

High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation

Maxime Gavage^{a,b}, Kaatje Van Vlierberghe^c, Christof Van Poucke^c, Marc De Loose^c, Kris Gevaert^{d,e}, Marc Dieu^{b,f}, Patsy Renard^{b,f}, Thierry Arnould^b, Nathalie Gillard^{a,*}

^a CER Groupe, Rue du Point du Jour 8, 6900 Marloie, Belgium

^b Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium

^c Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Brusselsesteenweg 370, 9090 Melle, Belgium

^d VIB Center for Medical Biotechnology, Albert Baertsoenkaai 3, B9000 Ghent, Belgium

^e Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, B9000 Ghent, Belgium

^f MaSUN, Mass Spectrometry Facility, University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium

ARTICLE INFO

Keywords:

High-resolution mass spectrometry
Peptide biomarker selection
Processed food products
Multiple allergen isoforms
Peanut origin

ABSTRACT

To protect allergic patients and guarantee correct food labeling, robust, specific and sensitive detection methods are urgently needed. Mass spectrometry (MS)-based methods could overcome the limitations of current detection techniques. The first step in the development of an MS-based method is the identification of biomarkers, which are, in the case of food allergens, peptides. Here, we implemented a strategy to identify the most salient peptide biomarkers in peanuts. Processed peanut matrices were prepared and analyzed using an untargeted approach via high-resolution MS. More than 300 identified peptides were further filtered using selection criteria to strengthen the analytical performance of a future, routine quantitative method. The resulting 16 peptides are robust to food processing, specific to peanuts, and satisfy sequence-based criteria. The aspect of multiple protein isoforms is also considered in the selection tree, an aspect that is essential for a quantitative method's robustness but seldom, if ever, considered.

1. Introduction

Food allergies are a global health problem. Several recent studies have indicated that the prevalence of food allergies has increased in the last few decades, and they primarily affect populations in industrialized countries. Children seem to be affected more often than adults as it is estimated that immunoglobulin-E (IgE)-associated food allergies affect 3–8% of children and 1–3% of adults (De Silva et al., 2014; Longo, Berti, Burks, Krauss, & Barbi, 2013; Sicherer & Sampson, 2014). Peanuts (*Arachis hypogaea*) are among the most common foods that induce severe allergic reactions and the most frequent cause of food-induced anaphylaxis (Husain & Schwartz, 2012). Sixteen protein allergens have been identified in peanuts and termed Ara h 1–17 according to the systematic allergen nomenclature approved by the World Health Organization and the International Union of Immunological Societies' Allergen Nomenclature Sub-committee; Ara h 3 and Ara h 4 were *a posteriori* considered to be the same allergen and are both referred to as Ara h 3 (Pomés et al., 2018). The most common peanut allergens

recognized by the serum IgE of more than 50% of allergic patients are Ara h 1–3 and Ara h 6. These four allergens are also the most abundant proteins in peanuts and are all seed storage proteins (Schmidt et al., 2009). Some of these 16 Ara h allergens do not correspond to unique protein sequences but include several protein isoforms. These multiple isoforms are expressed from different genes or are the result of different types of post-translational proteolytic processing of the same protein (Palladino, 2018).

Peanut species are classified into two subspecies, which are further classified into six botanical varieties based on morphology and growth habits (Moretzsohn et al., 2004). However, they are more commonly classified into four market types, including Runner, Spanish, Virginia and Valencia. The relative abundance of peanut allergens has been shown to vary depending on the growing conditions (Walczyk et al., 2013), as well as the specific cultivar (Wu et al., 2016). One study quantified the major allergen content of peanuts from different market types (Koppelman et al., 2016). In this report, the proportion of different Ara h allergens in peanuts was measured, and it was found that

* Corresponding author.

E-mail address: n.gillard@cergroupe.be (N. Gillard).

<https://doi.org/10.1016/j.foodchem.2019.125428>

Received 3 June 2019; Received in revised form 14 August 2019; Accepted 24 August 2019

Available online 26 August 2019

0308-8146/ © 2019 Elsevier Ltd. All rights reserved.

Ara h 1, 2, 3 and Ara h 6 account for $17.1 \pm 3.4\%$, $6.2 \pm 1.3\%$, $70.6 \pm 8.6\%$, and $5.8 \pm 1.8\%$ of the total of extracted protein, respectively. The extraction yield of the 20 peanut samples analyzed in that study was $74.7 \pm 8.5\%$.

Strict peanut avoidance remains the best option for allergic patients, despite progress being made in desensitization therapy (Bégin et al., 2014). This avoidance strategy requires rigorous and accurate food labeling and, consequently, sensitive and reliable detection methods to exclude/identify unintentional contamination. Commercially available detection methods are mainly based on immunological methods targeting specific proteins, such as enzyme-linked immunosorbent assays (ELISA). Other methods rely on polymerase chain reactions (PCRs) targeting allergen-specific DNA molecules. Despite their high sensitivity, both methods have their limitations. The antibodies used in immunological methods recognize epitopes. One drawback is the possible recognition by the antibodies of similar structures in non-allergenic ingredients, thus generating false positive results. In contrast, protein denaturation upon food processing can modify the conformation of the targeted proteins or the targeted epitopes, leading to false negative results (Van Hengel, 2007). Due to technical advances, MS-based methods have become the methods of choice for allergen detection (Monaci, Pilolli, De Angelis, & Mamone, 2015; Popping & Godefroy, 2011). Multi-allergen detection and quantification methods, including those for peanuts, were recently proposed (Boo, Parker, & Jackson, 2018; Henrottin et al., 2019; Planque et al., 2019). Studies on allergen detection by MS-based methods report detection limits in the same range (0.1–5 mg of allergenic ingredient protein per kg of matrix) as ELISAs (Croote & Quake, 2016). Lastly, the high specificity of MS-based methods is ensured by different technical elements including BLAST query of the selected peptides to public databases, chromatographic retention time or the analysis of multiple transitions.

The first step in the development of an MS-based method is the identification and further selection of possible peptides and, thus, proteins specific to the studied species (i.e., biomarkers). This selection is particularly important when developing a quantitative method since quantifying the allergenic ingredient will be directly linked to quantifying the selected biomarkers. Therefore, the method's overall accuracy, robustness, and precision rely on this selection. In a previous study (Gavage et al., 2018), we developed a strategy for identifying potential peptide biomarkers in processed food products. Our strategy, based on an untargeted approach using high-resolution mass spectrometry (HRMS), was applied to processed egg-containing matrices and allowed us to identify 16 potential egg peptide biomarkers.

The aim of the present study was to apply the same approach to identifying potential biomarkers of peanut peptides. Modifications needed to be implemented to adjust our approach to address peanuts and consider relevant food matrices that best represent processed foods that potentially contain peanuts. To include potential peanut origin variability, peanuts from two distinct geographical regions corresponding to two different market types were analyzed. We identified more than 300 peptides, which were further filtered using a series of selection criteria to ensure the adequate performance of a future, routine ultra-high-performance liquid chromatography (UHPLC)-MS/MS analytical method. These criteria include sequence-based features, as well as robustness to food processing, and address the issue of multiple protein isoforms. The latter is crucial in the development of a robust quantitative method but is seldom, if ever, considered in published studies. Eventually, 16 peanut peptide biomarkers were retained, covering all protein isoforms of Ara h 1 and Ara h 3, the two most abundant peanut allergens.

