Journal of Analysis and Testing
https://doi.org/10.1007/s41664-018-0064-6

ORIGINAL PAPER

Characterization of Structural Variability of the Allergenic 2S Albumin Ses i 1 Using Combinatorial Proteomics

Marlene Hummel[1] · Tina Wigger[1] · Jens Brockmeyer[2]

Received: 14 April 2018 / Accepted: 23 August 2018
© The Nonferrous Metals Society of China 2018

Abstract
Allergy to sesame seeds is a food allergy of high relevance due to its high abundance and the potential to elicit allergenic reactions, possibly resulting in anaphylaxis. The major sesame allergen Ses i 1 is a 2S albumin belonging to the family of seed storage proteins and despite its high allergenicity, comprehensive knowledge about Ses i 1 variants is still lacking. We, therefore, performed a detailed sequence analysis and characterized the C- and N-terminal ragged clipping of the small and large subunit of Ses i 1 using high-resolution mass spectrometry and the combination of bottom–up, middle–down, and top–down proteomic approaches. We detected extensive clipping at the C-terminus of the large (3–9 aa) and the small subunit (0–7 aa). In addition, the N-terminal conversion from glutamine to pyroglutamate and limited N-terminal clipping was confirmed for both subunits. Furthermore, we observed a sequence conflict at position 147 of Ses i 1 as well as a sequence shift of the large subunit compared to the reference sequence of the precursor.

Keywords Allergens · High-resolution mass spectrometry · Sesame · LC–MS · Ragged clipping · Ses i 1

1 Introduction

White sesame (*Sesamum indicum* L.) seeds are used worldwide for oil production and direct consumption, mainly as seasoning for different foods (bread, crackers and sweets). In addition, sesame oil is also used in pharmaceutical industry and cosmetics. The importance of sesame proteins as food allergens has been first reported in 2001 and it has been demonstrated that allergy to sesame can result in severe reactions such as anaphylaxis [1–3]. The seed storage protein Ses i 1 has been identified as the major allergen of sesame with a high frequency of reactions of sesame allergic patients against this allergen [1].

Ses i 1 belongs to the family of 2S albumins, that are widely distributed in seeds and tree nuts. 2S albumins are localized in protein bodies of the respective seed [4] and are synthesized from a single ~ 13 kDa precursor protein that is post-translational processed into a small subunit (3–4 kDa) and a large subunit (9–10 kDa). Both subunits are stabilized by four disulfide-bridges [5].

Studies investigating heterogeneity of 2S albumins demonstrate that these allergens are often highly polymorphic and that in general, several proteoforms are observed in the respective plant material [6, 7]. The occurrence of sequence polymorphisms is due to multiple genes encoding the respective 2S albumin and is highly pronounced in the hypervariable region that is in general neighboring epitope-binding sites. In addition, post-translational processing of the N- and C-terminal parts of both subunits can underlie 2S heterogeneity [8]. The clipping sites of termini can be heterogenic and results in variable length of the small and the large subunits leading to "ragged" termini. This ragged clipping has been observed for several 2S albumins, i.e., from rapeseed [9–11], brazil nut [12], castor oil seeds [13], hazelnut [14] and sesame [15]. Moreno et al. observed C-terminal clipping of the small subunit of Ses i 1 from white sesame seed as well as the modification of the N-terminal glutamines of both subunits to pyroglutamate [15]. Despite this important initial study, it is currently not clear, whether the C-terminus of the large subunit of Ses i 1 is also subjected to clipping and how the clipped subunits are combined in the intact protein. Using high-resolution mass spectrometry (HRMS) and

✉ Jens Brockmeyer
jens.brockmeyer@lc.uni-stuttgart.de

1 Institute of Food Chemistry, University of Muenster, Corrensstr. 45, 48149 Münster, Germany

2 Department of Food Chemistry, Institute of Biochemistry and Technical Biochemistry, University Stuttgart, Allmandring 5B, 70569 Stuttgart, Germany

Published online: 31 August 2018

the combination of bottom–up, middle–down and top–down strategies previously established in our group for this purpose [16], we aimed to provide a comprehensive overview of Ses i 1 sequence heterogeneity including C- and N-terminal clipping.

2 Materials and Methods

2.1 Materials and Reagents

White sesame (*Sesamum indicum* (L.)) seeds were purchased from a local supermarket, stored at − 20 ± 2 °C and were used within 3 months. Sodium acetate, tris(hydroxymethyl) aminomethane (TRIS), urea, 1,4-dithiothreitol and ammonium bicarbonate were purchased from Carl Roth (Karlsruhe, Germany). Iodoacetamide and acetonitrile were purchased from Sigma-Aldrich (Muenchen, Germany). Trypsin (sequencing grade) was purchased from Promega (Madison, USA).

2.2 Extraction of Sesame Proteins from Sesame Seeds

Due to its high fat content, sesame seeds were defatted prior analysis. The seeds were crushed in a knife mill (Grindomix GM 100, Retsch, Haan, Germany) for 2–3 s to obtain a coarse flour. The ten-fold amount of *n*-pentane was added to the flour and the slurry was stirred continuously for 2 h at room temperature, sieved and air-dried overnight. The pre-defatted, dry flour was subsequently milled to a fine powder in a nitrogen-cooled rotor-mill (Pulverisette, 1 mm mesh width, Fritsch, Idar Oberstein, Germany) and defatted a second time as described above. The defatted flour was stored at − 20 ± 2 °C and used within 3 months.

Sesame proteins were extracted from the defatted flour by vigorous shaking (Mini shaker, VWR, 550 rpm) for 2 h in a five-fold excess of sodium acetate buffer (100 mM, pH 3.6) or Tris buffer (100 mM, pH 8.0) at room temperature. Subsequently, the slurry was centrifuged at 9000×*g* (Universal 16, Hettich, Spenge, Germany) for 20 min at 23 °C and the supernatant was 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 sesame protein extract was stored at − 20 °C.

