

Separation of horse dander allergen proteins by two-dimensional electrophoresis

Molecular characterisation and identification of Equ c 2.0101 and Equ c 2.0102 as lipocalin proteins

Vincent BULONE, Tonje KROGSTAD-JOHNSSEN and Berit SMESTAD-PAULSEN

Institute of Pharmacy, University of Oslo, Norway

(Received 31 October 1997) – EJB 97 1545/3

The aim of this work was to identify which proteins in horse dander extracts are allergens and to characterise them. Two-dimensional PAGE showed that horse dander preparations are composed of up to 50 proteins, all having acidic isoelectric points in the pH range 3–4.5. Immunoblots of two-dimensional PAGE were used to compare the reactivity of the proteins with IgE from 23 allergic patients. Patient sera were divided into two main groups recognising either allergens of 18.5 kDa or proteins of 27–29 and 31 kDa. The proteins of 27–29 kDa and 31 kDa were all N-glycosylated and their glycan chains seem to play a role in the binding of IgE from allergic patients. The sugar composition of their carbohydrate moiety was determined and lectin-binding experiments indicated presence of terminal sialic acid linked α -(2→6) to galactose, galactose linked β -(1→4) to N-acetylglucosamine, and possibly presence of sialic acid linked α -(2→3) to galactose. The 27–29-kDa glycoproteins had heterogeneous isoelectric points, most probably due to different degrees of sialylation in their oligosaccharide chains. The two 18.5-kDa allergens exhibited slightly different isoelectric points and their N-terminal sequences were identical, showing that they most likely were isoforms of the same protein. Sequence analyses revealed that their N-terminal sequences are similar to proteins belonging to the lipocalin family. We named the two 18.5-kDa proteins Equ c 2.0101 and Equ c 2.0102, according to International Allergen Nomenclature recommendations [King, T. P., Hoffman, D., Löwenstein, H., Marsh, D. G., Platts-Mills, T. A. E. & Thomas, W. (1995) *J. Allergy Clin. Immunol.* 96, 5–14]. The N-terminal of the allergens of 27–29 kDa were blocked and their sequences were not determined. Their amino acid compositions were determined and comparison with acidic mammalian proteins in the Swiss-Prot database revealed high scores with lipocalin proteins. This suggests that the glycosylated horse dander allergens belong to the lipocalin family, like Equ c 2.0101 and Equ c 2.0102.

Keywords: horse dander allergen; two-dimensional PAGE; carbohydrate analysis; N-terminal sequencing; lipocalin protein.

Allergy to horse dander is limited to individuals directly in contact with the animal. The molecules responsible for this allergy have been studied for about 50 years. It has first been demonstrated that horse albumin has allergenic properties and belongs to allergens found in horse hair and dander preparations [1]. Since then, other horse proteins from serum and dander have been shown to be responsible for human allergies [2, 3].

Several groups have described partial purification and characterisation of allergens from horse hair and dander [4–8], usu-

ally by a combination of biochemical and immunochemical methods. However, there is so far only one report on the purification to homogeneity of a horse dander major allergen named Equ c 1 [9]. This allergen was purified by a straightforward method involving hydrophobic-interaction chromatography. This allowed the determination of several of its biochemical properties. For instance, it has been shown that it is a single polypeptide chain with a molecular mass of 21.5 kDa and a pI of 3.9. The results of Gregoire et al. [10] suggest that it is a glycoprotein bearing both O- and N-linked oligosaccharides. This glycoprotein contains 8.6% of carbohydrates, representing 1.85 kDa. The gene encoding Equ c 1 was recently cloned and the complete sequence of the protein is available [10]. Sequence analysis revealed that Equ c 1 is a member of the lipocalin family, a large group of small, predominantly extracellular proteins responsible for the transport of small hydrophobic molecules [11]. Equ c 1 was described by Dandeu et al. [9] as the horse dander major allergen. However, previous reports strongly suggest that there are other major allergens in horse dander preparations [4, 5, 7, 8]. No sequence of these molecules is available and much remains to be done in this field to allow their clear identification.

Correspondence to V. Bulone, Centre de Recherches sur les Macromolécules Végétales, CNRS-UPR 5301, BP 53, F-38041 Grenoble cedex 9, France

Fax: +33 4 76 54 72 03.

E-mail: bulone@cermav.cnrs.fr

Abbreviations. Blast, basic local alignment search tool; HCA, hydrophobic cluster analysis; HPAEC-PAD, high-performance anion-exchange chromatography pulsed amperometric detector; PNGase F, peptide N-glycosidase F.

Enzyme. Peptide N-glycosidase F (EC 3.2.1.96).

Note. The novel sequence data presented here have been submitted to the Swiss-Prot database and are available under accession numbers P81216 (Equ c 2.0101) and P81217 (Equ c 2.0102).

The aim of this study was to characterise horse dander major allergens other than Equ c 1. The proteins from horse dander extracts were separated by two-dimensional PAGE and 16 allergens were identified. Some of them were found to be glycosylated and their carbohydrate content was analysed. Also, two non-glycosylated major allergens were partially sequenced and identified as new members of the lipocalin family. They were named Equ c 2.0101 and Equ c 2.0102, according to International Allergen Nomenclature recommendations [12].

MATERIALS AND METHODS

Chemicals. Carrier ampholytes (Pharmalyte) and protein molecular mass standards containing phosphorylase *b* (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa) and α -lactalbumin (14 kDa) were from Pharmacia Biotech. Nitrocellulose (0.45 μ m), polyvinylidene difluoride membranes (0.2 μ m) and all chemicals used for two-dimensional PAGE were from Bio-Rad. Peroxidase-conjugated rabbit anti-human IgE antibodies specific for ϵ -chains were purchased from Dakopatts A/S. Recombinant peptide N-glycosidase F (PNGase F) cloned from *Flavobacterium meningosepticum* and expressed in *Escherichia coli* was from Boehringer Mannheim. All other chemicals were from Sigma Chemical Co.

Biological material. Horse dander was obtained from Allergon AB. The raw material was extracted (5%, mass/vol.) with 125 mM ammonium bicarbonate pH 8.0 at 4°C for 20 h. Insoluble material was removed by centrifugation (13 000 $\times g$, 10 min) and the supernatant was filtered through two layers of Miracloth (Behring Diagnostics). The extracts were freeze-dried and stored at -20°C. Sera from 23 patients were provided by Nycomed Pharma A/S. All patients had a history of allergy to horse only, and their sera were of the radioallergosorbent test (RAST) class 3 and higher.

Two-dimensional polyacrylamide gel electrophoresis. Two-dimensional PAGE was performed using either the Protean Minicell system from Bio-Rad, or both the Desaphor gel tube electrophoresis apparatus (Desaga) for IEF and the Protean II xi cell (Bio-Rad) for SDS/PAGE. The technique described by Dunbar [13] was modified as follows. IEF gels contained 8 M urea, 6% acrylamide/bisacrylamide (30:1.8), 7.5% carrier ampholytes (Pharmalyte, pH 2.5–5.0) and 3% Nonidet P-40. The protein content of the horse dander extract was determined by the Bradford dye-binding assay [14]. Samples were prepared by mixing one part of the extract dissolved in distilled water (10 mg/ml) with one part of sample buffer containing 8 M urea, 40 mM Tris, 3% Nonidet P-40 and 2.5% 2-mercaptoethanol. After a 1-h incubation at room temperature, 50 μ g protein was loaded on each minigel and 7–10 times more material was loaded on the large gels. The cathode and anode buffers consisted of 100 mM NaOH and 15 mM phosphoric acid respectively. Mini and large IEF gels were run until pH equilibrium was reached, at a constant voltage of 750 V and 1000 V, respectively. These gels were then equilibrated for 10 min in 125 mM Tris/HCl pH 6.8 containing 2% SDS, 10% glycerol and 0.8% 2-mercaptoethanol. For second-dimension electrophoresis, SDS/PAGE was performed as described by Laemmli [15], using 15% acrylamide gels. Proteins were stained with either Coomassie brilliant blue R-250 or silver [13], and glycosylated proteins were identified by staining with Schiff's reagent [16].

