

ORIGINAL ARTICLE

Allergens

Identification of the amino-terminal fragment of Ara h 1 as a major target of the IgE-binding activity in the basic peanut protein fraction

Rob C. Aalberse¹  | Geoffrey A. Mueller²  | Ninotska I. L. Derksen¹ |
Joost A. Aalberse³ | Lori L. Edwards²  | Anna Pomés⁴  | Jonas Lidholm⁵  |
Theo Rispens¹  | Peter Briza⁶ 

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

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

³Huisartsenpraktijk Postjesweg, Amsterdam, The Netherlands

⁴Indoor Biotechnologies, Inc, Charlottesville, VA, USA

⁵Thermo Fisher Scientific, Uppsala, Sweden

⁶Department of Biosciences, University of Salzburg, Salzburg, Austria

Correspondence

Rob C. Aalberse, Division of Research, Department of Immunopathology, Sanquin Blood Supply, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands.
Email: r.aalberse@sanquin.nl

Funding information

This study was supported in part by the Intramural Research Program of the National Institute of Environmental Health Sciences (Z01-ES102906-01, G Mueller).

Abstract

Background: Small, basic peanut proteins are often poorly extracted in pH-neutral buffers that are optimal for the extraction of peanut storage proteins such as Ara h 1. As a result, such proteins are easily missed as potential allergens.

Objective: To analyse the allergenic composition of the basic peanut protein (BPP) fraction.

Methods: A peanut extract prepared at pH 4 was fractionated by physicochemical procedures. Chemical analysis was performed by SDS-PAGE and mass spectrometry. Because immunoblotting was found to be inefficient for most of these small basic proteins, IgE-binding activity was measured by coupling the fractions to CNBr-activated Sepharose, followed by incubation with sera from 55 Dutch peanut-allergic children and ¹²⁵I-labelled anti-IgE.

Results: Most IgE reactivity of the BPP fraction was due to the 5-7 kDa amino-terminal fragment of Ara h 1. This finding was confirmed by the use of the fragment in recombinant form, to which 25/55 of the sera was IgE-positive.

Conclusion: The amino-terminal fragment of Ara h 1, a member of a family of small anti-microbial proteins, is an allergen independent of the carboxy-terminal fragment of Ara h 1.

KEYWORDS

allergens and epitopes, food allergy, IgE, immunologic tests

1 | INTRODUCTION

Peanut allergens largely belong to two physiological types: storage proteins and defence-related proteins. Some of the defence-related proteins, such as lipid transfer proteins (LTPs), are small (<10 kDa) basic peanut proteins (BPPs). The aim of our study was to investigate which of these BPPs were most relevant in relation to peanut sensitization (ie induction of sIgE, irrespective of clinical symptomatology upon exposure).

Immunoblotting, which reliably detects the known LTPs, was previously found to be inefficient in identifying other IgE-binding

proteins in the BPP fraction.¹ Therefore, we set out to identify IgE-binding BPPs using a traditional RAST-type assay based on covalent coupling of BPP fractions to CNBr-activated Sepharose beads and detection of bound IgE with ¹²⁵I-labelled anti-IgE. In addition, we used RAST inhibition to quantify IgE-binding allergenic activity. The IgE-binding activity was compared with mass spectrometric analyses, and we show that the amino-terminal fragment of Ara h 1 is a major IgE-binding component of peanut.

Some background information on the structure of Ara h 1, one of the major peanut allergens, may be desirable. It is a large (65 kDa)

protein. The recombinant allergen used in our diagnostic test is Ara h 1.0101. This is the full-length protein without the leader sequence and corresponds to the amino acid residues 26-626 encoded by the genomic sequence. We will refer to this full-length protein as rAra h 1. Ara h 1 purified from peanut extract largely lacks the amino-terminal amino acids 26-83 due to cleavage in the peanut by a vacuolar protease (for more information on vacuolar proteases producing multiple proteins from a precursor poly-protein, see Supporting information). We will refer to this processed carboxy-terminal fragment as natural Ara h 1 (nAra h 1, residues 84-626). The small protein corresponding to the amino-terminal amino acids 26-83 is the main topic of this paper. Since it is known in literature as a propeptide, we will refer to it as Ara h 1Pro. Its actual size in peanut extract proves to be smaller than predicted, because it is trimmed at both its termini.

2 | METHODS

2.1 | Patients, serum samples and serological tests

The human sera are referred to by an ID code preceded by the “#” sign. Most assays were performed with a BPP-positive reference serum. This BPP⁺ reference serum has been extensively used in the work presented in our preceding paper.¹ The BPP⁺ reference serum (total IgE 2743 kU/L) was obtained from a Dutch patient with a clinical peanut allergy. This serum was positive to rAra h 1, 2, 3, 6 and 8 and to rBet v 2 (84.7, 36.3, 5.8, 33.3, 10.0 and 1.26 kU_A/L, respectively). IgE reactivity to rAra h 9, Pru p 3 and CCD were <0.35 kU_A/L. The other reference serum used was an LTP-positive serum. The LTP⁺ reference serum (total IgE 131 kU/L) was obtained from a Dutch peach-allergic patient with a strong IgE reactivity to rPru p 3 (10.33 kU_A/L) and a weaker reactivity to rAra h 9 (6.20 kU_A/L), as measured using the ImmunoCAP (Thermo Fisher Scientific). IgE reactivity to rAra h 1, 2, 3, 6 and 8 and to rBet v 2 and CCD were all <0.35 kU_A/L. Regrettably, we have no information on the presence of symptoms upon peanut exposure. We used 55 sera from our panel of 64 Dutch paediatric patients to substantiate these results. These patients, with codes #01 to #64, were DBPCFC-tested, as described elsewhere.¹ For the ImmunoCAP results of these sera, see Ref. ¹ The study of these patients was approved by the local medical ethics review boards (METC, UMC Utrecht; project number 05/084), and informed consent was obtained for all subjects. IgE to the peanut extract and to the peanut fractions was measured using the Sanquin RAST¹ based on allergens coupled to CNBr-activated Sepharose and detected by ¹²⁵I-anti-IgE.

