

Determination of Isoforms, N-Linked Glycan Structure and Disulfide Bond Linkages of the Major Cat Allergen Fel d1 by a Mass Spectrometric Approach

Anne Kroll Kristensen¹, Carsten Schou² and Peter Roepstorff^{1,*}

¹ Department of Molecular Biology, Odense University, DK-5230 Odense M, Denmark

² ALK Abellø Group, Bøge Allé 10–12, DK-2970 Hørsholm, Denmark

* Corresponding author

The domestic cat (*Felis domesticus*) is an important source of indoor allergens, the major allergen being Fel d1 (formerly cat allergen 1). Fel d1 is responsible for cat allergy and has also been established to cause cat-induced asthma. The allergen is a 38 kDa dimer composed of two 19 kDa subunits. Each 19 kDa subunit comprises two disulfide linked polypeptide chains, a light α -chain and a heavy β -chain containing an N-linked oligosaccharide.

In this study a variety of endoproteinase digestions of the native allergen in combination with HPLC and matrix-assisted laser desorption mass spectrometry was used to determine the position of the disulfide bridges and to demonstrate that the peptide chains are linked in an anti parallel way. Enzymatic digestion of the reduced and alkylated peptides located the N-glycan to Asn₃₃. Moreover, Fel d1 is found to be partially truncated and to exist in several isoforms. Sequential degradation of the glycosylated peptide with specific glycosidases monitored by mass spectrometry, shows that the glycan is a heterogenous tri-antennary complex type structure. The heterogeneity is caused by terminal sialic acid and a fucose residue attached to a β -galactose residue.

Key words: Allergen / Disulfide bridges / Isoforms / N-linked glycan / Mass spectrometry.

Introduction

Cats are found in approximately 25% of the households in the Western communities, and allergy to cats is found in a large part of the population. In the United States for example, the frequency of cat allergy is in the order of 10%. Allergenic activity has been demonstrated in extracts from cat dander, pelt, serum, urine and saliva. The most important allergen from a clinical point of view is Fel d1 (*Felis domesticus*, latin for domestic cat), also known as Cat 1 (Leitermann and Ohman, 1984). This allergen has been de-

monstrated to be present in several tissues e.g., salivary glands, pelt, hair follicles and brain (reviewed by Schou, 1993). The Fel d1 molecule is responsible for a large proportion of the cat specific Ig E elicited in cat allergic patients and more than 60% of the histamine releasing activity of cat hair and dander extract was caused by Fel d1 (Kleine-Tebbe *et al.*, 1993). Fel d1 has furthermore been shown to be responsible for approximately 30% of the T-cell stimulatory activity of cat hair and dander extract. The molecule elicited a Th2-like response in cat allergic individuals (van Neerven *et al.*, 1994). The level of Fel d1 in houses with cats is often in excess of that required to provoke an asthmatic response. Despite the obvious abundance of this protein, its biological role is still unknown.

The Fel d1 molecule is a non-covalently linked dimer. Each subunit is composed of two covalently linked polypeptides, a light α -chain and a heavy β -chain containing an N-linked oligosaccharide. Monosaccharide analysis of the Fel d1 molecule showed high levels of fucose, glucosamine, galactose and mannose (Vailes *et al.*, 1994).

The complete primary structure of the two chains has been determined from immunoaffinity purified extracts of house dust (Morgenstern *et al.*, 1991). It was found that the α -chain shares a high degree of homology with the rabbit uteroglobin precursor which is a homodimer of anti parallel chains linked by disulfide bonds. Previous studies of Fel d1 indicated that the two chains composing the allergen are linked in an antiparallel way (Bond *et al.*, 1993). Fel d1 is the only allergen known so far to be a dimeric protein derived from two independent gene products. Both chains have been recombinantly expressed in high yields and purified to a high degree of purity (Rogers *et al.*, 1993).

The Ig E related epitopes are probably conformational and dependent on the presence of both chains (Duffort *et al.*, 1991). Ig E binding appeared to be specific to three synthetic peptides, aa 25–38 and 46–59 of the α -chain and 15–28 of the β -chain (van Milligen *et al.*, 1993). However, deglycosylation of the β -chain did not interfere with the antibody binding activity, suggesting that the glycan is not an important part of the epitope.

Several variants of Fel d1 have been observed. These include full-length and truncated versions of both chains as well as several amino acid substitutions. For the α -chain, the following amino acid substitutions have been reported: Lys₂₉→Arg, Val₃₃→Ser (Duffort *et al.*, 1991) and Lys₂₉→Asn (Griffith *et al.*, 1992). And for the β -chain: Asn₁₉→Ser, Gly₂₀→Leu (Duffort *et al.*, 1991). In addition, some less frequent amino acid substitutions in the β -chain are Ile₅₅→Val, Arg₅₇→Lys, Val₅₈→Phe, Glu₆₉→Val, Tyr₇₂→Asp, Gln₇₉→Glu and Asn₈₈→Lys (Griffith *et al.*, 1992).

The β -chain has been found to exist in two major forms, a long version consisting of the amino acid residues 1–92 (Morgenstern *et al.*, 1991), and a short version in which the residues 70 and 71 are deleted (Griffith *et al.*, 1992). In addition, truncated versions beginning at position 22 or 28 have been reported (Griffith *et al.*, 1992).

In this study we have determined the positions of the disulfide linkages and the position and structure of the N-linked glycan. In addition, we have found Fel d1 isoforms and truncations different from those hitherto reported.

Results

Analysis of Intact and Reduced Fel d1

Analysis by matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) of the intact protein isolated by immunoaffinity purification of extracts of cat hair showed a very broad peak with a maximum at m/z 19300, indicating that the protein was inhomogeneous. Enzymatic deglycosylation of the native protein was performed, but attempts to obtain a spectrum of the deglycosylated protein were unsuccessful.

The protein was further purified by reversed phase-HPLC (rp-HPLC) on a C_4 -column (Figure 1) resulting in a broad peak which was collected as three fractions.

The three fractions were subjected to analysis by MALDI MS before and after reduction on the mass spectrometric target. The reduction was not complete, as indicated by the presence of the singly (MH^+), doubly (MH_2^{2+}) and triply (MH_3^{3+}) charged molecular ions for the complete allergen, as well as the singly charged molecular ion for the α -chain (Figure 2). The β -chain could not be seen in the spectra. The molecular ion peak for the allergen is very broad and no distinct components can be identified, suggesting a complex heterogeneity consisting of a combination of sequence variants and heterogeneous secondary

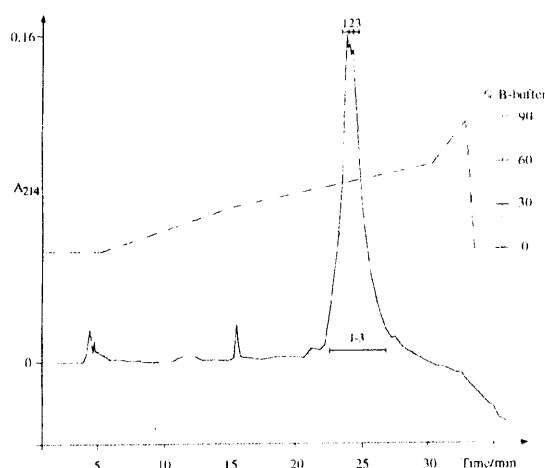


Fig. 1 Purification of Native Fel d1 by Reversed Phase-HPLC. The dotted line shows the gradient used. The upper bar shows the fractions corresponding to the spectra in Figure 2. The lower bar shows the fraction collected for all further studies.