2.3 Size Exclusion Chromatography

The isolation of the 2S albumin fraction from the sesame protein extract was performed by size exclusion chromatography (SEC) using an ÄKTA purifier FPLC system (GE Healthcare, Muenchen, Germany) and Superdex 200 10/300 GL column (GE Healthcare, Muenchen, Germany). The chromatographic separation was carried out with 100 mM Tris, pH 8.0 and a flowrate of 300 µL/min. Relevant fractions were collected, precipitated with a five-fold volume of ice-cold acetone and analyzed by SDS-PAGE. The identity and purity of sesame 2S albumins was confirmed by in-solution digest and subsequent LC-HRMS analysis. Identification of Ses i 1 was carried out on the basis of UniProt KB sequence Q9AUD1.

2.4 Tryptic Digest and Peptide Desalting

Acetone-precipitated pellets of relevant fractions containing about 1 mg of total protein were dissolved in 100 µL of 6 M urea solution and reduced with 200 mM dithiotreitol. To prevent re-oxidation of reduced cysteine residues, alkylation with 100 mM iodacetamide was carried out. The extract was diluted with 100 mM Tris–HCl buffer to a final urea concentration of 0.6 M, to allow tryptic digest. The digest was performed under slow shaking at 37 °C overnight after addition of 20 µg sequencing grade trypsin. The digest was stopped by adjusting the pH to < 6 and the peptide solution was desalted using Strata-X 33u polymeric reversed phase (RP) cartridges filled with 30 mg/mL C18 material according to the manual. Peptides were eluted from the cartridge with 70% MeOH/1% FA and the peptide solution was evaporated at 40 °C under a continuous nitrogen flow. Peptides were resuspended in solvent A (3% acetonitrile (ACN)/1% FA). The solution was used for subsequent MS analysis.

2.5 MS Analysis of Tryptic Peptides

The analysis of tryptic peptides was carried out by a separation step on a C18 HPLC column combined with subsequent high-resolution MS (HRMS) analysis. The samples were applied to an Accucore C18 column (Thermo Scientific, Dreieich, Germany) and peptides were separated with the following 120 min gradient: 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 H_2O, $B = 1\%$ FA in ACN). The flow rate was set to 250 µL/min. The MS analysis was performed on an LTQ Orbitrap XL Mass Spectrometer (Thermo Scientific, Dreieich, Germany) in positive ionization mode. Data acquisition was performed in both full MS mode and in the data dependent mode (DDMS). Survey full-scan MS spectra (m/z 300–1500) were acquired with a resolution of $r = 30{,}000$. The five most intense peptide ions in each survey scan were isolated and fragmented in the HCD cell (normalized collision energy of 35% and activation time of 100 ms). Fragment ions were detected with a resolution of $r = 7500$. Dynamic exclusion was activated and set to 60 s. 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.

Based on the MS analysis, a database search was performed using the software PEAKS (Bioinfor, Waterloo, Canada) including an enhanced target-decoy routine ("decoy fusion"). The database search was performed against a subset (*"Sesamum"*) of the UniProt KB database (203 search entries, no common contaminants). Database search parameters were set as follows: (a) trypsin as specific enzyme (two missed cleavages 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. PEAKS de novo sequencing and the PEAKS Spider algorithm were used to identify peptides with polymorphisms (insertion, deletion, mutation, PTM).

2.6 Middle–Down and Top–Down Approach

Acetone-precipitated pellets of relevant fractions from SEC fractionation were dissolved in 50 mM ammonium hydrogen carbonate buffer and without further preparation used for top–down analysis. For the middle–down approach (analysis of protein subunits), dissolved proteins (1–5 mg protein/mL) were reduced with 10 µL of 200 mM dithiothreitol for 30 min at 56 °C and, to prevent cysteine re-oxidation, alkylated with 10 µL of 200 mM iodacetamide for 30 min in the dark. The reduced and alkylated protein solution was diluted in a ratio of 1:4–1:100 with 5% (v/v) ACN/0.1% FA. The solution was used for subsequent MS analysis.

2.7 MS Analysis of Subunits and Intact Proteins

The analysis of subunits and intact proteins was carried out by a separation step on a C4 HPLC column combined with subsequent MS analysis using an LTQ Orbitrap system. The samples were applied to an Accucore C4 column (Thermo Scientific, Dreieich, Germany) and separated with the following 28 min gradient: 5% B for 1 min, 5–40% B in 16 min, 40–95% B in 3 min, 95% B for 2 min, 95–5% B in 1 min, 5% B for 5 min ($A = 0.1\%$ FA in H_2O, $B = 0.1\%$ FA in ACN). The flow rate was set to 400 µL/min. The mass spectrometrical analysis was performed on an LTQ Orbitrap XL Hybrid Ion Trap—Orbitrap Mass Spectrometer (Thermo Scientific, Dreieich, Germany) in positive ionization mode. Data acquisition was performed in full MS mode. Full MS spectra (m/z 500–2000) were acquired with a resolution of 30,000. 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.

To ensure highest mass accuracy we determined the monoisotopic mass for middle–down and top–down experiments according to the approach described by Hummel et al. [16].

3 Results and Discussion

We first analyzed the protein of interest on peptide level (bottom–up) with a special focus on sequence polymorphisms like insertion, deletion or point mutation possibly resulting in a change of amino acid sequence. In addition, we already identify at this step potential ragged clipping sites or sequence shifts. Next, we analyzed each subunit separately after reduction and alkylation and separation by liquid chromatography (middle–down), which allows the identification of clipping variants and confirms sequence shifts. To determine the combination of clipped subunit variants on intact protein level, we performed top–down proteomics.