IgE to purified natural Ara h 1 was measured by a modified RAST protocol, in which two mouse monoclonal antibodies to Ara h 1 were coupled to CNBr-activated Sepharose and subsequently loaded with purified nAra h 1 (100 ng/test). The monoclonal antibodies were directed to non-overlapping epitopes. Both the monoclonal antibodies and the purified nAra h 1 were provided by Indoor Biotechnologies. We loaded the anti-Ara h 1 coated beads separately with purified nAra h 1 and mixed these

two bead suspensions. Per test, we used the equivalent of 100 ng Ara h 1. We incubated these nAra h 1-coated beads with 50 µl patient serum overnight, washed, added ¹²⁵I-labelled affinity-purified sheep anti-IgE, incubated overnight and washed. Sepharose-bound radioactivity was measured and converted to kUa/L using a human-mouse chimeric IgE antibody to Der p 2 as reference. IgE to recombinant Ara h 1Pro (see below) was measured by coupling the recombinant protein to CNBr-activated Sepharose. Per test, we used the equivalent of 1 µg Ara h 1Pro.

2.2 | Fractionation of the basic peanut protein (BPP) fraction

Details on the production of the BPP fractions are presented in the Supporting information.

2.3 | Monitoring IgE-binding activity of peanut fractions by RAST and RAST inhibition

BPP fractions obtained by reversed-phase (RP) chromatography (for details on these fractions, see Figure 2A) were tested initially by direct RAST. Of each fraction, 10 µg protein was coupled to 50 mg CNBr-activated Sepharose. The IgE-binding activity of each fraction was tested by incubating 0.5 mg allergen-coated Sepharose with the BPP⁺ reference serum or with the LTP⁺ reference serum, followed by washing and incubation with ¹²⁵I anti-IgE.

For the initial screening also, two LTP-enriched fractions (fraction 53 and 56 in Figure 2A) were used in the direct RAST protocol to test for LTP-related IgE binding in sera from the panel of Dutch paediatric patients.¹

For RAST inhibition, the allergen sample under investigation was titrated and pre-incubated with the BPP⁺ reference serum. Residual IgE binding was measured by incubation with Sepharose-coupled fraction 45 (see in Figure 2A), followed by washing and incubation with ¹²⁵I anti-IgE.

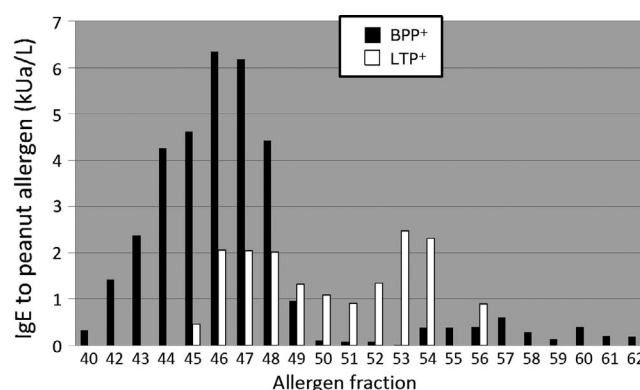


FIGURE 1 IgE-binding activity of Sepharose-coupled reversed-phase chromatography column fractions for the two reference sera BPP⁺ (black bars) and LTP⁺ (white bars)