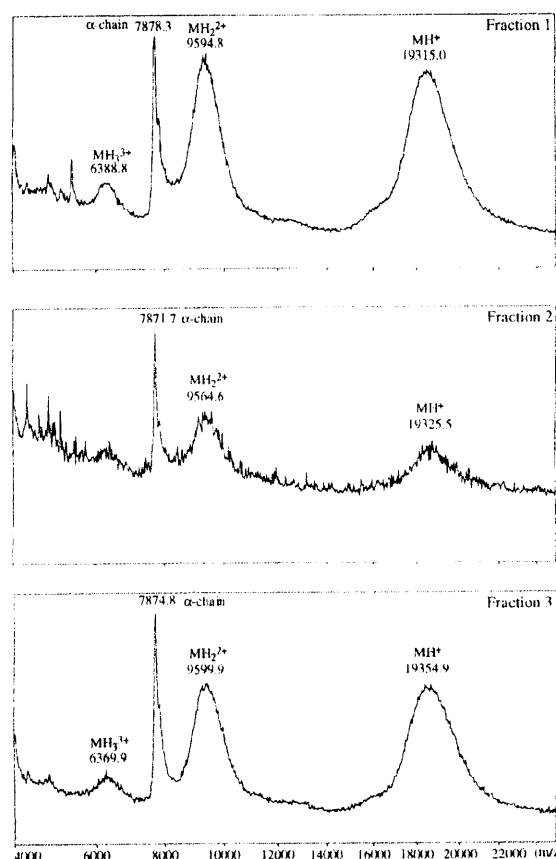


Fig. 2 MALDI Spectra of Fractions (1–3) (Figure 1), after Reduction on Target.

Both the singly (MH^+), doubly (MH_2^{2+}) and the triply (MH_3^{3+}) charged molecular ion of the allergen as well as the singly charged molecular ion of the α -chain are observed.

Table 1 Glycosidases Used in this Study.

Enzyme	Concentration	Specificity
N-acetyl neuraminidase	10 mU/ml	α 2–6 > α 2–3 > α 2–8
Exo β -galactosidase	2 mU/ml	β 1–3 > β 1–4
N-Acetyl β -D-Glucosaminidase	1 mU/ml	GlcNAc β 1–2Man
Endoglycosidase D	1 mU/ml	GlcNAc β 1–4 ↓ GlcNAc
Peptide N-glycosidase F (PNGase F)	50 mU/ml	Asn ↓ – GlcNAc

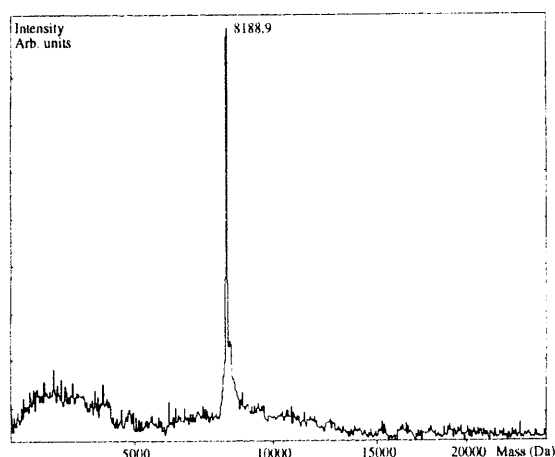
Specificities are based on data from the producers. ↓ indicates the site of cleavage.

modification, such as glycosylation. The observed molecular masses and those calculated according to the published sequence are given in Table 2. The calculated molecular mass of the α -chain corresponds to the observed molecular masses within an experimental error of 0.1%. The peak is, however rather broad at the base and the mass accuracy is not sufficient to exclude the presence of

Table 2 Measured and Calculated Molecular Masses of the Fel d1 Heterodimer and the α - and β -Chains.

Peak corresponding to	M_r observed	M_r calculated
α -chain	7874.9*	7869.0
β -chain	NI	9999.5
$\alpha + \beta$ -chain	19331.8*	17868.5

The calculated masses are based on sequence data from Morgenstern *et al.*, 1991. *: average of the three spectra in figure 2. NI: not identified.

**Fig. 3** MALDI Spectrum of the Reduced and Alkylated α -Chain. The spectrum shown is of an aliquot taken directly from the reduction/alkylation reaction mixture. Only the reduced and alkylated α -chain is observed.

the Lys-Asn isoform or other isoforms. The mass measured for the intact protein increases from fraction one to fraction three by +10 and +40 Da relative to fraction one. These mass differences are hardly significant considering the very broad peaks. It is, however, most likely that the observed differences in mass between the three chromatographic peaks represent sequence heterogeneity and not glycosylation heterogeneity, because the latter would result in larger mass differences. The molecular mass of the subunit based on the published sequence (Morgenstern *et al.*, 1991) is 17868.5 Da. The observed molecular mass is on average 19330 Da, which is 1460 Da higher than the calculated mass. This mass difference combined with the broad peak strongly indicates the presence of glycosylation. The correct mass assignment of the α -chain shows that the β -chain must be the glycosylated chain, as also supported by the presence of a consensus sequence for N-linked glycosylation in this chain. All following work was performed on a fraction containing all three components (Cf. Figure 1, fraction 1–3) because isolation of the individual components was not possible.

Upon reduction and alkylation of the allergen with 4-vinyl pyridine followed by HPLC, only the α -chain or a mixture of the α -chain and the heterodimer could be recovered. Similar results have been reported by Duffort *et al.*,

1991. A MALDI spectrum taken directly from the alkylation mixture of a fully reduced subunit also only showed the reduced and alkylated α -chain, Figure 3. The observed mass (8187.9 Da) is in perfect agreement with that calculated from the published sequence for the reduced and alkylated α -chain mass (8187.0 Da).

Sequencing of the intact protein for 6 cycles showed three different N-terminal sequences. Two corresponded to those expected for the two chains and the third indicated partial truncation of the β -chain by cleavage between position 3 and 4.