2. Material and methods

2.1. Reagent and materials

Urea, tris(hydroxymethyl)aminomethane (Tris-HCl), tris(2-

carboxyethyl)phosphine hydrochloride (TCEP), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB), and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), while Sep-Pak C18 1 cc Vac solid-phase extraction (SPE) cartridges were purchased from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with MS-grade Trypsin Gold from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase (Lys-C) was obtained from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Water, acetonitrile, methanol absolute, 2-propanol (ULC/MS-grade for all solvents), and formic acid were purchased from Biosolve (Valkenswaard, the Netherlands). ESI-L Low Concentration Tuning Mix was purchased from Agilent Technologies (Santa Clara, California, USA).

2.2. Food product preparation

Peanuts from two geographical regions and corresponding to two different market types were purchased from two providers. The first type of peanuts, provided by Quality Nuts BVBA (Zandhoven, Belgium), was cultivated in China and corresponds to the Spanish market type. The second type, provided by Versenoten.nl (Waddinxveen, The Netherlands), was cultivated in Israel and corresponds to the Virginia market type. Peanuts were purchased in an unprocessed, in-shell form and then shelled at the Food Pilot unit of the Flanders Research Institute for Agriculture, Fisheries, and Food (ILVO), producing five peanut matrices. Each peanut matrix was produced in duplicate, one for each peanut market type.

2.2.1. Preparation of a raw peanut matrix

Raw, whole peanuts were ground using a Kenwood AT286 Glass mini chopper/mill attachment KW714229 (Kenwood, New Hampshire, UK) for 20 s on pulsing mode using the maximum speed, vacuum-packed (vacuum packaging machine KN5, vc999, Missouri, USA), and stored at 4 °C in the dark.

2.2.2. Preparation of a heated peanut matrix

Raw, whole peanuts were roasted in batches of 500 g in an oven (baking oven condo CO 6.0608, Miwe, Germany) at 180 °C for 18 min. The peanuts were spread in a single layer on a baking plate and stirred with a spoon every 5 min to ensure homogenous roasting. The roasted peanuts were allowed to cool down at room temperature, ground into a peanut powder using ThermoResist Glazen Blender AT358 (Kenwood, New Hampshire, UK), vacuum-packed, and stored at 4 °C.

2.2.3. Preparation of a low-pH peanut matrix: fermented peanut milk

This matrix was produced by fermenting peanut milk. Peanut milk was produced by soaking the peanuts in reverse osmosis water overnight, followed by blending (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3500 TPM, Robot Coupe, Vincennes Cedex, France) for 8 min at the maximum speed, resulting in a milky substance. The peanut milk was subsequently sieved (Retsch Test Sieve, 250 µm pore size, Retsch, Aartselaar, Belgium) and then transferred to a sterile Erlenmeyer flask of 1 L, and a bacterial culture (nu-trish® BY-Mild, which contains *Bifidobacterium* species, *Lactobacillus delbrueckii* subsp. *Bulgaricus*, and *Streptococcus thermophilus*, CHR Hansen, Hørsholm, Denmark) was added (4 g of frozen culture for 600 mL of peanut milk). The flasks were aseptically sealed and incubated in a warm water bath (42 °C). Every 30 min, a sub-sample of the yogurt was checked for pH until a pH of 4.2 was reached. The yogurt was next divided into 20 g portions and stored at 4 °C in the dark. Given the lack of data concerning the stability of the produced yogurt, the analyses were performed within a month following the matrix production.

2.2.4. Preparation of a matrix subjected to the maillard reactions: Caramelized peanuts

An amount corresponding to 250 g of raw, whole peanuts was chopped (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3500 TPM, Robot Coupe, Vincennes Cedex, France) 3 times for 5 s duration at the maximum speed, and 200 g lactose (α -Lactose monohydrate, Sigma-Aldrich, Bornem, Belgium) was mixed with 100 mL reverse osmosis water to a smooth paste. The chopped peanuts were subsequently stirred into the lactose paste until all the peanut material was covered with the lactose paste. This peanut-lactose mixture was then spread over an oven plate and placed in a pre-heated (160 °C) convection oven for 15 min. Now, every 5 min, the mixture was stirred on the plate with a spoon to obtain homogeneous caramelization. The plate was left at room temperature to cool down for 16 h and, subsequently, ground into a fine powder (Kenwood Mini Chopper Mill Attachment AT320B). The caramelized peanut powder was divided into portions of 50 g, vacuum-packed (vacuum packaging machine KN5, vc999, Missouri, USA), and stored at 4 °C in the dark until use.

2.2.5. Preparation of a complex and fatty-rich peanut matrix: Chocolate incurred with peanuts

A peanut powder with particle sizes < 500 μ m in diameter was produced by grinding raw peanuts (Kenwood Mini Chopper Mill Attachment AT320B, pulsing speed, 3 min, Kenwood, New Hampshire, UK), resulting in a slushy substance. This peanut mash was then defatted by pressing the material between extra-thick cotton cloths, resulting in dry peanut powder. The peanut mash and powder were weighed before and after defatting, and the loss in mass was documented to determine the amount of weight lost during defatting. The fine, dry peanut powder was then sieved with a Retsch test sieve (pore size 500 μ m). Then, chocolate was prepared. First, a 'conche' was used to mix and smooth chocolate refiner flakes (90%) and cacao butter (10%). The refiner flakes (Callebaut, Lebbeke, Belgium) and cacao butter (Callebaut) were then warmed in a water bath in separate bowls until a temperature of 40–45 °C was reached. Next, the melted butter was slowly added to the warmed refiner flakes, and the mix was stirred for 3 min, followed by adding 1.2% ammonium phosphatide. This mixture was stirred for an additional 3 min. Finally, the chocolate incurred with peanuts (38,178 ppm [mg peanut proteins/kg chocolate] for Spanish and 36,792 ppm for Virginia market type peanuts) was produced by adding 70 g of defatted peanut powder to the liquid chocolate and stirring for 3 min while maintaining the temperature of the chocolate at 42 °C. The liquid chocolate was subsequently poured into sterile chocolate molds, covered with plastic foil, and left to cool and solidify overnight at 4 °C. The following day, the chocolate bars were rasped into chocolate flakes, which were then separated into aliquots of 50 g and stored at 4 °C in the dark. A proper control/blank chocolate material was prepared without peanut powder.

2.3. Sample preparation, mass spectrometry analysis and data treatment

The protocol used to prepare the samples is comparable to the method described in our previous study dedicated to eggs (Gavage et al., 2018). Briefly, proteins were extracted from food matrices, purified, and proteolytically digested. Two digestion protocols were compared for each processed food product. In the first protocol, the proteins were digested with trypsin alone. In the second protocol, tandem endoproteinase Lys-C/trypsin digestion was used. For each combination of digestion protocol and processed food product, three biological replicates were prepared and analyzed via HPLC-HRMS.

