, H., and Zuber, H. (1988) Biol. Chem. Hoppe-Seyler 369, 497–505
45. Kaumeyer, J. F., Polazzi, J. O., and Kotick, M. P. (1986) Nucleic Acids Res. 14, 7839–7850
46. Pevsner, J., Reed, R. R., Feinstein, P. G., and Snyder, S. H. (1988) Science 241, 336–339
47. MacInnes, J. I., Nozik, E. S., and Kurtz, D. T. (1986) Mol. Cell. Biol. 6, 3563–3567
48. Shahan, K., Denaro, M., Gilmartin, M., Shi, Y., and Derman, E. (1987) Mol. Cell. Biol. 7, 1947–1954
49. Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P. A. (1984) EMBO J. 3, 1451–1454
50. Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C., Findlay, J. B., Sivaprasadaro, R., Jones, T. A., Newcomer, M. E., and Kraulis, P. J. (1986) Nature 324, 383–385
51. Boczkai, Z., Green, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavaggioni, A., Findlay, J. B., and North, A. C. (1992) Nature 360, 186–188
52. North, A. C. (1989) J. Mol. Graph. 7, 67–70
53. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
54. Kabsch, W., and Sander, C. (1983) Biopolymers 22, 2577–2637