

Article

Characterization of mustard 2S albumin allergens by bottom-up, middle-down and top-down proteomics: A consensus set of isoforms of Sin a 1.

Marlene Hummel, Tina Wigger, and Jens Brockmeyer

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/pr5012262 • Publication Date (Web): 09 Feb 2015

Downloaded from <http://pubs.acs.org> on February 15, 2015

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications
High quality. High impact.

Journal of Proteome Research is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036
Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Characterization of mustard 2S albumin allergens by
bottom-up, middle-down and top-down proteomics:
A consensus set of isoforms of Sin a 1.

*Marlene Hummel, Tina Wigger, and Jens Brockmeyer**

Department of Chemistry, Institute of Food Chemistry, Westfälische Wilhelms-Universität
Münster, Corrensstr. 45, 48149 Münster, Germany

KEYWORDS

allergens, mass spectrometry, mustard, LC-MS, protein isoforms, protein modification, Sin a 1

ABSTRACT

The mustard allergen Sin a 1 belongs to the 2S-albumin family of seed storage proteins. Due to its high abundance in mustard seeds and the potential to elicit severe allergic reactions, Sin a 1 is considered a major allergen in mustard. Eight Sin a 1 isoforms have been identified using DNA cloning and sequencing and we aim in this study to thoroughly investigate sequence heterogeneity using a novel combination of bottom-up, middle-down and top-down proteomics. The characterization of purified Sin a 1 extract shows that sequence diversity is far more pronounced than previously assumed. We identified in total 24 sequence polymorphisms including 17 yet unpublished point mutations. Using middle-down and top-down approaches on the subunit and protein level of Sin a 1 we were able to detect eight consensus isoforms of Sin a 1 (including four novel isoforms), that we detect in the majority of the four different mustard samples included in this study. In addition, we provide for the first time data on relative abundance of the main Sin a 1 isoforms and identify phytic acid as a potential ligand of Sin a 1. Together, these data can form the basis for a more detailed investigation of the effect of Sin a 1 polymorphic sites on allergenicity of isoforms.

INTRODUCTION

Mustard seeds are derived from species in the genera *Brassica* and *Sinapis*, namely the mild yellow mustard (*Sinapis alba/Brassica hirta*), the brown oriental mustard (*Brassica juncea*) and the pungent black mustard (*Brassica nigra/Sinapis nigra*). The seeds are used worldwide as food ingredients, to season dishes, and are processed to table mustard. Due to its widespread use and high allergenic potential, mustard is one of the most important allergenic seed spices¹⁻³. Furthermore, clinical studies indicate that in regions with high consumption mustard is among the most important food allergens for children³ and deserves special attention as a potent “hidden allergen” in convenience products⁴.

Among the mustard seed storage proteins the 2S albumins are highly abundant and have been identified as major mustard allergens with significant higher response in allergy tests compared to other storage proteins⁵⁻⁷. 2S albumins belong to the prolamine superfamily and are widely distributed in seeds and tree nuts⁸. In general, they are considered major allergens in different foods, indicating their role as panallergen^{6,9}. The 2S are synthesized from a single ~13 kDa precursor protein that is subjected to extensive posttranslational proteolytic processing. The mature protein consists of two subunits of approximately 9-10 kDa (large subunit) and 3-4 kDa (small subunit), respectively and the highly conserved structure is stabilized by eight disulfide-bridged cysteine residues^{5,6,10}. 2S albumins are in general encoded by multiple genes, resulting in different isoforms that may show considerable differences in their structure and allergenicity¹¹.

For Sin a 1, the 2S allergen in yellow mustard, seven isoforms (Table 1) have been identified using recombinant cloning and DNA sequencing techniques^{7,12-14}. It has to be noted, that these studies did not aim to thoroughly investigate isoform pattern of Sin a 1 but rather provided the

initial characterization of this allergen. A comprehensive structural analysis of Sin a 1 polymorphisms on protein level *in planta* including abundance, posttranslational modifications and potential formation of complexes is still lacking. We have therefore used a combination of bottom-up, middle-down and top-down proteomics to first identify sequence polymorphisms and posttranslational modifications from the naturally occurring isoforms and to delineate the combination of polymorphisms on subunit and on protein level. In the bottom-up approach, proteins are proteolytically digested to a complex set of peptides which can be used to efficiently identify position and structure of sequence modifications by MS and MS/MS. The information how modifications are combined in the different isoforms is however lost in this approach. We therefore investigated the small and large subunits after reduction and alkylation (middle-down approach) and the intact Sin a 1 isoforms (top-down) by HRMS. This combination of three proteomic strategies allows us to identify for the first time the most abundant Sin a 1 isoforms in yellow mustard, characterize posttranslational modifications which are not detectable using DNA-based techniques and to identify potential interaction partners of Sin a 1 in mustard.

MATERIALS AND METHODS

Extraction of mustard proteins from mustard seeds

Yellow mustard seeds from 4 different brands (Fuchs Gewürze GmbH, Dissen; Hartkorn Gewürzmühle GmbH, Koblenz; Davert GmbH, Ascheberg; Ostmann Gewürze GmbH, Dissen a. T. W.) were purchased from local supermarkets and each sample was milled to a fine powder in a nitrogen-cooled rotor-mill (Pulverisette, Fritsch, Idar-Oberstein, Germany). The powder was defatted with *n*-Pentane (1 g : 10 mL) for two hours under continuous shaking, sieved and air-dried overnight. Proteins were extracted by shaking for 2 hours in a 5-fold excess of sodium

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

acetate buffer (100 mM, pH 3.5) The slurry was centrifuged at 9000 x g for 30 min at 23 °C and filtered through 0,22 µm pore size bottle-top filters (Steritop-GP, 0.22 µm, polyethersulfone, 150 mL, 45 mm, Merck Millipore, Darmstadt, Germany). The resulting protein extract was stored at -20 °C. In total, nine biological replicates were extracted and included in this study.

Size exclusion chromatography

To isolate the 2S albumin fraction of the mustard protein extract size exclusion chromatography (SEC) was applied using an ÄKTA™ purifier FPLC system (GE Healthcare, Munich, Germany) and a Superdex 200 10/300 GL column (GE Healthcare, Munich, Germany). Chromatography was performed using 100 mM Tris, pH 8.0 at a flowrate of 200 µL/min. Fractions were collected, precipitated by adding a 5-fold excess of ice-cold acetone and analyzed by SDS-PAGE. Identity and purity of 2S albumins was confirmed using tryptic in-solution digest followed by LC-HRMS analysis of resulting peptides. UniProt sequence AC P15322 was used for the confirmation of Sin a 1.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Samples were separated by Tricine SDS-PAGE (T = 16 %, C = 2,6 %) with a Mini-Protean cell (Bio-Rad, Munich, Germany) according to the protocol from Schaeffer et al.¹⁵ with minor modifications. The acetone-precipitated protein pellet was resuspended in 1x SDS-PAGE sample buffer and incubated at 95 °C for 5 min. Gels were run for about 90 min at 150 V and stained with Coomassie Blue staining solution.

Tryptic digestion

Acetone-precipitated pellets (containing about 1 mg protein) were resuspended in 100 μ L of 6 M urea solution and subjected to tryptic digest following a standard protocol with minor modifications¹⁶. Briefly, 1 mg of protein dissolved in 100 μ L urea solution was reduced with 200 mM dithiothreitol, alkylated with 100 mM iodoacetamide and diluted with 100 mM Tris-HCl buffer to a resulting urea concentration of 0.6 M. Digestion was performed under slow shaking overnight at 37 °C after adding 20 μ g of Serva sequencing grade trypsin (Serva, Heidelberg, Germany). The peptide solution was desalted using Strata-X 33u polymeric reversed phase (RP) cartridges filled with 30 mg/mL C18 material. Cartridges were activated with methanol (MeOH), equilibrated with 1% formic acid (FA), loaded with peptide-solution and washed (5% MeOH/1% FA). Peptides were eluted with 70% MeOH/1% FA. The solvent was removed at 40 °C under continuous nitrogen flow and peptides were resuspended in solvent A (3% ACN/1% FA) prior to MS-analysis.

Bottom-up MS

The HPLC-HRMS/MS system set up was as follows. All samples were injected by an autosampler of an Accela HPLC system (Thermo Scientific) and applied to an AccucoreTM C18 column (Thermo Scientific, Dreieich, Germany) (2,6 μ m Particle Size, 150 x 2.1 mm). The gradient of the 120 min chromatographic run was as follows: 3% B for 6 min, 3% to 12% B in 6 min, from 12% to 35% B in 79 min, 35% to 60% B in 9 min, 60% B for 8 min, 60% to 3% B in 2 min, 3% B for 10 min (A = 1% FA in H₂O, B = 1% FA in ACN). Flow rate was 250 μ L/min.