Mass spectrometry analysis and obtained data treatment were also performed as described in our previous article. For each sample, 3 μ g of peptides were analyzed using an ESI-MS/MS μ Axis Impact UHR-TOF (Bruker, Billerica, Massachusetts, USA) coupled to a Dionex UltiMate 3000 Standard LC Systems (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Peptides were separated by reverse-phase liquid

chromatography on an Acclaim PepMap100 C18 column (3 μ m, 100 Å, 1 mm $\times$ 15 cm nanoViper, Thermo Fisher Scientific, Waltham, Massachusetts, USA) with a 60 min gradient elution with a mixture of water/acetonitrile, each containing 0.1% formic acid. The data were processed using Mascot 2.5 (Matrix Science, London, United Kingdom) as the search engine and a homemade protein database, adapted to peanut samples, and downloaded from UniProt (uniprot.org). This database contains the *Arachis* genus proteome (which includes peanuts [*Arachis hypogaea*]) and some possible contaminants (i.e., cacao, bacteria from Nu-trish® BY-Mild [*Bifidobacterium animalis* subsp. *lactis*, *Lactobacillus delbrueckii* subsp. *Bulgarius*, and *Streptococcus thermophilus*] used for yogurt preparation, human keratin, trypsin, and Lys-C) to avoid misidentification. In total, this database contains 78,112 entries (1846 entries for the *Arachis* genus [1492 for peanut *Arachis hypogaea*], 34,951 for bacteria from Nu-trish® BY-Mild, 40,956 for cacao, and 350 for human keratin). Peptide and protein identifications were validated using Scaffold 4.8 (Proteome Software, Portland, Oregon, USA). Combination of the list of validated peptides and buckets generated by ProfileAnalysis (Bruker, Billerica, Massachusetts, USA) software allowed the generation of comparable results. Findings discussed below are based on signal intensities linked to the identified peptides and were obtained using this method.

2.4. Multiple sequence alignments and principal component analysis

Multiple sequence alignments were performed on the UniProt website (uniprot.org/align/) using Clustal Omega software (Sievers et al., 2011). Obtained alignments were transferred to Jalview software (Waterhouse, Procter, Martin, Clamp, & Barton, 2009) for analysis. Jalview was notably used to perform principal component analysis (PCA) on protein sequence alignments by selecting the BLOSUM62 scoring matrix.

3. Results

3.1. General strategy

Food allergies are an increasing problem not only for allergic patients and their social networks but also for the food industry and regulators. Undeclared allergens are, indeed, the primary cause of food product recalls and alerts in Western countries (Bucchini, Guzzon, Poms, & Senyuva, 2016). Therefore, an urgent need exists for reference methods to detect and quantify allergens in food products (Walker, Burns, Elliott, Gowland, & Mills, 2016).

The objective of this study was to identify potential peanut peptide biomarkers using an HRMS-based discovery proteomics approach. More than 300 peanut peptides were identified in the analyzed samples (Table 1), and a complete list of these peptides and their intensity in each peanut matrix can be found in Supplementary Material 1. The potential peanut peptide biomarkers were then used to develop a UHPLC-MS/MS method for quantifying peanuts in processed food products in routine settings.

Samples were prepared according to the procedure developed in a previous study for the identification of peptide biomarkers in eggs (Gavage et al., 2018). In that study, some egg proteins, such as ovalbumin, were found to be resistant to proteolytic digestion using a trypsin-only protocol. Therefore, we implemented a digestion protocol combining two enzymes, endoproteinase Lys-C and trypsin, which improved the digestion of resistant proteins as evident from the significantly higher MS signal intensities of peptides originating from such proteins. The two digestion protocols (i.e., trypsin-only and tandem Lys-C/trypsin) were applied in parallel to all peanut samples. Unlike egg proteins, no proteins appeared to be resistant to proteolytic digestion using only trypsin as only minor differences were observed on the peptide signal intensities when peanut samples digested with trypsin alone were compared to those digested with tandem Lys-C/trypsin. The

Table 1

Unique proteins and number of unique peptides for these proteins identified following HRMS analysis of peanut samples (raw peanuts, heated peanuts, low pH matrix, caramelized peanuts and fatty rich environment).

Identified protein	UniProt entry identifier(s)	Number of identified unique peptides	Coverage
Ara h 1	B3IXL2, E5G076, N1NG13, P43237, P43238, Q6PSU3	67	61.4%
Ara h 2	Q6PSU2	12	41.9%
Ara h 3	A1DZF0, B5TYU1, E5G077, Q0GM57, Q5I6T2, Q6T2T4, Q647H2, Q647H3, Q647H4, Q8LKN1, Q9FZ11, Q9SQH7	127	83.2%
Ara h 6	A1DZE9	5	24.8%
Ara h 8	B0YIU5, B1PYZ4, Q6VT83	6	31.4%
Ara h 10	Q647G5	3	23.1%
Ara h 13	C0HJZ1	4	36.1%
11S arachin	E9LFE8	7	30.8%
11S seed storage globulin B1	A1E2B0	6	25.9%
Annexin	A0A0F6VX63, F6KLJ6	7	23.2%
Bowman-Birk trypsin inhibitor	P01066, Q0PKR5, Q7X973	2	40%
Conarachin	Q647H1	10	33.9%
Galactose-binding lectin	A0A089ZXL7, Q38711	14	55%
Glyceraldehyde-3-phosphate dehydrogenase C2	A0A0A6ZDP1, A0A0A6ZDT0	4	38.2%
Late embryogenesis abundant protein 3	A0A1L5JJI4	2	7.0%
Late embryogenesis abundant protein group 5 protein	E5FHZ2	5	21.1%
Lipoxygenase	Q4JME6, Q4JME7	24	27.2%
Ribulose biphosphate carboxylase large chain	A0A191UJ50, O20356	4	11.8%
Seed maturation protein	N1NKG9	2	23.4%
Steroleosin-A	A7LB60	6	16.6%
Steroleosin-B	A7LB59	8	18.1%

results discussed below in further detail correspond to samples digested with trypsin alone; however, both digestion protocols' results are included in [Supplementary Material 1](#).

To ensure the robustness, specificity, sensitivity and trueness of a future quantitative method, selected peptide biomarkers have to fulfill several criteria. The peptides have to be specific to peanuts, robust to food processing, derived from abundant proteins, size-compatible with *m/z* analyzers (i.e., 8–25 amino acids), not prone to missed cleavages by the protease(s) used or to amino acid modifications, and present at comparable levels regardless of the peanut market type. Thus, to identify ideal peanut peptide biomarkers, the detected peptides were further filtered using these criteria as detailed in the following sections.

3.2. Peanut protein composition

Peanuts are seeds and, therefore, the vast majority of proteins recovered/extracted from them are seed storage proteins. These proteins serve as a reserve of amino acids for the growing embryo upon germination. The most relevant peanut allergens are storage proteins, including the four most abundant: Ara h 1 ($17.1 \pm 3.4\%$ of total extracted protein), Ara h 2 ($6.2 \pm 1.3\%$ of total extracted protein), Ara h 3 ($70.6 \pm 8.6\%$ of total extracted protein) and Ara h 6 ($5.8 \pm 1.8\%$ of total extracted protein) with an extraction yield of $74.7 \pm 8.5\%$ for the 20 analyzed peanut samples ([Koppelman et al., 2016](#)). As the ultimate aim of this study was to develop a sensitive, quantitative UHPLC-MS/MS-based detection method, the method, logically, must target these abundant proteins. This aspect was demonstrated here because, in the 60 analyses performed (i.e., 5 peanut food products, 2 market types, 2 digestion protocols, and 3 biological replicates), over 95% of the total MS signal intensities measured are associated with identified peanut peptides belonging to these 4 proteins. In fact, Ara h 1 peptides accounted for 21% of the signal, while Ara h 2, Ara h 3, and Ara h 6 peptides accounted for 1%, 74%, and 0.5%, respectively. Given these results, only peptides from Ara h 1 and Ara h 3 were further considered as potential peptide biomarkers for the remainder of this study.

