Electronic Supplement

Characterization of the IgE-binding activity in the basic peanut protein fraction as the amino-terminal fragment of Ara h 1

Rob C. Aalberse PhD[1], Geoffrey A. Mueller PhD.[2], Ninotska I.L. Derksen MSc[1], Joost A. Aalberse MD, PhD[3], Lori L. Edwards BSc[2], Anna Pomés PhD[4], Jonas. Lidholm[5], PhD, Theo Rispens PhD[1], Peter Briza PhD[6]

[1]Sanquin Research and Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, The Netherlands

[2]National Institute of Environmental Health Sciences, Genome Integrity and Structural Biology Laboratory, Research Triangle Park, NC, USA

[3] Huisartsenpraktijk Postjesweg Amsterdam, The Netherlands

[4]Indoor Biotechnologies, Inc., Charlottesville, VA, USA

[5]Thermo Fisher Scientific, Uppsala, Sweden

[66]Department of Biosciences, University of Salzburg, Salzburg, Austria

Abbreviations

GST: Gluthatione S-transferase
MES: 2-(N-morpholino)ethanesulfonic acid
MS: Mass Spectrometry
RP: Reversed Phase (chromatography)
SEC: Size exclusion chromatography
SP: SulphoProplyl (Cation exchange chromatography)
TEV (Tobacco Etch Virus) protease
TFA: trifluoroacetic acid

Fractionation of basic peanut proteins

Peanuts (100 g) were homogenized in 1 L 100 mM acetic acid (pH 4.1). If indicated, the extract was prepared in the presence of a protease inhibitor cocktail (cOmplete EDTA free, Merck, Germany 5 tablets per 50 ml extraction buffer). Initially, insoluble peanut material was removed by centrifugation (30 minutes at 10,000 rpm, 5000-15000 g for top and bottom of tubes, respectively, at 4^{o} C), which resulted in a milky supernatant with a lipid layer on top. However, the results reported in this communication were obtained with an HAc extract after overnight sedimentation during storage or very low-speed centrifugation at 4^{o} C (30 min at 100 rpm, 1.3 –2.0 g for top and bottom of tubes, respectively), in which all visible lipid material sedimented with the insoluble peanut matrix. With the available procedures and reagents we found no obvious differences in protein spectrum or IgE binding between the 2 extracts.

To remove remaining traces of the less basic allergens (e.g. Ara h 1, 2, 3 and 6) the supernatant was diluted with 3 volumes of water and adjusted to pH 8.0. This resulted in some visible aggregated material, which was removed by centrifugation and filtration. The clear fluid was incubated for 3 hours with DEAE Sephadex A50 (20 ml packed volume/10 g peanuts). The unbound fraction was adjusted to pH 6.0 and incubated with sulpho-propyl (SP) agarose, a strong cation exchanger. After transfer to a column, bound proteins were eluted by applying a gradient from 0 to 1 M NaCl in the 25 mM MES pH 6.0 running buffer. Two pools were made: 250-450 mM NaCl (pool A) and 450-650 mM NaCl (pool B). Most IgE-binding activity was recovered in pool A. These pools were concentrated by ultrafiltration over a 3.5 kDa cut-off filter and subjected to size-exclusion chromatography over Sephadex G50. Fractions with elution volumes corresponding to globular proteins < 15 kDa were fractionated by reversed-phase chromatography (rpHPLC) on a 4.6x50 mm ProSwift RP-1S column (Thermo Scientific) at room temperature, gradient: 5 – 95% acetonitrile in water + 0.1% TFA. Fractions of 250 µL were collected and neutralized by 50 µL 100 mM NaHCO3

The Ara h 1 polyprotein and its two derived allergens

The allergen referred to as Ara h 1 is actually a polyprotein that is cleaved into two products. The existence of polyproteins is not well integrated in the allergen nomenclature. Polyproteins are best known in virology (e.g. the Gag polyprotein of HIV) and endocrinology (e.g. pro-opiomelanocortin, which contains many polypeptide hormones). Polyproteins are cleaved into several functional proteins. In the case of plant polyproteins, this cleavage often occurs inside the vacuolar compartment by a specialized vacuolar processing enzyme (VPE) also known as legumain, which preferentially cleaves after asparagine [1]. In the case of Ara h 1, this cleavage results in the production of two very different proteins from the polyprotein precursor. The leader sequence is easily confused with the propeptide, so it may be relevant to clarify that the first 25 amino-terminal residues form the leader sequence, which is removed during the synthesis of the polyprotein and degraded. The remaining polyprotein is transferred to protein storage vacuoles and in due course cleaved by VPE.