3.1 Characterization of Protein Sequence Polymorphisms by Bottom–Up Analysis

The 2S albumin fraction of white sesame was isolated using size exclusion chromatography and the identity of the isolated proteins was confirmed by in-solution digest and subsequent LC-HRMS analysis. The UniProt KB reference sequence Q9AUD1 of Ses i 1 was used for data evaluation and potential polymorphisms were identified by PEAKS de novo sequencing and the PEAKS Spider algorithm. To ensure high reliability of results, seven biological replicates were analyzed, resulting in an average sequence coverage of 74% for the small subunit and 89% for the large subunit of Ses i 1, respectively. The summarized sequence coverage of the seven biological replicates is 93%. Unidentified Ses i 1 peptides were ≤ 4 amino acids in length and thus unlikely to identify by bottom–up MS. The overall sequence coverage of the large and small subunit, the processing sites of the precursor and observed sequence modifications are displayed in Fig. 1. Notably, in none of the seven biological replicates we observed point mutations, insertions or deletions compared to the reference sequence Q9AUD1, demonstrating that Ses i 1 is not encoded by multiple gene copies as described for other 2S albumins [14, 16]. Using bottom–up proteomics, the conversion of glutamine (*Q*) to pyroglutamate (pyroQ) at the N-terminus of the small and the large subunit was detected (Fig. 1), which has been described as a modification for different 2S albumins by several groups [11–13, 17, 18]. In addition, we detected clipped variants of the C-terminal peptides of the large subunit (see also and the middle–down approach for a more thorough analysis). Notably, we also

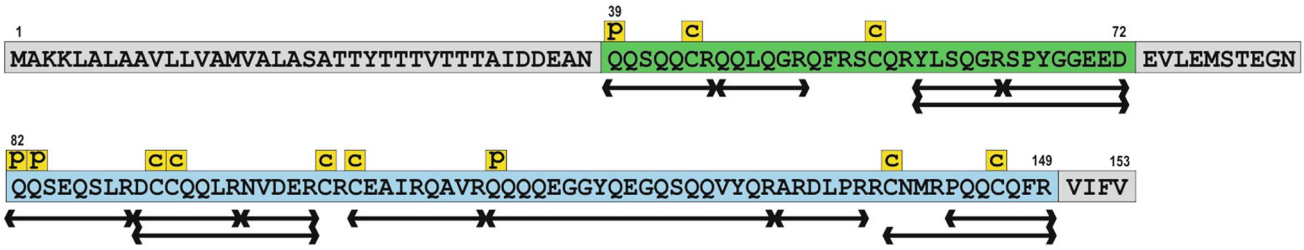


Fig. 1 Sequence coverage of Ses i 1 consisting of propeptide (M1–N38, shown in gray), small subunit (Q39–D72 shown in green), linker peptide (E73–N81, shown in gray) and large subunit (Q82–R149, shown in blue). The propeptide and the linker peptide are cleaved from the precursor upon post-translational processing resulting in the two subunits of the mature 2S albumin Ses i 1. Importantly, we observed a sequence shift in all of our samples compared to previously published Ses i 1 sequences. The C-terminal sequence VIFV (150–153, shown in gray) was not observed in any of our samples. Identification was carried out by in-solution digest, followed by LC–HRMS and database search using the software PEAKS. Arrows indicate sequence coverage by tryptic peptides. *c* carbamidomethylation, *p* pyroglutamate from *Q*. Small subunit is displayed in green, large subunit in blue

identified the sesame allergen Ses i 2 (Q9XHP1) and "2S albumin 3" (Q2XSW8) in our purified 2S fraction. The presence of these additional 2S albumins has been taken into account for further analysis.

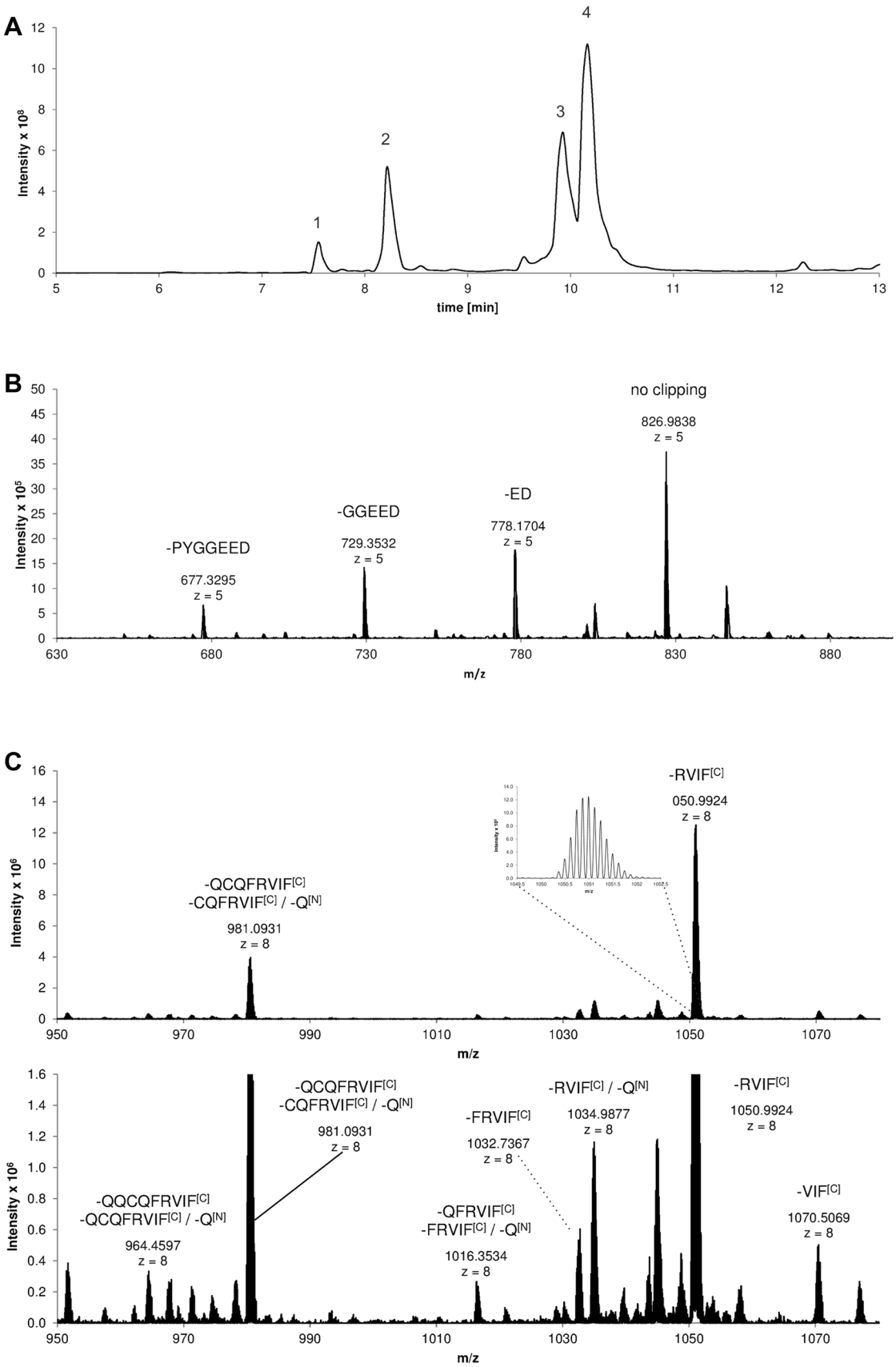
3.2 Identification and Characterization of C-/N-Terminal Clipping by Middle–Down Analysis

