

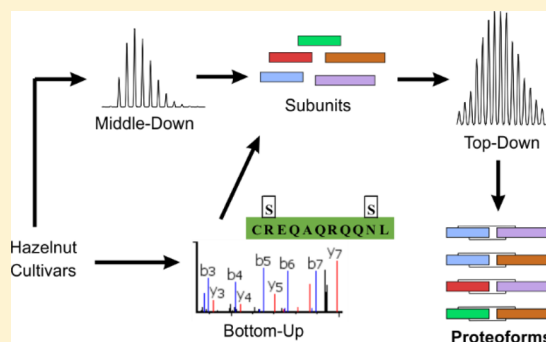
Structural Characterization of the Allergenic 2S Albumin Cor a 14: Comparing Proteoform Patterns across Hazelnut Cultivars

Robin Korte,[†] Jana Happe,[†] Ina Brümmer,[‡] and Jens Brockmeyer^{*,†,‡}[†]Institute of Food Chemistry, Westfälische Wilhelms-Universität Münster, Corrensstraße 45, 48149 Münster, Germany[‡]Analytical Food Chemistry, University of Stuttgart, Allmandring 5b, 70569 Stuttgart, Germany

S Supporting Information

ABSTRACT: The hazelnut allergen Cor a 14 belongs to the 2S albumins, a family of heterodimeric seed storage proteins exhibiting a high degree of structural diversity. Given its relevance as an allergen and the potential to elicit severe reactions, elucidation of the sequence heterogeneity of naturally occurring Cor a 14 is essential for the development of reliable diagnostics and risk evaluation. We therefore performed a comprehensive survey on the proteoforms of Cor a 14 and determined their quantitative distribution in three different hazelnut cultivars by a combinatory HPLC–HRMS approach including bottom-up and intact mass analysis. Compared with the Cor a 14 prototype sequence, we identified three sequence polymorphisms, two of the small and one of the large subunit, and elucidated their specific pairing on the protein level. Furthermore, we located a pronounced microheterogeneity on the protein termini and, for the first time, provide data on varying proteoform patterns between different cultivars of an allergenic seed. Together, these data present the basis for a more detailed investigation on the allergenicity of Cor a 14 in different cultivars and constitute, to be best of our knowledge, the largest set of proteoforms so far reported for a 2S albumin.

KEYWORDS: food allergens, 2S albumin, Cor a 14, hazelnut, HPLC–MS, mass spectrometry, proteoform, protein isoforms, clipping



1. INTRODUCTION

Allergy to hazelnut seeds ranks among the most prevalent food allergies in Europe and, like other tree nut allergies, has been associated with severe clinical symptoms.^{1,2} Eight proteins have so far been identified as elicitors for hazelnut allergy: Cor a 1 and Cor a 2 belong to the Bet v 1 and profilin families, respectively.^{3,4} Both are responsible for the majority of sensitizations⁵ but mainly cause rather mild reactions and are easily inactivated by thermal processes or proteolytic activity in the gastrointestinal tract.⁶ More severe symptoms have been associated with the lipid transfer protein Cor a 8,⁷ the 11S and 7S globulins Cor a 9 and Cor a 11,^{8,9} the oleosins Cor a 12 and Cor a 13,^{10,11} and, most recently identified, the 2S albumin Cor a 14.¹²

The 2S albumins comprise a family of seed storage proteins that are widely distributed in the plant kingdom. They consist of two subunits (MW approximately 8–10 and 3 to 4 kDa) connected via four conserved disulfide bridges and exhibit a high degree of α -helix elements.^{13,14} As a result of their compact 3D structure, the 2S albumins are extraordinarily resistant against heat and protease treatment, and this feature is considered to contribute to their high potential as allergenic sensitizers.^{15–18} 2S albumins are initially synthesized on the rough endoplasmic reticulum (rER) as a single precursor protein,¹⁹ which is cotranslationally transported into the ER-lumen. Following protein maturation, the ER signal sequence and a linker peptide are removed, the four disulfide bonds are

formed, and the folded protein is proteolytically processed, yielding the disulfide-associated heterodimer.^{14,20}

Because they are encoded by a multigene family and subjected to multiple stages of post-translational processing, 2S albumins exhibit a high degree of structural heterogeneity. Several studies have identified and described isoforms of 2S albumins from various seeds, which either differ through minor variations, for example, single amino acid polymorphisms or differential processing sites, or even show considerable variation over large parts of the sequence due to frameshift mutations.^{21–23} Recently, Pfeifer and coauthors reported on the mass spectrometric analysis of naturally occurring Cor a 14 as part of a larger study. In this context, they identified an isoform with a different amino acid sequence than reported in the initial characterization of the Cor a 14 gene and, furthermore, described clipping at the small subunit's C-terminus.²⁴

Although the high relevance of Cor a 14 as a sensitizer for hazelnut allergy has recently been confirmed as part of the EuroPrevall study,⁵ a comprehensive investigation of the structural heterogeneity, including relative quantitation of proteoforms, is still lacking. The focus of our study was therefore to systematically identify and describe the proteoforms of Cor a 14. In this context, the term proteoform

Received: October 24, 2016

Published: January 23, 2017

describes all molecular forms of a protein, including sequence polymorphisms, due to genetic variation, as well as post-translational modifications.²⁵ For this extensive characterization, our study combined three proteomics strategies (bottom-up analysis and intact mass analysis of subunits and of the unreduced proteins, respectively), an approach that has recently been applied for the characterization of Sin a 1 isoforms.²¹ Through the analysis of three different hazelnut cultivars, we aimed to identify and characterize a comprehensive consensus set of Cor a 14 proteoforms. In addition, we intended to investigate if available cultivars show differences in their Cor a 14 proteoform patterns, which may arise, for example, from differential expression or post-translational modification of the allergenic protein.

2. MATERIALS AND METHODS

Sample Material

Raw hazelnuts of different European cultivars, “Butler” (Slovenia), “Webb’s Prize Nut” (Leiselheim, Germany), and “Katalonski” (Gumulka, Poland), were kindly provided and certified by the Middle Bavarian Gardening Center (Amt für Ernährung, Landwirtschaft und Forsten, Fürth, Germany). All hazelnuts were harvested in 2011 and stored since at -20°C in unprocessed form.

Extraction and Purification of Natural Cor a 14

After the removal of the shell, hazelnuts were ground to a fine flour and homogenized with a cutting mill (Grindomix GM200, Retsch). The flour was defatted with *n*-pentane (1 g: 13 mL) for 2 h under continuous shaking, siphoned off by a suction strainer, and air-dried overnight. Proteins were extracted by 2 h shaking in Tris buffer (200 mM, pH 8.0, 1 g: 10 mL) at room temperature, then centrifuged at 3850g for 15 min, and filtered through a 0.22 μm poly(ether sulfone) Steritop-GP filter (Merck Millipore, Darmstadt, Germany).

Gel filtration was applied for isolation of the 2S albumin fraction, using an ÄKTA purifier FPLC system (GE Healthcare, Munich, Germany). A total of 100 μL of filtrated hazelnut extract was injected on a Superdex 200 10/300 GL column (GE Healthcare), the extraction buffer (200 mM Tris, pH 8.0) was run at a flow rate of 400 $\mu\text{L}/\text{min}$, and proteins were detected by their absorbance at 215, 254, and 280 nm. The identity of protein containing fractions was analyzed by Tricine SDS-PAGE (16%) under reducing conditions, according to the protocol from Schägger with slight modifications.²⁶ Gels were run for 30 min at 90 V, followed by 100 min at 150 V and stained with Coomassie Blue. Protein concentration in the Cor a 14 fraction was determined via Bradford assay.

Bottom-Up Analysis

Cor a 14 fractions from gel filtration (containing ~ 500 μg protein) were precipitated in a 4-fold volume of cold acetone (-20°C), resuspended in 50 μL of urea buffer (6 M, 100 mM Tris), and subjected to enzymatic digestion by either trypsin or endoproteinase Glu-C. A total of 2.5 μL of reduction reagent (200 mM dithiothreitol, 100 mM Tris) was added, and the sample was incubated for 1 h in the dark. After the addition of 10 μL of alkylation reagent (200 mM iodoacetamide, 100 mM Tris), the sample was incubated for 1 h, and another 10 μL of reduction reagent was added to inhibit further alkylation. The solution was diluted with 387.5 μL of H_2O , followed by the addition of 50 μL of enzyme solution, containing 10 μg of sequencing-grade trypsin (Serva, Heidelberg, Germany) or

endoproteinase Glu-C (MS approved, Serva) in 100 mM Tris. Digestion was carried out overnight under slow shaking at 37°C . For desalting and purification, the enzymatic digests were acidified to pH < 6 with acetic acid and applied to solid-phase extraction using Strata-X 33u RP 30 mg/1 mL columns (Phenomenex, Aschaffenburg, Germany). Cartridges were activated with methanol (MeOH) and equilibrated with 1% formic acid (FA). Samples were loaded, washed with 1 mL of 1% FA, and eluted with 1 mL of MeOH/ H_2O + 1% FA (7:3). The solvent was removed at 40°C under a gentle nitrogen flow, and peptides were redissolved in 100 μL of acetonitrile (ACN)/ H_2O (3:97), according to the HPLC starting conditions.

Data were acquired on an LTQ Orbitrap XL hybrid ion trap–orbitrap mass spectrometer coupled to an Accela HPLC system (both Thermo Scientific, Dreieich, Germany). The injection volume was 25 μL (corresponding to ~ 125 μg of digested protein), and LC separation of enzymatic digests was carried out on a 2.1×150 mm, 2.6 μm Accucore C18 column (Thermo Scientific) at a flow rate of 250 $\mu\text{L}/\text{min}$, employing the following gradient of ACN/ H_2O + 1% FA: 3% ACN for 6 min, 3 to 25% in 64 min, 25 to 60% in 10 min, 60 to 90% in 1 min, 90% for 4 min, 90 to 3% in 1 min, and reequilibration at 3% ACN for 5 min. The mass spectrometer was run in positive full-scan and data-dependent mode. Survey full-scan mass spectra (m/z 300–1500) were acquired in the Orbitrap ($r = 30\,000$), and the five most intense peptide ions were sequentially isolated and fragmented in the linear ion trap, using collision-induced dissociation (CID, normalized collision energy of 30% and activation time of 30 ms). Resulting fragments were detected at $r = 7500$, and dynamic exclusion was set to 30 s. Ions with unrecognized charge states were rejected. Electrospray ionization (ESI) conditions were: Vaporizer temperature 350°C ; capillary temperature 225°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.