Mass spectrometrical analysis was performed on an LTQ Orbitrap XLTM Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Scientific) in positive ion mode. The mass spectrometer was run in full MS mode and data dependent mode (DDMS5). For the DDMS5 method, survey full scan MS spectra (m/z 300-1500) were acquired in the Orbitrap ($r = 30.000$). The five most

intense peptide ions were sequentially isolated and fragmented in the HCD cell (normalized collision energy of 35% and activation time of 100 ms). Resulting fragments were detected at $r = 7500$. Dynamic exclusion was set to 60 s. Ions with unrecognized charge states were rejected.

Mass spectrometric conditions were:

source type: HESI; capillary temperature: 275 °C; vaporizer temperature: 350 °C; sheath gas flow: 40 (arbitrary units); auxiliary gas flow: 20 (arbitrary units); sweep gas flow: 5 (arbitrary units); source voltage: 3.5 kV and tube lens: 130 V.

Database search was performed with PEAKS software v7 (Bioinform, Waterloo, Canada) using the enhanced target-decoy method, (“decoy fusion”) for FDR estimation and result validation^{17,18}.

Database searching was performed against a subset (*Sinapis, Brassica*) of the UniProt KB database (870 search entries, no common contaminants, UniProt KB date of download: 2014-08-08). Search parameters were as follows: (a) trypsin as specific enzyme (one missed cleavage allowed); (b) precursor mass error tolerance 15 ppm; (c) fragment mass error tolerance 0.1 Da; (d) fixed modification: carbamidomethylation of Cys and variable modification: oxidation of Met, allowing a maximum of variable PTM per peptide of five. Following database search the PEAKS *de novo* sequencing and the PEAKS Spider algorithm were applied to identify peptides bearing polymorphisms (insertion, deletion, mutation, PTM) compared to published *Sin* a 1 sequences¹⁸. Only polymorphisms that were identified at least in 4 from 9 biological replicates were included in this study. In addition, the minimal -10lgP peptide score was set to 47 (respective spectrum shown in supplementary) and all spectra bearing potential polymorphisms were inspected manually to ensure unambiguous identification of sequence polymorphisms.

Middle-down and top-down approach

SEC fractions containing 2S albumins were precipitated as described above, resuspended in 50 mM ammonium hydrogen carbonate buffer and directly used for top-down measurements. For the analysis of subunits (middle-down), proteins (1-5 mg/mL) were reduced by adding 10 μ L of 200 mM dithiothreitol for 30 minutes at 56 °C and alkylated by addition of 10 μ L 200 mM iodoacetamide for 30 minutes in the absence of light. Reduced and alkylated samples were diluted in a ratio of 1:4 – 1:100 with 5 % (v/v) ACN / 0.1 % FA prior to analysis. Six biological replicates were used for top-down and middle-down analysis, respectively.

Middle-down and top-down MS

The HPLC-HRMS system set up was as follows. All samples were injected by an autosampler of an Accela HPLC system (Thermo Scientific) and applied to an AccucoreTM C4 column (Thermo Scientific, 2.6 μ m particle size, 100 x 3.0 mm). The gradient of the 28 min chromatographic run was as follows: 5% B for 1 min, 5% to 40% B in 16 min, 40% to 95% B in 3 min, 95% B for 2 min, 95% to 5% B in 1 min, 5% B for 5 min (A = 0.1% FA in H₂O, B = 0.1% FA in ACN) at a flow rate of 400 μ L/min. HPLC was coupled to an LTQ Orbitrap XLTM (Thermo Scientific) in positive ion mode. The mass spectrometer was run in full MS mode. Full MS spectra (m/z 500-2000) were acquired either with a resolution of $r = 30.000$ or $r = 60.000$. Mass spectrometric conditions were:

source type: HESI; capillary temperature: 225 °C; vaporizer temperature: 350 °C; sheath gas flow: 40 (arbitrary units); auxiliary gas flow: 20 (arbitrary units); sweep gas flow: 10 (arbitrary units); source voltage: 3.5 kV and tube lens: 135 V.

Determination of monoisotopic mass for middle-down and top-down experiments

Charge deconvolution of middle-down and top-down results was performed manually using Excel (Microsoft) according to the formula

$$M = (m/z - 1.0072765) \cdot z$$

where M is the deconvoluted mass, m/z is the mass to charge ratio, 1.0072765 is the monoisotopic mass of hydrogen and z is the charge state.

To ensure highest mass accuracy we developed an approach for the determination of monoisotopic masses instead of using average mass data. For middle-down and top-down data the identification of the monoisotopic mass is a problem, as the signal intensity is small compared to the most abundant isotopologs. In addition, we observed that available software-based approaches tend to shuffle isotopolog signals from different 2S isoforms leading to scrambling of mass data. We therefore used the isotope simulation tool included in the Xcalibur 2.0.7 SP1 software package (Thermo Scientific) to create the theoretical isotopic distribution of each subunit, and the intact protein of the Sin a 1 prototype (AC P15322), respectively.

Based on this theoretical pattern the mass difference between the most abundant isotopolog and the monoisotopic signal was calculated. Given the same charge state, this result can be transferred to the experimentally observed most abundant isotopolog to identify the mass of the experimentally observed monoisotopic signal. This mathematical correlation can be summarized in the formulas

$$M(\text{ma})_{\text{theo}} - M(\text{mi})_{\text{theo}} = \Delta M = M(\text{ma})_{\text{exp}} - M(\text{mi})_{\text{exp}}$$

$$M(mi)_{exp} = M(ma)_{exp} - (M(ma)_{theo} - M(mi)_{theo})$$

where $M(ma)_{theo}$ is the most abundant mass of the theoretical spectrum, $M(mi)_{theo}$ is the monoisotopic mass of the theoretical spectrum, $M(ma)_{exp}$ is the most abundant mass of the experimental spectrum and $M(mi)_{exp}$ is the monoisotopic mass of the experimental spectrum. For this approach it is one of the most important prerequisites that marginal changes in the amino acid sequences, as observed for isoforms, do not influence the isotopic distribution.

The determination of accurate protein monoisotopic mass by use of the signal of the most abundant isotopolog has been described by several groups^{19–21}. However, the comparison of experimental data with the theoretical isotopic distribution of a closely related isoform is a novel and promising approach that has, to the best of our knowledge, not been described before.

Quality criteria for data evaluation

To ensure highest reliability, we analyzed 6 samples for the small and the large subunit and the intact protein, respectively. For each sample at least three different charge states ($z = +7/+6/+5$ for small subunit, $z = +10/+9/+8$ for large subunit and $z = +11/+10/+9$ for intact protein) were analyzed and adducts of phytic acid were taken into account. All signals were inspected manually at least twice and only mass signals that were identified at least three times were included in the further analysis. Signals were de-isotoped, deconvoluted and mass signals of phytic acid adducts corrected to the free monoisotopic mass. In the respective mass window (14302.336–14828.065 Da for top-down, 10018.976–10354.109 Da for the large subunit, and 4217.070–4532.195 Da for the small subunit) signals were grouped when the mass difference was below 0.03 Da. Masses differing more than 0.5 Da were considered as potential isoforms. For the grouped masses the average mean was calculated and used as the monoisotopic mass of a

potential isoform in further analysis. The average standard deviation of all grouped masses was below 0.01 Da. Together, we observed 15, 22, and 17 monoisotopic masses for top-down, large subunit, and small subunit, respectively.

RESULTS

For the comprehensive characterization of Sin a 1 isoforms, we analyzed yellow mustard samples from 4 different brands. The 2S fraction of protein extracts was isolated using size exclusion chromatography to reduce sample complexity. Purification of the 2S fraction is shown in Fig. 1, and the identity of the isolated proteins as 2S was confirmed by in-solution digest and LC-HRMS resulting in an overall sequence coverage of 95 % for the small subunit (n=4) and 78-91 % coverage for the large subunit (n=9) (data not shown). Notably, no peptides specific for Bra j 1 (*Brassica juncea*, P80207), or 2S storage proteins from *Brassica nigra* or *Brassica juncea* (Q42491, Q42413, respectively) were identified in the isolated 2S fraction, demonstrating that the analyzed samples only contained yellow mustard and the respective 2S albumin Sin a 1.

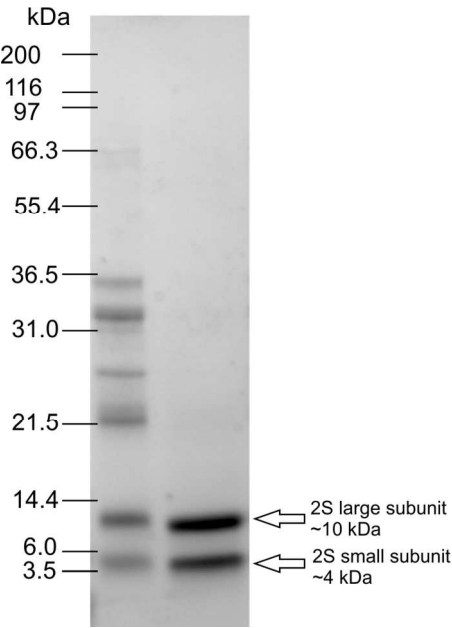


Figure 1. Tricine SDS-PAGE (T = 16 %, C = 2.6 %) of mustard extract. Lane 1 = crude extract, lane 2 = low molecular weight fraction (Sin a 1) purified by size exclusion chromatography.

