DOI 10.1002/mnfr.200500020 Mol. Nutr. Food Res. 2005, *49*, 744–755

Proteolytic processing of the peanut allergen Ara h 3

Sander R. Piersma[1], Marco Gaspari[2], Susan L. Hefle[3] and Stef J. Koppelman[2,4]*

[1]FOM Institute for Atomic and Molecular Physics, Amsterdam, The Netherlands
[2]TNO Nutrition and Food Research, Zeist, The Netherlands
[3]Food Allergy Research and Resource Program, University of Nebraska, Lincoln, NE, USA
[4]University Medical Center Utrecht, Dept. of Dermatology/Allergology, Utrecht, The Netherlands

The allergen Ara h 3 has been purified recently from peanuts. In contrast to recombinant Ara h 3, a 60 kDa single-chain polypeptide, the allergen isolated from its native source is extensively proteolytically processed. The characteristic proteolytic processing for 11S plant storage proteins of the glycinin family is observed for Ara h 3 yielding an acidic and a basic subunit, bound by a disulfide bridge. In addition to this, proteolytic truncation is observed for the acidic subunit but not for the basic subunit of Ara h 3. A series of Ara h 3 polypeptides ranging from 13–45 kDa was separated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and each band was digested by trypsin. Peptides related to the bands were identified and a scheme positioning the different polypeptides in the Ara h 3 sequence has been constructed. Peptide analysis showed sequence heterogeneity at two positions indicating the presence of multiple genes encoding variant, but highly homologous Ara h 3 proteins. The pool of Ara h 3 polypeptides from its native source illustrated that allergen from the peanut is much more complex than the recombinant protein used for epitope mapping experiments. From several Ara h 3 truncation products one or more immunoglobulin E (IgE) binding sites had been removed. Characterization of the allergenicity of Ara h 3 should therefore also include IgE-binding studies with peanut-derived Ara h 3, providing the high degree of variation in the Ara h 3 protein structure, as this is what peanut-allergic individuals are confronted with.

Keywords: Allergen / Ara h 3 / Mass spectrometry / Peanut / Proteolytic processing

Received: February 14, 2005; revised: April 30, 2005; accepted: May 27, 2005

1 Introduction

Peanut allergy is one of the most severe food allergies due to its persistency and life-threatening character [1]. The nature of the allergenic proteins in peanut has been studied extensively [2–4] and several allergens have been identified. Ara h 1, a 63 kDa glycoprotein [5], was identified as the major allergen in peanut [6] with distinct heat-stable immunoglobulin E (IgE) binding sites [7] both on the protein part [8, 9] as well as on the carbohydrate moiety [10]. Ara h 2 was identified as a second important allergen from peanuts and is characterized as a doublet of 17–20 kDa on SDS-PAGE [11–13]. Ara h 3, the third allergen described from peanuts, was first identified as a 14 kDa protein [14], but cloning its gene revealed a 60 kDa protein with IgE-binding sites throughout the recombinant gene product [13, 15]. Recently, the purification and biochemical characterization of Ara h 3 from peanuts, its native source, has been reported [16].

Ara h 3 belongs to the glycinin family of legume storage proteins with soy glycinin being the best-characterized representative of this protein family. After synthesis and removal of the signal peptide, the soy glycinin proprotein is stored as trimers in protein storage vacuoles (PSVs). In the PSVs the glycinin proprotein is proteolytically cleaved at the conserved N-G bond by an aspartic acid endoprotease [17] yielding an acidic and a basic subunit [18, 19]. The two subunits remain covalently linked by an intermolecular disulfide bridge [18–20] and associate into a very heat-stable hexameric structure [21]. From the analysis of the Ara h 3 gene and its recombinant gene product [15], and of peanut-derived Ara h 3 [16], it was concluded that Ara h 3 is a member of the glycinin family of the Leguminosae. Following the cloning of Ara h 3, epitope-scanning experiments

Correspondence: Dr. Sander R. Piersma, FOM Institute for Atomic and Molecular Physics (AMOLF), Kruislaan 407, NL-1098 SJ, Amsterdam, The Netherlands
E-mail: Piersma@amolf.nl
Fax: +31-20-6684106

Abbreviations: Ara h 3, arachis hypogaea allergen 3; **IgE**, immunoglobulin E

* Current addresses: HAL Allergy, Haarlem, The Netherlands and AllerTeQ, De Bilt, The Netherlands

identified four IgE-binding epitopes on the acidic subunit of Ara h 3 [15], but none on the basic subunit. In contrast to recombinant Ara h 3, SDS-PAGE analysis of purified peanut-derived Ara h 3 showed multiple polypeptide bands, including the expected bands of the acidic subunit and the basic subunit. The multiple band pattern was not related to proteolysis during the Ara h 3 extraction and isolation procedure and similar band patterns are found for Ara h 3 in different varieties of peanut [22]. Recently, it was shown that Ara h 3, Ara h 4, Ara h 3/4, Gly1, and iso-Ara h 3 are isoallergens [23]. The appearance of multiple Ara h 3-related polypeptides in peanuts indicates extensive processing of Ara h 3 in its native source. Indeed it has been speculated that the 14 kDa protein first identified as Ara h 3 is a *C*-terminally truncated acidic subunit of the mature Ara h 3 protein [14].

The availability of a method to purify Ara h 3 from peanuts opens a new route to a protein level characterization of this allergen and its post-translational processing products. Here, we have applied mass spectrometry to analyze the subunit size and composition of Ara h 3, and its proteolytic truncation products.

2 Materials and methods

2.1 Ara h 3 purification

The allergen Ara h 3 was isolated from peanuts (*Arachis hypochaea*; Runner cultivar) as described previously [16]. Briefly, 25 g ground peanuts were extracted with 250 mL 20 mM Tris-HCl buffer (pH 7.3), cleared by centrifugation at $3000 \times g$ (30 min) and subsequently at $10\,000 \times g$ (30 min). Next, 60 mL of this crude peanut extract was applied to a 300 mL Source Q anion exchange column (Pharmacia, Uppsala, Sweden) connected to an FPLC/AKTA protein purification system (Pharmacia). Peanut proteins were eluted with an NaCl gradient (0 – 1 M; 4000 mL, 100 mL/min) in 20 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS; pH 7.2). Eluting fractions were analyzed by SDS-PAGE; Ara h 3 eluted at approximately 400 – 500 mM NaCl.

2.2 Ara h 3 molecular mass

For determination of the intact molecular mass of Ara h 3, 1 μg/μL of the protein dissolved in 10 mM NH_4HCO_3 (pH 7.9) was mixed in a 1 : 10 v/v ratio with a matrix solution consisting of saturated sinapinic acid in 0.1% aqueous trifluoroacetic acid/acetonitrile 2 : 1 v/v. 1 μL of this mixture was pipetted on a stainless-steel Scout384 MALDI sample plate. Subunit molecular masses were determined by adding chaotrope and reducing agent to the protein. Ara h 3 dissolved in 8 M urea containing 10 mM dithiothreitol (DTT) was mixed in a 1:10 v/v ratio with matrix solution and also 1 μL of the mixture was pipetted on the sample plate. MALDI-TOF mass spectra were recorded on a Bruker Biflex III mass spectrometer (Bruker Daltonics, Bremen, Germany) in delayed extraction linear mode. The data was sampled with a 4 GHz digitizer (0.5 ns channel width). External calibration was performed using the software package XMASS 5.0 supplied with the instrument. Mass accuracy for the complex Ara h 3 sample was better than 0.2% for masses below 20 kDa and better than 0.5% for masses above 20 kDa.