The finding of an amino-terminal arginine in the mature Ara h 1 protein by Wichers et al [2] led to their hypothesis that an N-terminal fragment would exist that was produced by cleavage between Thr-Asn-Gln and Arg-Ser-Pro. Our findings essentially support this. However, we did not find the complete predicted fragment of 59 amino acids. This is presumably largely due to trimming by exopeptidases, both at the amino- and carboxy terminus of the initial fragment. The peanut VPE that is likely to be responsible for the initial cleavage step cleaves selectively after Asn and preferentially between Asn and Gln. Therefore, a more likely initial propeptide may be:
KSSPYQKKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDPRGHTGTTN.
In this case, the amino terminus of the carboxy-terminal fragment would be Gln-Arg-Ser-Pro. The finding of an N-terminal Arg rather than a Gln may be due to the tendency of N-terminal Gln to transform to pyroglutamate, which would have resisted N-terminal sequencing by the Edman degradation sequencing protocol used by Wichers et al [2]. Some protein with an N-terminal Arg would be produced by the action of pyroglutaminase. This activity is relatively high in seeds to promote the generation from seed storage proteins of free amino acids needed by the developing peanut embryo.

Based on MS analysis of a trypsin digest of purified nAra h 1, some of the propeptide is still present, either as part of the full-length protein or as complex with the major fragment. However, by measuring inhibition of IgE binding to recombinant propeptide by nAra h 1 we estimated that >90% of the propeptide is physically separated from the C terminal part *in planta*.

Production and purification of recombinant Arah1Pro

rArah1Pro was expressed in a pDest 565 vector with an N-terminal GST (Glutathione S-transferase) tag followed by a TEV (Tobacco Etch Virus) protease cleavage site. After sonication, the soluble cell lysate was flowed over a glutathione column and GST-Arah1Pro was eluted with PBS and 10 mM reduced glutathione. TEV protease was added and incubated at room temperature for 4 hours, or overnight at 4C. The resulting sequence was:
GKSSPYQKKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDPRGHTGTTN.
The N-terminal glycine, which replaces an alanine, is an artifact of TEV cleavage. Formally this protein should be called Ara h 1.0101 (25-83, A25G). The sample was concentrated to about 1 mL and then diluted to 5 mL with PBS to lower the reduced glutathione concentration to 2 mM, and 0.5 mM of oxidized glutathione was added, followed by another 30 minutes incubation at RT. NMR analysis demonstrated that this reduced heterogeneity in the Arah1Pro spectra (Nesbit, Maleki, Mueller et al, manuscript in preparation). The protein was further purified to homogeneity by SEC using Superdex G75 in PBS.

Sample characterization by Mass Spectrometry (MS)

The RP fractions were either reduced and alkylated using the reagents and protocol of the ProteoExtract All-in-One Trypsin Digestion Kit (EMD Millipore, Billerica, MA, USA) or analyzed without pretreatment. After desalting with C18 ZipTips (EMD Millipore), samples were loaded onto