We next applied three complementary proteomics strategies to identify polymorphisms and assign them to the different isoforms. The experimental workflow is displayed in Fig. 2.

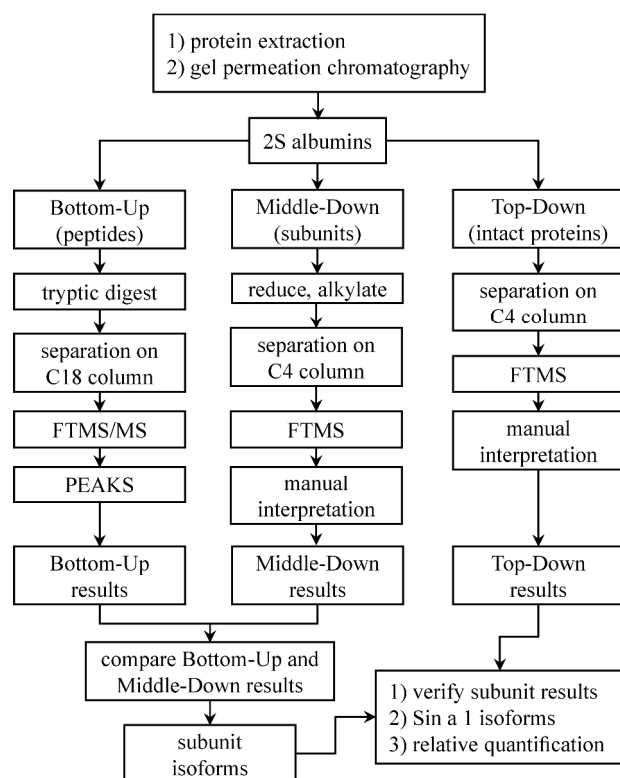


Figure 2. Experimental workflow of the combined approach using bottom-up, middle-down and top-down proteomics.

The 2S fraction was first analyzed on peptide level (bottom-up) following tryptic digest and sequence polymorphisms within peptides were identified using the PEAKS de novo and Spider algorithms. These results display the sum of polymorphisms identified in the analyzed samples.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

However, information about the combination of polymorphisms on subunit as well as on protein level is lost after tryptic digest. We therefore performed in addition LC-HRMS analysis of the 2S fraction on subunit level (middle-down). Reduction and alkylation resulted in the formation of the ~4 kDa small subunit and the ~10 kDa large subunit that were separated on a C4 column and analyzed by HRMS. These data allow mapping of identified polymorphisms on the different small and large subunits resulting in the identification of novel and known subunit sequences. Again, these data do not allow to delineate the pairing of subunits to the mature 2S isoforms. To this end, we analyzed the intact 2 S albumins using HRMS to identify isoforms from the combination of bottom-up, middle down and top-down data.

Bottom-up approach

To ensure confident identification of polymorphisms, we subjected the 2S fractions from 9 biological replicates derived from 4 different mustard samples to the bottom-up analysis. Polymorphisms that were detected in at least 4 replicates were included in this study. Together, we observed 24 point mutations. Eight mutations occurred in the small subunit and 16 in the large subunit of Sin a 1 when compared to the reference sequence (AC P15322) (Fig. 3, Table 2). Seven point mutations detected in this study are already described in known Sin a 1 isoforms which correspond to 70 % of all described mutations and demonstrates overall excellent recovery. In addition, 2 published mutations were detected but did not fulfill the inclusion criteria of our study (S36G was detected twice, K138Q once). On the other hand, we identified 17 novel mutations (70 % of all mutations), indicating a pronounced (and currently underestimated) degree of heterogeneity of Sin a 1. Notably, we did not observe C-terminal clipping of the small and large subunit which has been previously described for some 2S albumins from other plant species such as sesame²². We were able to confirm clipping for

sesame 2S albumins, demonstrating that we are in principle able to detect this modification with our approach, (data to be published elsewhere). This further underlines the lack of C-terminal clipping for Sin a 1. The position of the linker peptide that is excised between small and large subunit of the Sin a 1 precursor during posttranslational modification was so far only predicted using bioinformatics and has been confirmed in this study. Polymorphisms in the Sin a 1 sequence are especially pronounced in the hypervariable region which is assumed to be located between K⁹² and V¹¹²⁶. In this part of the large subunit, we also observed several accumulated point mutations where more than one amino acid was changed per tryptic peptide (Fig. 3, below prototype sequence). Notably, the only IgE-epitope of Sin a 1 identified yet overlaps with the hypervariable region⁶, indicating pronounced heterogeneity also in the epitope region of Sin a 1.

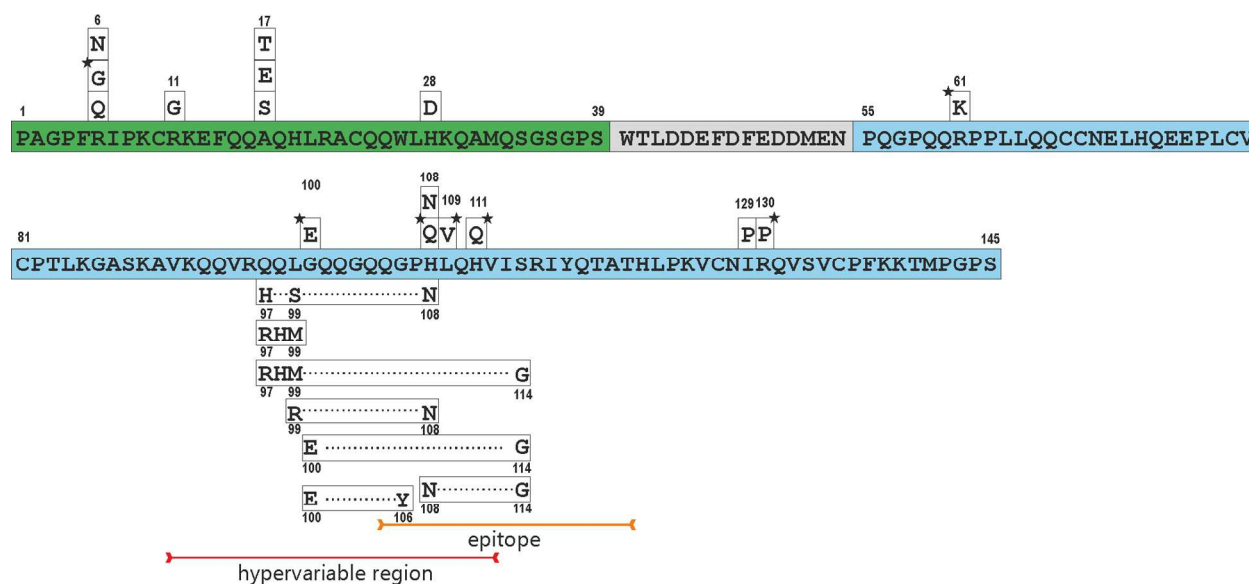


Figure 3. Mapping of sequence polymorphisms of Sin a 1. Modifications were identified after tryptic in-solution digest of the mustard 2S fraction. Sequence of the small subunit is highlighted in green, sequence of the large subunit is displayed in blue. Linker peptide (highlighted in grey) is excised during posttranslational processing. Modifications marked by asterisk are already published.

Middle-down approach

To allow assigning observed point mutations to Sin a 1 on subunit level, we applied direct HRMS analysis of reduced and alkylated small and large subunits after chromatographic separation. The total ion chromatogram shows two broad peaks representing the small and large subunit of Sin a 1 (Fig. 4a). Overlay of extracted ion chromatograms in the total ion chromatogram indicates partial chromatographic separation of subunit isoforms (Fig. 4a). In the corresponding MS-spectra signals of subunit isoforms were observed in different charge states and consequently at least three different charge states were included in the determination of exact isoform masses. Similarity of subunit isoform pattern at different charge states is exemplified for the large subunit in Fig. 4b and a more detailed view of charge state $z = +9$ is given. In addition, we observed adducts of both subunits with phytic acid (Fig. 5) which were also taken into account for the identification of subunit isoforms.

