Critical review on proteotypic peptide marker tracing for six allergenic ingredients in incurred foods by mass spectrometry

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

OPEN ACCESS

CRITICAL REVIEW ON PROTEOTYPIC PEPTIDE MARKER TRACING FOR SIX ALLERGENIC INGREDIENTS IN INCURRED FOODS BY MASS SPECTROMETRY

Rosa Pilolli[1], Chiara Nitride[2], Nathalie Gillard[3], Anne- Catherine Huet[3], Christof van Poucke[4], Marc de Loose[4], Olivier Tranquet[5], Collette Larré[5], Karine Adel-Patient[6], Hervé Bernard[6], E. N. Clare Mills[2], and Linda Monaci[1*]

1   Institute of Sciences of Food Production, National Research Council of Italy (ISPA-CNR), via Giovanni Amendola 122/O - 70126 Bari, Italy;

2   School of Biological Sciences, Division of Infection, Immunity and Respiratory Medicine, Manchester Academic Health Science Centre, Manchester Institute of Biotechnology, The University of Manchester, UK;

3   CER Groupe, Rue du point du Jour, 8, 6900 Marloie, Belgium;

4   Flanders Research Institute for Agriculture, Fisheries and Food, Brusselsesteenweg 370, 9170 Melle, Belgium;

5   UR1268 BIA, Rue de la Géraudière, BP 71627, 44316 Nantes, France;

6   INRA-CEA, Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Immuno-Allergie Alimentaire, Bât. 136-CEA de Saclay, 91191 - Gif-sur-Yvette, France.

* corresponding author: linda.monaci@ispa.cnr.it

Tel: +39 080 592 9343

ABSTRACT

Peptide marker identification is one of the most important steps in the development of a mass spectrometry (MS) based method for allergen detection, since the robustness and sensitivity of the overall analytical method will strictly depend on the reliability of the proteotypic peptides tracing for each allergen. The European legislation in place issues the mandatory labelling of fourteen allergenic ingredients whenever used in different food formulations. Among these, six allergenic ingredients, namely milk, egg, peanut, soybean, hazelnut and almond, can be prioritized in light of their higher occurrence in food recalls for undeclared presence with serious risk decision.

In this work, we described the results of a comprehensive evaluation of the current literature on MS-based allergen detection aiming at collecting all available information about proteins and peptide markers validated in independent studies for the six allergenic ingredients of interest. The main features of the targeted proteins were commented reviewing all details available about known isoforms and sequence homology particularly in plant-derived allergens. Several critical aspects affecting peptide markers reliability were discussed and according to this evaluation a final short-list of candidate markers was compiled likely to be standardized and implemented in MS methods for allergen analysis.

1. Introduction

IgE-mediated food allergies are steadily on the rise representing a huge health concern worldwide. Although firm prevalence data are lacking, there are extensive investigations suggesting that food allergies have increased and rates as high as approximately 10 % have been documented, with main prevalence in industrialized countries, and in children compared to adults (Scott et al. 2018). There are many manifestations of food allergy with different severities, and individual investigations rely on specific study populations, specific foods, and different methodologies, thus impairing the determination of undoubted statistics. Moreover, further constrains in obtaining solid prevalence data, arise from geographic variations, diet exposure effects, differences according to age, race, and ethnicity, and myriad other factors. Like all chronic disease, onset of food allergy is influenced by genetics, environment and genome-environment interactions including epigenetics effects. Its current management is based on allergen containing food avoidance. Food labelling legislative provisions are available in several countries and require a detailed declaration of allergen inclusion in foods, with number and types of priority allergens differing among countries and depending on specific dietary habits and allergy prevalence. On the other hand, a precautionary labeling system is also widespread, representing a safety measure put in place from industries to protect from any accidental cross-contamination likely to occur in foods expected to be allergen-free and that can take place at whatever stage along the food chain. All this raised the urgent need to develop reliable methodologies to trace allergens in foods with the highest confidence. Mass spectrometry (MS) techniques have been widely exploited in the last ten years to meet this goal and were specifically applied to food allergen detection, and to the food allergen identification and characterization, as well. Despite the need for expensive equipment and trained personnel,

66 the chance to provide multiplexing and unequivocal allergen identification accounts for the

67 overall strength of the MS based approaches compared to previously established methods,

68 such as immunoassays and DNA based methods. Nonetheless some gaps in the development

69 of MS-based quantitative methods still need to be addressed. The review paper authored by

70 Monaci et al., 2018 provides a comprehensive overview of the MS based methods developed

71 so far for allergen detection and also highlights the need of harmonization in method

72 development and validation.

73 In this frame, the European project ThRAll (Thresholds and Reference method for Allergen

74 detection method) titled "Detection and quantification of allergens in foods and minimum

75 eliciting doses in food allergic individuals", funded by the European Food Safety Agency will

76 actively contribute to the advancement in harmonization of MS-based method for food allergen

77 detection by developing a prototype quantitative reference method for the multiple detection of

78 food allergens in standardized incurred food matrices (Mills et al. 2019). Five main food

79 ingredients causing severe IgE-mediated reactions have been prioritized in light of their

80 occurrence in food recalls (Turner et al., 2015; Worm et al., 2014; Bucchini et al., 2016). Milk,

81 soybean, tree nuts, egg, and peanuts were selected as responsible for most alerts notified on

82 the RASFF (Rapid Alert System for Food and Feed) portal for undeclared presence with serious

83 risk decision. A recent notice from the European Commission (Commission Notice of

84 13.7.2017) relating to the provision of information on substances or products causing allergies

85 or intolerances as listed in Annex II of Regulation (EU) No 1169/2011 provides specifications

86 about how the Annex II of Regulation (EU) No 1169/2011 should be implemented. In particular,

87 as for milk and egg ingredients the notice clarifies that the legislation should be applied to milk

88 from all farmed animals and egg from all farmed birds. Given the timing of the ThRAll project,

4

89   and the urgent demand for harmonization of current analytical methods, the Consortium agreed

90   to prioritize cow's milk and hen's egg allergens, as first targets, also in light of their main role in

91   the primary production. Indeed according to recent statistical data (FaoStat, 2017), cow's milk

92   production represented in 2017 the 97% of total milk production in Europe and the 81% of the

93   production worldwide. Similarly, in the same year the hen's eggs accounted for the 99% of total

94   egg production in Europe and for the 82% worldwide. Whether the developed approach would

95   be applicable to all farmed animals/birds will be specified depending on the final validated

96   analytical markers. As for tree nuts allergens, two representative ingredients have been

97   selected, hazelnut and almond, which are widely used in food manufacturing and represent

98   foods of public health importance in Europe, causing severe reactions (Worm et al., 2014).

99   The allergenic ingredients will need to be incorporated into the food before being processed to

100   mimic as closely as possible the manufacturing process. Two model matrices will need to be

101   produced *ad hoc* in a food pilot plant, namely a chocolate bar and a broth powder. These latter

102   are the representative food commodities responsible for recalls, namely baked goods,

103   confectionary products and complex multiphase foods and pose specific challenges in MS

104   based analytical detection. Both matrices were selected because they are extremely complex,

105   with different features, having chocolate high contents of fat and polyphenols whilst broth

106   powder is an extensive processed food (cooking, boiling and drying), enriched with proteins.

107   One of first tasks scheduled in the ThRAll project (Objective 1) concerns the identification of

108   protein and peptide markers for the targeted allergenic ingredients. As a fact, markers

109   identification is one of the most important steps in the method development. Different workflows

110   are available to accomplish this objective and they have been reviewed in a recent paper

111 authored by Downs & Johnson in 2018. All issues besetting such choice including, various

112 ingredients formulations, extraction yield in thermally processed foods, and incomplete genomic

113 and proteomic sequence information, as well as all common issues associated with MS-based

114 protein quantitation were critically discussed. The main point raised by Downs & Johnson was

115 that the target selection cannot be accounted by *in-silico* only approach. Methodological details,

116 such as matrix composition, protein extraction and digestion, should be taken into consideration

117 for marker selection, affecting both the specificity and sensitivity of the final analytical method.

118 The assessment on a model matrix is required firstly for specificity issue because interfering

119 peaks from the matrix background may cause false-positives; in addition the processing degree

120 of the investigated food matrix may also differently affect the target protein/peptide detection

121 resulting in false-negative which can pose safety risk for the allergic consumer. None of these

122 issues can be tackled by the solely *in-silico* based approach, thus the need for an empirical

123 validation of the candidate peptide markers in incurred foods (Monaci et al., 2018).

124 An alternative approach could rely on the comprehensive assessment of the allergen detection

125 literature, supported by the experimental evidence that if the same peptides for given allergenic

126 proteins have been detected in independent studies and in multiple matrices, they can likely be

127 standardized (Croote & Quake, 2016). In order to promote this approach, an open-source

128 repository of discovery and targeted MS data on allergen detection and quantitation has been

129 developed and made available as public bioinformatic tool (Allergen Peptide Browser, 2018).

130 The database has been structured around proteins recognized by the World Health

131 Organization and International Union of Immunological Societies (WHO/IUIS) as food allergens.

After a proper screening of the different available options, the partners of ThRAll consortium agreed on undertaking peptide selection by applying a dual approach with different time scale: (i) critical evaluation of the peptide markers already reported in the existing literature; (ii) validation of the candidate peptide list by discovery analysis on the specific incurred matrices under investigation (chocolate bars and broth powder). Herein, we will describe the results of the first point, namely the comprehensive evaluation of the literature, providing a thorough discussion about the application of general acceptance criteria for harmonization of the markers selection step. The goal of this work is to draw a preliminary list of reliable peptide markers, which will be validated later on by untargeted HR-MS/MS analysis of incurred matrices, representing the bases for the selected reaction monitoring (SRM) based reference method under development within the ThRAll project.

2. Literature review based marker peptide selection

The analytical workflow displayed in Fig. 1 was designed to provide the comprehensive evaluation of the current literature. It consists of five main steps aimed at the critical evaluation of the information gathered from previous works and the proper selection of reported signature peptides according to specific criteria. We collected all information about the six allergens selected as targets in the ThRAll project (cow's milk, hen's egg, peanut, soybean, hazelnut, and almond) (step 1). Afterwards, the list of reported signature peptides have been evaluated and filtered according to specific features, namely sequence length, matrix similarity with the ThRAll project, kind of investigation (discovery only or targeted analysis) and occurrence of amino acid residues prone to endogenous and exogenous modifications (step 2). The third step

of the workflow consisted in retrieving all the information available for the relevant allergenic proteins (relative abundance, post-translational modification sites, protein isoforms and variants), from different informatics sources detailed in Figure 1. At step 4 peptide specificity was assessed by sequence alignment with other species searching not only for exact matches of tryptic peptides (performed for all candidate markers of the six allergenic ingredients) but also disclosing potential interference from sequence homology with single amino acid mismatches (performed only on the four plant-derived allergenic ingredients). Finally as last step of the analytical workflow, for the refined peptide list we collected experimental details about the MS-based detection, such as the instrumental platform, the precursor ions, and the selected transitions. Indeed, the peptide ionization efficiency and reproducibility, as well as the relative intensity of fragmentation pattern might be affected also by instrumental parameters which need to be assessed case-by-case. For the comprehensive assessment of the allergen detection literature (step 1 in Fig.1) we exploited the Allergen Peptide Browser (APB) database available online, and collected all information about the six allergens selected for the ThRAll project. The gathered information also integrated data fed by other studies based on MS/MS analysis with either ESI or MALDI ionization, encompassing the most recent publications not itemized yet on APB. A total of 89 papers were reviewed, published in the last twenty-five years, which resulted in an exhaustive list of tryptic peptides for milk (total 42 peptides), egg (total 62 peptides), peanut (total 50 peptides) and soybean (total 146 peptides), and a shorter, but still reasonable list for hazelnut (total 26 peptides) and almond (total 29 peptides) allergens (see Table S1). In the second step the full list was narrowed down according to specific constraints. A range of 7-20 amino acids (AA) in length was applied as good compromise between peptide specificity, ionization yield and reproducibility. About 84% of reported peptides were included

8

in this range. Afterwards, we searched for matrix similarity with the ThRAll project, and marker peptides that were never validated in either chocolate or highly processed incurred matrices were rejected. The resulting peptides list is displayed in Fig. 2 and Fig. 3 (and itemized in Table S2) together with details about the total number of citing papers per each peptide, and the kind of investigation (targeted vs discovery approaches), marked with different color (orange and blue, respectively).

The marker selection should comply with specific requirements that value the quantitative information retrieved. Indeed, the allergen quantification depends on the concentration of peptide markers, based on the assumption that the peptide is released completely from the original protein with a specific stoichiometric ratio. Any unpredictable distribution of the released peptide into different precursor ions, for example due to amino acid modification or multiple ionization states, should be traced and susceptible peptides should be excluded, in case of low reproducibility. Concerning the exclusion of signature peptides containing labile amino acids prone to modifications, it deserves to be noted that the number of potential endogenous and exogenous post-translational modifications is huge, but their occurrence is highly variable; therefore, it is not a realistic goal neither to evaluate nor to exclude them all strictly, rather an empirical validation of the specific case study should be preferable. In this investigation, we focused on the occurrence of specific amino acids of major concern, such as methionine, asparagine, glutamine, and cysteine, excluding only peptide sequences containing methionine residues prone to processing-induce oxidation, and asparagine-glycine motifs prone to deamidation (Li et al., 2005; Li et al., 2006). We deemed important for the robustness of the analytical method to avoid M containing peptides as analytical targets, because such residue would be highly prone to oxidation, in thermally processed matrices, in barely predictable

200 proportions. The NG motifs could be subjected to spontaneous deamidation *in-vitro*, however,

201 in case of need the asparagine deamidation might be forced in order to convert all asparagine

202 residues into aspartic acid. As for cysteine residues, most digestion protocols included a

203 reduction/alkylation step (mainly based on dithiothreitol and iodoacetamide), preliminary to

204 actual enzymatic proteolysis, which is specifically designed to break disulfide bridges and block

205 irreversibly the cysteine residues by carbamidomethylation. Common sense is to consider this

206 modification as complete however, in the perspective to avoid source of uncertainty, C

207 containing peptides should be preferentially excluded unless the reliability of the peptide is duly

208 assessed experimentally.

209 As third step of the analytical workflow reported in Fig. 1, starting from the refined list of peptide

210 markers, we sought for additional information available on the target proteins, which could affect

211 the detection sensitivity and reliability, such as relative abundance, natural post-translational

212 modification (PTM), protein isoforms and variants, gathering data from different sources (either

213 scientific publications or on-line databases, i.e. Uniprot, Allergome, WHO/IUIS).

214 Peptide specificity was assessed by BLAST search against the most common protein

215 databases disclosing also potential detection interferences in plant-derived proteins (single

216 amino acid mismatches) (step 4). The APB Database queried for this review of peptide markers,

217 already provided information about sequence specificity, retrieved by means of BLAST 2.2.31+

218 (BLAST DB, Sep 20, 2016). Only hits with 100% identity and no gaps were retained for a given

219 peptide, and the following constrains were applied to exclude hits: (i) bacterial super kingdom;

220 (ii) absent species name or kingdom; (iii) species / common name entry containing: 'synthetic

221 construct' or 'vector'; (iv) title containing 'partial' or 'fragment'. Since it was not clear whether

222   they considered or not the site specific trypsin cleavage, we also double-checked the specificity

223   for all the six allergenic ingredients by means of the Protein Prospector tool developed by the

224   University of California San Francisco (http://prospector.ucsf.edu/prospector/mshome.htm, accessed

225   on October 24th 2018). In particular, the MS-Homology option searching for 100% identity of

226   tryptic peptides within the main databases, NCBInr.2013.6.17 and UniProtKB.2017.11.01, was

227   applied. In addition, for plant-derived allergens, protein homology among different taxonomies

228   needs to be considered very carefully. Besides the previously discussed full matches (100%

229   identity), in low-resolution MS detection potential interferences might raise also under particular

230   conditions, from peptides with single amino acid mismatches. Therefore, further BLAST search

231   against non-redundant protein database was carried out (access in July 2018), seeking for

232   single amino acid substitutions, which could result in differences of peptide molecular weight

233   within ±1 Da (for example, I/L, D/N, E/Q). These kind of substitutions might generate interfering

234   precursor ions isolated and activated at least partially in the set m/z window, and might provide

235   several transitions, which are not distinguishable from the targeted peptides. Finally, as step 5

236   of the workflow we retrieved experimental details about the MS platform, the isolated ions and

237   the monitored transitions. As a fact, most tryptic peptides feature favorable ionization

238   properties, which provide high ionization efficiency and typically result in either double or triple

239   charged ions due to the occurrence of basic amino acid residues. However, some peptide

240   sequences may present multiple ionization routes resulting in a loss of sensitivity due to the

241   distribution of the total amount of released peptide in different precursor ions, each

242   characterized by a specific fragmentation pattern. Since we use as a calibrant the synthetic

243   natural version of the peptide and use as an anchor the ration measurement to the isotopic

244   labelled version of the peptide, the distribution among different charge state is accounted and

245 thereof can be considered negligible. From the quantitative point of view, by selecting the most

246 intense precursor ions loss of sensitivity can be limited. Besides the amino acid sequence,

247 several experimental features may affect this ionization behavior, e.g. the electrospray source

248 parameters, the MS platform, the composition of the food matrix, dragging competing species

249 to the ionization step. Therefore it is crucial that the theoretical selection of candidate markers,

250 based either on the *in-silico* approach or on the literature review approach should be confirmed

251 by experimental validation on the specific MS platform and on the main matrices of interest.

252 In the sections below, we present the results of the analytical workflow reviewing all information

253 collected for the six allergens of interest.

2.1. Proteins and peptide markers reported for cow's milk and hen's egg

255 The inclusion of milk from domesticated mammalian animals, mainly cattle, buffalo, sheep, goat

256 and camel, boasts a very long tradition in the human diet. According to the FAO data on primary

257 livestock production in 2017, about 96% of the world's fresh milk comes from cows and

258 buffaloes, while the remaining 4.0% is produced by goats, sheep and camels (FAOStat, 2017).

259 Noteworthy, milk from all ruminant species contains homologous proteins, which share the

260 same structural, functional, and biological properties (Monaci et al., 2006). Cow's milk contains

261 about 3 g of proteins per 100 mL and includes at least 25 different proteins classified in two

262 main categories according to their solubility at pH 4.6 and 20°C (Fox, 2001); the insoluble

263 protein fraction that precipitates are caseins (αS1-casein, αS2-casein, β-casein, κ-casein),

264 while the soluble fraction constitutes the serum characterized by whey proteins (β-lactoglobulin,

265 α-lactalbumin, bovine lactoferrin, bovine serum albumin, and immunoglobulins), corresponding

266 to 80% and 20%, respectively. All the caseins present genetic polymorphisms, which account

267　for the high heterogeneity of such proteins with several identified variants. The latter are

268　characterized by single amino acid substitution, by deletions of peptides fragments and by

269　occurrence of post-translational modifications, such as glycosylation, phosphorylation, or partial

270　hydrolysis, which may affect their specific allergenicity (Fox, 2001; Wal, 2001) and peptide utility

271　in LC-MS assays. The main allergenic proteins of whey are the globular proteins β-lactoglobulin

272　and α-lactalbumin, which are not phosphorylated and contain intramolecular disulfide bonds

273　that stabilize their structure.