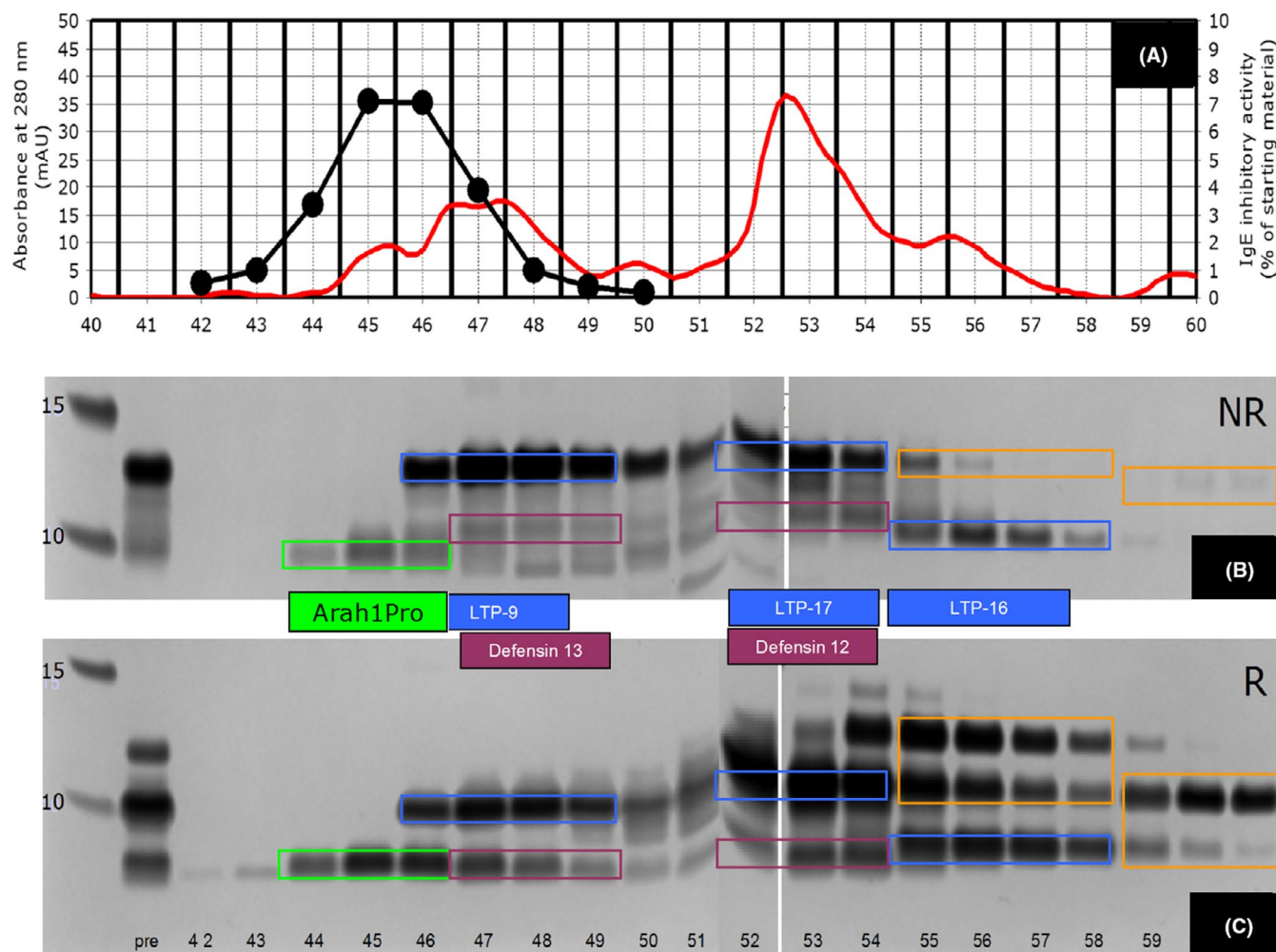


FIGURE 2 Frame A: Preparative reversed-phase (RP) chromatography of the peanut fraction purified by cation exchange and SEC. The black line indicates allergenic activity, as measured by RAST inhibition using the BPP⁺ reference serum. The allergenic activity is indicated on the Y-axis on the right using the starting material (ie the peanut fraction purified by cation exchange and SEC) as 100% value. The red line indicates the optical density at 280 nm. Frames B and C: Silver-stained SDS-PAGE (B: NR = non-reduced; C: R = Reduced) of the starting material ("pre") and the RP fractions. The green, blue and purple coloured rectangles indicate our provisional assignments of the allergens mentioned in the text. Arah1Pro is the propeptide of Ara h 1. LTP-9, 16 and 17 are Ara h 9, Ara h 16 and Ara h 17; defensin 12 and 13 are Ara h 12 and Ara h 13. The proteins indicated by the orange-coloured rectangle of fractions 59-61 have not been investigated

2.4 | Sample characterization by mass spectrometry

The methods used for mass spectrometry are specified in the Supporting information. Representative spectra are shown in Figures S3, S5 and S6 in the Supporting information.

2.5 | Production and purification of recombinant Arah1Pro

Recombinant Arah1Pro was prepared as described in the Supporting information. The connectivity of the disulphide bridges was investigated by nuclear magnetic resonance and mass spectrometry.

3 | RESULTS

By mass spectroscopic analyses (Wildner et al, manuscript in preparation), we identified two novel peanut LTPs, which have been provisionally accepted by the WHO/IUIS Allergen Nomenclature Committee in their database (www.allergen.org) as Ara h 16 (7011 Da, 35% identity to Ara h 9) and Ara h 17 (9349 Da, 62% identity to Ara h 9). Based on the high allergenicity of other LTPs, particularly peach LTP (Pru p 3),² we initially focused on the IgE-binding activity of the three identified peanut LTPs in the BPP fraction. However, as shown in Figure S1 in the Supporting information, in the Dutch paediatric panel, these LTPs were found to be only minor contributors to the IgE-binding activity among the group of small basic peanut proteins.

Upon reversed-phase (RP) fractionation, the strongest IgE-binding activity of the BPP fraction was found in fractions eluting before the LTPs (Figures 1 and 2). Upon fractionation by size-exclusion chromatography (SEC), most IgE-binding activity was found in fractions eluting slightly after the LTPs (results not shown). This chromatographic behaviour was similar to that of the two peanut defensins Ara h 12 and Ara h 13.³ However, analysis by mass spectrometry indicated that the strongest IgE-reactive RP fraction, which eluted just before the defensins, contained peptides derived from Ara h 1Pro, but were smaller.

To exclude the possibility of *in vitro* breakdown, we prepared peanut extract in the presence of a protease inhibitor cocktail and immediately isolated the 3.5–7 kDa fraction. The fractions with proteins smaller than LTPs (as judged by SDS-PAGE) were pooled, concentrated and separated by SP chromatography. The fractions were tested by RAST inhibition, using Sepharose-coupled fraction 45 (see Figure 2A) on the solid phase and the BPP⁺ reference serum as IgE source. Mass spectrometry analysis of peptides related to Ara h 1 indicated that these fractions contained trimmed Ara h 1Pro-derived peptides with calculated isoelectric points between 7.7 and 8.3 (Figures S4 and S7). The observations support the notion that not only the propeptide, but also the amino- and carboxy-terminally trimmed peptides, are formed *in planta*.

To exclude that the IgE reactivity to the fraction containing Ara h 1Pro was due to contaminating proteins, we produced the Ara h 1Pro as recombinant protein and tested it for IgE binding. Using a cut-off of 0.35 kUa/L, 25/55 sera were positive for rAra h 1Pro, compared to 17 for nAra h 1 (Figure 3). This is compatible with the conclusion that rAra h 1Pro and/or trimmed variants are a major source of the IgE-binding activity in the BPP fraction. Six of the 25 Ara h 1Pro-positive children (24%) passed the DBPCFC (ie were peanut-tolerant), compared to 2/17 = 11.8% for nAra h 1.

4 | DISCUSSION

Our results with Ara h 1Pro indicate that IgE-binding activity of small basic anti-microbial proteins is easily overlooked. Relevant factors for their optimal identification are extraction at low pH, IgE detection techniques other than immunoblotting and ways to deal with size heterogeneity due to proteolytic trimming. The latter feature makes it very hard to prepare a sufficiently pure natural protein, so the production of a recombinant protein is essential. With these cysteine-rich proteins, the connectivity of the disulphides needs to be established. The potential initial production as part of a poly-protein is also a complicating factor.

In the case of Ara h 1Pro, we find it likely that it acts as a predominantly independent molecular sensitizer and suggest that it in that sense may be relevant to regard it as a distinct peanut allergen. A similar conclusion was reached by Downs et al⁴ for the propeptide from walnut vicilin. It is, however, important to stress that the immunoblots shown in their paper do not provide proof of IgE-binding to a propeptide specifically. The walnut fraction that was tested for

