



Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma



Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry

M. Gavage^{a,b,*}, K. Van Vlierberghe^c, C. Van Poucke^c, M. De Loose^c, K. Gevaert^{d,e},
M. Dieu^{b,f}, P. Renard^b, T. Arnould^b, N. Gillard^a

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

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

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

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

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

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

ARTICLE INFO

Article history:

Received 3 August 2018

Received in revised form 5 November 2018

Accepted 18 November 2018

Available online xxx

Keywords:

High-resolution mass spectrometry

Allergen

Egg

Processed food products

Peptide biomarker selection

ABSTRACT

Food allergy is a growing health problem worldwide; thus, there is an urgent need for robust, specific, and sensitive analytical methods for detecting allergens. Mass spectrometry is an alternative to the existing methods, and it can overcome their limitations. One of the first steps in the development of any analytical method is the identification of the analytes to be further studied. In the case of allergen detection by mass spectrometry, the analytes are peptides. In this study, a strategy was developed for identifying potential peptide biomarkers in processed food products. This strategy was applied to processed egg matrices, and 16 potential peptide biomarkers were identified for the further detection and quantification of egg by means of mass spectrometry. With an empirical approach based on dedicated sample preparation, including tandem Lys-C/trypsin enzymatic digestion and high-resolution mass spectrometry analysis, hundreds of peptides from egg proteins were identified. This list of peptides was further refined with a series of criteria, obtained from empirical evidence, to identify the ideal biomarkers for the development of a quantitative method. These criteria include the resistance to food processing and the specificity of the peptides for eggs but also the effects of amino acid modifications and enzymatic digestion efficiency.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Food allergy is defined by the National Institute of Allergy and Infectious Diseases (NIAID) as “an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food.” It is a major global issue, and children, in particular, are being affected in increasing numbers [1]. Symptoms can affect multiple biological systems like the skin (hives), or the gastrointestinal (vomiting and/or stomach cramps) or respiratory (wheezing, hoarse throat) systems. They can become life-threatening in the case of anaphylaxis simultaneously affecting different systems [2]. Because there is no cure yet, the only solution for allergic patients is strict avoidance of exposure to food allergens [3]. To protect consumers, 14 ingredients that have the potential to cause allergic reactions are required by European leg-

islation (Regulation 1169/2011/CE) to be included in food labelling. Unfortunately, the incidental presence of allergens in food is possible, e.g., through cross-contamination, and this is not covered by legislation. Hence, specific and sensitive methods for the routine detection of food allergens are needed.

Some detection methods are commercially available, and they are based mainly on immunological methods targeting specific proteins, such as enzyme-linked immunosorbent assays (ELISA). Other methods rely on polymerase chain reactions (PCRs) targeting specific DNA molecules. However, such methods have limitations. For instance, with the PCR method, milk cannot be distinguished from cattle products nor egg from other chicken parts because they share the same DNA sequence. Furthermore, while the DNA is targeted, the proposed thresholds are expressed as total protein quantities; thus, the identification of the required conversion factor could be challenging [4]. The antibodies used in immunological methods recognize epitopes. One drawback, however, is the possible recognition by the antibodies of similar structures in non-allergenic ingredients, thus generating false positive results.

* Corresponding author at: CER Groupe, Rue du Point du Jour, 8, 6900, Marloie, Belgium.

E-mail address: n.gillard@cergroupe.be (M. Gavage).

In contrast, protein denaturation upon food processing can modify the conformation of the targeted proteins or epitopes, leading to false negative results [5].