3.3. Peptide sequence-based criteria

With the goal of developing a quantitative UHPLC-MS/MS method,

total peanut protein will be quantified through peanut peptides. To ensure method trueness, the quantification of these peptides should directly correlate with the protein amount. The ideal case would be a stoichiometric ratio between a given peptide and its corresponding protein, for example, x mole(s) of a given peptide corresponding to x mole(s) of its corresponding protein. To achieve this goal, peptides carrying amino acids that can potentially be modified or are subject to incomplete enzymatic digestion should be discarded ([Gavage et al., 2018](#)).

Potential amino acid modifications were first evaluated using a literature review ([Reimer et al., 2011](#)) and then empirically confirmed using Mascot's 'error-tolerant search'. The observed modifications are cysteine carbamidomethylation, methionine oxidation, *N*-terminal glutamine to pyroglutamate conversion and, asparagine and glutamine deamidation. Cysteine carbamidomethylation, inevitably, results from the reduction and alkylation step during sample preparation using TCEP and IAA. The other modifications are spontaneous and uncontrolled reactions. Therefore, peptides containing cysteine, methionine, or *N*-terminal glutamine were systematically discarded. Deamidation was treated differently, given the high occurrence of asparagines and glutamines ([Andersen, Hill, Gorbet, & Brodbeck, 1998](#)). Although both amino acids are prone to deamidation, the rate of glutamine deamidation was reported to be much lower than that of asparagine ([Robinson et al., 2004](#)). Glutamine deamidation was, therefore, not further considered. Moreover, no experimental evidence of glutamine deamidation was observed in our data. Deamidation is influenced by the nature of the amino acid at the carboxyl side of asparagine (Asn). The fastest reaction occurs when this amino acid is glycine (Gly), whereas for other amino acids, the reaction rates are at least 10-times lower ([Kosky, Dharmavaram, Ratnaswamy, & Manning, 2009](#); [Robinson & Robinson, 2004](#)). Thus, only peptides containing Asn-Gly were discarded. Finally, post-translational modifications documented in databases, such as UniProt, were also considered. Glycosylation on Asn⁵²¹ is, for example, documented for Ara h 1. We decided to reject peptides holding this glycosylation site to prevent that incomplete post-translational modification would negatively influence further results.

The types of amino acids surrounding the proteolytic recognition site are known to affect the efficiency of peptide bond hydrolysis ([Siepen, Keevil, Knight, & Hubbard, 2007](#)). Sequences that have a

negative effect on this efficiency must, therefore, be avoided. Protease recognition sites are described using the Schechter and Berger nomenclature (Schechter & Berger, 1967), in which amino acids surrounding the cleavage site are indicated as P4-P3-P2-P1-P1'-P2'-P3', where the peptide bond between P1 and P1' is the cleaved bond. Arginine, lysine, and proline negatively affect digestion when present at P1'. Peptides with one of these amino acids in P1' were also discarded. The acidic amino acids aspartate and glutamate also negatively influence digestion when they are present near the cleavage site. Given the relatively high occurrence of such amino acids, systematically discarding possibly-affected peptides is too restrictive. Peptides containing aspartate or glutamate close to the cleavage site were treated on a case-by-case basis, evaluating the importance of the potentially-missed cleavage. A peptide was rejected when the MS signal intensity of the associated peptide containing the missed cleavage exceeded 10% of the MS signal of the fully tryptic peptide.

3.4. Robustness to food processing

Peanuts are consumed in many forms, such as boiled, roasted, or added to snack foods. However, food processing and matrix interactions are known to affect the recovery and subsequent detectability of targeted compounds (Khuda, Jackson, Fu, & Williams, 2015; Sayers et al., 2016). To ensure the accuracy and robustness of our quantitative UHPLC-MS/MS method, the effects of food processing were considered when selecting peptide biomarkers. Matrices and processing conditions applied to peanut samples were selected to represent the conditions of daily food products. Selected peptide biomarkers should, ideally, not be affected in these conditions.

Matrix and food processing effects were evaluated comparing the MS signal intensities of the identified peptides under the five processing conditions assessed. Unaffected peptides should have comparable MS intensities in different conditions. Results for Ara h 1 and Ara h 3 are shown in Fig. 1. Only peptides fulfilling sequence-based criteria are displayed in this figure for both peanut market types, Spanish and Virginia. Some peptides, such as GTGNLELVAVR from Ara h 1 and SQSENFYVAFK from Ara h 3, seem to be largely resistant to the food processes selected in this study (i.e., heat treatment, caramelization, a low-pH matrix, and a complex matrix). Conversely, the intensity of some peptides depends on the food processing method, with significantly lower MS intensities under some conditions. For example, the intensity of the peptide IPSGFISYILNR from Ara h 1 was significantly lower in samples corresponding to the heat treatment, low pH matrix, and complex matrix (chocolate) conditions when compared to unprocessed samples (Student's *t*-test, $p < 0.05$). Potential peptide biomarkers have, thus, been refined based on these results as the detection of the selected peptides must be robust, even when various food processing methods are used, to ensure the future quantitative UHPLC-MS/MS method is robust.

3.5. Market type variation and protein isoforms

Differences in protein content and protein relative-abundance were shown to vary among the tested peanut varieties (Kottapalli et al., 2008; Schmidt et al., 2009). Moreover, for some allergens, Ara h corresponds to multiple protein isoforms with sequence variations, and the relative abundance of the different allergen isoforms is also likely to vary depending on the variety (Jin, Guo, Chen, Howard, & Zhang, 2009). To consider this potential variability in our peptide biomarkers selection, peanuts from two market types (i.e., Spanish and Virginia) were analyzed.

Peanut detection methods identified in other studies have rarely included aspects of market type variation and protein isoforms (Johnson et al., 2016). However, this additional degree of complexity must be taken into account when developing absolute quantification methods. Indeed, proposed thresholds for food allergens are expressed

as total protein quantities. Based on these thresholds, or reference doses (e.g., those proposed by VITAL® 2.0), most (i.e., 95–99%) of the allergic population should be protected, and appropriate measures may be taken when undeclared allergens are detected. Peptide-based approaches, therefore, require a conversion factor to translate results expressed in peptide quantities into total protein amounts. This conversion factor undeniably needs to include the proportion of the protein corresponding to the targeted peptide, compared to the total ingredient protein. Variability, in the proportion of a protein of interest based on the market type, increased the quantification method's inaccuracy in this study. Ideal peptide biomarkers should, thus, be present at a constant concentration, irrespective of the peanut variety, and include all protein isoforms.

Possible variation due to the peanut's origin was also evaluated by comparing peptide MS intensities. A comparison of the peptides fulfilling the sequence-based criteria is presented in Fig. 2. No significant differences are observed for peptides from Ara h 1, but the results are more complex for Ara h 3. While some peptides, TANDLNLILR and WLGLSAEYGNLYR, have comparable intensities when the two market types are compared, other peptides are present in higher levels in one of the market types. This was the case for SQSENFYVAFK and VYDEELQEGHVLVVPQNFVAVAGK, for which peak intensities are higher in the Spanish market type, whereas the peak intensities of other peptides (e.g., FYLAGNQEQEFLR and VYDEELQEGHVLVVPQNFVAAK) are more intense in the Virginia market type.