274　Besides the direct consumption of raw milk, a wide range of derived food products can be

275　manufactured from either raw milk or its condensed forms, such as cheese, butter products,

276　cream products, powdered formulations; in addition, casein and whey protein fractions can also

277　be used independently as ingredients in several commodities, including cheese, bakery

278　products and glues. Therefore, the design of robust methods of analysis tailored to detect

279　potential milk contaminations requires the selection of specific markers tracking independently

280　the two main milk fractions, caseins and whey.

281　The critical evaluation of data displayed in Fig. 2, pointed out that milk has not been only the

282　most investigated source of food allergens, but also the allergenic food where the highest

283　consensus was achieved in the selection of signature peptides. The milk proteotypic peptides

284　FFV, YLG and HQG (see Table S2 for peptides full sequences) belong to the αS1-casein (Bos

285　d 9), which is considered the most abundant protein in milk. They have been validated in many

286　independent studies for targeted analysis, including several kind of matrices, such as liquid

287　matrices (wine, ice cream), bakery products (bread, cookie, muffin), high fat and polyphenol

288　matrices (chocolate), acidic formulation (tomato sauce), high protein matrices (meat products)

289 etc. In addition, a few signature peptides were also reported for β-casein (Bos d 11), αS2-casein

290 (Bos d 10), and κ-casein (Bos d 12), in the perspective to monitor two different allergenic

291 proteins tracking for the same ingredient. As for whey proteins, main reliable signature peptides

292 belonged to the β-lactoglobulin (β-LG, MW=18.3 kDa), which is the most abundant protein in

293 whey. This protein belongs to the lipocalin superfamily and is one of the best characterized

294 lipid-binding proteins. It possesses three disulfide bridges and is present in two main variants

295 with punctual mutations. β-LG occurs naturally as a mixture of monomers and dimers, but the

296 proportion of monomers increases after heating to 70°C (Monaci et al., 2006). TPE, LVL and

297 LSF were highlighted as the most common signature peptides, already validated into incurred

298 matrices. Noteworthy, the LSF peptide corresponding to the protein C-terminal, despite

299 featuring a cysteine residue, has been selected for its robustness. In addition, some peptide

300 markers were proposed also for α-lactalbumin (VGINYWLAHK, FLDDDLTDDIMCVK,

301 DDQNPHSSNICNISCDK, LDQWLCEK, CEVFR), but none of them was validated in either

302 chocolate-based matrix or thermally processed incurred matrices; therefore, they were

303 excluded from the list of candidate markers at the first selection stage. The specificity of all

304 reported milk peptide markers was confirmed, and some of the sequences showed

305 conservation among proteins from other farmed species, still with no specificity issues with

306 other allergenic food ingredients.

307 Eggs are, together with milk the most common allergenic foods affecting European children.

308 Hen's eggs are very common in the human diet as inexpensive source of high quality proteins

309 and they represent a key ingredient in many food products, given its nutritional value and unique

310 functional properties, such as emulsifying, foaming and gelling. Several egg-based formulations

311 have been involved in food manufacturing, such as whole egg, egg white, lysozyme isolate,

14

312　egg yolk, all containing a different degree of allergenic potential (Benedé et al., 2015). The main

313　egg white allergenic proteins are ovomucoid (Gal d 1), ovalbumin (Gal d 2), lysozyme (Gal d 4)

314　and ovotransferrin (Gal d 3), all listed in Table 1, with specific mention to the known post-

315　translational modifications. Yolk, globally, it is significantly less allergenic than albumen,

316　containing two main allergenic proteins, namely Gal d 5 and Gal d 6, both belonging to the

317　livetins fraction. The latter accounts for about 30% of the yolk proteins, however, relative

318　abundance of Gal d 5 and Gal d 6, is hard to define. Gal d 5 is the main α-livetin, whereas Gal

319　d 6 (protein YPG 42) derives from major β-livetins as C-terminal fragment of vitellogenin 1.

320　Ovomucoid is one of the most important allergen in the egg white, however it was rarely

321　selected as reporter for egg contamination, in MS based methods. Indeed we retrieved

322　information about a single peptide marker (AFNPVCGTDGVTYDNECLLCAHK, Montowska &

323　Fornal, 2018) that was never validated in either chocolate-based matrix or thermally processed

324　incurred matrices, thus excluded from the list of candidate markers at the first stage of the

325　literature review. Ovalbumin, as most abundant protein in egg white, is the most common

326　protein used as marker forhen's egg contamination. It is a phosphoglycoprotein with a

327　molecular mass of 45 kDa belonging to the serpin superfamily, with an intra-molecular disulfide

328　bond between C residues in positions 74 and 121. Noteworthy, besides the most cited peptide

329　GGL, which was monitored in several independent investigations, eight of them including hard-

330　to analyze matrices, all the other peptides LTE, ELI and YPI contain amino acids prone to

331　modifications, methionine, asparagine (NG motif) and cysteine, respectively (Fig. 2). In

332　particular, the peptide LTE was alternatively detected either in its native or in the M-oxidized

333　forms depending on the matrix (De Angelis et al., 2017b; Pilolli et al., 2017a; Pilolli et al., 2017b;

334　Pilolli et al., 2018). The peptide ELI contains the dipeptide NG prone to spontaneous asparagine

335   deamidation. The peptide YPI contains the cysteine residue in position 121 involved in the

336   disulfide bridge. Alternative ovalbumin signature peptides have been also reported, such as

337   HIA and ISQ all displayed in Fig. 2.

338   Similarly to milk fractions, different egg-derived formulations used in food manufacturing require

339   specific markers to be selected for each part of the allergenic ingredient that could eventually

340   contaminate food commodities. Thus, in addition to egg white markers, it is important to select

341   specific markers of egg lysozyme, industrially applied as protein isolate for its antimicrobial

342   properties, and egg yolk. Only a limited percentage of the collected studies considered this

343   point. As for direct lysozyme detection, we disclosed three main markers, FES, NTD and GTD,

344   which have been validated at least once into incurred baked products, such as cookies and

345   muffins. The egg yolk has been rarely investigated so far, indeed only six studies were collected

346   detecting specific egg yolk proteins, four of them developing a targeted approach (Planque et

347   al,. 2016; Planque et al., 2017b; New et al., 2018; Planque et al., 2019), and two presenting

348   only the identification of potential markers by HR-MS discovery analysis (Lee & Kim, 2010;

349   Gavage et al., 2019). Luckily, most of them already tested a wide diversity of food matrices,

350   with different complexity degree, such as cookie, bread, chocolate, salad dressing, spices, ham

351   (see Table S2 for further details), also taking into consideration the ingredient inclusion before

352   food processing. Noteworthy, Gavage et al., 2019 proposed a systematic investigation of

353   potential egg markers (selected for both egg white and egg yolk), in raw and processed egg-

354   based commodities, promoting the use of a dual enzyme digestion protocol, which benefits

355   from the use of the lysyl endopeptidase in addition to trypsin to provide the same site-specific

356   cleavage with higher proteolytic yield. Focusing on the four targeted investigations (Planque et

357   al., 2016; Planque et al., 2017b; New et al., 2018; Planque et al., 2019), we disclosed that none

of the reported signature peptides traced directly the allergenic protein YGP42, which is the

carboxy-terminal portion (284 AAs) cleaved from the primary translation product of vitellogenin-

1. Indeed, the peptides YLL and ALL both belong to the amine terminal chain of vitellogenin-1,

known as lipovitellin-1. Three additional peptides included in the allergenic protein YGP42 were

proposed by Gavage et al., 2019, TVI, NVN, ATA, by mean of the dual enzyme digestion

protocol, which could be considered for further assessment in the development of targeted

investigations.

The sequence specificity control by BLAST proved that none of the reported peptide markers

was common to other allergenic ingredients, however some of them (especially lysozyme and

vitellogenin-2 peptides) presented conserved sequence in proteins from bacteria and from other

animal species, unlikely to occur in food commodities.

Further non-allergenic proteins were also proposed to trace for egg yolk contamination, such

as the vitellogenin-2, the apolipoprotein B and the apovitellenin-1.

2.2. Proteins and peptide markers for legumes: peanut and soybean.

Peanut and soybean are both members of the *Fabaceae* or *Leguminosae* family. Legumes

represent 27% of the primary crop production worldwide, highly diffused in the human diet as

excellent sources of proteins, water soluble fibers, numerous micronutrients, and

phytochemicals (Smýkal et al., 2015). Peanut is the best characterized legume in regard to its

allergenic content, having sixteen proteins, mainly seed storage proteins, classified as allergens

with specific isoforms identified and listed in Table 1. As for soybean, the number of

characterized allergens is relatively limited, with eight allergens officially registered by the

WHO/IUIS Allergen Nomenclature Subcommittee (Allergen Nomenclature, accessed in

September 2018). Legume allergens can be grouped into a restricted number of families and superfamilies, and the most relevant allergens are seed storage proteins belonging to the cupin and prolamin superfamilies. The cupin superfamily share a conserved β-barrel structural core domain, which can be single or double; the seed storage proteins 7S and 11S globulins are both bicupins. The prolamin superfamily includes proteins with conserved cysteine residues at specific positions, sharing a common tridimensional structure with four α-helices stabilized by disulfide bonds. This superfamily includes the 2S albumins and non-specific lipid transfer proteins (ns-LTPs). In addition, minor allergens belong to the profilins and Bet v (*Betula verrucosa*) 1-like superfamilies, which are associated with pollen allergy, the defensins, involved in plant defense against pathogens, and the oleosins, structural proteins of intracellular lipid storage organelles.

Current methods for MS-based peanut detection mainly target Ara h 1 (7S-globulin, vicilin-type), Ara h 2 (2S albumin, conglutin) and Ara h 3 (11S globulin, glycinin) proteins. All peptides reported for these three proteins are highly specific, being conserved only among specific peanut related taxonomies (Arachis ipaensis (130454), Arachis hypogaea (3818), Arachis duranensis (130453)). Ara h 1 is a glycoprotein that makes up 12-16% of the total protein content and forms stable trimers held by non-covalent interactions (Palladino & Breiteneder, 2018). Two isoforms has been identified, encoded by two different clones 41 B (Ara h 1.0101) and P17 (Ara h 1 clone P 17), with a sequence identity higher than 90%. Both proteins have an N-terminal 25 amino acid residue signal peptide and a single glycosylation site. The signal peptide drives the nascent protein to the storage vacuole where both the signal itself and a specific portion of the N-terminal is cleaved–off to yield the mature Ara h 1 found in peanut (Palladino & Breiteneder, 2018). All the Ara h 1 marker peptides reported so far are shared

403 between the two isoforms, and none of them include the glycosylation site (asparagine residue

404 at position 521 of clone 41B and position 516 of clone P17, see Fig. S1 for details). DLA, VLL

405 and GTG are the most frequently selected, already tested in several incurred and hard-to

406 analyze matrices.

407 Ara h 2 represents the 5.9-9.2% of the total protein content and can be expressed from different

408 genes into two isoforms Ara h 2.01, and Ara h 2.02, the latter differing for the insertion of a 12

409 amino acid motif at position 75 in comparison to the Ara h 2.01. In addition, the Allergome

410 database further distinguishes the two isoforms in four variants (Ara h 2.0101, Ara h 2.0102,

411 Ara h 2.0201, Ara h 2.0202), which, besides the signal sequence, differ for a single amino acid

412 mismatch (E vs D) close to the C-terminal portion of the full sequence (position 163 in the Ara

413 h 2.0201 sequence). Furthermore, Ara h 2 undergoes proteolytic processing by peanut

414 proteases resulting in the removal of the C-terminal dipeptide RY. Consequently, the Ara h 2

415 can be found as a mixture of all the four variants as well as their slightly truncated forms. Four

416 intra-molecular disulfide bonds hold the tridimensional structure of the Ara h 2 and three proline

417 residues are present in their hydroxylated form (see Fig. S2). Interestingly, none of the markers

418 reported in Fig. 3 for Ara h 2 contain the 4-hydroxylproline residues, however all of them contain

419 cysteine residues directly involved in the formation of the four disulfide bridges. In addition, the

420 CMC also contained two methionine residues, susceptible to oxidation. The first three peptides,

421 namely CCN, NLP and CMC are shared among all the isoforms (see Fig. S2), whereas the

422 peptide CDL includes the position affected by single amino acid substitution, thus both the

423 variants CDL*E*VESGGR (reported peptide) and CDL*D*VESGGR (potential alternative

424 sequence) can be found in nature with unpredictable proportions.

425 Ara h 3 is a glycinin-like protein (11S) with a molecular mass of 60 kDa for the monomer and

426 occurs in peanuts as hexamer of 360 kDa (Palladino & Breiteneder, 2018). The monomer is

427 post-translationally cleaved in 43 kDa acidic and 28 kDa basic subunits, covalently linked by a

428 single disulphide bond. Ara h 3 and Ara h 4, initially considered as different allergenic proteins,

429 have been identified as variants of the same gene, thus they were renamed as isoforms Ara h

430 3.01 and Ara h 3.02. These are the only two isoforms officially listed by the WHO/IUIS Allergen

431 Nomenclature Subcommittee, whereas further Uniprot entries are classified as Ara h 3 in the

432 Allergome database. A genomic clone encoding Ara h 3 (AF10854) (Viquez et al., 2004) was

433 identified as having four exons. The deduced protein (538 AA in length) showed 93% and 91%

434 identity with the Ara h 3 (isoform Ara h 3.01) and Ara h 4 (isoform Ara h 3.02) (Palladino &

435 Breiteneder, 2018), respectively. Furthermore, an additional isoform named iso-Ara h 3 was

436 reported as sharing only 70-85% of identity with previously cited sequences (Boldt et al., 2005).

437 Several peptide markers tracing for Ara h 3 protein were reported so far (see Fig. 3), however,

438 surprisingly for this particular case, the criterion of sequence sharing among known isoforms

439 has not been valued properly in the current literature (see Fig. S3). Indeed, some of proposed

440 markers were not conserved among protein variants, and according to common sense, they

441 should have been excluded, for their limited representativeness. For example, the peptide SPD,

442 which is cited in 13 investigations, is encrypted only in the Ara h 3.01 isoform (and in the

443 genomic clone). Quite surprising was also the selection of the peptide TAN_E_LNLLILR in two

444 very recent papers (Planque et al., 2017a; Planque et al., 2017b), which was not encoded by

445 the officially recognized isoforms Ara h 3.01 and Ara h 3.02, but only by the Ara h 3 genomic

446 clone. Noteworthy, the peptide LNA is the only marker fully conserved across the isoforms

447   while, the peptides RPF, QQP are common only to the two main isoforms Ara h 3.01 and Ara

448   h 3.02.

449   As previously stated, soybean is another edible legume belonging to the *Fabaceae* family,

450   widely consumed worldwide for its high protein content (approximately 38-40 %). It finds also

451   wide applicability as ingredient in meat/poultry products, bakery and pastry products, dairy

452   products, and edible spreads, as well as additive for a variety of pharmaceutical and industrial

453   applications (Verma et al., 2013). As such, the number of occurring soy based formulations is

454   very high, ranging from raw to highly processed and/or hydrolyzed commodities and this makes

455   the absolute quantification of soy contamination in complex matrices quite challenging.

456   The WHO/IUIS recognized eight soybean proteins as official allergens, including hydrophobic

457   protein (Gly m 1), defensin (Gly m 2), soy profilin (Gly m 3), pathogenesis-related protein (Gly

458   m 4), β-conglycinin (Gly m 5, vicilin, 7S globulin), glycinin (Gly m 6, legumin, 11S globulin),

459   seed biotinylated proteins (Gly m 7) and the 2S albumin protein group (Gly m 8). The major

460   storage proteins, β-conglycinin and glycinin, both belonging to the cupin superfamily, represent

461   alone the 70–80% of the total seed globulin fraction (De Angelis et al., 2017a). β-Conglycinin

462   is 7S globulin glycoprotein containing 5% of carbohydrate moieties and occur as trimers with

463   molecular masses of approximately 180 kDa. Therefore, the Gly m 5 allergen name refers to

464   the complex of three β-conglycinin subunits that were not characterized at the time the name

465   was assigned (Pomés et al., 2018). The three subunits were first described by Holzhauser et

466   al. (2009) labeled as α (67 kDa), α' (71 kDa) and β (50 kDa), and were assessed to combine at

467   different ratios to form the multimeric complexes. The α and α′ subunits are approximately 82%

468   identical, whereas the β-subunit is only 76% identical to the two α subunits. The individual

469   subunits were named as isoallergens Gly m 5.01, Gly m 5.02 and Gly m 5.03 in 2009, with

470   minor variants (e.g. Gly m 5.0301 and Gly m 5.0302 presenting two amino acid mismatches

471   L/F at positions 16 and 198). Interestingly, the *in-silico* simulation of tryptic digestion showed

472   that the isoallergens Gly m 5.01, Gly m 5.02 and Gly m 5.03 do not share any tryptic peptide

473   with minimum length of 7 AA, which could be selected as a common marker for the β-

474   conglycinin subunits. This issue makes the absolute quantification of soybean allergen by

475   synthetic peptide analysis quite challenging since the required conversion factors should base

476   on the knowledge of the specific combination of α, α′ and β-subunits occurring in the sample

477   under investigation. As expected by *in-silico* prediction, none of the reported peptide markers

478   for Gly m 5 displayed in Fig. 3 were conserved among the three isoallergens (see Fig. S4).

479   Interestingly, an alternative solution, in this case, would be to select unique markers for each

480   subunit in order to disclosethe specific combination of α, α′ and β-monomers within the trimeric

481   complex. In this frame, the peptide LIT would uniquely trace for Gly m 5.01, the peptides QQQ

482   and DSYNLQ would uniquely trace for Gly m 5.02, and the peptide DSYNLH would uniquely

483   trace for Gly m 5.03 (noteworthy, the latter is conserved between the two variants Gly m 5.0301

484   and Gly m 5.0302). Attention should be paid to the peptide AIV, even if theoretically unique for

485   Gly m 5.01, since an homologous peptide sequence is expressed in both variants of Gly m

486   5.03, with a single amino acid substitution L/I at position P19, which could not be discriminated

487   by SRM acquisition. None of the aforementioned peptides contain amino acid residues with

488   naturalpost-translational modification.