Peptide Mapping and Sequencing of the Chains

All attempts to isolate the reduced and alkylated β -chain failed. Therefore, it was decided to work on a mixture of the reduced and alkylated chains after dialysis against water to remove reagents. The mixture of the two chains was cleaved with endoproteinase Asp-N, and 22 fractions isolated by HPLC (Figure 4). All fractions were analyzed by MALDI MS, and the mass of the peptides correlated with the published sequence (Table 3). In addition, an endoproteinase Asp-N digestion using shorter digestion time (6 h) was also analyzed (Table 3). This resulted in identification of an additional number of peptides which covered some of the missing residues. Unexpectedly, cleavages N-terminally to glutamic acid residues as well as C-terminally to both aspartic acid and glutamic acid residues were also observed. In addition, cleavage between two leucine residues in the sequence DVFTAVASNGNEL ↓L, and between a Thr and a Val residue in the sequence DCMGEA-VQNT↓V was observed (↓ indicates site of cleavage).

Data from mass spectrometry and Edman degradation confirm the presence of a Lys₂₉→Asn substitution in the

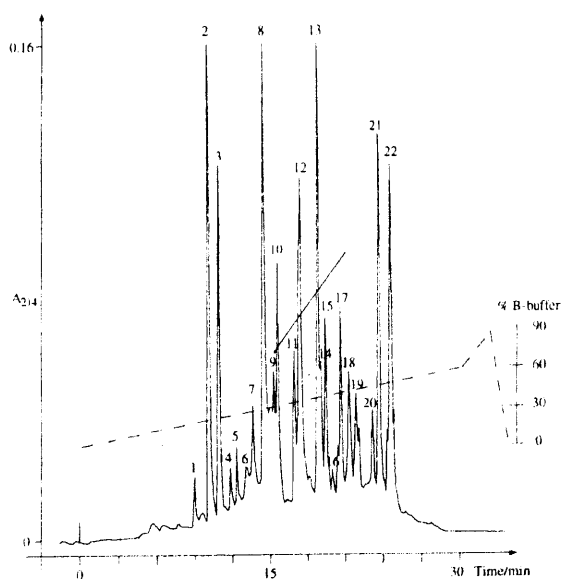
**Fig. 4** Reversed Phase-HPLC of 1 nmol Reduced and Alkylated Fel d1 Digested with Endoproteinase Asp-N. The glycosylated peptide is marked with an arrow. The gradient is shown by the dotted line.

Table 3 Molecular Mass and Sequence Information for Peptides Obtained by Endoproteinase Asp-N Digestion of a Mixture of the Reduced and Alkylated α - and β -Chains.

Amino acid residue	M_r observed	M_r calculated	Edman degradation	Fraction nr.
1–8 (α)	1018.9	1020.1	EICPAVKR	3
19–35 (α)	1963.9	1964.3	ND	19
19–39 (α)	2434.7	2434.7	DEYVEQVAQYKALPWLENA(R)	17
19–39 (α)	2421.0	2420.7	DEYVEQVAQYNALPWLENA(R)	18
19–45 (α)	3198.9	3196.7	DEYVEQVAQYNALPWLENA(RILKNVC)	22
19–45 (α)	3212.7	3210.7	ND	21
36–45 (α)	1263.2	1264.4	ENARILKNVC	7
46–52 (α)	821.9	822.8	DAKMTEE	14
46–61 (α)	1807.7	1806.9	ND	15
53–61 (α)	1001.9	1002.1	DKENALSLL	5
62–70 (α)	1144.3	1144.2	DKIYTSPLC	8
1–11 (β)	1406.8	1406.7	VKMAETCPIFY	13
5–11 (β)	977.1	976.9	ETCPIFY	12
12–23 (β)	1296.6	1295.4	ND	10
12–24 (β)	1336.2	1335.5	ND	6
12–25 (β)	1522.1	1521.7	ND	11
12–37 (β)	2789.8	2791.1	ND	20
26–35 (β)	1062.8	1061.2	ND	4
26–46 (β)	4–5000	2373.7	ND	9
47–51 (β)	ND	626.7	DCYVE	1
47–59 (β)	1587.2	1585.7	ND	16
72–81 (β)	ND	1172.1	DCMGEAVQNT	2
1–11 (β)	1423.4	1406.7	VKMAETCPIFY	*
4–11 (β)	1047.1	1048.9	AETCPIFY	*
12–25 (β)	1451.6	1448.6	DVFIAVASGNEILL	*
26–37 (β)	1187.2	1190.3	DLSLTQVATEP	*
26–38 (β)	4659	1416.5	DLSLTQVATEPE	*
47–57 (β)	1264.2	1267.4	DCYVENGLISR	*
12–23 (β)	1296.6	1295.4	DVFFAVANGNEL	*

Isoforms are underlined. N.D.: not determined, *: Endoproteinase Asp-N digestion for only 6 hours. () mass residue assigned according to the mass and the sequence published by Morgenstern *et al.*, 1991.

α -chain. In addition isoforms containing Phe₁₅→Thr and Asn₁₉→Ser substitutions in the β -chain are also identified (Table 3). The Asn₁₉/Ser isoform has been described by Duffort *et al.*, 1991, whereas the Phe₁₅/Thr isoform has not been described previously. The calculated and observed masses agree within the experimental error, with one exception: VKMAETCPIFY of the β -chain, for which a mass of + 16.7 Da is observed (Table 3). This is most likely due to oxidation of the methionine residue present.

A small peptide always co-eluted with the variant form of β 12–25. The sequence of this peptide (AETCPIFY) supported the previous mentioned presence of a truncated version of the β -chain, beginning at position 4.

The glycosylated peptide was located in fraction 9 based on a mass spectrum which showed the heterogeneity caused by glycosylation. In addition minor amounts of peptides containing the non-glycosylated sequence demonstrates that the allergen is not 100% glycosylated. Based on the data obtained, it is not possible to determine the amount of glycosylated versus non-glycosylated peptide.

Determination of the Disulfide Bond Linkages

To determine the positions of the disulfide bond linkages, the native allergen was treated with *Staphylococcus aureus* V8 protease and a combination of trypsin and chymotrypsin followed by separation of the obtained peptides by HPLC. Despite the presence of 1 M urea in the digestion buffer to denature the protein, none of the digestions were complete. All fractions were analyzed by MALDI MS and the relevant peptides were reduced on target and re-analyzed. The V8-derived peptides revealed the presence of one disulfide bond, α -chain Cys₇₀ linked to β -chain Cys₇. A peptide is observed at m/z 2507.5 Da corresponding to the residues 6–12 of the β -chain linked to the residues 54–70 of the α -chain (calculated mass: 2506.8). When reduced on target, this peptide disappears and a new peptide appears at 1650.3 Da, corresponding to the residues 54–70 of the α -chain (calculated mass: 1650.9). The β -chain peptide was not observed in the spectrum. This peptide has a rather low mass and may be masked by matrix ions or suppressed. It is not unusual that one of the two peptides is not seen in the spectrum upon reduction.