Allergens Ara h 1 and Ara h 3 do not necessarily correspond to unique proteins but, rather, to groups of protein isoforms. The UniProt database contains six entries (i.e., B3IXL2, E5G076, N1NG13, P43237, P43238, and Q6PSU3) corresponding to Ara h 1. These entries correspond to three isoforms as some have identical protein sequences (e.g., B3IXL2/P43237/Q6PSU3 and N1NG13/P43238). Concerning Ara h 3, 12 entries corresponding to 12 isoforms with different protein sequences can be found. The protein sequences of the different isoforms of Ara h 1 and Ara h 3 are not fully identical. Some sequence-based selected peptides were found to be specific to one or several of these isoforms. These peptides' specificity for the different isoforms is shown in Table 2 for Ara h 1 and Ara h 3. Some of the sequence-based selected peptides of Ara h 1, DQSSYLQGFSSR and GTGNLELVAVR, are common to all three Ara h 1 isoforms. However, other peptides, such as EHVQELTK and SSENNEGIVIVK, are specific to one or two of these isoforms. As previously mentioned, no significant differences in peptide MS signal intensity were observed between the two peanut market types, which means the abundance of Ara h 1 is comparable in Spanish and Virginia market type peanuts, considering the samples used in this study. Moreover, the abundance of each isoform seems to be similar in both market types.

Given the higher number of isoforms, the case of Ara h 3 is more complex. None of the 32 sequence-based selected peptides are common to all Ara h 3 isoforms. Thus, to cover all these isoforms and quantify the total abundance of this allergen, multiple peptides must be considered. The different isoforms were gathered in groups based on sequence similarities. Principal component analysis (PCA) was performed on the sequence alignment of the 12 protein isoforms of Ara h 3, which generated a spatial representation of the sequences' similarities. The two-dimensional PCA (Fig. 3) indicates the presence of three isoform groups arbitrarily called Group 1, Group 2 and Group 3. The selection of peptides should, therefore, contain peptides from each group. These 3 groups can be correlated to the variation observed in Fig. 2. Indeed, since the sequences of different isoforms of Ara h 3 are not fully identical, some peptides are specific to particular isoforms or groups of isoforms. The complete peptide distribution among the different isoforms and groups is displayed in Table 2.

Of the 32 selected peptides of Ara h 3, 5 peptides are specific to Group 3, which is composed of the sole Q647H2 isoform of Ara h 3. These peptides display weak MS signal intensities when compared to the intensity of other Ara h 3 peptides, such as AQSENYEYLAFK and



Fig. 1. Effect of food processing on Ara h 1 and Ara h 3. Exposed peptides fulfil the sequence based criteria. For each market type (Spanish and Virginia), five peanut matrices are compared. From left to right: no process, heated peanut, low pH matrix (yoghurt), caramelized peanut and fatty rich and complex environment (chocolate). The comparison is based on the MS signal intensity. The results are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 independent replicates. In each histogram, peptides are separated in two intensity-based groups with different axis scales to improve low intensity peptides visualisation.

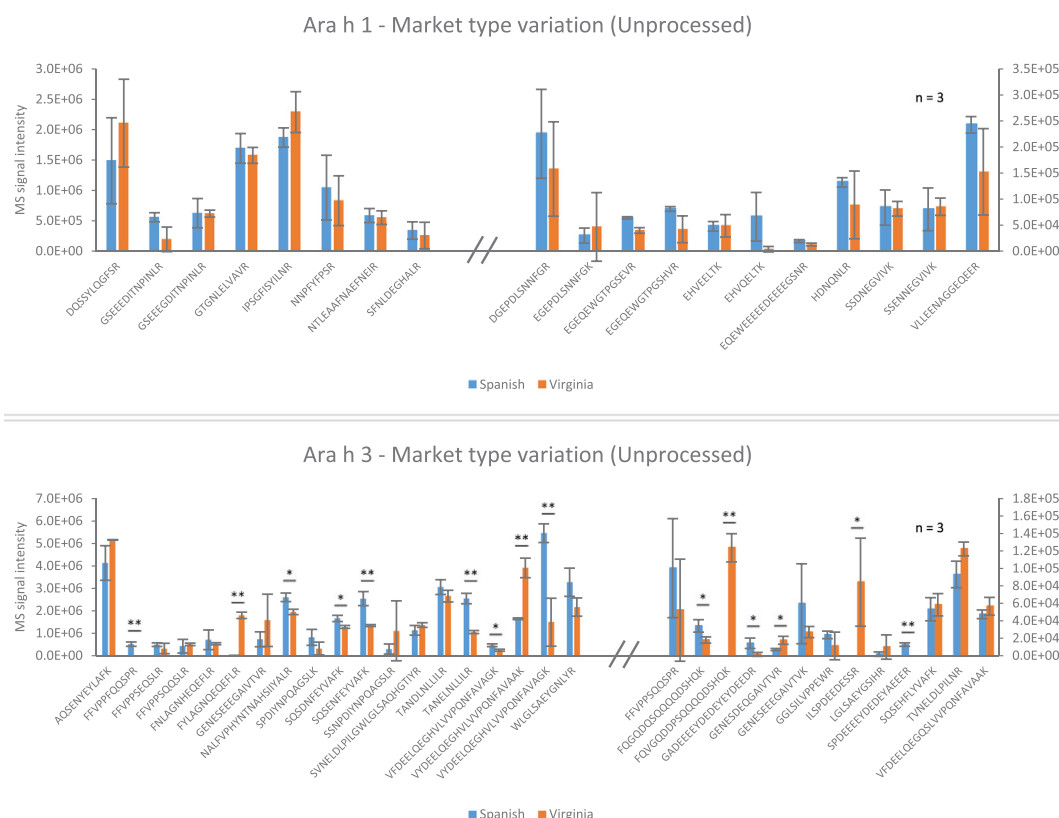


Fig. 2. Peptide MS signal intensity comparison between the two peanut market types (Spanish and Virginia) for Ara h 1 and Ara h 3. Only peptides fulfilling sequence based criteria were selected. Exposed results correspond to analysis of unprocessed peanuts and are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 independent replicates. In each histogram, peptides are separated in two intensity-based groups with different axis scales to improve low intensity peptides visualisation. Statistical significance is evaluated with Student's T-test (* = $p < 0.05$ and ** = $p < 0.01$).

TANDLNLILR, which may be due to a lower abundance of this isoform or, potentially, to technical effects (e.g., incomplete digestion or poor ionization). Given their very low MS intensities, these peptides were not considered as interesting biomarkers. However, the Q647H2 isoform of Ara h 3 could be present at higher levels in other peanut varieties or under different growth conditions. Therefore, at least one peptide of this isoform was considered in our selection. Based on its robustness to food processing and its MS signal intensity, TVNELDLPILNR seems the best candidate to represent this isoform by our selected peptides.

MS signal intensities associated with peptides from Group 2 seem to be higher in the Virginia market type samples. This Group 2 is composed of 2 protein isoforms, and 7 sequence-based selected peptides are specific to them. Among these peptides, 4 have a significantly higher MS signal intensity or are only detected in samples from the Virginia market type. In this case, since samples were prepared with the same protocol, the only different parameter is the peanut market type. It could, therefore, be concluded that protein isoforms from Group 2 are more abundant in Virginia market type peanuts.