A database search was performed with PEAKS 7 (Bioinform, Waterloo, Canada) against a hazelnut protein database (Uniprot, *Corylus avellana*, created 2015-01-29, 389 entries) using the PEAKS de novo algorithm and the enhanced target-decoy method (“decoy fusion”) for false discovery rate (FDR) estimation and result validation.^{27,28} Search parameters were: parent mass tolerance 5 ppm; fragment mass tolerance 0.02 ppm; monoisotopic precursor; max missed cleavages 1; max nonspecific cleavages 1; carbamidomethylation of Cys as fixed modification; oxidation of Met as variable modification; and max variable PTMs per peptide 5. Following database search, the PEAKS PTM and PEAKS Spider algorithms were applied to identify peptides bearing polymorphisms or PTMs compared with the published Cor a 14 sequence.¹² Peptide matches were limited to an FDR of 0.1%, and all spectra were inspected manually to ensure reliable identification of potential modifications.

Intact Mass Analysis of Protein Subunits

Fractions from gel filtration containing Cor a 14 (1.25 mL each, containing ~ 62.5 – 99 μg of protein) were precipitated as previously described and resuspended in 25 μL of Tris buffer (200 mM, pH 8.0). For the analysis of subunits, the native Cor a 14 was reduced by adding 35 μL of 50 mM ammonium bicarbonate buffer containing 14.3 mM dithiothreitol and incubating for 30 min at 56°C , followed by alkylation through

the addition of 10 μL of 150 mM iodoacetamide in 50 mM ammonium bicarbonate buffer for 30 min in the dark. Reduced and alkylated samples were diluted with 130 μL of ACN/0.1% FA (5:95), followed by HPLC–HRMS analysis.

Data acquisition was performed on the same system as described above for bottom-up analysis, applying an injection volume of 10 μL on a $3.0 \times 1 \text{ mm}$, $2.6 \mu\text{m}$ Accucore C4 column (Thermo Scientific). LC separation was carried out at a flow rate of 400 $\mu\text{L}/\text{min}$, employing the following gradient of ACN/ H_2O + 0.1% FA: 5% ACN for 1 min, 5 to 40% ACN in 16 min, 40 to 95% in 3 min, 95% for 2 min, 95 to 5% in 1 min, and reequilibration at 5% ACN for 5 min. Full-scan mass spectra (m/z 500–2000) were acquired with a resolution of $r = 30\,000$. ESI conditions were set as described above for bottom-up analysis.

Intact Mass Analysis of Unreduced Proteins

Gel filtration fractions containing Cor a 14 (1 mL each, containing $\sim 50\text{--}79 \mu\text{g}$ of protein) were precipitated as previously described and resuspended in 20 μL of Tris buffer (200 mM, pH 8.0). The native proteins were then diluted with 80 μL of ACN/0.1% FA (5:95) and directly applied to HPLC–HRMS.

For the acquisition of MS data, the same system and column as described for the intact mass analysis of protein subunits was applied. LC separation was carried out at a flow rate of 400 $\mu\text{L}/\text{min}$, employing the following gradient of ACN/ H_2O + 0.1% FA: 5% ACN for 1 min, 5 to 35% ACN in 19 min, 35 to 95% in 2 min, 95% for 5 min, 95 to 5% in 1 min, and reequilibration at 5% ACN for 3 min. Full-scan mass spectra (m/z 500–2000) were acquired with a resolution of $r = 60\,000$. ESI conditions were set as described above for bottom-up analysis.

Evaluation of Data from Intact Mass Analyses

Data evaluation was carried out manually for all intact mass measurements, using Xcalibur, version 2.0.7 SP1 (Thermo Scientific), and Microsoft Excel. To ensure the highest mass accuracy, we employed the following strategy for determination of the monoisotopic mass of each proteoform.

Using the isotope simulation tool included in Xcalibur, we calculated the theoretical isotopic distribution of each subunit and the unreduced protein according to the published Cor a 14 prototype sequence.¹² We then manually inspected the isotopic pattern of every experimentally observed proteoform and determined from it the monoisotopic mass (small subunit), the +5 isotopic mass (large subunit), or the +7 isotopic mass (unreduced proteins). Interpretation of the isotopic patterns was based on homology to the theoretical isotopic distribution. Monoisotopic masses were deduced by subtraction of m/z 5.0097 from the +5 isotopic mass (large subunit) and 7.01423 from the +7 isotopic mass (unreduced proteins), respectively, and charge deconvolution was performed to yield the monoisotopic protein mass as prior to ionization.

Because changes in the amino acid sequence, as observed for proteoforms, can influence the isotopic distribution, we did not strictly limit the data analysis to the theoretical isotopic distribution of the Cor a 14 prototype in case of every proteoform. In particular, proteoforms that show a high degree of clipping (truncation) at the termini may slightly deviate from the prototype isotopic distribution as a result of a lower molecular mass. Therefore, we also considered the theoretical isotopic pattern of a slightly truncated Cor a 14 sequence (C-terminal truncation by six and three amino acids at the small and large subunit, respectively) for the evaluation of peak

patterns at lower m/z . Altogether, this approach proved to be robust, and correct mass identification was furthermore assured through the quality criteria mentioned in the Results section, including the evaluation of multiple charge states.

3. RESULTS

To reduce sample complexity for mass spectrometry, the native Cor a 14 protein was purified from hazelnut extracts by gel filtration (Figure 1A). Fractions containing the 2S albumin

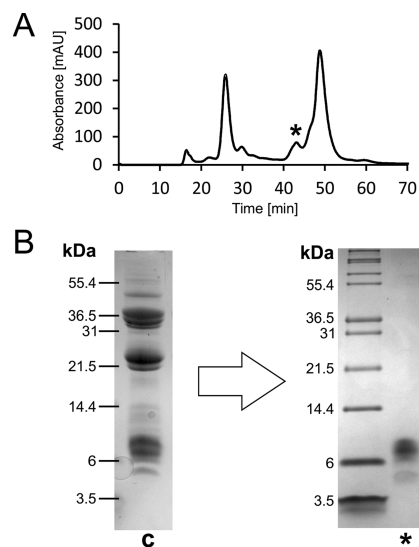


Figure 1. (A) Gel filtration profile of hazelnut extract, showing absorbance at 280 nm. The asterisk (*) indicates the Cor a 14 fraction. (B) Tricine SDS-PAGE (16%) analysis of crude hazelnut extract (C) and the purified SEC fraction (*).

were identified by reducing SDS-PAGE (Figure 1B), exhibiting the characteristic bands of the large (8 to 9 kDa) and small subunit (4 kDa) of Cor a 14. Protein concentrations were determined as 79 $\mu\text{g}/\text{mL}$ (Butler), 59 $\mu\text{g}/\text{mL}$ (Webb's Prize Nut), and 50 $\mu\text{g}/\text{mL}$ (Katalonski) via Bradford assay. Tryptic digestion and bottom-up HPLC–HRMS/MS analysis confirmed the identity and enrichment of Cor a 14, with an average sequence coverage of $78 \pm 8\%$ and an average -10lgP score of 464 ± 34 ($n = 6$, Cor a 14 prototype (UniProt D0PWG2), not including signal and linker peptides for calculation of coverage) but also indicated the presence of other remaining hazelnut allergens (Cor a 1, Cor a 8, Cor a 9, Cor a 11), however, at overall lower -10lgP scores.

The purified Cor a 14 fractions were then applied to mass spectrometric analysis, combining three complementary proteomics approaches, as displayed in Figure 2. Bottom-up analysis was used for the qualitative study of Cor a 14 heterogeneity on the peptide level. Applying the PEAKS de novo and Spider algorithms to HPLC–MS/MS data sets of enzymatic digests allowed us to identify potential sequence polymorphisms and PTMs and to assign them to the Cor a 14 sequence. However, bottom-up results yield no information about the combination of modifications. Consequently, the quantitative information on the bottom-up data is limited. We thus analyzed both subunits after reduction and alkylation of Cor a 14 by intact mass analysis via HPLC–HRMS to characterize the diversity of Cor a 14 proteoforms on the subunit level. Combining bottom-up data with the intact mass analysis of subunits allowed us to map polymorphisms and

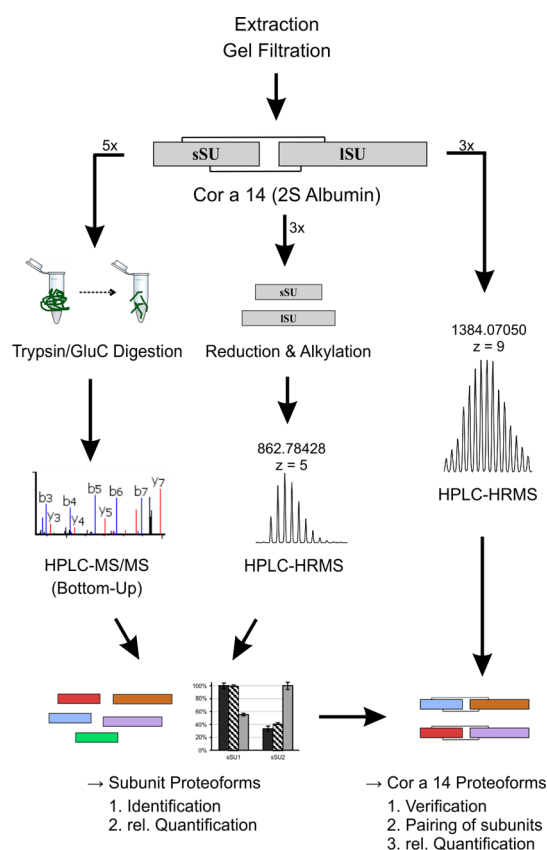


Figure 2. Complementary approach for the characterization of Cor a 14 proteoforms combining bottom-up and intact mass analysis.

PTMs and to identify the majority of the observed subunit proteoforms. Using intact mass analysis of the unreduced 2S albumins, we verified these results and identified the specific pairing of subunits to mature Cor a 14 proteoforms. Data from the intact mass analyses were also used for the relative quantification of these proteoforms across three different hazelnut cultivars.