an in-house prepared monolithic reverse phase capillary column (200 µm x 20 cm, column material Polystyrene/Divinylbenzene). Using a Dionex Ultimate 3000 nano HPLC (Thermo Fisher Scientific, Bremen, Germany), the column was developed with an acetonitrile gradient (Solvent A: 0.1% (v/v) FA/0.01% (v/v) TFA/5% (v/v) DMSO; solvent B: 0.1% (v/v) FA/0.01% (v/v) TFA/90% (v/v) ACN/5% (v/v) DMSO; 2–45% B in 60 min) at a flow rate of 800 nl/min at 55°C). The HPLC was acetonitrile gradient (Solvent A: 0.1% (v/v) FA/0.01% (v/v) TFA/5% (v/v) DMSO; solvent B: 0.1% (v/v) FA/0.01% (v/v) TFA/90% (v/v) ACN/5% (v/v) DMSO; 2–45% B in 60 min) at a flow rate of 800 nL/min at 55°C). The HPLC was directly coupled via nano electrospray (2 kV capillary voltage) to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific). The instrument was operated in Full-MS mode at maximum resolution (140,000) with both external and internal (lock-mass) calibration. Accuracy was in the 1-2 ppm range. For sequence analysis, a top 12 method MS/MS was used with the normalized fragmentation energy set to 27%. For the initial protein identification, proteins/peptides present in the samples were digested with trypsin and the obtained data were analyzed with the UniProt/Viridiplantae database (www.uniprot.org) and peanut allergen sequences from allergen.org, respectively. Only peptides with a probability score -10lgP>25 and proteins with a score -10lgP>100 were considered positive hits. Sequence identification was done with PEAKS Studio X (Bioinformatics Solutions), using ‘trypsin’ as digest tool, allowing non-specific cleavage at both ends of the peptides. Error tolerance was set at 3 ppm for peptides and 0.1 Da for fragment ions. Carbamidomethylation (C), oxidation (M) and deamidation (NQ) were used as variable modifications. Default values for de-novo sequencing and database searches were used. Undigested proteins/peptides present in the samples were identified in two ways. First, they were fragmented in a top-down approach and identified with PEAKS Studio X, using ‘none’ as digest tool. Peanut allergen sequences from allergen.org were used as database, all hits with a score -10lgP>15 were considered positive. In the second approach, multiple charged signals of the intact, unfragmented proteins/peptides were deconvoluted with the Xtract module of the Xcalibur Qual browser (Thermo Fisher Scientific). For identification of the naturally occurring allergen fragments, uncharged masses were matched to peanut proteins with the search function of the Protein/Peptide Editor of MassLynx v4.1 (Waters) with the maximum mass deviation set to 0.005 Dalton. In addition, resources of the SIB Bioinformatics Portal (www.expasy.org) were used for peptide/protein assignments.
For sequence analysis of the naturally occuring peptides, a top 12 method MS/MS was used with the normalized fragmentation energy set at 27%. Raw data were deconvoluted with the Xtract module of the Xcalibur Qual browser (Thermo Fisher Scientific). Peptide sequence identification was done with PEAKS Studio X (Bioinformatics Solutions).Resources of the SIB Bioinformatics Portal (www.expasy.org) were used for peptide/protein assignments.