Group 1 corresponds to 9 protein isoforms and 20 sequence-based selected peptides specific to this group. Two peptides (i.e., FNLAGNHEQEFLR and WLGLSAEYGNLYR) are found in all 9 isoforms, whereas others are specific to protein isoforms, display equivalent MS signal intensities in both market types (e.g., FNLAGNHEQEFLR and TANDLNLILR), or display a higher intensity in Spanish market type peanut samples (e.g., TANELNLILR and VYDEELQEGHVLVPQNFVAVAGK). None of the peptides in Group 1 are significantly more intense in Virginia market type samples. It can, therefore, be concluded that some protein isoforms of Group 1 are more abundant in Spanish market type peanuts.

Based on these considerations, peptides from Ara h 1 seem to be the

best potential biomarkers. Indeed, no variation between the two considered market types was observed. Moreover, the number of isoforms is limited, and some sequence-based selected peptides are well conserved in all isoforms. However, Ara h 3 is the most abundant protein in peanuts and, hence, a suitable target to ensure high sensitivity in a quantitative method. A combination of peptides could allow coverage of all Ara h 3 isoforms, as well as the global quantification of Ara h 3, so at this stage, peptide biomarker selection should not be limited to Ara h 1 peptides. Because of the high abundance of Ara h 3, peptides from this allergen were also selected despite the observed variabilities.

At this stage, considering sequence-based criteria, robustness to food processing, and the multiple isoforms, 17 peptides were pre-selected (i.e., 5 from Ara h 1 and 12 from Ara h 3). With this pre-selection, all isoforms of both Ara h 1 and Ara h 3 are covered, ensuring method robustness. The final selection of biomarkers will be performed during the development of the quantitative UHPLC-MS/MS method.

3.6. Specificity of selected peptide biomarkers and comparison with the literature

Via an empirical approach using HRMS, more than 300 peanut peptides were identified in the various peanut-based test materials. Applying the criteria extensively described in the previous sections, these peptides were refined to obtain a selection of potentially salient peptide biomarkers in peanuts. The last step involves assessing the selected peptides' specificity to peanuts. This verification was performed using basic local alignment search tool (BLAST) analyses considering the UniProt and NCBI protein databases. The selected peptides must be specific to peanuts and not found in the protein sequences of other food ingredients because, otherwise, false-positive results could occur.

Table 2

Specificity of the sequence based selected peptides to the different protein isoforms of Ara h 1 and Ara h 3 identified in UniProt database (uniprot.org). Isoforms of Ara h 3 are separated in the 3 artificial groups highlighted by the PCA performed on the protein sequences alignment.

Ara h 1 peptides	B3IXL2/P43237/Q6PSU3			N1NG13/P43238			E5G076		
DGEPDLSNNFGR	X								
DQSSYLQGFGR	X				X				X
EGEPDLSNNFGK					X				
EGEQEWGTPGSEVR	X								
EGEQEWGTPGSHVR					X				X
EHVEELTK					X				X
EHVQELTK	X								
EQEWEEEEDEEEGSNR	X								X
GSEEDITNPINLR	X								
GSEEDITNPINLR					X				X
GTGNLELVAVR	X				X				X
HDNQNLNR	X				X				
IPSGFISYILNR	X				X				
NNPFYFPSR	X				X				X
NTLEAAFAEFNEIR	X				X				X
SFNLDEGHALR	X				X				
SSDNEGIVVK	X								
SSENNEGIVVK					X				
VLEENAGGEQEER	X				X				X
WGPAEPR	X								

Ara h 3 peptides	Group 1									Group 2		Group 3
	A1DZF0	B5TYU1	Q5I6T2	Q6T2T4	Q647H3	Q647H4	Q8LKN1	Q9FZ11	Q9SQH7	E5G077	Q0GM57	Q647H2
AQSENYEYLAFK										X	X	
FFVPPFQQSPR	X								X			
FFVPPSEQSLR				X		X	X					
FFVPPSQQSLR								X				
FFVPPSQQSPR		X	X									
FNLAGNHEQEFLR	X	X	X	X	X	X	X	X	X			
FQGGDQSQQQDSHQK				X		X	X					
FQVGQDDPSQQQDSHQK										X	X	
FYLAGNQEQEFLR										X	X	
GADEEEYDEDEYDEEDR		X	X					X				
GENEDEQGAIVTVR				X	X	X	X					
GENESEEQGAIVTVK								X				
GENESEEQGAIVTVR			X						X			
GGLSILVPEWR												X
ILSPDEEDESRR										X	X	
LGLSAEYGSIRH												X
NALFVPHYNTNAHSIIYALR	X	X	X	X	X	X	X	X	X			
SPDEEEYDEDEYAEER	X											
SPDIYNPQAGSLK	X	X	X	X	X	X	X	X				
SQSDNFEYVAFK		X	X		X			X				
SQSEHFLYVAFK												X
SQSENFYVAFK	X			X		X	X		X			
SSNPDIYNPQAGSLR										X	X	
SVNELDLPIGLWGLSAQHGTIYR										X	X	
TANDLNLILR			X		X			X				
TANELNLILR	X	X		X		X						
TVNELDLPIILNR												X
VFDEELQEGHVLVVPQNFVAVGK				X		X	X					
VFDEELQEGQLVVPQNFVAAK												X
VYDEELQEGHVLVVPQNFVAAK										X	X	
VYDEELQEGHVLVVPQNFVAVGK	X	X	X		X			X	X			
WLGLSAEYGNLYR	X	X	X	X	X	X	X	X	X			

Among the 17 pre-selected peptides, only the peptide FYLAGNQEQEFLR from Ara h 3 was rejected based on BLAST results as this peptide is also present in faba beans (*Vicia faba* var. *minor*) protein. Since this bean is used not only in animal feed but also in human food (Crépon et al., 2010; Smith & Hardacre, 2011), this peptide was rejected because of its lack of specificity to peanuts. The final selected peptide biomarkers (Table 3) contain 16 peptides from the two most abundant proteins in peanuts, Ara h 1 (5 peptides) and Ara h 3 (11 peptides). In the future, these peptides will be used to develop a UHPLC-MS/MS routine method to quantify peanuts in processed food products.

The 16 selected peanut peptides were, finally, compared with biomarkers described in the literature. The literature review is summarized

in [Supplementary Material 2](#), where our potential peptide biomarkers are highlighted in bold, and the matrices used in the different papers are indicated. A total of 30 peanut biomarkers were identified from the literature, and most of them (28 out of 30) are from Ara h 1, Ara h 2, or Ara h 3. Some of our selected peptides, such as GTGNLELVAVR (Ara h 1) and WLGLSAEYGNLYR (Ara h 3), appear to be recurrent biomarkers. However, some cited peptides do not meet our selection criteria and were not considered as potential biomarkers in this study. For example, peptide CMCEALQQIMENQSDR, which was found in 3 publications, contains potentially modified cysteines and methionines. Furthermore, some of our selected peptides (3 from Ara h 1 and 3 from Ara h 3) were not identified as biomarkers in the considered literature. These peptides