Mass spectrometry (MS)-based methods may overcome these drawbacks. Typically, proteins are detected and quantified through their constitutive peptides, which are obtained after proteolytic digestion. Such an analysis is therefore based on peptide sequences rather than protein structures. Thus, MS-based methods are more robust than other methods. However, MS may be subject to matrix effects, defined as the combined effects of all the components of the sample other than the analyte on the measurement of the quantity of this specific analyte [6]. These matrix effects have to be minimized by dedicated sample preparation. The overall specificity of MS-based methods is ensured by the specificity of both the selected peptides and the multiple-reaction monitoring (MRM) transitions [7]. In addition, MRM is highly sensitive, and it is compatible with the detection of multiple analytes in a single experiment. Therefore, with MRM, multiple allergenic ingredients can be detected and quantified in a single analysis for a particular sample [8,9].

As was mentioned, with MS-based methods, the detection and the quantification of allergenic ingredients are performed by analyzing the constitutive peptides of the proteins from these food ingredients. The selection of these peptides is the first step in the development of the analytical method. This can be done with an *in silico* approach using software, e.g., Skyline[®], to generate MRM transitions that will be assessed by analyzing the samples with a triple quadrupole mass spectrometer (QqQ) [10].

In this study, peptides were selected with an empirical approach based on high-resolution mass spectrometry (HRMS). Such an untargeted approach allowed for the selection of the peptides, taking into account peptide-related information such as natural and artificial amino acid modifications, and proteolytic digestion efficiency through the detection of incompletely digested peptides. This important information was not readily available with the *in silico* approaches mentioned above. In practice, proteins are extracted from matrices that may or may not be contaminated with allergenic material. After proteolytic digestion, the resulting peptides are analyzed by HRMS. The obtained data are then used to select peptide biomarkers for the development of the actual quantitative detection method. To achieve this goal, a digestion protocol combining two enzymes, Lys-C and trypsin, was implemented because it was observed in preliminary tests that digestion using only trypsin was inefficient for some egg proteins.

It is clear that peptide selection is a crucial step in the development of a quantitative MS-based method. These peptides have to fulfil a series of criteria to ensure method accuracy, robustness, and precision [11–13]. Ideal peptide biomarkers have to be specific for the allergenic ingredient, robust to food processing, derivable from abundant proteins, size-compatible with *m/z*-analyzers (8–25 amino acids), and not prone to amino acid modifications or missed cleavages by the protease(s) used.

This study is part of a larger project aimed at developing MS-based methods for the quantification of four main allergens—egg, milk, peanut and hazelnut—in processed food products. This particular study was focused on egg, but the developed strategy should be applicable for detecting any allergenic ingredient. However, this strategy depends on the availability of protein databases. This could be a limitation for less studied allergenic ingredients like pecan (*Carya illinoensis*) with only 45 entries in UniProt protein database, for the moment.

Egg allergy is one of the most common food allergies, especially in children [14,15]. In addition, egg and its derived components are used a great deal by the food industry, thus making the complete avoidance of eggs difficult. To identify robust and specific peptides for egg, egg-based reference materials were prepared in a standardized way. The analytical criteria were also identified

and applied to the final selection of egg peptides by UHPLC-HRMS.

2. Material and methods

2.1. Reagent and materials

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

2.2. Food product preparation

Processed egg matrices were produced in the Food Pilot unit of the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO). The preparation of the egg matrices started with homogenized raw liquid whole egg (12.49% protein content) or homogenized pasteurized liquid whole egg (12.49% protein content). Neither egg product contained any additives. The eggs were purchased from chicken farms engaged in cage farming (Lodewijkx NV, Veerle, Belgium, egg code 3).

2.2.1. Preparation of a raw egg matrix

Raw whole liquid egg was freeze-dried upon arrival (freeze dryer Epsilon 2–10 D LSC, Martin-Christ, Osterode am Harz, Germany), vacuum packed (vacuum packaging machine KN5, VC999, North Kansas City, Missouri, USA), and stored at 4 °C in the dark.

2.2.2. Preparation of a heated egg matrix

Pasteurized whole liquid egg was freeze-dried upon arrival (freeze dryer Epsilon 2–10 D LSC), vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark. Egg pasteurization fulfils USDA requirements indicating that whole eggs must reach a minimum temperature of 60 °C for 3.5 min [16].