489   In most soybean varieties, the glycinin accounts for over 50% of seed storage proteins

490   representing the predominant protein fraction (De Angelis et al., 2017a). Glycinin has a complex

491   hexameric structure with a molecular weight ranging between 320 to 360 kDa. At least five

492   genes encode the monomeric subunits (Nielsen et al., 1989) and the Gly m 6 allergen can be

493   an arrangement of these five proteins, which have individual IgE binding properties (Holzhauser

494   et al., 2009). Each monomer consists of two specific polypeptide chains, one acidic (40 kDa)

495   (A) and one basic (20 kDa) (B), linked together by disulfide bonds. Each monomer can be one

496   of five subunits, namely glycinin G1 ($A_{1a}B_{1b}$), glycinin G2 (A_2B_{1a}), glycinin G3 ($A_{1b}B_{1a}$), glycinin

497   G4 ($A_5A_4B_3$) and glycinin (A_3B_4) (Wang et al., 2014), being renamed as five isoallergens Gly m

498   6.0101, Gly m 6.0201, Gly m 6.0301, Gly m 6.0401 and Gly m 6.0501, respectively, in 2009.

499   Similarly to what was already discussed for the quantification of Gly m 5 allergen, the

500   unpredictable composition of the Gly m 6 hexameric complex makes general conversion factors

501   a hard task to achieve. The peptide marker selection carried out so far in current literature

502   completely neglected the different sequence expression in known isoallergens. Indeed, the

503   most frequently used markers, VFD, VLI, and EAF, are expressed uniquely in Gly m 6.0101,

504   whereas the peptide SQS, cited in 8 independent investigations, is conserved across the three

505   isoforms Gly m 6.0101, Gly m 6.0201, Gly m 6.0301, and the peptide IST (cited in 7 independent

506   investigations) is common to the isoforms Gly m 6.0401 and Gly m 6.0501 (see Fig. S5).

507   ### 2.3. Proteins and peptide markers for tree nuts: hazelnut and almond

508   Hazelnut and almond allergens were selected in this project as most relevant members of the

509   tree nut category according to the European legislation. Similar to peanut and soybean, the

510   main proteins involved in tree nut allergy belong to cupin (legumins-11S globulin and vicilin-7S

511   globulin) and prolamin (conglutin-2S albumin and ns-LTPs) superfamilies. Several hazelnut

512   allergens have been identified and are already partly applied in component resolved diagnosis

513   (Geiselhart et al., 2018). However, only partial information have been collected so far about the

514 occurrence of potential allergen isoforms and their complete sequencing, even if their existence

515 is likely, given the high genetic diversity of plant organisms. Even the limited information

516 available for almond allergens, the provided following discussion about the peptide marker

517 selection was driven to the best of the current knowledge for these two allergens, and relevant

518 homologous plant proteins.

519 WHO/IUIS database included eight hazelnut allergens all listed in Table 1. According to the

520 EuroPrevall population-based survey, including allergic subjects from twelve European cities,

521 the predominance of specific hazelnut allergens is correlated to the geographical area

522 (Pastorello et al., 2002), as well as to the age of the sensitized individuals (Verweij et al., 2012).

523 In the Mediterranean areas, hazelnut allergy is mainly linked to Cor a 8, Cor a 14, Cor a 9, Cor

524 a 11 (Schocker et al., 2004; Blanc et al., 2015). Cor a 8 is a nsLTP type 1, as mature protein,

525 consisting of a unique polypeptide chain of 92 amino acid residues with eight strictly conserved

526 cysteines forming four intramolecular disulfide bridges. Cor a 14 (2S albumin) is a heterodimeric

527 protein with characteristic structural features, including four α-helices and eight cysteine

528 residues engaged in four disulfide bridges (Pfeifer et al., 2015). The precursor protein is post-

529 translationally cleaved into a large and a small subunits linked by disulfide bonds. The N-

530 terminal glutamine shows the cyclization to pyroglutamic acid and the small subunit is variably

531 truncated at the C-terminal, leading to a high micro-heterogeneity (Pfeifer et al., 2015). Cor a

532 9, as 11S legumin, exhibits some homology with other tree nut allergens, such as cashew Ana

533 o 2, peanut Ara h 3 and the soybean glycinin (Gly m 6). The monomer consists of two

534 polypeptide chains (acidic and basic chains), cleaved post-translationally by site specific

535 endopeptidases, and linked together by a single intermolecular disulfide bond. The WHO/IUIS

536 recorded a single isoform of the Cor a 9 protein corresponding to the Uniprot accession

537 Q8W1C2 and labelled as Cor a 9.0101, first identified in 2002 as IgE-binding protein by 2D-

538 Western blotting, using sera of hazelnut allergic patients (Beyer et al., 2002). In 2013, a new

539 isoform of the same protein, with a 514 AA full-length sequence, was registered (Uniprot

540 accession A0A0A0P7E3) sharing 96.5% homology (Grishina et al., 2013). This isoform has

541 been added to the Allergome database with the general label of Cor a 9. Additionally, at least

542 another 55 kDa protein, with high IgE-reactivity only in the alkaline chain (20.7 kDa), has been

543 identified by Nitride et al. (2013). In this investigation, shotgun proteomics allowed de novo

544 sequencing of six peptides of the basic chain and their assignment to 11S protein isoform

545 sharing high homology with 11S globulin-like proteins from several plant organisms, including

546 the canonical hazelnut one. Starting from this background, we reviewed the current literature

547 tracing the Cor a 9 selected markers. The peptide WLQ should be excluded by any further

548 discussion since conserved sequence among different tree nut species (hazelnut, pistachio,

549 pecan, and walnut). Besides this, we assessed that all the proposed signature peptides shared

550 100% identity between the official Cor a 9.0101 (Q8W1C2) sequence and its registered isoform

551 A0A0A0P7E3 (see Fig. S6). However, considering the further potential isoform only partially

552 sequenced by Nitride et al. (2013), three out of the seven reported peptides might be excluded,

553 as non-conserved regions and, noteworthy, this exclusion would involve two of the most cited

554 peptides INT and ALP. Therefore, a wider characterization of most common isoform occurring

555 by natural genetic diversity is urgently needed. In addition, the peptide VQV contains the NG

556 motif, which makes the asparagine residues highly susceptible to deamidation.

557 Cor a 11 is a 7S globulin like protein, featuring a trimeric structure of MW about 150/190kDa

558 with subunits of about 50kDa. The WHO/IUIS lists a single isoform labelled as Cor a 11.0101,

559 corresponding to the Uniprot entry Q8S4P9, which includes a single glycosylation site occurring

at the asparagine residue. Three peptide markers have been reported for this protein, namely LLS, AFS, and ELA, and none of them includes amino acid residues prone to modifications.

As for almond, six identified almond allergens, Pru du 3 (ns-LTP), Pru du 4 (profilin), Pru du 5 (60 S ribosomal protein), and Pru du 6 (legumin), are included in the WHO-IUIS allergen list. The specificity in the analytical detection of this allergen is particularly challenging due to both a limited availability of full-length sequences in official database and due to the high homology of almond proteins (taxonomy *Prunus dulcis*) within the *Prunus* taxonomy genus, mainly *Prunus persica* (peach) (Inman et al., 2018). For example, the almond allergen Pru du 3.0101 displays 99% amino acid sequence identity to a nsLTP from peach (UniProtKB M5W0S9, not listed in the IUIS) and 57% to two other peach nsLTP isoforms, Pru p 3.0101 and Pru p 3. 0102. Two native isoforms have been identified for the almond profilin (Pru du 4), but its relevance is mainly restricted to pollen-associated almond allergy. Pru du 5 has been identified as an 11.4 kDa 60S ribosomal protein involved in protein biosynthesis (Abolhassani & Roux, 2009); nevertheless, the clinical relevance of this allergen is unknown and requires further research. Finally, Pru du 6, a legumin-like protein, also called amandin, is the first and best biophysically characterized allergen in almond to date (Geiselhart et al., 2018). Two cDNA clones encoding almond legumins named prunin-1 (61.0 kDa) and prunin-2 (55.9 kDa) have been isolated, both included in the WHO/IUIS database as Pru du 6.0101 and Pru du 6.0201. By *in-silico* simulation of the enzymatic digestion with trypsin specific cleavage, it was assessed that no tryptic peptide with minimum length of 7 AA presents full identity between the two isoforms Pru du 6.0101 and Pru du 6.0201 (see Fig. S7). In addition, Allergome database also includes two variants of Pru du 6.0101 and Pru du 6.0201, registered with the Uniprot entries Q43607 and Q43608, respectively, which share 99% identity with the relevant official allergen isoform, featuring only

583 few amino acid substitutions (see Fig. S7). In Fig. 3, we presented the signature peptides

584 reported so far for almond detection by SRM analysis all belonging to the Pru du 6 protein,

585 either to the Pru du 6.0101 or to the Pru du 6.0201 isoallergens. Noteworthy, most of the

586 selected peptides highlighted with asterisks presented specificity issues, since the BLAST

587 search against non-redundant protein database showed 100% sequence identity with

588 homologous proteins from *Prunus persica*, *Prunus mume* or *Prunus avium,* (see Fig. S8 and

589 S9). In light of this, we kept in the list reported in Fig. 3 also the two markers QQG both VQV,

590 even if they were not validated in incurred or chocolate based matrices, as only candidates,

591 together with TDE, which feature uniqueness for *Prunus dulcis* species. Interestingly, none of

592 the reported peptides included cysteine residues involved in the disulfide bridges, whereas both

593 the unique peptides VQV and TDE contain the motif NG that make them susceptible to

594 spontaneous deamidation. In addition, the sequences VQVVNENGDPIL<u>D</u>DEVR and

595 ALPDEVLQ<u>N</u>AFR, belonging to the Pru du 6.0201 isoform were not conserved in the alternative

596 allergen variant registered as Q43608 in Uniprot database, both presenting single amino acid

597 substitutions (D/N and N/T, respectively) and the peptide variant VQVVNENGDPIL<u>N</u>DEVR is

598 shared with *Prunus persica* (see sequence alignment in Fig. S9).

599 **2.4. List of candidate markers agreed by ThRAll partners**

600 The thorough discussion presented in the previous section allowed the selection of a restricted

601 list of candidate signature peptides that can be fruitfully employed for allergen detection in

602 highly complex and processed foods with a high-confidence on identification accuracy. This list

603 representing the consensus table of the consortium is reported in Table 2 and will be further

604 validated for its reliability in the model matrices selected within the ThRAll project by discovery

experiments by untargeted HR-MS/MS analysis. As already mentioned, the preliminary marker selection here performed by means of the critical evaluation of current literature represents only the first step of the dual approach devised for the ThRAll project. The experimental validation of the candidate markers by MS/MS analysis is a mandatory step in method development because several empirical features, such as matrix composition, extraction/digestion conditions, and instrumental parameters might affect both the specificity and sensitivity of the final analytical method. Interferences in the chromatographic analysis from complex food background and information about processing effects on the ingredient itself or finished food products should be traced to avoid false-negative and false-positives. Such information cannot be disclosed merely by any *in-silico* based approach.

The Table 2 was populated with all the information about sequence sharing among isoforms/variants per each discussed peptide.

As for the protein homology check among different taxonomies carried out on plant-derived allergens, we concluded that according to the current knowledge of full-length protein sequences, none of the signature peptides selected for plant-derived allergens (Table 2) is affected by homology issues. To the best of the current knowledge, the refined list of peptides provided very high detection specificity. For sake of clarity in Table S3, we reported a short list of peptide markers identified that might experience this kind of interference from other homologous proteins, disclosing both the targeted and the interfering precursor ions, as well as the shared transitions.

As for Pru du 6 peptides, with high sequence identity with homologous proteins from *Prunus persica*, *Prunus mume* and *Prunus avium*, only two peptide markers were kept, including the

627 unique peptide TDE, even if encrypting the NG motif, since it can still be useful for confirmative

628 purposes.

629 Finally, in Table 2 we also collected information about the reported precursor ions and the

630 transitions monitored. Both the preferential ionization state of the target peptide and its typical

631 fragmentation pattern depend not only on the specific amino acid sequence, but also on

632 instrumental parameters such as the ionization source, the MS platform, and the activation

633 mode; therefore, after proper marker selection, the ionization and fragmentation features need

634 to be assessed on the specific MS platform involved in the study. From a general point of view,

635 we highlighted that most of the selected signature peptides were detected as double charged

636 ions, which indeed are very stable and reproducible both in their ionization yield and in the

637 fragmentation pattern, whereas an heterogeneous scenario was disclosed as for transitions

638 selection. Either two or three transitions per peptide were acquired, in some cases the two

639 monitored fragments are simply two different ionization states of the same transition (y_n^+ and

640 y_n^{++}). This behavior is typical of proline containing peptides that upon collisional induced

641 fragmentation preferentially break the peptide bond at the N-terminal proline site, releasing the

642 same transition as both single and double charged ions (Ma & Johnson, 2012). This transition

643 is usually very sensitive, but the ratio (y_n^+/y_n^{++}) between the two ions is not fixed, and may

644 change upon different instrumental conditions. Additionally, retrieving the SRM instrumental

645 details for the reported papers we also disclosed that in some investigations quite small

646 fragments have been selected, namely y_1^+, b_2^+, a_2^+, notwithstanding their low selectivity and

647 susceptibility to interference from the background.

648 In light of the harmonization of detection method, and given the heterogeneous picture retrieved

649 for instrumental set up, we will select at least three transitions for each marker, preferably

650 belonging to the y_n and b_n series. Indeed, y- and b- ion series together would allow read out of

651 the full peptide sequence from the MS/MS spectrum in the two directions starting from the C-

652 and the N-terminus, respectively (Ma & Johnson, 2012); in addition, the two series together

653 would account for most of the MS/MS spectrum ion intensity in both HCD and CID

654 fragmentations (Michalski et al., 2012), with very high stability of the y-type transitions.

655 Noteworthy, the monitored fragments should actually correspond to different transitions with a

656 minimum fragment length of 3 amino acids, given the minimum length for peptide markers of 7

657 amino acids, in order to guarantee the maximum specificity and sequence coverage to the final

658 developed method.

659 Future development of this work, will be the experimental validation of the reported preliminary

660 list of candidate markers. In particular, discovery high resolution MS/MS analysis will be carried

661 out directly on the two matrices, namely chocolate bar and broth powder as model hard-to

662 analyze food matrices, produced in a food pilot plant and incurred at the high contamination

663 level. The experimental list of identified peptides belonging to the allergenic ingredients will be

664 compared with the literature based selection either confirming or proposing new options for the

665 SRM set up.

666 **3. Conclusions**

667 In this work we described the results of our comprehensive evaluation of the existing literature

668 on food allergen detection by MS based methods, concerning the selection of peptide markers.

669 We collected all information about the six allergens selected for the ThRAll project, namely

670 cow's milk, hen's egg, peanut, soybean, hazelnut and almond. The selected peptides were

671 critically discussed according to specific considerations and recommended criteria for marker

672 selection that we promote in perspective harmonization of analytical methods development.

673 The in-depth analysis of current knowledge allowed the selection of a restricted list of candidate

674 signature peptides, already validated in several independent analyses, which could be

675 monitored with high confidence. This preliminary list will be validated in the two model matrices

676 selected within the ThRAll project, in a separate investigation for its final confirmation by specific

677 discovery experiments with untargeted HR-MS/MS analysis.

688 **Ethical approval**

689 This article does not contain any studies with human participants or animals performed by any

690 of the Authors.

691 **Competing interests**

692 The Authors declare they have no competing interest.

FIGURE AND TABLE CAPTIONS

Figure 1. Analytical workflow describing step-by-step the literature review and the critical evaluation of the proposed peptide markers in order to compile a final list of highly reliable signature peptides to be tested for the ThRAll project.

Figure 2. Citation overview of the peptides markers used for milk and egg allergens, which where validated at least once in hard-to-analyze matrices, highlighting the kind of investigation (targeted approach or discovery only approach) and the presence of amino acids prone to modifications. *peptide without C, M, E, N, or Q amino acid.

Figure 3. Citation overview of the peptides markers used for peanut, soybean, hazelnut and almond, which where validated at least once in hard-to-analyze matrices, highlighting the kind of investigation (targeted approach or discovery only approach) and the presence of amino acids prone to modifications. *peptide without C, M, E, N, or Q amino acid.

Table 1. Summary of the food allergens proteins recorded for the six allergen under investigations. *Acronyms*: PTM, post-translational modifications; IUIS, International Union of Immunological Societies. *information gathered from Uniprot Database (https://www.uniprot.org/, Accessed on September 19th 2018).

Table 2. Preliminary list of signature peptides selected by the Consortium according to the analytical workflow presented in Figure 1, which have been monitored in targeted investigations with relevant information on allergens isoforms/variants, peptide specificity, and acquisition related technical details (MS platform, precursor ion and selected transitions). The presence of amino acid residues prone to modification was highlighted with a different font (bold and underlined). The transition selected as quantifier ion (when specified) was marked with an

33

715 asterisk. *Acronyms*: QqQ, triple quadrupole; IT, ion trap; QToF, quadrupole time-of flight; Q-

716 Orbitrap, quadrupole-Orbitrap, LIT, linear ion trap.

717

REFERENCES

Abolhassani, M., & Roux, K. H. (2009). cDNA cloning, expression and characterization of an allergenic 60s ribosomal protein of almond (Prunus dulcis). *Iranian Journal of Allergy, Asthma and Immunology, 8,* 77–84. https://doi.org/08.02/ijaai.7784

Allergen Nomenclature, WHO/IUIS Allergen Nomenclature Sub-Committee. (2018). http://www.allergen.org/ Accessed in July-September 2018.

Allergen Peptide Browser. (2018). https://www.allergenpeptidebrowser.org/ Accessed July 2018.

Allergome. (2018). https://www.allergome.org/ Accessed July-September 2018.

Ansari, P., Stoppacher, N., Rudolf, J., Schuhmacher, R., & Baumgartner, S. (2011). Selection of possible marker peptides for the detection of major ruminant milk proteins in food by liquid chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry, 399*, 1105–1115 https://doi.org/10.1007/s00216-010-4422-0.

Ansari, P., Stoppacher, N., & Baumgartner, S. (2012). Marker peptide selection for the determination of hazelnut by LC–MS/MS and occurrence in other nuts *Analytical and Bioanalytical Chemistry, 402*, 2607–2615. https://doi.org/10.1007/s00216-011-5218-6.

Azarnia, S., Boye, J. I., Mongeon, V., & Sabik, H. (2013). Detection of ovalbumin in egg white, whole egg and incurred pasta using LC–ESI-MS/MS and ELISA. *Food Research International, 52,* 526–534. https://doi.org/10.1016/j.foodres.2013.02.039.

