

Mass Spectrometric Investigation of Molecular Variability of Grass Pollen Group 1 Allergens

François Fenaille,^{*,†} Emmanuel Nony,[§] Henri Chabre,[§] Aurélie Lautrette,[§]
Marie-Noëlle Couret,[§] Thierry Batard,[§] Philippe Moingeon,[§] and Eric Ezan[†]

CEA, iBitec-S Service de Pharmacologie et d'Immunoanalyse, 91191 Gif-sur-Yvette, France, and Stallergènes SA,
6 rue Alexis de Tocqueville, 92160 Antony, France

Received April 21, 2009

Natural grass pollen allergens exhibit a wide variety of isoforms. Precise characterization of such microheterogeneity is essential to improve diagnosis and design appropriate immunotherapies. Moreover, standardization of allergen vaccine production is a prerequisite for product safety and efficiency. Both qualitative and quantitative analytical methods are thus required to monitor and control the huge natural variability of pollens, as well as final product quality. A proteomic approach has been set up to investigate in depth the structural variability of five group 1 allergens originating from distinct grass species (Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1). Whereas group 1 is the most conserved grass pollen allergen, great variations were shown between the various isoforms found in these five species using mass spectrometry, with many amino acid exchanges, as well as variations in proline hydroxylation level and in main *N*-glycan motifs. The presence of O-linked pentose residues was also demonstrated, with up to three consecutive units on the first hydroxyproline of Ant o 1. In addition, species-specific peptides were identified that might be used for product authentication or individual allergen quantification. Lastly, natural or process-induced modifications (deamidation, oxidation, glycation) were evidenced, which might constitute useful indicators of product degradation.

Keywords: Group 1 grass pollen allergens • MALDI-TOF • ESI-Orbitrap • amino acid sequence • species-specific peptides • post-translational modifications • proline hydroxylation • N- and O-glycosylation

Introduction

Grass pollens constitute a major worldwide source of airborne allergens.¹ Grasses belong to the Poaceae family, which represents the fourth largest group of flowering plants. They are classified into five subfamilies, among which Pooideae can be found mainly in the temperate zones, such as the Mediterranean area.² Europeans and North Americans are simultaneously exposed to pollens from various species belonging to the Pooideae subfamily. Pooideae extracts demonstrate a high level of IgE cross-reactivity, likely due to protein sequence homologies and/or the existence of cross-reactive carbohydrate determinants (CCDs).^{3,4} Recently, an extract from pollen mixtures of five different species, namely, *Anthoxanthum odoratum*, *Dactylis glomerata*, *Lolium perenne*, *Phleum pratense*, and *Poa pratensis*, has been proposed as a combination for grass pollen-specific immunotherapy.⁵ Such a mixture constitutes a pertinent representation of exposure and sensitization conditions encountered in Western Europe by grass pollen-allergic patients.⁵ As for other health-oriented products, allergen vaccine production needs to be perfectly standardized,

to further guarantee product safety and efficiency. Accurate analytical methods are thus required to monitor and control the huge natural variability of pollens. The precise nature and composition of the extracts deserve to be thoroughly assessed both qualitatively and quantitatively, which still represents a major challenge in the field. The present study represents a first step toward product standardization, using purified group 1 allergens from the above-mentioned pollen mixtures as working material and mass spectrometry (MS) as analytical tool. In-depth qualitative evaluation of natural (e.g., amino acid variation, post-translational modifications (PTMs)) and potentially process-induced (e.g., oxidation, glycation) allergen molecular variability was performed. The resulting information was also used to highlight potential species-specific sequences, in order to obtain allergen identification.

Group 1 allergens are among the most prominent allergenic determinants in grass pollen extracts. They represent a family of glycoproteins of approximately 30 kDa that occur as cross-reactive antigens in all common grass species, with an amino acid sequence homology of ~90%.¹ Up to now, both their amino acid sequences and PTMs were only partially reported. Because of complexity of the allergen sources, their amino acid sequences were mainly deduced from cDNA. Table 1 summarizes the sequence information available from the Swiss-Prot database for the five group 1 allergens originating from species that compose the above-mentioned pollen mixtures,

* To whom correspondence should be addressed. Phone: 33-1-69-08-79-54. Fax: 33-1-69-08-59-07. E-mail: francois.fenaille@cea.fr. Mailing address: CEA, iBitec-S, Service de Pharmacologie et d'Immunoanalyse, Bât. 136, CEA Saclay, 91191 Gif-sur-Yvette France.

[†] CEA.

[§] Stallergènes SA.

Table 1. Summary of the Sequence Information Available on the Swiss-Prot Database for the Five Group 1 Allergens Studied

allergen	Swiss-Prot accession number	calculated average molecular mass (Da) ^a
Ant o 1	Q7M1X6	3405 ^b
	Q7M1Y0	3419 ^b
Dac g 1.01	Q7XAX8	26143
Dac g 1.02	Q7XAX7	26129
Lol p 1	P14946	26186
	Q9SC98	26109
Phl p 1	P43213	26151
	Q40967	25925
Poa p 1	Q9ZP03	25980

^a Average masses were calculated without signal peptide and considering three disulfide bonds. ^b Masses calculated from partial sequences with only the 32 N-terminal amino acids reported.

namely, Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1. As may be observed at first glance from Table 1, the sequence of Ant o 1 is restricted to its N-terminal part, which well reflects the frequent use of N-terminal Edman sequencing for allergen identification. For instance, to the best of our knowledge, the literature only reports analysis of the first 10–30 N-terminal amino acid residues for Dac g 1,⁶ Lol p 1,⁷ and Phl p 1,^{7,8} with almost no data available for Ant o 1 and Poa p 1. Predictions of the amino acid sequences of Lol p 1 and Phl p 1 from cDNA can also be found,^{9,10} while mass spectrometric analysis of these two proteins after gel electrophoresis from pollen extracts revealed only 20–50% sequence coverages.^{11,12}

Regarding PTMs, two hydroxyproline (Hyp) residues were detected by Edman sequencing at positions 5 and 8 for Lol p 1 and Phl p 1.^{7,13} An N-glycosylation site was suspected at position 9 for Dac g 1, Lol p 1, and Phl p 1 since almost no residue was detected at this site by Edman sequencing, which is also in agreement with the predicted N-glycosylation consensus sequence NIT.^{7,14} It was only shown for Phl p 1 allergen that the major N-glycan forms consisted of a Man₃GlcNAc₂ or Man₂GlcNAc₂ core with an α (1,3)-fucose attached to the proximal N-acetyl-glucosamine (GlcNAc) residue and/or a β (1,2)-xylose residue attached to the β -mannose (known as MMXF³ and M0XF³ motifs, respectively).¹⁵ These structures are consistent with previous ones from grass pollen extracts.¹⁶ N-glycan structures are known to be involved in IgE binding,^{3,17} thus, contributing to cross-reactivity between grass pollen allergens. These so-called CCDs do not constitute the single PTMs that are potential elicitors of IgE reactivity. Indeed, two different O-glycan structures were found on hydroxyproline residues of the major allergen of mugwort Art v 1.¹⁸ One of them, the β -arabinose moiety, was suggested to constitute a new, potentially cross-reactive, carbohydrate determinant in plant proteins.¹⁸ Such a structure was later identified on the Hyp residue at position 5 of Phl p 1.¹⁵ Noteworthy, a previous study has shown that hydroxylation of proline residues can already influence IgE reactivity of Phl p 1.¹⁹

Altogether, these data show a clear lack of precise structural information regarding the five different group 1 allergens constituting the above-mentioned pollen mixture, with almost no data available for Ant o 1 and Poa p 1 and only partial information on primary sequences and PTMs of Dac g 1, Lol p 1, and Phl p 1. In-depth structural characterization of their natural variability is required for a better understanding of allergic disorders and improvement of diagnostic methodolo-

gies. For instance, antibody-based assays, which are common analytical tools in the field, may react differently to distinct isoforms, thus, influencing test results. Regarding allergen extracts, allergenic potency as well as allergen content can also potentially lack batch-to-batch consistency.²⁰ Allergen quantification is a key issue that needs to be addressed. Furthermore, as already observed for other natural or recombinant therapeutic proteins, additional structural variations may also result from processing and/or storage conditions, thus, leading for example to peptide bond cleavage or primary structure modifications (e.g., oxidation, glycation).^{21–25} Such protein degradations obviously need to be assessed and controlled, to guarantee the integrity of the final product. Therefore, protein structure needs to be thoroughly characterized at all steps of the manufacturing process, in order to obtain well-standardized allergen products. In the case of pollen mixtures, additional species-specific information must also be obtained both for authentication and quantification purposes.

Nowadays, mass spectrometry is recognized as the technique of choice for both monitoring structural modifications and obtaining quantitative data on proteins.^{26,27} Moreover, mass spectrometry is also a useful tool for finding out whether an allergen isoform detected at the cDNA level is really expressed in quantities relevant for the patient sensitization, as shown recently for Bet v 1 and Mal d 1 allergens.²⁸ The present work was devoted to the detailed study of the five purified group 1 allergens Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1 allergens by mass spectrometric methods. In a first step, these five species were studied both at the protein and peptide levels by a combination of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and high resolution/high mass accuracy electrospray (ESI) mass spectrometry, to obtain structural information regarding primary sequence variations (e.g., variants, oxidized or modified residues). Particular attention was paid to the identification of species-specific peptides for each of the group 1 allergens to confirm their presence in the mixture. The second part of this work was devoted to characterization of the PTMs, and deeper insights were obtained regarding proline hydroxylation, and N- and O-glycosylation of the five group 1 allergens. Such structural information would be of potential interest to better understand differences in CD4⁺ T cell reactivity and IgE binding (Chabre et al., submitted).

Materials and Methods

Materials and Reagents. Dithiothreitol (DTT) and iodoacetamide were from Sigma-Aldrich (Saint Quentin Fallavier, France). Sequencing grade modified trypsin was from Promega (Madison, WI), whereas Endoproteinase Asp-N from *Pseudomonas fragi* and N-glycosidase A were from Sigma-Aldrich and Roche (Mannheim, Germany), respectively. RapiGest SF (lyophilized sodium-3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate) was from Waters (Milford, MA). The MALDI matrices 2,5-dihydroxybenzoic acid (DHB) and sinapinic acid (SA) were from Sigma-Aldrich, whereas α -cyano-4-hydroxycinnamic acid (HCCA) was from Laser BioLabs (Sophia Antipolis, France). Other chemicals and reagents were obtained from commercial sources at the highest level of purity available. All buffers were prepared using ultrapure water (Milli-Q, Millipore).

Allergen Purification. Group 1 allergens were purified from pollen extracts (Allergen AB, Sweden) obtained from five common allergenic grass species including *A. odoratum* (sweet

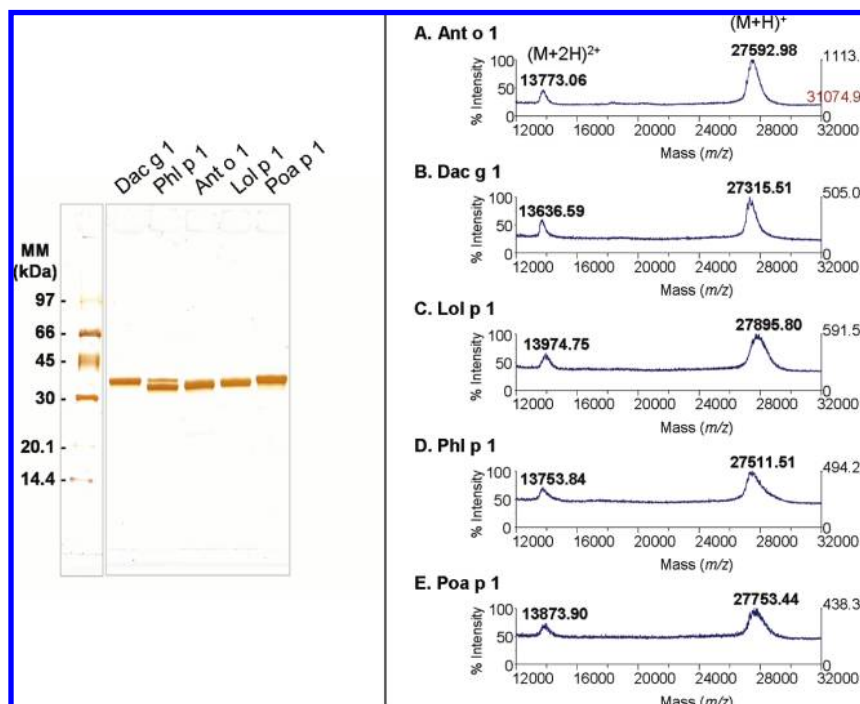


Figure 1. SDS-PAGE (left panel) and MALDI-TOF mass spectra (right panel) of natural group 1 allergens (A) Ant o 1, (B) Dac g 1, (C) Lol p 1, (D) Phl p 1, and (E) Poa p 1. SDS-PAGE analysis confirmed that purified group 1 allergens were >95% pure, while MALDI-TOF mass spectra exhibited broad peaks centered at ~27.5 kDa consistent with the coexistence of several isoforms.