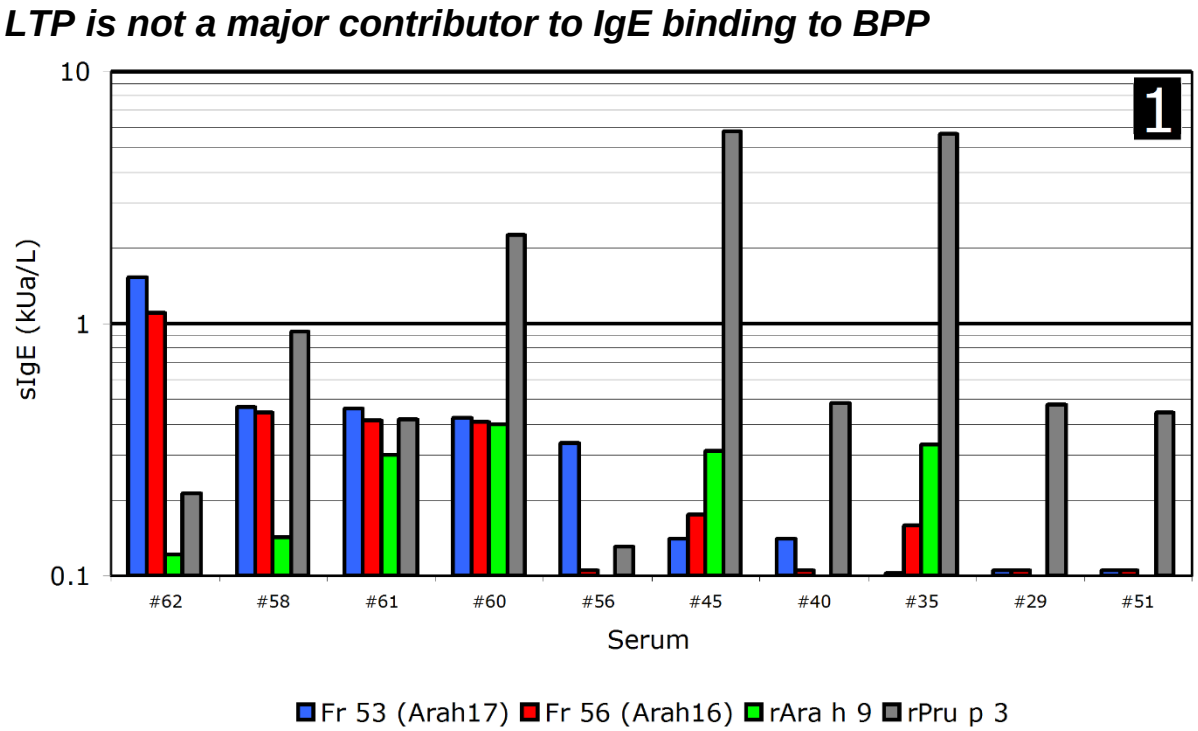


Figure E1. IgE binding by 10 selected sera from the pediatric cohort to Sepharose-coupled fractions 53 and 56 of the RP column (see figure 2 of the main text). For comparison, IgE binding to rAra h 9 (green bars) and peach LTP (grey bars) is also shown. The sera were included if a response >=0.35 kUa/L to at least one of the 4 LTPs was found. The sera were ranked by their response to the fraction that was enriched for the Ara h 17 (blue bars). The red bars represent IgE binding to the fraction enriched in Ara h 16.

Red/ox mobility shift in SDS-PAGE characteristic of nsLTPs

Using reducing conditions we found a strong band at 10 kDa, which we assumed to be LTP. Using non-reducing conditions, the 10 kDa band was very faint, which was possibly related to the presence of a strong 14 kDa band. Such a red/ox mobility shift in SDS-PAGE is characteristic of nsLTPs, including Ara h 9 (unpublished observations). The observed shift in mobility suggested the presence of disulphide-linked dimers, but no covalent LTP dimers were found by MS.