Identification and Characterization of Cor a 14 Small and Large Subunit Proteoforms

To provide for a comprehensive characterization of the Cor a 14 proteoform pattern, we particularly focused on the intact mass analysis of the 2S albumin subunits as the initial point of our systematic study. Given the direct measurement of protein subunits and the manual evaluation of the data, this approach features less abstraction from proteoform constitution, the object under investigation, than bottom-up. Furthermore, we achieved a distinct peak separation for most major proteoforms of both the small and large subunit by HPLC–HRMS on a C4 column, as shown in Figure 3A, allowing us to significantly reduce mass spectrometric complexity. Chromatograms were thus arbitrarily divided by retention time into sections of ~0.1 min; HRMS spectra were averaged for each section and inspected independently. Both small and large subunits were observed in multiple charge states exhibiting similar proteoform patterns, as demonstrated in Figure 3B, together with the isotopic patterns of two major large subunit proteoforms. To ensure the correct determination of subunit molecular mass, we manually inspected each signal in every observed charge state and extracted its monoisotopic mass. After subtraction of carbamidomethyl modifications resulting from the sample

preparation procedure, molecular masses with mass differences ≤ 0.05 Da from neighboring charge states were grouped and averaged. Only masses that were identified independently in ≥ 2 different charge states were considered for the following interpretation and characterization. The process was repeated with three different hazelnut cultivars (Butler, Webb's Prize Nut, and Katalonski), as interestingly we observed cultivar-specific differences in the proteoform pattern (Figure 3C).

For the structural characterization of these subunit proteoforms, the bottom-up analysis was employed. Either trypsin or endoproteinase Glu-C was used as a protease for enzymatic digestion to allow for an overall coverage of the whole Cor a 14 protein sequence. Altogether, 12 bottom-up experiments (two per protease per cultivar) were performed, and all fragment spectra suggested by the Spider algorithm for the identification of polymorphisms and PTMs were inspected manually. For both subunits, we observed single amino acid sequence polymorphisms as well as terminally truncated peptides originating from the N- and C-termini, presenting an explanatory basis for the heterogeneity of the Cor a 14 subunits observed in the intact mass analyses. Only such modifications that were found in both replicate experiments of a hazelnut cultivar were considered for interpretation of the intact mass results. Following this combination of the two approaches, we were able to assign 18 (Butler), 21 (Webb's Prize Nut), and 39 (Katalonski) proteoforms of the small and 17, 10, and 11 proteoforms of the large subunit with mass deviations < 3 ppm, taking into account the modifications presented in Figure 4. These included three sequence polymorphisms, two within the small subunit (R48S, N56S) and one within the large subunit (Q79H), subsequently referred to as isoforms sSU2, sSU3, and ISU2 (sSU1 and ISU1 are consecutively used to describe the original, unmodified subunits). sSU2 has been described as an isoform before by MALDI top-down sequencing mass spectrometry,²⁴ while ISU2 has been found as a mutation during the initial cloning of the Cor a 14 gene.¹² Both isoforms have, however, only been identified in one of the two cited studies and are therefore for the first time confirmed by our comprehensive approach. Furthermore, sSU3 represents a new, so far undescribed isoform of Cor a 14. In addition to sequence polymorphisms, we detected minor clipping at the N- and C-termini of both subunits (subsequently referred to as Nx/Cx, with x indicating the number of clipped amino acids), a modification that has already been observed for other 2S albumins.^{15,22,29–32} In the case of Cor a 14, only the C3, C5, C6, and C8 forms of the small and the C3 form of the large subunit have been described so far.²⁴ Similar to other mass spectrometric studies on 2S albumins,^{21,29} we also observed an ubiquitous transformation of N-terminal glutamine into pyroglutamic acid, which consistently occurred in bottom-up as well as in the intact mass analyses and has been described before as an ESI ionization artifact.³³ Most of the modifications given in Figure 4, except those only lowly abundant, were also found in combined forms (results discussed below), thus confirming their identity.

To verify the identified subunit proteoforms and to compare their distribution over the cultivars, we determined the peak areas of the monoisotopic (small subunit) and the +5 isotopic (large subunit) extracted ion traces (mass tolerance: 5 ppm) in three biological replicates of each cultivar. Because some proteoforms showed a different intensity over charge pattern, in particular, due to mutation or clipping of basic residues, we

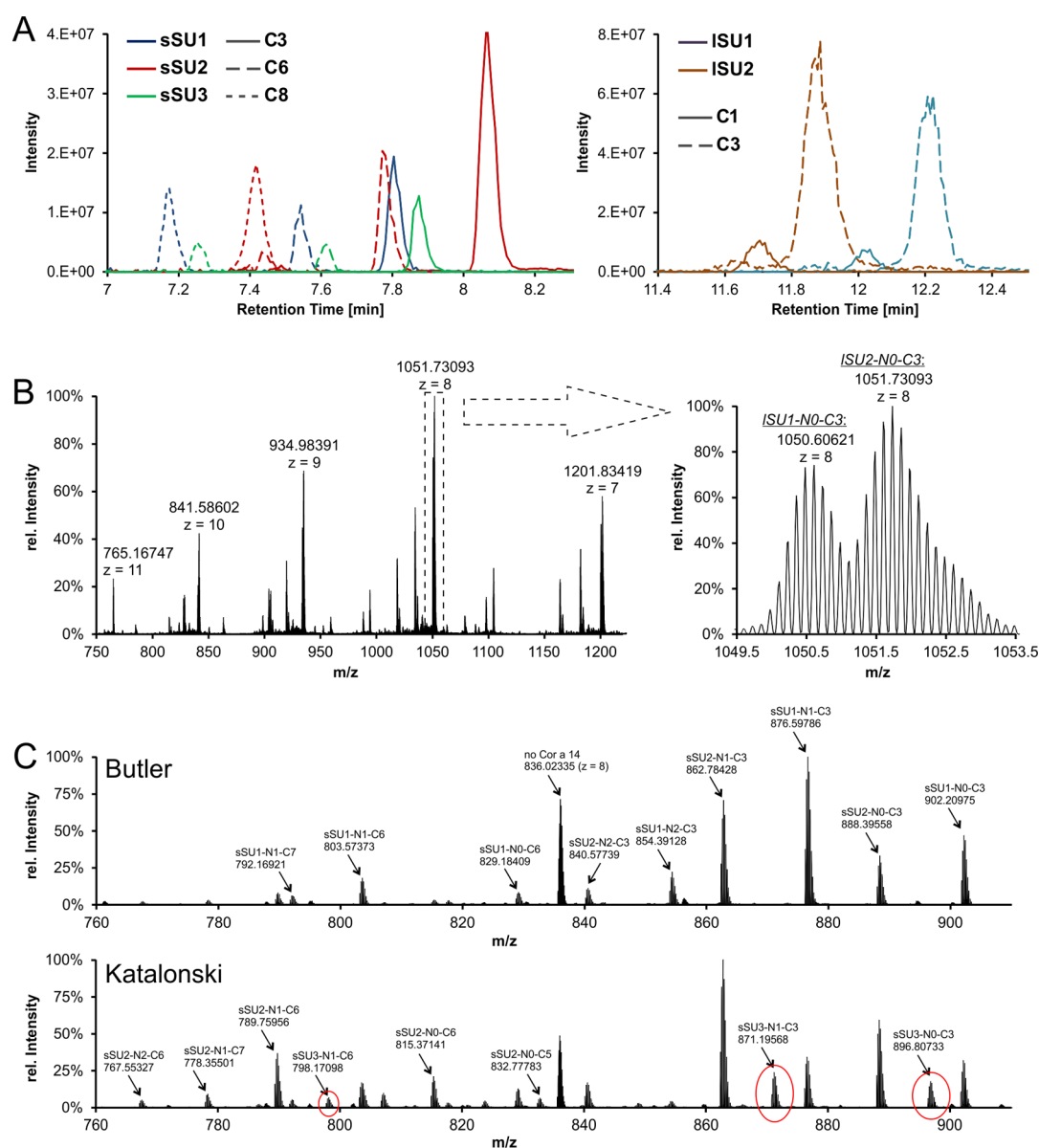


Figure 3. Intact mass analysis of Cor a 14 subunits in Katalonski hazelnuts. (A) Separation of small (left) and large (right) subunit proteoforms by HPLC. Extracted ion chromatograms of proteoforms carrying sequence polymorphisms are given in combination with C-terminal clipping. (B) HRMS spectra of the large subunit exhibiting similar proteoform patterns over multiple charge states. A detailed view of the overlapping isotopic patterns of two proteoforms is given to the right. (C) HRMS spectra of the small subunit from Butler and Katalonski hazelnut cultivars in direct contrast. (For assignment of abbreviations to protein sequence, see Figure 4.)

summed up the peak areas of all observable charge states (sSU: 4–7, ISU: 5–12) and calculated their relative values. The results given in Figure 5 demonstrate quantitative differences between the examined hazelnut cultivars, concerning all three types of modifications (polymorphisms, N-/C-terminal clipping) and both subunits (for the tabular data, the reader is referred to Tables S-1 and S-2 in the Supporting Information): Isoform sSU2 was found in all three cultivars, with the highest relative fraction in Katalonski, while sSU3 was detected exclusively in this cultivar. ISU2 occurred to be the dominating large subunit isoform, detectable in every cultivar but showing the highest degree in Webb's Prize Nut. N-terminal clipping was observed in all cultivars, with a moderately higher intensity in Butler for both subunits. A greater quantitative difference between cultivars was found in case of C-terminal clipping, where Butler exhibited the lowest and Katalonski the highest

degree of strongly clipped proteoforms. While the degree of N-terminal clipping to the small subunit appeared to be independent of sequence polymorphisms (see Figure S-1 and Table S-1 in the Supporting Information for quantitation of clipping in isoforms sSU2 and sSU3), a different relationship was observed for the large subunit: ISU2 exhibited a significantly lower relative share (21–31%) of the N1 form than ISU1 (63–84%), an effect that could be caused by the position of the polymorphism adjacent to the N-terminus, possibly influencing the activity of aminopeptidases. No such distinct isoform effect was observed for C-terminal clipping.

Confirmation and Pairing of Cor a 14 Subunit Proteoforms on Protein Level

Intact mass analysis of the unreduced 2S albumins was performed to verify the previously characterized subunit proteoforms and elucidate their pairing on the protein level.