2.3 Electrophoresis

SDS-PAGE was performed essentially according to Laemmli [24], using a BioRad Mini Protean II system (BioRad, Hercules, CA, USA) with 12% acrylamide gels (15×10 cm). Prestained molecular mass markers of 14.3, 21.5, 30, 46, 66, 97, and 220 kDa were used as reference. Ara h 3 samples (12.7 μg) were mixed in a 1 : 1 ratio with a sample buffer containing 63 mM Tris-HCl buffer (pH 6.8), 1% w/v DTT, 2% v/v SDS, 0.01% w/v bromophenol blue, and 20% v/v glycerol, and were boiled for 5 min. Gels were stained with Coomassie Brilliant Blue R-250 in 50% aqueous methanol containing 2% acetic acid. Gels were destained in the same solvent system and subsequently in 25% aqueous ethanol containing 8% acetic acid. Gels were scanned with an ImageMaster DTS (Pharmacia).

2.4 In-gel digestion

Protein bands were manually cut from the SDS-PAGE gel and transferred to Eppendorf tubes. Gel pieces were processed essentially according to Shevchenko [25]. Gel pieces (~1 mm^3) were dehydrated in 100 mM NH_4HCO_3/50% acetonitrile, liquid was removed and substituted with 100 mM NH_4HCO_3, and this procedure was repeated twice. Washed gel pieces were reduced in 10 mM DTT in 100 mM NH_4HCO_3 for 1 h at 56°C. Liquid was removed and gel pieces were cooled to room temperature, subsequently 55 mM iodoacetamide in 100 mM NH_4HCO_3 was added to alkylate the free cysteines and gel pieces were incubated in the dark at room temperature for 1 h. After removal of the liquid the gel pieces were washed with 100 mM NH_4HCO_3 and dehydrated with 100 mM NH_4HCO_3/50% acetonitrile. After removal of the liquid the gel pieces were incubated on ice with the digestion buffer consisting of 50 mM NH_4HCO_3 containing 0.1 mg/mL sequence-grade porcine trypsin (Promega, Madison, WI, USA). After removal of excess digestion buffer the gel pieces were incubated at 37°C for 18 h. Peptides were extracted by sonication (15 min) using subsequent extraction solvents consisting of Milli-Q water, 5% formic acid and 5% formic acid/50% acetonitrile. Extracts were pooled and lyophilized in a vacuum centri-

fuge. For MALDI and ESI-MS analyses the lyophilized peptides were reconstituted in 10 µL 0.1% formic acid.

2.5 Mass spectrometry

MALDI-TOF-MS peptide mass fingerprinting was performed on the same Biflex III instrument described above, and was performed in reflectron mode. 1 µL digest was mixed with 5 µL matrix solution (20 mg/mL dihydroxybenzoic acid in 0.1% aqueous trifluoroacetic acid/acetonitrile 2:1 v/v). 0.5 µL of this solution was transferred to the Scout 384 MALDI sample plate, and was allowed to crystallize. Mass spectra were calibrated using the trypsin autodigestion masses observed in all extracts. Electrospray-MS was performed on a quadrupole TOF Q-TOF 2 instrument (Micromass, Manchester, UK) coupled on-line with a capillary HPLC (CapLC) (Waters, Milford, MA, USA). Peptides were separated on a capillary C18 reversed-phase column (75 µm ID, 15 cm *L*) (LC Packings, Amsterdam, The Netherlands) using a 5–60% acetonitrile gradient in 0.1% formic acid (45 min; 200 nL/min). The column was connected to a metal-coated fused-silica electrospray needle (tip ID, 10 µm) (NewObjective, Woburn, MA, USA). Peptide mass fingerprinting and MS/MS ion searching was performed using Mascot (www.matrixscience.com). Ara h 3 was digested *in silico* using ProteinProspector (prospector.ucsf.edu) and the resulting peptide list was compared with measured peptide masses.

2.6 Sequence analysis

From the NCBI protein database the reported sequences of Ara h 3 were retrieved: AAC63045 Ara h 3 [*Arachis hypogaea*] [15], AAD47382 Ara h 4 [*Arachis hypogaea*] [13], AAM46958 allergen Arah 3/4 [*Arachis hypogaea*], AAG01363 Gly1 [*Arachis hypogaea*], and AAT39430 iso-Ara h 3 [*Arachis hypogaea*]. Multiple sequence alignment was performed using Clustal W (1.82) using default parameters (www.ebi.ac.uk/clustalw). Potential posttranslational modifications were analyzed using Findmod (www.expasy.org/tools/findmod) and NetPhos (www.cbs.dtu.dk/services/NetPhos).

3 Results

3.1 Ara h 3 subunit composition

In Fig. 1A the typical SDS-PAGE pattern for Ara h 3 purified from peanuts is shown. In addition to the major bands at 45, 42, and 23 kDa minor bands ranging from 12–35 kDa can be seen. Previous work (*N*-terminal sequencing) showed that the 45–42 bands are related to the acidic subunit and the band at 23 kDa is related to the basic subunit [16]. For mass spectrometric analysis the bands marked A–L were cut out of the gel and were reduced, alkylated, and digested by trypsin.

In Figs. 1B and C the MALDI mass spectrum of Ara h 3 is shown. Spectra were acquired in two mass windows for optimal sensitivity: 9–27 kDa and 27–80 kDa. In the high mass window the dominant peaks at 58 860, 57 926, 55 128, and 32 649 Da are found in addition to minor peaks at 45 225, 28 977, and 27 565 Da (the latter two are doubly charged Ara h 3 peaks). In the low mass window major peaks at 12 564 and 11 955 Da are detected in addition to minor peaks at 20 497, 13 889, and 10 725 Da. Upon reduction and addition of urea significant mass shifts occur. In Figs. 1D and E the MALDI mass spectrum of Ara h 3 after reduction is shown; spectra were acquired in the same mass windows as in panels B/C. In the high mass window the three peaks between 55 and 59 kDa are still visible as minor peaks, indicating incomplete reduction. The major peaks are now found at 40 659, 38 556, 37 721, and 34 671 Da. In the low-mass window the major peak is found at 20 505 Da in addition to minor peaks at 12 487, 11 956, and 10 250 Da. By comparing the SDS-PAGE band pattern with the reduced Ara h 3 masses, the major bands/peaks can be correlated. The intensely stained bands of the acidic subunit at 42–45 kDa (Fig. 1A) correspond to the MALDI peaks at 38 556, 37 721, and 34 671 Da. The intensely stained band at 23 kDa (Fig. 1A) corresponds to the dominant 20 505 Da peak in Fig. 1D. Considering that the acidic and basic subunit of Ara h 3 are covalently linked by a disulfide bridge, it can be calculated that the peaks at 58 860, 57 926, 55 128 Da in Fig. 1B correspond to the acidic-basic subunit pairs comprised of the peaks at 20 505 + 38 556, 37 721 and 34 671 Da, respectively, taking into account a mass accuracy better than 0.5% for the high mass region. The 32 649 Da peak in Fig. 1B can be correlated with the peaks at 12 487 and 20 505 Da giving rise to a small acidic-basic subunit pair. The 40 659 Da peak in Fig. 1C could not be correlated to any of the peaks shown in Fig 1B; association with the 20 505 Da basic subunit would yield an intact molecular mass exceeding 60 kDa. At the resolution obtained in Fig. 1B, a 60 kDa shoulder should be visible if present (not observed). Major peak molecular masses of basic (20.5–20.9 kDa) and acidic (37.8–38.8 kDa) polypeptides can be calculated from published sequences (Fig. 2) and are consistent with the MALDI signals observed.