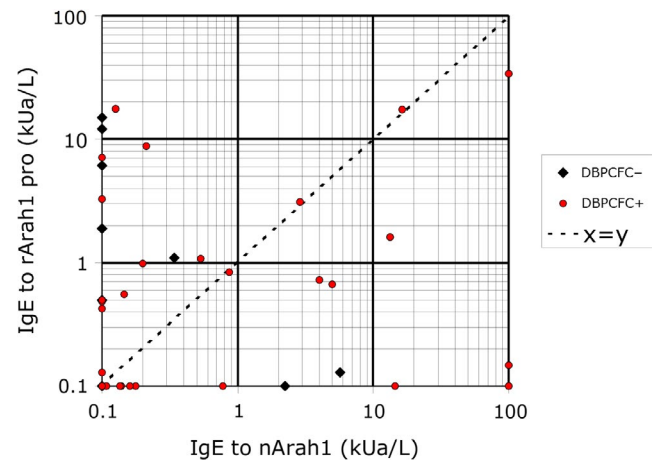


FIGURE 3 Comparison between IgE binding to the rAra h 1Pro and natural Ara h 1 purified from peanut extract. Of all sera tested, 17 were double-negative. The data are based on 55 sera from the Dutch paediatric panel,¹ of which 36 were positive in the DBPCFC

IgE binding was obtained by SEC fractionation, without further sub-fractionation. This fraction evidently contains the walnut vicilin pro-peptide, but presumably also other proteins such as defensin-related allergens which have a molecular weight of 5–6 kDa.³

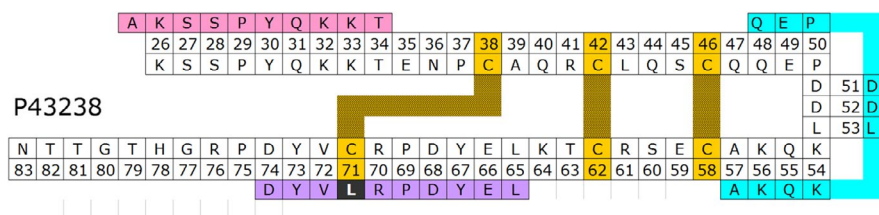
A potential weakness of the current data set is that the results for IgE to nAra h 1 could be influenced by trace amounts of potent contaminating allergens such as Ara h 2. We minimized such a contamination effect by using a modified RAST protocol based on capture of the purified allergen by monoclonal anti-Ara h 1 coupled to the solid phase and using a low dose of allergen (100 ng/test). As a consequence of this low dose, IgE levels may be underestimated in sera with a high level of IgG to nAra h 1.

Several small cysteine-rich BPPs that have been described as allergens are defence-related proteins belonging to the defensin-, LTP- or alpha-hairpinin family.⁵ Ara h 1Pro with its repeated CXXX motifs presumably belongs to the latter.⁶ Figure 4 shows a cartoon of the structure of Ara h 1Pro (based on mass spectrometry and a preliminary nuclear magnetic resonance analysis) and the three IgE epitopes predicted from synthetic peptides AKSSPYQKKT, QEPDDLKQKA and LEYDPRLVYD^(7,8); the first two peptides contained no cysteines; the single cysteine in the 3rd sequence was replaced for technical reasons by leucine). It will be interesting to see how efficient these peptides inhibit IgE binding to fully folded rAra h 1Pro. We expect that some IgE epitopes depend on the intact double-helical hairpin configuration.

Ye et al⁹ described an antifungal peptide from peanuts with sequence similarity to Ara h 1Pro, which they named Hypogin 1. Using the Edman degradation procedure, they reported a partial amino-terminal sequence KSPYYQKKTENPQAQRQLQSDDEPAKLKXKRGTSRDXTY. This sequence deviates from the Ara h 1Pro sequence, suggesting that even though the propeptide was present in their peanut fraction, it may not have been entirely pure. Further experiments are needed to substantiate the claim that Ara h 1Pro has anti-microbial activity.

The more we know about these small cysteine-rich basic proteins, the more it becomes evident that it is essential to show IgE-binding to recombinantly expressed proteins to establish the IgE-binding activity

FIGURE 4 Ara h1Pro cartoon showing the location of the three synthetic IgE-binding peptides (pink, cyan and purple) reported by Burks et al⁷



of such proteins. Our results with the recombinant protein show that Ara h1Pro has IgE-binding activity. In addition to IgE-binding, it will be important to establish allergenic activity in basophil/ mast cell systems. The small size of these tightly folded cysteine-rich proteins is likely to limit number of antibody molecules that fit on their surface, particularly if their N- and C-terminal termini are trimmed. Based on our results with rAra h1Pro, it is evident that it has a substantial IgE-binding activity. In view of its small size, it will be interesting to investigate if its activity in biological assays is inferior to that of the larger allergens. If so, this might explain the results shown in Figure 3, which suggest that sIgE to Ara h1Pro is more prevalent in peanut-tolerant children. Because of the high sensitivity of the bioassays, it is important to be aware of the effects of contamination by proteins of similar size in allergens preparations purified from natural source materials. Using recombinantly produced allergens is often crucial.

ACKNOWLEDGEMENTS

We thank Alexander CY Foo for the preparation of the batch of rAra h1Pro used for these studies. We thank Serge Versteeg for performing the confirmatory rAra h1 ImmunoCAP tests, Ronald van Ree, Laurian Jongejan, Jaap Akkerdaas and Serge Versteeg from the Academic Medical Center Amsterdam, Gabriele Gadermaier from the University of Salzburg and Stef Koppelman (University of Nebraska) for stimulating discussions.

CONFLICT OF INTEREST

RA, GM, ND, JA, LE, TR and PB have no conflict of interest related to the presented work. JL is an employee of Thermo Fisher Scientific. AP is an employee of Indoor Biotechnologies Inc

AUTHOR CONTRIBUTIONS

RA, PB, TR, GM, AP and JL contributed to the underlying concept and study design. ND, LE, PB, TR, GM and RA performed the experiments and data analyses. JA was responsible for the clinical work. RA wrote the initial draft of the manuscript. All coauthors approved the final version of the manuscript.

ETHICAL APPROVAL AND INFORMED CONSENT

The study was approved by the local medical ethics review boards (METC, UMC Utrecht; project number 05/084) and informed consent was obtained for all subjects.