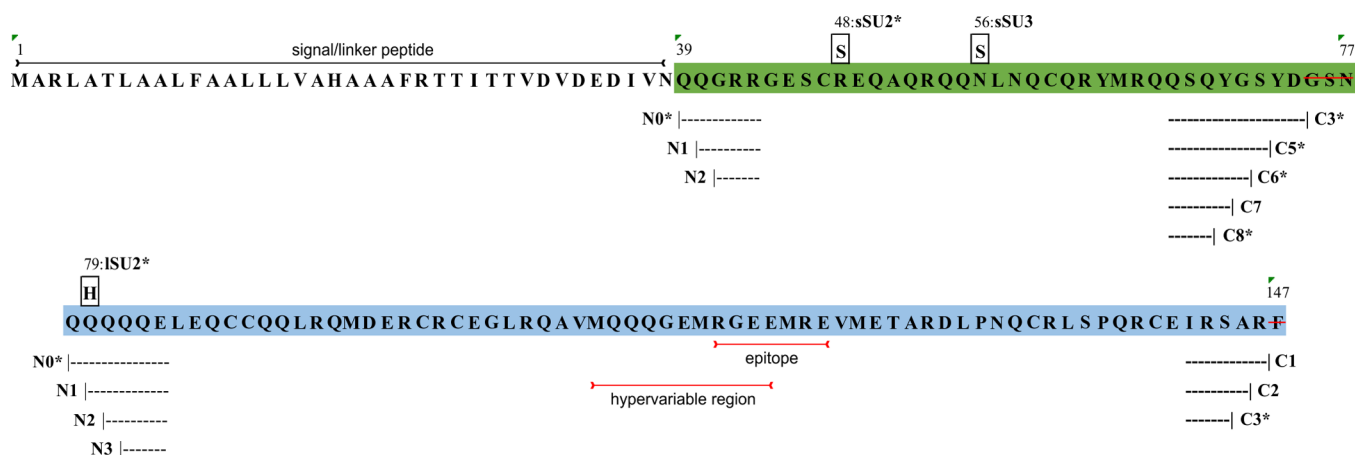


Figure 4. Mapping of modifications to Cor a 14 primary sequence, as described by Garino and coauthors. Sequences of the small and large subunit are highlighted in green and blue, respectively. Boxed letters indicate sequence polymorphisms, while Nx and Cx values indicate N- and C-terminal clipping. Amino acids crossed red were not confirmed in any hazelnut cultivar. All depicted modifications were reproducibly observed in bottom-up and verified by intact mass analysis. Modifications and clipping variants marked by asterisk (*) have been described before. The hypervariable region was assigned based on an alignment to the peanut 2S albumin Ara h 6.^{14,48}

As for the intact mass analysis of subunits, we manually inspected each signal in every observed charge state and extracted the monoisotopic mass, then grouped and averaged signals from neighboring charge states with mass differences ≤ 0.05 Da. All molecular masses identified independently in ≥ 2 charge states were compared with possible subunit pairings. This way, we identified 42 (Butler), 41 (Webb's Prize Nut), and 48 (Katalonski) proteoforms of Cor a 14 by exact mass, of which we were able to assign 37, 39, and 42 (88–95%) to combinations of the previously identified small and large subunits with a maximum error of 3 ppm. The results are shown in Figure 6, together with the relative quantitative distribution of the individual proteoforms, determined in triplicate analysis from the +7 isotopic extracted ion traces and summing up the peak areas of all observable charge states (7–11, mass tolerance: 5 ppm). Different from the analysis of subunits, we, however, did not achieve a chromatographic peak separation of the unreduced Cor a 14 proteoforms. Because several pairings of subunits, particularly those with different degrees of N-terminal clipping, show identical exact masses, we were not able to distinguish their peaks by HPLC–MS and therefore assigned these signals to both proteoforms equally. It must, however, be assumed that in most of these cases the relative abundance given in Figure 6 will be overrated and that possibly in some cases only one of the two isobaric proteoforms will actually be present in the hazelnut seeds.

In conclusion, all three sequence polymorphisms observed on the subunit level were confirmed by intact mass analysis of the unreduced proteins. The same applied to most variants of N- and C-terminal clipping. Two lowly abundant forms of the large subunit (N3 and C2) could, however, not be confirmed, possibly due to low signal intensities. Furthermore, the relative distribution of modifications over the investigated cultivars also matched the previously discussed results from subunit analysis: Butler hazelnuts exhibited the highest N-terminal and the lowest C-terminal clipping degree for both subunits. While sSU1 was the most abundant small subunit isoform in Butler and Webb's Prize Nut, sSU2 was again found to be the most abundant in Katalonski hazelnuts. Likewise, Katalonski featured the highest degree of C-terminal clipping at the small subunit

and proved to be the only cultivar that carried the sSU3 isoform.

Interestingly, analysis of the unreduced proteins also demonstrated that some subunit proteoforms were specifically paired with others based on their polymorphisms: While the small subunit proteoform sSU1 paired with both large subunit proteoforms (ISU1 and ISU2), sSU2 paired with ISU1 and sSU3 with ISU2 only, in each case exhibiting the characteristic pattern of N- and C-terminal clipping. We expect this specific pairing, as illustrated in Figure 7, to be identical to the genomic arrangement of nucleotide polymorphisms in the Cor a 14 multigene family.

4. DISCUSSION

The structural heterogeneity of 2S albumin seed storage proteins has been characterized before in several allergenic foods as a result of the high rates of polymorphisms and post-translational modifications in this protein family. So far, only one study focusing on molecular proteoform characterization of the hazelnut 2S albumin Cor a 14 has been published, which reported a single Arg/Ser exchange at position 48 of the precursor sequence and clipping of 3–8 residues at the small subunit's C-terminus.²⁴ In addition, the deletion of a Gln coding base triplet at position 79 and its mutation to His have been found as alternative variants during initial sequencing of the Cor a 14 genomic sequence.¹² On the basis of our comprehensive investigation on proteoform diversity, we were able to confirm the results from both previous studies and add further information to the pool of Cor a 14 heterogeneity. This includes an Asn/Ser exchange at position 56, C-terminal clipping of 1–3 residues at the large subunit, and minor N-terminal clipping of 1 to 2 residues at both subunits. Using a combination of three proteomics strategies, we hereby present a thoroughly validated consensus set of Cor a 14 proteoforms, by which we are able to characterize the clear majority (88–95% in each cultivar) of signals observed in the high-resolution intact mass analysis of the unreduced proteins, a technique that yields direct information on the native primary structure of each proteoform.²⁵ A minority of the observed proteoforms yet remains uncharacterized, and the detection of additional forms is likely to be expected using a more sensitive instrumentation.

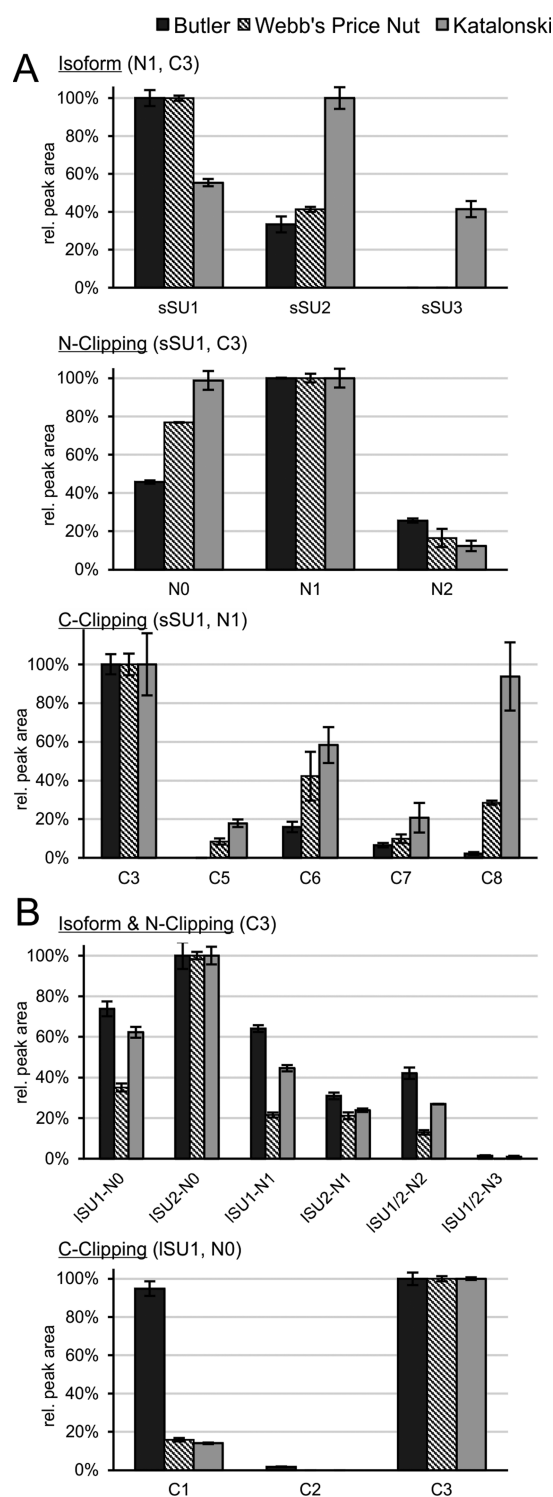


Figure 5. Relative quantification of the modifications described in Figure 4 (polymorphisms, N-/C-terminal clipping). Peak areas were determined from extracted ion traces of the small (part A) and large subunit (part B) and summed up over all observable charge states. The maximum value of each chart was defined as 100%. Error bars symbolize standard deviation from the analysis of three biological replicates. For an analogous representation of the N-/C-terminal clipping data of isoforms sSU2, sSU3, and ISU2 and for the tabular data the reader is referred to Figure S-1 and Tables S-1 and S-2 in the Supporting Information.

