

Model for Quality Control of Allergen Products with Mass Spectrometry

Jelena Spiric,[†] Thomas Schulenburg,[†] Luisa Schwaben,[†] Anna M. Engin,[†] Michael Karas,[‡] and Andreas Reuter^{*,†}

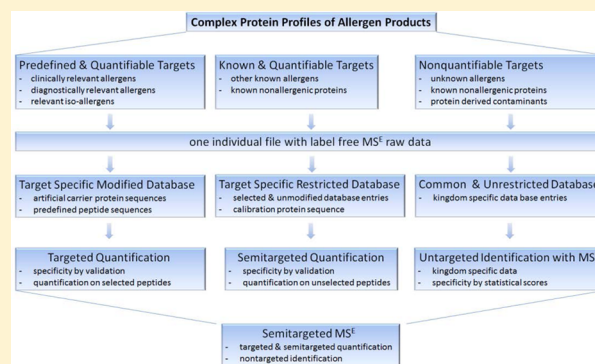
[†]Division of Allergology, Paul-Ehrlich-Institut, Langen, Germany

[‡]Institute of Pharmaceutical Chemistry, University of Frankfurt, Frankfurt, Germany

S Supporting Information

ABSTRACT: Birch pollen allergy is diagnosed and treated with aqueous extracts from birch pollen, which contain a mixture of allergens and nonallergenic proteins, including large numbers of closely related sequence variants, so-called iso-allergens of the major allergen, Bet v 1. The quality of therapeutic and diagnostic allergen products largely depends on the allergen and iso-allergen composition. Several biochemical methods are currently applied to detect and quantify allergens and to record protein profiles without differentiating between iso-allergens. Mass spectrometry (MS) may entirely replace these technologies, as it allows sequence specific identification and quantification of proteins and protein profiles including sequence variants in one run. However, the protein inference problem still hampers the automatic assignment of peptide sequences to proteins, consequently impeding the quantification of sequence variants. Therefore, the aim of the study was to set up semitargeted analyses of label-free MS data that allow unambiguous identification and quantification of birch pollen allergens and nonallergenic proteins. We combined data independent acquisition with manual assignment of predefined target sequences for quantification of iso-allergens and automatic quantification of other allergens and nonallergenic proteins. The quantitative data for birch pollen allergens and sequence variants of Bet v 1 were further confirmed by multiple reaction monitoring.

KEYWORDS: birch pollen, allergen quantification, untargeted, semitargeted, targeted mass spectrometry



INTRODUCTION

Protein derived allergen products are currently being characterized with a combination of methods. Each method targets an individual parameter that is relevant for the quality of the product. These parameters include sterility, impurities, and chemical parameters, such as pH and/or content of phenols, aldehydes, or adjuvants. The most important parameters are related to the protein fraction, as the active substances— allergens—are proteins. Governmental batch testing and quality control of a product by the pharmaceutical companies during manufacture and licensing procedures include the analysis of (1) the protein content; (2) the protein profile; (3) the identity of selected allergens; (4) and the IgE binding potency; and (5) it may also include the quantification of individual allergens. Allergen specific IgE antibodies trigger type I allergic reactions¹ and are utilized to measure the biological activity, i.e. the potency of allergen products. The IgE binding potency correlates with the allergen content, but it requires allergens to be biologically active i.e. correctly folded, chemically unmodified, and not degraded. Additionally, as serum IgE from allergic subjects is polyclonal, it does not allow a differentiation between allergens or closely related sequence

variants of an allergen. Consequently, IgE potency assay reports a quantitative value that is composed of all allergens recognized by the serum. Individual allergens are generally quantified with ELISA-like assays utilizing polyclonal sera or monoclonal antibodies raised against a purified allergen. The quantitative readout is therefore restricted to only one allergen with improved specificity compared to the IgE potency assay. Nevertheless, monoclonal antibodies cannot differentiate between closely related sequence variants, so-called iso-allergens or allergen variants.² Iso-allergens can be relevant because of differing diagnostic or therapeutic performance.³ Additionally, potency and ELISA-like assays are difficult to standardize, as they rely on biological reference material and calibrants, as well as on biological detectors, i.e. antibodies.

Therefore, it is desirable to develop a method that grants sequence specific qualitative and quantitative detection of all allergens and their sequence variants. The method should ideally rely on synthetic standards, and it should additionally include qualitative and quantitative information on the

Received: July 11, 2017

Published: August 18, 2017

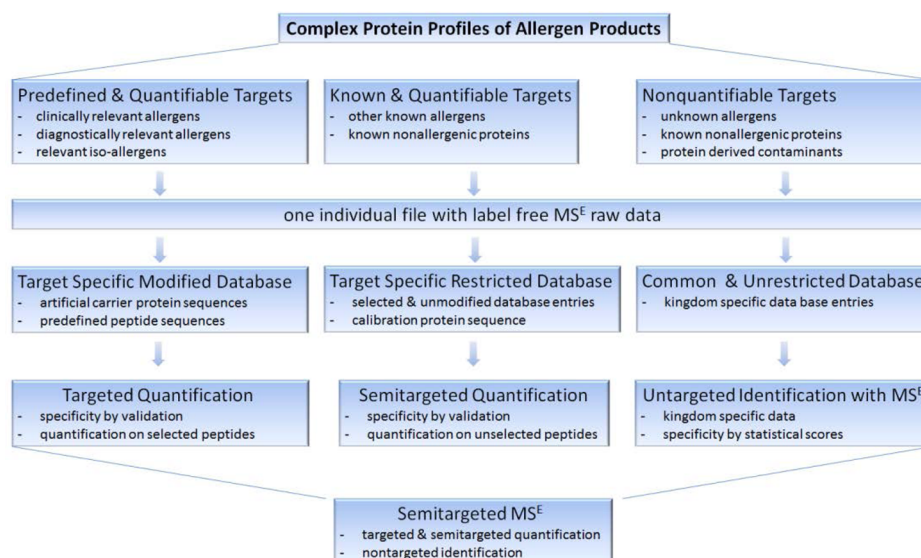


Figure 1. Semitargeted MS^E workflow.

nonallergenic protein profile. Such a method would be easier to standardize, and it would with one stroke replace several methods that are currently sequentially applied. Based on synthetic standards, several targeted and nontargeted LC-MS technologies would in part provide the desired information. Multiple reaction monitoring (MRM) would provide sequence specific data and absolute quantitative values, but MRM is restricted to predefined targets. High resolution MS with data dependent acquisition (DDA) would provide qualitative and quantitative data, but a large part of the protein profile would be missed. Data independent acquisition (DIA) modes such as SWATH, MS^E, HDMS^E, and All Ions MS/MS would provide quantitative or at least semiquantitative values to a larger portion of the protein profile.^{4–8} Nevertheless, DIA technologies also include database searches for identification which are still hampered by the protein inference problem.^{9–13} In our previous study, we described a partially manual workflow to reliably identify closely related sequence variants with MS^E and reported a new peptide grouping scenario¹⁴ that was not reported by Nesvizhskii and Aebersold, who published a comprehensive analysis of protein inference phenomenon.⁹ The results of a fully automated database search do not reach a necessary degree of reliability and unambiguous identification of sequence variants, when biomedical products are investigated. We tested this for MS^E data using the Protein Lynx Global Server (PLGS) for data analysis, but we concluded from the literature that other database search engines are affected as well.^{9,11–13,15,16} The protein inference problem affects the protein quantification, as quantitative values might be assigned to incorrect protein entries. Hence, the application of automated database searches for analysis of DIA based MS data is rendered impossible, if the quality of allergen products is to be controlled, e.g. within marketing licensing or batch control. However, we found it very appealing to apply a DIA (MS^E) based MS workflow to record qualitative and quantitative protein profiles of allergen products in parallel. We aimed to establish a data analysis workflow for a semitargeted evaluation of the protein profile of birch pollen, as birch pollen is one of the most important allergen sources in Central Europe. The major allergen, Bet v 1, is composed of at least 18 different sequence variants that are expressed as

proteins at the same time in the same tissue.¹⁴ Hence, it is impossible to directly apply the concept published by Silva et al.¹⁷ The authors applied MS^E, a vendor specific DIA mode, and reported a universal dose response factor by averaging intensity values for the three most intense peptides of a protein. An indispensable prerequisite for this concept is that each of the three most intense peptides does truly originate from the same protein sequence. However, each of the 18 different Bet v 1 sequence variants would theoretically release between 19 and 21 different peptides after tryptic digestion, summing up to a total number of 349 peptides. This number is reduced to 76 if all identical peptides are collapsed into one group. And, there are only 31 peptides that are unique i.e. that exist in only one individual protein (Figure S-1 and Table S-1, [Supporting Information](#)). The only two tryptic peptides (T19 and T121) that are 100% conserved were never detected in our previous study. All other peptides comprise between 2 and 9 different sequence variants. Different combinations of these peptide variants form 18 unique Bet v 1 sequence variants. It can easily be interpreted from the Bet v 1 sequence variant pattern (Figure S-1, [Supporting Information](#)) that the peptides T34 (blue variant) and T147 (red variant) would very likely be among the top three most intense peptides for any of the proteins displayed; nevertheless, these peptides could originate from 16 or 17 different proteins, respectively. Bet v 1.0301 is the only sequence variant that does not contain T34 (blue variant) and T147 (red variant) peptides; therefore, it would not be affected by this phenomenon. Recording quantitative protein profiles, which include nonallergenic proteins, to demonstrate batch to batch consistency is further complicated by the unambiguity of MS data. For example, as SDS-PAGE is a common tool used to demonstrate batch to batch consistency, a reduced intensity of an individual band of the nonallergenic part of the protein profiles might be accepted as a minor deviation. This decision might be made unaware whether the deviation is caused by altered quantity or the entire loss of an individual protein. MS would unambiguously report either the different quantity or the absence of a previously detected protein. The first case—the altered quantity—is uncritical, because limits for acceptance or rejection can be defined. The limits might well differ from the values set for diagnostically or

clinically relevant allergens and could be fairly broad. However, in the second case, it would be difficult to state that two batches are consistent if one or more nonallergenic proteins are entirely missing. Unfortunately, a low protein score during database search would lead to a quantity of zero even if the protein is not entirely absent. Consequently, the quantification of the nonallergenic part of the protein profile cannot entirely be untargated. The specific fraction of the entire nonallergenic protein profile should be defined whose protein scores should not be at risk of dropping below a set threshold, which could turn even a slight reduction in protein content into a negative result. However, irrevocably excluding all proteins except a predefined set of allergens and nonallergens, i.e. setting up an entirely targeted assay, would deprive MS^E of its greatest strength—the detection of the unexpected, i.e. contaminants and still unknown sequence variants.