2.2.3. Preparation of fatty rich environment egg matrix: type 1 = mayonnaise

A volume of 450 mL sunflower oil (Carrefour, Boulogne-Billancourt, France) was added slowly to 100 g raw liquid whole egg while being mixed (Bosch immersion blender MSM67170–750 W, Stuttgart, Germany) until a smooth mayonnaise was formed. The mayonnaise was divided into 50 g aliquots, vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark.

2.2.4. Preparation of fatty rich environment egg matrix: type 2 = chocolate

First, a fine egg powder was produced by mixing 10 g freeze-dried raw egg powder into a fine powder (Kenwood Mini Chopper Mill Attachment AT320B [Hampshire, United Kingdom], pulsing speed 3 min) followed by the sieving of this powder (Retsch test

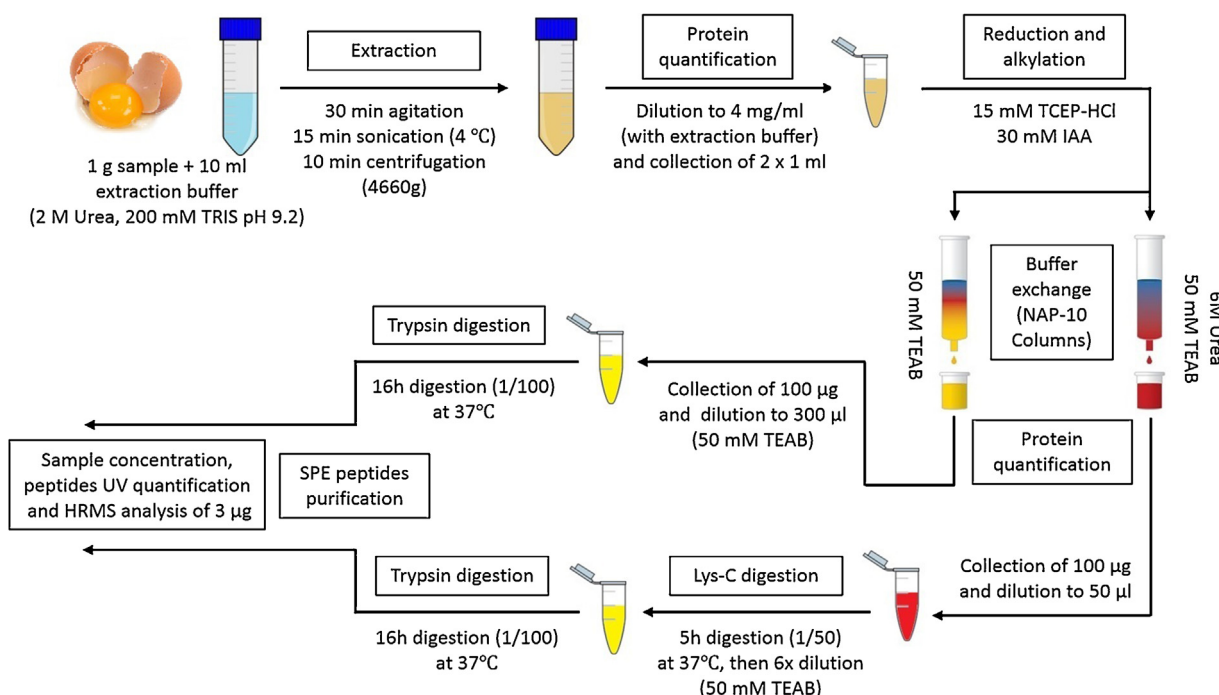


Fig. 1. Sample preparation flowchart. Proteins from a 1 g sample were extracted with 10 mL extraction buffer. The proteins were then reduced and alkylated. The buffer was exchanged with an appropriate buffer depending on the subsequent proteolytic digestion (trypsin alone or tandem Lys-C/trypsin). The resulting peptides were purified with the SPE before HRMS analysis was performed.