Immunoblotting with patient sera. Proteins separated by two-dimensional PAGE were transferred to nitrocellulose membranes using a Bio-Rad Trans-blot apparatus at 250 mA (constant current) for 16 h at 4°C, and 10 mM Tris buffer containing

1.1% glycine and 20% methanol. The following steps were performed at room temperature. The blots were blocked with 5% bovine serum albumin in 50 mM Tris/HCl pH 7.4, 150 mM NaCl (Tris/NaCl) for 1 h at room temperature, and incubated with a dilution (1:4 in blocking solution) of either an individual serum or a pool of sera from allergic patients. A serum from a non-allergic patient was used as a negative control. After a 16-h incubation, the membranes were washed three times for 10 min in Tris/NaCl and incubated for 4 h with peroxidase-conjugated rabbit anti-human IgE antibodies diluted (1:1000) in blocking solution. The blots were washed as described above and incubated with the peroxidase substrates 3,3'-diaminobenzidine (0.5 mg/ml) and H₂O₂ (0.015%).

Blotting with lectins. The carbohydrate structures of glycoproteins separated by two-dimensional PAGE were analysed by the specific binding of selected lectins, using the digoxigenin glycan differentiation kit from Boehringer Mannheim. Proteins separated by two-dimensional PAGE were transferred to nitrocellulose membranes as described above and probed with the following digoxigenin-labeled lectins according to the manufacturer's directions: *Galanthus nivalis* agglutinin, *Sambucus nigra* agglutinin, *Maackia amurensis* agglutinin, peanut agglutinin and *Datura stramonium* agglutinin. The lectins bound to the glycosylated proteins were detected using an alkaline phosphatase-conjugated anti-digoxigenin antibody and the 4-nitroblue tetrazolium chloride/5-bromo-4-chloro-3-indolyl-phosphate substrates following the manufacturer's instructions.

Deglycosylation of glycoproteins with PNGase F. For treatment with PNGase F, 1 vol. horse dander extract dissolved in water (10 mg/ml) was mixed with 1 vol. 40 mM sodium phosphate pH 7.2 containing 20 mM EDTA, 4% Nonidet P-40, 0.4% SDS and 2% 2-mercaptoethanol. PNGase F was added to the mixture at a final concentration of 40 U/ml. A control was prepared the same way by replacing the enzyme solution with phosphate buffer. The mixtures were then incubated for 24 h at 37°C and analysed by two-dimensional PAGE, immunoblotting, and blotting with lectins as described above.

N-terminal sequencing and determination of amino acid composition of horse dander proteins. Proteins from horse dander were separated by two-dimensional PAGE and transferred to polyvinylidene difluoride membranes using either 10 mM Caps/NaOH pH 11 containing 10% methanol for N-terminal sequencing, or 7.5 mM Tris buffer containing 1.2 mM boric acid for determination of amino acid composition. Proteins were stained with Coomassie brilliant blue R-250 solution (0.025% in 40% methanol) and spots from 10 membranes were excised and pooled for sequencing or amino acid analysis. Acetic acid was avoided in both the staining and the destaining (50% methanol) solutions as it might cause blockage of the amino terminus. Proteins were sequenced by the Protein Sequencing Facility (Biotechnology Centre of Oslo, University of Oslo), using an Applied Biosystems model 477A protein sequencer with an on-line analytical Applied Biosystems 120 PTH analyser.

Amino acid composition was also determined by the Protein Sequencing Facility at the Biotechnology Centre of Oslo. Proteins were hydrolysed under vacuum in 6 M HCl at 110°C and their amino acid compositions determined using a Biotronic LC 5000 automatic amino acid analyser.

Analyses of N-terminal sequences and amino acid compositions. Sequence data were analysed and compared with the National Centre for Biotechnology Information (Bethesda MD, USA) protein and DNA databases using Blast (Basic local alignment search tool) programs (BLASTP and TBLASTN) [17]. Amino acid compositions of horse dander proteins were compared with amino acid compositions of the proteins from the Swiss-Prot database using the AACompIdent program from the

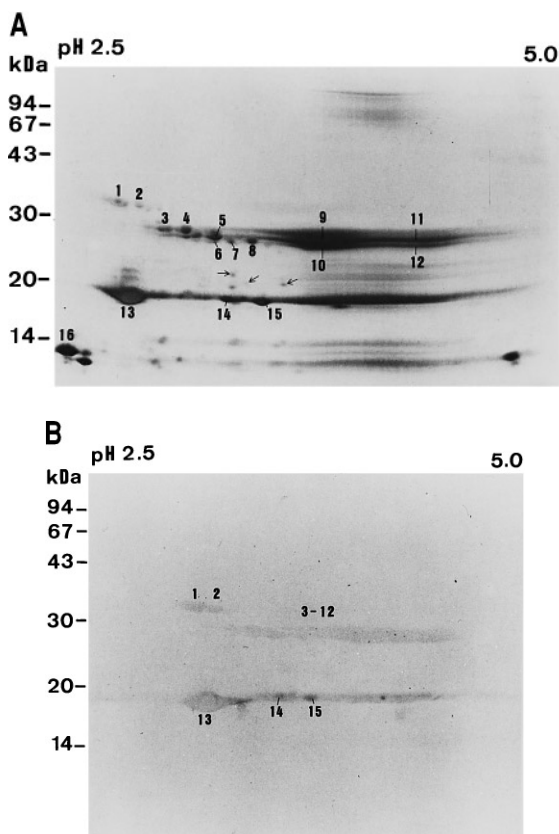


Fig. 1. Two-dimensional PAGE and western blot of horse dander proteins. (A) Proteins from the horse dander extract were separated by two-dimensional PAGE and stained with silver. The arrows indicate the position of minor components giving strong signals on immunoblots with IgE from several patient sera as illustrated in Fig. 2 B. (B) Western blot of horse dander proteins separated by two-dimensional PAGE and probed with a pool of allergic patient sera.

ExPASy Molecular Biology Server (University of Geneva). Hydrophobic cluster analysis (HCA) plots were made using the PCHCA program [18] and compared manually. HCA scores were calculated as described by Lemesle-Varloot et al. [19].

Sugar composition of horse dander glycoproteins. Sugar composition analysis was performed on individual components separated by two-dimensional PAGE, transferred to polyvinylidene difluoride membranes (Caps buffer) and stained with Coomassie brilliant blue R-250 as described above. Individual spots from 10 membranes were excised and pooled. The glycoproteins were separately incubated for 4 h at 100 °C in 2 M trifluoroacetic acid for the release of neutral and amino-monosaccharides. Release of sialic acid was performed under milder hydrolysis conditions, i.e. 0.1 M HCl at 80 °C for 1 h. Polyvinylidene difluoride membranes were discarded and the supernatant evaporated to dryness. Released monosaccharides were dissolved in a minimum volume of de-ionised water and analysed on a Dionex Bio LC system fitted with a CarboPac PA1 analytical anion-exchange column (4×250 mm) and a pulsed amperometric detector (Dionex). Neutral and amino-sugars were eluted isocratically using 16 mM NaOH at a flow rate of 1 ml/min. The samples containing sialic acid were eluted with a 50–180 mM sodium acetate gradient in 100 mM NaOH at 1 ml/min.