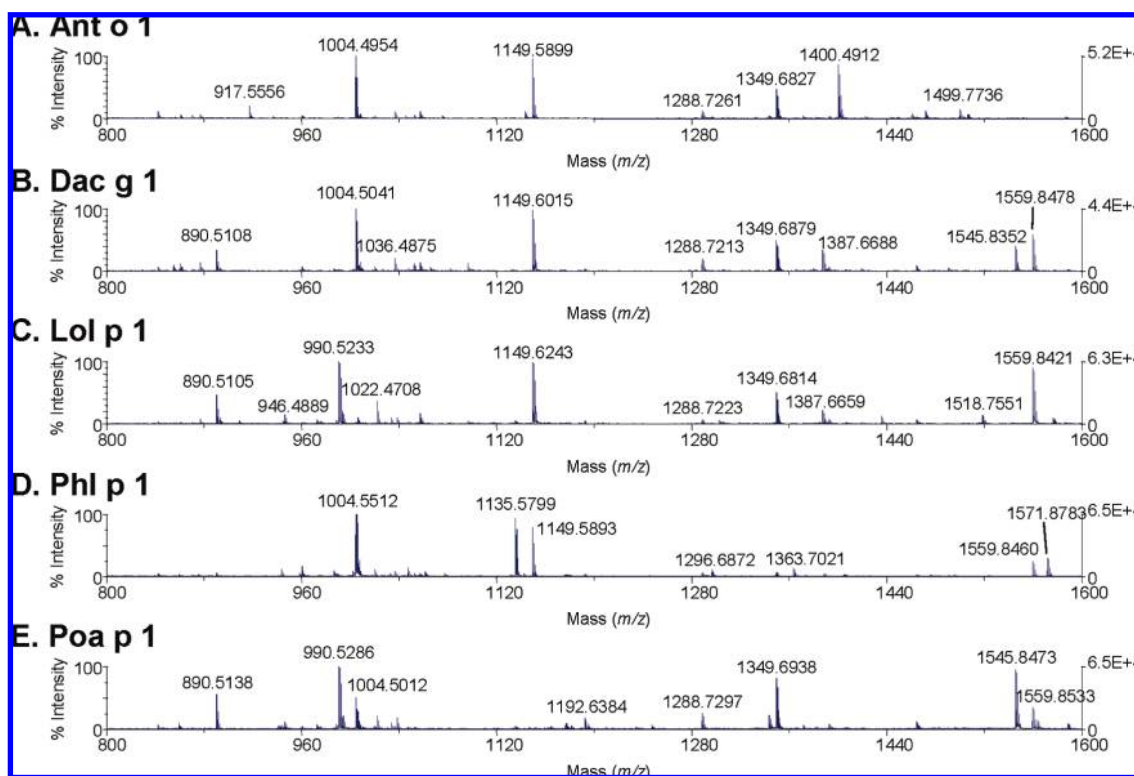


Figure 2. Part of MALDI-TOF mass spectra obtained for the tryptic peptides of group 1 allergens (A) Ant o 1, (B) Dac g 1, (C) Lol p 1, (D) Phl p 1, and (E) Poa p 1. Spectra were obtained from tryptic digests.

vernal), *D. glomerata* (orchard or cocksfoot), *L. perenne* (perennial rye), *P. pratense* (timothy) and *P. pratensis* (Kentucky blue grass). Purification was achieved using both hydrophobic interaction chromatography and size exclusion chromatography as described elsewhere (Chabre et al., submitted). Briefly, pollen extracts in 50 mM ammonium bicarbonate were first

purified using a phenyl-sepharose hydrophobic interaction column (GE Healthcare, Orsay, France), previously equilibrated with sodium phosphate 0.02 M and ammonium sulfate 1 M, pH 6.2. Group 1 elution was performed using sodium phosphate 0.02 M containing 5% of isopropanol (pH 7.6). Fractions were then concentrated by ultrafiltration, using an Amicon PM-

Pos. ^a	Sequence	Theor. Mono. Masses (M+H) ^{+b}	Measured Monoisotopic Masses (M+H) ^c									
			Ant o 1		Dac g 1		Lol p 1		Phl p 1		Poa p 1	
			LC-MS	MALDI-TOFMS	LC-MS	MALDI-TOFMS	LC-MS	MALDI-TOFMS	LC-MS	MALDI-TOFMS	LC-MS	MALDI-TOFMS
4-17	VPPGPNITATYGDK VPPGPNITAEYGDK	1429.7321 1457.7270	2454.0888 ^e	-	2470.0819 ^e	-	2498.0792 ^e	-	2454.0858 ^e	-	2470.0804 ^e	-
18-22	WLDAAK	632.3402	632.3397	632.3315	632.3401	632.3451	632.3397	632.3378	632.3400	632.3378	632.3397	632.3397
23-35	STWYGKPTGAGPK STWYGKPTAAGPK	1349.6848 1363.7004	1349.6847	1349.6827	1349.6831	1349.6879	1349.6833	1349.6814	1349.6816	1349.6924	1349.6818	1349.6938
36-44	DNNGACGYK DDGGACGYK	941.3781 942.3631	941.3765 942.3602	-	941.3755 942.3593	-	941.3767 942.3604	-	941.3750 942.3604	-	941.3758 942.3585	-
45-64	DVDKAPFNGMTGCGNTPIFK DVDKAPFDGMTGCGNTPIFK DVDKAPFNGMTGCGNTPIFK DVDKAPFDGMTGCGNTPIFK DVDKPPFSGMTGCGNTPIFK DVDKPPFSGMTGCGNTPIFK	2169.0103 2169.9943 2185.0052 2185.9892 2168.0150 2184.0099	2169.0096 2169.9915 - 2185.9875 2185.9898	- 2170.0027	2169.0057 2169.9900	- 2170.0016	2169.0090 2169.9921 2185.0014 2185.9857	2169.0065 2170.0084 - -	- -	- 2168.0124 2168.0597	2168.0115 2184.0039 1684.7758 1712.7966 1712.7792	- 2168.0399 - - -
49-64	APFSGMTGCGNTPIFK APFNGMTGCGNTPIFK APFDGMTGCGNTPIFK	1684.7821 1711.793 1712.7771	1711.7962 1712.7733	1711.7900 1712.7829	1711.7958 1712.7737	- 1712.7731	1711.7972 1712.7758	- 1712.7713	-	-	-	-
68-76	GCGSCFEIK	1057.4441	1057.4421	1057.4466	1057.4431	1057.4462	1057.4426	1057.4423	1057.4432	1057.4388	1057.4418	-
77-110	CTKPESCSGEPI/LLVHITDD NEEPIAAYHFDLSGK	3829.7683	-	-	-	-	-	-	-	-	3829.7635	-
87-117	AVTI/LTITDDNEEPIAPYHF DLSGHAFGAMAK	3331.5939	3331.5932	-	-	-	-	-	-	-	-	-
111-117	AFGAMAK	695.3545	695.3512	-	-	-	-	-	-	-	695.3527	695.3569
77-117	CTKPESCSGEAVTVHITDDN EEPIAPYHFDLSGHAFGSMAC CTKPESCSGEAVTVTITDDN EEPIAPYHFDLSGHAFGSMAC CTKPEACSGEPVVVHITDDN EEPIAPYHFDLSGHAFGAMAK CTKPEACSGEPVVVHITDDN EEPIAAYHFDLSGIAPGSMAC	4505.0118 4469.0006 4497.0583 4463.0627	-	-	4505.0005 4505.0179	-	4468.9980	-	-	-	4469.0035	-
126-133	SAGELELK	846.4567	-	-	-	-	-	-	-	-	846.4550	846.4654
126-135	SAGEVEIQFR SAGELEQFR	1135.5742 1149.5898	1149.5892	1149.5899	1149.5897	1149.6025	1149.5892	1149.6243	1135.5726 1149.5868	1135.5799 1149.5893	1135.5719 1149.5847	1135.5404 1149.5990
139-143	CEYPEK	696.3021	696.3000	-	-	-	-	-	-	-	696.3003	-
139-146	CKYPEGTK	982.4662	-	-	982.4637	-	-	-	-	-	-	-
141-146	YPEGTK	694.3406	-	-	694.3385	-	-	-	694.3393	-	-	-
147-153	VTFHVEK TFPHVEK	859.4672 861.4465	859.4675 861.446	-	859.467	859.4709	-	-	859.4661	859.4620	859.4657	859.4679
144-153	YPDDTKPTFFHVEK	1576.7642	-	-	-	-	1576.7642	1576.7622	-	-	-	-
144-153	GTKPTFHVEK	1143.6156	1143.6148	1143.6157	-	-	-	-	-	-	-	-
154-165	GSNPNYLALLVK ASNPNYLAILVK	1288.7259 1302.7416	1288.7257	1288.7261	1288.7247	1288.7213	1288.7244	1288.7223	1288.7247	1288.7323	1288.7233	1288.7297
166-179	YVGDGDVVAVDIK YVNGDGDVVAVDIK FVGDGDVVAVDIK FVAGDGDVVAVDIK FVGDGDVVAVADIK	1464.7216 1463.7376 1448.7267 1404.7369 1480.6988	1464.7219 - 1448.7272	1464.7209 - 1448.7400	1464.7205 1463.7414	1464.7273 -	1464.7206 -	1464.7178 -	1464.7197 1463.7333	1464.7322 -	1464.7181 -	1464.7251
186-190	WIALK WIELK	630.3973 688.4028	-</									

SDS-PAGE Analysis. Purified allergens (0.5 μ g) were diluted in sample buffer (62 mM Tris, 1% SDS, 0.01% bromophenol blue), boiled and applied onto a 12.5% Excel-gels (GE Health-

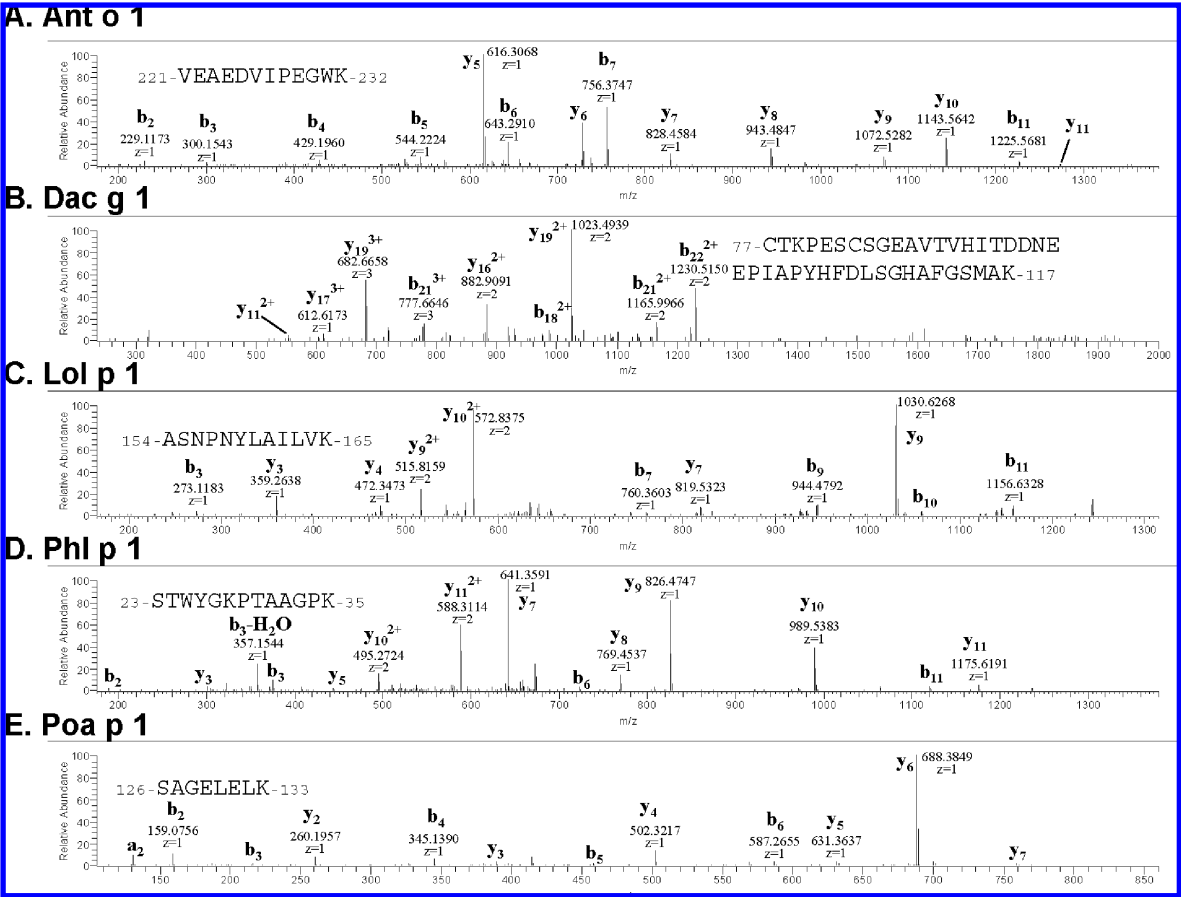


Figure 3. ESIMS and MS/MS spectra obtained for the species-specific peptides (A) Ant o 1, (B) Dac g 1, (C) Lol p 1, (D) Phl p 1, and (E) Poa p 1. Spectra were obtained from tryptic digests. Cysteine residues were alkylated with iodoacetamide.

Table 3. Potential Species-Specific Tryptic Peptides Identified for Group 1 Allergens

allergen	species-specific peptide	position	retention time (min)	(M + H) ⁺
Ant o 1	<i>(K)</i> VEAEDVIPEGWK ^a	(220)221–232^a	(17.7)18.5^a	(1499.7740)1371.6797^a
	TTFHVEK	147–153	11.2	861.4465
Dac g 1	CTKPESCSGEAVTVHITDDNEEPIAPYHFDLSGHAFGSMK	77–117	20.2	4505.0118
Lol p 1	ASNPNYLAILVK	154–165	21.0	1302.7416
	SEFEDVIPEGWK	221–232	20.5	1435.6739
Phl p 1	STWYGKPTAAGPK	23–35	13.3	1363.7004
	TEAEDVIPEGWK	221–232	18.0	1373.6583
Poa p 1	SAGELELK ^b	126–133^b	13.6	846.4567^b
	APFSGMTGCGNTPIFK	49–64	19.1	1684.7821

^aThe bracketed lysine residue accounts for the peptide bearing a missed cleavage site. ^bThis peptide is sometimes present in Ant o 1 at a concentration ~10-fold lower. The “most sensitive peptides” are shown in bold characters. Cysteine residues were alkylated with iodoacetamide.

care, Uppsala, Sweden). Electrophoresis was carried out under reducing conditions for 1 h at 80 V in parallel with molecular mass markers, and proteins were revealed using a silver staining kit (GE Healthcare).