sieve [Haan, Germany], pore size 500 µm). Chocolate was then prepared. First, a “conche” was made from chocolate refiner flakes (90%) and cacao butter (10%). An equivalent of 450 g chocolate refiner flakes (Callebaut, Lebbeke, Belgium) and 50 g cacao butter (Callebaut) were warmed in a double boiler in separate bowls until a temperature of 40–45 °C was reached. The temperature was controlled with a food thermometer. The melted butter was then added slowly to the melted chocolate refiner flakes, and the mix was stirred for 3 min with a spoon. Next, 1.2% ammonium phosphatide (Palsgaard A/S; Juelsminde, Denmark) was added, and the mixture was stirred for 3 more minutes. Finally, a chocolate incurred with egg (45,454 ppm egg proteins in matrix) was produced by adding 50 g egg powder to the liquid chocolate and stirring again for 3 min while maintaining the temperature of the chocolate at 42 °C. The chocolate was subsequently poured into sterile chocolate molds, covered with plastic foil, and left to cool at room temperature. Once it had cooled, the chocolate was kept overnight at 4 °C to cool further, to set, and to solidify. The next day, the chocolate bars were rasped manually into chocolate flakes in 50 g aliquots to avoid the melting of the chocolate. Chocolate flakes were kept in plastic jars in 50 g aliquots and stored at 4 °C in the dark. A proper blank chocolate was prepared following the same procedure without adding the egg powder.

2.2.5. Preparation of a fatty and low pH egg matrix: mayonnaise containing vinegar

A volume of 450 mL sunflower oil (Carrefour, Boulogne-Billancourt, France) was slowly added to 100 g raw liquid whole egg while being mixed (Bosch immersion blender MSM67170–750 W) until a smooth mayonnaise was formed. Next, 50 g vinegar (De Blauwe Hand, Puurs, Belgium) was mixed with the mayonnaise for 3 min (Bosch immersion blender MSM67170–750 W). The mayonnaise was divided into 50 g aliquots, vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark.

2.3. Sample preparation

The sample preparation protocol was based partly on the protocol of Planque et al. [10]. In short, proteins were extracted from food matrices, purified, and proteolytically digested. Two digestion protocols were compared for each processed food product. In the first protocol, the proteins were digested with trypsin alone. In the second, a tandem Lys-C/trypsin digestion was used. For each combination of digestion protocol and processed food product, three biological replicates were prepared and analyzed by UHPLC-HRMS (Fig. 1).

2.3.1. Protein extraction

The proteins from a 1 g sample were extracted in 50 mL conical centrifuge tubes with 10 mL extraction buffer (2 M urea, 200 mM Tris–HCl; pH 9.2) by shaking at 20 °C for 30 min (Agitelec, Toulouse, Mery-sur-Oise, France) followed by sonication (Elmasonic S 70, Elma Schmidbauer GmbH, Singen, Germany) for 15 min at 4 °C. The samples were centrifuged at 4660 g for 10 min at 4 °C. The middle liquid phase was then collected, and its protein concentration was measured using the Pierce 660 nm Assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

2.3.2. Protein reduction, alkylation and purification

The samples were diluted to a 4 mg/mL protein concentration with extraction buffer. The proteins from a 1 mL aliquot were reduced and alkylated using TCEP (15 mM final concentration) and IAA (30 mM final concentration); they were then mixed (500 rpm) for 15 min at 37 °C in the dark. The samples were desalted and buffer-exchanged by gel filtration chromatography (NAP-10 columns, GE Healthcare, United Kingdom). The protease used for protein digestion in the next step determined the nature of the exchange buffer. A 50 mM TEAB buffer was employed for trypsin only digestion; 6 M urea and 50 mM TEAB were employed for the tandem Lys-C/trypsin digestion. The protein concentration of the