The combined action of trypsin and chymotrypsin revealed the presence of a second disulfide bond between α -chain Cys₄₄ and to β -chain Cys₄₈ (Figure 5). The peptide observed at 2156.7 Da corresponds to residues 45–57 of the β -chain linked to residues 43–48 of the α -chain. When reduced on target, this peptide disappears and a peptide appears at 1510.1 Da corresponding to the residues 45–57 of the β -chain. The minor peptide at 1699.4 Da is probably the result of a reaction between the reduced peptide and the matrix (calculated mass: 1699.3). Again, only one of the two disulfide-linked peptides is observed after reduction.

The last disulfide bond between α -chain Cys₃ and β -chain Cys₇₂ was localized in a fraction containing a mixture of all three disulfide bonded peptides. A peptide at 2139.3 Da is seen which corresponds to residues 1–7 of the α -chain linked to residues 70–82 of the β -chain. After reduction this peptide disappears and a peptide at 758.6 Da, corresponding to residues 1–7 of the α -chain, appears. In addition peptides at 2158.3 Da and at 1666.3 Da corresponding to peptides containing the disulfide-linked α -

Table 4 Sequences Obtained by Automated Edman Degradation.

M_r cal- culated	M_r observed	Edman degradation	Corre- sponding amino acid
1020.0	1019.9	EICPAVKR	α 1–8
757.9	ND	ENALSLL	α 55–61
2430.6	2432.2	DVDLFLTGPDEYVEQVAQY	α 9–28
449.6	450.3	TAMK	β 40–43
1450.7	1451.9	VLDGLVMT	β 58–65

The peptides are the products of a trypsin and chymotrypsin digestion of the native protein. The molecular weights were determined using either PDMS or MALDI MS. N.D.: not determined.

chain Cys₄₄ to β -chain Cys₄₈ and α -chain Cys₇₀ to β -chain Cys₇ are seen. These peaks also disappear upon reduction and the corresponding reduced peptides appear. The unambiguous assignment of the α -Cys₃ to β -Cys₇₂ disulfide bond from the mixture was only possible because the two other disulfide bonds were already established.

The peptides isolated by digestion of the native protein were all, based on the determined mass, assigned to the published sequence of Feld 1 or the sequence variants identified from the endoproteinase Asp-N digestion. Based on their mass, some of the peptides were assigned to parts of the sequence not covered by the peptides derived by digestion with endoproteinase Asp-N. These peptides were sequenced to confirm the assignment (Table 4). As a result, the complete sequence of the α -chain and the sequence of the β -chain except the residues 84 to 92, could be confirmed.

Determination of the N-linked Glycan

Monosaccharide analysis by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC PAD) after hydrolysis of native Feld 1 showed that the allergen contained four different monosaccharides (Fuc, GlcNAc, Gal and Man) in agreement with a previous report (Vailes *et al.*, 1994).

Analysis of fraction 9 from the endoproteinase Asp-N digestion by MALDI MS showed a very heterogeneous peak pattern (Figure 6A), typical for glycosylation. The mass spectrum provides a structural 'fingerprint' of the glycan. A closer examination of the masses shows the presence of a mixture of 2 glycoforms. One is a triantennary complex structure with sialic acid and the other a triantennary complex structure with sialic acid and fucose (Table 5). The mass accuracy is not as high as for analysis of pure peptides because the molecular ion signals are distributed over several peaks resulting in a poor signal-to-noise ratio. Therefore somewhat larger mass deviations than normally acceptable has been allowed in the interpretation. The fraction was lyophilized and redissolved in 50 mM NH₄Ac, pH 6.0 and sequentially treated with the specific glycosidases listed in table 1, to determine the structure of the glycan.

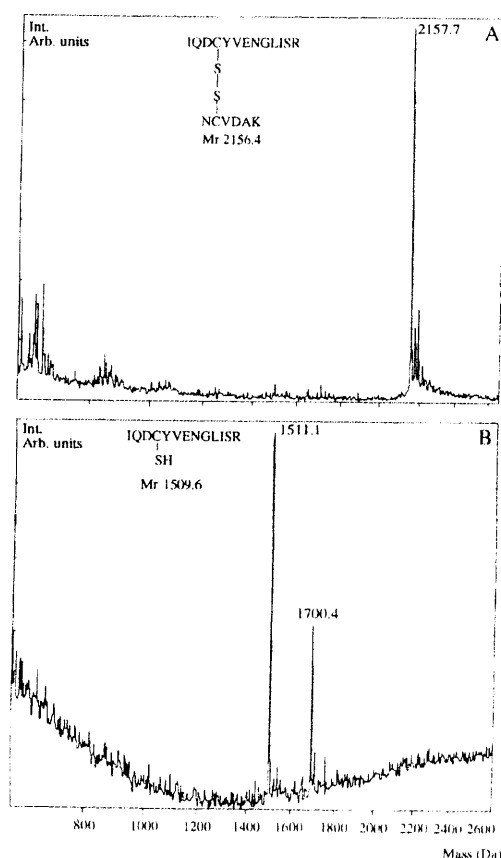


Fig. 5 MALDI Spectra Establishing the Cys₄₄-Cys₄₈ Disulfide Bond.

MALDI mass spectrum of the peptide_{45–57} of the β -chain linked to peptide_{43–48} of the α -chain by a disulfide bond (A): When reduced on target, the peak at m/z 2157.8 disappears and a peak at m/z 1511.1 corresponding to the peptide_{45–57} of the β -chain appears (B). The peak at m/z 1700.4 is most likely a reaction product between peptide_{45–57} and a matrix molecule.