To identify C- and N-terminal clipping in more detail on subunit level, we reduced and alkylated the protein to the small and large subunit and performed LC-HRMS analysis. The chromatographic separation resulted in four major signals (peak 1–4, Fig. 2a). Peaks 1 and 2 were identified as small subunit isoforms while peak 3 and 4 belonged to the large subunit. In the corresponding mass spectra, signals of clipped variants in different charge states were observed. Three charge states of the different clipped variants were analyzed for the determination of exact subunit masses. The clipping of the small subunit is exemplified in Fig. 2b. In addition, we observed cyclization of the N-terminus (*Q*) to pyroglutamate also on middle–down level.

All signals were inspected manually and were deisotoped and deconvoluted. Monoisotopic masses were determined as described previously [16]. Masses that were detected at least three times with $\sigma < 0.01$ in the eight biological replicates of the middle–down approach were included in the further analysis. In total, we obtained 209 and 139 mass signals for the small and the large subunit, respectively. Signals were grouped when mass differences were below 0.04 u, resulting in 12 potential subunit masses for the small subunit and 17 potential masses for the large subunit. Taking a C-terminal clipping of up to seven amino acids for the small subunit and up to ten amino acids for the large subunit into account and N-terminal clipping of up to two amino acids for both subunits, a theoretical monoisotopic mass list of all possible clipping combinations was created. Finally, all identified middle–down masses were compared with these theoretical masses.

With a maximum error of 3.4 ppm we identified C- and N-terminal clipping at both subunits summarized in Table 1. Notably, only minor clipping was observed at the N-terminus of both subunits. Besides the unclipped form, we detected clipping of one amino acid (-*Q*) at each of the subunits. With one exception, the N-terminal glutamine was always modified to pyroglutamate. A more extensive clipping was observed at the C-terminus of both subunits. At the small subunit 0, 1, 2, 4, 5, 6 or 7 amino acids were clipped from the C-terminus, with four abundant clipping variants (0, 2, 5 and 7) that have been described previously [15] and three novel variants (1, 4, 6) (Fig. 2b). The large subunit showed C-terminal clipping of 3–9 amino acids relative to the reference sequence with highest abundance of the −RVIF variant (Fig. 2c). Notably, C-terminal clipping of the large subunit has not been described before and we did not detect the unclipped variant of the large subunit in any of our experiments. In addition, we were able to confirm the processing sites of the precursor at D^{72} and N^{81} as identified in the bottom–up approach.

Fig. 2 Middle–down analysis of Ses i 1. **A** Separation of small (Peak 1 + 2) and large (Peak 3 + 4) subunit by HPLC after reduction and alkylation. **b** ESI–HRMS spectrum of the small subunit (signals from Peak 1 + 2 are included). For clarity, only charge state $z = 5$ is displayed showing the four most abundant clipping variants and clipped amino acids As a result of reduction and alkylation Cys is modified to carboxymethylcysteine. **c** Upper lane: ESI–HRMS spectrum of the large subunit (signals from peak 3 + 4 are included). For clarity, only charge state $z = 8$ is shown. As a result of reduction and alkylation Cys is modified to carboxymethylcysteine. Lower lane: zoom in of the spectrum and designation of all clipping variants ▶


A
Intensity x 10^8
time [min]
1
2
3
4
B
Intensity x 10^5
m/z
-PYGGEED
677.3295
z = 5
-GGEED
729.3532
z = 5
-ED
778.1704
z = 5
no clipping
826.9838
z = 5
C
Intensity x 10^6
m/z
-QCQFRVIF[C]
-CQFRVIF[C] / -Q[N]
981.0931
z = 8
-RVIF[C]
050.9924
z = 8
-QQCQFRVIF[C]
-QCQFRVIF[C] / -Q[N]
964.4597
z = 8
-QCQFRVIF[C]
-CQFRVIF[C] / -Q[N]
981.0931
z = 8
-QFRVIF[C]
-FRVIF[C] / -Q[N]
1016.3534
z = 8
-FRVIF[C]
1032.7367
z = 8
-RVIF[C] / -Q[N]
1034.9877
z = 8
-RVIF[C]
1050.9924
z = 8
-VIF[C]
1070.5069
z = 8

3.3 Confirmation of C-Terminal Clipping by Middle–Down MS/MS Experiments

To further confirm clipping positions of subunits, we performed fragmentation experiments and the obtained spectra were analyzed in comparison with theoretical spectra generated using mass. Figure 3 shows fragment spectra of $[M+5H]^{5+}$ of the four most abundant C-terminal clipping variants of the small subunit. Fragmentation experiments confirmed the postulated clipping. For the unclipped small subunit, we detected several b33 ions ($b33^{4+}$, $[b33+H_2O]^{5+}$, $[b33-NH_3]^{5+}$) indicative for fragmentation at the native C-terminus between E^{71} and D^{72}. The −ED clipping variant consequently showed b31 ions at the C-terminus, while in the −GGEED and the −PYGGEED variants only b28 and b26 ions, respectively, were detectable for the C-terminus (Fig. 3). Despite variation of different MS-parameter, we were not able to obtain efficient fragmentation of the large subunit in HCD and CID mode.