RESULTS

Two-dimensional electrophoresis. Two-dimensional PAGE using a pH gradient of 3–10 [8] showed that horse dander aller-

Table 1. Reactivity of IgE from individual patient sera on immunoblots with horse dander allergens. Horse dander proteins were separated by two-dimensional PAGE, transferred to nitrocellulose membranes and probed with allergic patients IgE. Proteins are numbered as described in Fig. 1 A; minor components are indicated by arrows in Fig. 1 A.

Patient serum number	Protein recognised	Signal intensity
1	1–12	++
2	13–15	++
3	1–12	+
4	16	+/-
	1–12	+
5	14, 15	++
6	14, 15	+/-
7	1–12	+
8	14, 15	+
9	14, 15	++
10	14, 15	+
11	14, 15	++
12	14, 15	++
	minor components	+/-
13	16	+/-
	13–15	+++
14	none	
15	none	
16	16	+
	13–15	+++
17	1–12	+
18	1–12	+
19	1–12	+
	minor components	++
20	14, 15	+++
	minor components	+
21	1–12	+
22	none	
23	none	

gens are clustered into two main groups having apparent molecular masses of 20 and 30 kDa, respectively, and acidic pI values in the range 3–5. However, this pH gradient did not allow identification of individual allergens because the components were not well separated. This was achieved by using two-dimensional PAGE with a pH gradient of 2.5–5.0 in the first dimension (Fig. 1 A). Fig. 1 A shows that the horse dander extract could be separated into up to 50 components as detected by silver staining.

IgE-binding components were identified after immunoblotting two-dimensional gels with a pool of sera from allergic patients (Fig. 1 B). Protein spots 1–15 gave a relatively strong signal and several of these are major components of the preparation. Their IgE-binding activity was not due to a non-specific binding of IgE antibodies since sera from non-allergic patients did not give any signal after immunoblotting (not shown). Also, some major components such as protein spot 16 and proteins detected in the same area (Fig. 1 A) did not give any positive signal with the pool of sera from allergic patients. These results suggest that the IgE-binding activity of horse dander proteins is not correlated with their relative abundance in the preparation.

Horse dander allergen reactivity with IgE in 23 patient sera was tested on immunoblots and results are presented in Table 1. About 50% of the patient sera contain IgE specifically recognising the 18.5-kDa components (proteins 14 and 15, pI of 3.4 and 3.5 respectively), as illustrated in Fig. 2 A. 35% of them recognised the 27–29-kDa and 31-kDa proteins (1–12), and IgE in these sera did not cross-react with the 18.5-kDa allergens

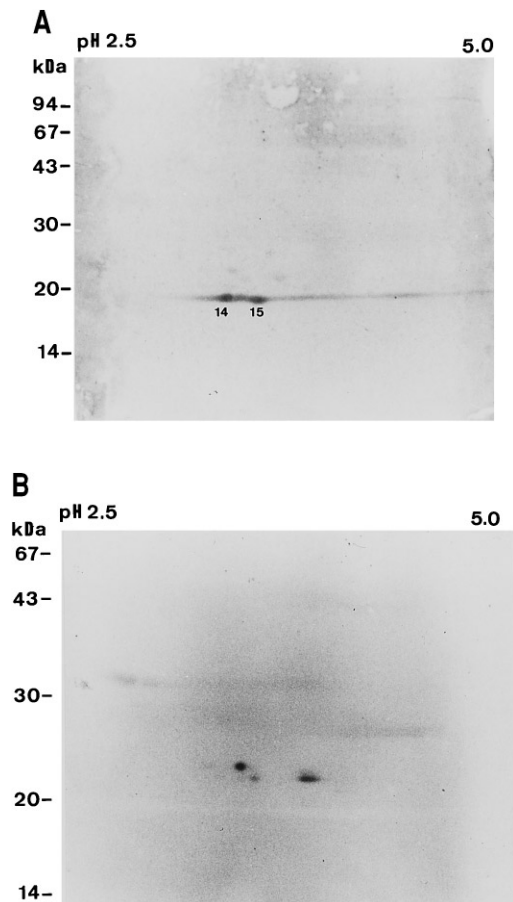


Fig. 2. Western blot analysis of horse dander proteins separated by two-dimensional PAGE and probed with sera from patients. (A) Detection of allergens 14 and 15 (see Fig. 1A) with specific IgE from one patient. (B) Detection of minor components (Fig. 1A, arrows) in the horse dander extract with specific IgE from another patient.

(Table 1). Reciprocally, the IgE from the sera recognising the 18.5-kDa allergens did not bind the 27–29-kDa and 31-kDa components. These results show that the allergens recognised by the largest number of patient sera have an apparent molecular mass of 18.5, 27–29 and 31 kDa, and that patient sera can be divided into two groups: the first one is composed of sera containing IgE specific for the 18.5-kDa allergens, the second group of sera specific for the 27–29-kDa and 31-kDa allergens. Interestingly, a few sera exhibited a strong reactivity on immunoblots with minor components (Fig. 1A, arrows), as shown in Fig. 2B. Again, this shows that the IgE-binding activity of horse dander allergens is not correlated with their relative abundance in the preparation.

Proteins from the horse dander extract separated by two-dimensional PAGE were stained with either Schiff's reagent and Coomassie brilliant blue, or Schiff's reagent alone to detect the presence of glycan chains on IgE-binding components. Fig. 3A shows that proteins 1–12 and 16 are glycosylated since they are clearly detected with Schiff's reagent. Their carbohydrate/protein ratios are relatively high compared with protein 13 which gives a very weak signal with Schiff's reagent but is strongly stained with Coomassie brilliant blue (Fig. 3). Components 14 and 15 were not or very poorly glycosylated since they could be detected with Coomassie brilliant blue only (Fig. 3). It is noteworthy that components 1–12 (27–29 and 31 kDa) are glycosylated and specifically recognised by the second group of pa-