Thus, we aimed to establish a workflow (Figures 1 and 2) that would allow (1) a reliable and unambiguous targeted, label-

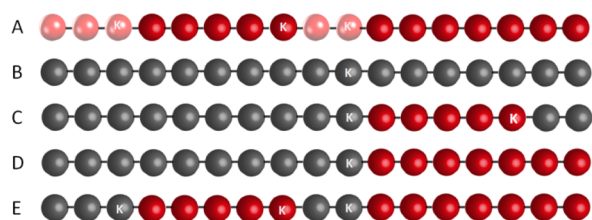


Figure 2. Creating artificial database entries for the targeted quantification of a predefined protein.

free identification and quantification of all birch pollen allergens and selected iso-allergens. (The targeting would specifically eliminate protein inference issues in identification and quantification of Bet v 1 sequence variants.); (2) a targeted quantification of the total Bet v 1 content including all Bet v 1 sequence variants; (3) a semitargeted and label-free quantification of predefined nonallergenic marker proteins (The nonallergenic part of the protein profile should be comprehensive enough to demonstrate batch to batch consistency.); (4) the untargated and label-free identification of as many as possible nonallergenic birch pollen proteins or nonbirch pollen derived contaminants. In order to confirm our workflow and MS^E data, we also report MRM derived values for quantification of selected birch pollen allergens and sequence variants of Bet v 1.

EXPERIMENTAL PROCEDURES

Study design

The study design to quantify proteins included three extraction replicates of birch pollen and three replicates of digestion, and each digestion was injected three times into the LC-MS system. In total, we recorded 27 data sets. To ensure assay's reproducibility, we also set the following acceptance criteria for protein quantification: (1) a protein had to be detected at least in two out of three technical replicates; (2) it had to be detected at least in two out of three digestion replicates; (3) criteria (1) and (2) had to be fulfilled in all three extraction replicates. For protein identification, only those proteins that were detected in at least two extraction replicates, one digestion replicate, and two out of three technical replicates were reported. All sequence variants and nonallergenic proteins that were at the detection limit were eliminated from data analyses

by these criteria, thus minimizing a risk of false negative results because of an insignificant protein score, although releasing peptides with intensities above the noise level. Protein content was calculated by converting the mole (fmol) amount reported by PLGS to weight (ng) using an online tool "Practical Molecular Biology".¹⁸ Standard deviation (s) were calculated using the following equation:

$$s = \sqrt{(\sum [(X_i - \bar{X})^2] / (n - 1))}$$

The coefficient of variation (CV) was calculated as the ratio of standard deviation (s) to average value (μ) ($n = 27$), and given in percent $CV = s / (\mu) * 100$.

Sample preparation

Proteins from birch pollen allergen product source material (*Betula Pendula*; Allergon AB, Ängelholm, Sweden) were extracted and stored as previously described.¹⁴

In-solution protein digestion

Twenty micrograms of the total soluble protein were adjusted to a total volume of 50 μ L in 50 mM NH_4HCO_3 . The samples were reduced with 20.8 μ L of 10 mM DTT in 50 mM NH_4HCO_3 for 15 min at 60 °C in a thermal cycler (Bibby Scientific, Staffordshire, UK) and alkylated with 20.8 μ L of 55 mM iodoacetamide in 50 mM NH_4HCO_3 for 15 min at ambient temperature in the dark. The samples were digested with trypsin at 1:10 protease to protein ratio (8.4 μ L of 0.238 μ g/ μ L in 25 mM NH_4HCO_3) and incubated for 8 h at 37 °C in a thermal cycler. The reaction was stopped by the addition of 5% formic acid (11 μ L). The resulting tryptic digests were stored at -80 °C until further MS analyses.

LC-MS^E analysis

Each digested sample was spiked with predigested Enolase (MassPREP Enolase digestion standard, Waters, Manchester, UK) that was used as an internal standard at a level of 50 fmol per μ L. For a sample analysis, 180 ng of digested birch pollen protein aliquots was loaded on the column and analyzed using the same MS parameters and the system as described by Spiric et al.¹⁴

Data processing and database searches for quantification and identification

Several specific databases were created for: (1) identification of Bet v 1 sequence variants; (2) targeted quantification of total Bet v 1 and Bet v 1 sequence variants; (3) semitargeted quantification of Bet v 2, 4, 6, and 7 and selected nonallergenic birch pollen proteins; (4) identification of other nonallergenic proteins. However, several common standard parameters were used for raw data processing and database searches and are described below.

Standard data processing parameters

The raw data files were processed using ProteinLynx Global Server (PLGS) version 2.4 (Waters). The processing parameters included: (1) automatic determination of Chromatographic Peak Width and MS TOF Resolution; (2) Lock Mass for Charge 1 (684.3469 Da/e), Lock Mass for Charge 2 (785.8426 Da/e), and Lock Mass Window (0.25 Da); (3) Low Energy Threshold (250 counts) and Elevated Energy Threshold (100 counts); (4) Retention Time Window (Automatic); (5) Intensity Threshold (1 500 counts).

Standard database search parameters

The following parameters were applied for all database searches: (1) a maximum of one missed cleavage site; (2) a minimum of three fragment ion matches per peptide; (3) a minimum of seven fragment matches per protein; (4) a minimum of one peptide match per protein; (5) a set false positive rate of 4%. A fixed modification included carbamidomethylation of cysteine (C), and variable modifications were restricted to deamination of asparagine (N) and glutamine (Q) and oxidation of methionine (M). The calibration protein was specified as Enolase (UniProt Acc.No. P00924), and the calibration protein concentration was set at 100 fmol.

Database for identification of Bet v 1 sequence variants. The workflow described by Spiric et al.¹⁴ was followed for identification of Bet v 1 sequence variants prior to their quantification.

Database for the targeted quantification of total Bet v 1 and individual Bet v 1 sequence variants. For quantification of total Bet v 1 and Bet v 1 sequence variants, a database consisting of a Uniprot *Homo sapiens* database (taxonomy 9609, downloaded on 16.02.2012.) was established with the addition of nine artificially created entries, following the concept outlined in Figures 1 and 2. These artificial entries consisted of a carrier sequence and one or two Bet v 1 sequence variant specific target peptides (Figure 2, Figure S-1, and Table S-1, Supporting Information). The carrier sequence was bovine serum albumin (UniProt Acc. No. P02769) that is not present in birch pollen. The artificial database entries and their individual Bet v 1 sequence variant specific target peptides are shown in Table S-2 (Supporting Information). We established one entry for the quantification of total Bet v 1 content (Bet v 1_T; in-house Acc. No. JS001) and for the quantification of eight individual Bet v 1 variants (In-house Acc. No. JS002, JS003, JS004, JS005, JS006, JS007, JS008, and JS009) and Enolase (Acc.No. P00924).

Total Bet v 1 content was quantified using a combination of two preselected peptides which were previously identified.¹⁴ These two peptides are most commonly shared and named T34 blue and T1147 red (Figure S-1 and Table S-1, Supporting Information) for all identified Bet v 1 sequences, and they were added to the carrier sequence creating the artificial entry, Bet v 1_T. This entry forced the PLGS to detect and quantify 16 out of 18 Bet v 1 variants only using the averaged intensity values of these two peptides. The values obtained for Bet v 1 sequence variant Q39415 (Bet v 1.0301, in-house Acc. No. JS005) and Q0QLS9 (Bet v 1.0119, in-house Acc.No. JS009) which do not share the above-mentioned peptides were added to the value of Bet v 1_T to get the total value for Bet v 1. It was not necessary to use bovine serum albumin as a template sequence to quantify Q39415 as this Bet v 1 sequence variant does not share any detectable tryptic peptides with other sequence variants.

Individual variants of Bet v 1 (Bet v 1.0102, Bet v 1.0104, Bet v 1.0203, Bet v 1.0115, Bet v 1.0113, Bet v 1.0116, and Bet v 1.0119) were quantified using only one preselected precursor, which belongs to a sequence variant specific peptide. This peptide was added, in the same way as peptides for Bet v 1_T, to the carrier sequence, creating an artificial database entry that forced PLGS to use only one peptide for detection and quantification of these sequence specific variants. Bet v 1.0301 was quantified using its original database entry (Q39415) as it does not share any peptide sequence with other variants.

Database for semitargeted quantification of Bet v 2, 4, 6, and 7 and nonallergenic birch pollen proteins. For

the quantification Bet v 2, 4, 6, and 7 and of nonallergenic proteins, we used the Uniprot database restricted to *Homo sapiens* (taxonomy 9609, downloaded on 02.08.2012.) with the addition of 82 accession numbers (Table S-3, Supporting Information) that were unambiguously identified in our previous study by 2D-MS analysis of birch pollen protein profile¹⁴ and in-solution optimization MS experiments in the current study, following the concept outlined in Figures 1 and 2. The database was also amended with Enolase (Acc.No. P00924), which served as calibrant. Quantification of proteins was achieved by applying the quantification concept developed by Silva et al.^{5,19}

Database for untargeted protein identification. All processed MS^E files were searched against a publicly available UniProt derived fasta format reviewed database restricted to green plants (*Viridiplantae*, taxonomy 33090) downloaded on 30.03.2012., with the addition of the calibration standard sequence (Enolase, Acc. No.P00924) using the search algorithm within the PLGS 2.4.

LC-MRM parameters

Study design, sample preparation, and in solution digestion. The same study design as mentioned above was carried out for quantification of Bet v 1 sequence variants and birch pollen allergens using a triple quadrupole system (Xevo TQ-S; Waters, Manchester, UK). Differing from this: Three extraction replicates were performed; each extraction was digested three times with a trypsin (0.53 $\mu\text{g}/\mu\text{L}$) to total protein ratio of 1:2.5 (w/w), and each digest was analyzed three times. The background matrix for the calibration curves was prepared by pooling equal volumes of three extraction replicates of birch pollen. One calibration curve was generated for each extraction replicate. Each calibration point was obtained from three technical replicates. Extraction digests were carried out on three separate days.

LC parameters

The chromatographic separation of the samples was performed using a reversed-phase ACQUITY UPLC BEH column (C18, 130 Å, 1.7 μm 2.1 mm $\times$ 50 mm, Waters, Manchester, UK). The binary solvent system consisted of an aqueous mobile phase (solvent A = water with 0.1% (v/v) formic acid) and an organic mobile phase (solvent B = acetonitrile with 0.1% (v/v) formic acid). Separation was accomplished using 10 μL injection volume at the flow rate 0.3 mL/min. The LC run was started with an initial hold of 0.5 min at 97% A, followed by a linear gradient to 60% A within 11.5 min continued with an isocratic step at 90% B for 1.2 min with the flow rate 0.5 mL/min, and equilibrated for 2.3 min at 97% A. The column was maintained at 40 °C during analysis, and the samples were kept at 12 °C.