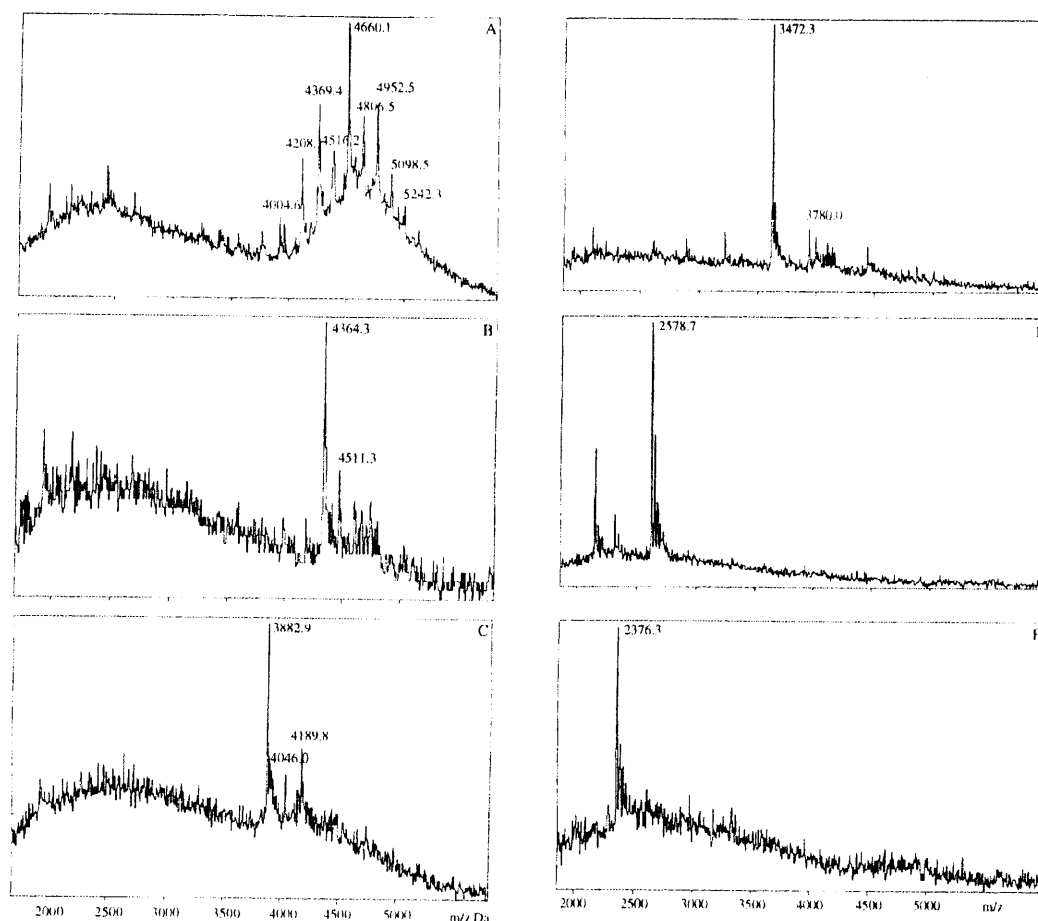


Fig. 6 MALDI Spectra of the Glycosylated Peptide and the Products of the Different Glycosidase Reactions.

MALDI spectra of the peptide present in fraction 9 (6A), treated with: Neuraminidase (6B), β -galactosidase (6C), N-acetylglucosaminidase (6D), Endo D (6E) and PNGase F (6F). Interpretation of the data is given in Table 5.

Treatment with neuraminidase (Figure 6B), reduced the number of peaks to one major and one minor peak, indicating that the major heterogeneity was caused by N-acetyl neuraminic acid. The major peak at m/z 4364.3 corresponds to a triantennary complex type, while the peak at m/z 4511.3 corresponds to the triantennary structure containing one fucose residue. The presence of additional fucose residues might be indicated by the small peaks located at multiples of 146 Da above the peak at m/z 4511.3. However, the signal-to-noise ratio is too low to unambiguously determine the total number of fucose residues.

Treatment with β -D-galactosidase (Figure 6C) gave rise to a major peak at m/z 3882.9, corresponding to the loss of three galactose units and a minor peak at m/z 4046.0 indicating that the removal of galactose was not complete. A peak at 4189.8 corresponds to the structure $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_3\text{GalFuc}$. This indicates that the enzyme could not remove a Gal with a fucose residue attached.

Treatment with N-acetylglucosaminidase (Figure 6D) gave rise to a major peak at m/z 3472.3 corresponding to the loss of two N-acetylglucosamines. The enzyme does not cleave β 1–4 or β 1–6-bound N-acetylglucosamine

which is present in tri- and tetra antennary structures. This strongly indicates that one of the N-acetylglucosamines is β 1–4 or β 1–6-bound to the core mannose. A peak at m/z 3780.0 corresponds to the β 1–4 or β 1–6-bound N-acetylglucosamine with a galactose and a fucose still attached.

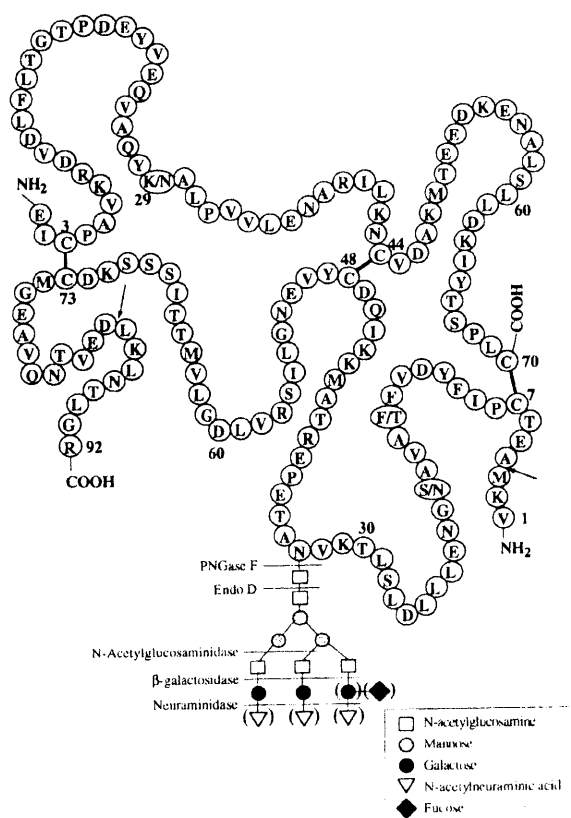
The mass difference between the Endo D (Figure 6E) and the PNGase F (Figure 6F) treated glycopeptide is 202.4 Da, corresponding to an N-acetylglucosamine residue. This proves that no core-bound fucose is present. The mass of the PNGase F treated peptide (2375.3) also locates the oligosaccharide to Asn₃₃ of the β -chain.

The combined data from glycosidase treatment and MALDI MS show that the N-linked glycan structure of the allergen is a mixture of several triantennary glycoforms due to heterogeneity with respect to fucose and sialic acid (Figure 7). The MALDI spectra do not allow quantification of the relative amounts of the different glycoforms because the intensities of the peaks do not necessarily reflect the amount of material present.

Table 5 Interpretation of the Oligosaccharide Structures Associated with the Peaks Observed in Figure 6.