Proteolytic Digestions. Ten microliters of purified group 1 allergen (~0.5–1 µg/µL) was mixed with 10 µL of a 0.2% (w/v) RapiGest SF²⁹ solution in 50 mM ammonium bicarbonate. Disulfide bridge reduction was accomplished by adding 10 µL of a 10 mM DTT solution in water and incubating the resulting mixture for 1 h at 60 °C. The solution was allowed to cool at room temperature, after which 10 µL of a 24 mM iodoacetamide solution in water was added, and the solution was incubated at room temperature for 45 min in the dark. After addition of 50 mM ammonium bicarbonate (15 µL), digestion with trypsin (5 µL of a 100 ng/µL aqueous solution) or Asp-N (5 µL of a 40 ng/µL aqueous solution) was performed by an overnight

incubation at 37 °C. Protein digests were acidified by addition of 5 µL of 1 M HCl, and further incubated at 37 °C for 45 min to hydrolyze RapiGest SF, and finally centrifuged at 13 000 rpm for 5 min to remove insoluble material.

Enrichment of Glycopeptides by Hydrophilic Interaction Chromatography (HILIC). Glycopeptide enrichment by HILIC microcolumns was adapted from a published procedure.³⁰ Briefly, 35 µL of enzymatic digest was mixed with 140 µL of acetonitrile (ACN) containing 0.5% of trifluoroacetic acid (TFA), and loaded on a HILIC Uptitip (Interchim, Montluçon, France) equilibrated with 150 µL of 80% ACN containing 0.1% TFA. The bound peptides were washed twice with the same buffer and eluted with 20 µL of 0.1% TFA.

Deglycosylation Procedure. Glycopeptide-enriched fraction obtained by HILIC was evaporated to dryness and further

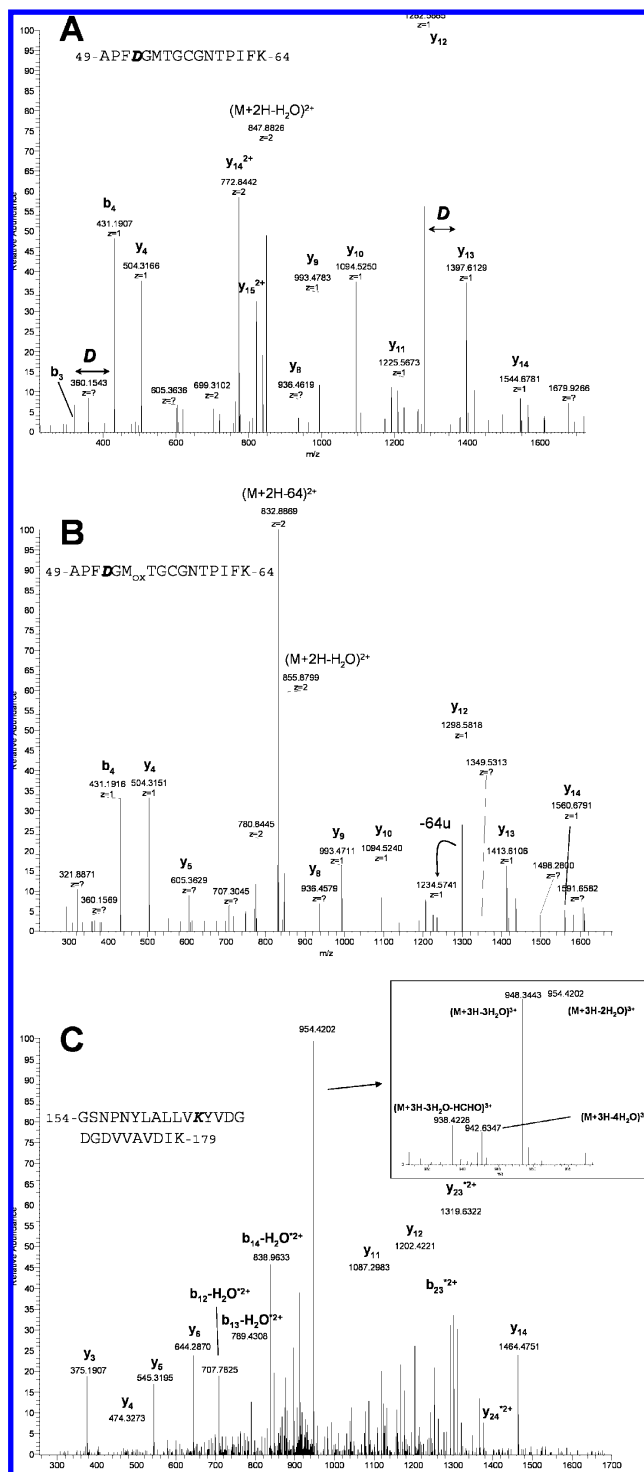


Figure 4. ESIMS/MS spectra of some modified peptides (A) deamidated, (B) oxidized, and (C) glycosylated. Spectra were obtained from Dac g 1 tryptic digest. For spectrum (C), fragment ions were monitored in the LTQ part. Cysteine residues were alkylated with iodoacetamide.

resuspended in 25 μ L of 50 mM sodium acetate pH 5 and digested with 0.05 mU of *N*-glycosidase A for 48 h at 37 $^{\circ}$ C.

MALDI-TOF MS Analysis. Peptide analyses were carried out on a Voyager-DE STR instrument (Applied Biosystems, Les Ulis, France) equipped with a pulsed nitrogen laser (337 nm). HCCA (10 mg/mL in 50% ACN containing 0.1% TFA) and DHB (saturated solution in ACN) matrices were used for the analysis

of proteolytic peptides and glycopeptides, respectively. Protein analysis was performed using SA as matrix (10 mg/mL in 50% ACN containing 0.1% TFA). In each case, 0.5 μ L of analyte solution and 0.5 μ L of matrix solution were deposited and thoroughly mixed on the target. For peptide analysis, the instrument was operated in the delayed extraction reflector mode with a 20 kV acceleration voltage and an extraction delay of 140 ns. The laser power was carefully monitored in order to be high enough to have sufficient signal/noise ratio, but low enough in order to avoid detector saturation and nonlinear response. Data were accumulated over the m/z ratio window of 700–5000 Th. All samples were prepared in duplicate and spotted, and spectra comprising 500 averaged laser shots each were acquired from two different regions of each spot to give four mass spectra per sample. Mass spectra were calibrated either internally with trypsin autolysis peaks at m/z 842.51 and m/z 2211.10 or externally using a mixture of standard peptides when these peaks were not observed. For protein analysis, ions were analyzed in the linear mode after acceleration at 25 kV, with an extraction delay of 750 ns. Mass calibration was performed externally.

LC-ESIMS/MS Conditions. Chromatographic separations were performed on a Zorbax 300SB C18 HPLC column (2.1 $\times$ 150 mm, 5 μ m, 300 $\text{\AA}$, Interchim, Montluçon, France) as previously described.³¹ Mobile phases consisted of 0.1% formic acid (A) and ACN containing 0.1% formic acid (B). Peptides were eluted from the column at 300 μ L/min using a 5–60% phase B gradient over 35 min.

The HPLC eluent was directly connected to the ESI probe of an LTQ-Orbitrap Discovery mass spectrometer (Thermo, San Jose, CA). The source conditions were as follows: capillary temperature, 275 $^{\circ}$ C; sheath gas flow, 30 arbitrary units; auxiliary gas flow, 3 arbitrary units; capillary voltage, 38 V; ESI spray voltage, 4.5 kV. The target was fixed at 1×10^7 ions and the automatic gain control was turned on. The instrument was operated in the data-dependent acquisition mode, allowing the automatic switching between MS and MS/MS. The MS survey scan was performed from m/z 300–2000 in the Orbitrap, using a resolution set at 30 000 (at m/z 400), and the ion population was held at 5×10^5 through the use of automatic gain control. The three most abundant ions (threshold 500 counts, charge states higher than +1) were further selected for collision-induced dissociation (CID) experiments. The CID fragment ions were detected in separate experiments, either in the linear ion trap or in the Orbitrap for *de novo* sequencing. In both cases, a normalized collision energy of 35% was used, with an activation q of 0.25 and an activation time of 30 ms. For tandem mass spectrometry in the linear ion trap, the ion population was set to 1×10^4 and the precursor isolation width was set to 3 m/z units. Detection in the Orbitrap was performed at a target resolution of 7500, whereas the ion population was set at 2×10^5 , and the precursor isolation width to 3 m/z units. *De novo* sequencing was performed both manually and occasionally with the help of the PepNovo program^{32,33} to generate peptide sequence tags. The PepNovo software is freely available at <http://bix.ucsd.edu>.

Results and Discussion

This in-depth investigation of the molecular variability of five purified group 1 allergens used mass spectrometric as an analytical tool. These different species were studied both at the protein and peptide levels by a combination of MALDI-TOF and high-resolution/high mass accuracy electrospray mass

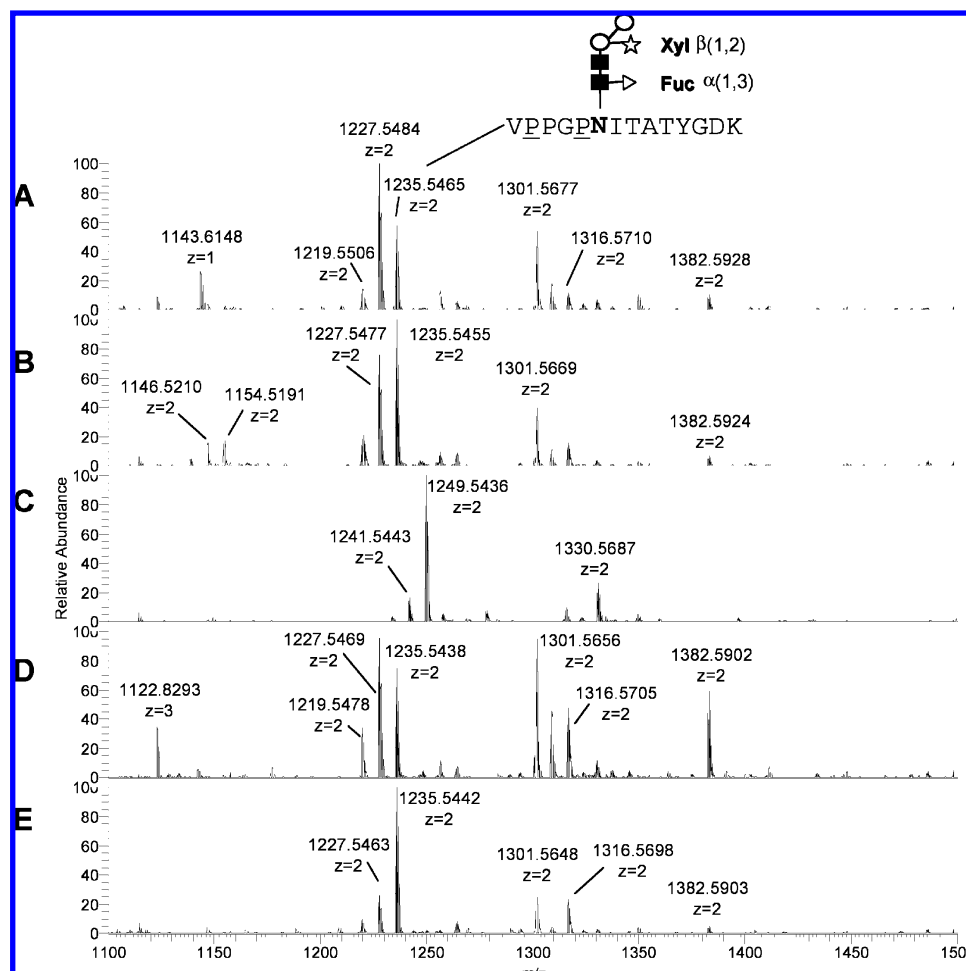


Figure 5. ESI mass spectra representing the glycopeptide population observed for each group 1 allergen. (A) Ant o 1, (B) Dac g 1, (C) Lol p 1, (D) Phl p 1, and (E) Poa p 1. Spectra were obtained from tryptic digests by averaging over a 5-min retention time window (from 10 to 15 min). (P) represents a hydroxyproline residue.

spectrometry, to obtain structural information regarding primary sequence variations and PTMs.

Global Characterization of Purified Group 1 Allergens by MALDI-TOFMS. Group 1 allergens were purified from pollen extracts obtained from 5 common grass species as described in the Materials and Methods. Before MS analyses, purity and global protein profile were verified using SDS-PAGE with silver staining. As shown in Figure 1 (left panel), all group 1 allergens resolve as a single band with similar molecular masses, with the exception of Phl p 1, which appears as a doublet potentially highlighting the presence of two distinct isoforms. Purified allergens were apparently not contaminated by other proteins.