MRM parameters

Sample analyses were performed in MRM mode using positive-ion electrospray ionization, with a capillary voltage of 2.5 kV, a source temperature of 150 °C, and a desolvation temperature of 200 °C. Cone voltage and collision energy were determined for each peptide individually. Table S-4 (Supporting Information) summarizes the optimized MS parameters for 13 target peptide sequences with their respective precursor and product ion m/z values, cone voltages, and collision energies. The description of the MRM method development can also be found in the Supporting Information. Briefly, all peptides (between 10 and 22 amino acids) were recorded as doubly charged precursor by

Table 1. Targeted, Semitargeted, and Untargeted Quantification of Birch Pollen Protein Profile by MS^E

Acc.No.	IUIS name	Description	S	P	C	E	Amount [μg/mg]	%CV	% Birch pollen protein
Allergens									
		Bet v 1_T		2		2.5	275.8	10.5	27.6
P43177	Bet v 1.0102	Pathogenesis-related proteins (PR)-10		1		2.0	65.0	11.3	6.5
P43179	Bet v 1.0104	Pathogenesis-related proteins (PR)-10		1		2.4	24.5	13.5	2.5
Q39415	Bet v 1.0301	Pathogenesis-related proteins (PR)-10		1		4.6			
Q96370	Bet v 1.0113	Pathogenesis-related proteins (PR)-10		1		7.6			
O23748	Bet v 1.0116	Pathogenesis-related proteins (PR)-11		1		4.9			
A4K9Z8	Bet v 2.0101	Profilin	4770	8	51	3.6	28.3	11.0	2.6
Q39419	Bet v 4.0101	Polcalcin	2804	5	64	1.9	7.0	9.2	0.7
Q9FUW6	Bet v 6.0102	Allergenic isoflavone reductase	4936	18	73	4.3	60.0	16.2	6.0
Q8L5T1	Bet v 7.0101	Peptidyl prolyl cis trans isomerase	9995	13	72	4.7	79.7	8.2	8.0
Nonallergenic proteins									
A8HTK0		Thioredoxin OS <i>Glycine max</i>	599	3	15	7.7	6.6	23.8	0.7
C4MF37		UDP glycosyltransferase <i>A.strigosa</i>	290	2	11	4.0	15.1	20.0	1.5
D6BR66		Glutathione S transferase omega <i>J.curcas</i>	349	3	10	2.7	12.8	19.5	1.3
D7T2W1		Putative un. protein <i>Vitis vinifera</i>	245	2	18	5.1	4.4	12.5	0.4
D7TQB6		Putative un. protein <i>Vitis vinifera</i>	777	4	17	4.5	49.3	16.6	4.9
G7JV45		Putative un. protein <i>M.truncatula</i>	1288	1	15	3.5	1.2	7.9	0.1
O49813		Olee1 like protein <i>B.pendula</i>	264	7	25	5.4	3.3	25.9	0.3
P42739		Polyubiquitin <i>Acetabularia peniculus</i>	903	6	10	3.3	4.3	33.4	0.4
Q676X7		Serine threonine kinase <i>H.orientalis</i>	469	4	25	2.2	8.1	13.0	0.8
Q7X7E8		Peptidyl prolyl cis trans isom. <i>T.aestivum</i>	312	1	8	5.0	10.7	18.6	1.1
A9NMR0		40S ribosomal protein <i>P.sitchensis</i>	160	2	7	5.1			
B9RB02		Glutaredoxin 1 <i>grx1</i> putative <i>R.communis</i>	1608	3	9	1.5			
D2X9U2		λclass glutathione transf. <i>P.trichocarpa</i>	138	3	8	3.8			
D7NJ40		Isopen. diphosphate isom. <i>S.lycopersicum</i>	167	3	14	2.8			
G7K7W9		Cytochrome c <i>M.truncatula</i>	352	4	28	7.5			
Q07796		Superoxide dismutase Cu Zn <i>I.batatas</i>	122	2	20	3.8			
Q8S3L1		Glutaredoxin <i>P. tremula</i> x <i>P. tremuloides</i>	296	3	14	5.5			

^aAcc. No.: Accession number; S: PLGS protein score; P: Number of assigned peptides; C: Protein sequence coverage %; E: Precursor RMS mass error (ppm); %CV: Coefficient of variation. Protein amounts reported are averaged values obtained from three extraction replicates.

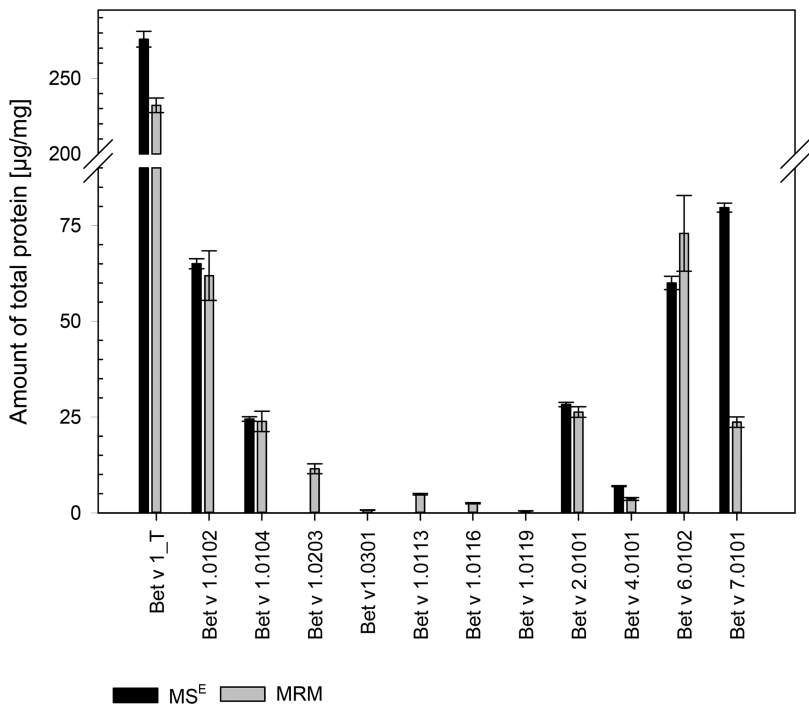


Figure 3. Quantification of allergenic proteins in birch pollen by MS^E and MRM.

using either y- or b-series fragments. For each peptide, the most intense transition was used for quantification and three more

transitions were recorded to ensure specificity. The MRM parameters, including amino acid sequence of the target peptide

and type and m/z value of each individual transition, are summarized in Table S-4. The data was analyzed with TargetLynx software version 4.1 (Waters, Manchester, UK).

RESULTS AND DISCUSSION

Targeted label-free quantification of Bet v 1 and Bet v 1 sequence variants

We quantified the total amount of Bet v 1, and the amounts of two sequence variants, Bet v 1.0102 and Bet v 1.0104. We did not assign quantitative values to the other three Bet v 1 sequence variants (Bet v 1.0301, Bet v 1.0113, and Bet v 1.0116) as they failed the acceptance criteria, even though they were identified in several samples. Total Bet v 1 content was determined on the averaged signals of only T34-blue and T147-red peptides. The quantitative value for Bet v 1.0301 was not added even though this iso-allergen is not represented by T34-blue and T147-red, as it failed the acceptance criteria for quantification, while Bet v 1.0119 was not detected at all. The quantitative readout of Bet v 1 sequence variants is summarized in Table 1. A concentration of 275 $\mu\text{g/mL}$ of the total protein extract was determined for total Bet v 1, representing app. 28% of the total protein content. The iso-allergens Bet v 1.0102 and Bet v 1.0104 represented 6.5 or 2.5%, respectively. The sequence coverage and protein score are not specified, as the identification and quantification of these proteins were done using artificial database entries. However, the average mass error calculated for all replicates was below 7.6 ppm. Studying only one biological replicate with replicates of extraction, digestion, and injection allowed assessing the reproducibility of the assay without a bias caused by biological variability. Relative standard deviation ranged from 10.5 to 13.5%. However, we applied the concept of calibration published by Silva et al.¹⁷ that is based on only one calibrant and poses a risk that individual values might be inaccurate even though this concept generally leads to precise and accurate values for many proteins. Thus, we quantified the allergens and iso-allergens by MRM using isotopically labeled synthetic variants of the same target peptides to confirm the quantitative values obtained by MS^E. The MRM derived values (see below) matched the MS^E data fairly well in most cases. Both values can be compared in Figure 3.

Semitargeted, label-free quantification of other birch pollen allergens and nonallergenic proteins

The values for the other birch pollen allergens and non-allergenic proteins are also shown in Table 1. Bet v 7.0101 was the most abundant allergen in this group followed by Bet v 6.0101, Bet v 2.0101, and Bet v 4.0101. In total, 10 nonallergenic proteins passed our acceptance criteria with a putative uncharacterized protein, described for *Vitis vinifera* (common grape vine) being the most abundant representative of this group. As unmodified protein entries were used for data processing and database search, protein scores, average mass error, number of peptides detected, and sequence coverage are specified for every protein. The reproducibility for four birch pollen allergens ranged between 8.2 and 16.2%, while relative standard deviation for nonallergenic proteins ranged from 7.9 and 33.4%.

Untargeted identification of nonallergenic proteins

A qualitative analysis of the label-free MS^E data was done by performing a database search with Database for untargeted identification (see Experimental Procedures section) consisting

of reviewed entries for green plants. The data are summarized in Table 1 (all rows without quantitative values). Mainly two pieces of information can be drawn from this set of data: (1) only 7 more proteins were identified as compared to the quantitative data; (2) three Bet v 1 sequence variants were additionally identified that failed the acceptance criteria for quantification. These sequence variants are Bet v 1.0301, Bet v 1.0113, and Bet v 1.0116.

Quantification of Bet v 1 variants and birch pollen allergens by MRM and correlation with MS data

In order to confirm MS^E data, Bet v 1 variants and the other birch allergens were quantified by the MRM method. Table S-5 (Supporting Information) shows the quantitative data for three extraction replicates and the overall mean values with their respective %CV. On the extraction level, the %CV values for all peptides ranged between 2.4% and 10.9%, except for Bet v 1-C (Bet v 1.0301). The %CV for Bet v 1-C ranged between 16.2% and 26%. The overall mean %CV values ranged from 4.1% for Bet v 1-E to 13.5% for Bet v 6. The Bet v 1-E (14 539 fmol) and Bet v 1-I (11 1726 fmol) peptides were the most abundant peptides, while Bet v 1-H (33 fmol; Bet v 1.0115) was the least abundant peptide. The quantitative values for all peptides spanned a range of three orders in magnitude (30 fmol to 14 629 fmol). It is noteworthy to mention that the individual values for Bet v 1-E and Bet v 1-I differ only by 20% and this data indicate that only one peptide could be used to quantify total Bet v 1 as the contribution of Bet v 1-C (Bet v 1.0301) and Bet v 1-H (Bet v 1.0115) to the total Bet v 1 amount is <1%. Seven sequence variants of Bet v 1 that can be quantified individually represent 46% of the total amount calculated for Bet v 1. Figure 3 and Table 2 illustrate the quantitative allergen