3.2 Ara h 3 sequence analysis

In Fig. 2 the Clustal W multiple sequence alignment of the amino acid sequences of the four Ara h 3 entries in the NCBI protein database, translated from their respective

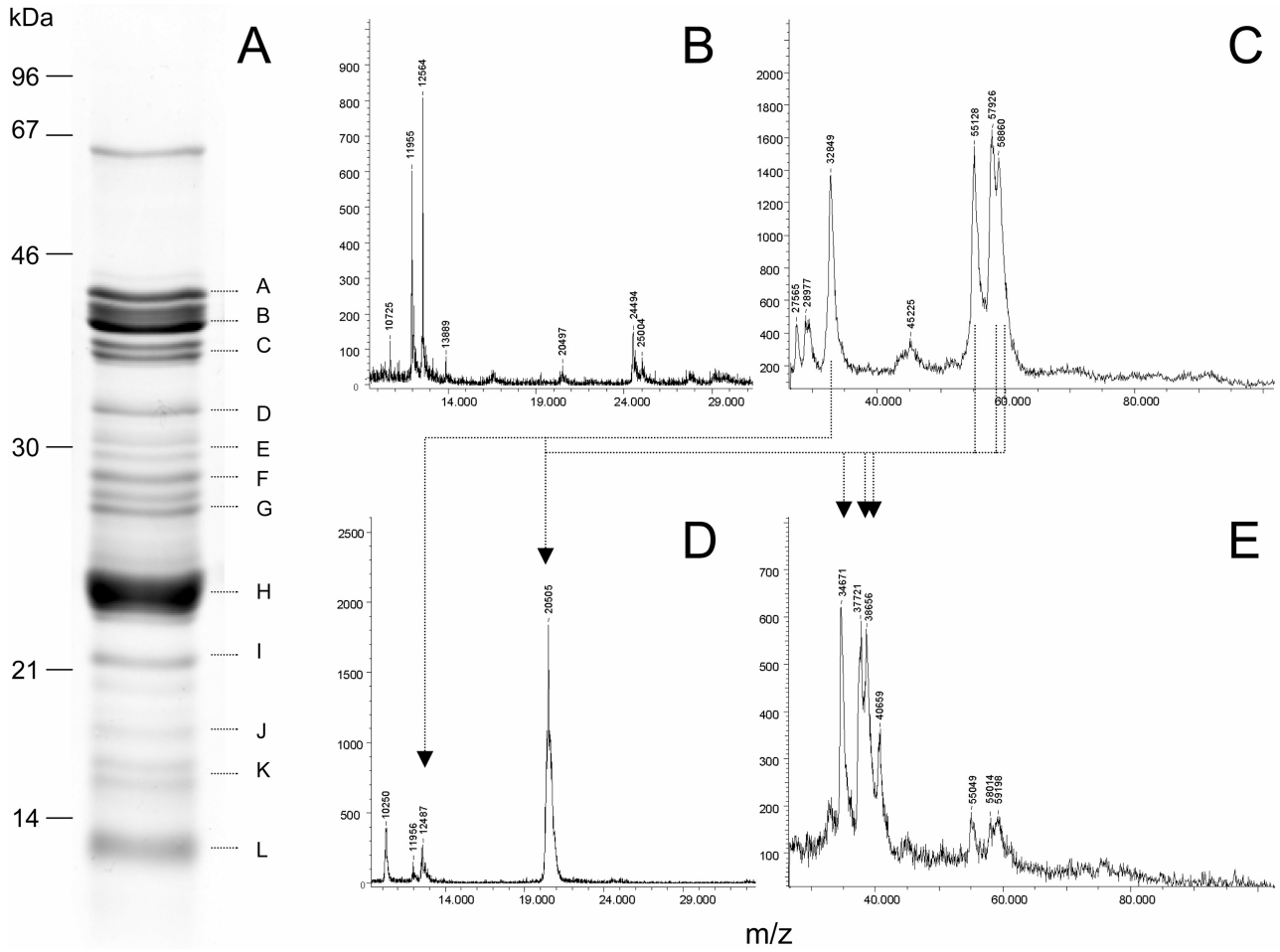

Figure 1. Subunit structure of Ara h 3. (A) SDS-PAGE gel showing the polypeptides of purified Ara h 3. Bands marked A – L were cut from the gel and processed further for peptide analysis. The band at ~65 kDa is a minor contamination by Ara h 1. (B), (C) MALDI mass spectrum of Ara h 3 dissolved in 10 mM ammonium bicarbonate, pH 7.9 (intact Ara h 3). (D), (E) MALDI mass spectrum of Ara h 3 dissolved in 8 M urea containing 10 mM DTT (reduced and dissociated Ara h 3).

nucleotide sequences, are shown. Accession numbers of the sequences are: AAC63045, AAM46958, AAD47382, and AAG01363. Before multiple sequence alignment signal peptides were removed, the start of the amino acid sequences shown is at the *N*-terminus of the mature protein (ISFRQQ...). The row marked MS shows the parts of the Ara h 3 sequence recovered by mass spectrometry and identified peptides are listed in Table 1. In the case of sequence heterogeneity, alternative peptides identified by MS are also shown at the appropriate positions. Overall sequence identity is high (83%), and most substitutions are homologous substitutions (63%). Four gaps are found ranging from 2–5 amino acids (mainly due to sequence AAM46958). In bold the two *N*-terminal peptides identified by Edman degradation are shown. The *N*-terminus ISFRQQ corresponds to the acidic subunit and the *N*-terminus GIEETI corresponds to the basic subunit.

3.3 Peptide identification

The SDS-PAGE bands marked A – L (Fig. 1A) were subjected to in-gel digestion by trypsin. The resulting peptides were eluted, lyophilized, redissolved, and analyzed by MALDI-MS and LC-MS. In Fig. 3 the MALDI-MS spectra of the tryptic digests of bands A and H, corresponding to the acidic and the basic subunit of Ara h 3 are shown. Identified peptides are labeled with their respective residue numbers (following the alignment shown in Fig. 2). In addition to 13 Ara h 3 peptides observed for band A, 8 autodigestion peptides of porcine trypsin are observed in the digestions, which served as internal calibrants in the MALDI experiments. The rich MALDI spectra obtained for gel bands A – L served as first pass for peptide identification. Following MALDI-MS analysis LC-MS/MS was performed for additional sequence coverage and peptide

Figure 2. Clustal W multiple sequence alignment of the four Ara h 3 sequences. Symbols used: * corresponds to amino acid identity, : is a conserved substitution (AVFPMILW, DE, RHK and STYHCNGQ), . is a homologous substitution. The rows marked MS show the Ara h 3 sequence recovered by MS. The *N*-termini of the acidic and the basic subunit determined by Edman degradation are shown in bold.