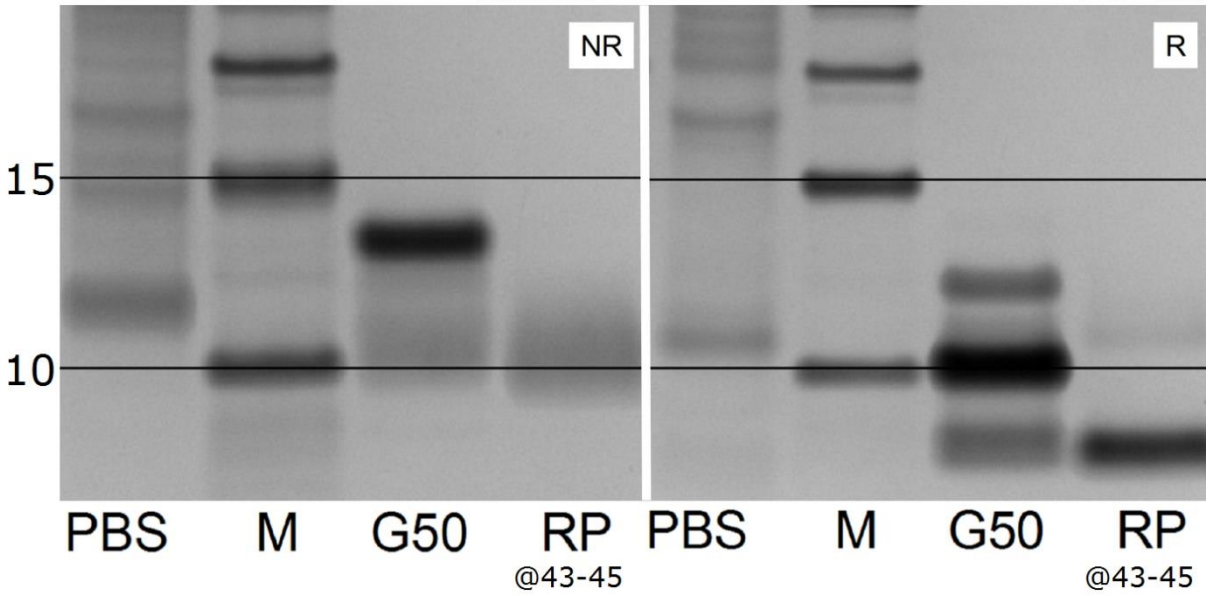


Figure E2. Anomalous effects of reduction on the electrophoretic mobility in SDS PAGE. In addition to the molecular mass marker M three peanut preparations are shown: a PBS peanut extract (**PBS**), a Sephadex G50 <15 kD fraction (**G50**) and the Reversed Phase sub-fraction pool 43-45 enriched in Arah1pro (**RP @43-45**). At left **NR**=Non-Reduced, at right **R**=Reduced.

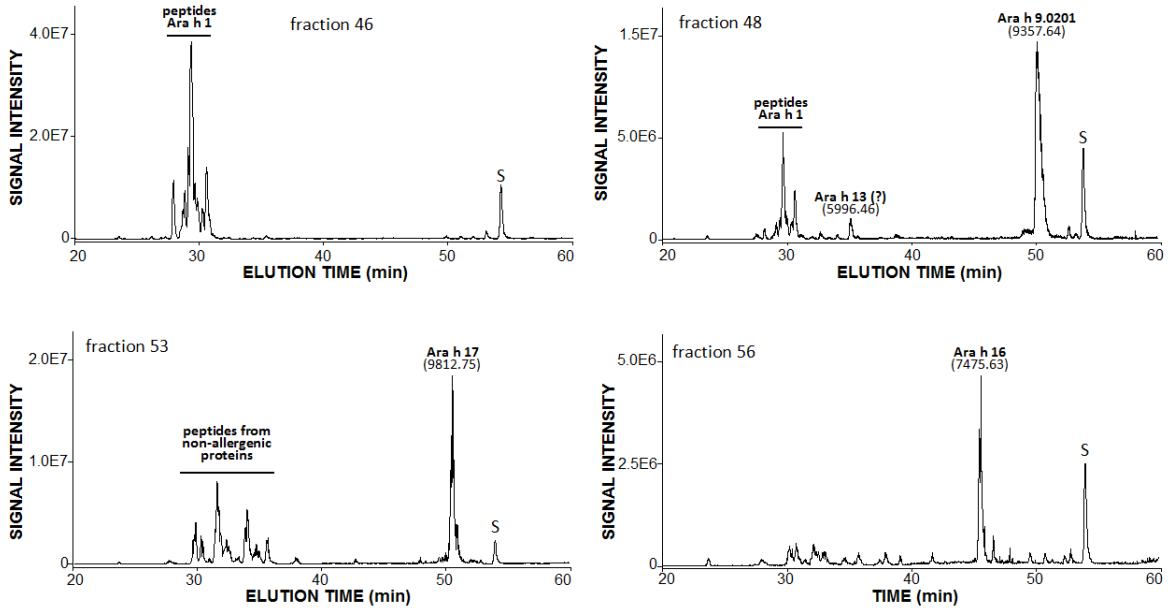


Figure E3. LC-MS/MS analysis of preparative RP chromatography fractions (see figure E6). The peak indicated by "S" is a system peak.