Table 2. Quantitative Profile of Allergens in Birch Pollen by MS^E and MRM^a

Protein	MS ^E	MRM
	% of birch pollen protein	
Bet v 1_T	27.6	23.2
Bet v 1.0102	6.5	6.2
Bet v 1.0104	2.5	2.4
Bet v 1.0203		1.1
Bet v 1.0301		0.1
Bet v 1.0113		0.5
Bet v 1.0116		0.3
Bet v 1.0119		0.1
Bet v 2.0101	2.8	2.6
Bet v 4.0101	0.7	0.4
Bet v 6.0102	6.0	7.3
Bet v 7.0101	8.0	2.4

^aProtein (represented by IUIS name or Bet v 1_T: total Bet v 1).

profile of birch pollen comparing MS^E with MRM data. Total Bet v 1 represented 28% and 23% of the total protein content when analyzed by MS^E and MRM, respectively. No significant difference in protein amount was observed for Bet v 1.0102 and Bet v 1.0104 variants. Two more iso-allergens (Bet v 1.0203 and Bet v 1.0301), and three variants Bet v 1.0113, Bet v 1.0116, and Bet v 1.0119 were quantified with MRM assay compared to MS^E. These five proteins represent 2% of the total protein amount and 9% of total Bet v 1. Almost the same quantitative values were determined for Bet v 2.0101 using both MS methods. The amount of Bet v 4 was lower by MRM

compared with MS^E data, while a slightly higher value for Bet v 6 was determined with MRM. Bet v 7 was found at a higher level using MS^E compared with the MRM method. From our data we cannot conclude why Bet v 7 content was overestimated with MS^E compared to MRM. We suspect that the problem originates from the natural Bet v 7 or the peptide amino acid sequence. These differences could also be attributed to different collision energies that are used during fragmentation of individual peptides, and the number of peptides that is used for obtaining the quantitative values for each protein. In the MS^E assay, precursor peptides that are used for quantification are recorded in the low-energy MS mode using constant collision energy, whereas in the MRM assay fragment ions of precursor peptides are used for quantification and collision energies are optimized for each target peptide individually. Additionally, the average intensity of the observed peptides/precursor ions for each individual protein is compared with the average intensity values of the three most abundant peptides/precursor ions of only one internal standard in the MS^E assay to quantify all proteins. Therefore, the same MS parameters were applied to all peptides and the standard in the MS^E assay, while in the MRM assay the MS parameters were optimized for every target peptide and its isotope-labeled standard. Still, the overall MS^E data correlated quite well with MRM data with only one exception, indicating that MS^E is precise and accurate in all cases except for Bet v 7. However, the correlation between MS^E and MRM values needs to be validated for each novel allergen that might be described in birch pollen in the future and also for other allergens from other allergenic sources, e.g. grass pollen. Bet v 2, Bet v 4, Bet v 6, and Bet v 7 comprised 18% and 13% of the total protein amount determined by MS^E and MRM assays, respectively.

Conceptual aspects

We established a workflow that comprises a targeted label-free, a semitargeted label-free, and untargeted label-free data analysis to comprehensively evaluate the protein pattern of a complex allergen product source material (Figure 1). The crucial aspect of the targeted approach to identify and quantify Bet v 1 and its sequence variants was a generation of artificial database entries. The application of such entries amended to a database that exclusively contains sequences of totally unrelated species facilitates two different things: (1) It allows detecting targets that require the assignment of quantitative values to specific and predefined i.e. IUIS approved and validated accession numbers. This approach avoids apparent reproducibility failure in replicate analysis due to varying database entries of sequence variants that are being used to describe the same MS and MS/MS spectra. (2) It essentially forces the software to detect preselected peptide targets and to assign quantitative values to them disabling any proper statistical scoring. This manipulation of the key step for the interpretation of MS data would be extremely prone to false positive results, if applied to an unknown sample. However, the intended use of this workflow is to evaluate the quality of a biomedical product that has been previously thoroughly characterized in the course of a marketing authorization procedure. This means that the presence of the proteins originally carrying these preselected target peptides was unambiguously confirmed before the targeted MS^E data analysis was applied. We have proven the suitability of this concept by studying a protein extract of birch pollen, which was extensively characterized by 2D-MS in our previous study.¹⁴ We quantified total Bet v 1 content by

selecting four peptides for the targeted approach, which together account for all Bet v 1 variants. We failed to detect Bet v 1.0119, and the results for Bet v 1.0301 did not pass our rigid inclusion criteria and were, consequently, disregarded from quantification. MRM data showed that these two isoallergens of Bet v 1 represent only 0.9% of total Bet v 1 content. Even though Bet v 1.0301 does not represent a major part of the total Bet v 1 content if the sole quantity is concerned, it may still represent a very significant part of the qualitative Bet v 1 sequence variant pattern, as its primary structure differs significantly from all other Bet v 1 sequence variants.¹⁴ Additionally, Bet v 1.0102 and Bet v 1.0104 were quantified by using artificial database entries for targeting the software to specific peptides. All other sequence variants were not detected because they were either below the detection limit (Bet v 1.0203, 1.0115 and 1.0119) giving consistent negative results, or they were at the detection limit (Bet v 1.0301, 1.0113 and 1.0116) giving positive and negative results in replicate analysis and failing the inclusion criteria. Our data clearly showed that the targeted analyses of MS^E data can reliably detect relevant predefined sequences in replicate analysis in cases where regular untargeted approaches would assign the same precursor to different entries looking at different replicates. The absence of any other entries of the species under investigation or the species of the same kingdom would also allow recognizing false positive hits easily. However, we never experienced a significant hit on a *Homo sapiens* protein entry. On the contrary, our semitargeted approach is less prone to false positive hits. We amended selected entries to a database with sequences of an unrelated species, but we did use unmodified protein entries. This allows complementing the quantitative data with protein scores. The PLGS protein scores obtained for birch pollen allergens were between approximately 2800 and 10 000, demonstrating statistical significance. However, these scores needed to be supported with additional information, such as a number of detected peptides, sequence coverage, and precursor mass errors, as the database is artificial in some respect, rendering the scores not meaningless but potentially biased. For example, the size and type of database contribute to the protein score, but the peptide scoring is unaffected, i.e. correct and unbiased. Additionally, this type of data analysis should be restricted to biomedical products that have been extensively characterized. In our current study, only entries that were first unambiguously confirmed to be present using the unrestricted database were used and added to a new database for semitargeted analysis. Therefore, false positive hits can only occur for *Homo sapiens* proteins, and thus can easily be recognized. Additionally, average mass errors between 1.9 and 4.7 ppm calculated on the basis of 5–18 individual peptides, representing 51–73% of the amino acid sequence, are sound values for a lock-mass corrected Q-TOF MS. A similar conclusion can be drawn from the data on nonallergenic proteins. Scores were from 312 to 1288 and the number of detected peptides was between 1 and 7, representing 8–25% of the amino acid sequences. Mass error ranged from 2.2 to 7.7 ppm. Two identifications based on a single peptide were only accepted because they were reproduced several times, passing our inclusion criteria described in the [Experimental Procedures](#) section. Unlike synthetic drugs or highly purified biologicals, complex biomedical products lack a clear-cut predefined and fully controllable composition. Therapeutic or diagnostic allergen products are mainly protein extracts and vary in their composition due to the natural variability of the source

material. For example, birch pollen is collected from field-grown trees and geographical differences, climate changes, environmental stress, etc. might cause significant variation in the protein pattern of the final products, even though the postharvest production process is tightly controlled. Both targeted and semitargeted approaches would only detect either the reduced amount or the total absence of one or more predefined proteins, but they would fail to detect proteins that have appeared for the first time. Therefore, we included an entirely untargeted approach to analyze such unexpected variations in the protein composition. Even though quantitative values were recorded for each protein, we only drew conclusions about their absence or presence because they failed the inclusion criteria for quantitative analysis. We found 43 nonallergenic proteins in our previous study using gel-digestion after 2D and picked up 17 of these proteins in-solution in this study. It is noteworthy that most nonallergenic proteins were reported as proteins from species other than birch, as birch is not yet fully sequenced. We have selected *Betula pendula* entries whenever possible and reported the other proteins anyway, as these types of proteins surely exist with partially identical amino acid sequences in other species and thus are a valuable description of the protein profile of birch pollen.

Allergological aspects

There is an ongoing debate whether Bet v 1 iso-allergens and other birch pollen minor allergens are relevant for clinical or diagnostic applications. Many scientists expressed concerns about the existence of multitude sequence variants and speculated whether many sequence variants are either not expressed *in vivo* for unknown reasons or are even cDNA cloning and sequencing artifacts. Two systematic studies showed that these speculations are unnecessary. Schenk et al.²⁰ demonstrated that different birch varieties do contain different and complex sequence variant patterns on the nucleic acid level, and we detected 18 sequence variants being expressed as proteins in parallel in one tissue.¹⁴ Hence, our overall aim was to determine total Bet v 1 content and to quantify individual Bet v 1 sequence variants. Seven Bet v 1 sequence variants were quantified by MRM, while with the label-free approach using targeted and semitargeted data analysis we could quantify 2 out of 5 variants. Nevertheless, several very interesting observations can be made from this data: (1) Total Bet v 1 content represents the largest protein portion; it accounts for approximately 28% of the total protein content; (2) Bet v 1.0102 (formerly known as Bet v 1d) and Bet v 1.0104 (formerly Bet v 1i) described as hypoallergenic, i.e. low IgE-binding^{21,22} variants account for 6.5 and 2.5% of the total protein content, respectively. This means that these two sequence variants may not significantly contribute to IgE-binding, that is the readout of IgE-potency assays as they account for approximately 32% of total Bet v 1 content. Analyzing the same birch pollen extract in our previous study, we additionally confirmed the presence of three more sequence variants that share between 96 and 98% of the amino acid sequence with Bet v 1.0102. One of these sequence variants (Bet v 1.0105; former Bet v 1g) was reported as low IgE-binding variant,²¹ indicating that the other two (Bet v 1.0113 and Bet v 1.0117) might exhibit similar IgE-binding capacity. Bet v 1.0113 represents 0.4% of total Bet v 1, while Bet v 1.0105 does not have a unique peptide i.e. cannot be individually quantified, and Bet v 1.0117 was not detected in-

solution experiments. Nonetheless, as they are present in this birch pollen extract, a certain part of total Bet v 1 content will not be detected by IgE-binding assays. Even though we could not exactly quantify this portion of total Bet v 1 content, it clearly exceeds one-third of total Bet v 1 content. While someone may doubt the clinical or diagnostic relevance of low IgE-binding sequence variants, Bet v 1.0102 does show T-cell reactivity;²¹ therefore, it is very unlikely that this sequence variant is entirely irrelevant for therapeutic applications. (3) The vast majority of scientific studies that refer to individual Bet v 1 sequences were done with Bet v 1.0101 (former Bet v 1a). As it turned out, its fraction of the total Bet v 1 content remains unknown. Bet v 1.0101 cannot be quantified individually, because it lacks a unique peptide.