Table 1. Trypic peptides identified from Ara h 3 subunits

Residues	*m/z* (obs)	*m/z* (calc)	+++	Sequence	MS	MS/MS	Gel bands
17–25	1083.57	1083.57	1+	LNAQRPDNR	M		ABC G L
62–73	1468.70	1468.72	1+	RPFYSNAPQEIF	M		AB G L
62–78	2051.03	2051.05	1+	RPFYSNAPQEIFIQQGR	M		ABC G L
62–78	2051.06	2051.05	3+	RPFYSNAPQEIFIQQGR	E	37	AB D I L
65–78	1650.81	1650.82	1+	YSNAPQEIFIQQGR	M		ABC FG L
65–78	1650.84	1650.82	2+	YSNAPQEIFIQQGR	E	50	ABCD L
66–78	1487.73	1487.76	1+	SNAPQEIFIQQGR	M		AB L
79–100	2488.19	2488.16	1+	GYFGLIFPGCPSTYEEPAQQGR	M		ABC
102–109	1005.53	1005.53	1+	HQSQRPPR	M		ABC
110–126	2073.00	2072.95	1+	RFQGQDQSQQQQDSHQK	M		B
111–126	1916.82	1916.85	1+	FQGQDQSQQQQDSHQK	M		ABC
176–188	1574.75	1574.77	1+	FNLAGNHEQEFLR	M		ABC EFG K
176–188	1574.76	1574.77	3+	FNLAGNHEQEFLR	E	30	ABC EFG
189–194	809.40	809.39	1+	YQQQSR	M		ABC FG
258–264	884.53	884.53	1+	QILQNLR	M		ABC JK
265–279	1603.73	1603.76	1+	GENESDEQGAIVTVR	M		AB JK
265–279	1603.76	1603.76	2+	GENESDEQGAIVTVR	E		ABC J
265–279	1590.78	1590.76	2+	GENESEEEGAIVTVK	E	87	BC JK
265–279	1618.77	1618.75	2+	GENESEEEGAIVTVR	E	58	ABC
265–283	1987.03	1986.98	1+	GENESDEQGAIVTVRGGLR	M		A
332–343	1356.62	1356.65	2+	GIEETICTASFK	E		H
351–363	1389.69	1389.70	1+	SPDIYNPQAGSLK	M		H
351–363	1389.64	1389.70	2+	SPDIYNPQAGSLK	E	54	H
364–376	1255.72	1255.73	2+	TANDLNLLILR	E	44	H
377–389	1541.71	1541.78	2+	WLGLSAEYGNLYR	E	66	H
377–389	1541.79	1541.78	1+	WLGLSAEYGNLYR	M		H
390–402	1497.68	1497.72	3+	NALFVPHYNTNAH	E	30	H
390–402	1497.71	1497.72	1+	NALFVPHYNTNAH	M		H
390–409	2314.19	2314.21	1+	NALFVPHYNTNAHSIIYALR	M		H
412–423	1295.63	1295.64	2+	AHVQVVDSNGNR	E	25	H
424–441	2099.00	2099.04	1+	VFDEELQEGHVLVVPQNF	M		H
424–446	2525.28	2525.30	1+	VFDEELQEGHVLVVPQNFAVAGK	M		H
424–446	2541.30	2541.30	1+	VYDEELQEGHVLVVPQNFAVAGK	M		H
424–446	2541.20	2541.30	3+	VYDEELQEGHVLVVPQNFAVAGK	E	41	H
447–458	1448.66	1448.67	1+	SQSENFEYVAFK	M		H
447–458	1448.63	1448.67	2+	SQSENFEYVAFK	E	77	H
459–491	3545.66	3545.74	3+	TDSRPSIANLAGENSFIDNLPEEVVANSYGLPR	E	50	H
464–491	3086.47	3086.54	3+	PSIANLAGENSFIDNLPEEVVANSYGLPR	E	64	H
499–504	733.38	733.36	1+	NNNPFK	M		H
505–515	1306.69	1306.68	1+	FFVPPSEQSLR	M		H
505–515	1306.67	1306.68	2+	FFVPPSEQSLR	E	22	H
508–515	912.46	912.47	2+	PPSEQSLR	E	32	H

Residue numbering follows that of NCBI entry AAM46958 starting from the protein *N*-terminus (ISFRQQ...). Peptides reported are tryptic and semi-tryptic peptides. Observed *m/z* values correspond to the mono-isotopic singly protonated peptides, *m/z* values are calculated using ProteinProspector. +++ : charge state of the peptide. MS techniques used: MALDI-MS (M) and LC-MS(/MS) (E). The number in the column MS/MS corresponds to the MASCOT ion-score for peptide MS/MS spectra. Gel bands correspond to the bands shown in Fig. 1A.

identification. In Table 1 all peptide identifications using MALDI-MS, LC-MS, and LC-MS/MS data are summarized. For each gel band analyzed the identified peptides are given. The residue-numbering scheme used corresponds to Ara h 3 sequence AAM46958. The total sequence coverage by MS is 54%, and is asymmetrically divided over the acidic subunit (335 amino acids; 35% coverage) and basic subunit (187 amino acids; 89% coverage). MALDI and LC-MS(/MS) data identify similar peptides, *e.g.*, res. 62–78, 265–279, 351–363 or 505–515 and complementary peptides, *e.g.*, res. 102–109, 189–194 or 499–504 (MALDI) and 332–343, 364–376 or 412–423 (LC-MS(/MS).) It can be observed that in addition to pure tryptic peptides (tryptic recognition at K/R-X, with X is not P), several half tryptic peptides are found due to aspecific cleavage at either the *N*- or *C*-terminal side of the peptide.

3.4 Sequence heterogeneities

At two sites in the sequence microheterogeneities are found: in the acidic subunit peptide encompassing res. 265–279 and in the basic subunit peptide encompassing res. 424–446. For the first peptide 3 variant peptides were identified at 1603.76, 1590.78, and 1618.75 Da corresponding to the sequences GENESDEQGAIVTVR, GENESEEEGAIVTVK and GENESEEEGAIVTVR, respectively (variants underlined). The observed variants, D → E, Q → E