Benedé, S., López-Expósito, I., Molina, E., & López-Fandiño, R. (2015). Egg proteins as allergens and the effects of the food matrix and processing. *Food and Function, 6,* 694–713. https://doi.org/10.1039/C4FO01104J.

Beyer,K., Grishina, G., Bardina, L., Grishin, A., Sampson, H., A. (2002). Identification of an 11S globulin as a major hazelnut food allergen in hazelnut induced systemic reaction. *Journal of Allergy and Clinical Immunology, 110,* 517-523. https://doi.org/10.1067/mai.2002.127434.

Bignardi, C., Elviri, L., Penna, A., Careri, M., & Mangia, A. (2010). Particle-packed column versus silica-based monolithic column for liquid chromatography–electrospray-linear ion trap-

745   tandem mass spectrometry multiallergen trace analysis in foods. *Journal of Chromatography*

746   *A, 1217,* 7579–7585. https://doi.org/10.1016/j.chroma.2010.10.037.

747   Bignardi, C., Mattarozzi, M., Penna, A., Sidoli, S., Elviri, L., Careri, M., & Mangia, A. (2013). A

748   Rapid Size-Exclusion Solid-Phase Extraction Step for Enhanced Sensitivity in Multi-Allergen

749   Determination in Dark Chocolate and Biscuits by Liquid Chromatography–Tandem Mass

750   Spectrometry. *Food Analytical Methods, 6,* 1144–1152. https://doi.org/10.1007/s12161-012-

751   9521-4

752   Blanc, F., Bernard, H., Ah-Leung, S., Przybylski-Nicaise, L., Stahl Skov, P., Purohit, A., de Blay,

753   F., Ballmer-Weber, B., Fritsche, P., Fernandez Rivas, M., Reig, I., Sinaniotis, A., Vassilopoulou,

754   E., Hoffmann-Sommergruber, K., Vieths, S., Rigby, N., Mills, E. N. C., & Adel-Patient, K. (2015).

755   Further studies on the biological activity of hazelnut allergens. *Clinical and Translational Allergy,*

756   *5,* 26. https://doi.org/10.1186/s13601-015-0066-7.

757   Boldt, A., Fortunato, D., Conti, A., Petersen, A., Ballmer-Weber, B., Lepp, U., Reese, G., &

758   Becker, W. M. (2005). Analysis of the composition of an immunoglobulin E reactive high

759   molecular weight protein complex of peanut extract containing Ara h 1 and Ara h 3/4.

760   *Proteomics, 5,* 675–686. https://doi.org/10.1002/pmic.200401150.

761   Boo, C. C., Parker, C. H., & Jackson, L. S. (2018). A Targeted LC-MS/MS Method for the

762   Simultaneous Detection and Quantitation of Egg, Milk, and Peanut Allergens in Sugar Cookies.

763   *Journal of AOAC International, 101,* 108-117. https://doi.org/10.5740/jaoacint.17-0400.

764   Bucchini, L., Guzzon, A., Poms, R., Senyuva, H. (2016). Analysis and critical comparison of

765   food allergen recalls from the European Union, USA, Canada, Hong Kong, Australia and New

766   Zealand. Food Additives and Contaminants: Part A, 33, 760-771.

767   https://doi.org/10.1080/19440049.2016.1169444.

768   Careri, M., Costa, A., Elviri, L., Lagos, J. B., Mangia, A., Terenghi, M., Cereti, A., & Garoffo, L.

769   P. (2007). Use of specific peptide biomarkers for quantitative confirmation of hidden allergenic

770   peanut proteins Ara h 2 and Ara h 3/4 for food control by liquid chromatography–tandem mass

771   spectrometry. *Analytical and Bioanalytical Chemistry, 389,* 1901–1907.

772   https://doi.org/10.1007/s00216-007-1595-2.

773 Careri, M., Elviri, L., Maffini, M., Mangia, A., Mucchino, C., & Terenghi, M. (2008). Determination
774 of peanut allergens in cereal-chocolate-based snacks: metal-tag inductively coupled plasma
775 mass spectrometry immunoassay versus liquid chromatography/electrospray ionization
776 tandem mass spectrometry. *Rapid Communications in Mass Spectrometry, 22,* 807–811.
777 https://doi.org/10.1002/rcm.3427.

778 Chassaigne, H., Nørgaard, J. V., & van Hengel, A. J. (2007). Proteomics-Based Approach To
779 Detect and Identify Major Allergens in Processed Peanuts by Capillary LC-Q-TOF (MS/MS).
780 *Journal of Agricultural and Food Chemistry, 55,* 4461–4473. https://doi.org/10.1021/jf063630e.

781 Chen, S., Yang, C.T., Downs, M. L. (2019). Detection of Six Commercially Processed Soy
782 Ingredients in an Incurred Food Matrix Using Parallel Reaction Monitoring. *Journal of Proteome*
783 *Research, 18,* 995−1005. https://pubs.acs.org/doi/10.1021/acs.jproteome.8b00689.

784 Costa, J., Ansari, P., Mafra, I., Oliveira, M. B., & Baumgartner, S. (2014). Assessing hazelnut
785 allergens by protein- and DNA-based approaches: LC-MS/MS, ELISA and real-time PCR.
786 *Analytical and Bioanalytical Chemistry,* 406, 2581–2590. https://doi.org/10.1007/s00216-014-
787 7679-x.

788 Croote, D., & Quake, S. R. (2016). Food allergen detection by mass spectrometry: the role of
789 systems biology. *NPJ Systems Biology Applications, 2,* 16022.
790 https://doi.org/10.1038/npjsba.2016.22.

791 Cryar, A., Pritchard, C., Burkitt, W., Walker, M., O'Connor, G., Burns, D. T., & Quaglia, M.
792 (2013). Towards absolute quantification of allergenic proteins in food--lysozyme in wine as a
793 model system for metrologically traceable mass spectrometric methods and certified reference
794 materials. *Journal of AOAC International, 96,* 1350-1361. https://doi.org/10.5740/jaoacint.12-
795 438.

796 De Angelis, E., Pilolli, R., Bavaro, S. L., & Monaci, L. (2017a). Insight into the gastro-duodenal
797 digestion resistance of soybean proteins and potential implications for residual immunogenicity.
798 *Food and Function, 8,* 1599-1610. https://doi.org/10.1039/c6fo01788f.

799 De Angelis, E., Pilolli, R., & Monaci, L. (2017b). Coupling SPE on-line pre-enrichment with
800 HPLC and MS/MS for the sensitive detection of multiple allergens in wine. *Food Control, 73,*
801 814-820. https://doi.org/10.1016/j.foodcont.2016.09.031.

802 Downs, M. L., & Johnson, P. (2018). Target selection strategies for LC-MS/MS food allergen
803 methods. *Journal of AOAC International, 101*, 146-151. https://doi.org/10.5740/jaoacint.17-
804 0404.

805 European Commission, COMMISSION NOTICE of 13.7.2017 relating to the provision of
806 information on substances or products causing allergies or intolerances as listed in Annex II of
807 Regulation (EU) No 1169/2011 on the provision of food information to consumers.

808 Eyers, C. E., Lawless, C., Wedge, D. C., Lau, K. W., Gaskell, S. J., & Hubbard, S. J. (2011).
809 CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using
810 consensus machine learning approaches. *Molecular and Cellular Proteomics, 10,*
811 M110.003384. https://doi.org/10.1074/mcp.M110.003384.

812 FAOSTAT (2017), Food and Agriculture Organization of the United Nations,
813 http://www.fao.org/faostat/en/#data/QL, accessed 9 January 2019.

814 Figeys, D., van Oostveen, I., Ducret, A., & Aebersold, R. (1996). Protein identification by
815 capillary zone electrophoresis/microelectrospray ionization-tandem mass spectrometry at the
816 subfemtomole level. *Analytical Chemistry, 68,* 1822-1828. https://doi.org/10.1021/ac960191h.

817 Fox, P. (2001). Milk proteins as food ingredients. *International Journal of Dairy Technology, 54,*
818 41–55. https://doi.org/10.1046/j.1471-0307.2001.00014.x.

819 Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., Renard,
820 P., Arnould, D. H., & Gillard, N. (2019). Selection of egg peptide biomarkers in processed food
821 products by high resolution mass spectrometry. *Journal of Chromatography A, 1584,* 115-125.
822 https://doi.org/10.1016/j.chroma.2018.11.036.

823 Geiselhart, S., Hoffmann-Sommergruber, K., & Bublin, M. (2018). Tree nut allergens. *Molecular*
824 *Immunology, 100,* 71–81. https://doi.org/10.1016/j.molimm.2018.03.011.

825 Gomaa, A., & Boye, J. (2015). Simultaneous detection of multi-allergens in an incurred food
826 matrix using ELISA, multiplex flow cytometry and liquid chromatography mass spectrometry
827 (LC-MS). *Food Chemistry, 175,* 585–592. https://doi.org/10.1016/j.foodchem.2014.12.017.

828 Grishina, G., Beyer, K., Bardina, L., Sampson, H. H. (2013). Isoform of a Major hazelnut food
829 allergen Cor a 9 (11S Globulin). EMBL/GenBank/DDBJ Databases.

Groves, K., Cryar, A., Walker, M., & Quaglia, M. (2018). Assessment of Recovery of Milk Protein Allergens from Processed Food for Mass Spectrometry Quantification. *Journal of AOAC International, 101*, 152-161. https://doi.org/10.5740/jaoacint.17-0214.

Gu, S., Chen, N., Zhou, Y., Zhao, C., Zhan, L., Qu, L., Cao, C., Han, L., Deng, X., Ding, T., Song, C., & Ding, Y. (2018). A rapid solid-phase extraction combined with liquid chromatography-tandem mass spectrometry for simultaneous screening of multiple allergens in chocolates. *Food Control, 84,* 89-96. https://doi.org/10.1016/j.foodcont.2017.07.033.

Heick, J., Fischer, M., Kerbach, S., Tamm, U., & Popping, B. (2011a). Application of a liquid chromatography tandem mass spectrometry method for the simultaneous detection of seven allergenic foods in flour and bread and comparison of the method with commercially available ELISA test kits. *Journal of AOAC International, 94,* 1060-1068.

Heick, J., Fischer, M., & Pöpping, B. (2011b). First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry. *Journal of Chromatography A, 1218,* 938–943. https://doi.org/10.1016/j.chroma.2010.12.067.

Hoffmann, B., Münch, S., Schwägele, F., Neusüß, C., Jira, W. (2017). A sensitive HPLC-MS/MS screening method for the simultaneous detection of lupine, pea, and soy proteins in meat products. *Food Control, 71,* 200-209. https://doi.org/10.1016/j.foodcont.2016.06.021.

Holzhauser, T., Wackermann, O., Ballmer-Weber, B., Bindslev-Jensen, C., Scibilia, J., Perono-Garoffo, L., Utsumi, S., Poulsen, L. K., & Vieths, S. (2009). Soybean (Glycine max) allergy in Europe: gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *Journal of Allergy and Clinical Immunology, 123,* 452–458. https://doi.org/10.1016/j.jaci.2008.09.034.

Houston, N. L., Lee, D. G., Stevenson, S. E., Ladics, G. S., Bannon, G. A., McClain, S., Privalle, L., Stagg, N., Herouet-Guicheney, C., MacIntosh, S. C., & Thelen, J. J. (2011). Quantitation of Soybean Allergens Using Tandem Mass Spectrometry. *Journal of Proteome Research, 10,* 763–773. https://doi.org/10.1021/pr100913w.

Huschek, G., Bönick, J., Löwenstein, Y., Sievers, S., & Rawel, H. (2016). Quantification of allergenic plant traces in baked products by targeted proteomics using isotope marked

peptides. *LWT - Food Science and Technology, 74,* 286-293. https://doi.org/10.1016/j.lwt.2016.07.057.

Inman, S. E., Groves, K., McCullough, B., Quaglia, M., & Hopley, C. (2018). Development of a LC-MS method for the discrimination between trace level Prunus contaminants of spices. *Food Chemistry, 245,* 289–296. https://doi.org/10.1016/j.foodchem.2017.10.101.

Ji, J., Zhu, P., Pi, F., Sun, C., Sun, J., Jia, M., Ying, C., Zhang, Y., & Sun, X. (2017). Development of a liquid chromatography-tandem mass spectrometry method for simultaneous detection of the main milk allergens. *Food Control, 74,* 79-88. https://doi.org/10.1016/j.foodcont.2016.11.030.

Ke, X., Zhang, J., Lai, S., Chen, Q., Zhang, Y., Jiang, Y., Mo, W., & Ren, Y. (2017). Quantitative analysis of cow whole milk and whey powder adulteration percentage in goat and sheep milk products by isotopic dilution-ultra-high performance liquid chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry, 409,* 213–224. https://doi.org/10.1007/s00216-016-9987-9.

Korte, R., Lepski, S., & Brockmeyer, J. (2016a). Comprehensive peptide marker identification for the detection of multiple nut allergens using a non-targeted LC–HRMS multi-method. *Analytical and Bioanalytical Chemistry, 408,* 3059-3069. https://doi.org/10.1007/s00216-016-9384-4.

Korte, R., & Brockmeyer, J. (2016b). MRM3-based LC-MS multi-method for the detection and quantification of nut allergens. *Analytical and Bioanalytical Chemistry,* 408, 7845-7855. https://doi.org/10.1007/s00216-016-9888-y.

Korte, R., Oberleitner, D., Brockmeyer, J. (2019). Determination of food allergens by LC-MS: Impacts of sample preparation, food matrix, and thermal processing on peptide detectability and quantification. *Journal of Proteomics, 196,* 131–140. https://doi.org/10.1016/j.jprot.2018.11.002.

Lamberti, C., Acquadro, E., Corpillo, D., Giribaldi, M., Decastelli, L., Garino, C., Arlorio, M., Ricciardi, C., Cavallarin, L., & Giuffrida, M. G. (2016). Validation of a mass spectrometry-based method for milk traces detection in baked food. *Food Chemistry, 199*, 119–127. https://doi.org/10.1016/j.foodchem.2015.11.130.

887   Lee, J. Y. & Kim, C. J. (2010). Determination of Allergenic Egg Proteins in Food by Protein-,
888   Mass Spectrometry-, and DNA-Based Methods *Journal of AOAC International, 93*, 462-477.

889   Li, B., Gorman, E. M., Moore, K. D., Williams, T., Schowen, R. L., Topp, E. M., & Borchardt, R.
890   T. (2005). Effects of acidic N+1 residues on asparagine deamidation rates in solution and in the
891   solid state. *Journal of Pharmaceutical Sciences, 94,* 666-675.
892   https://doi.org/10.1002/jps.20263.

893   Losito, I., Introna, B., Monaci, L., Minella, S., & Palmisano, F. (2013). Development of a Method
894   for the Quantification of Caseinate Traces in Italian Commercial White Wines Based on Liquid
895   Chromatography–Electrospray Ionization–Ion Trap–Mass Spectrometry *Journal of Agricultural*
896   *and Food Chemistry,* 61, 12436−12444. https://doi.org/10.1021/jf4034909.

897   Lutter, P., Parisod, V., & Weymuth, H. (2011). Development and Validation of a Method for the
898   Quantification of Milk Proteins in Food Products Based on Liquid Chromatography with Mass
899   Spectrometric Detection. *Journal of AOAC International, 94,* 1043-1059.

900   Ma, B., & Johnson, R. (2012). De novo sequencing and homology searching. Molecular &
901   Cellular Proteomics, 11, O111.014902. https://doi.org/10.1074/mcp.O111.014902.

902   Mattarozzi, M., Milioli, M., Bignardi, C., Elviri, L., Corradini, C., & Careri, M. (2014). Investigation
903   of different sample pre-treatment routes for liquid chromatography–tandem mass spectrometry
904   detection of caseins and ovalbumin in fortified red wine. *Food Control, 38,* 82-87.
905   https://doi.org/10.1016/j.foodcont.2013.10.015.

906   Mills, E.N.C., Adel-Patient, K., Bernard, H., De Loose, M., Gillard, N., Huet, A.-C., Larré, C.,
907   Nitride, C., Pilolli, R., Tranquet, O., Van Poucke, C., Monaci, L. (2019). Detection and
908   Quantification of Allergens in Foods and Minimum Eliciting Doses in Food-Allergic Individuals
909   (ThRAll). Journal of AOA C International, 102, 1-8. https://doi.org/10.5740/jaoacint.19-0063.

910   Monaci, L., Tregoat, V., van Hengel, A. J., & Anklam, E. (2006). Milk allergens, their
911   characteristics and their detection in food: a review. *European Food Research and Technology,*
912   *223,* 149–79. https://doi.org/10.1007/s00217-005-0178-8.

913   Monaci, L., Losito, I., Palmisano, F., & Visconti, A. (2010a). Identification of allergenic milk
914   proteins markers in fined white wines by capillary liquid chromatography–electrospray

915   ionization-tandem mass spectrometry. *Journal of Chromatography A, 1217,* 4300–4305.

916   https://doi.org/10.1016/j.chroma.2010.04.035.

917   Monaci L., Nørgaard, J.V., & van Hengel, A. J. (2010b). Feasibility of a capillary LC/ESI-Q-TOF

918   MS method for the detection of milk allergens in an incurred model food matrix. *Analytical*

919   *Methods, 2,* 967–972. https://doi.org/10.1039/C0AY00151A.

920   Monaci, L., Losito, I., Palmisano, F., Godula, M., & Visconti, A. (2011). Towards the

921   quantification of residual milk allergens in caseinate-fined white wines using HPLC coupled with

922   single-stage Orbitrap mass spectrometry. *Food additives and Contaminants: Part A,* 28, 1304-

923   1314. https://doi.org/10.1080/19440049.2011.593191.

924   Monaci, L., Losito, I., De Angelis, E., Pilolli, R., & Visconti, A. (2013). Multi-allergen

925   quantification of fining-related egg and milk proteins in white wines by high-resolution mass

926   spectrometry. *Rapid Communications in Mass Spectrometry, 27,* 2009–2018.

927   https://doi.org/10.1002/rcm.6662.

928   Monaci, L., Pilolli, R., De Angelis, E., Godula, M., & Visconti, A. (2014). Multi-allergen detection

929   in food by micro high-performance liquid chromatography coupled to a dual cell linear ion trap

930   mass spectrometry. *Journal of Chromatography A, 1358,* 136–144.

931   https://doi/org/10.1016/j.chroma.2014.06.092.

932   Monaci, L., De Angelis, E., Montemurro, N., & Pilolli, R. (2018). Comprehensive overview and

933   recent advances in proteomics MS based methods for food allergens analysis. *Trends in*

934   *Analytical Chemistry, 106*, 21-36. https://doi.org/10.1016/j.trac.2018.06.016.