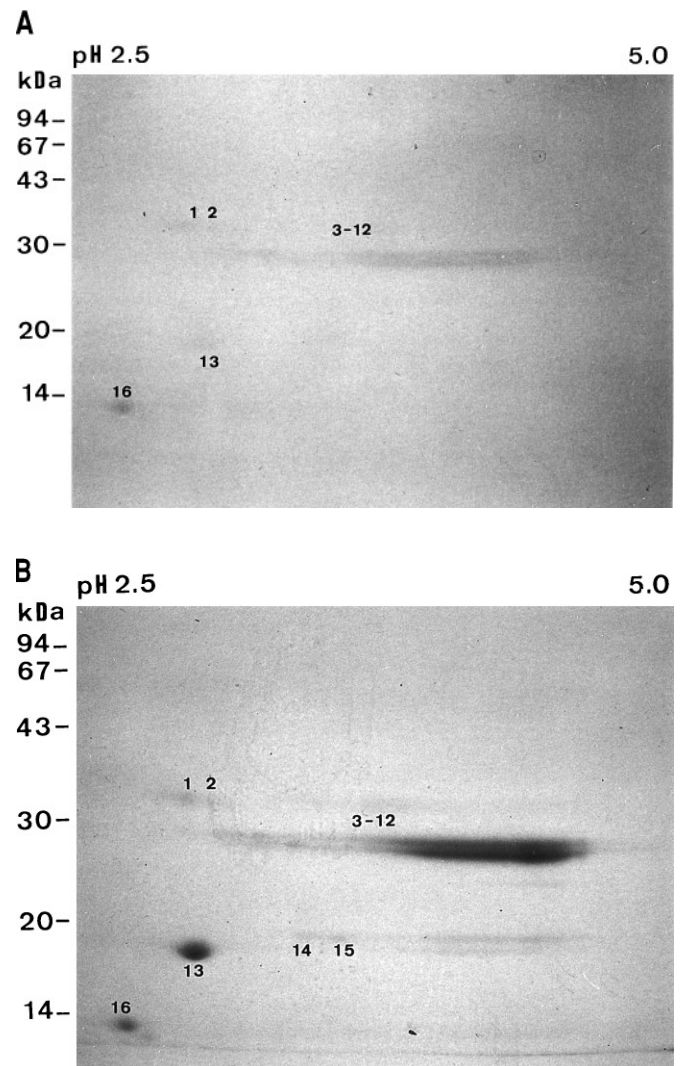


Fig. 3. Protein and glycoprotein staining of horse dander components separated by two-dimensional PAGE. (A) Carbohydrates were stained with Schiff's reagent. (B) Double staining of horse dander components with Schiff's reagent and Coomassie blue.

tient sera, whereas the first group of sera is specific for the non-glycosylated allergens (14 and 15 of 18.5 kDa).

Western blot immunoreactivity of patient sera from the second group was strongly reduced after periodate oxidation of the carbohydrate part of allergens 1–12 (data not shown). This indicates that the carbohydrate moiety of the glycosylated allergens might play a role in the binding of IgE from the second group of patient sera.

Carbohydrate analyses. After two-dimensional PAGE and blotting to polyvinylidene difluoride membranes, the sugar composition of the glycosylated allergens was determined by high-performance anion-exchange chromatography pulsed amperometric detector (HPAEC-PAD). The carbohydrate moiety of glycoproteins 1–5, 8–12 and 16 is composed of about 50% *N*-acetylglucosamine, 15–20% galactose and 15–30% mannose (Table 2). Amongst these allergens, some were found to be not (4, 5, 8) or only poorly (3) sialylated, whereas the carbohydrate moieties of others (1, 2, 9–12 and 16) contained up to 9.5% sialic acid. Components 6 and 7 have a much lower *N*-acetylglucosamine content and a much higher mannose content, by com-

Table 2. Sugar composition of horse dander glycoproteins. Sugar composition analysis was performed on individual components separated by two-dimensional PAGE and transferred to polyvinylidene difluoride membranes. Glycoproteins were stained with Coomassie blue and incubated at 100°C for 4 h in 2 M trifluoroacetic acid for release of neutral and amino-monosaccharides. Release of sialic acid was performed under milder hydrolysis conditions: 0.1 M HCl at 80°C for 1 h. Sugar composition of each glycoprotein was determined by HPAEC-PAD. SA, sialic acid.

Glycoprotein	Content of			
	GlcNAc	Gal	Man	SA
	%			
1	48	21	25	6.0
2	48	22	22	6.5
3	56	16	26	1.7
4	52	19	28	0
5	50	17	32	0
6	25	5	70	0
7	26	14	57	2.8
8	54	15	30	0
9	45	20	24	9.5
10	49	16	29	6.0
11	50	18	23	7.5
12	54	14	23	7.5
16	54	20	16	9.0

parison with the previous group of allergens (Table 2). Protein 6 was not found to be sialylated in contrast to allergen 7, and it contains three times less galactose.

Specific binding of several lectins was tested on blots obtained from two-dimensional PAGE of the horse dander extract. The lectins from *Galanthus nivalis* and peanut did not bind to any of the glycosylated allergens, whereas the lectins from *Sam-bucus nigra* and *Datura stramonium* gave a clear positive signal with some proteins (Table 3). This indicates the presence of terminal sialic acid linked α -(2→6) to galactose on allergens 1–12 and 16, and the presence of galactose linked β -(1→4) to N-acetylglucosamine on allergens 1–12. A much weaker signal was observed for glycoproteins 1–12 and 16 with the lectin from *Maackia amurensis*, indicating possible presence of terminal sialic acid linked α -(2→3) to galactose (Table 3).

The horse dander extract was treated with PNGase F and analysed by two-dimensional PAGE (Fig. 4). The originally gly-

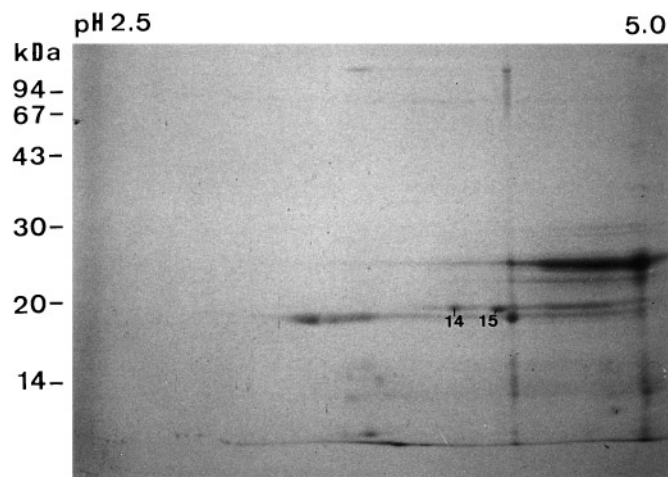


Fig. 4. Coomassie blue staining of horse dander components treated with PNGase F and separated by two-dimensional PAGE.

cosylated allergens then became completely deglycosylated as judged by the Schiff's reagent staining (not shown). Moreover, there was no positive signal when the PNGase-F-treated allergens were probed on blots with the lectins from *S. nigra*, *D. stramonium* and *M. amurensis* (not shown). These results indicate that the glycosylated components in the preparation all belong to the family of N-glycoproteins. The immunoreactivity of patient sera was strongly reduced when the allergens were treated with PNGase F (data not shown), confirming the role of the carbohydrate part in the interaction with IgE. Coomassie blue staining of PNGase-F-treated allergens showed shifts in the molecular mass and the pI of the glycosylated proteins (Fig. 4). Components 1–12 were clearly shifted to a less acidic pH range and to a lower molecular mass area (23–25 kDa instead of 27–29 and 31 kDa). These results show that the apparent acidic pI of these allergens is, at least partially, due to sialic acid in their carbohydrate moieties, and that the size of their N-glycan chains varies from 3–5 kDa (proteins 3–12) to 6 kDa (proteins 1 and 2). Surprisingly, the mobility of proteins 14 and 15 in two-dimensional gels was slightly modified after treatment with PNGase F, as illustrated in Fig. 4. The shift observed for these proteins in the less acidic pH range might be due to a low carbohydrate content, not detectable with Schiff's reagent. Their apparent molecular mass after treatment with PNGase F was not modified confirming that, if these proteins are glycosylated, their

Table 3. Lectin-binding activity of horse dander glycoproteins. The proteins were separated by two-dimensional PAGE, transferred to nitrocellulose membranes and probed with digoxigenin-labeled lectins, using the glycan differentiation kit from Boehringer Mannheim. Proteins are numbered as described in Fig. 1A.