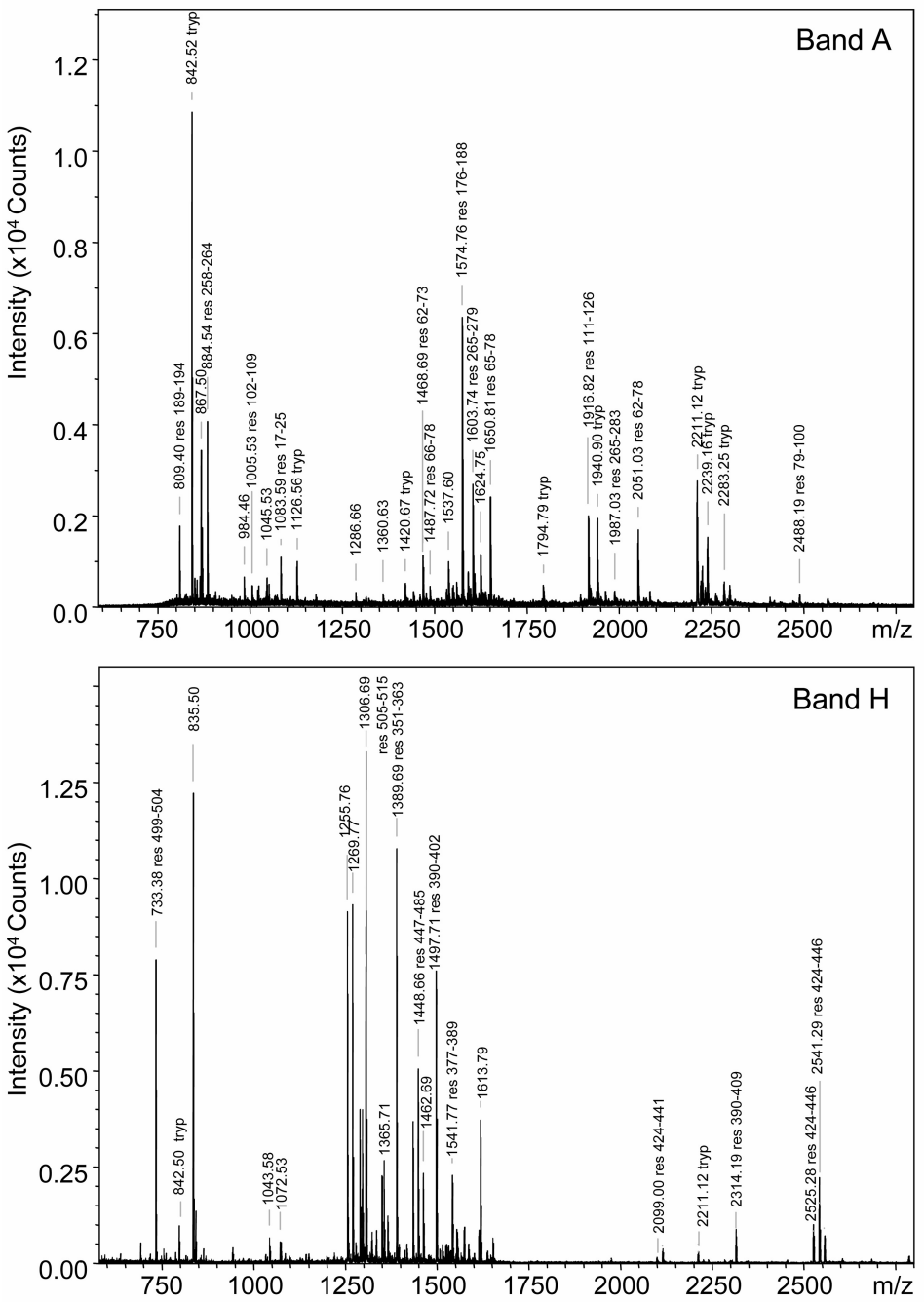

Figure 3. MALDI mass spectra of the tryptic digests of SDS-PAGE bands A and H. Identified peptides are labeled with residue numbers.

and R → K, are homologous substitutions. In Fig. 4 the MS/MS spectra of the 1590.78 and 1618.75 Da peptides are shown (795.90 Da and 809.89 Da in their doubly protonated forms, respectively). Both peptides display extensive y-ion fragmentation series ranging from y(3)–y(12/13). Fragment masses in both series are consistent with the proposed sequences and are offset by 28.01 Da, in line with the *C*-terminal residue being either R (156.10 Da) or K (128.09 Da). Localization of the other two microheterogeneities in this peptide, D → E and Q → E, are supported by multiple intact peptide data from both MALDI and LC-MS(/MS) experiments (Table 1). The other peptide displaying heterogeneity is found in the acidic subunit: V<u>F</u>DEEL-QEGHVLVVPQNFAVAGK (2525.30 Da) and V<u>Y</u>DEEL-

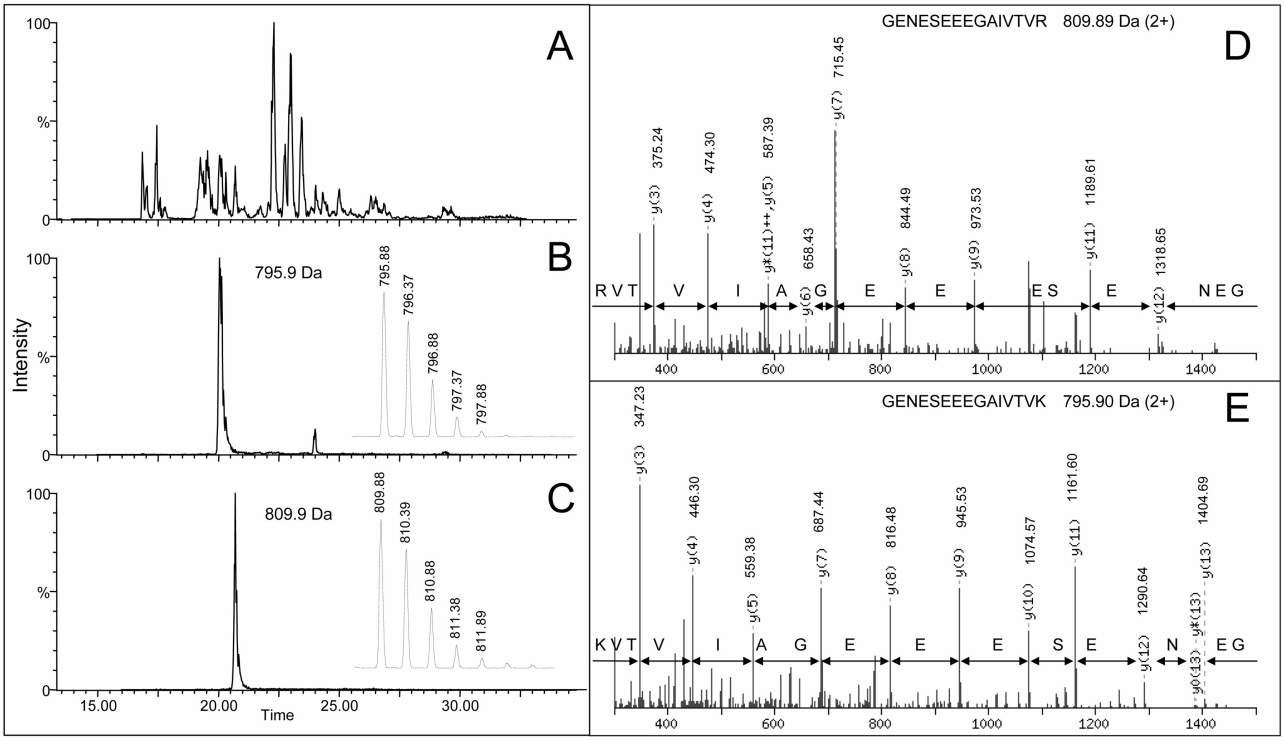

Figure 4. LC-MS/MS spectra of peptide (res. 265–279) showing sequence heterogeneity. (A) LC-MS chromatogram (basepeak intensity) of the digest of band B. (B) Selected ion chromatogram (SIC) of the mass window 795.82–795.94 Da, the insert shows the mass spectrum for this peak. (C) SIC of the mass window 809.82–809.95 Da, the insert shows the mass spectrum for this peak. (D) MS/MS spectrum of the peptide GENESEEEGAIVTVR. (E) MS/MS spectrum of the peptide GENESEEEGAIVTVK. For both peptides the amino acid sequence and the precursor ion mass and charge state are shown.