The mass spectra obtained for the different purified group 1 allergens exhibit broad peaks centered at ~27.5 kDa (Figure 1, right panel), consistent with the coexistence of several isoforms. Similar heterogeneity extents were observed for all the allergens. The determined average molecular masses are ~1000–1500 Da above those calculated for the published amino acid sequences (Table 1). These mass increments are likely to represent isoforms bearing one M0XF³ N-glycan motif on Asn₃ (i.e., +1008 Da), which is known to be the major one in grass pollens,¹⁶ along with other structural modifications such as proline hydroxylation.^{7,34} This hypothesis is supported by the observation of ~160 Da mass increments (i.e., hexose additions) in the mass spectrum of some species (e.g., Dac g

1, Figure 1B), that would well fit the N-glycan structures previously reported in various grass pollens.¹⁶ N-glycosylation motifs have been shown to be well-conserved and structurally homogeneous in various pollen allergens,¹⁶ which tends to indicate the presence of amino acid substitutions to explain the broad unresolved peaks observed in the MALDI-TOF mass spectra. To obtain more precise information regarding structural heterogeneity, group 1 allergens were then studied after enzymatic digestion by both MALDI-TOFMS and liquid chromatography coupled with high-resolution/high mass accuracy electrospray mass spectrometry (LC-ESIMS/MS) using an LTQ-Orbitrap mass spectrometer.

Characterization of Amino Acid Sequences from Group 1 Allergens after Enzymatic Digestion. To obtain more insights into group 1 allergen structural microheterogeneity, purified allergens were digested with trypsin (and Asp-N for confirmation purposes) and the resulting proteolytic peptides were first fingerprinted by MALDI-TOFMS and then thoroughly analyzed and sequenced by LC-ESIMS/MS.

1. Peptide Mass Fingerprinting by MALDI-TOFMS. In a first set of experiments, peptides resulting from trypsin and Asp-N digestion of group 1 allergens were analyzed by MALDI-TOFMS. For instance, Figure 2 shows the mass spectra obtained for the Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1 tryptic digests. For clarity, Figure 2 shows only the most informative *m/z*

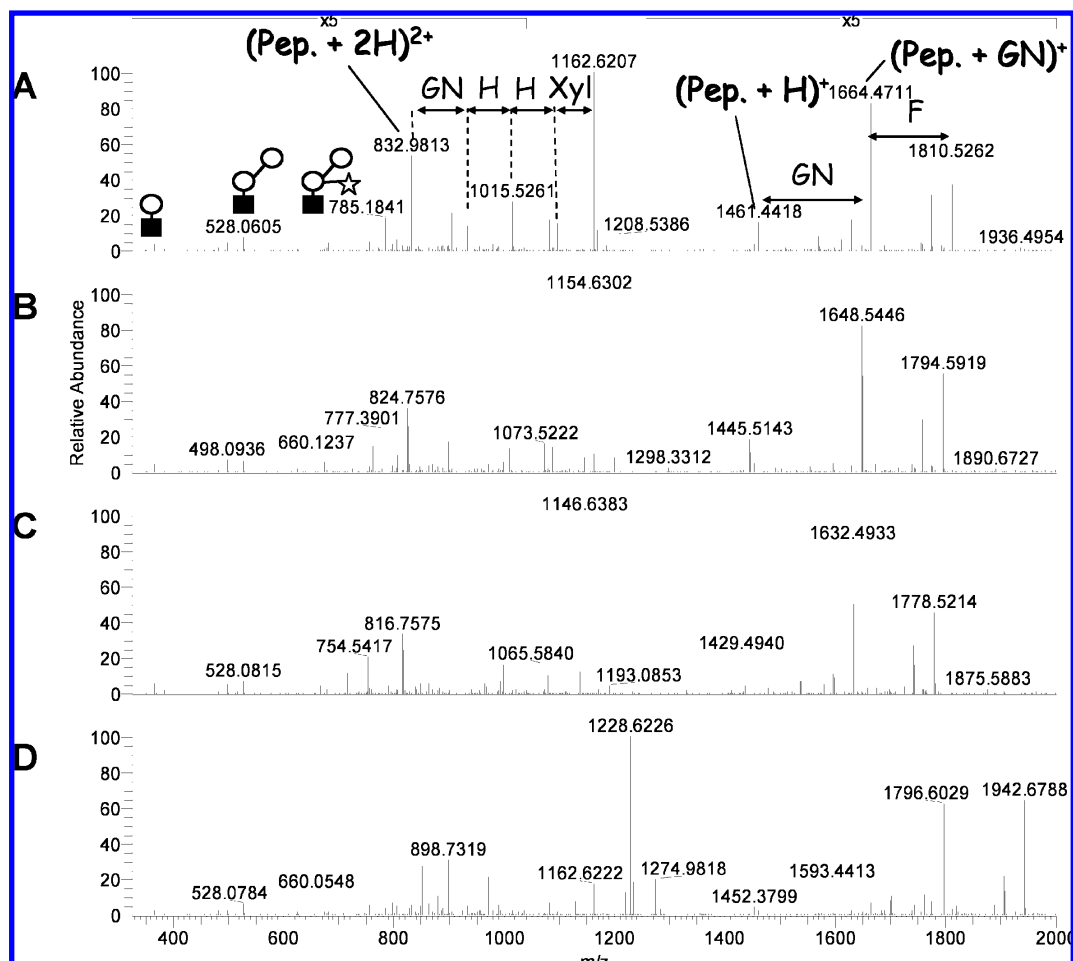


Figure 6. MS/MS spectra of (Val₄–Lys₁₇) glycopeptide bearing the M0XF³ motif and (A) two Hyp residues (*m/z* 1235.55), (B) one Hyp residue (i.e., Hyp₈, *m/z* 1227.55), (C) no Hyp residue (*m/z* 1219.55), and (D) two Hyp and one O-linked pentose residues (on Hyp₅, *m/z* 1301.57). All fragmented ions are doubly charged ions. Spectra were obtained from Dac g 1 tryptic digest.

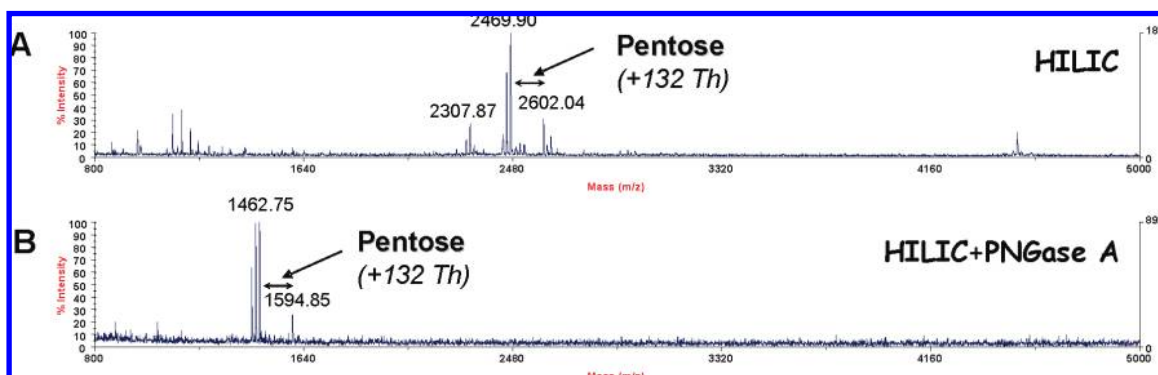


Figure 7. MALDI-TOF mass spectra of HILIC-enriched glycopeptides (A) before and (B) after PNGase A digestion. Spectra were obtained from Dac g 1 tryptic digest.

800–1600 portion of the spectra. Although most of the observed peaks can be putatively identified by mass matching with the theoretical peptides derived from the sequences reported in databases, sequences were further confirmed/elucidated by LC-MS/MS (*vide infra*). The main identified peptides along with their respective sequences are summarized in Table 2. As expected, strong homologies were observed between the tryptic peptides of the five group 1 allergens. According to these data, ~5 peptides observed in MALDI-TOF mass spectra were shared by all the allergens. Nevertheless, it must be mentioned that some fluctuations in peak intensity can be observed from one

allergen to another, which potentially reflects the coexistence of isoforms. For example, the ion at *m/z* 1349.7, which corresponds to the peptide 23-STWYGKPTGAGPK-35, is among the most prominent peaks for Ant o 1, Dac g 1, Lol p 1, and Poa p 1, while it is far less intense for Phl p 1. In the latter case, an accompanying peak appears at *m/z* 1363.7, which accounts for an amino acid substitution of the second glycine by an alanine residue (Table 2). The peaks common to all species are represented on a gray background in Table 2.

2. Peptide Analysis and *de Novo* Sequencing of Unassigned Peptides by LC-MS/MS. Data generated upon LC-MS/

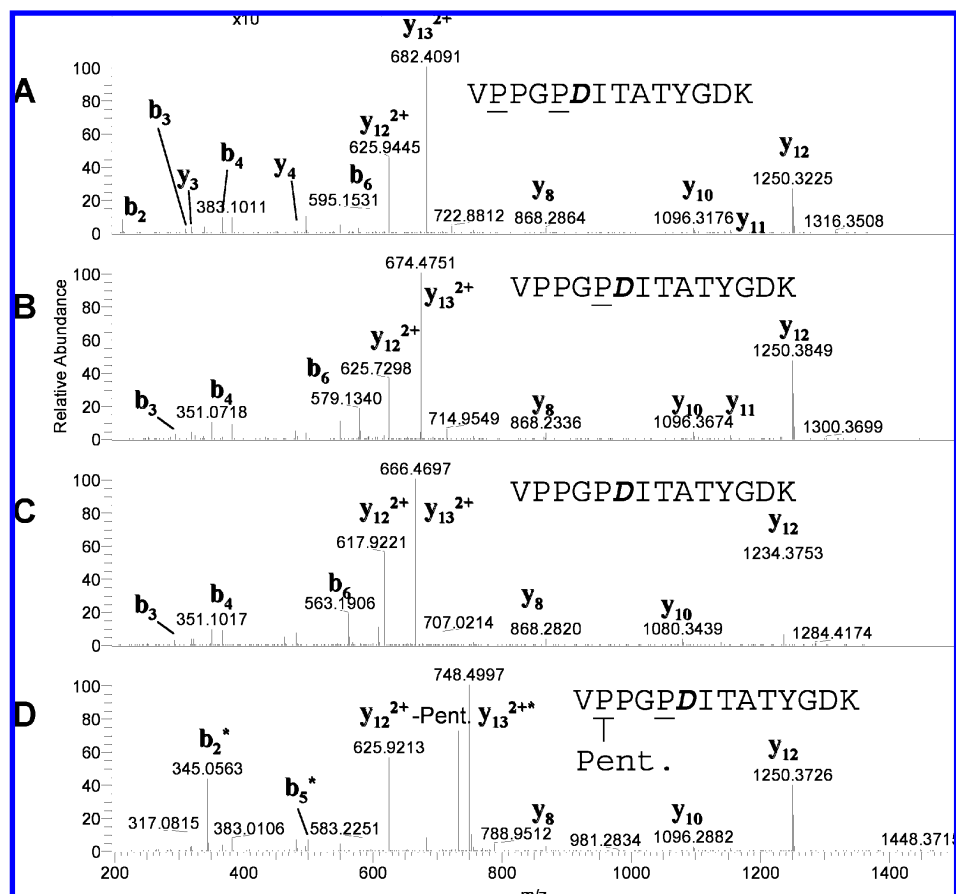


Figure 8. MS/MS spectra of HILIC-enriched and deglycosylated (Val₄–Lys₁₇) glycopeptide bearing (A) two Hyp residues (m/z 731.80), (B) one Hyp residue (i.e., Hyp₈, m/z 723.80), (C) no Hyp residue (m/z 715.80), and (D) two Hyp and one pentose residues (on Hyp₅, m/z 797.80). All fragmented ions are doubly charged ions. Spectra were obtained from Dac g 1 tryptic digest. Fragment ions bearing one pentose residue are labeled with an asterisk. (-Pent.) accounts for a loss of a pentose residue from the parent ion, while (P) represents a hydroxyproline residue. Because of the chemical lability of the *O*-glycosidic bond under MS/MS conditions, neutral losses of pentose residues can be observed.

MS analysis of allergen digests were first matched with sequences in the databases (Table 1) and the unmatched MS/MS spectra were submitted to manual and computer-aided *de novo* sequencing. The PepNovo software was used to generate peptide sequence tags. Although successful *de novo* sequencing results can be obtained with low-resolution ion trap instruments,^{35,36} both high resolution and high mass accuracy in both MS and MS/MS mode are desirable for obtaining high-confidence sequences.^{37,38} Thus, in this study, allergen digests (trypsin and Asp-N) were analyzed by LC-MS/MS at high resolution and high mass accuracy on an LTQ-Orbitrap instrument. LC-MS chromatograms obtained for the five tryptic digests are presented in Figure S1 of Supporting Information, while examples of MS and MS/MS spectra obtained with ion detection in the Orbitrap analyzer are given in Figure 3 regarding species-specific peptides (*vide infra*). This approach enabled the identification of up to 11 unreported amino acid substitutions (Table 2). By combining the data from trypsin and Asp-N digestion, sequence coverages of 100% were obtained for Dac g 1, Lol p 1, Phl p 1, and Poa p 1, while ~86% of the Ant o 1 amino acid sequence was covered.

All the data gathered in Table 2 indicate that despite a high amino acid sequence homology (~90%),¹ group 1 allergens exhibit a considerable molecular variability with only 7 tryptic peptides shared by all of them. These peptides are rather equally distributed over the amino acid sequence, which tends

to indicate the absence of a highly variable part of the protein primary structure. These structural differences raise the possibility that major allergens from different grass pollens may elicit distinct immune responses, underlining that some grass-positive sera will not be detected by *in vitro* diagnostics with the use of a single grass species.³⁹ Moreover, it is worth mentioning that, considering a given allergen, several isoforms with different amino acid variants coexist within natural material. Such wide natural variability has potential consequences for immunotherapy, indicating that the use of recombinant allergens may not accurately reflect natural exposure and sensitization conditions.