3.4 Confirmation of C-/N-Terminal Clipping and Subunit Pairing by Top–Down Analysis

To confirm the identified clipping and delineate the combination of clipped subunits on protein level, LC-HRMS of intact Ses i 1 was performed. Seven biological replicates were analyzed and, again, three different charges states included in the data evaluation. Mass signals observed at least three times with $\sigma < 0.01$ were used for further characterization. In total, we identified 123 mass signals that were grouped when mass differences were below 0.04 u, resulting in eight possible variants for the intact protein. Taking all subunit variants observed by middle–down approach into account, a mass list of all combinations was created and compared to the included mass signals. All top–down masses were assigned with a maximum error of 1.88 ppm.

Notably, the eight Ses i 1 variants included five different C-terminal clipping variants of the small subunit, but only the most abundant C-terminal clipping variant (-RVIF) of the large subunit, although seven clipping variants of the large subunit have been identified by middle–down. We assume that only the variants with high abundance are detectable top–down and that further yet unidentified and less abundant clipped Ses i 1 variants occur. Concerning the N-terminal clipping, both clipping variants (-Q and no clipping) were identified for the large and the small subunit. Results of the top–down analysis and the identified clipping variants are given in Fig. 4.

Table 1 Clipping variants of small and large subunit identified by middle–down approach

Subunit	Monoisotopic obsM	Monoisotopic theoM	Error [ppm]	N-term. clipping	C-term. clipping
Small	4013.840	4013.836	1.08	No clipping[a]	No clipping
Small	3898.814	3898.809	1.20	No clipping[a]	−D
Small	3769.770	3769.766	0.95	No clipping[a]	−ED
Small	3641.716	3641.708	2.24	−Q[a]	−ED
Small	3583.706	3583.702	1.18	No clipping[a]	−GEED
Small	3526.684	3526.681	0.85	No clipping[a]	−GGEED
Small	3398.626	3398.622	1.06	−Q[a]	−GGEED
Small	3363.617	3363.617	0.24	No clipping[a]	−YGGEED
Small	3266.566	3266.565	0.37	No clipping[a]	−PYGGEED
Small	3138.509	3138.506	0.96	−Q[a]	−PYGGEED
Large	8208.865	8208.841	2.99	No clipping[a]	−VIF
Large	8069.756	8069.766	1.26	No clipping	−RVIF
Large	8052.756	8052.740	2.02	No clipping[a]	−RVIF
Large	7924.705	7924.681	2.96	−Q[a]	−RVIF
Large	7905.696	7905.671	3.16	No clipping[a]	−FRVIF
Large	7777.636[b]	7777.613	2.94	No clipping[a]	−QFRVIF
Large	7777.636[b]	7777.613	2.94	−Q[a]	−FRVIF
Large	7546.566[b]	7546.545	2.52	No clipping[a]	−QCQFRVIF
Large	7546.566[b]	7546.545	2.52	−Q[a]	−CQFRVIF
Large	7418.511[b]	7418.486	3.36	No clipping[a]	−QQCQFRVIF
Large	7418.511[b]	7418.486	3.36	−Q[a]	−QCQFRVIF

[a]Conversion from glutamate into pyroglutamate

[b]Different clipping variants resulting from identical masses, and therefore, not distinguishable

3.5 Structural Heterogeneity of Ses i 1

Prior to our study, structural aspects of Ses i 1 have been investigated in three publications [1, 15, 19] and available data already show a certain degree of heterogeneity especially in the large subunit. Interestingly, none of the published (partial) Ses i 1 sequences is entirely confirmed by our data, further underlining breed-specific differences and sequence heterogeneity. Published Ses i 1 sequences are summarized in Fig. 5 and are compared to our results.

The sequence of the small subunit has been described in three publications including ours and the sequence of the unclipped variant that is released from the precursor is identical in all studies (Q^{39}–D^{72}). In addition, we were able to confirm published C-terminal clipping variants of the small subunit and identified three additional and less abundant clipping variants.

Results concerning the structure of the large subunit are more divergent. Moreno et al. postulated a large subunit ranging from Q^{86} to F^{152} (67 aa) of the precursor sequence [15]. Using bottom–up and middle–down approaches, we identified a large subunit comprising of 68 aa. Importantly, the N-terminus of the unclipped large subunit in our study shows a sequence shift of four amino acids (and 3 aa at the C-terminus), thus covering Q^{82}–R^{149} of the precursor sequence. The structure of the large subunit was identified with high probability as we identified the relevant peptide ^{82}QQSEQSLR^{89} in all seven replicates and in four of seven samples with a $-10\lg P > 47$ (55.78 on average) and confirmed this sequence in the middle–down approach (see Table 1). In addition, we did not find an indication in any of our experiments that the C-terminal amino acids ^{150}VIFV^{153} of the precursor are part of the large subunit.

Available literature indicates that the processing site for the release of the N-terminus of the large subunit from the precursor is highly specific. Hara-Nishimura et al. identified a protease in *Ricinus communis* involved in release of the large subunit from the precursor by cleavage C-terminal of Asn [20]. Notably, this protease also mediated cleavage with identical specificity for a variety of seed storage proteins from other plants and a homologue protease with identical specificity was identified in soy (*Glycine max*) [21, 22], indicating a highly conserved N-terminal processing site for the large subunits of 2S albumins. The release of the large subunit C-terminal to Asn was confirmed for a number of 2S albumins from different species [11, 13, 16, 23–25]. The processing site for the large subunit of Ses i 1 identified in this study is also C-terminal to Asn, while the sequence proposed by Moreno et al. starts C-terminal to Glu [15].