QEGHVLVVPQNFAVAGK (2541.30 Da). Also here the heterogeneity is due to a homologous substitution: F → Y, and is supported by multiple peptides detected by MALDI and LC-MS/MS (Table 1). The sequence microheterogeneities are found in multiple bands (A, B, C, J and K for res. 265–279 and band H for res. 424–446) indicating that the different bands on SDS-PAGE do not correspond to distinct genetic variant forms of Ara h 3 (like observed for soy glycinin).

3.5 Proteolytic truncation

For each identified peptide in Table 1 the corresponding band cut from the SDS-PAGE gel (Fig. 1A) can be found in the rightmost column of the table. It can be observed that the *C*-terminal 187 amino acids are only identified in band H (the basic subunit), whereas the *N*-terminal 331 amino acids are localized in multiple bands varying from A–L (but not in H). Even though sequence coverage is around 35% for the acidic subunit, a pattern is discernable. In Fig. 5 a schematic overview is given for the localization of the different proteolytic products observed for Ara h 3. Bands A, B, and C (∼42–45 kDa) encompass the whole *N*-terminal region corresponding to the complete basic subunit with minor *C*-terminal truncation of band B and C. The smallest band on SDS-PAGE at ∼13kDa, band L, is a truncated *N*-terminal fragment of the acidic subunit including at least res. 1–78. The same pattern is found for band G (∼27 kDa), starting at the *N*-terminus and including at least res. 1–194. Band F (∼29 kDa) follows the pattern of band G, however, fewer peptides are identified (3), limiting interpretation. A complementary picture arises for band K (∼16 kDa), with the first peptide identified at res. 176–188 and ranging upto res. 279. Band K corresponds to an *N*-terminally truncated acidic subunit, which most likely extends until the *C*-terminus of the acidic subunit at N331. The other bands D, E, I, and J are only defined by a minimal set of peptides (1 or 2) and therefore assignment into the sequence can only be done with limited confidence (hatched bars in Fig. 5). However, the peptides identified do correlate these bands to the acidic subunit. No processing of the basic subunit is observed. Sequence coverage is partially related to protein abundance in each band. The SDS-PAGE intensities of bands A, B, C, and H are reflected in their extensive peptide yields, whereas minor SDS-PAGE bands such as D, E, I, and J bands give lower sequence coverage.

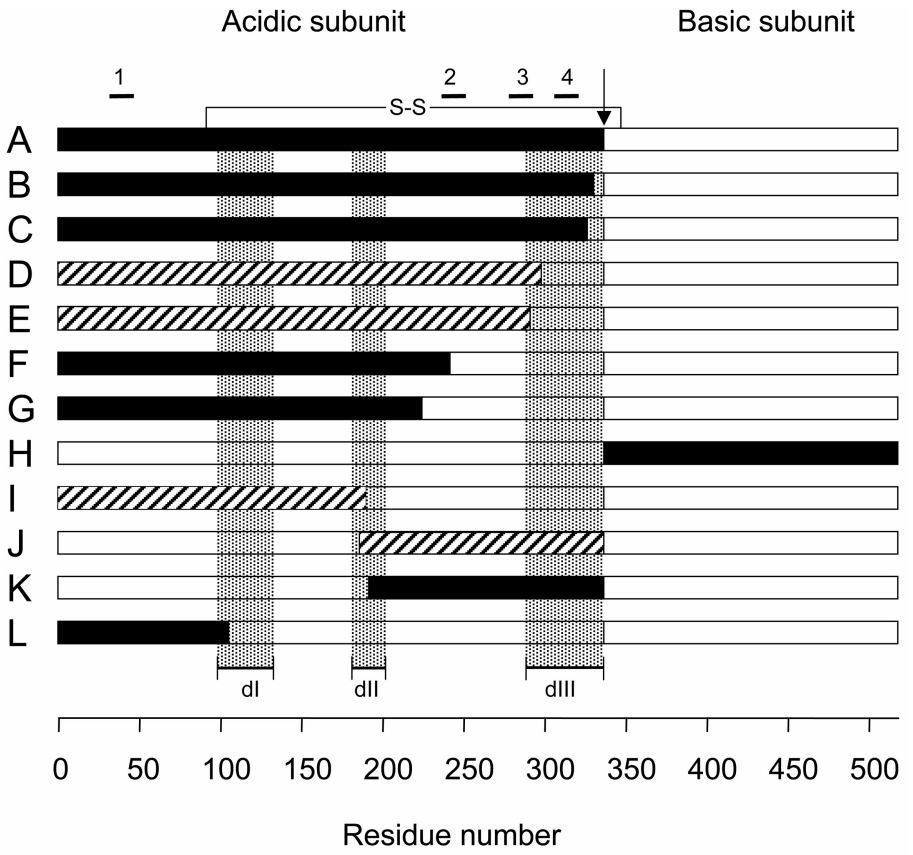

Figure 5. Schematic overview of the proteolytic truncation products A–L (black horizontal bars) in relation to the amino acid sequence of Ara h 3. Hatched bars indicate tentative assignment (only 1 or 2 peptides recovered). Vertical grey areas marked dI, dII, and dIII indicate regions of disordered structure. On top the acidic subunit and basic subunit are indicated, and the location of epitopes 1, 2, 3, and 4 [13, 15] and the interchain disulfide bridge are shown. An arrow indicates he cleavage site between the acidic and basic subunit.

4 Discussion

4.1 Subunit composition

SDS-PAGE analysis and mass spectrometric analysis give a consistent picture of the Ara h 3 acidic/basic subunit composition linked by a disulfide bridge (acidic subunit C88–basic subunit C338) analogous to the subunit organization of soy glycinin. The basic subunit (band H ~23 kDa on SDS-PAGE, 20505 Da by MALDI) appears essentially as a single polypeptide in Ara h 3. For the acidic subunit multiple bands are observed ranging from 13–45 kDa on SDS-PAGE, indicating proteolytic truncation of the (pro)protein in the peanut. Although relative intensities in MALDI mass spectra can generally not be taken as quantitative measure for protein abundance, relative intensities of highly homologous proteins in a mixture can be interpreted semiquantitatively. Based on SDS-PAGE and MALDI-MS the most abundant polypeptides comprising the mature Ara h 3 in peanuts are A, B, C, and H.

Under nonreducing conditions, the heterodimers comprised of H–L, H–A, H–B, and H–C are ~equally abundant and are responsible for >90% of the integrated MALDI signal intensity in the high-molecular-mass range (Fig. 1B). The same can be concluded from SDS-PAGE, for bands A, B, C, and H. The disproportionately low staining intensity of Coomassie blue for low-molecular-mass proteins (<15 kDa) may account for the faintness of band L. The other bands are present in smaller quantities relative to polypeptides A, B, C, H, and L.

4.2 Localization of proteolytic fragments

In Fig. 5 a schematic overview correlating the Ara h 3 sequence with the different proteolytic products is shown. Bands A, B, C, (F), G, and L are *N*-terminal polypeptides of the acidic subunit of the mature Ara h 3 molecule truncated *C*-terminally at different positions, and are characterized by an apparent molecular weight on SDS-PAGE of 45, 43, 42

(28), 27, and 13 kDa, respectively. Conversely, band K (and J) are *C*-terminal polypeptides of the acidic subunit truncated *N*-terminally, and are characterized by an apparent molecular mass of 16 (and 17) kDa.