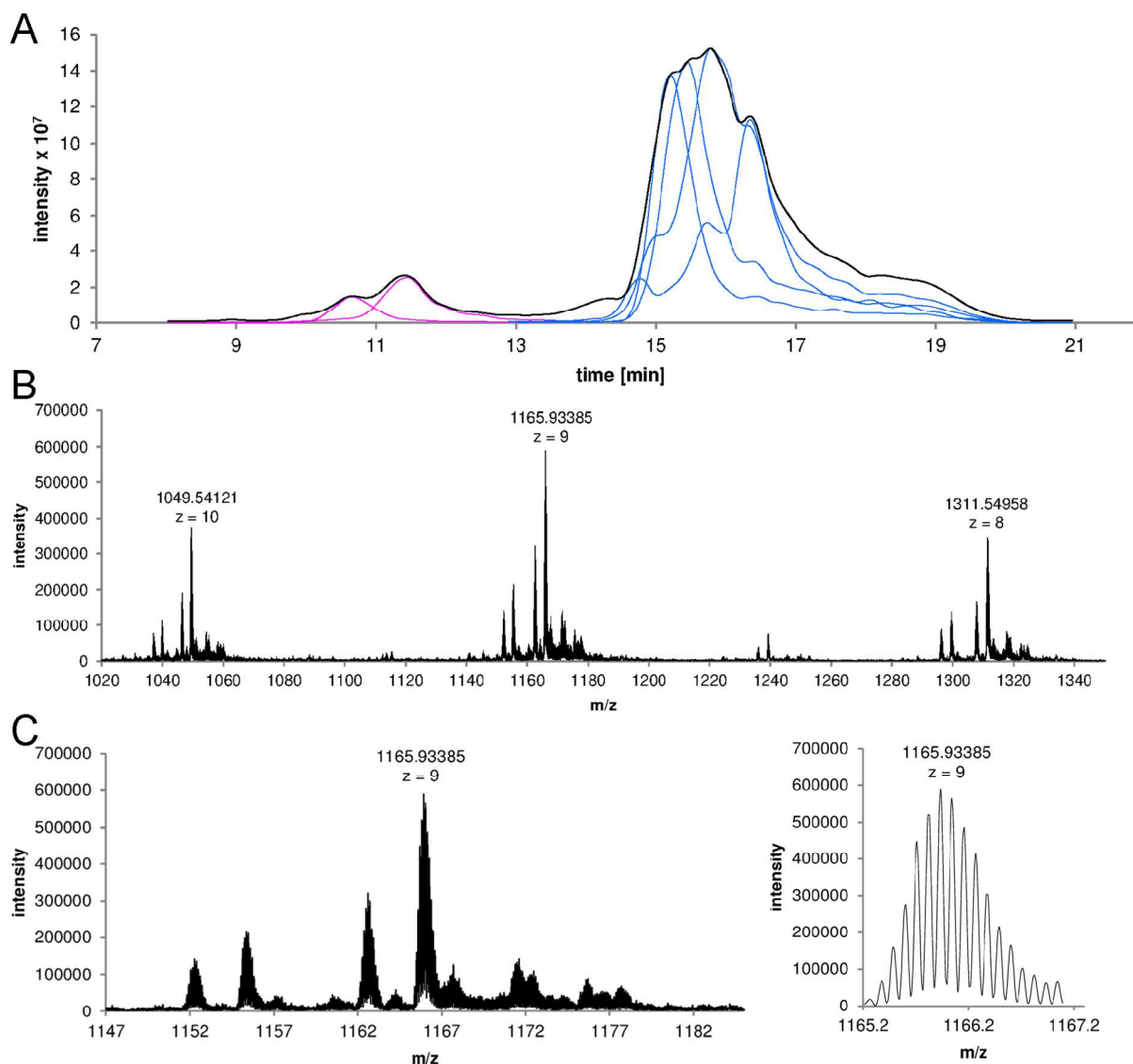


Figure 4. Middle-down analysis of Sin a 1. a. Separation of subunits by HPLC after reduction and alkylation. Ion traces of the small subunits are given in purple and ion traces of the large subunits are displayed in blue. (see Fig 4b). b. ESI-HRMS of the large subunit. Three charge states ($z = 8-10$) are displayed showing signal heterogeneity of isoforms of the large subunit. c. left part: detailed view of charge state $z = 9$ showing isoform distribution. Right part: Isotopic pattern of one isoform in charge state $z = 9$.

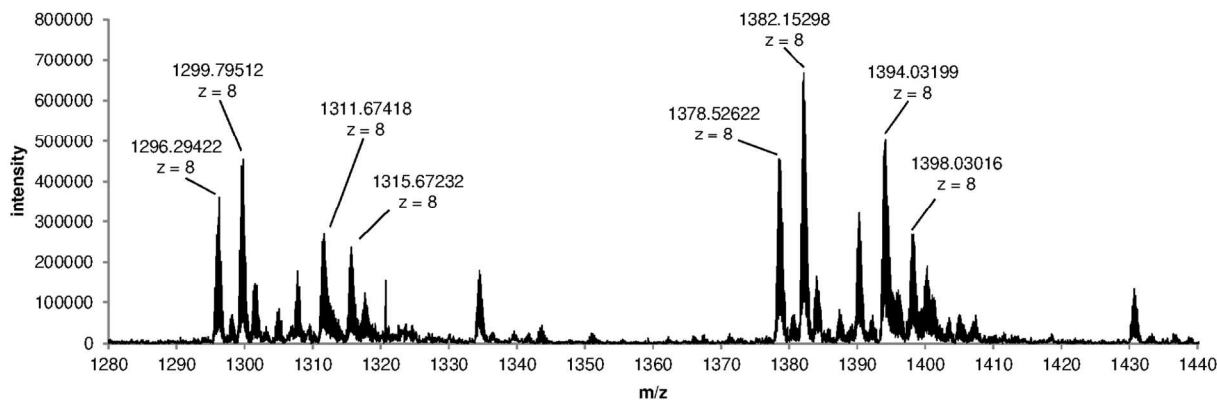


Figure 5. Phytic acid adducts observed of the large subunit in ESI-HRMS ($z = 8$). Corresponding subunits without phytic acid (m/z 1296 - 1334) and with one molecule of phytic acid (right hand side, m/z 1378 - 1430).

De-isotoped and deconvoluted monoisotopic masses that were detected at least three times with $\sigma < 0.01$ in the 6 biological replicates were included in the comparative analysis of middle-down data. We obtained 327 and 328 mass signals for the small and the large subunit, respectively that also included adducts with phytic acid (up to two molecules per large subunit and one molecule per small subunit). Mass signals of adducts were converted to the monoisotopic mass of the free subunit and signals were grouped when mass differences were below 0.04 Da. This resulted in the identification of 17 potential subunit masses (detected in average ~ 20 times in the 6 replicates) for the small subunit (4217.070 Da - 4532.195 Da) and 22 potential subunit masses for the large subunit (10018.976 Da - 10354.109 Da). Next, all identified middle-down masses were compared with the theoretical mass of all possible combinations of bottom-up identified polymorphisms.

Taking the resulting error as well as a minimal signal to noise ratio of 3:1 into account, we identified four isoforms of the small subunit with a maximum error of 13.25 ppm. Both isoforms

described in literature were confirmed and two novel isoforms were identified (Table 3).

Applying the same criteria as mentioned above, four isoforms of the large subunit have been identified (maximum error of 15.34 ppm) that are all published in literature (Table 4).

Top-down results and combination of data

To elucidate how identified subunits pair on protein level to the Sin a 1 isoforms (and to further confirm identified polymorphisms), we performed top-down experiments. Again, only masses observed at least three times in 6 biological replicates were included in the further analysis. We observed 304 mass signals (including adducts of phytic acid in molar ratio of (1:1 to 1:3) and reduced signal complexity as described for middle down data resulting in 15 mass groups of potential isoforms. Possible subunit pairings were compared with observed exact masses of top-down experiments. Using this approach, we identified seven isoforms (Fig.6, Table 5) with a maximum error of 83.92 ppm and a signal to noise ratio of not less than 3:1. 50 % of known isoforms (4 of 8) were detected in the analyzed samples and four novel isoforms were identified. The four new identified isoforms were designated Sin a 1 isoform 9 to isoform 12, respectively. All identified isoforms resulting from the combination of bottom-up, middle-down and top-down approaches are summarized in Figure 6. Based on our results we assessed the relative quantitative distribution of the identified isoforms using signal intensity of the most abundant isotopolog of each isoform. To this end, all samples have been included and the mean averaged is given as a percentage of the most abundant isoform (Table 6).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Prototype	1	11	21	31	55	61	71	81
Sin a 1.0105**	PAGPFRIPKC	RKEFQQAQHL	RACQQWLHKQ	AMQSGSGPS	(-)PQGPQQ	RPPLLQQCCN	ELHQEEPLCV	CPTLKGASKA
Sin a 1.0106	-----G-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1.0107**	-----G-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1.0108	-----G-----	-----	-----	-----	(-)-----	K-----	-----	-----
Sin a 1 isoform 9	-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 10	-----	-----S-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 11	-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 12 vs1	-----	G-----S-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 12 vs2	-----G-----	-----S-----	-----	-----	(-)-----	-----	-----	-----

Prototype	91	101	111	121	131	141
Sin a 1.0105**	VKQQVRQQLG	QQGQQGPHLQ	HVISRIYQTA	THLPKVCNIR	QVSVCPFKKT	MPGPS
Sin a 1.0106	-----E-----	-----	-----	-----P-----	-----	-----
Sin a 1.0107**	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1.0108	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 9	-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 10	-----E-----	-----	-----	-----P-----	-----	-----
Sin a 1 isoform 11	-----E-----	-----	Q-----	-----	-----	-----
Sin a 1 isoform 12 vs1	-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 12 vs2	-----	-----QV-----	-----	-----P-----	-----	-----

Figure 6. Published and unpublished isoforms of Sin a 1 based on the prototype sequence (Sin a 1.0102) published in UniProtKB. Seven isoforms of Sin a 1 have already been published and four isoforms (Sin a 1.0105 to Sin a 1.0108) were confirmed in this study. Four yet unpublished isoforms were identified in this study and designated Sin a 1, isoform 9 to isoform 12, respectively. **Isoform 1.0105 and 1.0107 differ only in the linker peptide that is removed during processing.