ORCID

Rob C. Aalberse <https://orcid.org/0000-0002-3902-6596>

Geoffrey A. Mueller <https://orcid.org/0000-0001-8361-5323>

Lori L. Edwards <https://orcid.org/0000-0002-4930-8191>

Anna Pomés <https://orcid.org/0000-0002-8729-1829>

Jonas Lidholm <https://orcid.org/0000-0002-7779-3411>

Theo Rispens <https://orcid.org/0000-0001-9600-1312>

Peter Briza <https://orcid.org/0000-0002-4941-6782>

REFERENCES

1. Aalberse JA, Meijer Y, Derksen N, van der Palen-Merkus T, Knol E, Aalberse RC. Moving from peanut extract to peanut components: towards validation of component-resolved IgE tests. *Allergy*. 2013;68:748-756.
2. Pastorello EA, Farioli L, Stafylaraki C, et al. Anti-rPru p 3 IgE levels are inversely related to the age at onset of peach-induced severe symptoms reported by peach-allergic adults. *Int Arch Allergy Immunol*. 2013;162:45-49.
3. Petersen A, Kull S, Rennert S, et al. Peanut defensins: novel allergens isolated from lipophilic peanut extract. *J Allergy Clin Immunol*. 2015;136:1295-1301.
4. Downs ML, Semic-Jusufagic A, Simpson A, et al. Characterization of low molecular weight allergens from English walnut (*Juglans regia*). *J Agric Food Chem*. 2014;62:11767-11775.
5. Nolde SB, Vassilevski AA, Rogozhin EA, et al. Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). *J Biol Chem*. 2011;286:25145-25153.
6. Zhang J, Payne CD, Pouvreau B, et al. An ancient peptide family buried within vicilin precursors. *ACS Chem Biol*. 2019;14:979-993.
7. Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA. 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.
8. Flinterman AE, Knol EF, Lencer DA, et al. Peanut epitopes for IgE and IgG4 in peanut-sensitized children in relation to severity of peanut allergy. *J Allergy Clin Immunol*. 2008;121:737-743.
9. Ye XY, Ng TB. Hypogin, a novel antifungal peptide from peanuts with sequence similarity to peanut allergen. *J Pept Res*. 2001;57:330-336.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Aalberse RC, Mueller GA, Derksen NIL, et al. Identification of the amino-terminal fragment of Ara h 1 as a major target of the IgE-binding activity in the basic peanut protein fraction. *Clin Exp Allergy*. 2020;50:401-405. <https://doi.org/10.1111/cea.13554>

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 PhD¹, Geoffrey A. Mueller PhD.², Ninotska I.L. Derksen MSc¹, Joost A. Aalberse MD, PhD³, Lori L. Edwards BSc², Anna Pomés PhD⁴, Jonas. Lidholm⁵, PhD, Theo Rispens PhD¹, Peter Briza PhD⁶

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

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

³Huisartsenpraktijk Postjesweg Amsterdam, The Netherlands

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

⁵Thermo Fisher Scientific, Uppsala, Sweden

⁶Department of Biosciences, University of Salzburg, Salzburg, Austria

Abbreviations

GST: Glutathione S-transferase

MES: 2-(N-morpholino)ethanesulfonic acid

MS: Mass Spectrometry

RP: Reversed Phase (chromatography)

SEC: Size exclusion chromatography

SP: SulphoPropyl (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° 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° 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:

KSSPYQKKTENPCAQRCLQSCQEPDDLKQKACESRCKLEYDPRCVYDPRGHTGTTN.

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 Ara h 1Pro

rAra h 1Pro 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-Ara h 1Pro was eluted with PBS and 10 mM reduced glutathione. TEV protease was added and incubated at room temperature for 4 hours, or overnight at 4°C. The resulting sequence was:

GKSSPYQKKTENPCAQRCLQSCQQEPDDLKQKACESRCKLEYDPRCVYDPRGHTGTTN.

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 Ara h 1Pro 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 $-10\lg P > 25$ and proteins with a score $-10\lg P > 100$ 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

Electronic Supplement

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 $-10\lg P > 15$ 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 occurring 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.

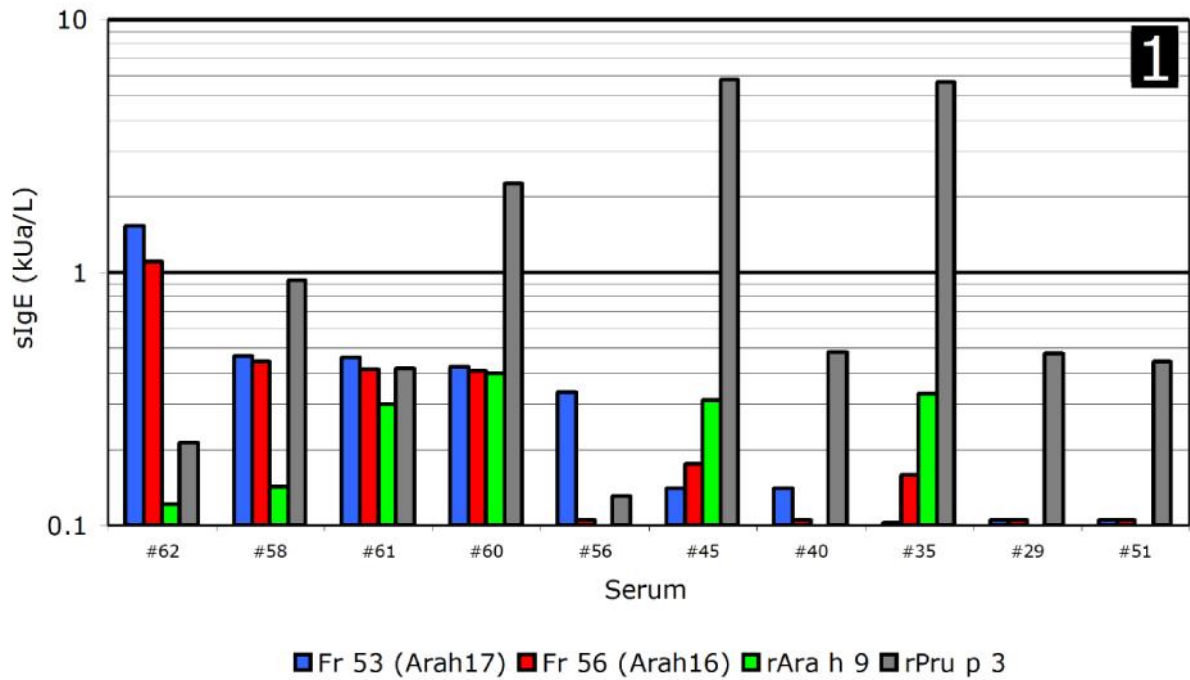
LTP is not a major contributor to IgE binding to BPP

Figure S1. 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.