Although the primary protein sequence predicted by Garino and coauthors¹² from the cloned Cor a 14 gene (GenBank ID FJ358504) could be confirmed, we found conflicting evidence of the position of the cleavage site between the linker peptide and the small subunit, indicating that the mature Cor a 14 sequence is in fact four amino acids longer at the N-terminus than previously reported (not considering terminal clipping). This is also in agreement with information available on the cleavage sites of the 2S albumins from mustard,²¹ brazil nut,³⁴ walnut,^{35,36} sesame,³⁷ and sunflower,²³ which are all located after an asparagine residue. Furthermore, we demonstrated that clipping at the N- and C-termini may lead to a significant truncation, particularly of the small subunit. Together with the observed polymorphisms, our findings suggest that the original putative protein sequence, which was recently used in the large European multicenter study on food allergies, EuroPrevall, for in vitro component-resolved diagnosis,⁵ does not properly fit the structure of naturally occurring Cor a 14. We therefore recommend that this novel information be considered for future applications and production of recombinant allergens.

Compared with the information available on other 2S albumins, Cor a 14 follows the common structural features of the protein family. Like most 2S albumins, it is post-translationally cleaved into two disulfide-bridged subunits. It does, however, not carry a linker peptide of significant length between the small and large subunits, which in several other 2S albumins is excised from the proprotein during post-translational processing.^{15,21,30,37} Three single amino acid mutations, two of the small and one of the large subunit, were identified. Such sequence polymorphisms have been described for 2S albumins from other plants before^{21,38,39} and have been traced back to a multigene family.¹⁴ Notably, all amino acid exchanges observed in our study can be interpreted through single nucleotide polymorphisms in the genomic sequence of Cor a 14, as demonstrated in Figure 8. There is no evidence of genomic mutations (like insertions or deletions) that may cause a reading-frame shift and therefore lead to extreme differences in the protein's primary sequence, as it was described for the isoforms of the peanut 2S albumin Ara h 7.⁴⁰ Interestingly, subunits carrying polymorphisms are paired specifically on the native protein level, as we were able to show through the combination of our results from the intact mass analyses of the protein subunits and of the unreduced proteins, respectively. Thus, at least four copies of the Cor a 14 gene must be assumed in the hazelnut genome to explain the observed proteoform diversity.

Many previous studies on 2S albumins have reported clipping at either the N- or the C-terminus in varying degrees from one up to seven amino acids.^{15,22,29–32} Recently, our group has shown that this modification does not apply to all members of the protein family, as no such clipping could be observed for the mustard 2S albumin Sin a 1.²¹ The clipping observed for Cor a 14 in our study, leading to highly abundant proteoforms counting up to eight missing residues, must therefore be rated as considerably pronounced in comparison to other 2S albumins. Basically, three different origins have been discussed for the formation of terminal microheterogeneity: different genetically derived precursors, a shift in the position of the cleavage site during protein maturation, or the presence of exopeptidases in the protein bodies of the seeds.^{14,29} Recently, C-terminal microheterogeneity was observed for the albumin SESA2-1 from sunflower seeds and attributed to differential endopeptidase-processing, as SESA2-1

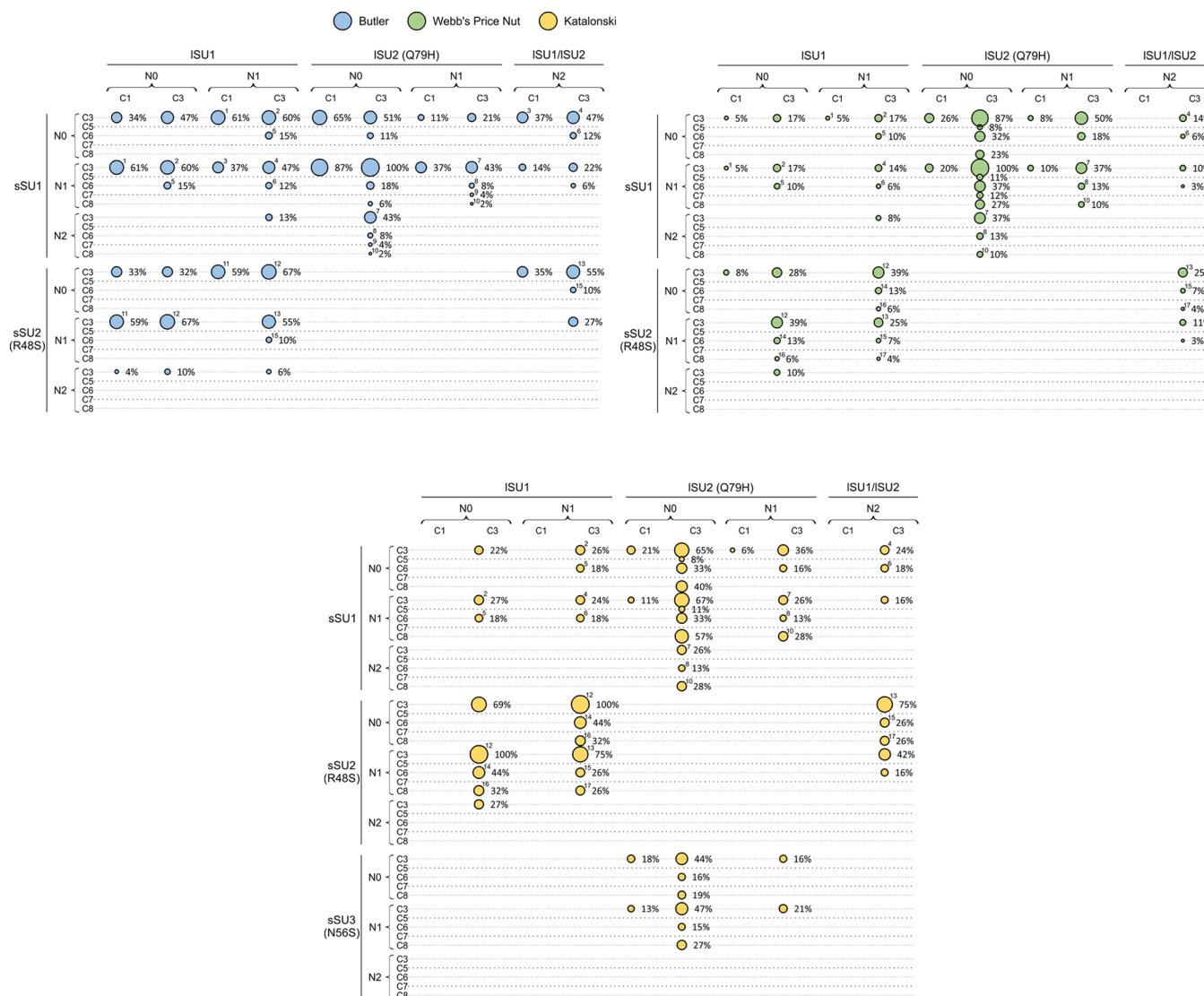


Figure 6. Proteoforms of Cor a 14, identified by intact mass analysis of the unreduced proteins and characterized through pairing of the small and large subunit. Bubble size and percentage values indicate the relative fraction of proteoforms, determined from peak areas of extracted ion traces summed up over all observable charge states (7–11). The maximum value of each cultivar was defined as 100%. Superscript numbers indicate isobaric proteoforms. For the tabular data the reader is referred to Table S-3 in the [Supporting Information](#).

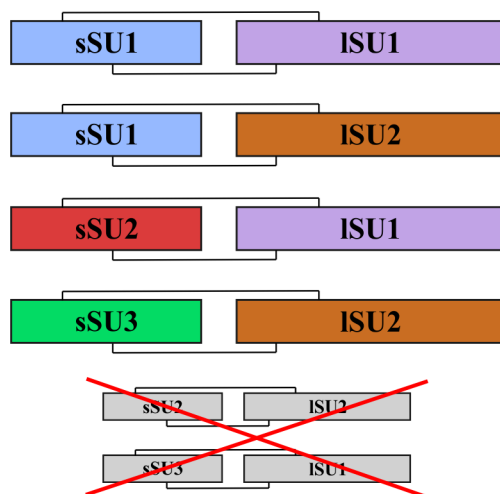


Figure 7. Pairing of small and large subunit proteoforms differing in sequence polymorphisms, as observed in intact mass analysis.

exhibits only two such proteoforms differing in the cleavage site of a Thr-Thr-Thr motif.²³ Considering the high diversity of differently clipped proteoforms in the case of Cor a 14, the variable residues at the cleavage sites, and the rather statistical pairing of N- and C-terminal clipping and polymorphisms, there is evidence of an unspecific post-translational modification through the activity of amino- and carboxypeptidases, yet single C- or N-clipped proteoforms could also be of genetic origin, taking into account that Garino and coauthors reported a copy of the Cor a 14 gene carrying a gap at amino acid position 79, which, in our study, would have been identified as clipping.¹² It must be, furthermore, noted that the degree of processing caused by exopeptidases may depend on the duration and conditions of storage of the nuts prior to analysis. However, no substantial enzymatic activity is assumed under the applied conditions (storage at -20°C for 3 to 4 years before nuts were processed and measured). Taking a closer look at the polymorphisms presented in [Figure 4](#), notably none is located within the assumed “hypervariable region”, a shared structural element of 2S albumins, commonly described to

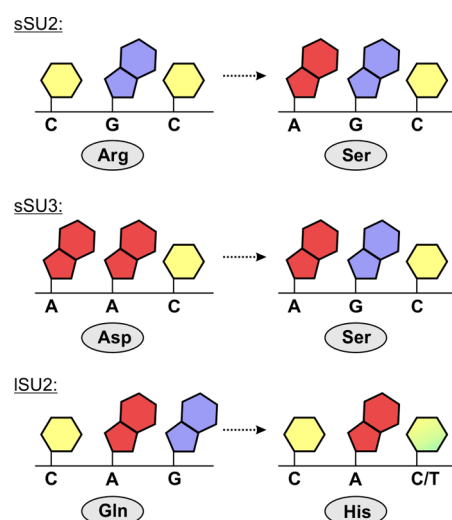


Figure 8. Single nucleotide polymorphisms explaining the heterogeneity observed on the level of the Cor a 14 protein sequence (Cor a 14 genomic sequence taken from GenBank ID FJ358504).

feature a high variability and to represent the most important immunogenic region.^{14,41} Also, the putative IgE-epitope RGEExxE at positions 115–121, which was identified in the mutational epitope analysis of the homologous walnut 2S albumin Jug r 1,³⁶ is not affected by either type of the observed modifications. We can therefore not conclude a difference in allergenicity of the identified proteoforms on the basis of the available data. However, no systematic epitope mapping survey has been reported for Cor a 14 so far, and such a study would probably reveal additional epitopes along the sequences of both subunits, comparable to the 2S albumins from peanut,^{42,43} cashew,⁴⁴ and brazil nut.^{45,46} We also expect that the application of our approach on additional hazelnut cultivars would probably result in the identification of more polymorphisms.