DISCUSSION

In the current paper we present the identification of isoforms of the mustard allergen Sin a 1 by a novel approach combining high resolution mass spectrometry on peptide-, subunit-, and protein level after fractionation directly from plant material. Isoforms of Sin a 1 have hitherto only been identified using cloning and sequencing of mustard cDNA^{7,13,14} and were characterized on protein level after recombinant expression. This strategy has huge advantages when single target proteins are purified from complex matrices e.g. for functional characterization as in the above mentioned studies. However, essential information is only obtained when the proteins are directly analyzed from natural sources. The 2S albumins are in general subjected to pronounced posttranslational processing^{7,13,14} that is not detectable using DNA-based techniques. In addition,

the relative abundance of (posttranslationally modified) isoforms can only be determined by direct analysis of plant material as well as the identification of potential physiological ligands in the plant.

The bottom-up approach on peptide level is best suited to identify sites of posttranslational modifications and sequence polymorphisms. Using this approach, the theoretical clipping site between the small and large Sin a 1 subunit was confirmed in this study and we demonstrate that this 2S albumin is not further modified by “ragged clipping” as described for other 2S²². One of our major findings is that Sin a 1 polymorphisms are far more frequent than previously estimated. We have identified in total 24 polymorphic sites, including 17 yet unpublished point mutations. It has to be noted, that we only included point mutations that occurred at least in 4 out of 9 replicates from four different mustard samples. Sequence polymorphisms described in this study are therefore “consensus polymorphisms” occurring in the majority of the four different mustard samples. We detected further point mutations (including two published polymorphisms) that occur only in single mustard samples, indicating that also breed-specific Sin a 1 isoforms do exist.

As we analyzed polymorphisms directly in Sin a 1 fractions from plant material, it is necessary to assign identified polymorphic sites to the different small and large subunits and subsequently to Sin a 1 isoforms. This approach further enhances reliability of identified Sin a 1 isoforms, as polymorphisms are detected on peptide, subunit and protein level. We observed however that the confident determination of monoisotopic subunit and protein masses works best with the more abundant signals and consequently not all polymorphisms identified in bottom-up experiments can be assigned to the respective subunits in middle-down experiments. We were able to assign ~42 % (10 of 24) of polymorphisms on subunit level and identified 8 Sin a 1 “consensus

isoforms” on protein level. Taking into account the overall analytical strategy, these isoforms occur frequently and with high abundance in the mustard samples included in our study. Notably, our data demonstrate that further yet unidentified and less abundant Sin a 1 isoforms exist, that also might contribute to mustard allergenicity.

Isoforms have by definition a sequence homology of more than 90 %²³ and can show considerable differences in their allergenic potential. Several studies reported that small changes in the primary structure can significantly alter IgE binding^{24–27}. Mutational analysis of the peanut allergen Ara h 3 showed that single amino acid changes diminished IgE binding by 35-85 %²⁵, leading to a ~25-fold increase of the required amount of modified Ara h 3 in the IgE-inhibition assay compared to wild-type Ara h 3^{24,25}. Similarly, Ferreira et al. investigated the effect of point mutations on the IgE-binding capacity of Bet v 1, the major birch pollen allergen by site-directed mutagenesis²⁶. Exchange of six amino acids led to ~20-fold reduced IgE-binding of mutant Bet v 1 compared to the wild-type protein in the IgE-inhibition assay²⁶.

Early studies on epitope mapping of 2S albumins propose several linear IgE-binding epitopes which are all located in the “hypervariable region”^{6,28–31}. This region has been identified between two α -helices in several 2S albumins and is located between Lys⁹²-Val¹¹² in Sin a 1⁶. Taking a closer look at the identified polymorphisms (Fig. 3) in our study, this region clearly is a hot-spot for sequence modification and we hypothesize that modifications in this region might have a considerable effect on allergenicity of isoforms. Future studies will therefore address the identification of less abundant Sin a 1 isoforms with polymorphisms in the hypervariable region using more sensitive instrumentation and further pre-fractionation and enrichment steps as well as functional investigation of allergenic potential of the Sin a 1 isoforms.

1
2
3 In conclusion, we present here a consensus panel of Sin a 1 isoforms that we detected in the
4 majority of four different mustard samples and provide evidence that the heterogeneity of the
5 mustard allergen Sin a 1 is much more pronounced than previously expected. In addition, we
6 provide for the first time data on relative abundance of the main Sin a 1 isoforms and identify
7 phytic acid as a potential ligand of Sin a 1. Complex formation of phytic acid with proteins has
8 been described for various seed storage proteins and it has to be noted that bioavailability and
9 digestibility can be affected by protein complexes with phytic acids³². For peanut allergens the
10 interaction with phytic acid resulted in reduction of solubility and IgE-reactivity, indicating that
11 this complex might reduce allergenicity^{33,34}. Together, these data can form the basis for a more
12 thorough investigation of the effect of Sin a 1 polymorphic sites on allergenicity of isoforms.
13
14 From the analytical point of view, we present to the best of our knowledge for the first time the
15 combination of bottom-up, middle-down and top-down strategies for the direct analysis of
16 protein isoforms in complex matrices.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TABLES.

Table 1. Published isoforms of mustard 2S allergen Sin a 1. * Isoform 1.0105 and 1.0107 differ only in the linker peptide that is removed during processing.

isoform	Polymorphism (amino acid level)	Ref.
Sin a 1	incomplete sequence (AA 109-111 missing), obsolete	7
Sin a 1.0102 (SA2S1)	prototype sequence (Uniprot AC P15322)	13
Sin a 1.0101	R6G, G100E, R130P	7
Sin a 1.0103 (SA2S2)	S36G, G100E, H111Q	13
Sin a 1.0104 (SIN1)	R6G, G100E, K125R, K138Q	14
Sin a 1.0105 (SIN2*)	G100E, R130P	14
Sin a 1.0106 (SIN3)	R6G, H108Q, L109V, R130P	14
Sin a 1.0107 (SIN4*)	G100E, R130P	14
Sin a 1.0108 (SIN5)	R6G, R61K, H108Q, L109V, R130P	14

Table 2. Mutations detected by HRMS in the Sin a 1 fraction by bottom-up approach after tryptic digestion. For mutations already described in literature, isoform and reference are given, novel mutations are given in italics.

amino acid mutation	Isoform and reference (if published)
<i>R6Q</i> , R6G, <i>R6N</i>	1.0101 ⁷ , 1.0104 ¹⁴ , 1.0106 ¹⁴ , 1.0108 ¹⁴
<i>R11G</i>	
<i>A17S</i> , <i>A17E</i> , <i>A17T</i>	
<i>H28D</i>	
S36G	1.0103 ¹³
R61K	1.0108 ¹⁴
<i>Q97H</i> , <i>Q97R</i>	
<i>Q98H</i>	
<i>L99S</i> , <i>L99M</i> , <i>L99R</i>	
G100E	1.0101 ⁷ , 1.0103 ¹³ , 1.0104 ¹⁴ , 1.0105 ¹⁴ , 1.0107 ¹⁴
<i>G106Y</i>	
H108Q, <i>H108N</i>	1.0106 ¹⁴ , 1.0108 ¹⁴
L109V	1.0106 ¹⁴ , 1.0108 ¹⁴
H111Q	1.0103 ¹⁴
<i>S114G</i> , <i>S114Y</i>	
K125R	1.0104 ¹⁴
<i>I129V</i> , <i>I129P</i>	-
R130P, <i>R130H</i>	1.0101 ⁷ , 1.0105 ¹⁴ , 1.0106 ¹⁴ , 1.0107 ¹⁴ , 1.0108 ¹⁴
K138Q	1.0104 ¹⁴

Table 3. Isoforms of the small subunit identified by comparison of middle-down (Monoisotopic obsM) and bottom-up (Monoisotopic theoM) results. The table shows the difference between observed and theoretical masses in parts per million (ppm) as well as the publication status of each identified isoform.