In addition, Moreno et al. and Pastorello et al. [1, 15] proposed a point mutation at position 147 (Q147A) relative to the reference sequence Q9AUD1, which we did not detect in any of our samples. The corresponding peptides ^{143}PQQCQFR^{149}, ^{139}CNMRPQQCQFR^{149} and

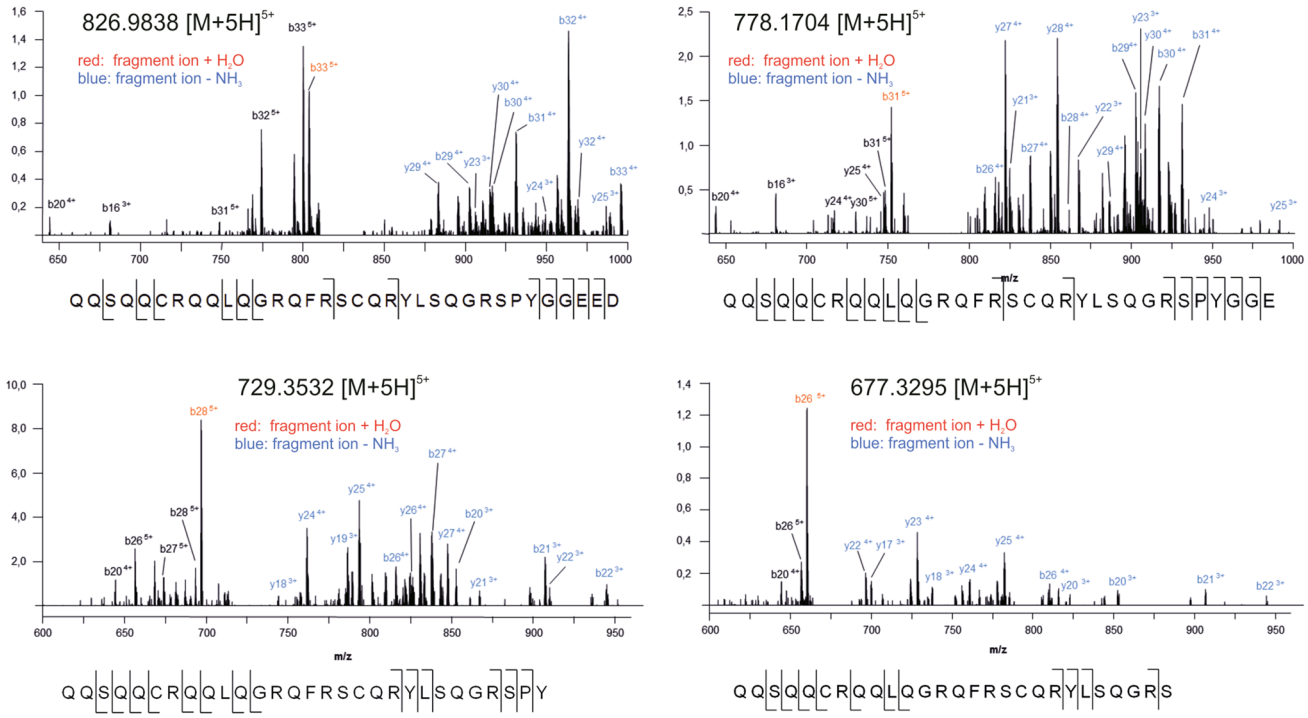


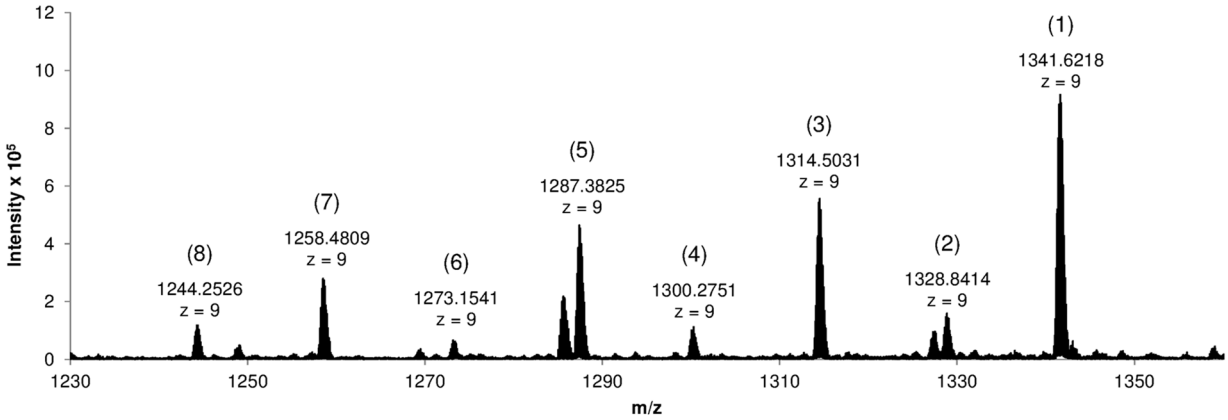
Fig. 3 Confirmation of clipping variants by ESI–HRMS/MS fragmentation experiments. Fragmentation of four clipping variants of the small subunit are exemplified. All precursor ions have $z = 5$, m/z of precursors are given in the respective spectra. Identified b-, and y-ions are highlighted and the sequence coverage of the clipping variant is given below the spectrum

^{138}RCNMRPQQCQFR^{149} containing the mentioned amino acid Q^{147} have been identified seven times in six different samples with an average − 10lgP score of 80.1. In addition, we confirmed Q^{147} in several middle–down and top–down analyses with high confidence (see Table 1 and 2). Together, these findings indicate a certain degree of breed-specific differences at position 147.

4 Conclusion

In this study we present a comprehensive overview of Ses i 1 structural heterogeneity by HRMS using a combination of bottom–up, middle–down and top–down proteomics. The existence of numerous sequence polymorphisms and the occurrence of a "hypervariable region" due to the existence of multigene families is a phenomenon that has been described for several 2S albumins such as mustard [16], hazelnut [14], brazil nut [12] and rapeseed [17]. In contrast to that, we did not identify any polymorphisms for Ses i 1 except a sequence conflict at position 147 that most probably reflects breed-specific differences in sesame. One of our major results is the identification and characterization of extensive clipping at both C-termini of Ses i 1 resulting in eight major protein variants that are first described on protein level in this study. Though in principle ragged clipping can underlie the presence of different precursor proteins, we hypothesize that plant-internal carboxypeptidases with limited specificity mediate modification of C-termini of the small and large Ses i 1 subunits. This hypothesis is supported by the fact, that we did not identify Ses i 1 sequence polymorphisms indicating that only one precursor peptide is present in sesame. In addition, available literature indicates the presence of carboxy- and aminopeptidases that mediate ragged clipping [10, 23, 26]. The isolation and characterization of plant-internal carboxypeptidases might, therefore, be a promising approach for further studies.