To date, only little information is available on the difference in allergen profiles between hazelnut cultivars. Recently, Garino and coauthors reported differences in the overall expression of Cor a 8 and Cor a 11 between four hazelnut cultivars on the level of mRNA transcripts using real-time PCR.⁴⁷ The study did, however, not analyze Cor a 14 and did also not include the cultivars from our study, for which we found distinct differences in the proteoform patterns. Both overall expression (analyzed as protein concentration) and sequence modifications may influence the allergenicity of hazelnut or hazelnut-processed foods, and further investigations should therefore be made on the allergen and proteoform profiles of different cultivars. Because only one batch of each cultivar was available for our study, we suggest that the relative quantification of proteoforms be repeated with samples from different areas and harvest years to clarify if the observed differences, in particular, the differential activity of amino- and carboxypeptidases, are reproducible.

5. CONCLUSIONS

We present here a comprehensive survey on the proteoforms of the hazelnut allergen Cor a 14. Our study demonstrates a significantly higher heterogeneity than previously expected. This data set constitutes, to the best of our knowledge, the largest set of proteoforms so far reported for a 2S albumin. Moreover, evidence is provided that the naturally occurring Cor

a 14 deviates from the originally proposed protein sequence, which is commonly used for component resolved diagnosis. For the first time, we report varying proteoform patterns between cultivars of an allergenic seed and provide data on the relative abundance of proteoforms in each cultivar. Together, these data present a basis for future studies on the allergenicity of Cor a 14 in hazelnuts, which should involve epitope mapping, proteoform investigations in different cultivars, and immunological studies on IgE-affinity and thus allergenicity of individual proteoforms.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00924.

Figure S-1: Relative quantification of N-/C-terminal clipping in isoforms sSU2, sSU3, and ISU2. Figure S-2: Fragment spectrum of tryptic peptide G44-R53, indicating the R48S polymorphism. Figure S-3: Fragment spectrum of tryptic peptide Q54-R62, indicating the N56S polymorphism. Figure S-4: Fragment spectrum of tryptic peptide H79-R92, indicating the Q79H polymorphism. (PDF)

Table S-1: Relative peak areas of small subunit proteoforms. Table S-2: Relative peak areas of large subunit proteoforms. Table S-3: Relative peak areas of Cor a 14 proteoforms. Table S-4: Peptide sequences assigned to Cor a 14 identified by bottom-up approach. Table S-5: Identification of proteoforms by intact mass analysis. (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jens.brockmeyer@lc.uni-stuttgart.de. Tel: 0049-711-685-64359.

ORCID

Robin Korte: 0000-0002-0442-6020

Author Contributions

The research strategy of this project was devised by R.K. and J.B. Experiments were performed by R.K., J.H., and I.B., and data analysis and interpretation were performed by the contribution of all authors. R.K. and J.B. have written the manuscript. All authors provided input to drafting and revising the manuscript and gave approval to the final version.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Robin Korte received a Ph.D. scholarship from the German National Academic Foundation (Studienstiftung des Deutschen Volkes).

■ ABBREVIATIONS

HRMS, high-resolution mass spectrometry; MW, molecular weight; ER, endoplasmic reticulum; MeOH, methanol; FA, formic acid; ACN, acetonitrile; ESI, electrospray ionization; PTM, post-translational modification; FDR, false discovery rate; IgE, immunoglobulin E

REFERENCES

- (1) Burney, P. G. J.; Potts, J.; Kummeling, I.; Mills, E. N. C.; Clausen, M.; Dubakiene, R.; Barreales, L.; Fernandez-Perez, C.; Fernandez-Rivas, M.; Le, T.-M.; Knulst, A. C.; Kowalski, M. L.; Lidholm, J.; Ballmer-Weber, B. K.; Braun-Fahlander, C.; Mustakov, T.; Kralimarkova, T.; Popov, T.; Sakellariou, A.; Papadopoulos, N. G.; Versteeg, S. A.; Zuidmeer, L.; Akkerdaas, J. H.; Hoffmann-Sommergruber, K.; van Ree, R. The prevalence and distribution of food sensitization in European adults. *Allergy* **2014**, *69*, 365–371.
- (2) McWilliam, V.; Koplin, J.; Lodge, C.; Tang, M.; Dharmage, S.; Allen, K. The Prevalence of Tree Nut Allergy: A Systematic Review. *Curr. Allergy Asthma Rep.* **2015**, *15*, 1–13.
- (3) Breiteneder, H.; Ferreira, F.; Hoffmann-Sommergruber, K.; Ebner, C.; Breitenbach, M.; Rumpold, H.; Kraft, D.; Scheiner, O. Four recombinant isoforms of Cor a I, the major allergen of hazel pollen, show different IgE-binding properties. *Eur. J. Biochem.* **1993**, *212*, 355–362.
- (4) Jimenez-Lopez, J. C.; Morales, S.; Castro, A. J.; Volkmann, D.; Rodríguez-García, M. I.; Alché, J. D. D. Characterization of Profilin Polymorphism in Pollen with a Focus on Multifunctionality. *PLoS One* **2012**, *7*, e30878.
- (5) Datema, M. R.; Zuidmeer-jongejan, L.; Asero, R.; Barreales, L.; Belohlavkova, S.; De Blay, F.; Bures, P.; Clausen, M.; Dubakiene, R.; Gislason, D.; Jedrzejczak-Czechowicz, M.; Kowalski, M. L.; Knulst, A. C.; Kralimarkova, T.; Le, T.-M.; Lovegrove, A.; Marsh, J.; Papadopoulos, N. G.; Popov, T.; Del Prado, N.; Purohit, A.; Reese, G.; Reig, I.; Seneviratne, S. L.; Sinaniotis, A.; Versteeg, S. A.; Vieths, S.; Zwinderman, A. H.; Mills, C.; Lidholm, J.; Hoffmann-Sommergruber, K.; Fernández-Rivas, M.; Ballmer-Weber, B.; van Ree, R. Van. Hazelnut allergy across Europe dissected molecularly: A EuroPrevall outpatient clinic survey. *J. Allergy Clin. Immunol.* **2015**, *136*, 382–391.
- (6) Vieths, S.; Scheurer, S.; Ballmer-Weber, B. Current Understanding of Cross-Reactivity of Food Allergens and Pollen. *Ann. N. Y. Acad. Sci.* **2002**, *964*, 47–68.
- (7) Schocker, F.; Lüttkopf, D.; Scheurer, S.; Petersen, A.; Cisteró-Bahima, A.; Enrique, E.; San Miguel-Moncín, M.; Akkerdaas, J.; van Ree, R.; Vieths, S. Recombinant lipid transfer protein Cor a 8 from hazelnut: A new tool for in vitro diagnosis of potentially severe hazelnut allergy. *J. Allergy Clin. Immunol.* **2004**, *113*, 141–147.
- (8) Beyer, K.; Grishina, G.; Bardina, L.; Grishin, A.; Sampson, H. a. Identification of an 11S globulin as a major hazelnut food allergen in hazelnut-induced systemic reactions. *J. Allergy Clin. Immunol.* **2002**, *110*, 517–523.
- (9) Lauer, I.; Foetisch, K.; Kolarich, D.; Ballmer-Weber, B. K.; Conti, A.; Altmann, F.; Vieths, S.; Scheurer, S. Hazelnut (*Corylus avellana*) vicilin Cor a 11: molecular characterization of a glycoprotein and its allergenic activity. *Biochem. J.* **2004**, *383*, 327–334.
- (10) Akkerdaas, J. H.; Schocker, F.; Vieths, S.; Versteeg, S.; Zuidmeer, L.; Hefle, S. L.; Aalberse, R. C.; Richter, K.; Ferreira, F.; van Ree, R. Cloning of oleosin, a putative new hazelnut allergen, using a hazelnut cDNA library. *Mol. Nutr. Food Res.* **2006**, *50*, 18–23.
- (11) Zuidmeer-Jongejan, L.; Fernández-Rivas, M.; Winter, M. G.; Akkerdaas, J. H.; Summers, C.; Lebens, A.; Knulst, A. C.; Schilte, P.; Briza, P.; Gadermaier, G.; van Ree, R. Oil body-associated hazelnut allergens including oleosins are underrepresented in diagnostic extracts but associated with severe symptoms. *Clin. Transl. Allergy* **2014**, *4*, 4.
- (12) Garino, C.; Zuidmeer, L.; Marsh, J.; Lovegrove, A.; Morati, M.; Versteeg, S.; Schilte, P.; Shewry, P.; Arlorio, M.; van Ree, R. Isolation, cloning, and characterization of the 2S albumin: a new allergen from hazelnut. *Mol. Nutr. Food Res.* **2010**, *54*, 1257–1265.
- (13) Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 821–830.
- (14) Moreno, F. J.; Clemente, A. 2S Albumin Storage Proteins: What Makes them Food Allergens? *Open Biochem. J.* **2008**, *2*, 16–28.
- (15) 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.). *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, *1752*, 142–153.
- (16) Berecz, B.; Clare Mills, E. N.; Paradi, I.; Lang, F.; Tamas, L.; Shewry, P. R.; MacKie, A. R. Stability of sunflower 2S albumins and LTP to physiologically relevant in vitro gastrointestinal digestion. *Food Chem.* **2013**, *138*, 2374–2381.
- (17) Mattison, C. P.; Grimm, C. C.; Wasserman, R. L. In vitro digestion of soluble cashew proteins and characterization of surviving IgE-reactive peptides. *Mol. Nutr. Food Res.* **2014**, *58*, 884–893.
- (18) Hazebrouck, S.; Guillon, B.; Drumare, M. F.; Paty, E.; Wal, J. M.; Bernard, H. Trypsin resistance of the major peanut allergen Ara h 6 and allergenicity of the digestion products are abolished after selective disruption of disulfide bonds. *Mol. Nutr. Food Res.* **2012**, *56*, 548–557.
- (19) Bollini, R.; Chrispeels, M. J. The Rough Endoplasmic Reticulum Is the Site of Reserve-Protein Synthesis in Developing Phaseolus vulgaris Cotyledons. *Planta* **1979**, *146*, 487–501.
- (20) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed Storage Proteins: Structures and Biosynthesis. *Plant Cell* **1995**, *7*, 945–956.
- (21) Hummel, M.; Wigger, T.; Brockmeyer, J. Characterization of Mustard 2S Albumin Allergens by Bottom-up, Middle-down, and Top-down Proteomics: A Consensus Set of Isoforms of Sin a 1. *J. Proteome Res.* **2015**, *14*, 1547–1556.
- (22) Murén, E.; Ek, B.; Rask, L. Processing of the 2S storage protein pronapin in *Brassica napus* and in transformed tobacco. *Eur. J. Biochem.* **1995**, *227*, 316–321.
- (23) Franke, B.; Colgrave, M. L.; Mylne, J. S.; Rosengren, K. J. Mature forms of the major seed storage albumins in sunflower: A mass spectrometric approach. *J. Proteomics* **2016**, *147*, 177–186.
- (24) Pfeifer, S.; Bublin, M.; Dubiela, P.; Hummel, K.; Wortmann, J.; Hofer, G.; Keller, W.; Radauer, C.; Hoffmann-Sommergruber, K. Cor a 14, the allergenic 2S albumin from hazelnut, is highly thermostable and resistant to gastrointestinal digestion. *Mol. Nutr. Food Res.* **2015**, *59*, 2077–2086.
- (25) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–521.
- (26) Schagger, H. Tricine-SDS-PAGE. *Nat. Protoc.* **2006**, *1*, 16–22.
- (27) 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.
- (28) 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–M111.010587.
- (29) Moreno, F. J.; Jenkins, J. A.; Mellon, F. A.; Rigby, N. M.; Robertson, J. A.; Wellner, N.; Mills, E. N. C. *Biochim. Biophys. Acta, Proteins Proteomics* **2004**, *1698*, 175–186.
- (30) da Silva, J. G., Jr.; Machado, O. L. T.; Izumi, C.; Padovan, J. C.; Chait, B. T.; Mirza, U. A.; Greene, L. J. Amino Acid Sequence of a New 2S Albumin from *Ricinus communis* Which Is Part of a 29-kDa Precursor Protein. *Arch. Biochem. Biophys.* **1996**, *336*, 10–18.
- (31) Radosavljevic, J.; Dobrijevic, D.; Jadranin, M.; Blanus, M.; Vukmirica, J.; Cirkovic Velickovic, T. Insights into proteolytic processing of the major peanut allergen Ara h 2 by endogenous peanut proteases. *J. Sci. Food Agric.* **2010**, *90*, 1702–1708.
- (32) Gehrig, P. M.; Krzyzaniak, A.; Barciszewski, J.; Biemann, K. Mass spectrometric amino acid sequencing of a mixture of seed storage proteins (napin) from *Brassica napus*, products of a multigene family. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 3647–3652.
- (33) Purwaha, P.; Silva, L. P.; Hawke, D. H.; Weinstein, J. N.; Lorenzi, P. L. An Artifact in LC-MS/MS Measurement of Glutamine and Glutamic Acid: In-Source Cyclization to Pyroglutamic Acid. *Anal. Chem.* **2014**, *86*, 5633–5637.
- (34) Gander, E. S.; Holmstroem, K.-O.; De Paiva, G. R.; De Castro, L. A. B.; Carneiro, M.; Grossi de Sá, M.-F. Isolation, characterization and expression of a gene coding for a 2S albumin from *Bertholletia excelsa* (Brazil nut). *Plant Mol. Biol.* **1991**, *16*, 437–448.