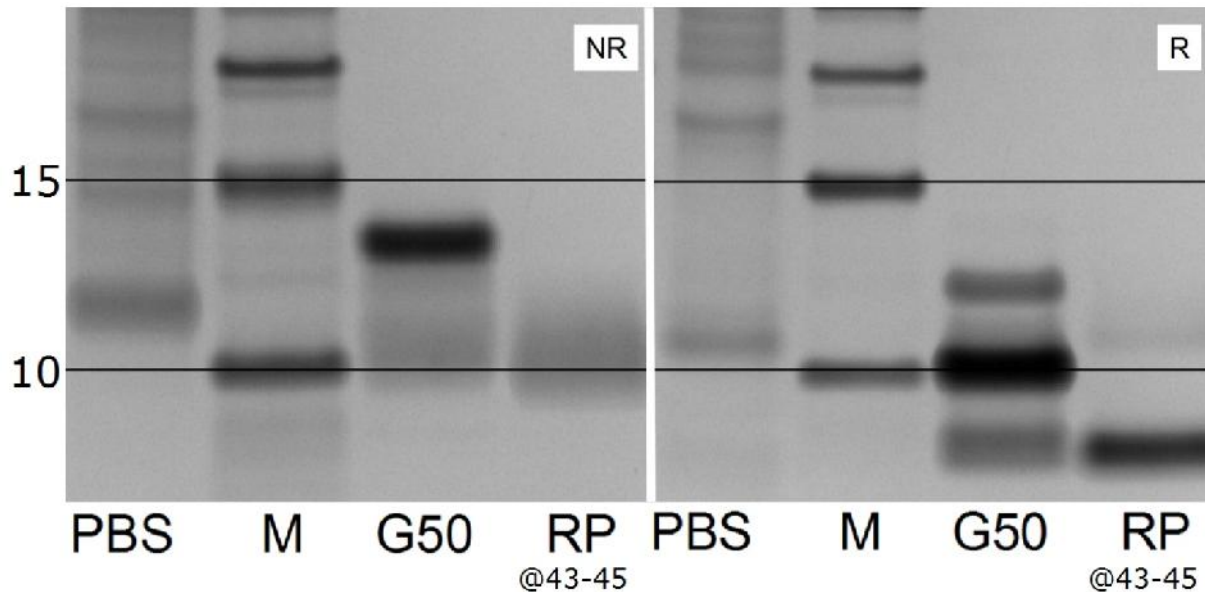


Figure S2. Anomalous effects of reduction on the electrophoretic mobility in SDS PAGE. In addition to the molecular mass marker M three peanut preparations are shown: a PBS peanut extract (**PBS**), a Sephadex G50 <15 kD fraction (**G50**) and the Reversed Phase sub-fraction pool 43-45 enriched in Ara h 1pro (**RP @43-45**). At left **NR**=Non-Reduced, at right **R**=Reduced.

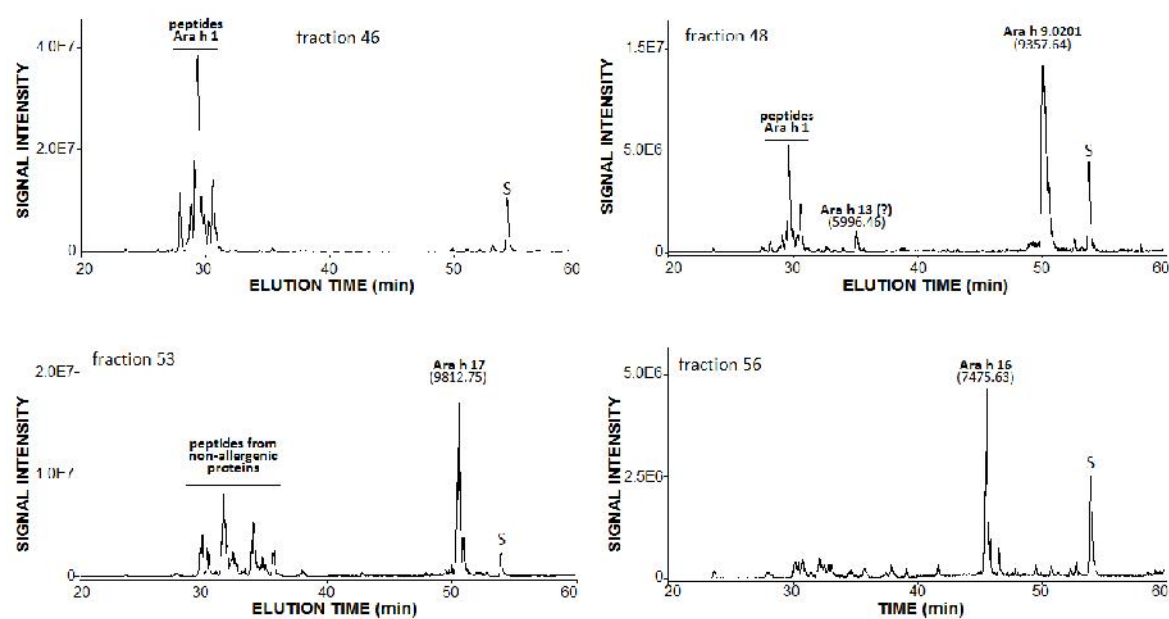


Figure S3. LC-MS/MS analysis of preparative RP chromatography fractions (see figure S6). The peak indicated by "S" is a system peak.