Looking at the content of all other birch allergens, Bet v 1 (28%), as major constituent of birch pollen, is followed by the Bet v 7 (8%), Bet v 6 (6%), Bet v 2 (2.6%), and Bet v 4 (0.7%). Thus, the minor allergens of birch pollen represent approximately 17% of the total protein content and all allergens together sum up to 45%. This means that only half of the total protein content is not allergenic. Interestingly, only one nonallergenic protein (D7TQB6) is among the five most abundant proteins, confirming previous observations that allergens are often the most abundant proteins in allergenic source materials with the exception of Bet v 4.

Quality control and batch testing

We developed a workflow that for the first time facilitates parallel identification and quantification of all IUIS approved allergens of an allergen source, including sequence variants of Bet v 1. Additionally, quantitative values were obtained for 10 nonallergenic proteins and qualitative data were recorded for several Bet v 1 sequence variants and 7 nonallergenic proteins. Nevertheless, our data raise several interesting questions regarding a potential application for batch testing and/or allergen product standardization.

The targeted and semitargeted quantification of specific allergens, sequence variants, and nonallergenic proteins is valid from a mass spectrometric point of view. The careful selection of peptide-, protein-, and carrier-protein sequences facilitated the complete avoidance of false positive or false negative results. However, the application within any regulatory procedure requires going beyond technical correctness: (1) How is the regulatory relevance of iso-allergens and variants defined/determined? We have used a set of peptides to quantify total Bet v 1, which picks up all Bet v 1 variants and isoforms expect one, including several putative low IgE-binding variants. A lack of IgE-binding capacity may indicate irrelevance for diagnostic preparations under investigation, where the ability to bind IgE-antibodies is a key feature. Thus, qualitative and/or quantitative data for these variants should not be reported for diagnostic products. Nevertheless, these variants could have T-Cell reactivity and play a major role during specific immunotherapy. On the contrary to diagnostic products data, therapeutic products data should therefore contain information about these sequences. (2) How should redundant accession numbers be removed or selected from public databases? We have detected several variants at the protein level for the first time. For example, we selected autonomously one out of several sequentially identical accession numbers for the IUIS approval. But a different researcher might have selected a different accession number that would have been as correct as ours. This might not be an issue in scientific research, but for example, a

differing accession number might be interpreted as a differing active substance in a regulatory procedure. (3) How to select or define a panel of allergens for quantitative analysis with MS? IgE-potency assays may include IgE-reactivity to the other birch pollen allergens Bet v 2, 4, 6, and 7 representing a combined value of the total allergen content while MS offers the option to look either at all allergens or only at individual allergens. The MS assays are reproducible enough to meet regulatory requirements illustrated in Figure 4.

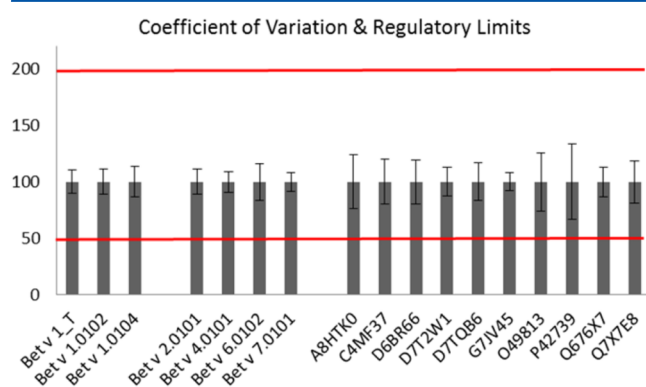


Figure 4. Coefficients of variation for all allergenic and nonallergenic targets determined by MS^E.

The allergen content needs to be within 50–200% of the stated amount (in the EU), and the CV values for all allergens were <17%. Such data might be very valuable for personalized medicine, e.g. to select products matching the specific sensitization profile of an individual patient. However, the quantification of multiple allergens, and even more challenging, the standardization of the content of multiple allergens, is so far not required. (4) How to define the fraction of the entire protein profile that needs to be identified to provide sufficient information on the entire composition of a product? We only detected 17 and quantified 10 out of 47 nonallergenic proteins that were detected by 2D-PAGE MS in our previous study, which is only 36 or 21%, respectively. This ratio improves to 26 detected (38%) and 17 quantified (25%) proteins by including Bet v 1 sequence variants and other birch pollen allergens (Bet v 2, 4, 6, and 7). This translates to 56% if total protein content is considered as a reference value. However, currently within regulatory procedures only SDS-PAGE is applied to evaluate consistent protein composition. A SDS-PAGE derived protein pattern of birch pollen is shown in Figure 5.

We identified between 23 and 26 individual bands by manual analysis and between 18 and 20 bands by a software aided approach, indicating that our approach identifies a larger fraction of the birch pollen protein profile. Additionally, it adds quantitative values which cannot be achieved with SDS-PAGE or 2D-PAGE, as each individual band or spot contains more than one protein.

Standardization of allergen products by synthetic reagents versus biological reference material

Allergen products are most commonly standardized focusing on total IgE-binding potency in comparison to a manufacturer specific in-house reference.²³ This allows the comparison of batches from one manufacturer but not between different manufacturers. Himly and co-workers published data on characterization of the recombinant grass pollen allergen Phl p 5.0109 as European pharmacopoeia standard where a

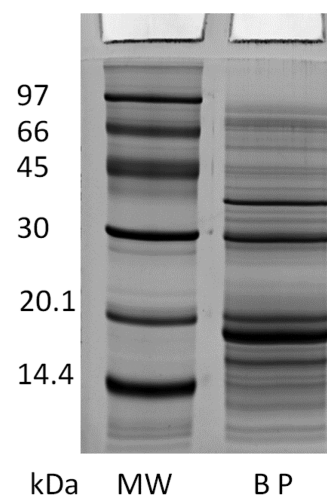


Figure 5. Protein pattern of birch pollen (BP) recorded with SDS-PAGE.

complex array of methods was applied to ensure identity, quantity, biochemical quality, and biological activity.²⁴ In 2016, an ELISA for quantification of Bet v 1 was published as one result of the CREATE project that commenced in 2001 and its successor BSP090, a joint program that is included in the Biological Standardization Program of the European directory for the Quality of Medicines & Healthcare (EDQM).²⁵ This ELISA was validated as a candidate European Pharmacopoeia standard. Both rPhl p 5 standard and Bet v 1 ELISA can be utilized for determining mass units of allergens in different products, i.e. facilitating cross manufacturer comparison. However, it seems to be unfeasible to develop biological reference materials or antibody based quantitative assays, considering the large number of allergen sources and the multiplicity of different allergens and allergen variants within these allergen sources. Additionally, it remains unknown which part of the Bet v 1 or Phl p 5 isoform pattern is addressed. On the contrary, the label-free MS method relies on synthetic standards, i.e. one synthetic calibration peptide, which can be produced at constant quality in virtually unlimited amounts that does not require extensive characterization. The MRM method involving stable isotope-labeled, target-specific synthetic calibration peptides can be applied for absolute quantification. Thus, applying MS for allergen quantification makes it more likely that standardization of individual allergen content could be expanded to most allergen products on the market. MS also offers more flexibility regarding specificity. It facilitates targeting either individual sequence variants or total allergen values. Even though we could not quantify all individual sequence variants of Bet v 1, we can conclude that these sequences are not very likely highly abundant. Nevertheless, we developed two assays that for the first time allow quantification of all known allergens in a complex allergen source, offering a possibility to consider the content of minor allergens for standardization of allergen products.

CONCLUSIONS

This study describes the first comprehensive quantification of the birch pollen protein profile, including all allergens at iso-allergen and variant levels, and nonallergenic proteins in parallel. The label-free MS method has gained popularity due to its simplicity. However, it fails in application if a reproducible, unambiguous identification, and quantification of sequence

variants and specifically predefined allergens is needed. By combining targeted, semitargeted, and untargeted data analysis we gained reproducibility and unambiguity without compromising specificity. This approach can only be applied on well-characterized samples, thus being ideal for characterization or standardization of complex allergen products. Another advantage of data-independent acquisition mode is the option to retrospectively mine the data for formerly unknown allergens when they become relevant. After reviewing our data, we found that we have quantified Glutathione-S-Transferase (Acc. No. D6BR66) as a nonallergenic protein that was afterward suggested by Deiff et al.²⁶ as a new minor allergen (IUIS approval has not yet been obtained).

The assay provides more comprehensive information on the nonallergenic protein profile than currently applied methods. However, it remains to be decided by regulatory bodies how to interpret such data and how to address the following questions: (1) How large should be the protein fraction detected by MS? (2) Are 10 out of 17 detected proteins enough? (3) Do 17 proteins represent a meaningful fraction out of 47 nonallergenic proteins? (4) How large is the protein fraction of birch pollen that is not represented by these 47 proteins? The same questions also apply to other allergen sources that are not yet completely sequenced.

Nonetheless, this study provides a model for development of advanced methods for batch consistency testing in order to improve the safety and efficacy of birch pollen allergen products. Data independent acquisition modes such as MS^E could further be applied to study the natural variability in protein and allergen expression of different birch pollen varieties and other allergen sources, which is a critical step in characterization and development of diagnostic and therapeutic allergen products. Even though the targeted MRM method might be preferred when the accuracy of absolute amounts is essential, we have shown that MS^E derived quantitative values for birch pollen allergens are in most cases correct. However, further studies on the IgE- or T-cell reactivity of sequence variants and isoallergens would provide information about which allergens, iso-allergens, or variants should be considered for quality control of allergen products. Finally, this information would be extremely helpful in order to implement MS^E or MRM methods in quality control of allergen products for which source materials are less comprehensively characterized than birch pollen.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00490.

Figures and tables that are originally mentioned in experimental procedures with intention to describe the workflow and the databases that were used to analyze MSE data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Andreas.Reuter@pei.de. Phone: +49 6103 775210.