Table 1

Unique proteins and number of unique peptides for these proteins identified with HRMS analysis of egg samples (raw egg, heated egg, fatty rich environment, and low pH matrix). These unique peptides and proteins are identified in at least one sample using the following search parameters: "trypsin digestion", "allow up to three missed cleavages", "masses of monoisotopic peptide ions with a tolerance of 7 ppm", and "a fragment tolerance of 0.05 Da." Carbamidomethylation of the cysteines was set as a fixed modification. Oxidation of the methionine, conversion of the N-terminal glutamine into pyroglutamate, and deamidation of the asparagine and glutamine were allowed as variable modifications. The MS/MS-based protein and peptide identifications were validated using Scaffold 4.8. The peptide identification was accepted if a peptide could be established at >95% probability by the Peptide Prophet algorithms. The proteins were identified assuming a false discovery rate of 1% and accepted if they contained at least two identified peptides.

Allergenic ingredient	Identified protein	UniProt entry identifier	Number of identified unique peptides	Coverage
Egg white	Alpha-1-acid glycoprotein	A7UEB0	6	27%
	Clusterin	Q9YGP0	6	10%
	Hep21 protein	Q8AV77	3	34%
	Lysozyme C (Gal d 4)	P00698	22	81%
	Mucin-5B	Q98UI9	24	12%
	Ovalbumin (Gal d 2)	P01012	36	59%
	Ovalbumin-related protein X	R9TNA6	14	34%
	Ovalbumin-related protein Y	P01014	24	34%
	OvoglobulinG2	I0J170	6	10%
	Ovomucoid (Gal d 1)	P01005	5	41%
	Ovostatin	P20740	6	3%
	Ovotransferrin (Gal d 3)	P02789	86	71%
	Apolipoprotein B	F1NV02	166	27%
	Apovitellenin-1	P02659	4	33%
Egg yolk	Ig lambda chain C region	P20763	4	20%
	Serum albumin (Gal d 5)	P19121	38	33%
	Vitellogenin-1 (Gal d 6)	P87498	81	33%
	Vitellogenin-2	P02845	144	54%

1.5 mL volume samples was again measured with the Pierce 660 nm Assay.

2.3.3. Proteolytic digestion

Two types of proteolytic digestion were used. For the trypsin-only digestion, the samples were diluted with 50 mM TEAB buffer to a 0.33 mg/mL protein concentration, and 300 μ L of these samples was transferred to a 1.5 mL microcentrifuge tube. A 16 h digestion at 37 °C under 300 rpm agitation (Stuart® Microtiter plate shaker incubator SI505, Bibby Scientific, Stone, United Kingdom) was performed after 1 μ g of trypsin (protein:trypsin ratio 1:100) was added. For the tandem Lys-C/trypsin digestion, the samples were diluted with 6 M urea and 50 mM TEAB buffer to 2 mg/mL protein concentration; 50 μ L was transferred to a 1.5 mL microcentrifuge tube. A first 5 h digestion was performed at 37 °C under 300 rpm agitation by adding 2 μ g of Lys-C. The samples were then diluted to a 300 μ L final volume (to reduce the urea concentration to 1 M). A second 16 h digestion was performed at 37 °C under 300 rpm agitation by adding 1 μ g of trypsin. For both digestion types, the reaction was stopped by the addition of 10 μ L of 33% TFA.

To minimize potential carbamylation of proteins and peptides, urea solutions were freshly prepared. Extraction and digestion steps were also performed during the same day.

2.3.4. SPE purification

The peptides resulting from the digestion were purified using C18 SPE cartridges. The cartridges were first washed with 1 mL acetonitrile/water (50/50, v/v) and equilibrated with 1 mL wash solvent (acetonitrile/water/TFA, 2/97.9/0.1, v/v/v). The samples were diluted 10 times with water, loaded onto the cartridges, and washed with 1 mL wash solvent. The peptides were eluted with 2 mL acetonitrile/water (80/20, v/v). The samples were subsequently evaporated under a nitrogen flow at 40 °C. The pellets were dissolved in 500 μ L of 0.1% formic acid and concentrated approximately 10 times with a vacuum concentrator (Heto, Denmark).