Lectin	Specificity	Glycoprotein recognised	
		number	signal intensity
<i>G. nivalis</i> agglutinin	mannose terminally linked	—	
<i>S. nigra</i> agglutinin	terminal sialic acid linked α -(2→6) to galactose or N-acetylgalactosamine	16 1–12	+/- ++
<i>M. amurensis</i> agglutinin	terminal sialic acid linked α -(2→3) to galactose	16 3–12 1–2	+/- +/- +
Peanut agglutinin	galactose linked β -(1→3) to N-acetylgalactosamine	—	
<i>D. stramonium</i> agglutinin	galactose linked β -(1→4) to N-acetylglucosamine	1–12	++

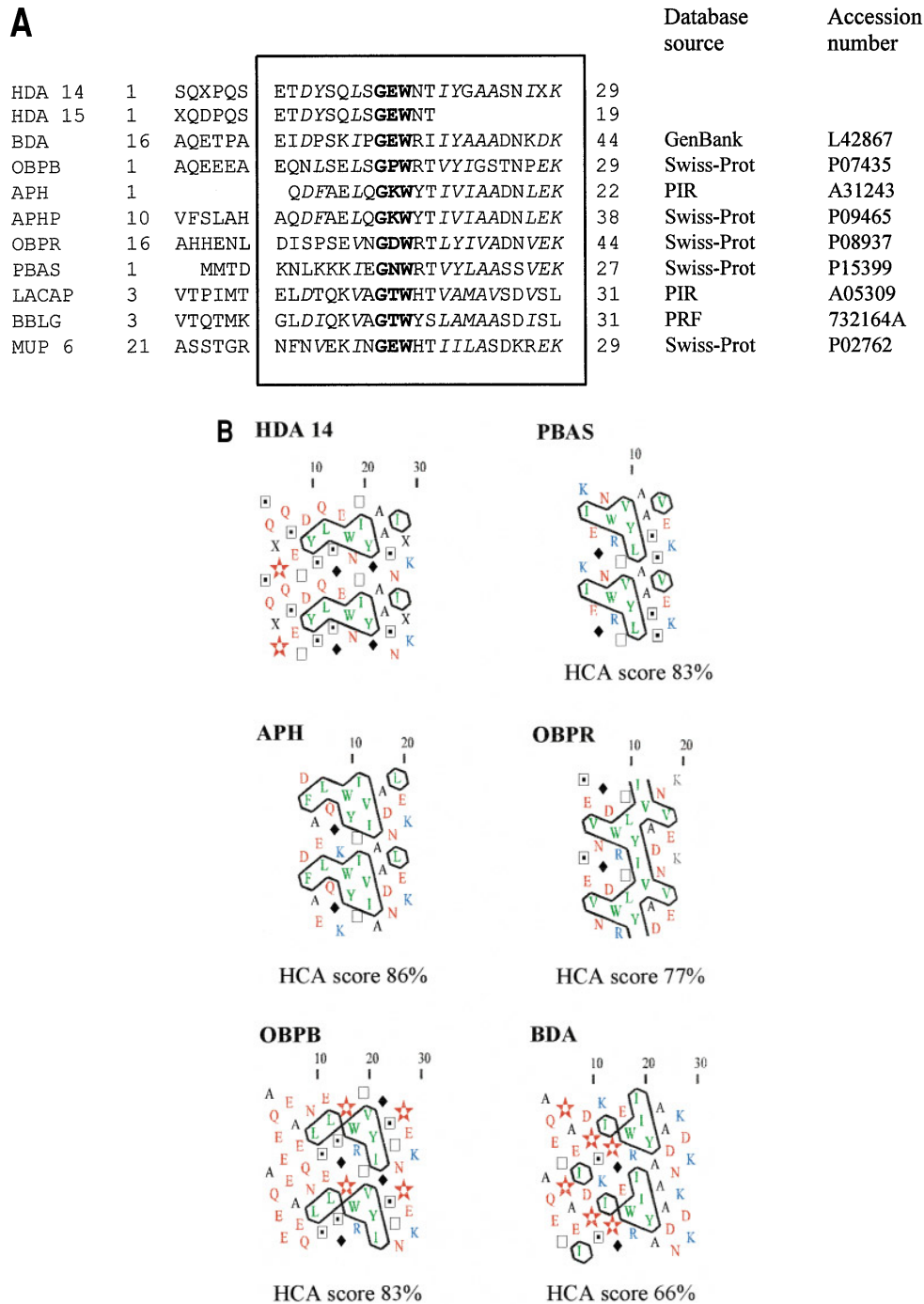


Fig. 5. Sequence analyses of allergens 14 and 15. (A) Amino acid sequence alignments of horse dander proteins 14 (HDA 14) and 15 (HDA 15) and nine similar proteins. The amino acid residues of each protein are numbered. Abbreviations and sources of protein sequence data are BDA, bovine dander allergen [20, 21]; OBPB, bovine odorant-binding protein [23]; APH [24] and APHP [25], hamster aphrodisin and aphrodisin precursor, respectively, from black-bellied hamster; OBPR, rat odorant-binding protein [26]; PBAS, rat probasin [27]; LACAP, porcine β -lactoglobulin A [29]; BBLG, bovine β -lactoglobulin [22]; MUP 6, mouse major urinary protein 6 precursor [28]. A consensus sequence of lipocalin proteins is boxed. The highly conserved GXW motif is bold, and the amino acid residues are in italics where seven or more residues in a column are either acidic (D or E), basic (R or K) or hydrophobic (A, I, L, M, F, W, Y, V). (B) HCA plots drawn from the N-terminal sequences of horse dander protein 14 and five similar proteins abbreviated as in (A). HCA scores were calculated manually as described by Lemesle-Varloot et al. [19]. The amino acid color code of Lemesle-Varloot et al. [19] was used: hydrophilic residues are either red (D, E, N, Q) or blue (H, K, R), and hydrophobic residues are either green (V, I, L, F, W, M, Y) or black (A, C). ♦, G; □, T; □, S; ☆ (red), P.

carbohydrate/protein ratios are low. Allergen 13 was hardly shifted after incubation with PNGase F, confirming the low carbohydrate content of this component. Conversely, the polypeptide corresponding to allergen 16 cannot be located on the gel shown in Fig. 4, indicating that treatment with PNGase F of this

component dramatically affects its mobility, and confirming that it has a relatively high carbohydrate/protein ratio.

N-terminal sequencing and amino acid analysis. IgE-binding components were separated by two-dimensional PAGE, transfer-

Table 4. Amino acid composition of glycoproteins 9 and 10. The proteins were separated by two-dimensional PAGE, transferred to polyvinylidene difluoride membranes, hydrolysed with 6 M HCl and subjected to amino acid analysis. The amino acid compositions of these proteins were compared with those in the Swiss-Prot database, using the AACompIdent program (Expasy Molecular Biology server). The compositions of two proteins giving high scores and belonging to the lipocalin family is compared with the compositions of horse proteins 9 and 10. These are the rat odorant-binding protein (OBP rat [26]) and the mouse major urinary protein 5 (MUP5 mouse [32]). The amounts of amino acids in horse proteins correspond to the average values found for proteins 9 and 10; nd = not determined.