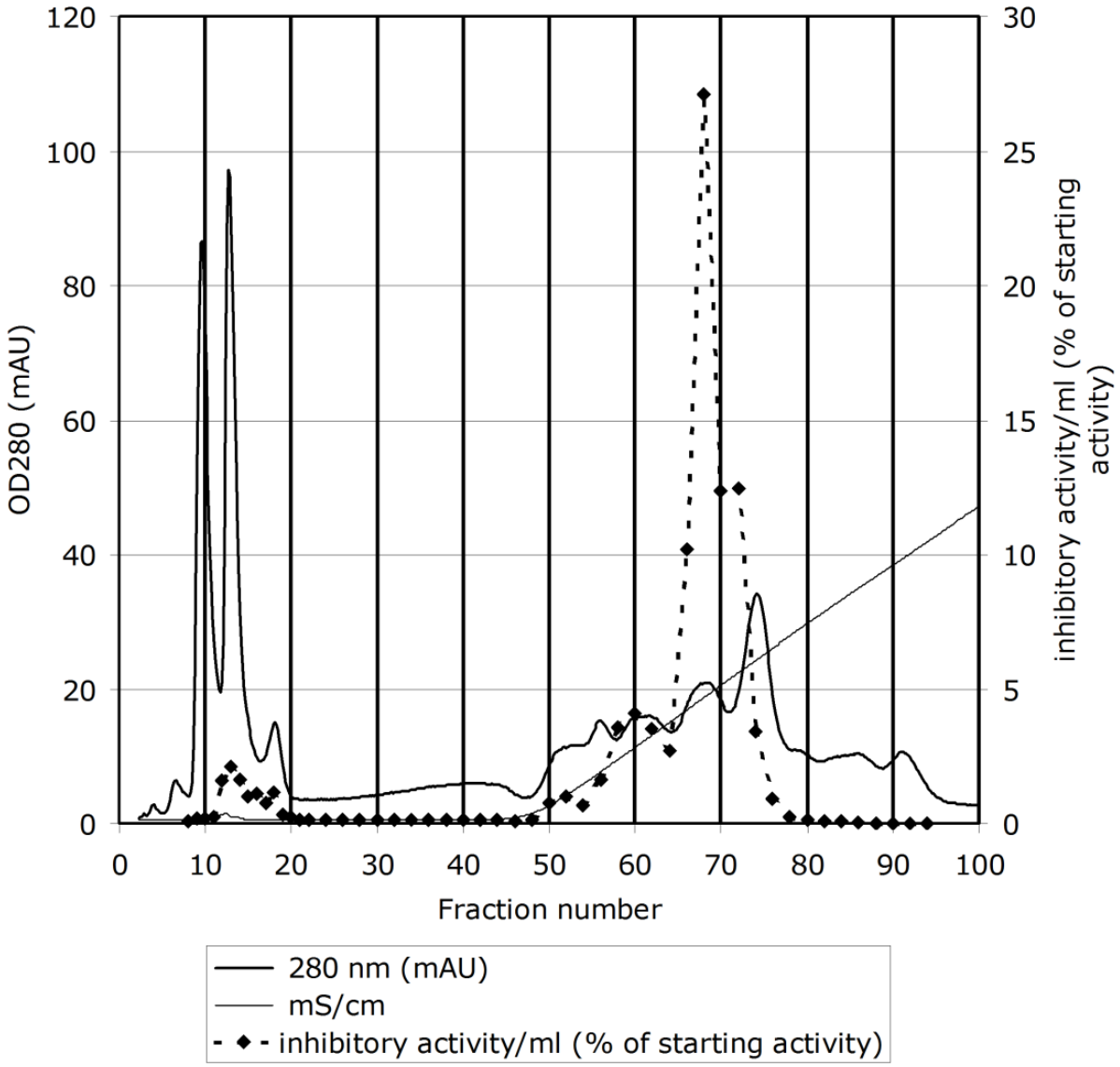


Figure E4. Elution profile from an analytical sulphopropyl cation exchange column of the 3.5 - 7 kDa fraction of BPP prepared in the presence of a protease inhibitor cocktail. Most of the allergenic activity eluted in fractions 66-72, at a conductivity of 19 mS/cm, which contained predominantly peptides with calculated isoelectric points of 7.7 - 8.3.