ORCID

Jelena Spiric: 0000-0001-9027-4774

Andreas Reuter: 0000-0002-6082-9000

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) *Immunology: Chapter 4: Antibodies: Structure and Function*; Goldsby, R. A. K. i., Ed., 5th ed.; W.H. Freeman: New York, Houndmills, 2003.
- (2) van Ree, R.; Chapman, M. D.; Ferreira, F.; Vieths, S.; Bryan, D.; Cromwell, O.; Villalba, M.; Durham, S. R.; Becker, W. M.; Aalbers, M.; et al. The CREATE project: Development of certified reference materials for allergenic products and validation of methods for their quantification. *Allergy* **2008**, *63*, 310–326.
- (3) Ballmer-Weber, B. K.; Wangorsch, A.; Bohle, B.; Kaul, S.; Kündig, T.; Fötisch, K.; van Ree, R.; Vieths, S. Component-resolved in vitro diagnosis in carrot allergy: does the use of recombinant carrot allergens improve the reliability of the diagnostic procedure? *Clin. Exp. Allergy* **2005**, *35*, 970–978.
- (4) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.O111.016717.10.1074/mcp.O111.016717
- (5) Silva, J. C.; Gorenstein, M. V.; Li, G.-Z.; Vissers, J. P. C.; Geromanos, S. J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **2006**, *5*, 144–156.
- (6) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **2007**, *261*, 1–12.
- (7) Kinyua, J.; Negreira, N.; Ibáñez, M.; Bijlsma, L.; Hernández, F.; Covaci, A.; van Nuijs, A. L. N. A data-independent acquisition workflow for qualitative screening of new psychoactive substances in biological samples. *Anal. Bioanal. Chem.* **2015**, *407*, 8773–8785.
- (8) Agilent Technologies. Agilent | Mass Spectrometry All Ions Technology. <http://www.agilent.com/en-us/promotions/allionstechnology> (accessed June 27, 2017).
- (9) Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419–1440.
- (10) He, Z.; Huang, T.; Liu, X.; Zhu, P.; Teng, B.; Deng, S. Protein inference: A protein quantification perspective. *Comput. Biol. Chem.* **2016**, *63*, 21–29.
- (11) Claassen, M.; Reiter, L.; Hengartner, M. O.; Buhmann, J. M.; Aebersold, R. Generic comparison of protein inference engines. *Mol. Cell. Proteomics* **2012**, *11*, O110.007088.O110.007088.10.1074/mcp.O110.007088
- (12) Ma, Z.-Q.; Dasari, S.; Chambers, M. C.; Litton, M. D.; Sobecki, S. M.; Zimmerman, L. J.; Halvey, P. J.; Schilling, B.; Drake, P. M.; Gibson, B. W.; et al. IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J. Proteome Res.* **2009**, *8*, 3872–3881.
- (13) He, Z.; Yang, C.; Yu, W. A partial set covering model for protein mixture identification using mass spectrometry data. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2011**, *8*, 368–380.
- (14) Spiric, J.; Engin, A. M.; Karas, M.; Reuter, A. Quality Control of Biomedical Allergen Products - Highly Complex Isoallergen Composition Challenges Standard MS Database Search and Requires Manual Data Analyses. *PLoS One* **2015**, *10*, e0142404.
- (15) Meyer-Arendt, K.; Old, W. M.; Houel, S.; Renganathan, K.; Eichelberger, B.; Resing, K. A.; Ahn, N. G. IsoformResolver: A peptide-centric algorithm for protein inference. *J. Proteome Res.* **2011**, *10*, 3060–3075.
- (16) Lukasse, P. N. J.; America, A. H. P. Protein inference using Peptide quantification patterns. *J. Proteome Res.* **2014**, *13*, 3191–3199.
- (17) Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Li, G.-Z.; Richardson, K.; Wall, D.; Geromanos, S. J. Simultaneous qualitative

and quantitative analysis of the Escherichia coli proteome: A sweet tale. *Mol. Cell. Proteomics* **2006**, *5*, 589–607.

(18) Practical Molecular Biology Conversion. *Conversion: weight - moles (for proteins)* | Practical Molecular Biology. http://www.molbiol.edu.ru/eng/scripts/01_04.html (accessed June 27, 2017).

(19) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G.-Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; et al. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **2005**, *77*, 2187–2200.

(20) Schenk, M. F.; Cordewener, J. H. G.; America, A. H. P.; Van't Westende, W. P. C.; Smulders, M. J. M.; Gilissen, L. J. W. J. Characterization of PR-10 genes from eight Betula species and detection of Bet v 1 isoforms in birch pollen. *BMC Plant Biol.* **2009**, *9*, 24.

(21) Ferreira, F.; Hirtenlehner, K.; Jilek, A.; Godnik-Cvar, J.; Breiteneder, H.; Grimm, R.; Hoffmann-Sommergruber, K.; Scheiner, O.; Kraft, D.; Breitenbach, M.; et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: Potential use of hypoallergenic isoforms for immunotherapy. *Journal of experimental medicine* **1996**, *183*, 599–609.

(22) Breitenbach, M.; Ferreira, F.; Jilek, A.; Swoboda, I.; Ebner, C.; Hoffmann-Sommergruber, K.; Briza, P.; Scheiner, O.; Kraft, D. Biological and immunological importance of Bet v 1 isoforms. *Advances in experimental medicine and biology* **1996**, *409*, 117–126.

(23) Zimmer, J.; Vieths, S.; Kaul, S. Standardization and Regulation of Allergen Products in the European Union. *Curr. Allergy Asthma Rep.* **2016**, *16*, 21.

(24) Himly, M.; Nandy, A.; Kahlert, H.; Thilker, M.; Steiner, M.; Briza, P.; Neubauer, A.; Klysner, S.; van Ree, R.; Buchheit, K.-H.; et al. Standardization of allergen products: 2. Detailed characterization of GMP-produced recombinant Phl p 5.0109 as European Pharmacopoeia reference standard. *Allergy* **2016**, *71*, 495–504.

(25) Kaul, S.; Zimmer, J.; Dehus, O.; Costanzo, A.; Daas, A.; Buchheit, K. H.; Asturias, J. A.; Barber, D.; Carnés, J.; Chapman, M.; et al. Standardization of allergen products: 3. Validation of candidate European Pharmacopoeia standard methods for quantification of major birch allergen Bet v 1. *Allergy* **2016**, *71*, 1414–1424.

(26) Deifl, S.; Zwicker, C.; Vejvar, E.; Kitzmüller, C.; Gadermaier, G.; Nagl, B.; Vrtala, S.; Briza, P.; Zlabinger, G. J.; Jahn-Schmid, B.; et al. Glutathione-S-transferase: a minor allergen in birch pollen due to limited release from hydrated pollen. *PLoS One* **2014**, *9*, e109075.

Supporting Information

Title: Model for Quality Control of Allergen Products with Mass Spectrometry

Authors: Jelena Spiric,^{*,†} Thomas Schulenburg,[†] Luisa Schwaben,[†] Anna M. Engin,[†] Michael Karas,[‡] Andreas Reuter[†]

[†]Division of Allergology, Paul-Ehrlich-Institut, Langen, 63225 Germany

[‡]Institute of Pharmaceutical Chemistry, University of Frankfurt, 60438 Frankfurt, Germany

Table of contents:

Figure S-1. Sequence alignment of 18 Bet v 1 Iso-allergens that are expressed in parallel.

Table S-1. Tryptic peptides of Bet v 1 sequence variants.

Table S-2. Artificial database entries.

Table S-3. List of accession numbers added to Uniprot Homosapiens database for quantification of birch pollen protein profile.

Table S-4. Optimized MRM parameters for allergen-specific peptides

Table S-5. Quantification of peptides by MRM

Table S-6. Specificity testing in-silico by BLAST

MRM method development

Bet v	Acc.No.	T1	T19	T22	T34	T56	T57	T67	T70	T72	T79		
1.0101	P15494	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0103	P43178	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ IPFK	YVK	GR	VDEV ~~D~~ HNTFK			
1.0106	P43183	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0104	P43179	MGVFN ~~Y~~ EIEATSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0102	P43177	MGVFN ~~Y~~ EIEETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ VPK	VAPQAISSVENIEGNGGPGTIK	K	INFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0105	P43180	MGVFN ~~Y~~ ESETTSVIPAAAR	L ~~F~~ PK	AFILEGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	INFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0204	P43186	MGVFN ~~Y~~ ESETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ITFPEG ~~G~~ SFFK	YVK	ER	VDEV ~~D~~ HANFK			
1.0203	P43184	MGVFN ~~Y~~ ESETTSVIPAAAR	L ~~F~~ PK	AFILEGDT ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ITFPEG ~~G~~ SFFK	YVK	ER	VDEV ~~D~~ HANFK			
1.0301	Q39415	MGVFN ~~Y~~ EDEATSVIPAAAR	L ~~F~~ PK	SVFLDADN ~~L~~ IPK	VAPENVSSAENIEGNGGPGTIK	K	ITFPEG ~~G~~ SHFK	YMK	HR	VDEIDHANFK			
1.0205	Q39427	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ITFPEG ~~G~~ SFFK	YVK	ER	VDEV ~~D~~ HANFK			
1.0115	Q39431	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~A~~ HK	NEK		
1.0113	Q96370	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ VPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0116	O23748	-GVFN ~~Y~~ EIEATSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ IPFK	YVK	DR	VDEV ~~D~~ HANFK			
1.0119	Q0QLS9	---FN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVQNIENGGGPGTIK	K	INFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0206	O23749	-GVFN ~~Y~~ ESETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ITFPEG ~~G~~ SFFK	YVK	ER	VDEV ~~D~~ HANFK			
1.0117	Q9SC10	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ FPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
1.0207	Q0QLV2	---FN ~~Y~~ ESETTSVIPAAAR	L ~~F~~ PK	AFILEGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ITFPEG ~~G~~ SFFK	YVK	ER	VDEV ~~D~~ HANFK			
1.0118	Q9SCH6	MGVFN ~~Y~~ ETETTSVIPAAAR	L ~~F~~ PK	AFILDGDN ~~L~~ IPK	VAPQAISSVENIEGNGGPGTIK	K	ISFPEG ~~G~~ FFK	YVK	DR	VDEV ~~D~~ HNTFK			
Peptide variants		6	1	6	3	1	5	2	4	4	1		
Bet v	UniProt	T82	T99	T105	T113	T117	T121	T125	T131	T136	T139	T141	T147
1.0101	P15494	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0103	P43178	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQIK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0106	P43183	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQIK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0104	P43179	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQIK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0102	P43177	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0105	P43180	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0204	P43186	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0203	P43184	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0301	Q39415	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0205	Q39427	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0115	Q39431	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0113	Q96370	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0116	O23748	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GDHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0119	Q0QLS9	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0206	O23749	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0117	Q9SC10	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0207	Q0QLV2	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
1.0118	Q9SCH6	YNSVIEGGPIGDTLEK	ISNEIK	IVATPDGGSILK	ISNK	YHTK	GNHEVK	AEQVK	ASK	EMGETLLR	AVESYLLAHSDAYN		
Peptide variants		8	3	9	1	3	1	4	3	3	2	4	2

Figure S-1. Sequence alignment of 18 Bet v 1 Iso-allergens that are expressed in parallel. Gaps indicate trypsin cleavage sites. Tryptic peptides are enumerated according to the amino acid position. For example, T99 starts at position 99 and exists in all proteins, while T105 exists in all proteins but is longer for Bet v 1.0118 and T113 exclusively exits for Bet v 1.0115. Peptides with 100% sequence identity that are conserved for all proteins are printed in black. Identical peptides are grouped by identical colors. Unique sequence stretches are underlined. A – indicates amino acid positions with a lack of genomic sequence data. Those peptides were color coded by assuming that MGV is highly conserved for T1 peptides and that AVESYLL is always followed by AHSDAYN. (Modified Figure 3 from Spiric et al.¹⁴)

Table S-1. Tryptic peptides of Bet v 1 sequence variants.