The Ara h 3 molecule as isolated has predominantly a hexameric structure but also a minor fraction in the trimeric form is observed (<10%, judged from gel permeation chromatography) [16]. The Ara h 3 used in this study is from the same batch as used in [16]. The potential truncation sites of Ara h 3 were analyzed using the recently published 3-D structure of the soy glycinin A3B4 homohexamer [21]. Proteolytic cleavage by an Asn-specific endoprotease yielding the acidic and the basic subunit is required for hexamer formation. In the acidic subunit of glycinin, three binding regions are essential for proper assembly of the trimeric proproteins into the mature hexameric structure. The Ara h 3 equivalents of these regions are PGCPSTYEEP (region I; res. 86–95), KVR(/H)R (region II; res. 127–130) and LSPD (region III; res. 284–287), respectively. The residues between binding regions I and II (res. 96–126) show a highly disordered structure; the same is observed for the residues observed between res. 180 and 200. A third disordered region is found between binding region III and the *C*-terminus of the acidic subunit, which is a region of high variability in the glycinin protein family. We hypothesized that proteolytic truncation most likely occurs in the disordered regions that are not critically involved in hexamer formation. The disordered regions are shown in Fig. 5 (dI–dIII). Indeed, the molecular masses of the large polypeptides A, B, and C (and D and E) indicate truncation in the disordered region following binding region III. Similarly, polypeptide L corresponds to truncation in the disordered region following binding region I. This small polypeptide is covalently linked by the intermolecular disulfide bond to the basic subunit at C88–C338, consistent with the L–H MALDI signal at 32.6 kDa in Fig. 1C. Given the high mass accuracy (<0.2%) in the low-mass window and the high sequence homology for the *N*-terminal ~110 amino acids a tentative assignment of the *C*-terminus of polypeptide L can be made. Calculation gives a mass of 12 484 Da for the polypeptide spanning ISFRQQ...SQRP res. 1–107, indicating proteolysis by a proline specific endoprotease. The sequence of peptide L includes binding region I, therefore association in the hexameric structure can be rationalized. The proteolysis products K, I, and J are truncated in the second disordered region following res. 180. Polypeptide I is *C*-terminally processed analogous to polypeptides A–G and L. Polypeptides J and K are interesting since C88 is not included in these *N*-terminally truncated polypeptides, therefore they cannot form the disulfide bond with the basic subunit. Apparently, the noncovalent interactions in the trimer/hexamer are sufficient to retain these polypeptides in the oligomeric Ara h 3 structure. Proteolysis in the disordered regions of the acidic subunit accounts for the majority of the polypeptides observed including the high abundance polypeptides A, B, C, H, and L. Only polypeptides F and G have molecular masses that do not correlate with any of the three disordered regions in the acidic subunit. Interactions in the hexamer for these polypeptides are limited to binding regions I and II.

Taken together, the most abundant polypeptides comprising the mature Ara h 3 in peanuts, A, B, C, and H, follow the canonical glycinin folding pattern that make-up the hexameric structure. Indeed the relative abundances of the different truncation products, observed by both MALDI and SDS-PAGE, are consistent with this rationale.

4.3 Sequence heterogeneity, variants, and processing products

Sequence analysis of the five (nucleotide) database entries for Ara h 3 indicates sequence variability mainly due to homologous substitutions and minor gaps. Protein level analysis of Ara h 3 by MS shows a set of peptides that does not correspond uniquely to one of the four sequences, but instead constitutes a sixth sequence homologous to the other five Ara h 3 sequences. In addition to the proteolytic truncation of polypeptides originating from a single sequence, also microheterogeneities are observed at two positions (in both the acidic and basic subunit) indicating the presence of at least five Ara h 3 genes for the single peanut variety (Runner cultivar) in this study. The primary sequence was also analyzed with respect to potential post-translational modifications including glycosilation, glycation, phosphorylation, and sulfation. The Ara h 3 sequence contains three potential glycosilation sites having the N-X-S/T/C (with X is not P) motif: QQPEENACQFQR (res. 10–12), GENESDEQGAIVTVR (res. 267–269) and RNRSPDI (res. 349–351). The first and third tryptic peptides located at/near the *N*-termini of either the acidic or basic subunit are not recovered by MS, even taking into account the different possible modifications. The second potentially modified tryptic peptide is recovered by MS in its native form (in different varieties, seeTable 1), excluding glycosylation at this site. Peptides containing S or T were analyzed for phosphorylation, however, no phosphorylated peptides were recovered. Taken together, we conclude that Ara h 3 is mainly proteolytically modified, with possible glycosylation at residues 10–12 and/or 349–351.

4.4 Consequences for immunological research

Epitope scanning by IgE-binding to synthetic peptides has been shown to be a very powerful technique when applied to peanut allergens in general (reviewed by Burks *et al.* [4]). Four IgE-epitopes have been reported for Ara h 3 and all epitopes were located on the acidic subunit [15]. The dominant epitope VTVRGG**L**R**IL**S**P**DRK (res. 276–290,

amino acids critical for IgE binding in boldface) is recognized by all Ara h 3-allergic patients ($N = 8$) [15]. A recent report showed that 87% of symptomatic peanut allergic patients had specific IgE directed against this dominant epitope [26]. Bands A, B, C, and K, and probably D and J contain the *C*-terminal part of the acidic subunit rendering this dominant epitope. It is interesting to note that the 14 kDa *N*-terminal part of the acidic subunit that was originally identified as Ara h 3 [14] only contains a minor IgE-epitope [15]. Our data show that some of the mapped IgE-epitopes are only partially present on the *C*-terminally processed acidic subunit (bands D, E, F, and G) and these peptides can therefore not be the sensitizing allergen giving rise to IgE. Probably, the larger forms of the acidic subunit (band A, B, and C) are the primary sensitizing allergens for these particular IgE-epitopes. In a previous paper we showed by IgE-blotting that peanut-derived Ara h 3 has IgE-binding sites on the acidic as well as on the basic subunit, with the acidic subunits stained more intensely [16]. For soy glycinin, IgE-binding sites were found mainly on the acidic subunit [27 – 29], while the basic subunit was reported to contain a minority of the IgE epitopes [29], supporting the suggestion that the acidic subunit is the most allergenic subunit for legume glycinins.

For peanut-derived Ara h 3 we observed multiple peptides for the *N*-terminal part of the dominant epitope. Indeed the tryptic peptides identified (1603.76, 1590.78, and 1618.75 Da) showed the sequence microheterogeneity of this epitope including the R $\rightarrow$ K variation (VTVRGGLRILSPDRK, variation underlined). Although the variation does not affect amino acids identified to be critical for IgE-binding [15], it does show that the DNA level sequence and resulting recombinant protein do not necessarily correlate with the protein as isolated from its native environment. In the peanut, a much richer Ara h 3 protein pool is found than for the *Escherichia coli*-derived recombinant protein/peptides used for epitope scanning experiments. The presence of multiple Ara h 3 genes and gene products in the peanut, in addition to proteolytic *N*- and *C*–terminal truncation, yields a much wider repertoire of allergens than a single-chain *E. coli*-derived recombinant protein, probably bearing potential IgE-epitopes that are not picked-up with recombinant protein- or synthetic-peptide approaches. The amino acid sequence of allergens purified from their native source gives the molecular information required for detailed analysis of allergen structure, and structure variation due to DNA sequence variation and post-translational processing.