<i>M_r</i> observed	<i>M_r</i> calculated	Composition
<i>Untreated peak 9 (Figure 6A):</i>		
4004.6	3997.2	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₂ Gal ₂
4208.1	4200.4	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₂
4369.4	4362.5	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃
4516.2	4508.7	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Fuc
4660.2	4654.8	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Sia
4806.5	4800.9	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ SiaFuc
4952.5	4946.1	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Sia ₂
5098.5	5092.3	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Sia ₂ Fuc
5242.3	5236.4	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Sia ₃
<i>Neuraminidase treated (Figure 6B):</i>		
4364.3	4362.5	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃
4511.3	4508.7	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal ₃ Fuc
<i>β-Galactosidase treated (Figure 6C):</i>		
3882.9	3876.1	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃
4046.0	4038.3	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ Gal
4189.8	4184.4	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc ₃ GalFuc
<i>N-Acetyl Glucosaminidase treated (Figure 6D):</i>		
3472.3	3469.7	(-Asn-) GlcNAc ₂ Man ₃ GlcNAc
3780.0	3777.7	(-Asn-) GlcNAc ₂ Man ₃ GlcNAcGalFuc
<i>Endo D treated (Figure 6E):</i>		
2578.7	2576.9	(-Asn-) GlcNAc
<i>PNGase F treated (Fig. 6F):</i>		
2376.3	2374.7	-Asp

(-Asn-) or (-Asp-) represents the peptide β 26–46.



Discussion

Figure 7 gives a summary of the covalent structure of Fel d 1 as determined in this study. The allergen is glycosylated at Asn₃₃ of the β-chain and shows several amino acid variations and truncations. The disulfide bonds were unambiguously assigned, demonstrating that the two chains are anti parallel.

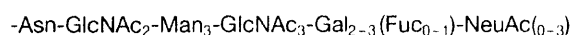
Two isoforms of the α-chain were observed, one with Lys and one with Asn at position 29, as also reported by Griffith *et al.*, 1992. The presence of two isoforms of the β-chain was also established. In position 15 a Phe to Thr and in position 19 an Asn to Ser substitution was seen in the same peptide. The Asn₁₉/Ser isoform has been described by Duffort *et al.*, 1991, whereas the Phe₁₅/Thr isoform has not been described previously. In addition, an N-terminally truncated form starting at position 4 of the β-chain was indicated from the sequence data.

Fig. 7 Structure of Fel d 1.

The structure of Fel d 1, showing the structure of the glycan, the disulfide bonds and the observed isoforms. Possible truncations are indicated by arrows. The structure of the glycan including heterogeneity as well as the mode of action of the glycosidases mentioned in Table 1 are indicated.

MALDI MS in combination with Edman degradation of the endoproteinase Asp-N derived peptides provided a 100% sequence coverage of the α -chain and an 88% coverage of the β -chain. The C-terminal peptide of the β -chain was not observed in the Asp-N digestion, neither by MALDI MS nor by Edman degradation.

Endoproteinase Asp-N digestion performed on the mixture of the reduced and alkylated chains showed that the protein is glycosylated on Asn₃₃ of the β -chain. Sequential treatment with different glycosidases monitored by MALDI MS showed the structure of the glycan to be a mixture of triantennary complex types:



The determination of the structure of the glycan is based on the specificity of the glycosidases employed. A detailed structure and linkage analysis was beyond the scope of this work. α -fucosyl linkages occur in various positions. N-acetylglucosamine residues at the reducing termini of N-linked sugar chains are often substituted by 1–6 linked fucose residues. Galactose and N-acetylglucosamine residues included in two backbone structures of complex-type sugar chains are also substituted by α -fucosyl residues resulting in the formation of the blood group antigens and their related antigenic structures. Recently fucose linked α 1,2 to galactose or α 1,3 to GlcNAc has been reported in glycoproteins (Lis and Sharon, 1993). Thus the observation of the Gal-linked fucose in Fel d1, albeit unexpected, is not unique.

The molecular mass calculated for the prevailing form of the glycan is approximately 2280 Da (with one sialic acid and no fucose). The calculated molecular mass of the allergen (average of the different isoforms) is approximately 17950 Da. Therefore, an estimate of the calculated molecular mass of the glycosylated allergen is approximately 20230 Da. Comparison of the determined molecular mass (19330 Da) with the calculated (value 20230 Da) reveals a mass deficit of 900 Da. In combination with the failure to observe peptides containing the C-terminal part of the β -chain, this strongly indicates that the β -chain is C-terminally truncated; most likely from position 85.

It is noteworthy that the peptide corresponding to the α -chain 9–29 was always obtained in high yield upon digestion of the native protein with trypsin, indicating that this sequence is exposed in the native allergen. Ig E binding has been demonstrated to be specific to one synthetic peptide covering the sequence 25–38 (van Milligen *et al.*, 1993), supporting that this region is exposed.

The elucidation of the structure of Fel d1 is likely to advance the rational development of future specific strategies for treatment of cat allergy. It is furthermore thought that the basic mechanisms underlying allergic diseases could be more readily addressed once the structure of Fel d1, the major cat allergen, has been worked out in detail.

Materials and Methods

Chemicals and Enzymes

PNGase F (EC 3.5.1.52), Endoglycosidase D (EC 3.2.1.96), Neuraminidase (EC 3.2.1.18), N-acetyl β -D-glucosaminidase (EC 3.2.1.52), β -D-galactosidase (EC 3.2.1.23), Endoproteinase Asp-N (EC 3.4.24.33) and chymotrypsin (EC 3.4.21.1) were obtained from Boehringer Mannheim (Mannheim, Germany). *Staphylococcus aureus* V8 protease (EC 3.4.21.19) was from ICN Biochemicals (Cleveland, USA). Trypsin (EC 3.4.21.4) was a gift from Novo Nordisk A/S (Bagsværd, Denmark). HPLC-grade acetonitrile was from Rathburn Chemicals (Walkerburn, UK). Nucleosil C₁₈ column material from Machery Nagel (Düren, Germany), Vydac C4 column material was from Hesperia (CA, USA), reagents for the sequencer from Knauer (Berlin, Germany), 50% NaOH solution from J. T. Baker (BV Deventer, The Netherlands). All other chemicals were of sequencing grade from Sigma (St. Louis, USA). All water used was obtained from a MilliQ water purification system (Millipore, GB).

Fel d1 Allergen Preparation

The major cat allergen Fel d1 was purified by two-step immunoaffinity chromatography.

Step 1: Murine monoclonal antibody (Cat 1, ALK-Abelló, Hørsholm, Denmark) (32.5 mg) with specificity for Fel d1 was coupled to 2.5 g of CNBr-activated Sepharose 4B (Pharmacia, Uppsala, Sweden) overnight at 5 °C according to manufacturers instructions. Subsequently, 50 mg of freeze-dried cat hair and dander extract (ALK-Abelló, Hørsholm, Denmark) was redissolved at 20 mg/ml in 0.05 M sodium carbonate buffer (pH 9.0) and applied to the immunoaffinity matrix at 30 ml/h and washed with 50 ml of application buffer. Fractions (5 ml) were collected throughout. The bound material was eluted with 50 ml 0.05 M glycine/NaOH, pH 11.0 and the fractions were immediately neutralized by addition of 2 M HCl. All purifications were conducted at 5 °C and all of the fractions collected were analyzed by fused rocket immunoelectrophoresis (Svendsen, 1973) with a rabbit antibody raised against cat hair and dander extract (Ra@Fel d, ALK-Abelló, Hørsholm, Denmark). At this stage, the purified Fel d1 contained slight contamination from albumin. The eluted fractions containing Fel d1 were pooled and dialyzed overnight at 5 °C against 1000 volumes of phosphate buffered saline, pH 7.4 (PBS). The dialyzed material was then subjected to step 2.