- (35) Teuber, S. S.; Dandekar, A. M.; Peterson, W. R.; Sellers, C. L. Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut (*Juglans regia*), a major food allergen. *J. Allergy Clin. Immunol.* **1998**, *101*, 807–814.
- (36) 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.
- (37) Tai, S. S. K.; Lee, T. T. T.; Tsai, C. C. Y.; Yiu, T.-J.; Tzen, J. T. C. Expression pattern and deposition of three storage proteins, 11S globulin, 2S albumin and 7S globulin in maturing sesame seeds. *Plant Physiol. Biochem.* **2001**, *39*, 981–992.
- (38) Ampe, C.; Van Damme, J.; de Castro, L. A. B.; Sampaio, M. J. A. M.; van Montagu, M.; Vandekerckhove, J. The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). *Eur. J. Biochem.* **1986**, *159*, 597–601.
- (39) Bernard, H.; Mondoulet, L.; Drumare, M. F.; Paty, E.; Scheinmann, P.; Thäi, R.; Wal, J. M. Identification of a New Natural Ara h 6 Isoform and of Its Proteolytic Product as Major Allergens in Peanut. *J. Agric. Food Chem.* **2007**, *55*, 9663–9669.
- (40) Schmidt, H.; Krause, S.; Gelhaus, C.; Petersen, A.; Janssen, O.; Becker, W. Detection and Structural Characterization of Natural Ara h 7, the Third Peanut Allergen of the 2S Albumin Family. *J. Proteome Res.* **2010**, *9*, 3701–3709.
- (41) 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.
- (42) Stanley, J. S.; King, N.; Burks, A. W.; Huang, S. K.; Sampson, H.; Cockrell, G.; Helm, R. M.; West, C. M.; Bannon, G. A. Identification and Mutational Analysis of the Immunodominant IgE Binding Epitopes of the Major Peanut Allergen Ara h 2. *Arch. Biochem. Biophys.* **1997**, *342*, 244–253.
- (43) Shreffler, W. G.; Lencer, D. A.; Bardina, L.; Sampson, H. A. IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2. *J. Allergy Clin. Immunol.* **2005**, *116*, 893–899.
- (44) Robotham, J. M.; Wang, F.; Seamon, V.; Teuber, S. S.; Sathe, S. K.; Sampson, H. A.; Beyer, K.; Seavy, M.; Roux, K. H. Ana o 3, an important cashew nut (*Anacardium occidentale* L.) allergen of the 2S albumin family. *J. Allergy Clin. Immunol.* **2005**, *115*, 1284–1290.
- (45) Stickler, M.; Mucha, J.; Estell, D.; Power, S.; Harding, F. A Human Dendritic Cell–Based Method to Identify CD4⁺ T-Cell Epitopes in Potential Protein Allergens. *Environ. Health Perspect.* **2003**, *111*, 251–254.
- (46) Alcocer, M. J. C.; Murtagh, G. J.; Wilson, P. B.; Progijs, P.; Lin, J.; Archer, D. B. The Major Human Structural IgE Epitope of the Brazil Nut Allergen Ber e 1: A Chimaeric and Protein Microarray Approach. *J. Mol. Biol.* **2004**, *343*, 759–769.
- (47) Garino, C.; Locatelli, M.; Coisson, J. D.; D’Andrea, M.; Cereti, E.; Travaglia, F.; Arlorio, M. Gene transcription analysis of hazelnut (*Corylus avellana* L.) allergens Cor a 1, Cor a 8 and Cor a 11: a comparative study. *Int. J. Food Sci. Technol.* **2013**, *48*, 1208–1217.
- (48) Lehmann, K.; Schweimer, K.; Neudecker, P.; Rösch, P. Sequence-specific ¹H, ¹³C and ¹⁵N resonance assignments of Ara h 6, an allergenic 2S albumin from peanut. *J. Biomol. NMR* **2004**, *29*, 93–94.

Structural Characterization of the Allergenic 2S Albumin Cor a 14: Comparing Proteoform Patterns Across Hazelnut Cultivars

Robin Korte[†], Jana Happe[†], Ina Brümmer[§] and Jens Brockmeyer^{*†§}

[†]Institute of Food Chemistry, Westfälische Wilhelms-Universität Münster, Corrensstraße 45,
48149 Münster

[§]Analytical Food Chemistry, University of Stuttgart, Allmandring 5b, 70569 Stuttgart

Contents

Figure S-1: Relative quantification of N- and C-terminal clipping in isoforms sSU2, sSU3 and ISU2.

Figure S-2: Fragment spectrum of the tryptic peptide G44-R53, indicating the R48S polymorphism.

Figure S-3: Fragment spectrum of the tryptic peptide Q54-R62, indicating the N56S polymorphism.

Figure S-4: Fragment spectrum of the tryptic peptide H79-R92, indicating the Q79H polymorphism.

Table S-1: Relative peak areas of small subunit proteoforms.

Table S-2: Relative peak areas of large subunit proteoforms.

Table S-3: Relative peak areas of Cor a 14 proteoforms.

Table S-4: Peptide sequences assigned to Cor a 14 identified by bottom-up approach.

Table S-5: Identification of proteoforms by intact mass analysis.

(all tables stand available for download as excel spreadsheets)