935   Montowska, M., & Fornal, E. (2018). Detection of peptide markers of soy, milk and egg white

936   allergenic proteins in poultry products by LC-Q-TOF-MS/MS. *LWT - Food Science and*

937   *Technology, 87*, 310-317. https://doi.org/10.1016/j.lwt.2017.08.091.

938   Montowska, M., & Fornal, E. (2019). Absolute quantification of targeted meat and allergenic

939   protein additive peptide markers in meat products. *Food Chemistry, 274,* 857–864.

940   https://doi.org/10.1016/j.foodchem.2018.08.131.

941 New, L. S., Schreiber, A., Stahl-Zeng, J., & Liu, H. F. (2018). Simultaneous Analysis of Multiple
942 Allergens in Food Products by LC-MS/MS. *Journal of AOAC International, 101,* 132-145.
943 https://doi.org/10.5740/jaoacint.17-0403.

944 Newsome, G. A., & Scholl, P. F. (2013). Quantification of allergenic bovine milk α(S1)-casein
945 in baked goods using an intact ^{15}N-labeled protein internal standard. *Journal of Agricultural and*
946 *Food Chemistry, 61,* 5659−5668. https://doi.org/10.1021/jf3015238.

947 Nielsen, N. C., Dickinson, C. D., Cho, T. J., Thanh, V. H., Scallon, B. J., Fischer, R. L., Sims,
948 T. L., Drews, G. N., Goldberg, R. B. (1989). Characterization of the glycinin gene family in
949 soybean. *The Plant Cell, 1,* 313-328. https://doi.org/10.1105/tpc.1.3.313.

950 Nitride, C., Mamone, G., Picariello, G., Mills, E.N.C., Nocerino, R., Berni Canani, R., & Ferranti,
951 P. (2013). Proteomic and immunological characterization of a new food allergen from hazelnut
952 (Corylus avellana). *Journal of proteomics, 86,* 16–26.
953 https://doi.org/10.1016/j.jprot.2013.05.001.

954 Palladino, C., & Breiteneder H. (2018). Peanut allergens. *Molecular Immunology, 100,* 58–70.
955 https://doi.org/10.1016/j.molimm.2018.04.005.

956 Parker, C. H., Khuda, S. E., Pereira, M., Ross, M. M., Fu, T.J., Fan, X., Wu, Y., Williams, K. M.,
957 DeVries, J., Pulvermacher, B., Bedford, B., Zhang, X., Jackson, L. S. (2015). Multi-allergen
958 Quantitation and the Impact of Thermal Treatment in Industry-Processed Baked Goods by
959 ELISA and Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Agricultural and*
960 *Food Chemistry, 63,* 10669−10680. https://doi.org/10.1021/acs.jafc.5b04287.

961 Pastorello, E. A., Vieths, S., Pravettoni, V., Farioli, L., Trambaioli, C., Fortunato, D., Lüttkopf,
962 D., Calamari, M., Ansaloni, R., Scibilia, J., Ballmer-Weber, B. K., Poulsen, L. K., Wütrich, B.,
963 Hansen, K. S., Robino, A. M., Ortolani, C., & Conti, A. (2002). Identification of hazelnut major
964 allergens in sensitive patients with positive double-blind, placebo-controlled food challenge
965 results. *Journal of Allergy and Clinical Immunology, 109,* 563-570.
966 https://doi.org/10.1067/mai.2002.121946.

967 Pavón-Pérez, J., Henriquez-Aedo, K., Aranda, M. (2019). Mass Spectrometry Determination of
968 Fining-Related Allergen Proteins in Chilean Wines. *Food Analytical Methods, 12*, 827–837.
969 https://doi.org/10.1007/s12161-018-01416-0.

970 Pedreschi R., Nørgaard, J., & Maquet, A. (2012). Current Challenges in Detecting Food
971 Allergens by Shotgun and Targeted Proteomic Approaches: A Case Study on Traces of Peanut
972 Allergens in Baked Cookies. *Nutrients, 4,* 132-150. https://doi.org/10.3390/nu4020132.

973 Pfeifer, S., Bublin, M., Dubiela, P., Hummel, K., Wortmann, J., Hofer, G., Keller, W., Radauer,
974 C., & Hoffmann-Sommergruber, K. (2015). Cor a 14, the allergenic 2S albumin from hazelnut,
975 is highly thermostable and resistant to gastrointestinal digestion. *Molecular Nutrition & Food
976 Research, 59,* 2077–2086. https://doi.org/10.1002/mnfr.201500071.

977 Pilolli, R., De Angelis, E., Godula, M., Visconti, A., & Monaci, L. (2014). Orbitrap™ monostage
978 MS versus hybrid linear ion trap MS: application to multi-allergen screening in wine. *Journal of
979 Mass Spectrometry, 49,* 1254–1263. https://doi.org/10.1002/jms.3453.

980 Pilolli, R., De Angelis, E., & Monaci, L. (2017a). Streamlining the analytical workflow for
981 multiplex MS/MS allergen detection in processed foods. *Food Chemistry, 221,* 1747–1753.
982 https://doi.org/10.1016/j.foodchem.2016.10.110.

983 Pilolli, R., Chaudhari, R., Palmisano, F., & Monaci, L. (2017b). Development of a mass
984 spectrometry immunoassay for unambiguous detection of egg allergen traces in wines.
985 *Analytical and Bioanalytical Chemistry, 409,* 1581-1589. https://doi.org/10.1007/s00216-016-
986 0099-3.

987 Pilolli, R., De Angelis, E., & Monaci, L. (2018). In house validation of a high resolution mass
988 spectrometry Orbitrap-based method for multiple allergen detection in a processed model food.
989 *Analytical and Bioanalytical Chemistry, 410,* 5653-5662. https://doi.org/10.1007/s00216-018-
990 0927-8.

991 Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2016). Advances in
992 ultra-high performance liquid chromatography coupled to tandem mass spectrometry for
993 sensitive detection of several food allergens in complex and processed foodstuffs. *Journal of
994 Chromatography A, 1464,* 115–123. https://doi.org/10.1016/j.chroma.2016.08.033.

995 Planque, M., Arnould, T., Renard, P., Delahaut, P., Dieu, M., & Gillard, N. (2017a) Highlight on
996 Bottlenecks in Food Allergen Analysis: Detection and Quantification by Mass Spectrometry.
997 *Journal of AOA International, 100,* 1-5. https://doi.org/10.5740/jaoacint.17-0005.

Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2017b). Liquid chromatography coupled to tandem mass spectrometry fordetecting ten allergens in complex and incurred foodstuffs. *Journal of Chromatography A, 1530,* 138–151. https://doi.org/10.1016/j.chroma.2017.11.039.

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 Chem. 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., Lopata, A. L., Mueller, G. A., Nandy, A., Radauer, C., Chan, S. K., Jappe, U., Kleine-Tebbe, J., Thomas, W. R., Chapman, M. D., van Hage, M., van Ree, R., Vieths, S., Raulf, M., Goodman, R. E., & WHO IUIS Allergen Nomenclature Sub-Committee. (2018). WHO/IUIS Allergen Nomenclature: Providing a common language. *Molecular Immunology 100,* 3–13. https://doi.org/10.1016/j.molimm.2018.03.003.

Qi, K., Liu, T., Yang, Y., Zhang, J., Yin, J., Ding, X., Qinc, W., Yang, Y. (2019). A rapid immobilized trypsin digestion combined with liquid chromatography – Tandem mass spectrometry for the detection of milk allergens in baked food. *Food Control, 102,* 179–187. https://doi.org/10.1016/j.foodcont.2019.03.017

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,* 4130–4141. https://doi.org/10.1039/c6an00359a.

Sayers, R. L., Gethings, L. A., Lee, V., Balasundaram, A., Johnson, P. E., Marsh, J. A., Wallace, A., Brown, H., Rogers, A., Langridge, J. I., & Mills, E. N. C. (2018). Microfluidic separation coupled to mass spectrometry for quantification of peanut allergens in a complex food matrix. *Journal of Proteome Research, 17,* 647−655. https://doi.org/10.1021/acs.jproteome.7b00714.

Schocker, F., Lüttkopf, D., Scheurer, S., Petersen, A., Cisteró-Bahima, A., Enrique, E., San Miguel-Moncín, M., Akkerdaas, J., Van Ree, R., Vieths, S., & Becker, W. M. (2004). Recombinant lipid transfer protein Cor a 8 from hazelnut: a new tool for in vitro diagnosis of

potentially severe hazelnut allergy. *Journal of Allergy and Clinical Immunology, 113,* 141–147. https://doi.org/10.1016/j.jaci.2003.09.013.

Scott, H., Sicherer, M.D., High, A., Sampson, M.D. (2018). Food allergy: A review and update on epidemiology, pathogenesis, diagnosis, prevention, and management. *Journal of Allergy and Clinical Immunology*, 141, 41-58. https://doi.org/10.1016/j.jaci.2017.11.003.

Shefcheck, K. J., Callahan, J. H., & Musser, S. M. (2006). Confirmation of Peanut Protein Using Peptide Markers in Dark Chocolate Using Liquid Chromatography−Tandem Mass Spectrometry (LC-MS/MS). *Journal of Agricultural and Food Chemistry, 54,* 7953-7959. https://doi.org/10.1021/jf060714e.

Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M. W., Berger, J., Greene, S. L., Nelson, M. N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R. K., Roorkiwal, M., Pandey, M. K., Hu, J., Li, Y. H., Wang, L. X., Guo, Y., Qiu, L. J., Redden, R. J., & Varshney, R. K. (2015). Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. *Critical Reviews in Plant Sciences, 34,* 43-104. https://doi.org/10.1080/07352689.2014.897904.

Tolin, S., Pasini, G., Simonato, B., Mainente, F., & Arrigoni, G. (2012a). Analysis of commercial wines by LC-MS/MS reveals the presence of residual milk and egg white allergens. *Food Control, 28,* 321-326. https://doi.org/10.1016/j.foodcont.2012.05.015.

Tolin, S., Pasini, G., Curioni, A., Arrigoni, G., Masi, F., Mainente, F., & Simonato, B. (2012b). Mass spectrometry detection of egg proteins in red wines treated with egg white. *Food Control, 23,* 87-94. https://doi.org/10.1016/j.foodcont.2011.06.016.

Turner, P. J., Gowland, M. H., Sharma, V, Ierodiakonou, D., Harper, N., Garcez, T., Pumphrey, R., & Boyle, R. J. (2015). Increase in anaphylaxis-related hospitalizations but no increase in fatalities: an analysis of United Kingdom national anaphylaxis data, 1992-2012. *Journal of Allergy and Clinical Immunology, 135*, 956-63. https://doi.org/10.1016/j.jaci.2014.10.021.

Vandekerckhove, M., Van Droogenbroeck, B., De Loose, M., Taverniers, I., Daeseleire, E., Gevaert, P., Lapeere, H., & Van Poucke, C. (2017). Development of an LC-MS/MS method for the detection of traces of peanut allergens in chili pepper. *Analytical and Bioanalytical Chemistry, 409,* 5201–5207. https://doi.org/10.1007/s00216-017-0506-4.

1055   Verma, A. K., Kumar, S., Das, M., & Dwivedi, P. D. (2013). A Comprehensive Review of
1056   Legume Allergy. Clinical Review in Allergy and Immunology, 45, 30–46.
1057   https://doi.org/10.1007/s12016-012-8310-6.

1058   Verweij, M. M., Hagendorens, M. M., Trashin, S., Cucu, T., De Meulenaer, B., Devreese, B.,
1059   Bridts, C. H., De Clerck, L. S., & Ebo, D. G. (2012). Age-dependent sensitization to the 7S-
1060   vicilin like protein Cor a 11 from hazelnut (Corylus avellana) in a birch-endemic region. *Journal*
1061   *of Investigational Allergology and Clinical Immunology, 22,* 245-251.

1062   Viquez, O. M., Konan, K. N., & Dodo, H. W. (2004). Genomic organization of peanut allergen
1063   gene, Ara h 3. *Molecular Immunology, 41,* 1235–1240.
1064   https://doi.org/10.1016/j.molimm.2004.06.033.

1065   Wal, J. (2001). Structure and function of milk allergens. *Allergy, 56,* 35-38.
1066   https://doi.org/10.1034/j.1398-9995.2001.00911.x

1067   Wang, T., Qin, G.-X., Sun, Z.-W., & Zhao, Y. (2014). Advances of Research on Glycinin and β-
1068   Conglycinin: A Review of Two Major Soybean Allergenic Proteins. *Critical Review in Food*
1069   *Science and Nutrition, 54,* 850–862. https://doi.org/10.1080/10408398.2011.613534.

1070   Worm, M., Moneret-Vautrin, A., Scherer, K., Lang, R., Fernandez-Rivas, M., Cardona, V.,
1071   Kowalski, M. L., Jutel, M., Poziomkowska-Gesicka, I., Papadopoulos, N. G., Beyer, K.,
1072   Mustakov, T., Christoff, G., Bilò, M. B., Muraro, A., Hourihane, J. O., & Grabenhenrich, L. B.
1073   (2014). First European data from the network of severe allergic reactions (NORA). *Allergy, 69,*
1074   1397-404. https://doi.org/10.1111/all.12475.

1075   Yang, W., Liqing, W., Fei, D., Bin, Y., Yi, Y., Jing, W. (2014). Development of an SI-traceable
1076   HPLC–isotope dilution mass spectrometry method to quantify β-lactoglobulin in milk powders.
1077   *Journal of Agricultural and Food Chemistry, 62,* 3073−3080. https://doi.org/10.1021/jf4054337.

1078   Zhang, J., Hong, Y., Cai, Z., Huang, B., Wang, J., Ren, Y. (2019). Simultaneous determination
1079   of major peanut allergens Ara h1 and Ara h2 in baked foodstuffs based on their signature
1080   peptides using ultraperformance liquid chromatography coupled to tandem mass spectrometry.
1081   *Analytical Methods, 11*, 1689-1696. https://doi.org/10.1039/c9ay00256a.