Step 2: Rabbit anti-human albumin (A001, DAKO, Glostrup, Denmark) (600 μ l) was coupled to 1 g of CNBr-activated Sepharose 4B as described above. This antibody is known to react with feline albumin. The pooled Fel d1 from step 1 was applied to the immunoaffinity matrix in PBS at 30 ml/h and 4 ml fractions of the effluent were collected. The immunoaffinity matrix was washed with 16 ml PBS. Bound material was then eluted with 16 ml 0.05 M glycine/NaOH, pH 11.0. All operations were conducted at 5 °C and the fractions recovered were analyzed as in step 1. The effluent and the first wash fraction contained Fel d1 without any detectable contamination with albumin. These fractions were pooled, dialyzed against 1000 volumes of 0.05 M ammonium-bicarbonate buffer and lyophilized. As analyzed by SDS-PAGE with silver staining (Schäfer-Nielsen and Rose, 1982), the Fel d1 preparation did not contain any contaminants.

Reduction and Alkylation

For reduction and alkylation the lyophilized protein (2 nmol) was dissolved in 200 μ l 0.5 M Tris-HCl, 6 M Guanidine-HCl, pH 7.6, 5 mM EDTA. 5 μ l 1.4 M DTT was added. After 10 min at room temperature, 1 μ l 4-vinyl pyridine was added, and after another 10

min the reaction was stopped by the addition 5 μ l 1.4 M DTT. Excess reagents were removed by dialysis against ultra-high quality (UHQ) water for 2–3 days after which the protein was lyophilized.

Proteolytic Digestions

Digestion with Endoproteinase Asp-N The reduced and alkylated protein was dissolved in 200 μ l 50 mM Tris, 0.2 M NaCl, 1 M urea, pH 8.5. 0.4 μ g endoproteinase Asp-N was added and the reaction stopped after 18 hours at 37 °C by freezing. The obtained peptides were separated by rp-HPLC. All fractions were analyzed by MALDI MS and relevant peptides were sequenced by automatic Edman degradation.

Digestion with Trypsin and Chymotrypsin The native protein (1 nmol) was dissolved in 3–4 μ l 8 M urea and vortexed for one min. 100 μ l of 50 mM NH₄Ac pH 6.0 and 5% (w/w) of trypsin was added. After 18 hours at 37 °C the progress of the digestion was checked by MALDI MS followed by the addition of 2% (w/w) chymotrypsin. After an additional 4 hours the reaction was stopped by freezing and the peptides separated by HPLC.

Digestion with V8 Protease The native protein (1 nmol) was dissolved in 200 μ l of 100 mM NH₄Ac, pH 4.0, followed by addition of 2.5 μ g *Staphylococcus aureus* V8-protease. After incubation at 37 °C for 18 hours, the reaction was stopped by freezing. The obtained peptides were separated by rp-HPLC and all fractions analyzed by MALDI MS. The peptides of interest were reduced on target and re-analyzed.

Digestions with Glycosidases

The glycosidases listed in Table 1 were used.

The fraction containing the glycopeptide was lyophilized, redissolved in 10 μ l of 100 mM NH₄Ac, pH 6.0, and 0.2 μ l of the appropriate glycosidase solution was added. The reaction took place at 37 °C for 18 hours after which the resulting product was analyzed by MALDI MS. Treatment with the different glycosidases was performed sequentially, i.e., the product of one glycosidase treatment was treated with the next.

Reversed phase-HPLC

The HPLC system consisted of two LKB 2150 HPLC-pumps (Pharmacia, Sweden) controlled by a PC and an LKB 2151 variable wavelength UV detector operating at 214 nm (Pharmacia, Sweden). The intact protein was purified using a Vydac C₄ column (Hesperia, CA, USA) while all digestions were purified using a Nucleosil C₁₈ column (Machery-Nagel, Düren, Germany). Gradients are shown in the Figures. Gradient elution was performed at a flow rate of 1 ml/min. Buffer A was 0.1% trifluoroacetic acid (TFA) and buffer B was 0.08% TFA in 90% acetonitrile.

Mass spectrometric analysis

Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI MS)

All mass spectra were obtained on a Bruker Reflex TOF instrument (Bruker-Franzen Analytik, Bremen, Germany). Data were acquired by a LeCroy 9450A 400M samples/sec digital storage oscilloscope (LeCroy Corporation, Chestnut Ridge, NY, USA) from which single shot spectra were transferred to a Macintosh Quadra 950 computer (Apple Computer Inc., Cupertino, CA, USA) via a National Instruments NI-DAQ GPIB controller board (National Instruments, Austin, TX, USA). Control of data acquisition parameters as well as all further data processing was carried out using the computer program 'Laser-One' (M. Mann and P. Mortensen, EMBL, Heidelberg, Germany). All mass spectra were obtained in the linear mode and calibrated

using the singly charged matrix ion, resulting in a mass accuracy of $\pm 0.1\%$ or better.

Samples were dissolved in 0.1% TFA to a concentration of approximately 0.05 mg/ml. 1 μ l of sample solution was applied to a stainless steel probe tip pre-coated with 1 μ l matrix solution (α -cyano 4-hydroxy cinnamic acid dissolved in acetone, 15 mg/ml) and thoroughly washed with 3×10 μ l 0.1% TFA before introduction into the mass spectrometer (Vorm *et al.*, 1994).

Reduction on target Relevant peptides were reduced on the probe tip by placing 3 μ l of DTT solution (155 mM DTT in 100 mM NH₄HCO₃, pH 7.8) as a droplet on top of the sample. The sample was left to dry for 10 min and a new layer of matrix applied, followed by washing as described above (Kusmann *et al.*, 1997).

Plasma desorption mass spectrometry (PDMS) The peptides below approximately 800 Da were analyzed by PDMS using a ²⁵²Cf time-of-flight plasma desorption mass spectrometer (Bio-Ion 20, Applied Biosystems AB, Uppsala, Sweden). The lyophilized HPLC-fractions were dissolved in 5 μ l 0.1% TFA and applied to nitrocellulose coated targets using the spin-cast technique (Nielsen *et al.*, 1988). The spectra of the peptides were accumulated for $0.5 - 1.0 \times 10^6$ fission events at an accelerating voltage of 15 kV.