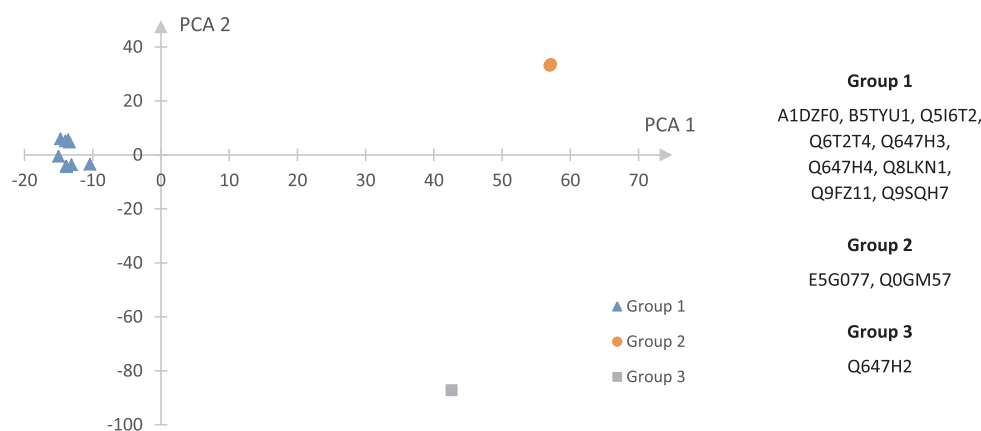


Fig. 3. Two dimensional principal component analysis (PCA) performed on the protein sequence alignment of the 12 isoforms of Ara h 3 found in the UniProt database. PCA highlight the presence of 3 groups of isoforms with close protein sequences and called Group 1, Group 2 and Group 3.

Table 3

Final potential peptide biomarkers selection fulfilling the above-mentioned criteria. The peptides are linked to their respective allergen (Ara h 1 and Ara h 3).

Allergen	Selected peanut peptide biomarkers
Ara h 1	DQSSYLQGFSR GSEEDITNPINLR GSEEGDITNPINLR GTGNLELVAVR NNPFYFPSR
Ara h 3	AQSENYEYLAFAK FNLAGNHEQEFLR NALFVPHYNTNAHSIIYALR SQSDNFEYVAFK SQSENFYVAFK TANDLNLILR TANELNLILR TVNELDLPIILNR VYDEELQEGHVLVVPQNFVAFAAK VYDEELQEGHVLVVPQNFVAVAGK WLGLSAEYGNLYR

are DQSSYLQGFSR, GSEEDITNPINLR, and GSEEGDITNPINLR for Ara h 1, and NALFVPHYNTNAHSIIYALR, SQSDNFEYVAFK, and TVNELDLPIILNR for Ara h 3. Hence, they could be considered as new peptide biomarkers for peanuts as they meet our established selection criteria. We noticed that none of the 15 Ara h 3 peptide biomarkers identified in the literature are present in the Q647H2 Ara h 3 isoform sequence. This omission could lead to bias and total Ara h 3 under-estimation in a future quantitative method.

4. Conclusions

An allergic reaction to peanuts is one of the most serious types of food allergies, considering their prevalence and the potential severity of this reaction, which can lead to anaphylactic shock. To provide more protection for allergic patients via improved food labeling, accurate detection and quantification methods are needed for peanuts and, more broadly, all food allergens. Mass spectrometry-based methods are promising and can overcome the limitations of existing methods.

Identifying potential allergen biomarkers is the first step in the development of a mass spectrometry-based method. Using an empirical approach based on high-resolution mass spectrometry, this study presents the identification of potential peptide biomarkers to detect and quantify peanuts in processed food.

More than 300 peptides were identified during the analysis of various processed peanut matrices (i.e., raw peanuts, heated peanuts, a low-pH peanut matrix, caramelized peanuts, and peanuts in a fat-rich

environment). These peptides were filtered using a set of selection criteria to ensure method specificity, sensitivity, trueness, and robustness. Ideal peptide biomarkers must be specific to peanuts, belong to abundant proteins, and be robust to food processing but not prone to missed cleavages by the protease used or to amino acid modifications. Multiple protein isoforms and origin variation issues were also considered by analyzing peanuts from two geographical regions corresponding to two market types. Ara h 1 includes 3 isoforms with relatively close sequences and several common tryptic peptides. No significant differences were observed between the two market types (i.e., Virginia and Spanish) concerning the amount of Ara h 1. However, Ara h 3 corresponds to 12 protein isoforms with more significant sequence variations, and no tryptic peptide was found conserved between these 12 isoforms. The abundance of some Ara h 3 isoforms was found to vary by peanut market type. These aspects are hardly considered in peanut peptide biomarkers identification studies or in detection and quantification method development. They are, however, essential to ensure the robustness of the corresponding method. Our selection includes peptides covering all protein isoforms of Ara h 1 and Ara h 3. Our approach led to the identification of 16 potential peanut peptide biomarkers. These peptides will be used as targets in the future development of a routine UHPLC-MS/MS method to detect and quantify allergens in processed food products.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS. The authors also thank MaSUN, the Mass Spectrometry facility of the University of Namur (URBC-Narilis).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2019.125428>.

References

- Andersen, P. C., Hill, K., Gorbett, D. W., & Brodbeck, B. V. (1998). Fatty acid and amino acid profiles of selected peanut cultivars and breeding lines. *Journal of Food*