1082

FIGURES AND TABLES

Figure 1 (color online version only)
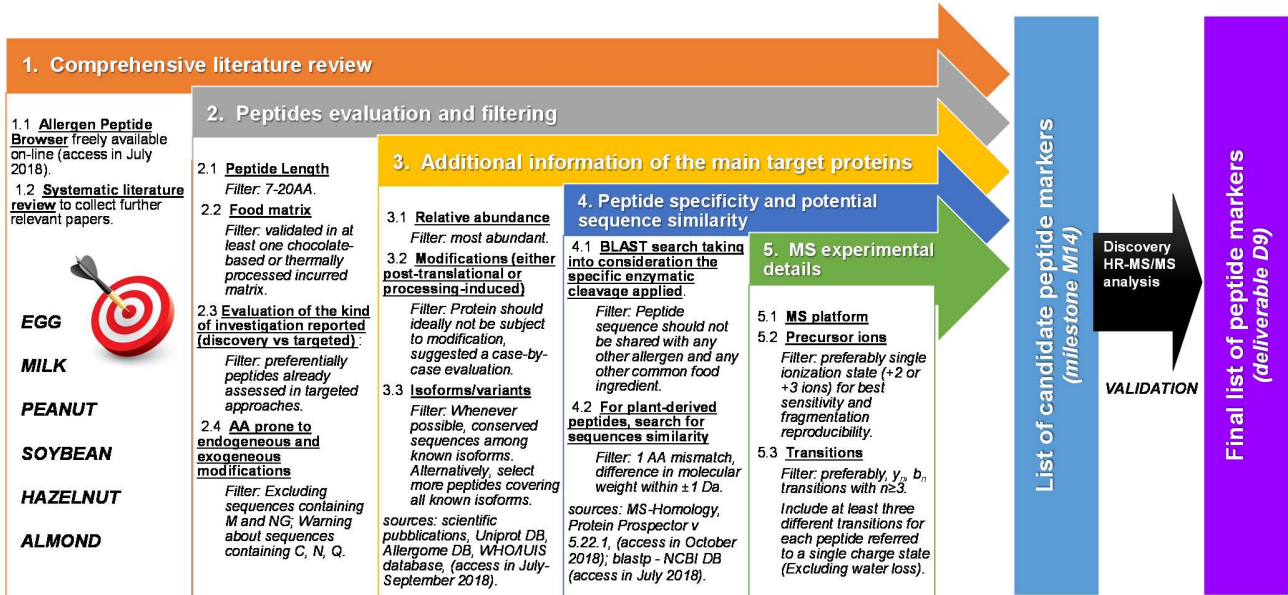
Figure 2 (color online version only)
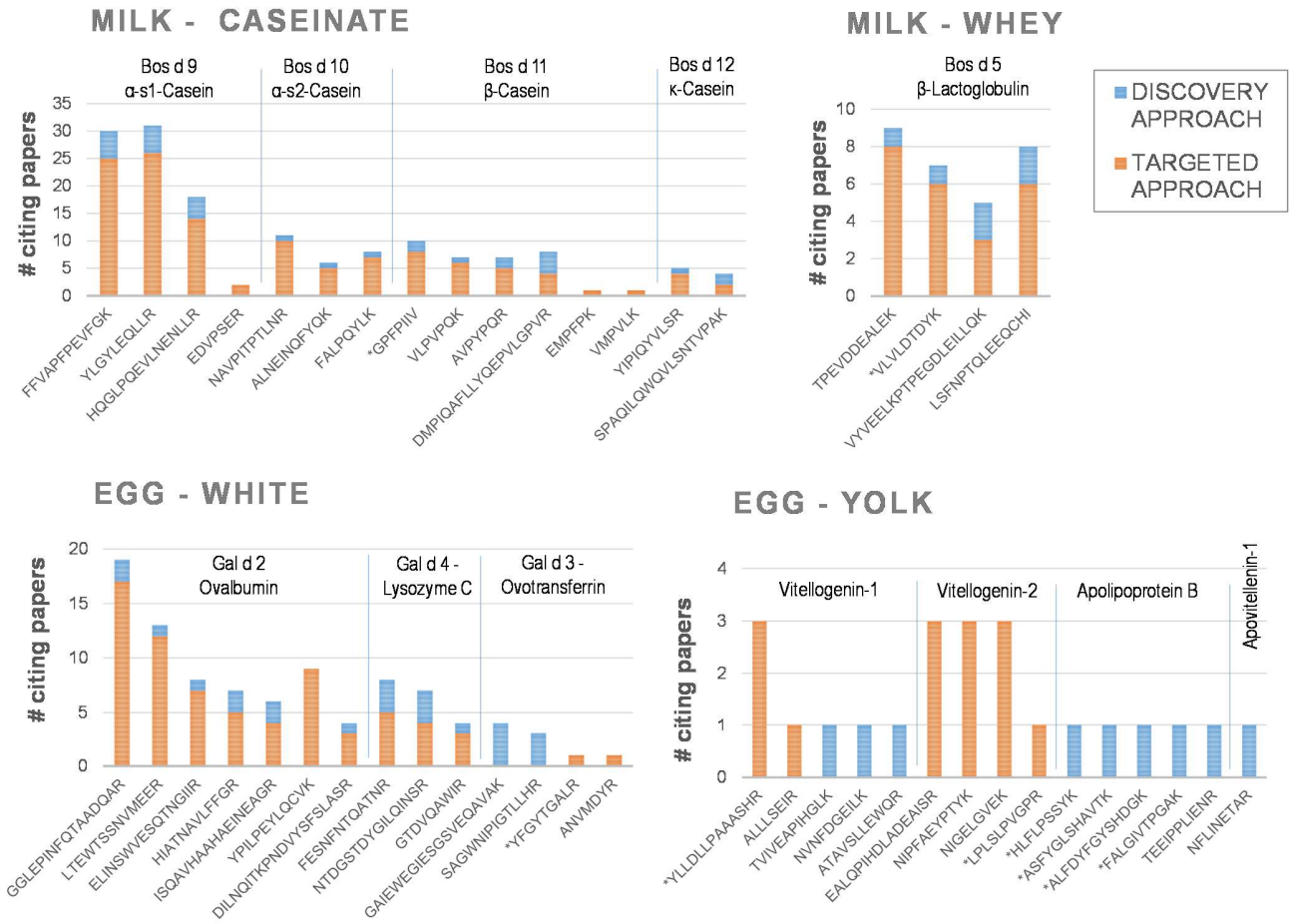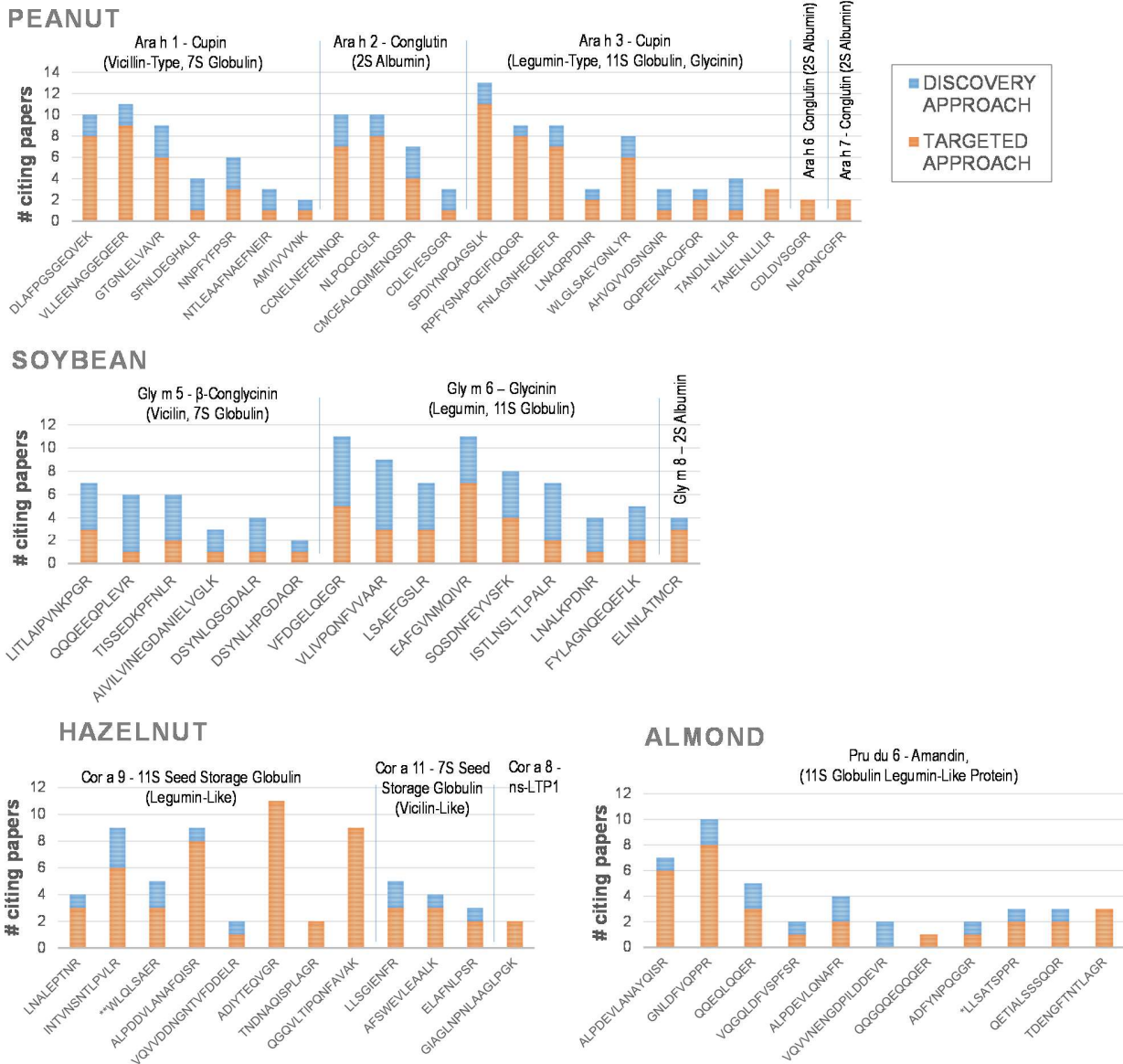
Figure 3

Table 1

Allergenic ingredient	Protein	Allergen Name	MW (kDa)	Number of known PTMs*	Isoforms and Variants	IUIS (YES/NO)	GenBank Protein	UniProt
Milk	αS1-Casein	Bos d 9	23.6	n° 9 Ser phosphorylationsphosphorylations	Bos d 9.0101	Yes	NP_851372	P02662
	αS2-Casein	Bos d 10	25.2	n° 13 Ser phosphorylationsphosphorylations	Bos d 10.0101	Yes	NP_776953	P02663
	β-Casein	Bos d 11	24	n° 5 Ser phosphorylationsphosphorylations	Bos d 11.0101	Yes	XP_005902099	P02666
	κ-Casein	Bos d 12	19	n°1 N-terminal Gln cyclization to pyroglutamate n° 3 Disulfide Bonds n° 3 Ser phosphorylationsphosphorylations n° 1 Thr phosphorylation n° 6 Thr glycosylations n° 2 Ser glycosylations	Bos d 12.0101	Yes	NP_776719	P02668
	α-Lactalbumin	Bos d 4	14.2	n° 4 Disulfide Bonds n° 1 Asn glycosylation	Bos d 4.0101	Yes	AAA30615	P00711
	β-Lactoglobulin	Bos d 5	18.3	n° 3 Disulfide Bonds	Bos d 5.0101	Yes	CAA32835	P02754
	Bovine serum albumin	Bos d 6	66.3	n° 17 Disulfide Bonds n° 6 Ser phosphorylations n° 4 Thr phosphorylations n° 2 N6-succinyllysine n° 1 Lys methylation	Bos d 6.0101	Yes	AAA51411	P02769
	Immunoglobulins	Bos d 7	160		Bos d 7.0101	Yes		
Egg	Ovalbumin	Gal d 2	44	Initial Met Removal n°1 Gly acetylation n° 2 Ser phosphorylations n° 1 Disulfide Bond n° 1 Asn glycosylation	Gal d 2.0101	Yes	CAA23682	P01012
	Ovomucoid	Gal d 1	28	n° 9 Disulfide Bonds n° 5 Asn glycosylations	Gal d 1.0101	Yes	P01005	P01005
	Ovotransferrin	Gal d 3	78	n° 15 Disulfide Bonds n° 1 Asn glycosylation	Gal d 3.0101	Yes	CAA26040	P02789
	Lysozyme C	Gal d 4	14	n° 4 Disulfide Bonds	Gal d 4.0101	Yes	CAA23711	P00698
	Serum Albumin	Gal d 5	69	n° 17 Disulfide Bonds n° 1 Asn glycosylation	Gal d 5.0101	Yes	CAA43098	P19121
	YGP42	Gal d 6	35	n° 3 Asn glycosylations	Gal d 6.0101	Yes	BAA13973	chain 1628-1912 of the P87498
	Myosin light chain 1f	Gal d 7	22	n° 1 Ala trimethylation	Gal d 7.0101	Yes	K02608.1, K02609.1 and K02610.1	P02604
	alpha-parvalbumin	Gal d 8	11.8	NA	Gal d 8.0101	Yes	CAX32963	C1L370
	Beta-enolase	Gal d 9	50	n°1 Ser acetylation	Gal d 9.0101	Yes	NP_990450	P07322

	Aldolase	Gal d 10		NA	Gal d 10.0101	Yes		
Peanut	Vicilin-type, 7S globulin (Cupin)	Ara h 1	64	n° 1 Asn glycosylation	Ara h 1.0101 (clone P41B)	Yes	AAB00861	P43238
					Ara h 1 - Clone P17	No	AAA60336	P43237
	Conglutin, 2S albumin (Prolamin)	Ara h 2	17	n° 3 Pro hydroxylation n° 4 Disulfide Bonds	Ara h 2.0101	Yes	AAK96887	
					Ara h 2.0102	No	see Allergome	
					Ara h 2.0201	Yes	AAN77576	
					Ara h 2.0202	No	see Allergome	
	Legumin-type, Glycinin, 11S globulin (Cupin)	Ara h 3	60, 37 (fragment)	NA	Ara h 3.0101	Yes	AAC63045	O82580
					Ara h 3.0201	Yes	AAD47382	Q9SQH7
					Ara h 3 genomic clone	No	AAM46958	
					Iso Ara h 3	No	AAT39430	
	Profilin	Ara h 5	15	NA	Ara h 5.0101	Yes	AAD55587	Q9SQI9
	Conglutin, 2S albumin (Prolamin)	Ara h 6	15	n° 5 Disulfide Bonds	Ara h 6.0101	Yes	AAD56337	Q647G9
	Conglutin, 2S albumin (Prolamin)	Ara h 7	15	NA	Ara h 7.0101	Yes	AAD56719	Q9SQH1
					Ara h 7.0201	Yes	ABW17159	B4XID4
					Ara h 7.0301	Yes	AAU21496	Q647G8
	Pathogenesis-related protein, PR-10, (Bet v 1 like)	Ara h 8	17	NA	Ara h 8.0101	Yes	AAQ91847	Q6VT83
					Ara h 8.0201	Yes	ABP97433	B0YIU5
	Nonspecific lipid-transfer protein type 1 (Prolamin)	Ara h 9	9.8	Disulfide bond	Ara h 9.0101	Yes	ABX56711	B6CEX8
					Ara h 9.0201	Yes	ABX75045	B6CG41
	16 kDa oleosin (Glycosyl transferase GT-C)	Ara h 10	16	NA	Ara h 10.0101	Yes	AAU21499	Q647G5
					Ara h 10.0102	Yes	AAU21500	Q647G4
	14 kDa oleosin (Glycosyl transferase GT-C)	Ara h 11	14	NA	Ara h 11.0101	Yes	AAZ20276	Q45W87
					Ara h 11.0102	Yes	AAZ20277	Q45W86
	Defensin (Scorpion toxin-like knottin)	Ara h 12	8 kDa (reducing), 12 kDa (non-reducing), 5.184 kDa (mass)	NA	Ara h 12.0101	Yes		
	Defensin (Scorpion toxin-like knottin)	Ara h 13	8 kDa (reducing), 11 kDa (non-reducing), 5.472 kDa (mass)	NA	Ara h 13.0101	Yes		
					Ara h 13.0201	Yes		
	Oleosin (Glycosyl transferase GT-C)	Ara h 14	17.5	NA	Ara h 14.0101	Yes	AAK13449	Q9AXI1
					Ara h 14.0102	Yes	AAK13450	Q9AXI0
					Ara h 14.0103	Yes	AAT11925	Q6J1J8
	Oleosin (Glycosyl transferase GT-C)	Ara h 15	17	NA	Ara h 15.0101	Yes	AAU21501	Q647G3
	non-specific Lipid Transfer Protein 2 (Prolamin)	Ara h 16	8.5 (reducing)	NA	Ara h 16.0101	Yes		
	non-specific Lipid Transfer Protein 1 (Prolamin)	Ara h 17	11 (reducing)	NA	Ara h 17.0101	Yes		
Hazelnut		Cor a 1	17	Initial Met Removal	Cor a 1.0101	Yes	CAA50327	Q08407

	Pathogenesis-related protein, PR-10, (Bet v 1 like)				Cor a 1.0102	Yes	CAA5032	Q08407
					Cor a 1.0103	Yes	CAA50325	Q08407
					Cor a 1.0104	Yes	CAA50326	Q08407
					Cor a 1.0201	Yes	CAA96548	Q39453
					Cor a 1.0301	Yes	CAA96549	Q39454
					Cor a 1.0401	Yes	AAD48405	Q9SWR4
					Cor a 1.0402	Yes	AAG40329	Q9FPK4
					Cor a 1.0403	Yes	AAG40330	Q9FPK3
					Cor a 1.0404	Yes	AAG40331	Q9FPK2
	Profilin	Cor a 2	14	NA	Cor a 2.0101	Yes	AAK01235	Q9AXH5
					Cor a 2.0102	Yes	AAK01236	Q9AXH4
	Non-specific lipid transfer protein type 1 (Prolamin)	Cor a 8	9	n° 4 Disulfide Bonds	Cor a 8.0101	Yes	AAK28533	Q9ATH2
	Legumin-like, 11S globulin (Cupin)	Cor a 9	40	Disulfide bond	Cor a 9.0101	Yes	AAL73404	Q8W1C2
					-	No	AHA36627	A0A0A0P7E3
	Vicilin-like, 7S globulin (Cupin)	Cor a 11	48	NA	Cor a 11.0101	Yes	AAL86739	Q8S4P9
	Oleosin	Cor a 12	17	NA	Cor a 12.0101	Yes	AAO67349	Q84T21
	Oleosin	Cor a 13	14-16	NA	Cor a 13.0101	Yes	AAO65960	Q84T91
	2S albumin (Prolamin)	Cor a 14	10	NA	Cor a 14.0101	Yes	ACO56333	D0PWG2
Almond	Non-specific lipid transfer protein 1 (Prolamin)	Pru du 3	9	Disulfide bond	Pru du 3.0101	Yes	ACN11576	C0L0I5
	Profilin	Pru du 4	14	NA	Pru du 4.0101	Yes	AAL91662	Q8GSL5
					Pru du 4.0102	Yes	AAL91664	Q8GSL5
	60s acidic ribosomal prot. P2	Pru du 5	10	NA	Pru du 5.0101	Yes	ABH03379	Q8H2B9
	Legumin-like, Amandin, 11S globulin (Cupin)	Pru du 6	ca. 360	Disulfide bond	Pru du 6.0101	Yes	ADN39440	E3SH28
					Pru du 6.01 variant	No	CAA55009	Q43607
					Pru du 6.0201	Yes	ADN39441	E3SH29
					Pru du 6.02 variant	No	CAA55010	Q43608
Soybean	Profilin	Gly m 3	14	Initial Met Removal	Gly m 3.0101	Yes	CAA11756	O65809
					Gly m 3.0102	Yes	CAA11755	O65810
	Pathogenesis-related protein, PR-10, (Bet v 1 like)	Gly m 4	17	NA	Gly m 4.0101	Yes	CAA42646	P26987
	Vicilin, β-conglycinin, 7S globulin (Cupin)	Gly m 5	63	NA	Gly m 5.0101	Yes	BAA23360	O22120
			65		Gly m 5.0201	Yes	BAA74452	Q9FZP9
			48	n° 1 Asn glycosylation	Gly m 5.0301	Yes	AAB23463	P25974 (variant F36L V51G F197L)
			48	n° 1 Asn glycosylation	Gly m 5.0302	Yes		P25974 (variant V51G)
	Legumin, Glycinin, 11S globulin (Cupin)	Gly m 6	54	n° 2 Disulfide Bonds n° 2 Propeptides	Gly m 6.0101	Yes	AAA33966 BAC78522	P04776
			52	n° 2 Disulfide Bonds n° 1 Propeptide	Gly m 6.0201	Yes	BAA00154	P04405
			52	n° 2 Disulfide Bonds n° 1 Propeptide	Gly m 6.0301	Yes	CAA33217	P11828

		61	Disulfide bond	Gly m 6.0401	Yes	BAA74953	Q9SB11
		56	n° 2 Disulfide Bonds	Gly m 6.0501	Yes	BAB15802	Q7GC77
Seed biotinylated protein	Gly m 7	76	NA	Gly m 7.0101	Yes	ACS49840	C6K8D1
2S albumin (Prolamin)	Gly m 8	14	Disulfide Bond n° 1 Propeptide	Gly m 8.0101	Yes	AAB71140	P19594