	Monoisotopic obsM	Monoisotopic theoM	mutation	error [ppm]	published isoforms
s-SU1	4431.1577	4431.2164	prototype	13.25	1.0105, 1.0107
s-SU2	4332.0827	4332.1368	R6G	12.49	1.0101, 1.0104, 1.0106, 1.0108
s-SU3	4348.0845	4348.1317	R11G, A17S	10.86	-
s-SU4	4447.1642	4447.1551	A17S	2.05	-

Table 4. Isoforms of the large subunit identified by comparison of middle-down (Monoisotopic obsM) and Bottom-Up (Monoisotopic theoM) results. The table shows the difference between observed and theoretical masses in parts per million (ppm) as well as the publication status of each identified isoform.

	Monoisotopic obsM	Monoisotopic theoM	mutation	error [ppm]	published isoforms
I-SU1	10136.0209	10136.1730	G100E, R130P	15.00	1.0101, 1.0105, 1.0107
I-SU2	10040.9819	10041.1358	H108Q, L109V, R130P	15.33	1.0106
I-SU3	10012.9761	10013.1297	R61K, H108Q, L109V, R130P	15.34	1.0108
I-SU4	10186.2609	10186.2210	G100E, H111Q	3.92	1.0103

Table 5. Isoforms of Sin a 1 identified by comparison of top-down (Monoisotopic obsM) and bottom-up (Monoisotopic theoM) results. The table shows the difference between observed and theoretical masses in parts per million (ppm) as well as the publication status of each identified isoform.

	Monoisotopic obsM	Monoisotopic theoM	mutation	error [ppm]	published isoforms
s-SU1 + l-SU1	14567.4176	14567.3894	G100E, R130P	1.94	1.0105
s-SU2 + l-SU2	14373.3049	14373.2726	R6G, H108Q, L109V, R130P	2.25	1.0106
s-SU1 + l-SU1	14567.4176	14567.3894	G100E, R130P	1.94	1.0107
s-SU2 + l-SU3	14345.2976	14345.2665	R6G, R61K, H108Q, L109V, R130P	2.17	1.0108
s-SU3 + l-SU2 Var. 1	14388.0600	14389.2675	R11G, A17S, H108Q, L109V, R130P	83.92	-
s-SU3 + l-SU2 Var. 2	14388.0600	14389.2675	R11G, A17S, H108Q, L109V, R130P	83.92	-
s-SU1 + l-SU2	14472.3634	14472.3522	H108Q, L109V, R130P	0.77	-
s-SU4 + l-SU1	14582.4144	14583.3281	A17S, G100E, R130P	62.66	-
s-SU1 + l-SU4	14617.4186	14617.4374	G100E, H111Q	1.29	-

Table 6. Relative quantitative distribution of identified Sin a 1 isoforms (assessed by relative signal height).

isoform	monoisotopic mass	relative intensity [%]
1.0108	14345.2665	~57
1.0106	14373.2726	~76
isoform 12	14389.2675	~26
isoform 9	14472.3522	~25
prototype	14536.4060	~67
1.0107	14567.3894	~100
isoform 10	14583.3281	~34
isoform 11	14617.4374	~36

ASSOCIATED CONTENT

Supporting information

- Table 1: Peptide sequences identified by bottom-up approach using PEAKS v7.
- Table 2: Set of relevant peptide sequences bearing sequence modifications identified by bottom-up and confirmed by middle-down and top-down.
- Figure 1: HPLC-HRMS positive ion spectrum of phytic acid.

This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jbrockm@uni-muenster.de. Phone: +49-251-8333392. Fax: +49-251-8333396

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Marlene Hummel was supported by a grant from the Heinrich-Stockmeyer-Stiftung, Bad Rothenfelde.

ABBREVIATIONS

1
2
3 SEC, size exclusion chromatography; HRMS, high resolution mass spectrometry; MeOH,
4
5 methanol; FA, formic acid; RP, reversed phase; ACN, acetonitrile; PTM, posttranslational
6
7 modification.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

(1) Figueroa, J.; Blanco, C.; Dumpiérrez, A. G.; Almeida, L.; Ortega, N.; Castillo, R.; Navarro, L.; Pérez, E.; Gallego, M. D.; Carrillo, T. Mustard allergy confirmed by double-blind placebo-controlled food challenges: clinical features and cross-reactivity with mugwort pollen and plant-derived foods. *Allergy* **2005**, *60*, 48–55.

(2) Rancé, F. Mustard allergy as a new food allergy. *Allergy* **2003**, *58*, 287–288.

(3) Rancé, F.; Dutau, G.; Abbal, M. Mustard allergy in children. *Allergy* **2000**, *55*, 496–500.

(4) Monsalve, R. I. Allergy to Mustard Seeds: The Importance of 2S Albumins as Food Allergens. *Internet Symposium on Food Allergens* **2001**, *3*, 57–69.

(5) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed storage proteins: structures and biosynthesis. *Plant Cell* **1995**, *7*, 945–956.

(6) Moreno, F. J.; Clemente, A. 2S Albumin Storage Proteins: What Makes them Food Allergens? *Open Biochem. J.* **2008**, *2*, 16–28.

(7) Menéndez-Arias, L.; Moneo, I.; Domínguez, J.; Rodríguez, R. Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, Sin a I. *Eur. J. Biochem.* **1988**, *177*, 159–166.

(8) Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 821–830.

(9) Sathe, S. K.; Kshirsagar, H. H.; Roux, K. H. Advances in Seed Protein Research: A Perspective on Seed Allergens. *J. Food Sci.* **2005**, *70*, 93–120.

- (10) Moreno, F. J.; Jenkins, J. A.; Mellon, F. A.; Rigby, N. M.; Robertson, J. A.; Wellner, N.; Clare Mills, E. N. Mass spectrometry and structural characterization of 2S albumin isoforms from Brazil nuts (*Bertholletia excelsa*). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2004**, *1698*, 175–186.
- (11) Guéguen, J.; Popineau, Y.; Anisimova, I. N.; Fido, R. J.; Shewry, P. R.; Tatham, A. S. Functionality of the 2S Albumin Seed Storage Proteins from Sunflower (*Helianthus annuus L.*). *J. Agric. Food Chem.* **1996**, *44*, 1184–1189.
- (12) Menéndez-Arias, L.; Monsalve, R. I.; Gavilanes, J.; Rodríguez, R. Molecular and spectroscopic characterisation of a low molecular weight seed storage protein from yellow mustard (*Sinapis alba L.*). *Int. J. Biochem.* **1987**, *19*, 899–907.
- (13) González De La Peña, M A; Villalba, M.; García-López, J. L.; Rodríguez, R. Cloning and expression of the major allergen from yellow mustard seeds, Sin a I. *Biochem. Biophys. Res. Commun.* **1993**, *190*, 648–653.
- (14) González De La Peña, M A; Monsalve, R. I.; Batanero, E.; Villalba, M.; Rodríguez, R. Expression in *Escherichia coli* of Sin a 1, the major allergen from mustard. *Eur. J. Biochem.* **1996**, *237*, 827–832.
- (15) Schägger, H. Tricine–SDS–PAGE. *Nat. Protoc.* **2006**, *1*, 16–22.
- (16) Kinter, M.; Shermann, N. E. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, 1st Edition; John Wiley & Sons: New York, 2000.

- (17) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.
- (18) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell Proteomics* **2012**, *11*, M111.010587.
- (19) Chen, Y.-F.; Chang, C. A.; Lin, Y.-H.; Tsay, Y.-G. Determination of accurate protein monoisotopic mass with the most abundant mass measurable using high-resolution mass spectrometry. *Anal. Biochem.* **2013**, *440*, 108–113.
- (20) Zubarev, R. A.; Bonddarenko, P. V. An A priori relationship between the average and monoisotopic masses of peptides and oligonucleotides. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 276–277.
- (21) Senko, M. W.; Beu, S. C.; McLafferty, F. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
- (22) Moreno, F. J.; Maldonado, B. M.; Wellner, N.; Mills, E. N. C. Thermostability and in vitro digestibility of a purified major allergen 2S albumin (Ses i 1) from white sesame seeds (*Sesamum indicum* L.). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005**, *1752*, 142–153.
- (23) Chapman, M. D.; Pomés, A.; Breiteneder, H.; Ferreira, F. Nomenclature and structural biology of allergens. *J. Allergy Clin. Immunol.* **2007**, *119*, 414–420.