3. Identification of Species-Specific Peptides. In the case of pollen mixtures, additional species-specific information must also be obtained for authentication and potential quantification purposes. This task is rendered difficult for group 1 allergens due to their high degree of sequence homology. In the framework of this study, peptides were both selected according to their allergen specificity, and their detection sensitivity under ESIMS conditions and chromatographic properties (peptides eluting in the column dead volume were rejected). Because of its high specificity and robustness, trypsin was chosen as digestion agent. The original objective was to obtain two tryptic peptides per allergen, to reach an appropriate level of specificity. This goal was only partially reached, since only one specific peptide was found for Dac g 1. Nevertheless, its relatively high

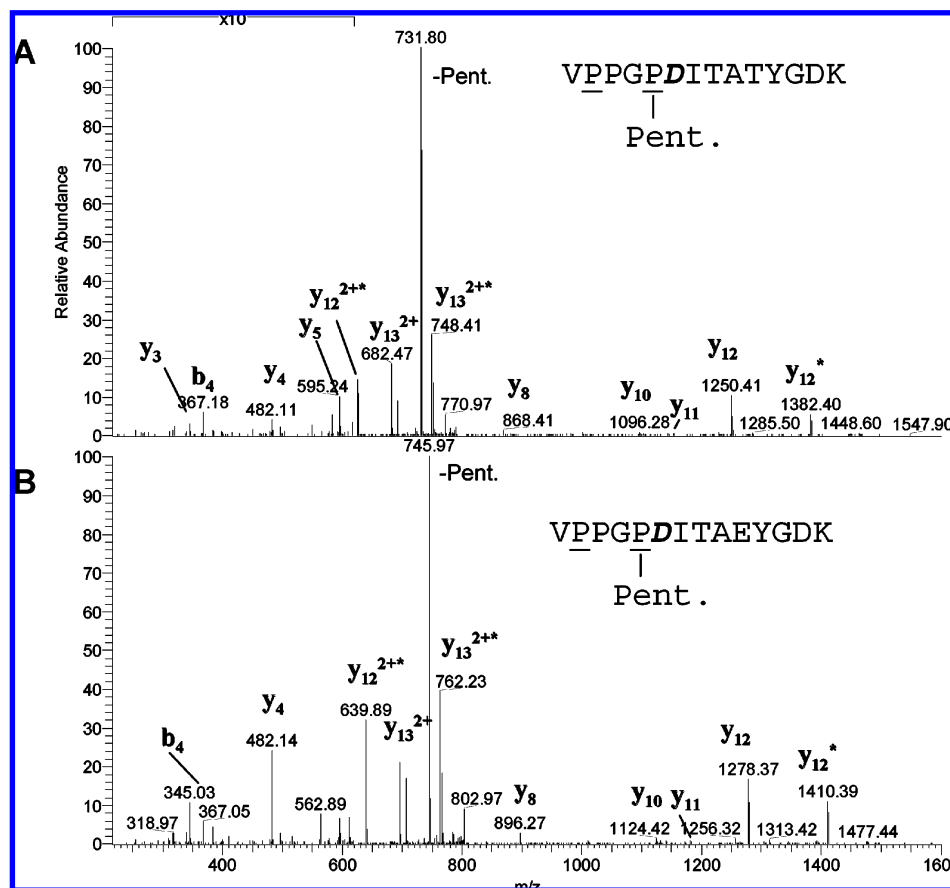


Figure 9. MS/MS spectra of (Val₄–Lys₁₇) glycopeptides from (A) Ant o 1, and (B) Lol p 1 bearing one pentose residue on Hyp₈ residue. Fragment ions bearing one pentose residue are labeled with an asterisk. (-Pent.) accounts for a loss of a pentose residue from the parent ion, while (P) represents a hydroxyproline residue. Because of the chemical lability of the O-glycosidic bond under MS/MS conditions, neutral losses of pentose residues can be observed.

molecular mass and sequence uniqueness may be regarded as providing enough specificity. Table 3 indicates the peptides that may be potentially used as identification tools for group 1 allergens, the “most sensitive peptides” being represented in bold characters. Figure 3 shows the corresponding MS/MS data. These chosen species-specific peptides may not be regarded as ideal according to classical MS-based protein quantification schemes, since traditionally peptides bearing known variants as well as reactive or labile amino acid residues (e.g., methionine or cysteine residues, Asp-Pro bonds) are avoided.⁴⁰ However, regarding homologous proteins, amino acid variants provide specificity and thus constitute valuable tools. Moreover, although much attention was paid to labile amino acids, it was not always possible to avoid their occurrence within the selected peptide sequences. In addition, it has been recently demonstrated that high-MS responding peptides often bear such labile residues or bonds.⁴¹

Peptide specificity has been further confirmed by analyzing three separate batches of purified allergens (data not shown). Although useful and efficient for specifically tracing the presence of a given allergen in pollen mixtures, the present set of peptides cannot be used without precautions for quantitative issues (*vide infra*). For instance, the most sensitive peptide specifically belonging to Ant o 1 was always observed in the digests in two forms, either fully cleaved or with one missed cleavage site (Table 3), the latter always being at least ~5-fold more intense. This is primarily due to the presence of an additional Lys residue in position 219 of Ant o 1, thus,

generating a -Lys–Lys- motif which is known to be responsible for the appearance of missed cleavage sites.⁴² Moreover, the resulting peptide, which has a Lys residue as N-terminal residue, is more resistant to further proteolysis by trypsin.⁴² A somewhat similar situation was also observed for the marker of Phl p 1, which has a -Lys–Pro- bond. The commonly accepted rule is that trypsin cleaves next to lysine and arginine but not before proline residues.⁴² However, recently published data indicated that this rule is often violated.⁴³ Under our conditions, it was observed that this specific bond was cleaved to a low extent, with peptide STWYGK found at trace levels with more than a ~10-fold lower intensity than the “parent peptide” STWYGKPTAAGPK (data not shown). These two peptides were identified at two distinct retention times, which exclude a possible artifactual generation by in-source fragmentation. It is worth noticing that using alternative digestion protocols did not modify these digestion patterns (data not shown). If a quantitative measurement could be envisioned, these drawbacks can be corrected, at least in part, through the use of an external calibration curve established with pure protein and/or labeled peptides incorporating a cleavage site as surrogates, to mimic cleavage efficiency.⁴⁴ The latter, of course, does not take into account the influence of protein folding on trypsin efficiency.

Identification and Characterization of Modified Peptides.

1. Peptides Incorporating Deamidation, Oxidation, or Glycation Sites. This part of the work was devoted to the preliminary investigation of the main degradation products of

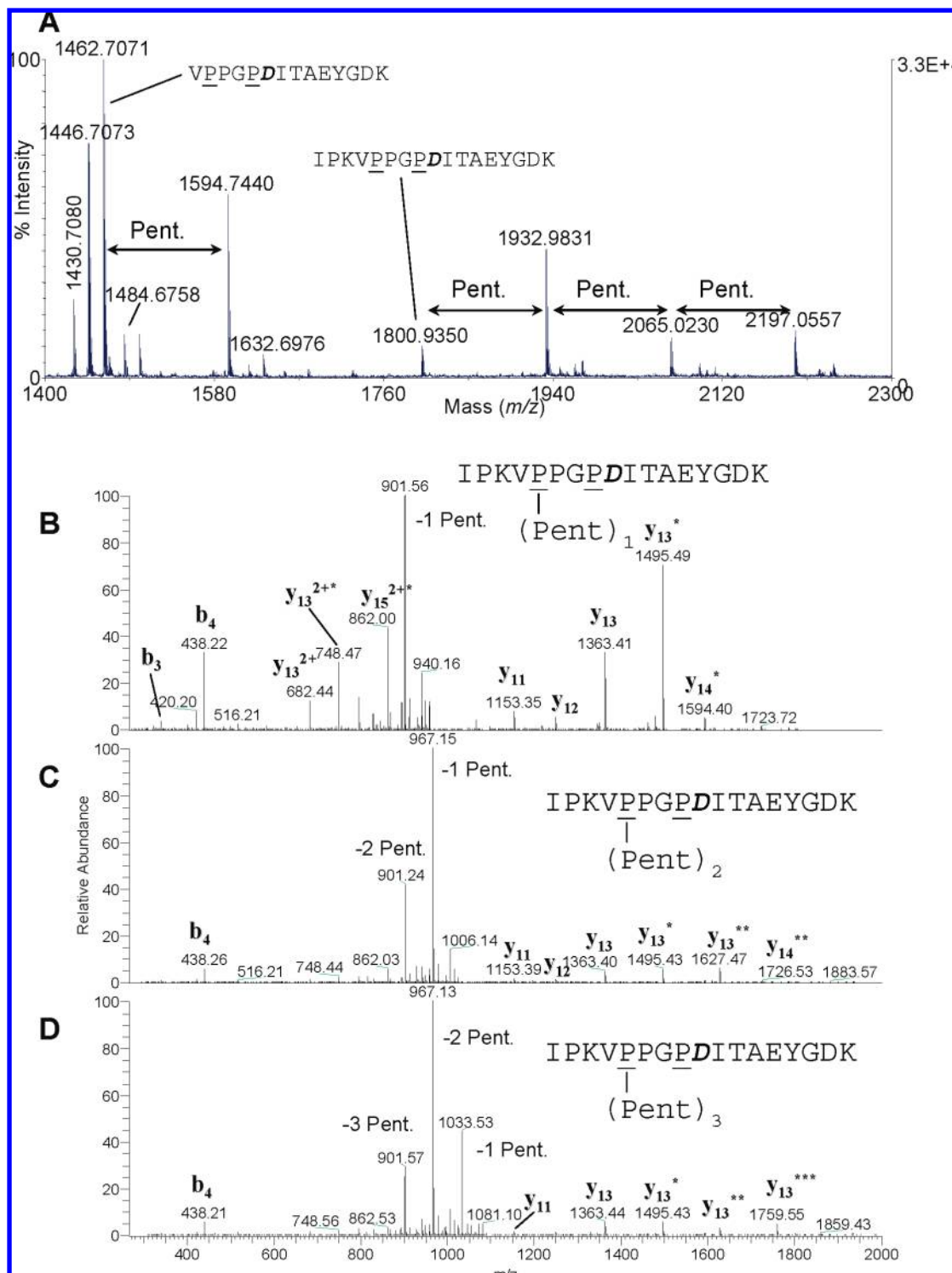


Figure 10. Demonstration of oligopentoseylation of Ant o 1 allergen, (A) MALDI-TOF mass spectrum of the HILIC-enriched and deglycosylated fraction, MS/MS spectra of (Ile₁-Lys₁₇) glycopeptide bearing two Hyp residues and (B) one (m/z 967.0), (C) two (m/z 1033.0), and (D) three pentose residues (m/z 1099.0). All fragmented ions are doubly charged ions. Fragment ions bearing n pentose residues are labeled with n asterisks. ($-n$ Pent.) accounts for a loss of n pentose residues from the parent ion, while (P) represents a hydroxyproline residue. Because of the chemical lability of the O -glycosidic bond under MS/MS conditions, neutral losses of pentose residues can be observed.

group 1 allergens, even though their process-induced and/or natural origin was not addressed and thus remains questionable. For instance, Asn₃₇ and Asn₅₂ were found to be deamidated, with the corresponding peptides mainly bearing an Asp residue (Table 2). MS/MS analysis confirmed the presence of

an Asp instead of an Asn residue at position 52 of the tryptic peptide (Ala₄₉-Lys₆₄), as may be seen in Figure 4A. It is not so surprising to find these deamidated residues since the corresponding peptides both have the -Asn-Gly- sequence, which is known to be especially prone to deamidation.⁴⁵ Moreover,

Characterization of Group 1 Grass Pollen Allergens

since deamidation introduces a new Asp-N cleavage site within the peptide backbone with the Asn to Asp conversion, Asp-N digestion was used to further confirm these results (data not shown). The precise origin of this modification remains to be clearly identified, since recent studies have shown that deamidation can readily occur during sample preparation.^{45,46}

Methionine residues are well-known to be preferred oxidation sites. Out of the two methionine residues present in group 1 allergen sequences, only Met₅₄ was oxidized. MS/MS data confirmed the oxidation of this residue, with the characteristic loss of 64 mass units corresponding to the loss of the CH₃SOH moiety,⁴⁷ for some of the major fragments bearing the oxidized methionine residue (Figure 4B).

Last, some peptides bearing one missed cleavage site and exhibiting strong consecutive water losses (up to 4, or 4 combined with the loss of a formaldehyde moiety) under MS/MS conditions were observed and further assigned as glycosylated peptides. Corresponding multiply charged ions were consistent with the presence of a hexose moiety. An example is given in Figure 4C for peptide (Gly₁₅₄–Lys₁₇₉). Such particular fragmentation behavior is characteristic of glycosylated peptides,^{48–50} with strong opposition to glycosylated peptides that exhibit preferential sugar losses.^{51,52} Thus, Lys₃, Lys₂₂, Lys₄₈, Lys₁₆₅, Lys₂₀₄, and Lys₂₃₂ were identified as putative glycation sites. This glycation site identification would greatly benefit from methodological improvements such as reduction of the Amadori product with sodium borohydride product prior to MS/MS acquisition,⁴⁹ or even from the employment of electron transfer dissociation instead of collision-induced dissociation.⁵⁰ Complementary experiments will also be performed in order to identify the origin of these glycosylated peptides, which can be “naturally” occurring or processed-induced modifications.