Whether the different clipping variants have influence on the allergenicity of Ses i 1 requires further investigation. To the best of our knowledge, no positional information is available for linear or conformational epitopes. Assuming that the



clipping variant	small subunit	large subunit
1	Q^{P}QSQQCR___SQGRSPYGGEED	Q^{P}QSEQSL___QQCQF
2	Q^{P}QSQQCR___SQGRSPYGGEE	Q^{P}QSEQSL___QQCQF
3	Q^{P}QSQQCR___SQGRSPYGGE	Q^{P}QSEQSL___QQCQF
4 var1	Q^{P}QSQQCR___SQGRSPYGGE	Q^{P}SEQSL___QQCQF
4 var2	Q^{P}SQQCR___SQGRSPYGGE	Q^{P}QSEQSL___QQCQF
5	Q^{P}QSQQCR___SQGRSPY	Q^{P}QSEQSL___QQCQF
6 var1	Q^{P}QSQQCR___SQGRSPY	Q^{P}SEQSL___QQCQF
6 var2	Q^{P}SQQCR___SQGRSPY	Q^{P}QSEQSL___QQCQF
7	Q^{P}QSQQCR___SQGRS	Q^{P}QSEQSL___QQCQF
8 var1	Q^{P}QSQQCR___SQGRS	Q^{P}SEQSL___QQCQF
8 var2	Q^{P}SQQCR___SQGRS	Q^{P}QSEQSL___QQCQF

Small: ^{39}QQSQQCRQQLQGRQFRSCQRYLSQGRSPYGGEED^{72}
Large: ^{82}QQSEQSLRDCCQQLRNVDERCRCEAIRQAVRQQQQEGGYQEGQSQQVYQRARDLPRRCNMRPQQCQFRVIF^{150}

Fig. 4 ESI-HRMS of intact Ses i 1 proteoforms. Clipping variants of Ses i 1 were identified by combination of results from middle–down and top–down analysis by LC–HRMS. Upper lane: ESI–HRMS spectrum of abundant Ses i 1 proteoforms. Reference sequences are given below the clipping combinations. Unclipped "core region" of subunits are given in gray. Different clipping variants resulting from identical masses are not distinguishable, and therefore, designated var1/var2

```
                    1          11         21         31         41         51
Uniprot Q9AUD1      MAKKLALAAV LLVAMVALAS ATTYTTTVTT TAIDDEANQQ SQQCRQQLQG RQFRSCQRYL
Moreno et al.       ---------- ---------- ---------- --------QQ SQQCRQQLQG RQFRSCQRYL
Hummel et al.       ---------- ---------- ---------- --------QQ SQQCRQQLQG RQFRSCQRYL

                    61         71         81         91         101        111
Uniprot Q9AUD1      SQGRSPYGGE EDEVLEMSTG NQQSEQSLRD CCQQLRNVDE RCRCEAIRQA VRQQQQEGGY
Moreno et al.       SQGRSPYGGE ED-------- -----QSLRD CCQQLRNVDE RCRCEAIRQA VRQQQQEGGY
Hummel et al.       SQGRSPYGGE ED-------- -QQSEQSLRD CCQQLRNVDE RCRCEAIRQA VRQQQQEGGY

                    121        131        141        151
Uniprot Q9AUD1      QEGQSQQVYQ RARDLPRRCN MRPQQCQFRV IFV
Moreno et al.       QEGQSQQVYQ RARDLPRRCN MRPQQCAFRV IF-
Hummel et al.       QEGQSQQVYQ RARDLPRRCN MRPQQCQFR- ---
```

Fig. 5 Published sequences of Ses i 1. Tai et al. [19] have first described the sequence using cDNA cloning and sequencing techniques (Uniprot KB entry AC Q9AUD1). Moreno et al. [15] extracted Ses i 1 and analyzed the protein using MALDI-TOF–MS. Signal peptide and linker peptide not belonging to the mature Ses i 1 are displayed in gray. The sequence shift of the large subunit as identified in this study compared to [15] is highlighted in red. The sequence conflict at position 147 is displayed in gray and green. Notably, Pastorello et al. [1] published a partial 17 aa Ses i 1 sequence also harboring A^{147} indicating a breed-specific polymorphisms at this position

Table 2 Clipping variants of intact proteins identified by top–down approach

	Monoisotopic obsM	Monoisotopic theoM	Error [ppm]	Small subunit		Large subunit	
				N-term. clipping	C-term. clipping	N-term. clipping	C-term. clipping
1	12066.60	12066.58	1.62	No clipping[a]	No clipping	No clipping[a]	−RVIF
2	11951.57	11951.55	1.76	No clipping[a]	−D	No clipping[a]	−RVIF
3	11822.53	11822.51	1.65	No clipping[a]	−ED	No clipping[a]	−RVIF
4	11694.47[b]	11694.45	1.88	No clipping[a]	−ED	−Q[a]	−RVIF
4	11694.47[b]	11694.45	1.88	−Q[a]	−ED	No clipping[a]	−RVIF
5	11579.44	11579.42	1.74	No clipping[a]	−GGEED	No clipping[a]	−RVIF
6	11451.38[b]	11451.36	1.20	No clipping[a]	−GGEED	−Q[a]	−RVIF
6	11451.38[b]	11451.36	1.20	−Q[a]	−GGEED	No clipping[a]	−RVIF
7	11319.33	11319.30	1.87	No clipping[a]	−PYGEEED	No clipping[a]	−RVIF
8	11191.27[b]	11191.25	1.84	No clipping[a]	−PYGEEED	−Q[a]	−RVIF
8	11191.27[b]	11191.25	1.84	−Q[a]	−PYGGEED	No clipping[a]	−RVIF

[a]Conversion from glutamate into pyroglutamate

[b]Different clipping variants resulting from identical masses, and therefore, not distinguishable

epitopes are not located at the termini, a direct effect of clipping on the integrity of epitopes is rather unlikely. Nevertheless, clipping of Cys^{146} at the C-terminus of the large subunit results in break of a disulfide bridge, leading to considerable change of the protein conformation and stability.