- (24) Rabjohn, P.; Helm, E. M.; Stanley, J. S.; West, C. M.; Sampson, H. A.; Burks, A. W.; Bannon, G. A. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J. Clin. Invest.* **1999**, *103*, 535–542.
- (25) Rabjohn, P.; West, C. M.; Connaughton, C.; Sampson, H. A.; Helm, R. M.; Burks, A. W.; Bannon, G. A. Modification of peanut allergen Ara h 3: effects on IgE binding and T cell stimulation. *Int. Arch. Allergy Immunol.* **2002**, *128*, 15–23.
- (26) Ferreira, F.; Ebner, C.; Kramer, B.; Casari, G.; Briza, P.; Kungl, A. J.; Grimm, R.; Jahn-Schmid, B.; Breiteneder, H.; Kraft, D.; Breitenbach, M.; Rheinberger, H. J.; Scheiner, O. Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. *FASEB J.* **1998**, *12*, 231–242.
- (27) Takai, T.; Yokota, T.; Yasue, M.; Nishiyama, C.; Yuuki, T.; Mori, A.; Okudaira, H.; Okumura, Y. Engineering of the major house dust mite allergen Der f 2 for allergen-specific immunotherapy. *Nat. Biotechnol.* **1997**, *15*, 754–758.
- (28) Menéndez-Arias, L.; Domínguez, J.; Moneo, I.; Rodríguez, R. Epitope mapping of the major allergen from yellow mustard seeds, Sin a I. *Mol. Immunol.* **1990**, *27*, 143–150.
- (29) Monsalve, R. I.; Gonzalez de la Peña, M A; Menendez-Arias, L.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R. Characterization of a new oriental-mustard (*Brassica juncea*) allergen, Bra j IE: detection of an allergenic epitope. *Biochem. J.* **1993**, *293*, 625–632.
- (30) Robotham, J. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Linear IgE epitope mapping of the English walnut (*Juglans regia*) major food allergen, Jug r 1. *J. Allergy Clin. Immunol.* **2002**, *109*, 143–149.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(31) Pantoja-Uceda, D.; Bruix, M.; Giménez-Gallego, G.; Rico, M.; Santoro, J. Solution Structure of RicC3, a 2S Albumin Storage Protein from *Ricinus communis*. *Biochemistry* **2003**, *42*, 13839–13847.

(32) Reddy, N. R.; Sathe, S. K.;Salunkhe, D. K. Phytates in legumes and cereals. *Adv. Food Res.* **1982**, *28*, 1–92.

(33) Chung, S. Y.; Champagne, E. T. Effects of phytic acid on peanut allergens and allergenic proterties of extracts. *J. Agric. Food Chem.* **2007**, *55*, 9054-9058.

(34) Lepski, S.; Brockmeyer, J. Impact of dietary factors and food processing on food allergy. *Mol. Nutr. Food Res.* **2013**, *57*, 145-152.

Figure 1: Identification of the epitope on the Hsp70 protein. The top panel shows a mass spectrum with a major peak at m/z 16322, identified as Hsp70. The bottom-left panel shows a 1D NMR spectrum with peaks at 1.076, 2.002, 2 = 0 and 1.036, 1.019, 0 = 0. The bottom-right panel is a schematic of the Hsp70 protein structure, highlighting the hypervariable region and the specific epitope (residues 158-164) in red.

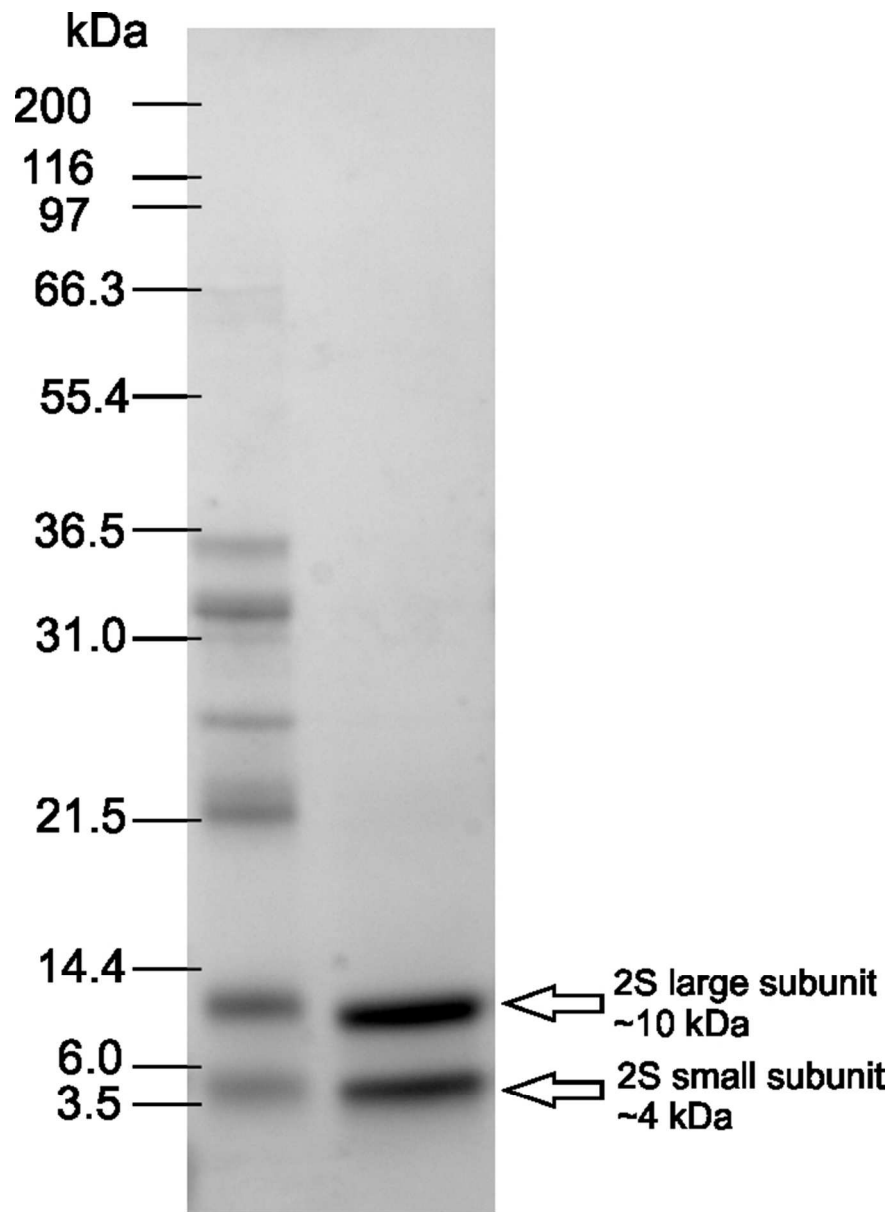
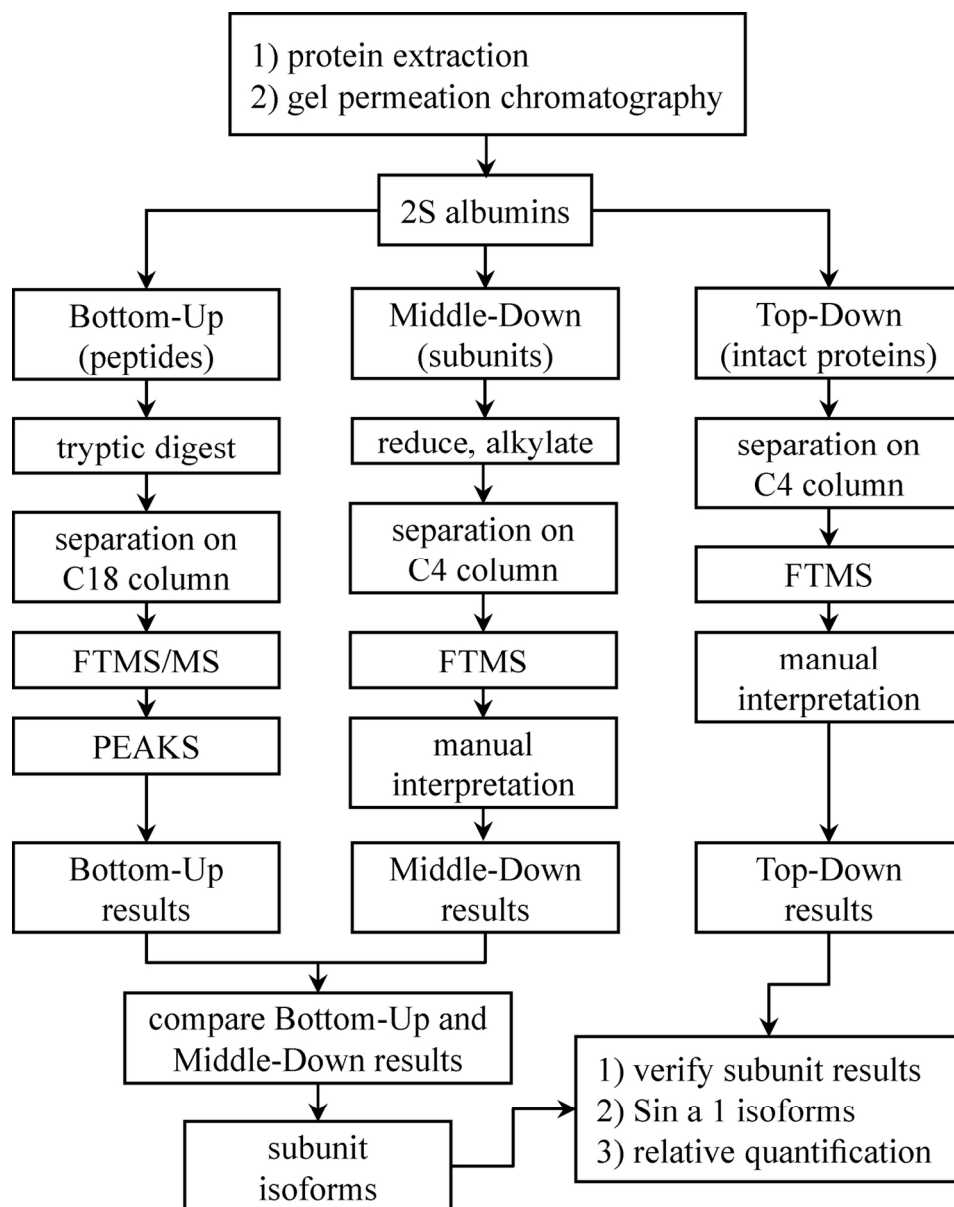
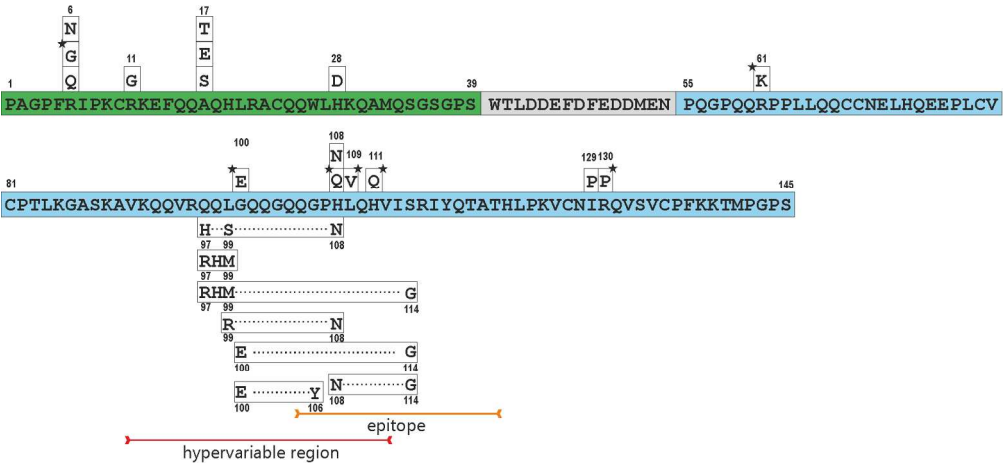