It should be pointed out that the tryptic peptide (Ala₄₉–Lys₆₄), or (Asp₄₅–Lys₆₄) considering one missed cleavage, bears deamidated, oxidized and glycosylated residues. Thus, this part of the sequence, which is presumably surface exposed, can be regarded as a useful indicator of product modification/degradation.

2. Peptides Incorporating Hydroxylation, and/or N-/O-Glycosylation Sites. Both hydroxylation and N-glycosylation sites are located on the same tryptic peptide, that is, peptide 4-VPPGPNITATYGDK-17, with the exception of Lol p 1 for which Thr₁₃ is replaced by a Glu residue (Table 2). Glycopeptides were located within the LC-MS/MS runs by monitoring low molecular weight carbohydrate fragment ions.⁵¹ Under our analytical conditions, due to the low mass cutoff of the ion trap, the oxonium ion at m/z 366.1 proved to be the most useful diagnostic ion.

Figure 5 shows the mass spectra obtained for the different glycopeptides composed of the five purified allergens. All detected ions were doubly charged, with the numerous ions observed reflecting the isoform heterogeneity. At first glance, one can note that roughly the same species exist in the different allergens, but with major relative intensity fluctuations. The ion at m/z 1235.54 (m/z 1249.54 for Lol p 1) is consistent with the peptide (Val₄–Lys₁₇) bearing two Hyp residues and the M0XF³ N-glycan motif. Isoform heterogeneity is further evidenced by the presence of other prominent ions at m/z 1227.55 and m/z 1301.57 corresponding to –16, and +132 Da mass differences when compared with M0XF³-bearing peptide. The former mass shift can either correspond to the substitution of a hexose by a fucose residue or, more probably, to the lack of one Hyp residue. The 132 Da mass shift is likely to account for

an additional pentose residue, but its linkage to either peptide backbone or to the N-glycan moiety remains questionable. These structural hypotheses were further investigated by MS/MS.

In most cases, MS/MS spectra of N-glycopeptides are dominated by B- and Y-type fragmentation of glycosidic linkages, thereby predominantly revealing information regarding the nature and sequence of the glycan moiety.^{51,53} In this particular case, MS/MS fragment ions were monitored in the LTQ part, since their detection in the Orbitrap proved to be less sensitive. The MS/MS spectrum of the ion at m/z 1235.55 (from Dac g 1, Figure 6A) confirms the occurrence of the M0XF³ motif as the major one among group 1 allergen glycoforms. Indeed, the oxonium ions at m/z 366.1, m/z 528.1 and m/z 660.1 account for (GlcNAc–Man⁺), (GlcNAc–Man₂⁺), and (GlcNAc–Man₂–Xyl⁺) structures, respectively, while other diagnostic ions corresponding to (peptide + H)⁺, (peptide + GlcNAc + H)⁺, and (peptide + GlcNAc + Fucose + H)⁺ can also be observed at m/z 1461.4, m/z 1667.5, and m/z 1810.5, respectively (Figure 6A). Moreover, the (peptide + H)⁺ ion observed at m/z 1461.4 is consistent with the peptide (Val₄–Lys₁₇) bearing two Hyp residues. The occurrence of a (Val₄–Lys₁₇) peptide bearing only one and no Hyp residue, as well as an O-linked pentose residue, can be further assumed due the presence of corresponding (peptide + H)⁺ ions at m/z 1445.5, m/z 1429.5, and m/z 1593.4, respectively (Figure 6B–D). When the same MS/MS interpretation scheme was used, low levels of glycoforms with plus or minus one hexose residue with respect to the M0XF³ structure (m/z 1317.1 and m/z 1155.02, respectively) were evidenced.

Hydroxylation and potential O-glycosylation of the peptide backbone were further investigated by performing another stage of fragmentation (MS³ experiments) with the (peptide + GlcNAc + H)⁺ ions as precursors.^{51,53} Although enabling peptide sequence confirmation and confirming the previous assumptions, the resulting MS³ spectra were not informative enough to precisely localize the modification sites (Figure S2). Thus, it was decided to pursue our investigations on deglycosylated peptides.

To improve sensitivity, tryptic glycopeptides from group 1 allergens were first specifically enriched by HILIC^{30,54} before being treated by N-glycosidase A to remove the N-glycans from the peptide moiety. Figure 7, panels A and B, shows the MALDI mass spectra of a HILIC-enriched tryptic digest of Dac g 1 before and after PNGase A digestion, respectively. While the glycopeptides were almost invisible in the crude digest, they appeared as dominant peaks after HILIC enrichment. It should also be noticed that the relative distribution of the different glycoforms is similar to what has been previously observed by LC-ESIMS (Figure 5B), thus, demonstrating that HILIC did not preferentially select one glycoform or the other. After treatment with N-glycosidase A, the corresponding MALDI mass spectrum demonstrates that the –16, and +132 Da mass differences still exist (Figure 7B). Corresponding isoforms bearing 0, 1, and 2 Hyp residues were clearly evidenced through the use of ESIMS/MS, while the presence of an O-linked pentose on the first Hyp residue was also demonstrated (i.e., Hyp₅, Figure 8A–D). Moreover, it is worth noticing that the isoform bearing only one Hyp almost exclusively lacks the first Hyp. This is in excellent agreement with the absence of a pentose residue on peptides bearing only one Hyp, and strongly supports the presence of the pentose moiety on the first Hyp residue. By analogy with published data, it can be assumed that this pentose is an arabinose (Ara) residue.^{15,18,55} Marginal O-

glycosylation of Hyp₈ was also observed on Ant o 1 and Lol p 1, but not on the other group 1 allergens (Figure 9).

Polyarabinylation of Hyp residues has been reported for Hyp-rich plant glycoproteins.^{18,55} Such modification was observed only at trace levels on peptide (Val₄–Lys₁₇), with up to two pentose residues on Ant o 1-derived peptide. However, after HILIC enrichment and further N-deglycosylation of Ant o 1 tryptic digest, the presence of up to three Ara residues on (Ile₁–Lys₁₇) peptide can readily be postulated according to the corresponding MALDI-TOF mass spectrum (Figure 10A). The presence of 1, 2, and 3 Ara on Hyp₅ was further confirmed by ESIMS/MS experiments performed on the corresponding doubly charged ions (Figure 10B–D). These particular O-glycosylated peptides were not clearly identified from the original LC-MS (prior to HILIC enrichment and N-deglycosylation), due to both the high complexity of their corresponding MS/MS spectra and their low detection yield. It should also be mentioned that the presence of more than one Ara residue on Hyp₅ tends to inhibit trypsin cleavage after Lys₃. Such oligoarabinylation was not clearly observed for the other group 1 allergens.

Altogether, these data show, for the first time, that great variations exist between the isoforms found in Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1. Indeed, both variations in proline hydroxylation as well as the nature of the main N-glycans, or the presence of multiple O-linked Ara residues were not reported previously for these group 1 allergens. This is of prime interest since knowledge of such post-translational modifications, which influence IgE binding, can help understanding diagnostic discrepancies and allergic disorders. For instance, within group 1, we have recently shown that out of the 14 potential MHC class II restricted T cell epitopes identified within group 1, only one is conserved among the 5 grass species (Chabre et al., submitted). Moreover, it was also shown that no single extract can totally inhibit IgE binding for several sera of patients, while a five grass mix always fully blocks IgE reactivity (Chabre et al., submitted).

Conclusion

The present study describes the in-depth mass spectrometric characterization of the five grass pollen group 1 allergens Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1. Both intra- and interspecies variations of amino acid sequences and post-translational modifications were highlighted. All the data from MALDI-TOF and LC-MS/MS analysis of digests indicated that, despite a high amino acid sequence homology (~90%), group 1 allergens exhibit considerable molecular variability with a small number of tryptic peptides shared by all of them. In the case of pollen mixtures, additional species-specific information must also be obtained for authentication and possibly quantification purposes. Thus, several peptides that may be potentially used as identification tools for group 1 allergens were highlighted, and their specificity further was confirmed by analyzing several batches of purified allergens (3 batches of each species collected in 2 different years and 2 geographic areas).

The second part of this work was devoted to the characterization of the post-translational modifications of the five group 1 allergens. Both hydroxylated proline and glycosylated residues were found on the N-terminal part of the amino acid sequence. HILIC enrichment proved to be of great benefit in studying this particular peptide. Thus, variations in proline hydroxylation N-glycosylation motifs between the different group 1 allergens

were found, and also the presence of O-glycosylation was demonstrated on one hydroxyproline residue. To the best of our knowledge, this is the first study reporting such a high level of structural information on pollen group 1 allergens.

Such structural information of these group 1 allergens is of great interest for a better understanding of allergic disorders and improvement of diagnostic methodologies or immunotherapies. It also suggests that patients are concomitantly exposed to multiple allergen isoforms bearing shared but also distinct structural determinants some of which are species-specific. As a consequence, diagnosis and immunotherapy should be performed with extracts made from mixtures of pollens obtained from several grass species, in order to recapitulate natural exposure and sensitization conditions.

Supporting Information Available: Figure S1, base peak LC-MS chromatograms obtained for group 1 allergen tryptic digests Ant o 1, Dac g 1, Lol p 1, Phl p 1, and Poa p 1; Figure S2, MS³ spectra of (Val₄–Lys₁₇) glycopeptide bearing the M0XF³ motif and two Hyp residues, one Hyp residue (i.e., Hyp₈), no Hyp residue, and two Hyp and one O-linked pentose residues (on Hyp₅). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Andersson, K.; Lidholm, J. Characteristics and immunobiology of grass pollen allergens. *Int. Arch. Allergy Immunol.* **2003**, *130* (2), 87–107.
- (2) Mohapatra, S. S.; Lockey, R. F.; Shirley, S. Immunobiology of grass pollen allergens. *Curr. Allergy Asthma Rep.* **2005**, *5* (5), 381–387.
- (3) Altmann, F. The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* **2007**, *142* (2), 99–115.
- (4) Chardin, H.; Senechal, H.; Wal, J. M.; Desvaux, F. X.; Godfrin, D.; Peltre, G. Characterization of peptidic and carbohydrate cross-reactive determinants in pollen polysensitization. *Clin. Exp. Allergy* **2008**, *38* (4), 680–685.
- (5) Moingeon, P.; Hrabina, M.; Bergmann, K. C.; Jaeger, S.; Frati, F.; Bordas, V.; Peltre, G. Specific immunotherapy for common grass pollen allergies: pertinence of a five grass pollen vaccine. *Int. Arch. Allergy Immunol.* **2008**, *146* (4), 338–342.
- (6) van Oort, E.; de Heer, P. G.; Dieker, M.; van Leeuwen, A. W.; Aalberse, R. C.; van Ree, R. Characterization of natural Dac g 1 variants: an alternative to recombinant group 1 allergens. *J. Allergy Clin. Immunol.* **2004**, *114* (5), 1124–1130.
- (7) Petersen, A.; Grobe, K.; Lindner, B.; Schlaak, M.; Becker, W. M. Comparison of natural and recombinant isoforms of grass pollen allergens. *Electrophoresis* **1997**, *18* (5), 819–825.
- (8) Petersen, A.; Becker, W. M.; Schlaak, M. Characterization of grass group I allergens in timothy grass pollen. *J. Allergy Clin. Immunol.* **1993**, *92* (6), 789–796.
- (9) Griffith, I. J.; Smith, P. M.; Pollock, J.; Theerakulpisut, P.; Avcioglu, A.; Davies, S.; Hough, T.; Singh, M. B.; Simpson, R. J.; Ward, L. D. Cloning and sequencing of Lol pl, the major allergenic protein of rye-grass pollen. *FEBS Lett.* **1991**, *279* (2), 210–215.
- (10) Petersen, A.; Schramm, G.; Bufer, A.; Schlaak, M.; Becker, W. M. Structural investigations of the major allergen Phl p I on the complementary DNA and protein level. *J. Allergy Clin. Immunol.* **1995**, *95* (5), 987–994.
- (11) Corti, V.; Cattaneo, A.; Bachi, A.; Rossi, R. E.; Monasterolo, G.; Paolucci, C.; Burastero, S. E.; Alessio, M. Identification of grass pollen allergens by two-dimensional gel electrophoresis and serological screening. *Proteomics* **2005**, *5* (3), 729–736.
- (12) Iraneta, S. G.; Acosta, D. M.; Duran, R.; Apicella, C.; Orlando, U. D.; Seoane, M. A.; Alonso, A.; Duschak, V. G. MALDI-TOF MS analysis of labile Lolium perenne major allergens in mixes. *Clin. Exp. Allergy* **2008**, *38* (8), 1391–1399.
- (13) van Oort, E.; de Heer, P. G.; Dieker, M.; van Leeuwen, A. W.; Aalberse, R. C.; van Ree, R. Characterization of natural Dac g 1 variants: an alternative to recombinant group 1 allergens. *J. Allergy Clin. Immunol.* **2004**, *114* (5), 1124–1130.
- (14) van Oort, E.; de Heer, P. G.; Dieker, M.; van Leeuwen, A. W.; Aalberse, R. C.; van Ree, R. Characterization of natural Dac g 1 variants: an alternative to recombinant group 1 allergens. *J. Allergy Clin. Immunol.* **2004**, *114* (5), 1124–1130.