- Composition and Analysis, 11(2), 100–111. <https://doi.org/10.1006/JFCA.1998.0565>.
- Bégin, P., Winterroth, L. C., Dominguez, T., Wilson, S. P., Bacal, L., Mehrotra, A., ... Nadeau, K. C. (2014). Safety and feasibility of oral immunotherapy to multiple allergens for food allergy. *Allergy, Asthma & Clinical Immunology*, 10(1), 1. <https://doi.org/10.1186/1710-1492-10-1>.
- Boo, C. C., Parker, C. H., & Jackson, L. S. (2018). A targeted LC-MS/MS method for the simultaneous detection and quantitation of egg, milk, and peanut allergens in sugar cookies. *Journal of AOAC International*, 101(1), 108–117. <https://doi.org/10.5740/jaoacint.17-0400>.
- Bucchini, L., Guzzon, A., Poms, R., & Senyuva, H. (2016). Analysis and critical comparison of food allergen recalls from the European Union, USA, Canada, Hong Kong, Australia and New Zealand. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, 33(5), 760–771. <https://doi.org/10.1080/19440049.2016.1169444>.
- Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., & Duc, G. (2010). Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Research*, 115, 329–339. <https://doi.org/10.1016/j.fcr.2009.09.016>.
- Croote, D., & Quake, S. R. (2016). Food allergen detection by mass spectrometry: The role of systems biology. *NPJ Systems Biology and Applications*, 2, 16022. <https://doi.org/10.1038/npsba.2016.22>.
- De Silva, D., Geromi, M., Panesar, S. S., Muraro, A., Werfel, T., Hoffmann-Sommergruber, K., ... Sheikh, A. (2014). Acute and long-term management of food allergy: Systematic review. *European Journal of Allergy and Clinical Immunology*, 69, 159–167. <https://doi.org/10.1111/all.12314>.
- Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N. (2018). Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *Journal of Chromatography A*. <https://doi.org/10.1016/J.CHROMA.2018.11.036>.
- Henrotin, J., Planque, M., Huet, A.-C., Marega, R., Lamote, A., & Gillard, N. (2019). Gluten analysis in processed foodstuffs by a multi-allergens and grain-specific UHPLC-MS/MS method: One method to detect them all. *Journal of AOAC International*. <https://doi.org/10.5740/jaoacint.19-0057>.
- Husain, Z., & Schwartz, R. A. (2012). Peanut allergy: An increasingly common life-threatening disorder. *Journal of the American Academy of Dermatology*, 66(1), 136–143. <https://doi.org/10.1016/J.JAAD.2011.02.031>.
- Jin, T., Guo, F., Chen, Y. wei., Howard, A., & Zhang, Y. Z. (2009). Crystal structure of Ara h 3, a major allergen in peanut. *Molecular Immunology*, 46(8–9), 1796–1804. <https://doi.org/10.1016/j.molimm.2009.01.023>.
- Johnson, P. E., Sayers, R. L., Gethings, L. A., Balasundaram, A., Marsh, J. T., Langridge, J. I., & Mills, E. N. C. (2016). Quantitative proteomic profiling of peanut allergens in food ingredients used for oral food challenges. *Analytical Chemistry*, 88(11), 5689–5695. <https://doi.org/10.1021/acs.analchem.5b04466>.
- Khuda, S. E., Jackson, L. S., Fu, T. J., & Williams, K. M. (2015). Effects of processing on the recovery of food allergens from a model dark chocolate matrix. *Food Chemistry*, 168, 580–587. <https://doi.org/10.1016/j.foodchem.2014.07.084>.
- Koppelman, S. J., Jayasena, S., Luykx, D., Schepens, E., Apostolovic, D., de Jong, G. A. H., ... Maleki, S. (2016). Allergenicity attributes of different peanut market types. *Food and Chemical Toxicology*, 91, 82–90. <https://doi.org/10.1016/J.FCT.2016.02.016>.
- Kosky, A. A., Dharmavaram, V., Ratnaswamy, G., & Manning, M. C. (2009). Multivariate analysis of the sequence dependence of asparagine deamidation rates in peptides. *Pharmaceutical Research*, 26(11), 2417–2428. <https://doi.org/10.1007/s11095-009-9953-8>.
- Kottapalli, K. R., Payton, P., Rakwal, R., Agrawal, G. K., Shibato, J., Burow, M., & Puppala, N. (2008). Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. *Plant Science*, 175(3), 321–329. <https://doi.org/10.1016/J.PLANTSCI.2008.05.005>.
- Longo, G., Berti, I., Burks, A. W., Krauss, B., & Barbi, E. (2013). IgE-mediated food allergy in children. *The Lancet*, 382, 1656–1664. [https://doi.org/10.1016/S0140-6736\(13\)60309-8](https://doi.org/10.1016/S0140-6736(13)60309-8).
- Monaci, L., Pilolli, R., De Angelis, E., & Mamone, G. (2015). Mass spectrometry in food allergen research. *Comprehensive Analytical Chemistry*, 68, 359–393. <https://doi.org/10.1016/B978-0-444-63340-8.00007-8>.
- Moretzsohn, M. de C., Hopkins, M. S., Mitchell, S. E., Kresovich, S., Valls, J. F. M., & Ferreira, M. E. (2004). Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology*, 4, 11. <https://doi.org/10.1186/1471-2229-4-11>.
- Palladino, C. (2018). Peanut allergens. *Molecular Immunology*, 100, 58–70. <https://doi.org/10.1016/J.MOLIMM.2018.04.005>.
- Planque, M., Arnould, T., Delahaut, P., Renard, P., Dieu, M., & Gillard, N. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chemistry*, 274, 35–45. <https://doi.org/10.1016/j.foodchem.2018.08.095>.
- Pomés, A., Davies, J. M., Gadermaier, G., Hilger, C., Holzhauser, T., Lidholm, J., ... Goodman, R. E. (2018). WHO/IUIS Allergen Nomenclature: Providing a common language. *Molecular Immunology*, 100, 3–13. <https://doi.org/10.1016/J.MOLIMM.2018.03.003>.
- Popping, B., & Godefroy, S. B. (2011). Allergen detection by mass spectrometry-The new way forward. *Journal of AOAC International*.
- Reimer, J., Shamshurin, D., Harder, M., Yamchuk, A., Spicer, V., & Krokhin, O. V. (2011). Effect of cyclization of N-terminal glutamine and carbamidomethyl-cysteine (residues) on the chromatographic behavior of peptides in reversed-phase chromatography. *Journal of Chromatography A*, 1218(31), 5101–5107. <https://doi.org/10.1016/J.CHROMA.2011.05.079>.
- Robinson, N. E., & Robinson, A. B. (2004). Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects. *Journal of Peptide Research*, 63(5), 437–448. <https://doi.org/10.1111/j.1399-3011.2004.00148.x>.
- Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L., & Robinson, A. B. (2004). Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *The Journal of Peptide Research*, 63(5), 426–436. <https://doi.org/10.1111/j.1399-3011.2004.00151.x>.
- Sayers, R. L., Johnson, P. E., Marsh, J. T., Barran, P., Brown, H., & Mills, E. N. C. (2016). The effect of thermal processing on the behaviour of peanut allergen peptide targets used in multiple reaction monitoring mass spectrometry experiments. *Analyst*, 141(13), 4130–4141. <https://doi.org/10.1039/C6AN00359A>.
- Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I Papain. *Biochemical and Biophysical Research Communications*, 27(2), 157–162. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X).
- Schmidt, H., Gelhaus, C., Latendorf, T., Nebendahl, M., Petersen, A., Krause, S., ... Janssen, O. (2009). 2-D DIGE analysis of the proteome of extracts from peanut variants reveals striking differences in major allergen contents. *Proteomics*, 9(13), 3507–3521. <https://doi.org/10.1002/pmic.200800938>.
- Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology*, 133. <https://doi.org/10.1016/j.jaci.2013.11.020>.
- Siepen, J. A., Keevil, E.-J., Knight, D., & Hubbard, S. J. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of Proteome Research*, 6(1), 399–408. <https://doi.org/10.1021/pr060507u>.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Retrieved from *Molecular Systems Biology*, 7(1)<http://msb.embopress.org/content/7/1/539.abstract>.
- Smith, J., & Hardacre, A. (2011). Development of an extruded snack product from the legume *Vicia faba* minor. *Procedia Food Science*, 1, 1573–1580. <https://doi.org/10.1016/J.PROFOO.2011.09.233>.
- Van Hengel, A. J. (2007). Food allergen detection methods and the challenge to protect food-allergic consumers. *Analytical and Bioanalytical Chemistry*, 389, 111–118. <https://doi.org/10.1007/s00216-007-1353-5>.
- Walczyk, N. E., Smith, P. M. C., Tovey, E., Wright, G. C., Fleischfresser, D. B., & Roberts, T. H. (2013). Analysis of crude protein and allergen abundance in peanuts (*Arachis hypogaea* cv. Walter) from three growing regions in Australia. *Journal of Agricultural and Food Chemistry*, 61(15), 3714–3725. <https://doi.org/10.1021/jf305347r>.
- Walker, M. J., Burns, D. T., Elliott, C. T., Gowland, M. H., & Mills, E. N. C. (2016). Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs. *The Analyst*, 141(1), 24–35. <https://doi.org/10.1039/c5an01457c>.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
- Wu, Z., Zhou, N., Xiong, F., Li, X., Yang, A., Tong, P., ... Chen, H. (2016). Allergen composition analysis and allergenicity assessment of Chinese peanut cultivars. *Food Chemistry*, 196, 459–465. <https://doi.org/10.1016/j.foodchem.2015.09.070>.