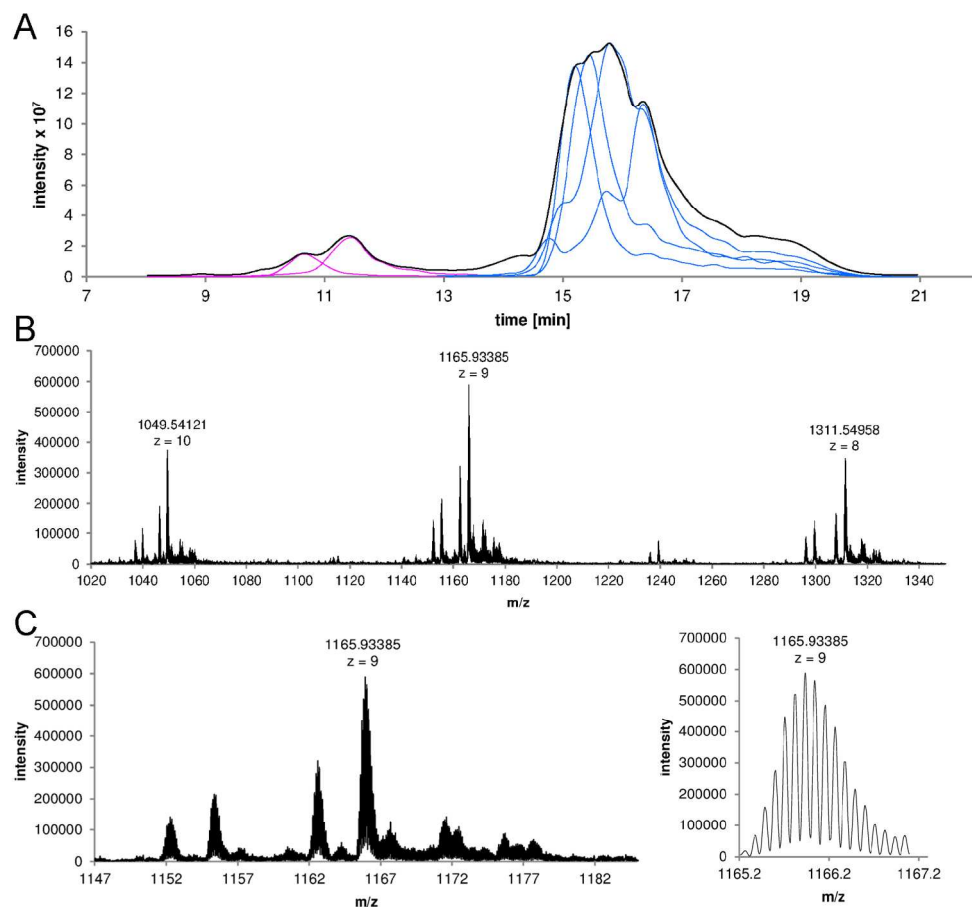
Figure 1. Tricine SDS-PAGE (T = 16 %, C = 2.6 %) of mustard extract. Lane 1 = crude extract, lane 2 = low molecular weight fraction (Sin a 1) purified by size exclusion chromatography.
59x81mm (300 x 300 DPI)



131x165mm (300 x 300 DPI)

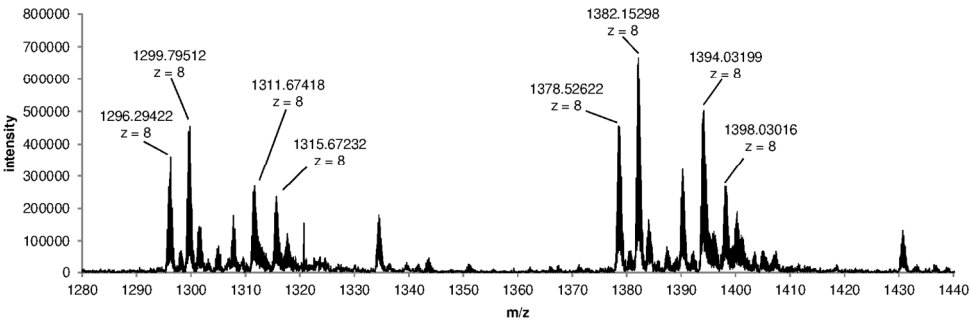


177x81mm (300 x 300 DPI)



177x163mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



175x60mm (300 x 300 DPI)

```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

Prototype	1	11	21	31	55	61	71	81
Sin a 1.0105**	PAGPFRIPKC	RKEFQQAQHL	RACQQWLHKQ	AMQSGSGSPS	(-)PQGPQQ	RPQLLQQCCN	ELHQEEPLCV	CPTLKGASKA
Sin a 1.0106	-----G-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1.0107**	-----G-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1.0108	-----G-----	-----	-----	-----	(-)-----	K-----	-----	-----
Sin a 1 isoform 9	-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 10	-----	-----S-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 11	-----	-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 12 vs1	-----	G-----S-----	-----	-----	(-)-----	-----	-----	-----
Sin a 1 isoform 12 vs2	-----G-----	-----S-----	-----	-----	(-)-----	-----	-----	-----

Prototype	91	101	111	121	131	141
Sin a 1.0105**	VKQOVRQOLG	OQGOOGPHLQ	HVISRIYOTA	THLPKVCNIR	QVSVCPEKKT	MPGPS
Sin a 1.0106	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1.0107**	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1.0108	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 9	-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 10	-----E-----	-----	-----	-----P-----	-----	-----
Sin a 1 isoform 11	-----E-----	-----Q-----	-----	-----P-----	-----	-----
Sin a 1 isoform 12 vs1	-----E-----	-----QV-----	-----	-----P-----	-----	-----
Sin a 1 isoform 12 vs2	-----E-----	-----QV-----	-----	-----P-----	-----	-----

350x156mm (300 x 300 DPI)