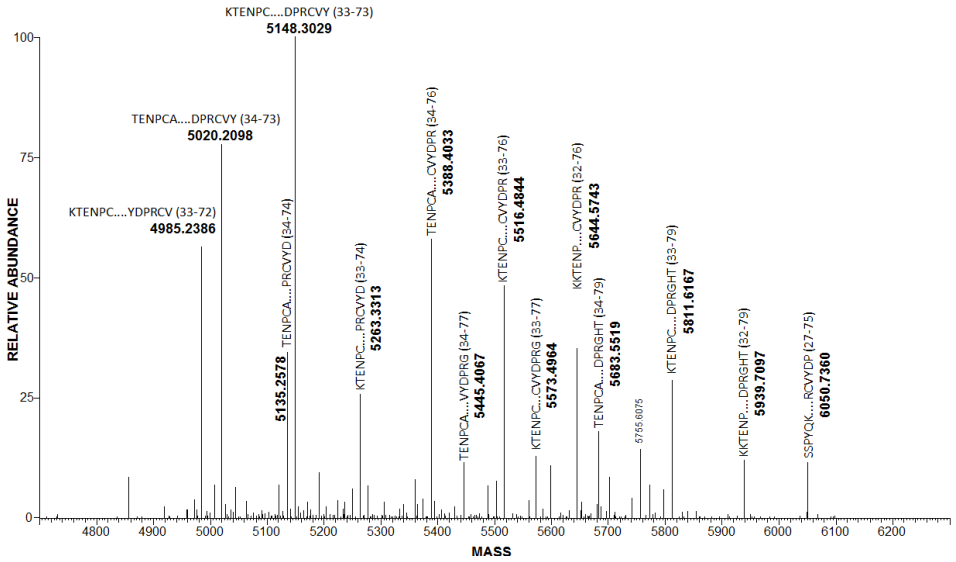


Figure E5. Spectrum of Ara h 1-peptides in RP-fraction 46 (see figure E3). Raw data of the Ara h 1-containg fractions (27 to 32 minutes elution time) were merged and deconvoluted to obtain uncharged signals. Peptide identification was based on both mass matching and MS/MS sequencing.

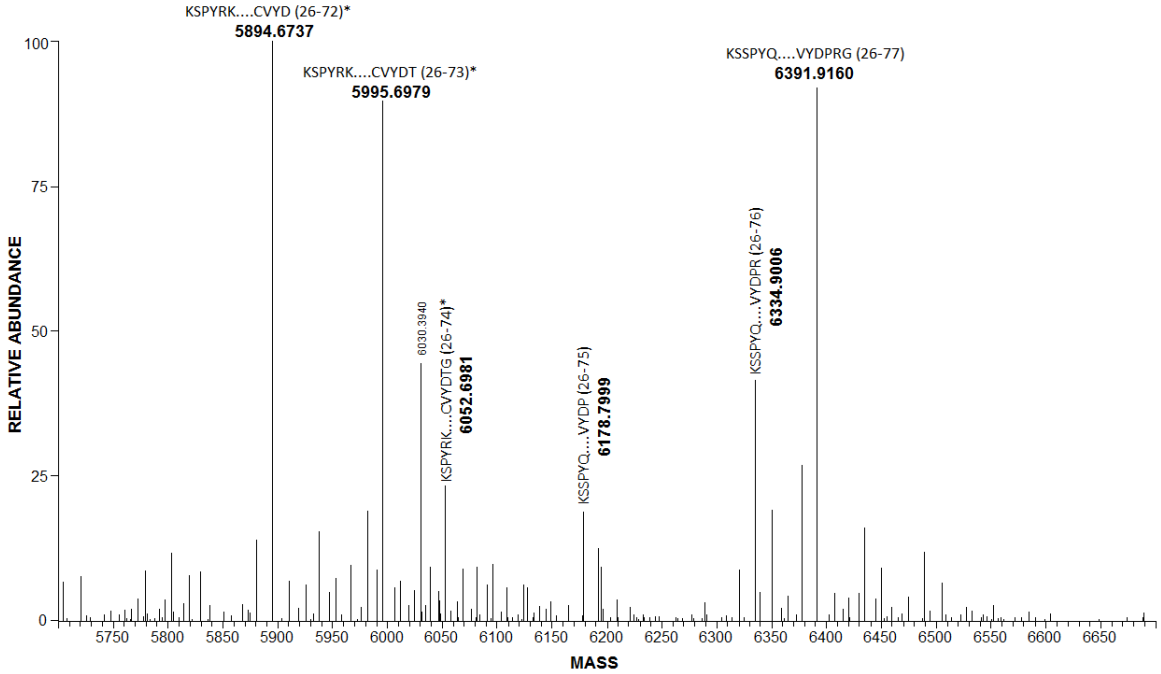
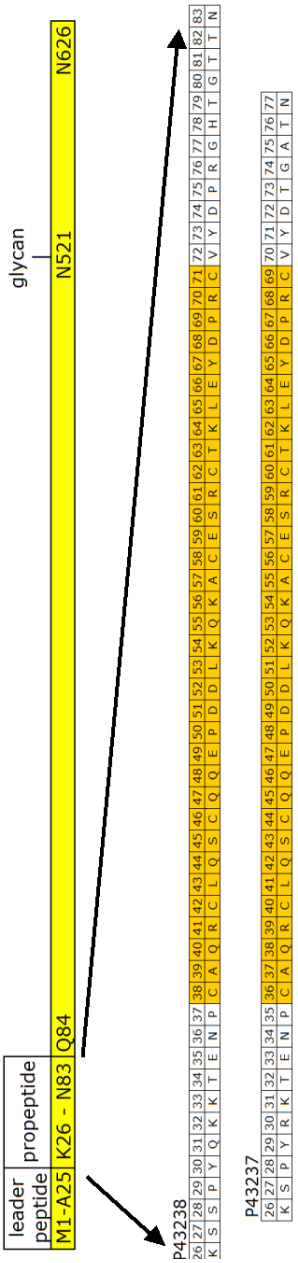