Allergen*	Acc.No. [#]	T1 [§]	T19	T22	T34	T56	T57	T67	T70	T72	T79	T82	T99	T105	T113	T117	T121	T125	T131	T136	T139	T141	T147
Bet v 1.0101	P15494	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0103	P43178	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0106	P43183	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0104	P43179	<u>+</u>	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0102	P43177	<u>+</u>	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0105	P43180	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0204	P43186	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	+	-	-	+	+
Bet v 1.0203	P43184	+	-	<u>+</u>	<u>+</u>	-	+	-	-	+	ne	+	+	+	ne	-	-	+	+	-	-	+	+
Bet v 1.0301	Q39415	<u>+</u>	-	<u>+</u>	<u>+</u>	-	-	-	-	<u>+</u>	ne	-	-	<u>+</u>	ne	-	-	-	ne	-	-	-	-
Bet v 1.0205	Q39427	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	+	-	-	-	+
Bet v 1.0115	Q39431	+	-	+	+	-	+	-	-	-	-	-	+	-	-	-	-	+	-	-	+	ne	+
Bet v 1.0113	Q96370	+	-	+	+	-	+	-	-	+	ne	<u>+</u>	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0116	O23748	+	-	+	+	-	+	-	-	+	ne	<u>+</u>	+	+	ne	-	-	+	-	-	+	ne	+
Bet v 1.0119	Q0QLS9	+	-	+	-	-	+	-	-	+	ne	+	+	<u>+</u>	ne	-	-	+	-	-	+	ne	+
Bet v 1.0206	O23749	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	+	-	-	<u>+</u>	+
Bet v 1.0117	Q9SCI0	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	-	-	-	+	ne	+
Bet v 1.0207	Q0QLV2	+	-	+	+	-	+	-	-	+	ne	+	+	+	ne	-	-	+	+	-	-	+	+
Bet v 1.0118	Q9SCH6	+	-	+	+	-	+	-	-	+	ne	+	+	-	ne	ne	-	+	+	-	-	+	+

*: IUIS approved iso-allergen and variant names of major birch pollen allergen Bet v 1. #: UniProt accession numbers. §: Start position of tryptic peptides according to amino acid sequences as presented in supplementary figure 1. +: detectable from in-gel digestion in previous study, -: not detectable from in-gel digestion in previous study, underlined: sequence variant specific i.e unique peptide, na: non-existent, red: detectable in-solution in current study and selected for quantification.

Table S-2. Artificial database entries.

Accession Numbers*	Description	Target Bet v 1 variant	Target Peptide [#]
- P02769	carrier sequence	-	-
MKWVFISILLSSAYSIRGVFRDRTHKSEIAHRFDLGEHFHKLVLIAFSQYLQCPDFDEHVKLVLNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCEKQEPERNECFLSHKDDSPDLPKLPDPNTLCDEFKADEKFKFWGKLYEIAARRHPYFYAPELLYANKYNGVGFQECQAEADKGACLLPKIETMREKVLASSARQLRCSAQKFGGERALKAWSVARLSQKFPKAEFVETKLVDTLTKVHKECHGDLLECADDRAADLAKYICDNQDITSSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAQDAFLGSLFYEYSRRHPYAVSVLLRLAKEYEATLECCAKDDPHACYSTVFDKLHLVDEPQNLKQNCDOFEKLGVEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCTKPESEIRMPCTEDYLSULNRLCVLHEKTPVSEKVTCKCTESLVNRRPCFSALTPDETIVPKAFDEKLFTHADICTLPDTEKQIKKQIALVELLKHKPKATEEQQLKVTMENFVAFVDKCAADDEKACFAVEGPKLVSTQTALA			
JS001	- Bet v 1_Total quantifier	-	T34 blue / T147 red
MKWVTF.....	AVEGPKVAPQAISVENIEGNGGPGTIKLVVSTQTALAKAVESYLLAHSDAYN [§]		
JS002	P43179 variant specific quantifier	Bet v 1.0104	T1
MKMGVFNIEATSVIPAAARWVTFI.....	LVVSTQTALA		
JS003	P43177 variant specific quantifier	Bet v 1.0102	T1
MKMGVFNIEITTSVIPAAARWVTFI.....	LVVSTQTALA		
JS004	P43184 variant specific quantifier	Bet v 1.0203	T22
MKAFILEGDTLIPKWVTFI.....	LVVSTQTALA		
JS005	Q39415 original specific sequence	Bet v 1.0301	-
MGVFNIEDEATSVIAPARLFKSVLDADNLIPKAPENVSAAENIEGNGGPGTIKITFPESGSHFYMKHRVDEIDHANFKYCYSIIIEGGLGDTLEKISYEIKIVAAPGGGSILKITSKYHTKGDISLNEEEKAGKEKGAGLFAVENYLVHNPAYN			
JS006	Q39431 variant specific quantifier	Bet v 1.0115	T82
MKYSYSIEGGPIGDTLEKVVTFI.....	LVVSTQTALA		
JS007	Q96370 variant specific quantifier	Bet v 1.0113	T82
MKYNYSIEGGPMGDTLEKVVTFI.....	LVVSTQTALA		
JS008	O23748 variant specific quantifier	Bet v 1.0116	T82
MKYSYSIEGGPVGDTLEKVVTFI.....	LVVSTQTALA		
JS009	Q0QLS9 variant specific quantifier	Bet v 1.0119	T105
MKIVATPDGGCILKVVTFI.....	LVVSTQTALA		

*: Accession numbers are either in-house (JS...) or from UniProt. #: Target peptide positions refer to table S1 and figure S1. The figure S1 colour code is given for T 34 & T147 as they do not origin from a specific sequence. §: Only the amino acid residues flanking the target peptides are shown. The dots represent the unmodified amino acid residues of the remaining carrier sequence.

Table S-3. List of accession numbers added to Uniprot Homosapiens database for quantification of birch pollen protein profile.

UniProt Accession Numbers			
A4K9Z8	B6T7G7	D7TQB6	Q20CC1
A2Y7R5	B9GCF2	D7U461	Q2MCJ5
A2YSU3	B9MX28	D8SHY6	Q2XNS9
A2ZNJ7	B9RB02	D8SZX6	Q5TIQ0
A4RR85	B9REC4	D8TAA3	Q676X7
A5AH06	B9SB56	D8TBQ5	Q69JG0
A5BNR6	C4MF37	D9ZIQ8	Q6YWB9
A5BWC9	C5WQ78	E1Z4S1	Q7X7E8
A8HTK0	C5YD96	E1ZFX4	Q7X9E5
A9NMR0	C6SZN7	F2CUL2	Q8L8A7
A9P208	C6T9X5	F6GVW7	Q8LEB8
Q39419	C6TJB9	F6H7L5	Q8S3L1
A4K9Z8	D1LWT8	G7JV45	Q8W2G9
A9SJE6	D2D2Z8	G7K7W9	Q96447
B4FI58	D2X9U2	O49813	Q9FHQ6
B4YB12	D3WFE0	O65735	Q9SWA8
B5M1Y7	D6BR66	P42644	Q8L5T1
B5SVG6	D7KLT8	P42739	Q9FUW6
B5TV62	D7M9C8	P59232	Q0QLS9
B6SWQ7	D7NJ40	P86104	
B6T748	D7T2W1	Q07796	

Table S-4. Optimized MRM parameters for allergen-specific peptides

Peptide	IUIS Name	Amino acid sequence	Peptide Length	Fragment ions	RT (min)	Precursor ion (m/z)	CoV (V)	Product ions							
								Quan. (m/z)	CE (V)	1st Qual. (m/z)	CE (V)	2nd Qual. (m/z)	CE (V)	3rd Qual. (m/z)	CE (V)
Bet v 1-A	Bet v 1.0102	GVFN Y IEITTSVIPAA	17	y9, y10 , y11, y12	6, 41	934, 55 939, 5	20	1044, 6	20	915, 55	20	1157, 65	20	1286, 7	20
		GVFN Y IEITTSVIPAA-R*						1054, 6		925, 55		1167, 6		1296, 7	
Bet v 1-B	Bet v 1.0104	GVFN Y IEATSVIPAA	17	y8, y9 , y10, y11	6, 69	919, 55 924, 5	30	885, 5	30	814, 5	30	1014, 55	30	1127, 7	30
		GVFN Y IEATSVIPAA-R*						895, 5		824, 5		1024, 55		1137, 7	
Bet v 1-C	Bet v 1.0301	SFV L DADNLIPK	12	y7 , y8, y9, y10	6, 86	666, 8 670, 75	10	770, 45	19	1097, 6	19	885, 45	20	998, 55	19
		SFV L DADNLIP-K*						778, 45		1105, 6		893, 45		1006, 6	
Bet v 1-D	Bet v 1.0203	AFI L EGDTLIPK	12	b5, y5, y6, y7	6, 67	659, 3 663, 25	20	743, 45	16	574, 3	16	571, 4	25	686, 4	16
		AFI L EGDTLIP-K*						751, 45		574, 3		579, 4		694, 4	
Bet v 1-E**		VAPQAISSVENI E GNGGPGTIK	22	y7 , y8, y9, y10	5, 21	1070, 2 1074, 15	40	629, 35	34	800, 45	37	929, 45	31	743, 4	31
		VAPQAISSVENI E GNGGPGTI-K*						637, 35		808, 45		937, 45		751, 4	
Bet v 1-F	Bet v 1.0113	YNSV I EGGPMGDTLEK	17	y9, y10 , y11, y12	5, 66	937, 55 941, 5	20	1004, 5	31	1246, 6	28	1133, 5	28	947, 45	22
		YNSV I EGGPMGDTLE-K*						1012, 5		1254, 6		1141, 55		955, 45	
Bet v 1-G	Bet v 1.0116	YSY S LIEGGPVGDTLEK	17	y8, y10 , y11, y12	5, 98	915 919	20	972, 5	28	1101, 5	28	858, 45	28	1214, 6	28
		YSY S LIEGGPVGDTLE-K*						980, 5		1109, 6		866, 45		1222, 7	
Bet v 1-H	Bet v 1.0119	IVAT P DGG-Camc-ILK	12	y7, y8 , y9, y10	4, 66	622, 75 626, 75	20	859, 45	19	1031, 5	19	960, 5	19	762, 4	25
		IVAT P DGG-Camc-IL-K*						867, 45		1039, 6		968, 5		770, 4	
Bet v 1-I**		AVE S YLLAHSDAYN	14	y8 , y9, y10, y11	5, 1	777, 35 780, 8	30	890, 4	21	1003, 5	21	1253, 6	25	1166, 6	25
		AVE S Y-L*-AHSDAYN						897, 4		1010, 5		1260, 6		1173, 6	
Bet v 2	Bet v 2.0101	LGDY L IDQGL	10	b4, b5, b6 , b7	6, 61	554, 13 557, 6	20	675, 35	13	449, 2	21	562, 3	25	790, 4	13
		LGDY L IDQG-L*						675, 35		449, 2		562, 3		790, 4	
Bet v 4	Bet v 4.0101	ISAA E LGEALK	11	y6 , y7, y8, y9	5, 33	551, 65 555, 6	40	630, 4	25	901, 5	25	830, 45	25	759, 4	25
		ISAA E LGEAL-K*						638, 4		909, 5		838, 5		767, 45	
Bet v 6	Bet v 6.0101	NIYSF N EIVALWEK	14	y5 , y6, y7, y8	8, 27	863, 45 867, 95	40	646, 35	25	745, 4	31	858, 5	25	987, 55	25
		NIYSF N EIVALWE-K*						654, 35		753, 45		866, 5		995, 55	
Bet v 7	Bet v 7.0101	TSKPVV V AD-Camc-GQLS	14	b8 , b9, b10, b11	3, 97	731, 35 734, 8	20	782, 5	22	897, 5	25	1057, 55	25	1114, 6	25
		TSKPVV V AD-Camc-GQ-L*-S						782, 5		897, 5		1057, 55		1114, 6	

*: Isotope-labeled sequence; **: Two most commonly shared sequences of Bet v 1 variants; CoV: Cone voltage; CE: Collision energy; Quant.: Quantifier ion; Qual.: Qualifier ion
Red colored amino acids are identified with respected MRM transitions. Red colored fragment ions are used as quantifier ions.