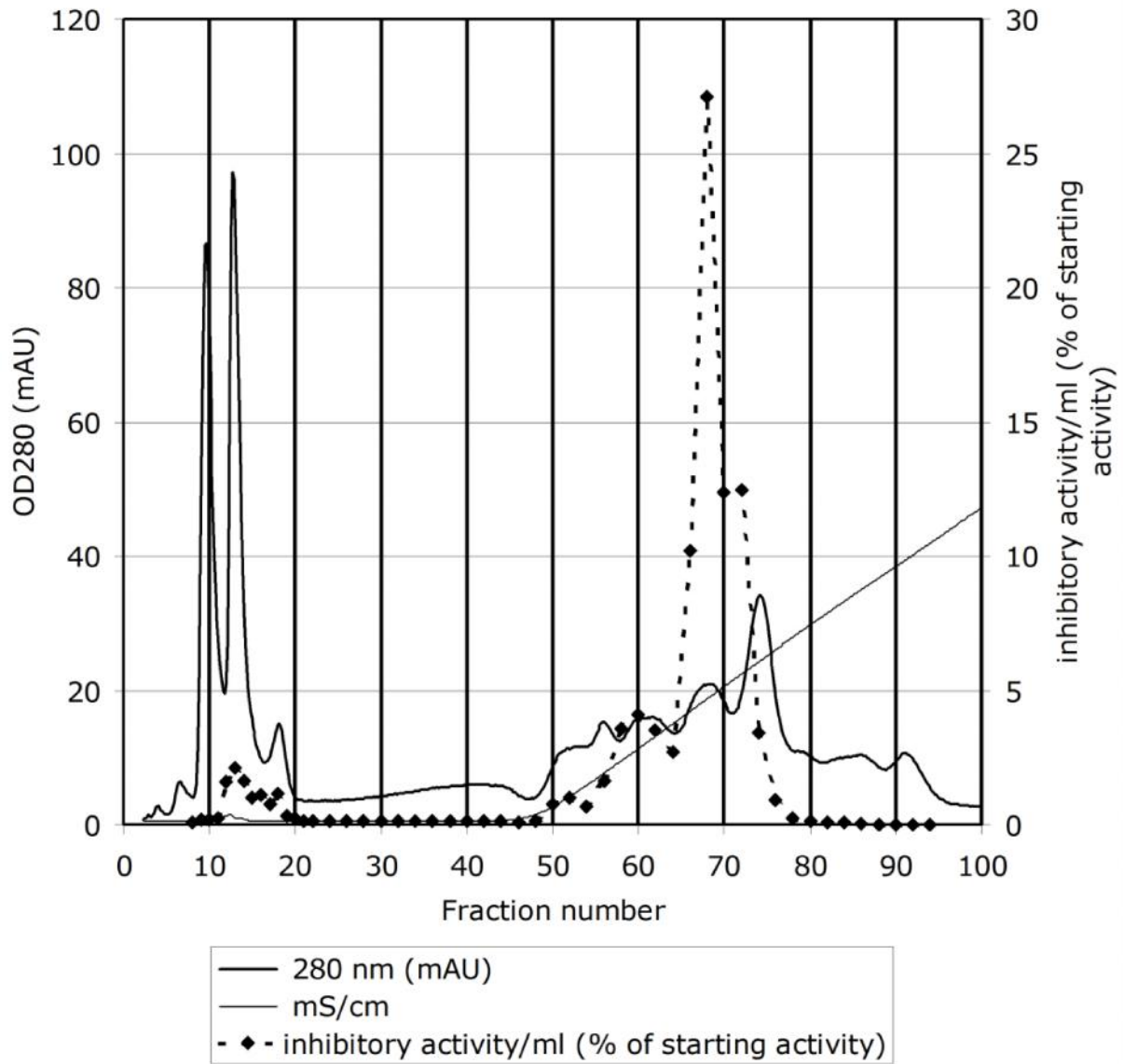


Figure S4. Elution profile from an analytical sulphopropyl cation exchange column of the 3.5 - 7 kDa fraction of BPP prepared in the presence of a protease inhibitor cocktail. Most of the allergenic activity eluted in fractions 66-72, at a conductivity of 19 mS/cm, which contained predominantly peptides with calculated isoelectric points of 7.7 - 8.3.

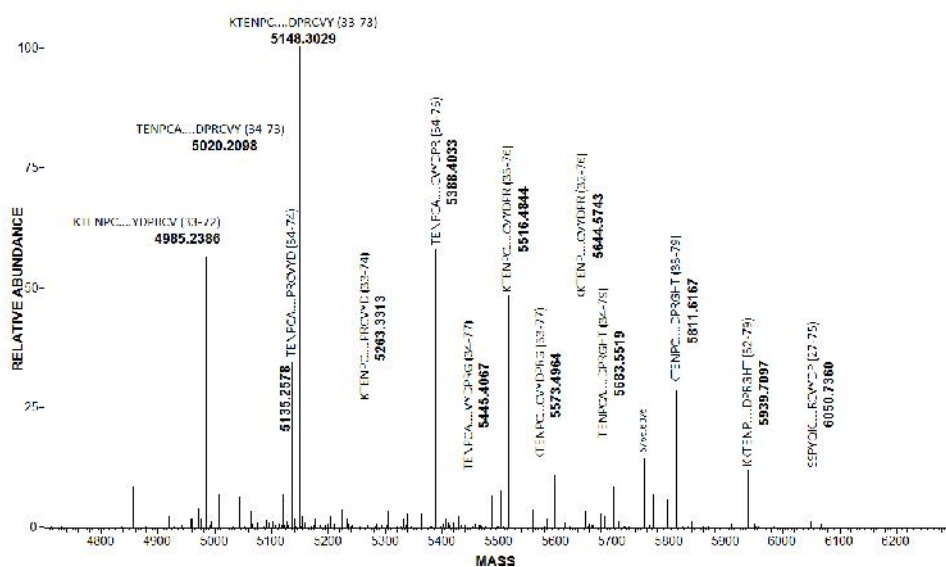


Figure S5. Spectrum of Ara h 1-peptides in RP-fraction 46 (see figure S3). Raw data of the Ara h 1-containing fractions (27 to 32 minutes elution time) were merged and deconvoluted to obtain uncharged signals. Peptide identification was based on both mass matching and MS/MS sequencing.