Edman Degradation

Automated Edman degradation was performed by a Knauer 910 pulsed-liquid gasphase sequencer. The peptide samples were dissolved in 0.1% TFA or 70% formic acid and spotted onto polybrene coated PVDF membranes. The phenylthiohydantoin (PTH)-derivatives were identified by gradient elution on a narrow-bore HPLC using a 2×250 mm column packed with 5 mm SuperSpher C18 (Merck, Darmstadt, Germany). Buffer A was sodiumacetate, pH 4.6/7% acetonitrile, and buffer B 100% acetonitrile.

Monosaccharide Analysis

Approximately 360 pmol of the native allergen was hydrolyzed in 2 M TFA for 5 hours at 100 °C to release the monosaccharides from the protein. The sample was lyophilized and the released monosaccharides separated from the remaining protein by HPLC using a Vydac C₄ column (monosaccharides are not retained by the column and elute in the void volume). The isolated monosaccharides were lyophilized and dissolved in 150 μ l UHQ water. 50 μ l (approximately 50 pmol) were analyzed by HPAEC PAD.

The system used for monosaccharide analysis consisted of two gradient pumps (LKB-Pharmacia, Uppsala, Sweden) and a pulsed amperometric detector. Monosaccharides were separated using a CarboPac PA1 column with a CarboPac PA1 guard column (Dionex, California, USA). The flow rate was 0.2 ml/min and all samples were eluted with 16 mM NaOH using isocratic conditions. The separated monosaccharides were detected by a gold working electrode. To minimize baseline drift 300 mM NaOH was added to the postcolumn effluent in a mixing chamber at a flow rate of 1 ml/min. The following pulse potentials and durations were used: E₁ = 0.05 V (*t*₁ = 360 ms); E₂ = 0.75 V (*t*₂ = 120 ms) and E₃ = -0.15 V (*t*₃ = 420 ms).

Buffers were prepared by dilution of 50% NaOH solution with UHQ water. Dissolved carbon dioxide was removed by a continuous sparge of helium. The column was washed after each separation for 15 min with 300 mM NaOH and allowed to equilibrate for another 15 min with 16 mM NaOH.

References

- Bond, J.F., Brauer, A.W., Segal, D., Nault, A.K., Rogers, B.L., and Kuo, M.-C. (1993). Native and recombinant Fel d1 as probes into the relationship of allergen structure to human IgE immunoreactivity. *Mol. Immunol.* 30, 1529–1541.
- Duffort, O.A., Carreira, J., Nitti, G., Polo, F., and Lombardero, M. (1991). Studies on the biochemical structure of the major cat allergen *Felis domesticus* 1. *Mol. Immunol.* 28, 301–309.
- Griffith, I.J., Craig, S., Pollock, J., Yu, X., Morgenstern, J.P., and Rogers, B.L. (1992). Expression and genomic structure of the genes encoding Fd1, the major allergen from the domestic cat. *Gene* 113, 263–268.
- Kleine-Tebbe, J., Kleine-Tebbe, A., Jeep, S., Schou, C., Löwenstein, H., and Kunkel, G. (1993). Role of the major allergen (Fel d1) in patients sensitized to cat allergens. *Int. Arch. Allergy Immunol.* 100, 256–262.
- Kusmann, M., Nordhoff, E., Nielsen, H.R., Larsen, M.R., Haebel, S., Mirgorodskaya, E., Jensen, C., and Roepstorff, P. (1997). Sample preparation in MALDI-MS designed for various peptide and protein analytes. *J. Mass. Spectrom.*, in press.
- Leitermann, K., and Ohmann, J.L. (1984). Cat allergen 1: Biochemical, antigenic and allergic properties. *J. Allergy Clin. Immunol.* 74, 147–153.
- Lis, H., and Sharon, N. (1993). Protein glycosylation. Structural and functional aspects. Review. *Eur. J. Biochem.* 218, 1–27.
- Morgenstern, J.P., Griffith, I.J., Brauer, A.W., Rogers, B.L., Bond, J.F., Chapman, M.D., and Kuo, M. (1991). Amino acid sequence of Fel d1, the major allergen of the domestic cat: Protein sequence analysis and cDNA cloning. *Proc. Natl. Acad. Sci. USA* 88, 9690–9694.
- Nielsen, P.F., Klarskov, K., Højrup, P., and Roepstorff, P. (1988). Optimization of sample preparation for plasma desorption mass spectrometry of peptides and proteins using a nitrocellulose matrix. *Biomed. Environ. Mass. Spectrom.* 17, 355–362.
- Nielsen, P.F., and Roepstorff, P. (1989). Suppression effects in peptide mapping by plasma desorption mass spectrometry. *Biomed. Environ. Mass. Spectrom.* 18, 131–137.
- Rogers, B.L., Morgenstern, J.P., Garman, R.D., Bond, J.F., and Kuo, M. (1993). Recombinant Fel d1: Expression, purification, IgE binding and reaction with cat-allergic human T-cells. *Mol. Immunol.* 30, 559–568.
- Schou, C. (1993). Defining allergens of mammalian origin. *Clin. Exp. Allergy* 23, 7–14.
- Schäfer-Nielsen, C., and Rose, C. (1982). Separation of nucleic acids and chromatin proteins by hydrophobic interaction chromatography. *Biochem. Biophys. Acta* 696, 323–331.
- Svendsen, P.J. (1973). Fused rocket immunoelectrophoresis. *Scand. J. Immunol.* 1, 69–70.
- Vailes, L.D., Li, Y., Bao, Y., DeGroot, H., Aalberse, R.C., and Chapman, M.D. (1994). Fine specificity of B-cell epitopes on *Felis domesticus* allergen 1 (Fel d1): Effect of reduction and alkylation or deglycosylation on Fel d1 structure and antibody binding. *J. Allergy Clin. Immunol.* 93, 22–33.
- van Milligen, F.J., van't Hoof, W., van den Berg, M., and Aalberse, R.C. (1993). IgE epitopes on the cat (*Felis domesticus*) major allergen Fel d1: A study with overlapping synthetic peptides. *J. Allergy Clin. Immunol.* 93, 34–43.
- van Neerven, R.J.J., van de Pol, M.M., van Milligen, F.J., Jansen, H.M., Aalberse, R., and Kapsenberg, M.L. (1994). Characterization of cat dander specific T lymphocytes from atopic patients. *J. Immunol.* 152, 1203–1210.
- Vorm, O., Roepstorff, P., Mann, M. (1994). Matrix surfaces made by fast evaporation yield improved resolution and very high sensitivity in MALDI TOF. *Anal. Chem.* 66, 3281–3287.

Received April 10, 1997; accepted June 11, 1997