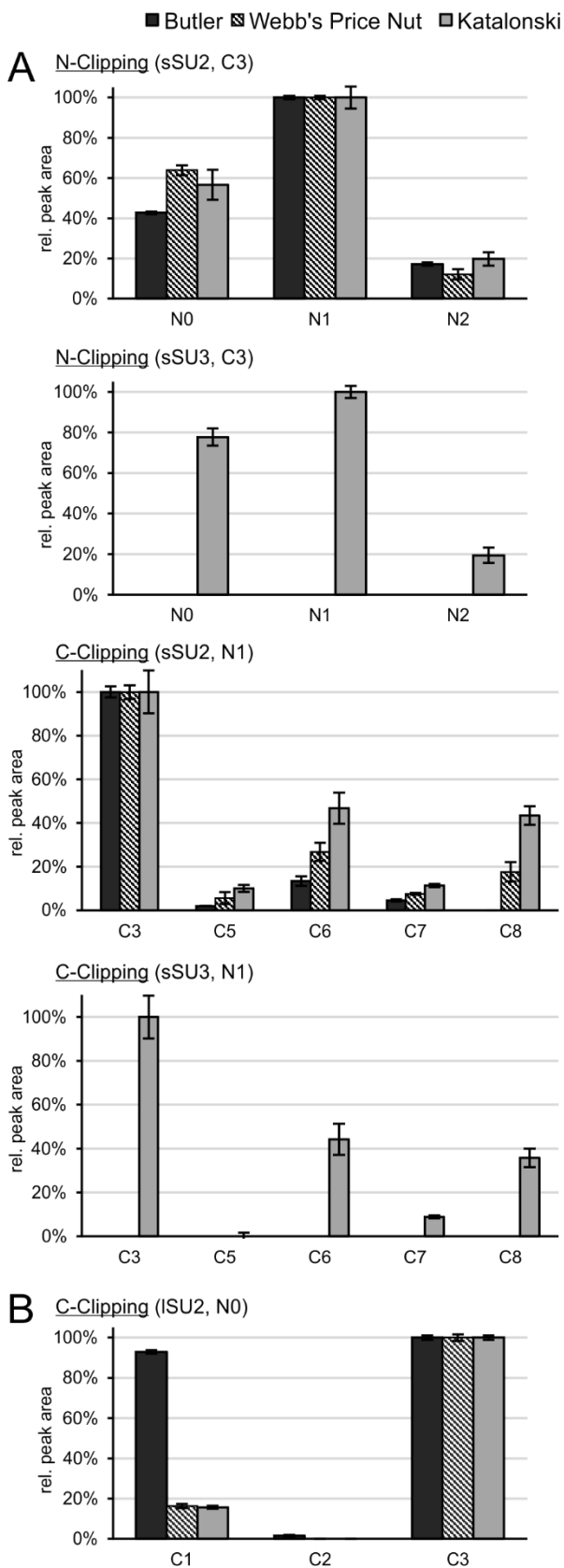


Figure S-1: Relative quantification of N- and C-terminal clipping in isoforms sSU2, sSU3 and lSU2. Peak areas were determined from extracted ion traces of the small (part A) and large subunit (part B) and summed up over all observable charge states. The maximum value of each chart was defined as 100%. Error bars symbolize standard deviation from the analysis of three biological replicates.

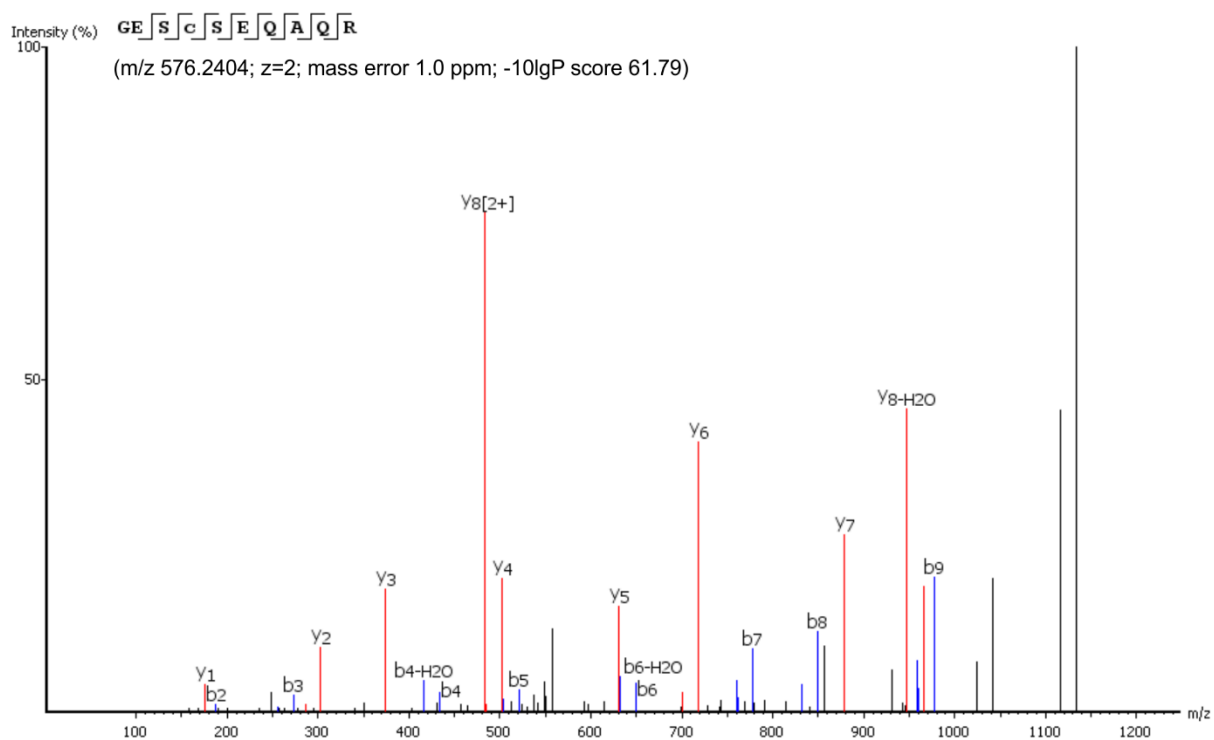


Figure S-2: Fragment spectrum of the tryptic peptide G44-R53, indicating the R48S polymorphism.

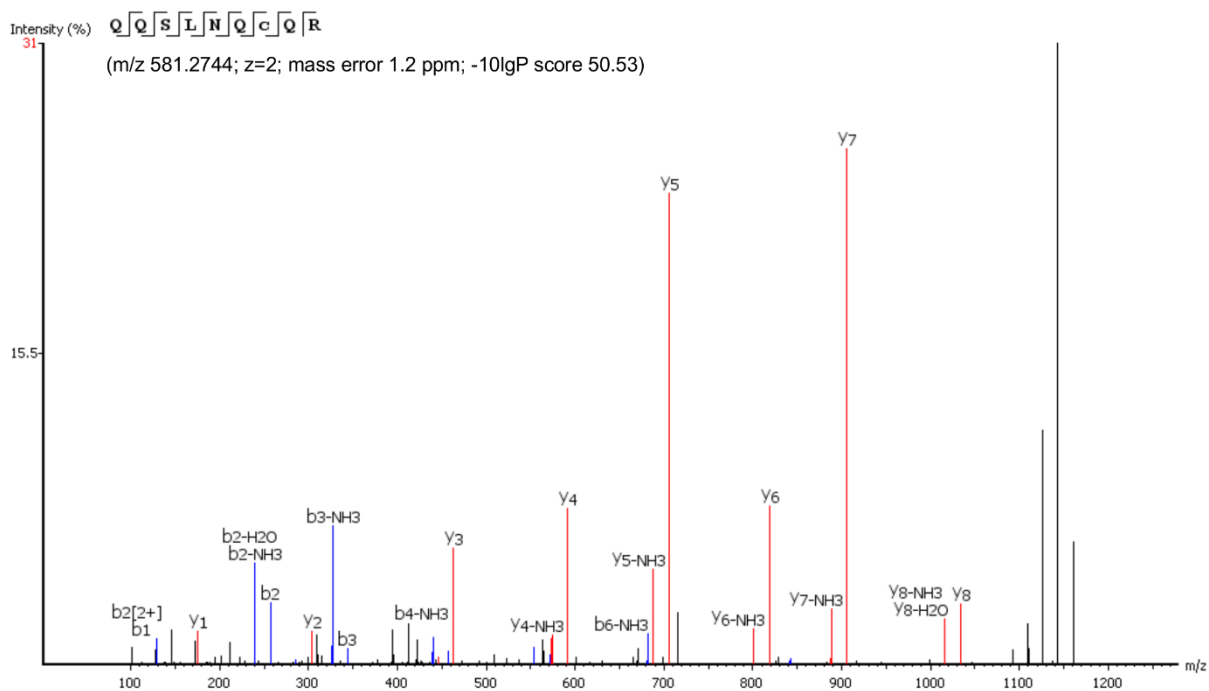


Figure S-3: Fragment spectrum of the tryptic peptide Q54-R62, indicating the N56S polymorphism.

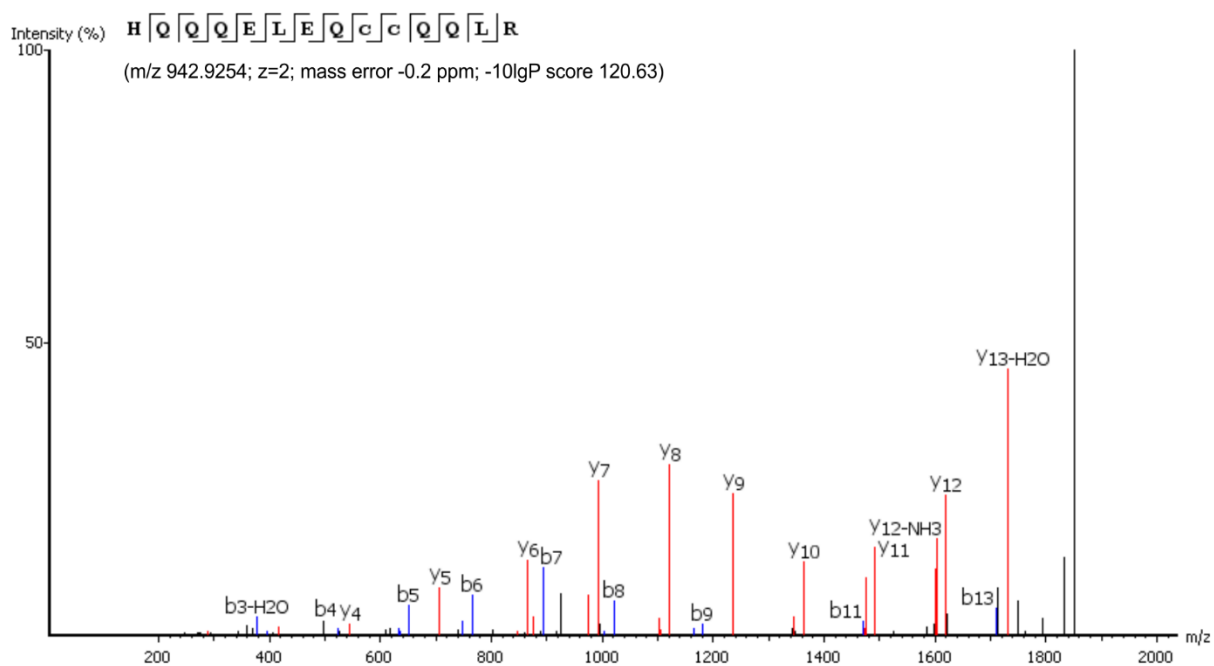


Figure S-4: Fragment spectrum of the tryptic peptide H79-R92, indicating the Q79H polymorphism.