- (15) Wicklein, D.; Lindner, B.; Moll, H.; Kolarich, D.; Altmann, F.; Becker, W. M.; Petersen, A. Carbohydrate moieties can induce mediator release: a detailed characterization of two major timothy grass pollen allergens. *Biol. Chem.* **2004**, *385* (5), 397–407.
- (16) Wilson, I. B.; Altmann, F. Structural analysis of N-glycans from allergenic grass, ragweed and tree pollens: core alpha1,3-linked fucose and xylose present in all pollens examined. *Glycoconjugate J.* **1998**, *15* (11), 1055–1070.
- (17) van Ree, R.; Cabanes-Macheteau, M.; Akkerdaas, J.; Milazzo, J. P.; Loutelier-Bourhis, C.; Rayon, C.; Villalba, M.; Koppelman, S.; Aalberse, R.; Rodriguez, R.; Faye, L.; Lerouge, P. Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J. Biol. Chem.* **2000**, *275* (15), 11451–11458.
- (18) Leonard, R.; Petersen, B. O.; Himly, M.; Kaar, W.; Wopfner, N.; Kolarich, D.; van, Ree, R.; Ebner, C.; Duus, J. O.; Ferreira, F.; Altmann, F. Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. *J. Biol. Chem.* **2005**, *280* (9), 7932–7940.
- (19) Petersen, A.; Schramm, G.; Schlaak, M.; Becker, W. M. Post-translational modifications influence IgE reactivity to the major allergen Phl p 1 of timothy grass pollen. *Clin. Exp. Allergy* **1998**, *28* (3), 315–321.
- (20) Esch, R. E. Allergen source materials and quality control of allergenic extracts. *Methods* **1997**, *13* (1), 2–13.
- (21) Hermeling, S.; Aranha, L.; Damen, J. M.; Slijper, M.; Schellekens, H.; Crommelin, D. J.; Jiskoot, W. Structural characterization and immunogenicity in wild-type and immune tolerant mice of degraded recombinant human interferon alpha2b. *Pharm. Res.* **2005**, *22* (12), 1997–2006.
- (22) Houde, D.; Kauppinen, P.; Mhatre, R.; Lyubarskaya, Y. Determination of protein oxidation by mass spectrometry and method transfer to quality control. *J. Chromatogr., A* **2006**, *1123* (2), 189–198.
- (23) Cohen, S. L.; Price, C.; Vlasak, J. Beta-elimination and peptide bond hydrolysis: two distinct mechanisms of human IgG1 hinge fragmentation upon storage. *J. Am. Chem. Soc.* **2007**, *129* (22), 6976–6977.
- (24) Fenaille, F.; Groseil, C.; Ramon, C.; Riande, S.; Siret, L.; Chtourou, S.; Bihoreau, N. Mass spectrometric characterization of N- and O-glycans of plasma-derived coagulation factor VII. *Glycoconjugate J.* **2008**, *25* (9), 827–842.
- (25) Gadgil, H. S.; Bondarenko, P. V.; Pipes, G.; Rehder, D.; McAuley, A.; Perico, N.; Dillon, T.; Ricci, M.; Treuheit, M. The LC/MS analysis of glycation of IgG molecules in sucrose containing formulations. *J. Pharm. Sci.* **2007**, *96* (10), 2607–2621.
- (26) Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212–217.
- (27) Srebalus Barnes, C. A.; Lim, A. Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrom. Rev.* **2007**, *26* (3), 370–388.
- (28) Helsper, J. P.; Gilissen, L. J.; van Ree, R.; America, A. H.; Cordewener, J. H.; Bosch, D. Quadrupole time-of-flight mass spectrometry: a method to study the actual expression of allergen isoforms identified by PCR cloning. *J. Allergy Clin. Immunol.* **2002**, *110* (1), 131–138.
- (29) Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S.; Gebler, J. C. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003**, *75* (21), 6023–6028.
- (30) Thaysen-Andersen, M.; Thøgersen, I. B.; Nielsen, H. J.; Lademann, U.; Brunner, N.; Enghild, J. J.; Hojrup, P. Rapid and individual-specific glycoproteomics of the low abundance N-glycosylated protein tissue inhibitor of metalloproteinases-1. *Mol. Cell. Proteomics* **2007**, *6* (4), 638–647.
- (31) Duriez, E.; Fenaille, F.; Tabet, J. C.; Lamourette, P.; Hilaire, D.; Becher, F.; Ezan, E. Detection of ricin in complex samples by immunocapture and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Proteome Res.* **2008**, *7* (9), 4154–4163.
- (32) Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.
- (33) Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* **2007**, *6* (1), 114–123.
- (34) van Oort, E.; de Heer, P. G.; Dieker, M.; van Leeuwen, A. W.; Aalberse, R. C.; van Ree, R. Characterization of natural Dac g 1 variants: an alternative to recombinant group 1 allergens. *J. Allergy Clin. Immunol.* **2004**, *114* (5), 1124–1130.
- (35) Carrera, M.; Canas, B.; Pineiro, C.; Vazquez, J.; Gallardo, J. M. De novo mass spectrometry sequencing and characterization of species-specific peptides from nucleoside diphosphate kinase B for the classification of commercial fish species belonging to the family Merlucciidae. *J. Proteome Res.* **2007**, *6* (8), 3070–3080.
- (36) Selles-Marchart, S.; Luque, I.; Casado-Vela, J.; Martinez-Esteso, M. J.; Bru-Martinez, R. Proteomics of multigenic families from species underrepresented in databases: the case of loquat (*Eriobotrya japonica* Lindl.) polyphenol oxidases. *J. Proteome Res.* **2008**, *7* (9), 4095–4106.
- (37) Zubarev, R.; Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* **2007**, *6* (3), 377–381.
- (38) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (47), 18132–18138.
- (39) van Ree, R.; van Leeuwen, W. A.; Aalberse, R. C. How far can we simplify in vitro diagnostics for grass pollen allergy? A study with 17 whole pollen extracts and purified natural and recombinant major allergens. *J. Allergy Clin. Immunol.* **1998**, *102* (2), 184–190.
- (40) Pan, S.; Aebersold, R.; Chen, R.; Rush, J.; Goodlett, D. R.; McIntosh, M. W.; Zhang, J.; Brentnall, T. A. Mass spectrometry based targeted protein quantification: Methods and applications. *J. Proteome Res.* **2009**, *8* (2), 787–797.
- (41) Fusaro, V. A.; Mani, D. R.; Mesirov, J. P.; Carr, S. A. Prediction of high-resolving peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.* **2009**, *27* (2), 191–198.
- (42) Yen, C. Y.; Russell, S.; Mendoza, A. M.; Meyer-Arendt, K.; Sun, S.; Cios, K. J.; Ahn, N. G.; Resing, K. A. Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra. *Anal. Chem.* **2006**, *78* (4), 1071–1084.
- (43) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A. Does trypsin cut before proline? *J. Proteome Res.* **2008**, *7* (1), 300–305.
- (44) Barnidge, D. R.; Hall, G. D.; Stocker, J. L.; Muddiman, D. C. Evaluation of a cleavable stable isotope labeled synthetic peptide for absolute protein quantification using LC-MS/MS. *J. Proteome Res.* **2004**, *3* (3), 658–661.
- (45) Krokhn, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal. Chem.* **2006**, *78* (18), 6645–6650.
- (46) Giza-Bulsec, G.; Li, B.; Bulsec, A.; Liu, H. Method to differentiate Asn deamidation that occurred prior to and during sample preparation of a monoclonal antibody. *Anal. Chem.* **2008**, *80* (24), 9491–9498.
- (47) Swiderek, K. M.; Davis, M. T.; Lee, T. D. The identification of peptide modifications derived from gel-separated proteins using electrospray triple quadrupole and ion trap analyses. *Electrophoresis* **1998**, *19* (6), 989–997.
- (48) Lapolla, D.; Fedele, D.; Reitano, R.; Arico, N. C.; Seraglia, R.; Traldi, P.; Marotta, E.; Tonani, R. Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (4), 496–509.
- (49) Fenaille, F.; Morgan, F.; Parisot, V.; Tabet, J. C.; Guy, P. A. Solid-state glycation of beta-lactoglobulin by lactose and galactose: localization of the modified amino acids using mass spectrometric techniques. *J. Mass Spectrom.* **2004**, *39* (1), 16–28.
- (50) Zhang, Q.; Ames, J. M.; Smith, R. D.; Baynes, J. W.; Metz, T. O. A Perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009**, *8* (2), 754–769.
- (51) Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr., B* **2007**, *849* (1–2), 115–128.
- (52) Macek, B.; Hofsteenge, J.; Peter-Katalinic, J. Direct determination of glycosylation sites in O-fucosylated glycopeptides using nano-electrospray quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15* (10), 771–777.
- (53) Dalpathado, D. S.; Desaire, H. Glycopeptide analysis by mass spectrometry. *Analyst* **2008**, *133* (6), 731–738.
- (54) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **2004**, *3* (3), 556–566.
- (55) Kieliszewski, M. J.; O'Neill, M.; Leykam, J.; Orlando, R. Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline O-arabinylation. *J. Biol. Chem.* **1995**, *270* (6), 2541–2549.

PR900359P

SUPPORTING INFORMATION

Figure S1. Base peak LC-MS chromatograms obtained for group 1 allergen tryptic digests (A) Ant o 1, (B) Dac g 1, (C) Lol p 1, (D) Phl p 1, and (E) Poa p 1.

Figure S2. MS³ spectra of (Val₄-Lys₁₇) glycopeptide bearing the M0XF³ motif and (A) two Hyp residues, (B) one Hyp residue (*i.e.*, Hyp₈), (C) no Hyp residue, and (D) two Hyp and one *O*-linked pentose residues (on Hyp₅). All fragmented ions correspond to (peptide + GlcNAc)⁺ species. Spectra were obtained from Dac g 1 tryptic digest.

Figure S1.