Amino acid	Residues found (nearest integer)		Amount in		
	for horse proteins		horse proteins	OBP rat	MUP5 mouse
	9	10			
			%		
Asx	31	29	16.7	14.5	12.2
Thr	6	6	3.1	5.2	5.0
Ser	14	15	6.4	4.0	5.0
Thr + Ser	20	21	9.5	9.2	10.0
Glx	32	31	19.1	14.0	15.0
Pro	4	4	1.8	1.7	0.6
Gly	18	19	5.8	5.8	3.3
Ala	11	12	4.2	4.7	5.0
Cys	1	1	0.2	3.5	2.8
Val	17	17	8.0	8.7	5.6
Met	0	0	0	1.2	2.2
Ile	12	12	6.6	4.7	7.2
Leu	12	12	6.7	8.7	11.7
Ile + Leu	24	24	13.3	13.4	18.9
Tyr	5	5	3.6	3.5	2.8
Phe	9	9	5.9	3.5	6.1
Tyr + Phe	14	14	9.5	7.0	8.9
His	1	1	0.6	4.7	3.3
Lys	10	10	6.1	6.4	7.8
Arg	7	8	5.3	4.7	3.9
Lys + Arg	17	18	11.4	11.1	11.7
Trp	nd	nd	nd	0.6	0.6
Total residues	190	191			

red to polyvinylidene difluoride membranes, and ten spots stained with Coomassie brilliant blue corresponding to the same protein were pooled and used for N-terminal sequencing to identify the horse dander major allergens. Respectively 27 and 18 amino acids of proteins 14 and 15 were positively identified (Fig. 5A). N-terminal sequences of these two allergens showed complete identity. This result strongly suggests that components 14 and 15 are two isoelectric variants of the same protein. Analysis of the longest N-terminal amino acid sequence (protein 14) showed in the first 29 amino acids a 44% identity and a 58% similarity to the Bos d 2 bovine major allergen [20, 21]. Searches in protein and DNA databases using BLASTP and TBLASTN programs [17] revealed a 31–47% identity and 60–71% similarity between protein 14 and a number of proteins from cow (β -lactoglobulin [22], odorant-binding protein [23]), hamster (aphrodisin [24], aphrodisin precursor [25]), rat (odorant-binding protein [26], probasin [27]), mouse (major urinary protein 6 precursor [28]) and pig (β -lactoglobulin A [29]), all belonging to the lipocalin family (Fig. 5A). One of the consensus sequences of this family (boxed in Fig. 5A) begins at amino acid residue 7 of the horse dander major allergens and includes

the highly conserved Gly-Xaa-Trp motif [30]. Results obtained with Blast programs and amino acid alignments revealed that proteins 14 and 15 do not show any significant similarity to the horse major allergen called Equ c 1 [10].

Sequences were further compared by hydrophobic cluster analysis (HCA) [19, 31], a very efficient method that can handle several sequences simultaneously and combines similarity detection with secondary structure analysis. The results are shown in Fig. 5B. HCA plots made from sequences of protein 14 and the proteins from the lipocalin family mentioned above revealed the presence of a common hydrophobic cluster at the N-terminal. HCA scores of proteins similar to protein 14 vary from 66% to 86%. This indicates a high percentage of topographically conserved hydrophobic residues in the sequences of these proteins by comparison to the total number of hydrophobic residues in the compared N-terminal segments.

Attempts to sequence the N-terminal of the glycosylated allergens of 30 kDa were unsuccessful both before and after deglycosylation with PNGase F. The N-terminal of these proteins might be blocked, making their direct sequencing impossible. However, amino acid compositions of the glycosylated allergens 9 and 10 could be determined as illustrated in Table 4. The composition of these proteins were almost identical, suggesting that they might correspond to isoforms of the same protein. When these compositions were compared to amino acid compositions of the mammalian proteins exhibiting acidic pI values in the Swiss-Prot database (AACompIdent tool, Expasy Molecular Biology server), high scores were obtained with mouse urinary proteins (see for example MUP5, MUP1, MUP2 [32] and MUP6 [28]) and rat [26] and bovine [33] odorant-binding proteins, all belonging to the lipocalin family. Table 4 compares the compositions of proteins 9 and 10 with the mouse major urinary protein 5 [32] and the rat odorant-binding protein [26]. It shows similarities between these proteins but also major differences, especially in the content of Cys, Met and His. The horse dander allergens also exhibit slightly higher content in Asx and Glx, compared to the mouse and rat proteins. The bovine dander major allergen [20, 21], another member of the lipocalin family, was also amongst the proteins giving the highest scores. Although results from such comparisons must be interpreted with care, our data suggest that the glycosylated horse dander major allergens, like the non-glycosylated ones, belong to the lipocalin family. Results from amino acid analyses of the glycosylated allergens give a total number of residues of less than 200 (Table 4), which correspond to a molecular mass of about 21–23 kDa. This molecular mass is comparable to the apparent molecular mass of the corresponding deglycosylated proteins observed after two-dimensional PAGE analysis (Fig. 4).

DISCUSSION

Up to now several papers have described antigenic and allergenic composition of partially purified extracts from horse dander [2–8]. However, these reports do not contain any sequence data, making impossible the clear identification of the horse dander allergens. In 1993, Dandeu et al. [9] reported that they had achieved a 99% purification of what was referred to as the horse dander major allergen using a straightforward procedure involving hydrophobic-interaction chromatography. The data presented in our paper show that the situation is actually complex and that horse dander extracts contain a mixture of several major allergens rather than a single one. Since 1993, there have been no more reports on the purification to homogeneity and characterisation of other horse dander allergens. For the first time, we have been able to separate by two-dimensional PAGE the different

components of horse dander extracts. Up to 50 proteins exhibiting acidic pI in the range 2.5–5.0 could be detected by silver staining. Amongst these proteins, 16 are allergens and belong to one of the two families we have identified in this work. Of the patient sera tested, 50% had IgE recognising allergens of the first family, i.e. allergens that were not found to be glycosylated. The second family is composed of glycosylated allergens that were recognised by 35% of the patient sera tested. None of the patient sera that bound antigens from the first family recognised antigens from the second family, and reciprocally. On this basis, IgE of patient sera can also be divided in two groups.

When vicinal hydroxyl groups of the carbohydrate moieties of the glycosylated allergens were oxidized and destroyed by periodate oxidation, their immunoreactivity were strongly reduced on western blots (not shown). A comparable result was obtained when the glycan chains of the allergens were removed by treatment with PNGase F (not shown). Taken together, these data indicate that the carbohydrate moiety of allergens belonging to the second family is likely to interact at least partially with IgE from patient sera. Reduced immunoreactivity after periodate treatment has been previously found for other allergens such as timothy pollen [34], yeast [35] and fungal allergens [36, 37]. However, other reports on the cat major allergen [38, 39] have shown that the protein moiety rather than the N-glycan chain of the allergen is involved in the binding of IgE. To our knowledge, data from our laboratory (this work and [8]) provide the first indication that the glycosidic part of animal allergens might also participate in the interaction with IgE from allergic patients.

It is important to determine the structures recognised by IgE to understand what differentiates an allergen from a classical antigen. We have started to analyse the carbohydrate moiety of several horse dander allergens since it seems to play a role in the binding of IgE. These analyses show that allergens from the second family are all N-glycosylated and that the size of the glycan chains varies over 3–6 kDa. Some of these allergens have heterogeneous pI values, which might be due to different degree of sialylation of their glycan chains. Sialic acid is at least partially responsible for the apparent acidic pI of some allergens. When these were treated with PNGase F, their pI values were clearly shifted to a less acidic pH range. Amongst the N-glycans analysed, those from allergens 6 and 7 are likely to be of the high-mannose type, as judged by their mannose content. The other allergens contain a lower proportion of mannose and a much higher amount (50%) of *N*-acetylglucosamine. Their carbohydrate parts are most likely complex or hybrid-type N-linked oligosaccharides. Experiments conducted with lectins provided information on the kind of linkages involved in the glycan chains: terminal sialic acid linked α -(2→6) and possibly α -(2→3) to galactose, and galactose linked β -(1→4) to *N*-acetylglucosamine. The results presented in this paper provide important information on the structure of the glycan chains. However, the complete structural characterisation of the carbohydrate moieties of each major allergen will need more powerful techniques, such as mass spectrometry. This is underway and we expect to have resolved the structure of the glycan chains of most glycosylated allergens in the near future.