Figure E6. As for Figure E5, but now for SP-fraction 70 (see figure E4).



Fr	IEP	MS	Nterm	Cterm
64	7.75	900	YQKKTENP	- VYDPRG
64	***7.81***	***200***	***RKTENP***	***- VY***
64	7.75	200	KKTENP	- VYDPR
64	6.54	80	P	- VYDPRG
64	6.99	70	P	- VYDPRGHTGTTN
64	7.75	60	YQKKTENP	- VYDPR
66	***7.80***	***2000***	***KSPYRKTENP***	***- VYD***
68	7.74	700	KSSPYQKKTENP	- VYD
68	7.74	200	KSSPYQKKTENP	- VYDP
68	***6.27***	***100***	***RKTENP***	***- VYDT***
68	***6.27***	***30***	***YRKTENP***	***- VYD***
70	***7.80***	***1000***	***KSPYRKTENP***	***- VYDT***
70	8.29	500	KSSPYQKKTENP	- VYDPR
70	8.29	100	KSSPYQKKTENP	-
70	***7.80***	***20***	***KSPYRKTENP***	***- VYDTG***
72	8.29	2000	KSSPYQKKTENP	- VYDPRGHTGTTN
72	8.29	900	KSSPYQKKTENP	- VYDPRG
72	5.05	100	P	- VYD
74	***8.38***	***3000***	***KSPYRKTENP***	***- VY***
74	***8.39***	***500***	***KSPYRKTENP***	***-***
74	8.29	100	KSSPYQKKTENP	- VYDPRGHT
74	***8.39***	***90***	***KSPYRKTENP***	***- V***

Figure E7. Mass-spectrometric analysis of Ara1Pro-derived peptides in the fractions 64, 66, 68, 70, 72 and 74 of the sulphopropyl column of the <9 kDa fraction of BPP prepared in the presence of a protease inhibitor cocktail (see figure E5). First column Fr= fraction number; 2nd column: IEP = calculated isoeletric point of the peptide; MS=relative signal strength in the mass spectrum; N-term= sequence predeeding the 6-cysteine core region; Cterm= sequence following the 6-cysteine core region. The sequences in bold italics are from Ara h 1 P43237 (provisional WHO-IUIS Ara h 1.0102); the other sequences are from P43238 (provisional WHO-IUIS Ara h 1.0101) or from both subtypes.

References

1. Yamada K, Shimada T, Kondo M, Nishimura M, Hara-Nishimura I. Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme. J Biol Chem. 1999 22;274:2563-70.
2. Wichers HJ, De Beijer T, Savelkoul HF, Van Amerongen A. The major peanut allergen Ara h 1 and its cleaved-off N-terminal peptide; possible implications for peanut allergen detection. J Agric Food Chem. 2004;52:4903-7.