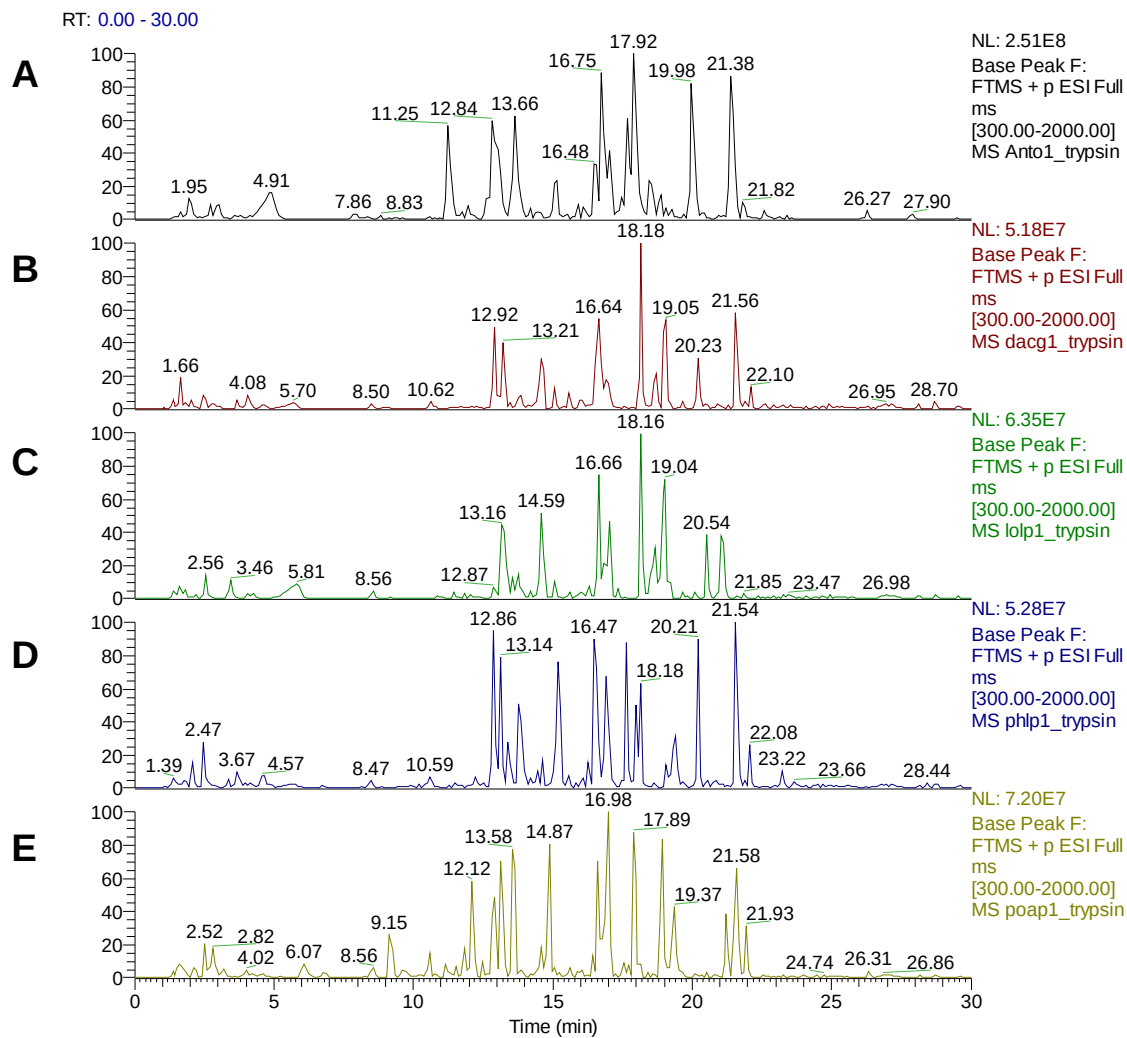
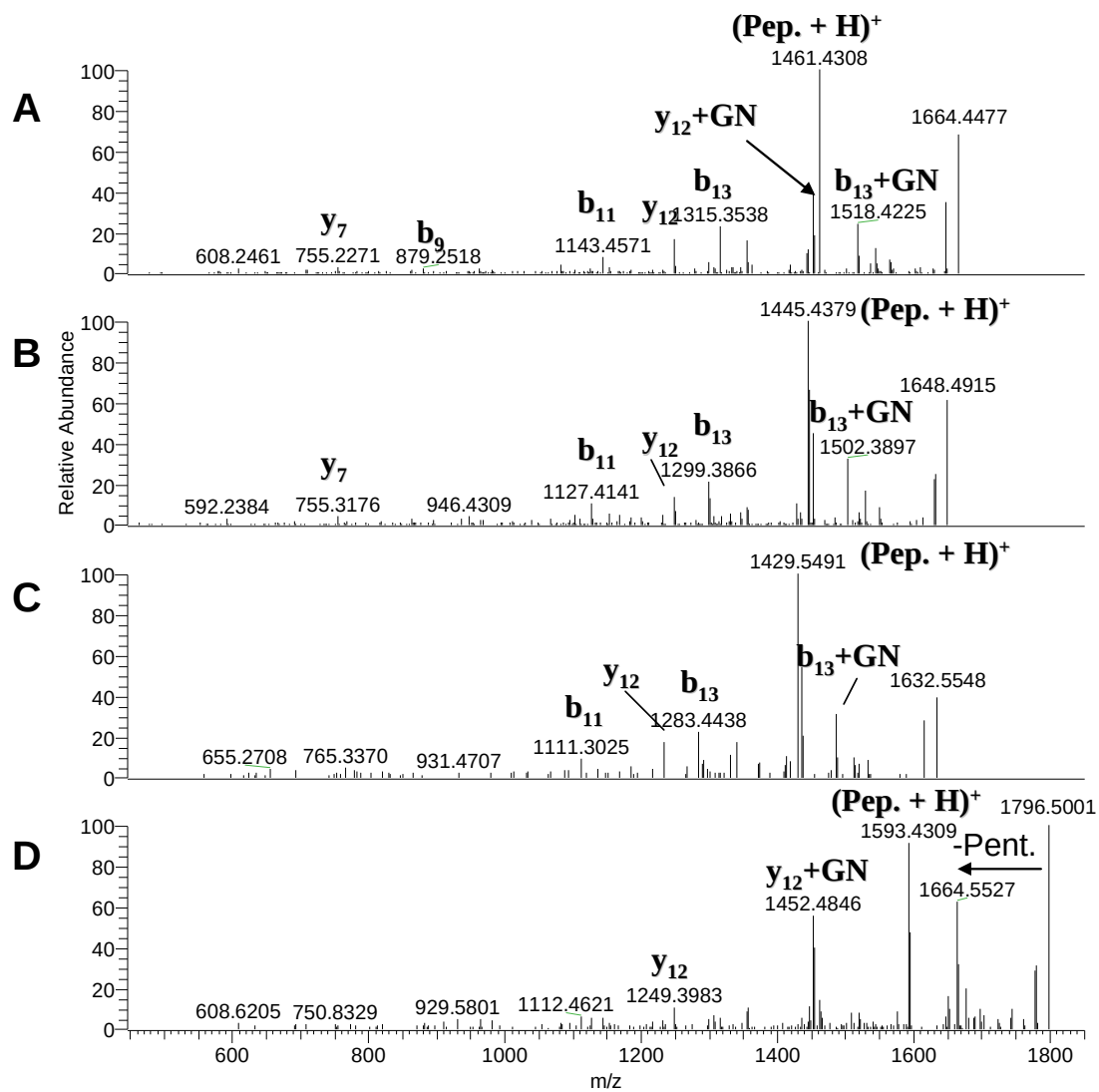
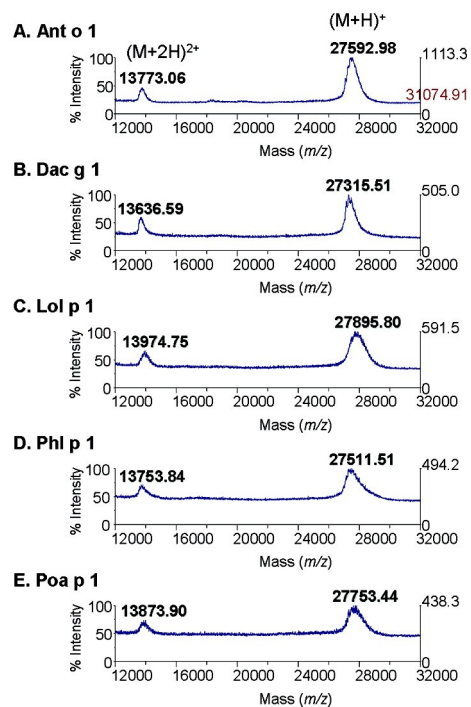
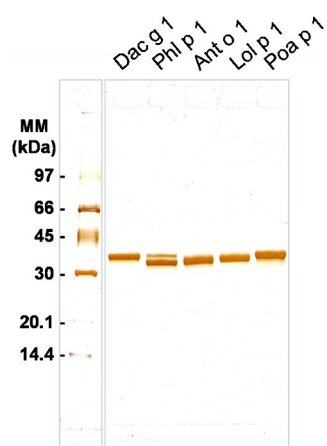
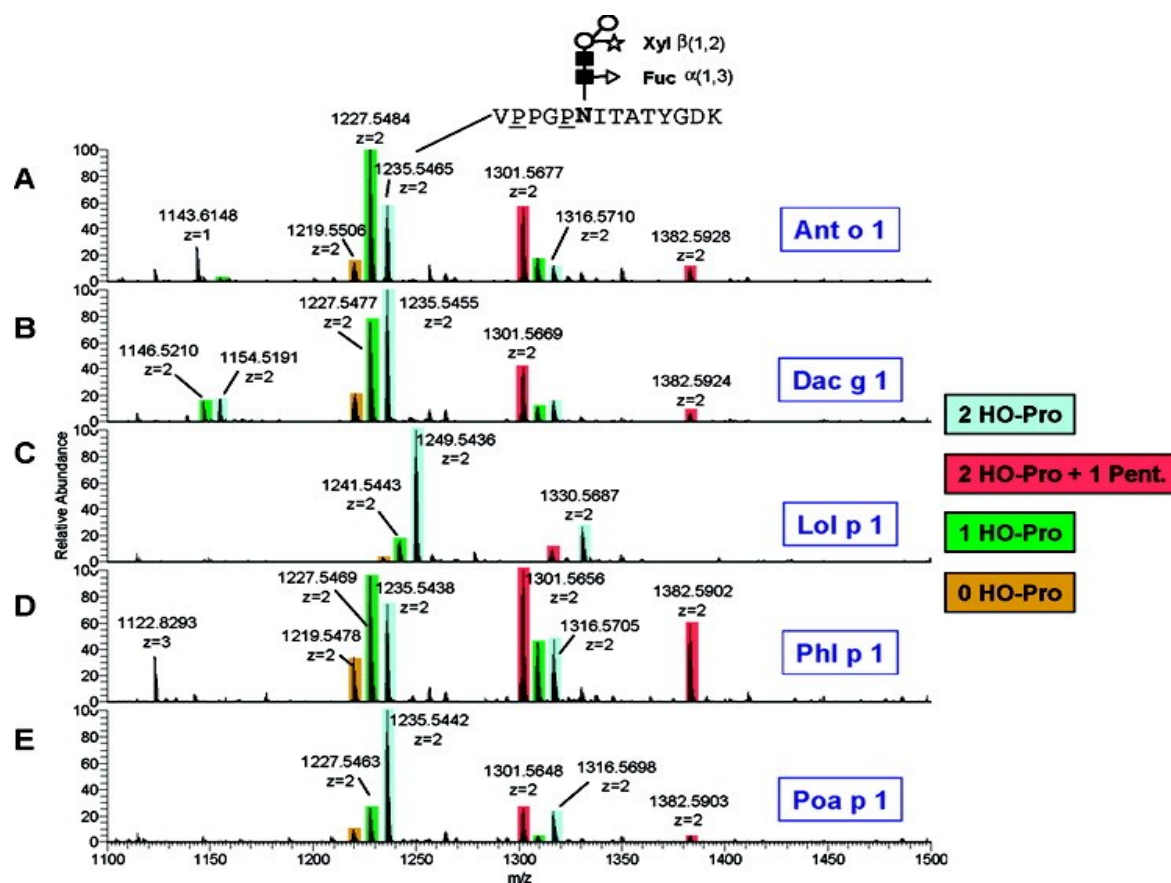
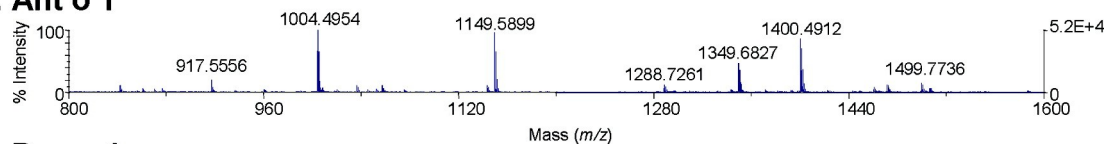


Figure S2.

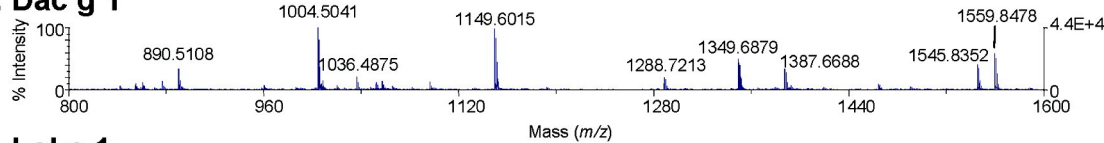




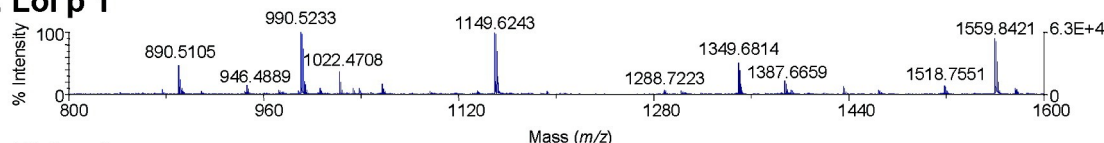
A. Ant o 1



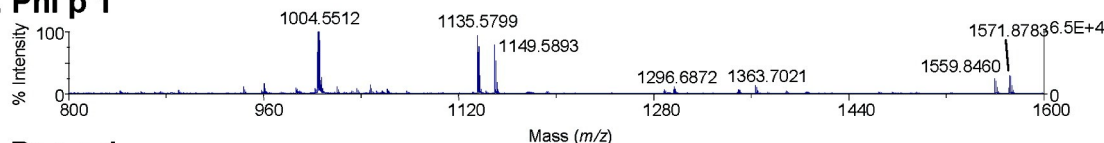
B. Dac g 1



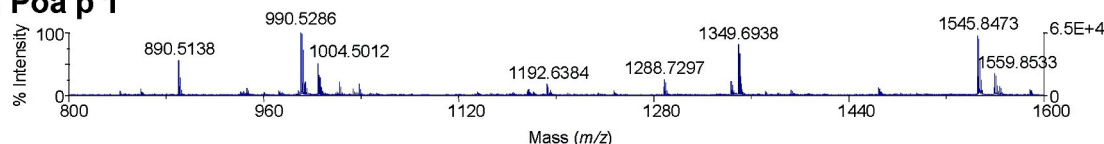
C. Lol p 1



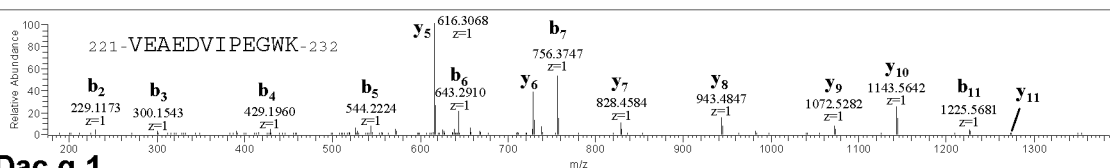
D. Phl p 1



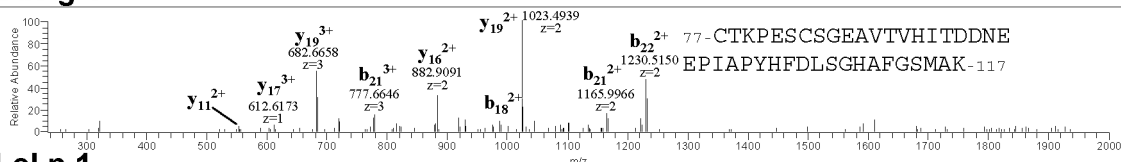
E. Poa p 1



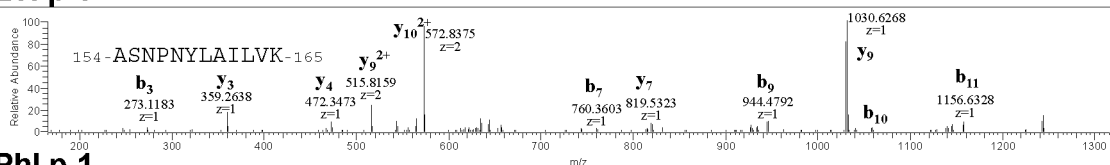
A. Ant o 1



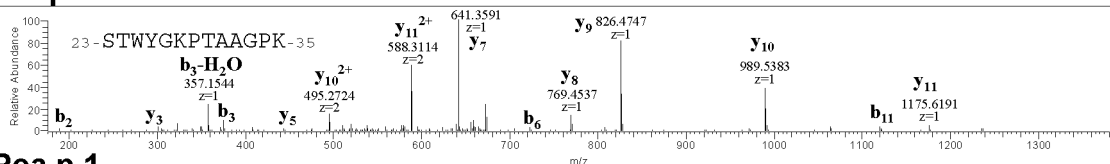
B. Dac g 1



C. Lol p 1



D. Phl p 1



E. Poa p 1