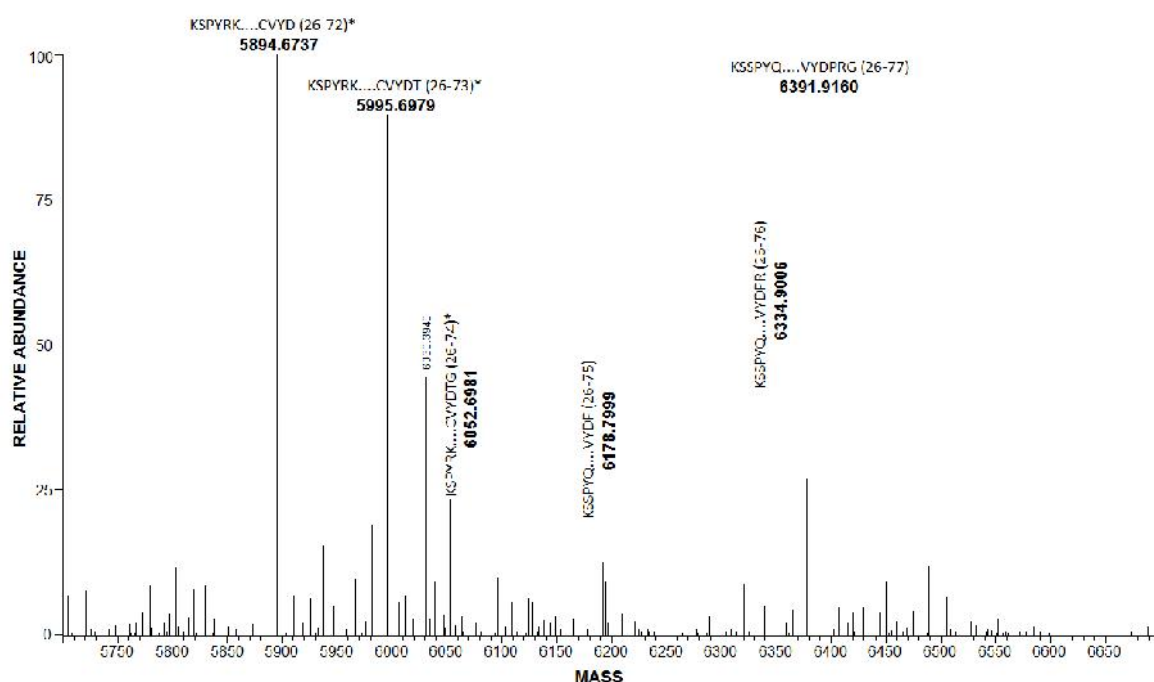


Figure S6. As for Figure S5, but now for SP-fraction 70 (see figure S4).

Electronic Supplement

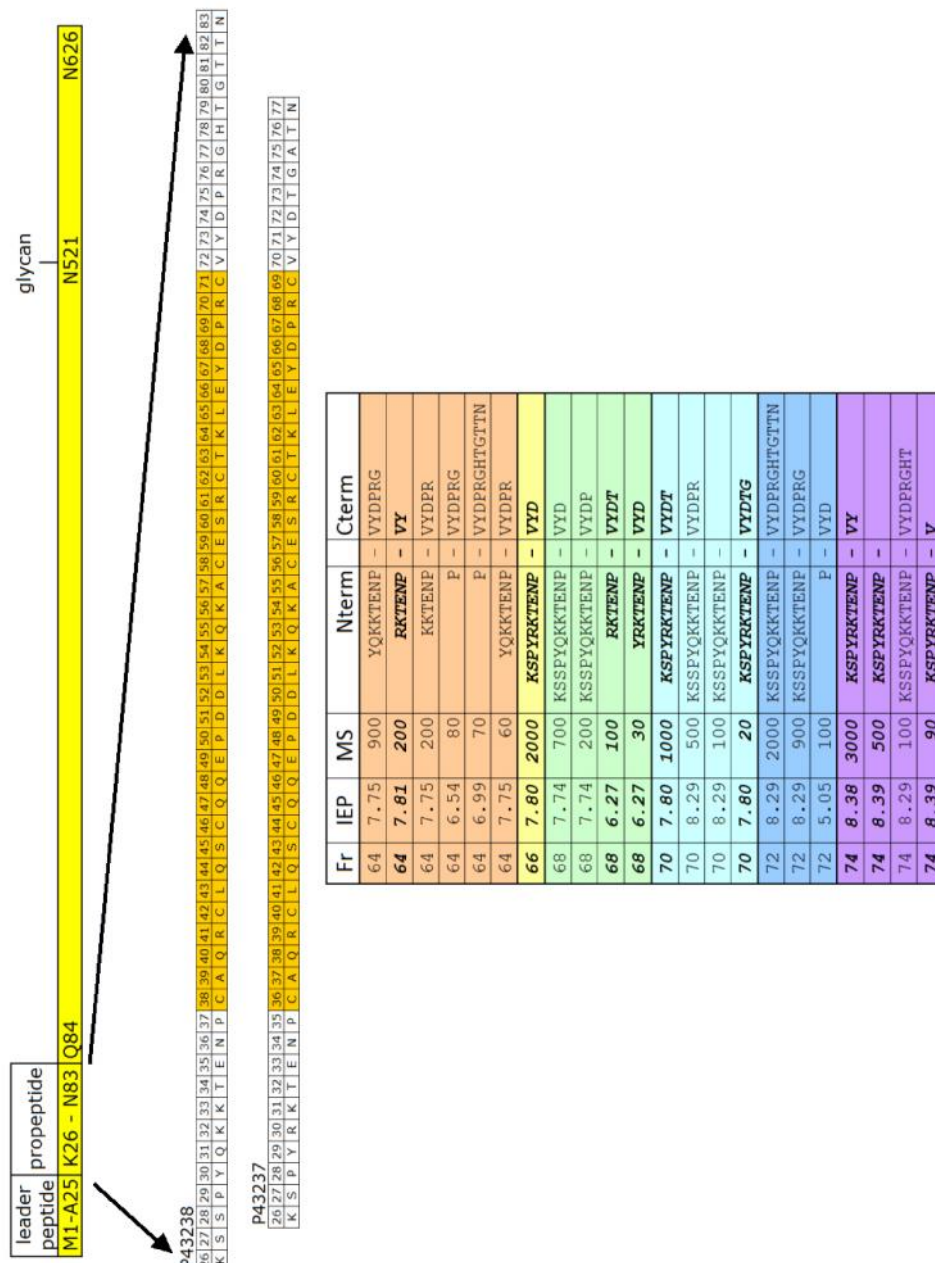


Figure S7. Mass-spectrometric analysis of Ara1Pro-derived peptides in the fractions 64, 66, 68, 70, 72 and 74 of the sulphopropyl column of the <9 kDa fraction of BPP prepared in the presence of a protease inhibitor cocktail (see figure S5). First column Fr= fraction number; 2nd column: IEP = calculated isoelectric point of the peptide; MS=relative signal strength in the mass spectrum; N-term= sequence preceeding 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.

Reference

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.