Table S-5. Quantification of peptides by MRM

Peptide	Cal. curve range (fmol)	E-I (fmol)	%CV	E-II (fmol)	%CV	E-III (fmol)	%CV	Avg (fmol)	%CV	Amount [µg/mg]	% Birch pollen protein
Bet v 1-A	1000-16000	3662	8,2	3811	7,2	3109	3,2	3527	10,5	61,9	6,2
Bet v 1-B	725-12000	1450	9,4	1443	7,2	1183	5,9	1359	11,2	23,9	2,4
Bet v 1-C	25-400	52	16,2	57	12,8	48	25,9	52	8,0	0,7	0,1
Bet v 1-D	150-2400	665	6,9	726	8,3	578	7,4	656	11,3	11,5	1,1
Bet v 1-E	3750-60000	14629	8,0	15085	6,0	13902	5,2	14539	4,1	255,5	25,5
Bet v 1-F	100-1600	274	8,8	290	6,4	265	6,6	276	4,6	4,9	0,5
Bet v 1-G	62,5-1000	146	6,3	151	6,8	134	7,0	144	6,2	2,5	0,3
Bet v 1-H	7,5-120	30	10,9	33	7,9	34	5,0	33	6,5	0,5	0,1
Bet v 1-I	3750-60000	11036	3,9	11621	2,4	12521	3,4	11726	6,4	206,0	20,6
Bet v 1_T		12915		13443		13294		13217	2,0	232,0	23,2
Bet v 2	375-6000	1874	3,8	1983	3,0	1785	7,6	1881	5,3	26,3	2,6
Bet v 4	125-2000	390	5,6	424	4,0	348	4,7	387	9,8	3,7	0,4
Bet v 6	500-8000	1834	3,5	2409	4,5	2197	6,6	2147	13,5	73,0	7,3
Bet v 7	300-4800	1216	5,2	1302	3,8	1366	5,2	1294	5,8	23,7	2,4

E: mean value for each extraction, Avg: Mean value of three extraction replicates with corresponding %CV values.

Molecular mass of Bet v 1.0101 (17.57 kDa) was used for calculating amount of Bet v 1-E and Bet v 1-I peptides.

Table S-6. Specificity testing in-silico by BLAST

Peptide	Ogranisms with 100% identical peptide	Peptide	Ogranisms with 100% identical peptide
Bet v 1-A	Betula chichibuensis	Bet v 2	Amborella trichopoda
	Betula lenta (Sweet birch)		Apium graveolens (Celery)
	Betula nigra (River birch)		Beta vulgaris (Sugar beet)
	Betula platyphylla (Asian white birch)		Citrus clementina
	Betula populifolia (Gray birch)		Corylus avellana (European hazel)
Bet v 1-B	Betula platyphylla (Asian white birch)		Cucumis melo (Muskmelon)
	Betula costata		Cucumis melo var. Cantalupensis
Bet v 1-C	Betula nigra (River birch)		Elaeis guineensis var. tenera (Oil palm)
	Betula lenta (Sweet birch)		Eucalyptus grandis (Flooded gum)
	Betula costata		Fraxinus excelsior (European ash)
	Betula chichibuensis		Genlisea aurea
	Betula schmidtii		Glycine max (Soybean) (Glycine hispida)
	Betula platyphylla (Asian white birch)		Humulus scandens (Hop)
	Betula populifolia (Gray birch)		Jatropha curcas (Barbados nut)
	Corylus avellana (European hazel)		Lotus japonicus
	Corylus heterophylla		Malus domestica (Apple) (Pyrus malus)
			Medicago truncatula (Barrel medic)
Bet v 1-D	Betula platyphylla (Asian white birch)		Musa acuminata (Wild banana)
	Betula costata		Petroselinum crispum (Parsley)
Bet v 1-E	All Betula spicies		Phaseolus vulgaris (Kidney bean)
			Plantago major (Common plantain)
Bet v 1-F	Betula platyphylla (Asian white birch)		Populus trichocarpa (Western balsam poplar)
			Prunus persica (Peach) (Amygdalus persica)
Bet v 1-G	-		Pyrus communis (Pear) (Pyrus domestica)
			Ricinus communis (Castor bean)
Bet v 1-H	-		Vitis vinifera (Grape)
		Bet v 4	-
Bet v 1-I	All Betula spicies	Bet v 6	-
		Bet v 7	Gossypium hirsutum (Upland cotton)
			Mimulus guttatus (Spotted monkey flower)
			Populus trichocarpa (Western balsam poplar)

MRM method development:

Peptide standards

A list of peptide sequences used for development of the MRM method for quantification of birch allergens and iso-allergens was generated from the data obtained in 2D and in solution analysis of birch pollen proteome using Q-TOF MS instrument. We selected peptides for quantification of: 1) total Bet v 1; 2) individual quantification of seven Bet v 1 sequence variants; 3) other four birch allergens. To quantify total Bet v 1, as it was done for the MSE assay, we averaged intensity values for the two most commonly shared Bet v 1 peptides (T34-blue and T147-red), and added the values obtained for Bet v 1.0301 and Bet v 1.0119 peptides, because these two Bet v 1 sequence variants are not represented by the T34-blue and T147-red peptides. For individual quantification of Bet v 1 sequence variants, we selected seven variant-specific peptides. In the case of the other four birch allergens we selected peptides that gave the best fragmentation pattern and highest signal intensity on the high resolution instrument.

We determined the specificity of each peptide in-silico using Basic Logical Alignment Search Tool (BLAST). Each peptide was BLAST searched against all Uniprot database entries, and the list of organisms with identical peptide sequence is presented in Table S-6. All Bet v 1 peptides are shared among many members of *Betula genus*. In addition to *Betula genus*, Bet v 1-C peptide is found in four variants of Cor a 1.04, which is a major allergen in hazelnut (*Corylus avellana*). Bet v 1-G and Bet v 1-H peptides are unique to *B.pendula*. The Bet v 2 peptide is shared in more than 20 different plant species and the majority of them are food sources. Peptides for Bet v 4 and Bet v 6 are unique for birch pollen. The Bet v 7 peptide is present in three additional plant species. Even though we identified species with identical peptides, this finding does not compromise the specificity of the assay as the majority of the identical peptides belongs to *Betula genus* or is found in species that one would not expect to be possible contaminants of birch pollen allergen products. The results of in silico specificity testing are listed above in supplementary table S-4.

To further develop and optimize MRM assay two types of synthetic peptides were purchased from JPT Peptide Technologies (Berlin, Germany): 1) SpikeTides-unpurified non-labeled peptides with or without C-terminal lysine or arginine; 2) SpikeTides TQL peptides labeled with stable isotopes (C terminal Arg U-13C6; U-15N4 or Lys U-13C6; U-15N2). Peptides were reconstituted according to manufacturer's instructions. Briefly, peptides were solubilized in 20% acetonitrile, 80% 100 mM NH₄HCO₃ solution, mixed for 30 min under gentle agitation at 500 rpm using Thermomixer (Eppendorf, Hamburg, Germany), and aliquots were stored at -80°C.

Determination of the charge states

Initially, we recorded full scan data for all non-labeled peptides using MS^E mode and the high-resolution instrument to determine the predominant charge states and fragmentation pattern of the peptides. To further study and determine the predominant charge states at the quantitative level, we monitored each peptide at m/z values representing singly, doubly and triply charged peptides. This was done in SIR data acquisition mode on UPLC-Xevo-TQ-S system using non-labeled synthetic peptides. All peptides were observed as a mixture of doubly and triply charged species. In addition to doubly and triply charged ions, Bet v 2 peptide was also detected as singly charged ion. However, the highest intensity values were observed for doubly charged peptides. This data is in agreement with

charged states observed using MS^E data acquisition mode. Taken together, all peptides detected in doubly charged state were selected as precursor ions.

Selecting fragment ions and optimizing cone voltage and collision energy

We recorded all fragment ions (b and y series ions) for doubly charged precursor ion using non-labeled peptides and the birch pollen extract. The data for fragment ions of non-labeled peptides and natural peptides need to be in concordance in order to be considered as suitable MRM transitions. We selected transitions for peptides based on the following criteria: 1) four consecutive transitions whenever possible; 2) signal intensities of MRM transitions; 3) absence of interferences. We optimized the cone voltage (CoV) and collision energies (CE) using non-labeled peptides. Optimal CoVs were determined by monitoring the difference between signal intensities of doubly charged precursor ion in increments of ten Volts between 10 and 80 Volts using the SIR data acquisition mode. Following the CoV optimization, we determined the CEs for each peptide by monitoring signal intensities of fragment ions in increments of three between 10 and 37 V using MRM data acquisition mode. Table S-4 summarizes the optimized MS parameters for thirteen target peptide sequences with their respective precursor and product ions m/z values, cone voltages, and collision energies.

Following the selection and optimization of MRM transitions, we monitored all MRM transitions of isotope-labeled (TQL) peptides for the presence of unspecific signals that could arise from other components that are present in birch pollen extract. The MRM parameter for all peptides are listed above in table S-5

Calibration design

The matching concentrations (1X) of isotope-labeled peptides to natural peptides were determined in order to set up calibration curves spanning a 4 fold concentration range above and below the matching concentrations. Therefore, the matching concentration is the concentration of the isotope-labeled peptide that gives a similar intensity as the natural counterpart peptide in the extract. Moreover, the matching concentration defines the middle concentration that is used to set up a five point calibration curve. The 4 fold calibration range for characterization of the assay was chosen, because variations in protein amounts due to the extraction and digestion of the birch pollen extract were not expected to exceed 4 fold values. The signal intensities for each tested value of isotope-labeled peptides gave a linear response and the corresponding linear equations were used to determine the approximate amounts of natural peptides in the extract. The amounts determined for natural peptides were rounded to the nearest integer and these values were set as the approximate matching concentration of isotope-labeled peptides. Therefore, we used five-point external calibration curve with two fold dilution between each concentration point i.e. 4X, 2X, 1X, 0.5X, and 0.25X of isotope-labeled peptides for quantification of proteins in birch pollen extract.