4.5 Conclusions

In addition to the normal processing of proteins in the glycinin family, Ara h 3 is further proteolytically truncated in the acidic subunit at multiple sites. Sequence variants are observed at the peptide level and indicate the presence of multiple Ara h 3 genes in peanuts. Characterization of the allergenicity of Ara h 3 should therefore also include IgE-binding studies with peanut-derived Ara h 3, providing the high degree of variation in the Ara h 3 protein structure, as this is what peanut-allergic individuals are confronted with.

This work is part of research program No. 49 "Mass spectrometric imaging and structural analysis of biomacromolecules" of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)", which is financially supported by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)". A contribution of the University of Nebraska Agricultural Research Division, Lincoln, Nebraska, USA 68583. Journal Series No. 14543.

5 References

[1] Sampson, H. A., McCaskill, C. C., Food hypersensitivity and atopic dermatitis: Evaluation of 113 patients. *J. Pediatr.* 1985, *107*, 669 – 675.

[2] Sachs, M. I., Jones, R. T., Yunginger, J. W., Isolation and partial characterization of a major peanut allergen. *J. Allergy Clin. Immunol.* 1981, *67*, 27 – 34.

[3] Barnett, D., Howden, M. E., Partial characterization of an allergenic glycoprotein from peanut (*Arachis hypogaea* I.). *Biochim. Biophys. Acta* 1986, *882*, 97 – 105.

[4] Burks, W., Sampson, H. A., Bannon, G. A., Peanut allergens. *Allergy* 1998, *53*, 725 – 730.

[5] Buschmann, L., Petersen, A., Schlaak, M., Becker, W. M., Reinvestigation of the major peanut allergen Ara h 1 on molecular level. *Monogr. Allergy* 1996, *32*, 92 – 98.

[6] Burks, A. W., Williams, L. W., Helm, R. M., Connaughton, C., *et al.*, Identification of a major peanut allergen, Ara h 1, in patients with atopic dermatitis and positive peanut challenges. *J. Allergy Clin. Immunol.* 1991, *88*, 172 – 179.

[7] Koppelman, S. J., Bruijnzeel-Koomen, C. A., Hessing, M., de Jongh, H. H., Heat-induced conformational changes of Ara h 1, a major peanut allergen, do not affect its allergenic properties. *J. Biol. Chem.* 1999, *274*, 4770 – 4777.

[8] Burks, A.W., Shin, D., Cockrell, G., Stanley, J. S., *et al.*, Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur. J. Biochem.* 1997, *245*, 334 – 339.

[9] Shin, D. S., Compadre, C. M., Maleki, S. J., Kopper, R. A., *et al.*, Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.* 1998, *273*, 13753 – 13759.

[10] van der Veen, M. J., van Ree, R., Aalberse, R. C., Akkerdaas, J., *et al.*, Poor biologic activity of cross-reactive IgE directed to carbohydrate determinants of glycoproteins. *J. Allergy Clin. Immunol.* 1997, *100*, 327 – 334.

[11] de Jong, E. C., Van Zijverden, M., Spanhaak, S., Koppelman, S. J., *et al.*, Identification and partial characterization of multiple major allergens in peanut proteins. *Clin. Exp. Allergy* 1998, *28*, 743 – 751.

[12] Burks, A. W., Williams, L. W., Connaughton, C., Cockrell, G., *et al.*, Identification and characterization of a second major peanut allergen, Ara h 2, with use of the sera of patients with atopic dermatitis and positive peanut challenge. *J. Allergy Clin. Immunol.* 1992, *90*, 962 – 969.

[13] Kleber-Janke, T., Crameri, R., Appenzeller, U., Schlaak, M., Becker, W. M., Selective cloning of peanut allergens, including profilin and 2s albumins, by phage display technology. *Int. Arch. Allergy Immunol.* 1999, *119*, 265 – 274.

[14] Eigenmann, P. A., Burks, A. W., Bannon, G. A., Sampson, H. A., Identification of unique peanut and soy allergens in sera adsorbed with cross-reacting antibodies. *J. Allergy Clin. Immunol.* 1996, *98*, 969 – 978.

[15] Rabjohn, P., Helm, E. M., Stanley, J. S., West, C. M., *et al.*, Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J. Clin. Invest.* 1999, *103*, 535 – 542.

[16] Koppelman, S. J., Knol, E. F., Vlooswijk, R. A., Wensing, M., *et al.*, Peanut allergen Ara h 3: Isolation from peanut and biochemical characterization. *Allergy* 2003, *58*, 1144 – 1151.

[17] Scott, M. P., Jung, R., Muntz, K., Nielsen, N. C., 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 – 662.

[18] Jung, R., Scott, M. P., Nam, Y. W., Beaman, T. W., *et al.*, The role of proteolysis in the processing and assembly of 11s seed globulins. *Plant Cell* 1998, *10*, 343 – 357.

[19] Dickinson, C. D., Hussein, E. H., Nielsen, N. C., Role of post-translational cleavage in glycinin assembly. *Plant Cell* 1989, *1*, 459 – 469.

[20] Staswick, P. E., Hermodson, M. A., Nielsen, N. C., Identification of the cystines which link the acidic and basic components of the glycinin subunits. *J. Biol. Chem.* 1984, *259*, 13431 – 13435.

[21] Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., *et al.*, Crystal structure of soybean 11s globulin: Glycinin A3B4 homohexamer. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 7395 – 7400.

[22] Koppelman, S. J., Vlooswijk, R. A., Knippels, L. M., Hessing, M., *et al.*, Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties runner, spanish, virginia, and valencia, bred in different parts of the world. *Allergy* 2001, *56*, 132 – 137.

[23] Boldt, A., Fortunato, D., Conti, A., Petersen, A., *et al.*, Analysis of the composition of an immunoglobulin E reactive high molecular weight protein complex of peanut extract containing Ara h 1 and Ara h 3/4. *Proteomics* 2005, *5*, 675 – 686.

[24] Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, *227*, 680 – 685.

[25] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 1996, *68*, 850 – 858.

[26] Beyer, K., Ellman-Grunther, L., Jarvinen, K. M., Wood, R. A., *et al.*, Measurement of peptide-specific IgE as an additional tool in identifying patients with clinical reactivity to peanuts. *J. Allergy Clin. Immunol.* 2003, *112*, 202 – 207.

[27] Pedersen, H. S., Djurtoft, R., Antigenic and allergenic properties of acidic and basic peptide chains from glycinin. *Food Agric. Immunol.* 1989, *1*, 101 – 109.

[28] Herian, A. M., Taylor, S. L., Bush, R. K., Identification of soybean allergens by immunoblotting with sera from soy-allergic adults. *Int. Arch. Allergy Appl. Immunol.* 1990, *92*, 193 – 198.

[29] Helm, R. M., Cockrell, G., Connaughton, C., Sampson, H. A., *et al.*, A soybean G2 glycinin allergen. 2. Epitope mapping and three-dimensional modeling. *Int. Arch. Allergy Immunol.* 2000, *123*, 213 – 219.