2.4. Mass spectrometry

The peptides were analyzed using an ESI-MS/MS maXis Impact UHR-TOF (Bruker, Billerica, Massachusetts, USA) coupled to a Dionex UltiMate 3000 Standard LC Systems (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

The peptide concentrations were measured using NanoDrop ultra-violet absorbance [17] (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Each injected sample volume was adapted such that each time, 3 μ g peptides were analyzed. The peptides were separated by reverse-phase liquid chromatography on an Acclaim PepMap100 C18 column (3 μ m, 100 Å, 1 mm $\times$ 15 cm nanoViper, Thermo Fisher Scientific, Waltham, Massachusetts, USA) using the Dionex UltiMate 3000 liquid chromatography system. The column temperature was set at 40 °C and A and B mobile phases were water and acetonitrile, respectively, each containing 0.1% formic acid. The solvent gradient and the flow rate (FR) were set as follows: 0–3 min: FR 30 μ L/min and 5% B; 3–40 min: FR 30 μ L/min and 5%–40% B; 40–45 min: FR 40 μ L/min and 40%–95% B; 45–50 min: FR 50 μ L/min and 95% B; 50–50.1 min: FR 30 μ L/min and 95% to 5% B; and 50.1–60 min: FR 30 μ L/min and 5% B. After each sample analysis, stringent washing was performed using water/acetonitrile/methanol/2-propanol (1/1/1/1) for 20 min to elute the residual components. Peptides of interest are mainly eluted within the 3%–40% acetonitrile gradient. But, to avoid contamination between the different samples, this strategy, including variable flow rate and stringent wash, was developed to elute the stickier peptides that remained on the liquid chromatography column.

The column effluent was connected to an electrospray ionization source. The MS and MS/MS data acquisition was carried out through 3 s cycles. Each cycle started with a 0.5 s survey scan during which MS spectra were acquired in a mass-to-charge (m/z) range of 150–2200. The most intense peptide ions containing 2–4 positive charges were selected for fragmentation during the remaining 2.5 s. The collision-induced dissociation (CID) energy was automatically set according to the m/z ratio and the charge state of the precursor ion. Dynamic exclusion is used in the acquisition method: after a first spectrum acquisition, the ion is excluded during 2 min except if its intensity increases 3 times. The mass spectrometer and liquid chromatography systems were piloted by Compass Hystar 3.2 (Bruker, Billerica, Massachusetts, USA).

2.5. Data analysis

The peak lists were generated using DataAnalysis 4.2 (Bruker, Billerica, Massachusetts, USA). They were saved as MGF files. To

improve mass accuracy, the data were processed with a lock mass calibration [hexakis(1 h,1 h,4h-hexafluorobutyl)phosphazine, principal positive ion at $m/z = 1221.9906$]. The MGF files were then treated with ProteinScape 3.1 (Bruker, Billerica, Massachusetts, USA) and Mascot 2.5 (Matrix Science, London, United Kingdom) as the search engine. The peak lists were searched against a home-made protein database downloaded from UniProt (uniprot.org). This database contains hen's proteome (*Gallus gallus*) and some possible contaminants (cacao, human keratin, trypsin and Lys-C) to avoid misidentification for a total of 77,682 entries (36,371 entries for hen, 40,959 for cacao, and 350 for human keratin). The following search parameters were used: "trypsin digestion", "allow up to three missed cleavages", "masses of monoisotopic peptide ions with a tolerance of 7 ppm", and "a fragment tolerance of 0.05 Da." Carbamidomethylation of the cysteines was set as a fixed modification. Oxidation of the methionine, conversion of the N-terminal glutamine into pyroglutamate, and deamidation of the asparagine and glutamine were allowed as variable modifications. Last, the peak lists were searched against the homemade database with an automatic decoy database search.