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

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing financial interest.

References

1. Pastorello EA, Varin E, Farioli L, Pravettoni V, Ortolani C, Trambaioli C, et al. The major allergen of sesame seeds (Sesamum indicum) is a 2S albumin. J Chromatogr B Biomed Sci Appl. 2001;756:85–93.
2. Beyer K, Bardina L, Grishina G, Sampson HA. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. J Allergy Clin Immunol. 2002;110:154–9.
3. Navuluri L, Parvataneni S, Hassan H, Birmingham NP, Kelly C, Gangur V. Allergic and anaphylactic response to sesame seeds in mice: identification of Ses i 3 and basic subunit of 11S globulins as allergens. Int Arch Allergy Immunol. 2006;140:270–6. https://doi.org/10.1159/000093284.

4. Hara-Hishimura I, Takeuchi Y, Inoue K, Nishimura M. Vesicle transport and processing of the precursor to 2S albumin in pumpkin. Plant J. 1993;4:793–800.
5. Moreno FJ, Clemente A. 2S albumin storage proteins: what makes them food allergens? Open Biochem J. 2008;2:16–28. https://doi.org/10.2174/1874091X00802010016.
6. Delapena MAG, Villalba M, Garcialopez JL, Rodriguez R. Cloning and expression of the major allergen from yellow mustard seeds, Sin a I. Biochem Biophys Res Commun. 1993;190:648–53. https://doi.org/10.1006/bbrc.1993.1097.
7. Pena MAG, Monsalve RI, Batanero E, Villalba M, Rodriguez R. Expression in *Escherichia coli* of Sin a 1, the major allergen from mustard. Eur J Biochem. 1996;237:827–32. https://doi.org/10.1111/j.1432-1033.1996.0827p.x.
8. Shewry PR, Napier JA, Tatham AS. Seed storage proteins: structures and biosynthesis. Plant Cell. 1995;7:945–56. https://doi.org/10.1105/tpc.7.7.945.
9. Monsalve RI, Lopez-Otin C, Villalba M, Rodríguez R. A new distinct group of 2S albumins from rapeseed. Amino acid sequence of two low molecular weight napins. FEBS Lett. 1991;295:207–10.
10. Murén E, Ek B, Rask L. Processing of the 2S storage protein pronapin in Brassica napus and in transformed tobacco. Eur J Biochem. 1995;227:316–21.
11. Gehrig PM, 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 USA. 1996;93:3647–52.
12. Moreno FJ, Jenkins JA, Mellon FA, Rigby NM, Robertson JA, Wellner N, Clare Mills EN. Mass spectrometry and structural characterization of 2S albumin isoforms from Brazil nuts (*Bertholletia excelsa*). Biochim Biophys Acta. 2004;1698:175–86. https://doi.org/10.1016/j.bbapap.2003.11.007.
13. da Silva JG, Machado OL, Izumi C, Padovan JC, Chait BT, Mirza UA, Greene LJ. Amino acid sequence of a new 2S albumin from *Ricinus communis* which is part of a 29-kDa precursor protein. Arch Biochem Biophys. 1996;336:10–8.
14. Korte R, Happe J, Brümmer I, Brockmeyer J. Structural characterization of the allergenic 2S albumin Cor a 14: comparing proteoform patterns across hazelnut cultivars. J Proteome Res. 2017;16:988–98. https://doi.org/10.1021/acs.jproteome.6b00924.
15. Moreno FJ, Maldonado BM, Wellner N, Mills ENC. 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. 2005;1752:142–53. https://doi.org/10.1016/j.bbapap.2005.07.022.
16. 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;14:1547–56. https://doi.org/10.1021/pr5012262.
17. Monsalve RI, Menéndez-Arias L, López-Otin C, Rodríguez R. β-Turns as structural motifs for the proteolytic processing of seed proteins. FEBS Lett. 1990;263:209–12. https://doi.org/10.1016/0014-5793(90)81375-X.
18. 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;177:159–66.
19. Tai SSK, Lee TTT, Tsai CCY, Yiu T-J, Tzen JTC. Expression pattern and deposition of three storage proteins, 11S globulin, 2S albumin and 7S globulin in maturing sesame seeds. Plant Physiol Biochem. 2001;39:981–92. https://doi.org/10.1016/S0981-9428(01)01314-6.
20. Hara-Nishimura I, Inoue K, Nishimura M. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. FEBS Lett. 1991;294:89–93.
21. Scott MP, Jung R, Muntz K, Nielsen NC. A protease responsible for post-translational cleavage of a conserved Asn–Gly linkage in glycinin, the major seed storage protein of soybean. Proc Natl Acad Sci USA. 1992;89:658–62.
22. Shimada T, Hiraiwa N, Nishimura M, Hara-Nishimura I. Vacuolar processing enzyme of soybean that converts proproteins to the corresponding mature forms. Plant Cell Physiol. 1994;35:713–8.
23. Rini JM, Hofmann T, Carver JP. Amino acid sequence differences in the alpha chains of pea seed isolectins: C-terminal processing. Biochem Cell Biol. 1987;65:338–44.
24. Pantoja-Uceda D, Bruix M, Santoro J, Rico M, Monsalve R, Villalba M. Solution structure of allergenic 2S albumins. Biochem Soc Trans. 2002;30:919–24.
25. Kortt AA, Caldwell JB, Lilley GG, Higgins TJ. Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). Eur J Biochem. 1991;195:329–34.
26. Ampe C, van Damme J, de Castro LA, Sampaio MJ, 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;159:597–604.