We sequenced non-glycosylated major allergens from the first family, i.e. proteins 14 and 15. These proteins do not correspond to Equ c 1 [10], since they exhibit different sequences. Therefore, we propose to call them Equ c 2.0101 and Equ c 2.0102 respectively, according to the International Nomenclature System [12]. Equ c 2.0101 and Equ c 2.0102 are most likely two variants of the same protein since their N-terminal sequence is identical. They show the same apparent molecular mass after two-dimensional PAGE (18.5 kDa) and they have slightly different acidic pI (3.4 for Equ c 2.0101 and 3.5 for

Equ c 2.0102). Searches in protein and DNA databases revealed a significant sequence identity (up to 47%) and similarity (up to 71%) of Equ c 2.0101 and Equ c 2.0102 to a group of mammalian proteins belonging to the lipocalin family. They did not show any similarity between Equ c 2.0101/Equ c 2.0102 and the allergen Equ c 1, another member of the lipocalin family that is found in horse hair and dander. Equ c 1 is an N-glycosylated allergen that exhibits an apparent molecular mass of about 25 kDa after SDS/PAGE under reducing conditions [9, 10]. It is likely that it corresponds to one of the 27-kDa glycoproteins that we have grouped in the second family of horse dander allergens. Equ c 2.0101 and Equ c 2.0102 N-terminal sequences are characterised by the presence of a motif (Gly-Xaa-Trp) that is highly conserved at the N-terminal of the lipocalin proteins [30]. Interestingly, one of the similar proteins, the bovine dander major allergen (Bos d 2 [20, 21]), has a similar origin as Equ c 2.0101 and Equ c 2.0102. It is important to note that the other proteins showing similarity with Equ c 2.0101 and Equ c 2.0102 are low molecular mass (18–19 kDa) mammalian proteins found in nasal mucosa or epithelial cells and their secretions (Fig. 5). Like the other members of the lipocalin family, they are extracellular proteins responsible for the binding and transport of small hydrophobic molecules [11]. Lipocalins can also bind to specific cellular receptors and, in a quite general way, a common role for these molecules in the mediation of cell regulation has been proposed [11]. Equ c 2.0101 and Equ c 2.0102 have in common with most members of the lipocalin protein family a number of properties such as their molecular mass, their total number of residues and their acidic pI [11]. HCA [19, 31] was used to further compare the N-terminal sequences of Equ c 2.0101 and Equ c 2.0102 with the sequences from several lipocalins. HCA plots revealed the presence of a common N-terminal hydrophobic cluster and a high percentage of topographically conserved hydrophobic residues in the segments examined. The similarity found between Equ c 2.0101/Equ c 2.0102 and proteins from the lipocalin family after HCA confirms at the two-dimensional level the significant scores obtained with the classical Blast alignment method [17]. Although the sequences compared are relatively short, these data suggest a common secondary structure for the N-terminal of the proteins analysed.

In summary, the results presented in this paper strongly suggest that the allergens Equ c 2.0101 and Equ c 2.0102 are new members of the lipocalin family. However, the biological function of Equ c 2.0101 and Equ c 2.0102, especially their role in the transport of small hydrophobic molecules, has still to be demonstrated. It is noteworthy that several other members of the lipocalin family have been included in the official list of allergens published by the IUIS Allergen Nomenclature Subcommittee (Hørsholm, Denmark). This is the case, for example, for the recently described horse (Equ c 1 [9, 10]) and bovine (Bos d 2 [20, 21]) dander major allergens, for the bovine α -lactalbumin (Bos d 4) and β -lactoglobulin (Bos d 5 [40]), and the mouse major urinary protein (Mus m 1 [41]).

We were unable to sequence the N-glycosylated major allergens belonging to the second family. Their N-terminal is probably blocked, making impossible their direct sequencing. Amino acid compositions of some of these proteins could be determined and when they were compared to the compositions of known acidic mammalian proteins in the Swiss-Prot database, high scores were obtained with proteins belonging to the lipocalin family. The glycosylated lipocalin Equ c 1 might correspond to one of the glycosylated allergens we have studied, although it was not amongst the proteins exhibiting high scores. The biochemical properties of the glycosylated allergens from the second family (apparent molecular mass, size of the glycan chains, number of residues) are similar to those of known glycosylated

lipocalin proteins [11]. These results suggest that several allergens from the second family are also lipocalins. However, this must be confirmed by internal sequences of these proteins after treatment with a protease and separation of the resulting oligopeptides.

Although much remains to be discovered on the horse dander major allergens, the results presented in this paper represent important steps towards their complete structural characterisation. Accumulation of such information is essential for detailed studies of the molecular interactions in allergenic reactions, as well as for clinical diagnosis and therapy.

This work was supported by a grant from the European Union Commission (ERBCHRXCT 940598). We thank Mrs Ellen Hanne Cohen (Institute of Pharmacy, University of Oslo) for technical assistance with HPAEC-PAD, Dr Knut Sletten (Biotechnology Centre of Oslo, University of Oslo) for help with protein sequencing and amino acid analyses, and Dr Bernard Henrissat (CERMAV-CNRS, Grenoble, France) for helpful comments on HCA plots.