The MS/MS-based protein and peptide identifications were validated using Scaffold 4.8 (Proteome Software, Portland, Oregon, USA). The peptide identification was accepted if a peptide could be established at >95% probability by the Peptide Prophet algorithms [18,19]. The proteins were identified assuming a false discovery rate of 1% and accepted if they contained at least two identified peptides.

In parallel, the ProfileAnalysis (Bruker, Billerica, Massachusetts, USA) software was used to align the UHPLC-MS/MS raw data to compare the presence and the intensity of the peptides through the various samples. Using the "find molecular features" algorithm, the ions belonging to the same compound were combined into one feature (e.g., isotopes and charge states). The retention times of the obtained features were then aligned to rectify any potential retention shift that could appear during the run. Buckets were created, and a quantile normalization was applied. These buckets corresponded to the pairs of feature masses and the retention times. Comparable results were built by combining the list of peptides identified and validated by Scaffold and the buckets generated by ProfileAnalysis. The signal intensity linked to the identified peptides obtained this way was the basis for the results discussed below.

3. Results and discussion

3.1. Global strategy

Given the urgent need for a reference method for detecting and quantifying allergens in food [20], the objective of this study was to identify robust and specific egg peptide biomarkers that can be further used in a UHPLC-MS/MS method to quantify egg in processed food products. With the HRMS-based discovery proteomics approach used in this study, more than 600 egg peptides were identified in the samples. Unique identified proteins and the number of unique peptides for these proteins are listed in Table 1. The complete list of peptides associated with their intensity in each egg matrix can be found in Supplementary material 1. To ensure the sensitivity, specificity, robustness, and trueness of the future method, it was necessary that the selected analytical targets fulfil a series of criteria detailed in the following sections. The strategy was thus to filter, using these criteria, all of the identified peptides and to withhold the ideal biomarkers for the quantitative method.

3.2. Selection criterion 1: protein distribution in the egg

Eggs are composed of three main parts: egg white or albumen, yolk, and eggshell, which includes egg membrane. These three parts

are used together by the food industry, but they can also be found individually in some recipes. To be able to detect egg in any situation, this issue needed to be considered in the selection of the biomarkers. Eggshell, generally considered a waste product, can be used as a food additive because of its high calcium content (>90%) [21]. Eggshell also contains about 6% protein. These proteins correspond to eggshell-specific proteins and to proteins found in egg white. The eggshell membrane is composed mainly of collagen and egg white proteins [22–24]. Because both the eggshell and the eggshell membrane contain egg white proteins, such as ovalbumin or lysozyme, and no specific allergen has been identified in these parts [25], the selected biomarkers in this study did not include shell-specific peptides. In contrast, both the egg white and the yolk, which can be used individually by the food industry, contain specific allergens [26]; thus, the peptide biomarkers had to be selected from these two parts.

3.3. Selection criterion 2: the amino acid composition of targeted peptides

For the quantitative UHPLC-MS/MS method, the peptide biomarker selection has to comply with specific rules to ensure method trueness. Indeed, the total egg protein quantification is directly linked to the concentration of the selected peptides through their respective protein quantities in the egg. The targeted peptides should therefore be found in a single form: a single mass. Thus, the peptides containing amino acids that are prone to modification needed to be omitted because such peptides could be present in non-modified or modified versions in not very predictable proportions. Based on a review of the literature, potential modifications were empirically identified using Mascot's "error tolerant search." The identified modifications are summarized in Supplementary material 2.

Cysteine modification is intentionally introduced by the sample preparation. Indeed, cysteine are reduced and alkylated to disrupt disulfide bridges and increase protein denaturation before the enzymatic digestion step. TCEP and IAA used for this purpose induce cysteine carbamidomethylation. This modification was set as a fixed modification for peptide identification in the data analysis. This was done because of the observed high modification rate and the need to limit the number of modifications set as a variable. This in turn increased the database search time and reduced its specificity [27]. Nevertheless, the decision was made to reject peptides containing cysteine to prevent the effects of possible incomplete carbamidomethylation, which could induce a bias in the additional quantitative method.