Table 2

Target protein	Peptide Sequence *Isoallergens/variants*	Specificity (Tax ID)	MS platform	Precursor	Transition 1		Transition 2		Transition 3		References
Bos d 9 – α-S1-Casein	FFVAPFPEVFGK *Bos d 9.0101*	*Bos taurus (9913)*	QqQ	693.3 (+2)	920.8	y8+	676.6	y6+	267.3	a2+	Ansari et al., 2011
		Bos mutus (72004)	QqQ	692.9 (+2)	920.3	y8+	991.3	y9+	-	-	Heick et al., 2011a
		Bubalus bubalis (89462)	QqQ	692.9 (+2)	920.3	y8+	991.3	y9+	-	-	Heick et al., 2011b
			IT	692.9 (+2)	920.5	y8+	991.5	y9+	1090.6	y10+	Lamberti et al., 2016
			IT	1384.6 (+1)	-	-	-	-	-	-	Losito et al., 2013
			Orbitrap	692.9 (+2)	-	-	-	-	-	-	Monaci et al., 2013
			Orbitrap	692.9 (+2)	-	-	-	-	-	-	Monaci et al., 2011
			Q-ToF	692.6 (+2)	-	-	-	-	-	-	Monaci et al., 2010a
			Q-ToF	692.9 (+2)	920.5	y8+	676.4	y6+	295.2	b2+	Monaci et al., 2010b
			Dual LIT	692.9 (+2)	920.5	y8+	991.5	y9+	-	-	Monaci et al., 2014
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	1090.6	y10+	Newsome & Scholl, 2013
			QqQ	692.9 (+2)	920.5*	y8+	991.5	y9+	1090.6	y10+	Parker et al., 2015
			Dual LIT / Orbitrap	692.9 (+2)	920.5	y8+	991.5	y9+	-	-	Pilolli et al., 2014
			Dual LIT	692.9 (+2)	920.5*	y8+	991.5	y9+	545.8	y10++	Pilolli et al., 2017a
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	676.4	y6+	Planque et al., 2016
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	676.4	y6+	Planque et al., 2017a
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	676.4	y6+	Planque et al., 2017b
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	676.4	y6+	Gu et al., 2018
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	1090.6	y10+	Boo et al., 2018
			QqQ	692.9 (+2)	920.5	y8+	-	-	-	-	Groves et al., 2018
			QqQ	692.9 (+2)	295.1*	b2+	394.2	b3+	-	-	Ke et al., 2017
			Q-Orbitrap	692.9 (+2)	-	-	-	-	-	-	Pilolli et al., 2018
			Dual LIT	692.9 (+2)	920.5	y8+	991.5	y9+	1090.6	y10+	De Angelis et al., 2017b
			QqQ	692.9 (+2)	920.5	y8+	991.5	y9+	-	-	Planque et al., 2019
			QqQ	692.9 (+2)	460.8*	y8++	496.3	y9++	-	-	Qi et al., 2019
			QqQ	692.8 (+2)	1090.6	y10+	676.4	y6+	450.3	y4+	Montowska & Fornal, 2019
	YLGYLEQLLR *Bos d 9.0101*	*Bos taurus (9913)*	QqQ	634.8 (+2)	991.7	y8+	771.8	y6+	249.4	a2+	Ansari et al., 2011
		Capra hircus (9925)	QqQ	634.3 (+2)	991.3	y8+	249.2	a2+			Heick et al., 2011a
		Ovis aries (9940)	QqQ	634.3 (+2)	991.3	y8+	249.2	a2+			Heick et al., 2011b
		Bubalus bubalis (89462)	IT	634.4 (+2)	991.6	y8+	771.5	y6+	658.4	y5+	Lamberti et al., 2016
		Bos motus (72004)	IT	1267.6 (+1)	-	-	-	-	-	-	Losito et al., 2013
			LIT	634.8 (+2)	991.8	y8+	771.4	y6+			Mattarozzi et al., 2014
			Orbitrap	634.4 (+2)	-	-	-	-	-	-	Monaci et al., 2013
			Orbitrap	634.4 (+2)	-	-	-	-	-	-	Monaci et al., 2011
			Q-ToF	634.2 (+2)	-	-	-	-	-	-	Monaci et al., 2010a
			Q-ToF	634.3 (+2)	991.6	y8+	771.5	y6+	249.2	a2+	Monaci et al., 2010b
			Dual LIT	634.4 (+2)	991.6	y8+	771.5	y6+	-	-	Monaci et al., 2014
			QqQ	634.4 (+2)	991.6	y8+	771.5	y6+	-	-	Newsome & Scholl, 2013
			QqQ	634.4 (+2)	991.6*	y8+	771.5	y6+	658.4	y5+	Parker et al., 2015
			Dual LIT / Orbitrap	634.4 (+2)	991.6	y8+	771.5	y6+	-	-	Pilolli et al., 2014
			Dual LIT	634.4 (+2)	991.6	y8+	771.5	y6+	658.4	y5+	Pilolli et al., 2017a
			QqQ	634.4 (+2)	934.5	y7+	771.5	y6+	658.4	y5+	Planque et al., 2016
			QqQ	634.4 (+2)	934.5	y7+	771.5	y6+	658.4	y5+	Planque et al., 2017b
			QqQ-ToF	634.6 (+2)	991.3	y8+	249.2	a2+	-	-	Ji et al., 2017
			QqQ	634.4 (+2)	991.6	y8+	771.5	y6+	658.4	y5+	Gu et al., 2018
			QqQ	634.4 (+2)	991.6	y8+	771.5	y6+	658.4	y5+	Boo et al., 2018
			QqQ	634.4 (+2)	991.6	y8+	249.2	a2+	-	-	New et al., 2018
			QqQ	634.4 (+2)	991.6	y8+			-	-	Groves et al.,2018
			Q-Orbitrap	634.4 (+2)	-	-	-	-	-	-	Pilolli et al., 2018
			Dual LIT	634.4 (+2)	991.6	y8+	771.5	y6+	658.4	y5+	De Angelis et al., 2017b
			QqQ	634.4 (+2)	934.5	y7+	771.5	y6+	-	-	Planque et al., 2019
			QqQ	634.4 (+2)	552.8	y9++	249.1	b4++	-	-	Qi et al., 2019

Allergen / Isoform	Peptide	Species	MS platform	Precursor m/z (charge)	Product ion 1	Type	Product ion 2	Type	Product ion 3	Type	Reference
			QqQ	634.4 (+2)	991.6	y_8^+	771.5	y_6^+	529.3	y_4^+	Montowska & Fornal, 2019
			QqQ	634.6 (+2)	991.8	y_8^+	771.4	y_6^+			Pavón-Pérez et al., 2019
Bos d 10 – αS2-Casein	NAVPITPTLNR Bos d 10.0101	Bos taurus (9913) Bubalus bubalis (89462) Bos motus (72004)	Q-ToF	1195.7 (+1)	911.4	y_8^+	701.4	y_6^+	600.3	y_5^+	Gomaa & Boye, 2015
			QqQ	598.3 (+2)	911.4	y_8^+	158.3	a_2^+	-	-	Heick et al., 2011a
			QqQ	598.3 (+2)	911.4	y_8^+	158.3	a_2^+	-	-	Heick et al., 2011b
			Orbitrap	598.3 (+2)	-	-	-	-	-	-	Monaci et al., 2013
			Q-ToF	598.4 (+2)	-	-	-	-	-	-	Monaci et al., 2010a
			QqQ	598.3 (+2)	911.5	y_8^+	456.3	y_8^{++}	285.2	b_3^+	Planque et al., 2016
			QqQ	598.3 (+2)	911.5	y_8^+	456.3	y_8^{++}	285.2	b_3^+	Planque et al., 2017b
			QqQ	598.3 (+2)	911.5	y_8^+	701.4	y_6^+	600.3	y_5^+	Gu et al., 2018
			QqQ	598.3 (+2)	912.0	y_8^+	285.5	b_3^+	-	-	Ke et al., 2017
			QqQ	598.3 (+2)	911.5	y_8^+	285.2	b_3^+	-	-	Planque et al., 2019
			QqQ	598.4 (+2)	911.6*	y_8^+	456.3	y_8^{++}	-	-	Qi et al., 2019
	FALPQYLK Bos d 10.0101	Bos taurus (9913) Bos motus (72004)	QqQ	490.3 (+2)	120.1	a_1^+	648.4	y_5^+	-	-	Heick et al., 2011a
			QqQ	490.3 (+2)	120.1	a_1^+	648.4	y_5^+	-	-	Heick et al., 2011b
			IT	979.3 (+1)	-	-	-	-	-	-	Losito et al., 2013
			QqQ	490.1 (+2)	761.5		648.4*	y_5^+	-	-	Lutter et al., 2011
			Orbitrap	979.6 (+1)	-	-	-	-	-	-	Monaci et al., 2013
			QqQ	490.3 (+2)	761.5	y_6^+	648.4	y_5^+	423.3	y_3^+	Gu et al., 2018
			QqQ	490.2 (+2)	761.5	y_6^+	219.2	b_2^+	-	-	Ke et al., 2017
			QqQ	490.3 (+3)	648.4	y_5^+	219.1	b_2^+	-	-	Qi et al. 2019
Bos d 5 - β-Lactoglobulin	TPEVDDEALEK Bos d 5.0101	Bos taurus (9913) Bubalus bubalis (89462) Bos mutus (72004)	QqQ	623.3 (+2)	918.4	y_8^+	819.4*	y_7^+	-	-	Lutter et al., 2011
			Dual LIT	623.3 (+2)	572.8	y_{10}^{++}	819.4	y_7^+	-	-	Monaci et al., 2014
			QqQ	623.3 (+2)	572.8	y_{10}^{++}	819.4	y_7^+	918.4	y_8^+	Parker et al., 2015
			QqQ	623.3 (+2)	-	-	-	-	-	-	Yang et al., 2014
			QqQ-ToF	623.3 (+2)	1048.2	y_9^+	199.2	b_2^+	-	-	Ji et al. 2017
			QqQ	623.3 (+2)	572.8	y_{10}^{++}	819.4	y_7^+	918.4	y_8^+	Boo et al., 2018
			Q-Orbitrap	623.3 (+2)	-	-	-	-	-	-	Pilolli et al., 2018
			QqQ	623.3 (+2)	572.8		y_{10}^{++}	819.4	y_7^+	-	New et al., 2018
	VLVLDTDYK Bos d 5.0101	Bos taurus (9913) Bubalus bubalis (89462) Capra hircus (9925) Ovis aries musimon (9938) Ovis aries (9940) Rangifer tarandus tarandus (86329) Bos indicus (9915) Bos grunniens (30521) Bos motus (72004) Ovis sp. (9939)	QqQ	533.6 (+2)	-	-	-	-	-	-	Figeys et al., 1996
			QqQ	533.3 (+2)	853.4*	y_7^+	754.4	y_6^+	-	-	Lutter et al., 2011
			QqQ	533.3 (+2)	853.4	y_7^+	754.4	y_6^+	641.3	y_5^+	Parker et al., 2015
			QqQ	533.3 (+2)	853.4	y_7^+	754.4	y_6^+	641.3	y_5^+	Planque et al., 2016
			QqQ	533.3 (+2)	853.4	y_7^+	754.4	y_6^+	641.3	y_5^+	Planque et al., 2017b
			QqQ	533.3 (+2)	853.4	y_7^+	754.4	y_6^+	-	-	Planque et al., 2019
	VYVEELKPTPEGDLEILLQK Bos d 5.0101	Bos taurus (9913) Bubalus bubalis (89462) Bos grunniens (30521) Bos motus (72004)	QqQ	771.5 (+3)	-	-	-	-	-	-	Figeys et al., 1996
			QqQ	771.8 (+3)	912.0	y_{16}^{++}	790.9	y_{14}^{++}	627.9	y_{11}^{++}	Planque et al., 2016
			QqQ	771.8 (+3)	912.0	y_{16}^{++}	790.9	y_{14}^{++}	627.9	y_{11}^{++}	Planque et al., 2017b
	LSFNPTQLEEQ**C**HI Bos d 5.0101	Bos taurus (9913) Bubalus bubalis (89462)	QqQ	858.4 (+2)	928.4	y_7^+	815.3	y_6^+	627.8	y_{10}^{++}	Parker et al., 2015
			QqQ	858.4 (+2)	928.4	y_7^+	1254.6	y_{10}^+	627.8	y_{10}^{++}	Planque et al., 2016
			QqQ	858.4 (+2)	1254.6	y_{10}^+	815.3	y_6^+	627.8	y_{10}^{++}	Planque et al., 2017b
			QqQ	858.4 (+2)	928.4	y_7^+	627.8	y_{10}^{++}	-	-	Boo et al., 2018
			QqQ	858.6 (+2)	462.2*	b_4^+	685.1	y_{11}^{++}	-	-	Ke et al., 2017
			QqQ	858.4 (+2)	1254.6		y_{10}^+	627.8	y_{10}^{++}		Planque et al., 2019
Gal d 2 - Ovalbumin	GGLEPINFQTAADQAR Gal d 2.0101	Gallus gallus (9031) Alcaligenes xylosoxydans xylosoxydans (85698)	Q-ToF	-	-	-	-	-	-	-	Azarnia et al., 2013
			LIT	844.4 (+2)	666.3	y_{12}^{++}	1331.7	y_{12}^+	-	-	Mattarozzi et al., 2014
			Orbitrap	844.4 (+2)	-	-	-	-	-	-	Monaci et al., 2013
			Dual LIT	844.4 (+2)	666.3	y_{12}^{++}	1121.5	y_{10}^+	-	-	Monaci et al., 2014
			QqQ	844.4 (+2)	1007.5	y_9^+	1121.5	y_{10}^+	860.4	y_8^+	Parker et al., 2015
			Dual LIT / Orbitrap	844.4 (+2)	666.3	y_{12}^{++}	1331.7	y_{12}^+	-	-	Pilolli et al., 2014

Protein	Peptide	Species	Instrument	Precursor (m/z)	m/z	ion	m/z	ion	m/z	ion	Reference
			Dual LIT	844.4 (+2)	666.3	y12++	1121.5	y10+	732.4	y7+	Pilolli et al., 2017a
			QqQ	844.4 (+2)	666.3	y12++	1121.5	y10+	1331.7	y12+	Planque et al., 2016
			Q-ToF	844.4 (+2)	-	-	-	-	-	-	Tolin et al., 2012b
			QqQ	844.4 (+2)	666.3	y12++	-	-	-	-	Planque et al., 2017a
			QqQ	844.4 (+2)	666.3	y12++	1121.5	y10+	1331.7	y12+	Planque et al., 2017b
			QqQ	844.4 (+2)	1007.5	y9+	1121.5	y10+	860.4	y8+	Boo et al., 2018
			Dual LIT	844.4 (+2)	666.3	y12++	1121.5	y10+	732.4	y7+	Pilolli et al., 2017b
			QqQ	844.4 (+2)	666.3	y12++	1121.5	y10+	-	-	New et al., 2018
			Q-Orbitrap	844.4 (+2)	-	-	-	-	-	-	Pilolli et al., 2018
			Dual LIT	844.4 (+2)	666.3	y12++	1121.5	y11+	1331.7	y12+	De Angelis et al., 2017b
			QqQ	844.4 (+2)	666.3	y12++	1121.5	y11+			Planque et al., 2019
	HIATNAVLFFGR Gal d 2.0101	Gallus gallus (9031)	Q-ToF	-	-	-	-	-	-	-	Azarnia et al., 2013
		Alcaligenes xylosoxydans xylosoxydans (85698)	QqQ	673.4 (+2)	1095.6	y10+	223.2	a2+	-	-	Heick et al., 2011a
			QqQ	673.4 (+2)	1095.6	y10+	223.2	a2+	-	-	Heick et al., 2011b
			Orbitrap	673.4 (+2)	-	-	-	-	-	-	Monaci et al., 2013
			QqQ	449.3 (+3)	639.4*	y5+	526.3	y4+	608.3	b6+	Parker et al., 2015
			Q-ToF	-	-	-	-	-	-	-	Tolin et al., 2012°
	ISQAVHAAHAEINEAGR Gal d 2.0101	Gallus gallus (9031)	Q-ToF	-	-	-	-	-	-	-	Azarnia et al., 2013
		Coturnix japonica (93934)	Orbitrap	887.4 (+2)	-	-	-	-	-	-	Monaci et al., 2013
		Alcaligenes xylosoxydans xylosoxydans (85698)	QqQ	887.5 (+2)	1138.6	y11+	1067.5	y10+	996.5	y9+	Planque et al., 2016
			QqQ	887.5 (+2)	1138.6	y11+	1067.5	y10+	996.5	y9+	Planque et al., 2017b
Gal d 4 - Lysozyme C	FESNFNTQATNR Gal d 4.0101	Gallus gallus (9031)	QqQ	714.8 (+2)	277.1	b2+	-	-	-	-	Cryar et al., 2013
		Coturnix japonica (93934)	Orbitrap	714.8 (+2)	-	-	-	-	-	-	Monaci et al., 2013
		Rattus norvegicus (10116)	QqQ	714.8 (+2)	1152.5	y10+	951.5	y8+	804.4	y7+	Parker et al., 2015
		Gallus lafayetii (9032)	QqQ	714.8 (+2)	589.3	y5+	690.4	y6+	-	-	Pilolli et al., 2014
		Gallus sonneratii (9033)	QqQ	714.8 (+2)	1152.5	y10+	951.5	y8+	804.4	y7+	Boo et al., 2018
	NTDGSTDYGILQINSR Gal d 4.0101	Gallus gallus (9031)	Orbitrap	877.4 (+2)	-	-	-	-	-	-	Monaci et al., 2013
		Anas platyrhynchos (8839)	QqQ	585.3 (+3)	489.3	y4+	617.3	y5+	730.4	y6+	Parker et al., 2015
		Catreus wallichii (9085)	Dual LIT / Orbitrap	877.4 (+2)	489.3	y4+	617.3	y5+	-	-	Pilolli et al., 2014
		Chrysolophus amherstiae (9088) Coturnix japonica (93934) Lophophorus impejanus (9040) Lophura leucomelanos (140445) Meleagris gallopavo (9103) Pavo cristatus (9049) Phasianus colchicus colchicus (9057) Phasianus versicolor (9055) Syrmaticus soemmerringii (9067) Syrmaticus reevesii (9066) Tragopan satyra (9070) Tragopan temminckii (9071) Rattus norvegicus (10116) Pavo cristatus (9049) Phasianus colchicus (9054) Phasianus colchicus colchicus (9057) Gallus lafayetii (9032) Gallus sonneratii (9033) Alopochen aegyptiaca (30382)	Dual LIT	877.4 (+2)	489.3	y4+	617.3	y5+	730.4	y6+	De Angelis et al., 2017b
Vitellogenin-1	YLLDLLPAAASHR (lipovitellin-1 chain)	Gallus gallus (9031)	QqQ	480.6 (+3)	709.4	y7+	582.3	y11++	355.2	y7++	Planque et al., 2016
			QqQ	480.6 (+3)	709.4	y7+	582.3	y11++	355.2	y7++	Planque et al., 2017b
			QqQ	480.6 (+3)	709.4	y7+	355.2	y7++	-	-	New et al., 2018
			QqQ	480.6 (+3)	709.4	y7+	582.3	y11++			Planque et al., 2019
	ALLLSEIR (lipovitellin-1 chain)	Gallus gallus (9031) Fundulus heteroclitus (8078)	QqQ	457.8 (+2)	617.4	y5+	730.5	y6+	-	-	New et al., 2018