REFERENCES

1. Squire, J. R. (1950) The relationship between horse dandruff and horse serum antigen in asthma, *Clin. Sci.* 9, 127–150.
2. Stanworth, D. R. (1957) The isolation and identification of horse dandruff allergen, *Biochem. J.* 65, 582–598.
3. Ponterius, G., Brandt, R., Hulten, E. & Yman, L. (1973) Comparative studies on the allergens of horse dandruff and horse serum, *Int. Arch. Allergy Appl. Immunol.* 44, 679–691.
4. Løwenstein, H., Markussen, B. & Weeke, B. (1976) Isolation and partial characterization of three major allergens of horse hair and dandruff, *Int. Arch. Allergy Appl. Immunol.* 51, 48–67.
5. Markussen, B., Løwenstein, H. & Weeke, B. (1976) Allergen extracts of horse hair and dandruff. Quantitative immunoelectrophoretic characterization of the antigens, *Int. Arch. Allergy Appl. Immunol.* 51, 25–37.
6. Franke, D., Maasch, H. J., Wahl, R., Schultze-Werninghaus, G. & Bretting, H. (1990) Allergens of horse epithelium, *Int. Arch. Allergy Appl. Immunol.* 92, 309–317.
7. Fjeldsgaard, B. E. & Smestad Paulsen, B. (1993) Comparison of IgE-binding antigens in horse dander and a mixture of horse hair and skin scrapings, *Allergy* 48, 535–541.
8. Johnsen, T. K., Thanh, D. B., Ly, Q., Smestad Paulsen, B. & Wold, J. K. (1996) Further characterization of IgE-binding antigens in horse dander, with particular emphasis on glycoprotein allergens, *Allergy* 51, 608–613.
9. Dandeu, J. P., Rabillon, J., Divanovic, A., Carmi-Leroy, A. & David, B. (1993) Hydrophobic interaction chromatography for isolation and purification of Equ c 1, the horse major allergen, *J. Chromatogr.* 621, 23–31.
10. Gregoire, C., Rosinski-Chupin, I., Rabillon, J., Alzari, P. M., David, B. & Dandeu, J. P. (1996) cDNA cloning and sequencing reveal the major horse allergen Equ c 1 to be a glycoprotein member of the lipocalin superfamily, *J. Biol. Chem.* 271, 32951–32959.
11. Flower, D. R. (1994) The lipocalin protein family: a role in cell regulation, *FEBS Lett.* 354, 7–11.
12. King, T. P., Hoffman, D., Løwenstein, H., Marsh, D. G., Platts-Mills, T. A. E. & Thomas, W. (1995) Allergen nomenclature, *J. Allergy Clin. Immunol.* 96, 5–14.
13. Dunbar, B. S. (1987) *Two-dimensional electrophoresis and immunological techniques*, Plenum Press, New York and London.
14. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding, *Anal. Biochem.* 72, 248–254.
15. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227, 680–685.
16. Zacharius, R. M., Zell, T. E., Morrison, J. H. & Woodlock, J. J. (1969) Glycoprotein staining following electrophoresis on acrylamide gels, *Anal. Biochem.* 30, 148–152.
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol.* 215, 403–410.
18. Bouterin, B., Lavaitte, S. & Henrissat, B. (1995) *PCHCA: a HCA plot program for the IBM-PC and Postscript laser printers*, Centre de Recherches sur les Macromolécules Végétales, CNRS, Grenoble.
19. Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. & Mornon, J. P. (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences, *Biochimie* 72, 555–574.
20. Mantjarvi, R., Parkkinen, S., Ryttonen, M., Pentikainen, J., Pelkonen, J., Rautiainen, J., Zeiler, T. & Virtanen, T. (1996) cDNA cloning of the predominant allergen of bovine dander: a new member in the lipocalin family, *J. Allergy Clin. Immunol.* 97, 1297–1303.
21. Rautiainen, J., Ryttonen, M., Virtanen, T., Pentikainen, J., Zeiler, T. & Mantjarvi, R. (1997) BDA20, a major bovine dander allergen characterized at the sequence level is Bos d 2, *J. Allergy Clin. Immunol.* 100, 251–252.
22. Braunitzer, G., Chen, R., Schrank, B. & Stangl, A. (1973) Die sequenzanalyse des β -lactoglobulins, *Hoppe-Seyler's Z. Physiol. Chem.* 354, 867–878.
23. Cavaggioni, A., Sorbi, R. T., Keen, J. N., Pappin, D. J. & Findlay, J. B. (1987) Homology between the pyrazine-binding protein from nasal mucosa and major urinary proteins, *FEBS Lett.* 212, 225–228.
24. Henzel, W. J., Rodriguez, H., Singer, A. G., Stults, J. T., Macrides, F., Agosta, W. C. & Niall, H. (1988) The primary structure of aphrodisin, *J. Biol. Chem.* 263, 16682–16687.
25. Magert, H. J., Hadrys, T., Cieslak, A., Groger, A., Feller, S. & Forssmann, W. G. (1995) cDNA sequence and expression pattern of the putative pheromone carrier aphrodisin, *Proc. Natl Acad. Sci. USA* 92, 2091–2095.
26. Pevsner, J., Reed, R. R., Feinstein, P. G. & Snyder, S. H. (1988) Molecular cloning of odorant-binding protein: member of a ligand carrier family, *Science* 241, 336–339.
27. Spence, A. M., Sheppard, P. C., Davie, J. R., Matuo, Y., Nishi, N., McKeen, W. L., Dodd, J. G. & Matusik, R. J. (1989) Regulation of a bifunctional mRNA results in synthesis of secreted and nuclear probasin, *Proc. Natl Acad. Sci. USA* 86, 7843–7847.
28. Clark, A. J., Clissold, P. M., Al Shawi, R., Beattie, P. & Bishop, J. (1984) Structure of mouse major urinary protein genes: different splicing configurations in the 3'-non-coding-region, *EMBO J.* 3, 1045–1052.
29. Bell, K., McKenzie, H. A. & Shaw, D. C. (1981) Porcine β -lactoglobulin A and C. Occurrence, isolation and chemical properties, *Mol. Cell. Biochem.* 35, 103–111.
30. Flower, D. R., North, A. C. T. & Attwood, T. K. (1993) Structure and sequence relationships in the lipocalins and related proteins, *Protein Sci.* 2, 753–761.
31. Gaboriaud, C., Bissery, V., Benchetrit, T. & Mornon, J. P. (1987) Hydrophobic Cluster Analysis: an efficient new way to compare and analyse amino acid sequences, *FEBS Lett.* 224, 149–155.
32. Shahan, K., Gilmartin, M. & Derman, E. (1987) Nucleotide sequences of liver, lachrymal, and submaxillary gland mouse major urinary protein mRNAs: mosaic structure and construction of panels of gene-specific synthetic oligonucleotide probes, *Mol. Cell. Biol.* 7, 1938–1946.
33. Tirindelli, R., Keen, J. N., Cavaggioni, A., Eliopoulos, E. E. & Findlay, J. B. (1989) Complete amino acid sequence of pyrazine-binding protein from cow nasal mucosa, *Eur. J. Biochem.* 185, 569–572.
34. Haavik, S., Smestad Paulsen, B. & Wold, J. K. (1987) Glycoprotein allergens in pollen of timothy. V. Significance of the carbohydrate moiety for the immunological activity of a basic glycoprotein allergen, *Int. Arch. Allergy Appl. Immunol.* 83, 231–237.
35. Heelan, B. T., Allan, S. & Barnes, R. M. R. (1991) Identification of a 200 kDa glycoprotein antigen of *Saccharomyces cerevisiae*, *Immunol. Lett.* 23, 181–186.
36. Hansen, M. Y., Wold, J. K., Smestad Paulsen, B., Cohen, E. H. & Karlsson-Borgå, Å. (1994) Allergens in *Aspergillus fumigatus*. I.

- Characterization of two different allergen extracts and evaluation of their stability and the importance of the carbohydrate for IgE binding, *Allergy* 49, 235–241.
37. Swärd-Nordmo, M., Smestad Paulsen, B., Wold, J. K. & Grimmer, Ø. (1984) Characterization of the carbohydrate moiety in a partly purified allergen preparation from the mould *Cladosporium herbarum* and its possible importance for allergenic activity as tested by RAST-inhibition, *Int. Arch. Allergy Appl. Immunol.* 75, 149–156.
38. Duffort, O. A., Carreira, J., Nitti, G., Polo, F. & Lombardero, M. (1991) Studies on the biochemical structure of the major cat allergen *Felis domesticus* 1, *Mol. Immunol.* 28, 301–309.
39. Bond, J. F., Brauer, A. W., Segal, D. B., Nault, A. K., Rogers, B. L. & Kuo, M. C. (1993) Native and recombinant Fel d 1 as probes into the relationship of allergen structure to human IgE immunoreactivity, *Mol. Immunol.* 30, 1529–1541.
40. Ball, G., Shelton, M. J., Walsh, B. J., Hill, D. J., Hosking, C. S. & Howden, M. E. (1994) A major continuous allergenic epitope of bovine β -lactoglobulin recognized by human IgE binding, *Clin. Exp. Allergy* 24, 758–764.
41. Lorusso, J. R., Moffat, S. & Ohman, J. L. Jr (1986) Immunologic and biochemical properties of the major mouse urinary allergen (Mus m 1), *J. Allergy Clin. Immunol.* 78, 928–937.