Vitellogenin-2	NIPFAEYPTYK (lipovitellin-1 chain)	Galdieria sulphuraria (130081) Nicrophorus vespilloides (110193)									
Vitellogenin-2	NIPFAEYPTYK (lipovitellin-1 chain)	Gallus gallus (9031) Larus argentatus (35669) Anas platyrhynchos (8839) Cuculus canorus (55661) Buceros rhinoceros silvestris (175836) Charadrius vociferus (50402) Merops nubicus (57421) Pelecanus crispus (36300) Mesitornis unicolor (54374) Colinus virginianus (9014) Meleagris gallopavo (9103) Callipepla squamata (9009) Haliaeetus albicilla (8969) Mesitornis unicolor (54374)	QqQ QqQ QqQ	671.8 (+2) 671.8 (+2) 671.8 (+2)	1115.5 1115.5 1115.5	y_9^+ y_9^+ y_9^+	508.3 508.3 558.3	y_4^+ y_4^+ y_9^{++}	558.3 558.3 -	y_9^{++} y_9^{++} -	Planque et al., 2016 Planque et al., 2017b Planque et al., 2019
	NIGELGVEK (lipovitellin-1 chain)	Gallus gallus (9031) Larus argentatus (35669) Anas platyrhynchos (8839) Nipponia nippon (128390) Calypte anna (9244) Nestor notabilis (176057) Buceros rhinoceros silvestris (175836) Charadrius vociferus (50402) Pelecanus crispus (36300) Merops nubicus (57421) Aptenodytes forsteri (9233) Patagioenas fasciata monilis (372326) Phalacrocorax carbo (9209) Antrostomus carolinensis (279965) Chlamydotis macqueenii (187382) Opisthocomus hoazin (30419)	QqQ QqQ QqQ	479.8 (+2) 479.8 (+2) 479.8 (+2)	731.4 731.4 731.4	y_7^+ y_7^+ y_7^+	674.4 674.4 228.1	y_6^+ y_6^+ b_2^+	545.3 545.3 -	y_5^+ y_5^+ -	Planque et al., 2016 Planque et al., 2017b New et al., 2018
	LPLSLPVGPR (lipovitellin-2 chain)	Gallus gallus (9031) Hydroprogne caspia (425641) Taeniopygia guttata (59729) Sterna hirundo (108405) Falco sparverius (56350) Larus argentatus (35669) Meleagris gallopavo (9103) Nipponia nippon (128390) Nestor notabilis (176057) Callipepla squamata (9009) Phaethon lepturus (97097) Falco sparverius (56350) Buceros rhinoceros silvestris (175836) Charadrius vociferous (50402) Amazona aestiva (12930) Ficedula albicollis (59894) Merops nubicus (57421) Aptenodytes forsteri (9233) Hydroprogne caspia (425641)	QqQ	524.8 (+2)	468.3	y_9^{++}	725.4	y_7^+	-	-	New et al., 2018

Allergen	Peptide	Species	MS	Precursor m/z (charge)	Q1	ion	Q2	ion	Q3	ion	Reference
		Patagioenas fasciata monilis (372326)									
		Mesitornis unicolor (54374)									
		Balearica regulorum gibbericeps (100784)									
		Phalacrocorax carbo (9209)									
		Antrostomus carolinensis (279965)									
		Chlamydotis macqueenii (187382)									
		Colinus virginianus (9014)									
		Opisthocomus hoazin (30419)									
Ara h 1 | Cupin (Vicillin-Type, 7S Globulin)	DLAFPGSGEQVEK *Ara h 1.0101 (clone P41B) Ara h 1 - clone P17*	*Arachis hypogaea (3818) Arachis duranensis (130453)*	Q-ToF	688.9 (+2)	-		-		-		Chassaigne et al., 2007
			QqQ	688.8 (+2)	930.6	y_9^+	300.2	a_3^+	-		Heick et al., 2011a
			QqQ	688.8 (+2)	930.6	y_9^+	300.2	a_3^+	-		Heick et al., 2011b
			QqQ	688.8 (+2)	930.5	y_9^+	833.4	y_8^+	1077.5	y_{10}^+	Pedreschi et al., 2012
			QqQ	688.8 (+2)	930.5	y_9^+	833.4	y_8^+	776.4	y_7^+	Sayers et al., 2016
			QqQ	688.8 (+2)	930.5	y_9^+	833.4	y_8^+	776.4	y_7^+	Sayers et al., 2018
			QqQ	688.9 (+2)	930.5	y_9^+	300.2	b_3^+	229.1	b_2^+	Shefcheck et al., 2006
			QqQ	688.8 (+2)	930.5	y_9^+	300.2	b_3^+	-		New et al., 2018
			QqQ	688.8 (+2)	929.4	y_9^+	447.2	b_4^+	-		Zhang et al., 2019
	VLLEENAGGEQEER *Ara h 1.0101 (clone P41B) Ara h 1 - clone P17*	*Arachis hypogaea (3818) Arachis duranensis (130453)*	QqQ	786.9 (+2)	804.3	y_7^+	875.4	y_8^+	989.4	y_9^+	Pedreschi et al., 2012
			QqQ	786.9 (+2)	804.3	y_7^+	875.4	y_8^+	989.4	y_9^+	Sayers et al., 2016
			Dual LIT	786.9 (+2)	804.3	y_7^+	875.4	y_8^+	680.8	y_{12}^{++}	Pilolli et al., 2017a
			QqQ	786.9 (+2)	804.3	y_7^+	875.4	y_8^+	989.4	y_9^+	Sayers et al., 2018
			QqQ	786.9 (+2)	804.4	y_7^+	213.2	b_2^+	989.4	y_9^+	Shefcheck et al., 2006
			Q-Orbitrap	786.8 (+2)	-		-		-		Pilolli et al., 2018
			QqQ	786.9 (+2)	804.4	y_7^+	1118.5	y_{10}^+	-		New et al., 2018
			Q-ToF	786.9 (+2)	-		-		-		Chassaigne et al., 2007
			LIT-Orbitrap	786.9 (+2)	-		-		-		Korte et al., 2016°
	GTGNLELVAVR *Ara h 1.0101 (clone P41B) Ara h 1 - clone P17*	*Arachis hypogaea (3818) Arachis duranensis (130453)*	QqQ	564.4 (+2)	686.6	y_6^+	557.5	y_5^+			Heick et al., 2011a
			QqQ	564.4 (+2)	686.6	y_6^+	557.5	y_5^+			Heick et al., 2011b
			QqQ	564.8 (+2)	686.4	y_6^+	557.4	y_5^+	799.5	y_7^+	Parker et al., 2015
			QqQ	564.8 (+2)	686.4	y_6^+	557.4	y_5^+	444.3	y_4^+	Gu et al., 2018
			QqQ	564.8 (+2)	686.4	y_6^+	557.4	y_5^+	799.5	y_7^+	Boo et al., 2018
			Q-Orbitrap	564.8 (+2)	-		-		-		Pilolli et al., 2018
Ara h 2 - Conglutin (2S Albumin)	**CC**NELNEFENNQR *Ara h 2.0101, Ara h 2.0201 Ara h 2.0202*	*Arachis ipaensis (130454) Arachis hypogaea (3818) Arachis duranensis (130453)*	QqQ	807.0 (+2)	1050.5	y_8^+	-		-		Careri et al., 2007
			QqQ								Careri et al., 2008
			QqQ	576.2 (+3)	920.3	b_7^+	807.4	y_6^+	660.3	y_5^+	Parker et al., 2015
			QqQ	863.8 (+2)	807.4	y_6^+	1050.5	y_8^+	660.3	y_5^+	Pedreschi et al., 2012
			QqQ	863.9 (+2)	1163.5	y_9^+	1050.5	y_8^+	936.4	y_7^+	Sayers et al., 2016
			QqQ	863.8 (+2)	1163.5	y_9^+	1050.5	y_8^+	936.4	y_7^+	Sayers et al., 2018
			QqQ	863.8 (+2)	1163.5	y_9^+	1050.5	y_8^+	936.4	y_7^+	Boo et al., 2018
	NLPQQ**C**GLR *Ara h 2.0101, Ara h 2.0201 Ara h 2.0202*	*Arachis ipaensis (130454) Arachis hypogaea (3818) Arachis duranensis (130453)*	QqQ	543.3 (+2)	858.4	y_7^+	761.4	y_6^+	633.3	y_5^+	Parker et al., 2015
			QqQ	543.3 (+2)	858.4	y_7^+	761.4	y_6^+	633.3	y_5^+	Pedreschi et al., 2012
			QqQ	543.3 (+2)	858.4	y_7^+	761.4	y_6^+	633.3	y_5^+	Sayers et al., 2016
			QqQ	543.3 (+2)	858.4	y_7^+	761.4	y_6^+	633.3	y_5^+	Sayers et al., 2018
			QqQ	543.6 (+2)	858.2	y_7^+	429.8*	y_7^{++}	-		Vandekerckhove et al 2017
			QqQ	543.3 (+2)	858.4	y_7^+	429.7	y_7^{++}	633.3	y_5^+	Planque et al., 2017b
			QqQ	543.3 (+2)	858.4	y_7^+	761.4	y_6^+	633.3	y_5^+	Boo et al., 2018
			QqQ	543.3 (+2)	858.4	y_7^+	429.7	y_7^{++}	-		Planque et al., 2019
			QqQ	543.8 (+2)	858.4	y_7^+	633.3	y_5^+	-		Zhang et al., 2019
Gly m 5 - β-Conglycinin (Vicilin, 7S Globulin)	LITLAIPVNKPGR *Gly m 5.0101*	*Glycine max (3847) Glycine soja (3848)*	QqQ	464.7 (+3)	767.5*	y_7^+	583.4*	y_{11}^{++}	476.3*	y_9^{++}	Houston et al., 2011
			QqQ	464.6 (+3)	767.5	y_7^+	583.4	y_{11}^{++}	476.3	y_9^{++}	Planque et al., 2016
			QqQ	464.6 (+3)	767.5	y_7^+	583.4	y_{11}^{++}	476.3	y_9^{++}	Planque et al., 2017b
			Q-Orbitrap	464.63 (+3)	-		-		-		Chen et al., 2019

Allergen	Peptide	Species	Instrument	Precursor m/z	Q1	Ion	Q2	Ion	Q3	Ion	Reference
	QQQEEQPLEVR *Gly m 5.0201*	*Glycine max (3847)* *Glycine soja (3848)*	QqQ Q-Orbitrap	692.3 (+2) 692.34 (+2)	1127.6 -	y9+ -	870.5 -	y7+ -	613.4 -	y5+ -	Gu et al., 2018 Chen et al., 2019
	DSYNLQSGDALR *Gly m 5.0201*	*Glycine max (3847)* *Glycine soja (3848)*	QqQ	669.8 (+2)	859.5	y8+	746.4	y7+	618.3	y6+	Gu et al., 2018
	DSYNLHPGDAQR *Gly m 5.0301, Gly m 5.0302*	*Glycine max (3847)*	QqQ	458.2 (+3)	643.3	y6+	546.3	y5+	374.2	y3+	Gu et al., 2018
Gly m 6 - Glycinin (Legumin, 11S Globulin)	VFDGELQEGR *Gly m 6.0101*	*Glycine max (3847)*	QqQ QqQ QqQ QqQ QqQ Q-Orbitrap	575.2 (+2) 575.2 (+2) 575.3 (+2) 575.3 (+2) 575.3 (+2) 575.28 (+2)	903.2 903.2 903.2 788.4 788.4 -	y8+ y8+ y8+ y7+ y7+ -	219.2 219.2 489.2 602.3 602.3 -	a2+ a2+ y4+ y5+ y5+ -	788.5 789.4 - -	y7+ b7+ - -	Heick et al., 2011b Heick et al., 2011a Huschek et al., 2016 Planque et al., 2017b Planque et al., 2019 Chen et al., 2019
	SQSDNFEYVSFK *Gly m 6.0101, Gly m 6.0201, Gly m 6.0301*	*Glycine max (3847)*	QqQ QqQ Dual LIT QqQ Q-Orbitrap	725.7 (+2) 725.7 (+2) 725.8 (+2) 725.8 (+2) 725.83 (+2)	381.2 381.2 381.2 772.4 -	y3+ y3+ y3+ y6+ -	1235.4 1235.4 716.8 643.3 -	y10+ y10+ [M+2H]++−H2O y5+ -	1235.4 480.3 -	y10+ y4+ -	Heick et al., 2011a Heick et al., 2011b Pilolli et al., 2017a Gu et al., 2018 Chen et al., 2019
	ISTLNSLTLPALR *Gly m 6.0401, Gly m 6.0501*	*Glycine max (3847)* *Glycine soja (3848)* *Glycine microphylla (45693)*	QqQ QqQQqQ Q-Orbitrap	699.9 (+2) 699.9 (+2) 699.9 (+2) 699.92 (+2)	984.6 984.6 1097.7 -	y9+ y9+ y10+ -	870.5 870.5 670.4 -	y8+ y8+ y6+ -	783.5 783.5 456.4 -	y7+ y7+ y4+ -	Planque et al., 2016 Planque et al., 2017b Montowska et al., 2019 Chen et al., 2019
	FYLAGNQEQEFLK *Gly m 6.0101, Gly m 6.0201*	*Glycine max (3847)*	Dual LIT QqQ	793.9 (+2) 793.9 (+2)	638.8 638.8	y11++ y11++	1163.6 424.2	y10+ b3+	1092.5 283.1	y9+ a2+	Pilolli et al., 2017a Hoffmann et al., 2017
Cor a 9 - 11S Seed Storage Globulin (Legumin-Like)	INTVNSNTLPVLR *Cor a 9.0101 (Q8W1C2) Cor a 9 (A0A0A0P7E3)*	*Corylus avellana (13451)*	QqQ QqQ QqQ QqQ QqQ QqQ LIT-Orbitrap	720.9 (+2) 720.9 (+2) 721.1 (+2) 721.1 (+2) 720.9 (+2) 720.9 (+2) 720.9 (+2)	1013.6 1013.6 1013.7 1013.7 1013.6 1013.6 -	y9+ y9+ y9+ y9+ y9+ y9+ -	484.4 484.4 484.4 484.4 899.5 484.3 -	y4+ y4+ y4+ y4+ y8+ y4+ -	228.2 228.2 812.5 - -	b2+ b2+ y7+ - -	Heick et al., 2011a Heick et al., 2011b Ansari et al., 2012 Costa et al., 2014 Planque et al., 2017b New et al., 2018 Korte et al., 2016°
	ALPDDVLANAFQISR *Cor a 9.0101 (Q8W1C2) Cor a 9 (A0A0A0P7E3)*	*Corylus avellana (13451)*	QqQ QqQ QqQ QqQ Dual LIT Q-Orbitrap QqQ QqQ	815.5 (+2) 815.5 (+2) 815.6 (+2) 815.6 (+2) 815.4 (+2) 815.4 (+2) 815.4 (+2) 815.4 (+2)	906.6 906.6 906.5 906.5 906.5 - 906.5 906.5	y8+ y8+ y8+ y8+ y8+ - y8+ y8+	1019.5 1019.5 185.2 185.2 835.4 - 1019.6 1019.6	y9+ y9+ b2+ b2+ y7+ - y9+ y9+	175.2 175.2 723.4 - 723.4	y1+ y1+ y13++ - y13++	Heick et al., 2011a Heick et al., 2011b Ansari et al., 2012 Costa et al., 2014 Pilolli et al., 2017a Pilolli et al., 2018 Planque et al., 2017b Planque et al., 2019
	ADIYTEQVGR *Cor a 9.0101 (Q8W1C2) Cor a 9 (A0A0A0P7E3)*	*Corylus avellana (13451)*	QqQ QqQ LIT LIT Dual LIT QqQ QqQ Q-Orbitrap QqQ LIT-Orbitrap QqQ	576.3 (+2) 576.3 (+2) 577.0 (+2) 577.0 (+2) 576.3 (+2) 576.3 (+2) 576.3 (+2) 576.3 (+2) 576.3 (+2) 576.3 (+2) 576.3 (+2)	689.4 689.4 689.0 689.0 689.4 689.4 689.4 - 689.4 - 689.4	y6+ y6+ y6+ y6+ y6+ y6+ y6+ - y6+ - y6+	852.5 852.5 567.0 567.0 567.3 852.4 852.4 - 852.4 - 588.3	y7+ y7+ [M+2H]++−H2O [M+2H]++−H2O [M+2H]++−H2O y7+ y7+ - y7+ - y5+	- - - 852.5 588.3 588.3 - - -	y7+ y5+ y5+ - -	Heick et al., 2011a Heick et al., 2011b Bignardi et al., 2010 Bignardi et al., 2013 Pilolli et al., 2017a Planque et al., 2017b Gu et al., 2018 Pilolli et al., 2018 New et al., 2018 Korte et al., 2016a Planque et al., 2019
	QGQVLTIPQNFAVAK *Cor a 9.0101 (Q8W1C2) Cor a 9 (A0A0A0P7E3)*	*Corylus avellana (13451)*	QqQ QqQ QqQ QqQ QqQ Q-Orbitrap	807.5 (+2) 807.5 (+2) 807.8 (+2) 807.8 (+2) 807.5 (+2) 807.5 (+2)	1088.6 1088.6 874.6 874.6 1088.6 -	y10+ y10+ y8+ y8+ y10+ -	874.6 874.6 186.2 186.2 437.7 -	y8+ y8+ b2+ b2+ y8++ -	314.2 314.2 - -	b3+ b3+ - -	Heick et al., 2011a Heick et al., 2011b Ansari et al., 2012 Costa et al., 2014 Korte & Brockmeyer, 2016b Pilolli et al., 2018

			QqQ	807.5 (+2)	1088.6	y_{10}^+	987.6	y_9^+	874.5	y_8^+	Planque et al., 2017b
			LIT-Orbitrap	807.5 (+2)	-	-	-	-	-	-	Korte et al., 2016a
			QqQ	807.5 (+2)	1088.6	y_{10}^+	874.5	y_8^+			Planque et al., 2019
			QqQ								Korte et al. 2019
Cor a 11 - 7S Seed Storage Globulin (Vicilin-Like)	LLSGIENFR *Cor a 11.0101*	*Corylus avellana (13451)*	QqQ	524.9 (+2)	822.4	y_7^+	565.4	y_4^+	199.3	a_2^+	Ansari et al., 2012
			QqQ	524.9 (+2)	822.4	y_7^+	565.4	y_4^+	199.3	a_2^+	Costa et al., 2014
			QqQ	524.8 (+2)	822.4	y_7^+	227.2	b_2^+			New et al., 2018
	AFSWEVLEAALK *Cor a 11.0101*	*Corylus avellana (13451)*	QqQ	682.7 (+2)	644.4	y_6^+	402.3	y_4^+	191.3	a_2^+	Ansari et al., 2012
			QqQ	682.7 (+2)	644.4	y_6^+	402.3	y_4^+	191.3	a_2^+	Costa et al., 2014
			QqQ	682.4 (+2)	872.5	y_8^+	743.5	y_7^+			New et al., 2018
Pru du 6 - Amandin, 11S Globulin Legumin-Like Protein	QQGQQEQQQER *Pru du 6.0101*	*Prunus dulcis (3755)*	LIT	694.0 (+2)	677.0		685.0	$[M+2H]^{++}-H_2O$	-	-	Bignardi et al., 2010
	TDE**NG**FTNTLAGR *Pru du 6.0201*	*Prunus dulcis (3755)*	QqQ	698.3 (+2)	936.5	y_9^+	879.5	y_8^+	732.4	y_7^+	Planque et al., 2017b
			LIT-Orbitrap	698.3 (+2)	-	-	-	-	-	-	Korte et al., 2016a
			QqQ	698.3 (+3)	879.5	y_8^+	732.4	y_7^+	-	-	Planque et al., 2019