The other possible amino acid modifications that were identified were methionine oxidation, asparagine and glutamine deamidation and N-terminal glutamine-to-pyroglutamate conversion. Therefore, the peptides containing cysteine, methionine, or N-terminal glutamine were systematically rejected.

Deamidation was treated differently, given the high occurrence of asparagine and glutamine [28]. Deamidation is typically a spontaneous reaction converting asparagine (Asn) to aspartic acid or isoaspartic acid and glutamine (Gln) to glutamic acid. The reaction rate is influenced by temperature and pH and also by the nature of the amino acid on the carboxyl side adjacent to either the asparagine or glutamine [29–32]. The fastest reaction occurs when this carboxyl side adjacent amino acid is a glycine (Gly). For other amino acids, the reaction rates are at least 10 times slower [29,31]; thus, only the peptides containing Asn-Gly or Gln-Gly sequences were rejected. Last, post-translational modifications must also be taken into account. For example, serine phosphorylation occurs in egg proteins, such as ovalbumin. The peptides known to contain such modified amino acids were rejected. This information

has been documented in proteomics databases, such as UniProt (uniprot.org).

To assess the importance of considering these peptide modifications for the MS quantitative analysis, a systematic analysis of the signal intensity of the vitellogenin-2 peptides either unmodified (hereafter called native peptides) or modified by methionine oxidation, glutamine-to-pyroglutamate conversion, or glutamine and asparagine deamidation was performed. As was shown in the graph in Supplementary material 3, the proportion of native and modified peptides was hardly predictable. Variations were also observed among the different modifications, and food processing also seemed to have an effect. For example, methionine oxidation was observed in mayonnaise more than in unprocessed eggs. However, no general trend was identified. Peptide modifications were observed in the various egg proteins and for the different food processes considered in this study without predictable ratios between the native and modified peptides. Therefore, the peptides containing the methionine, cysteine, N-terminal glutamine, or Asn-Gly or Gln-Gly sequences were systemically rejected to ensure method trueness.

3.4. Selection criterion 3: protease(s) used and cleavage site composition

In the protocol of Planque et al., as in many other studies, the proteins were enzymatically digested with trypsin. To improve the digestion of the trypsin-resistant proteins, the use of Lys-C in combination with trypsin was tested. Trypsin is the most used protease in bottom-up mass spectrometry-based proteomics [33,34]. The reason is its high catalytic activity and specificity in peptide bonds hydrolysis at the C-terminal side of the arginine and lysine residues. Tryptic peptides are ideal for mass spectrometry analysis (length, hydrophobicity and ionization). Lys-C can be used in combination with trypsin, given their partly overlapping specificities (cleavage at the Cterminal side of lysine). One interesting advantage of Lys-C is that it works efficiently under highly denaturing conditions, thus also limiting the risk of missed cleavages that otherwise form a source of the underestimation of the peptide quantitation [35,36].

As was described in the material and methods section, the proteins were extracted and digested with trypsin alone or with a Lys-C/trypsin combination. Both digestion protocols were applied parallel to each sample. As is shown in Fig. 2, the improvement associated with Lys-C/trypsin digestion varied according to the specific protein being considered. The peptide MS signal intensity following the tandem Lys-C/trypsin digestion was significantly higher (compared to the trypsin alone) for most of the proteins from the egg white (ovalbumin [Fig. 2A], ovomucoid, ovotransferrin [Supplementary material 1]), and for the egg yolk proteins (Apolipoprotein B, Apovitellenin-1 [Supplementary material 1]). For some proteins, however, the digestion efficiency improvement was less obvious. No improvement was observed for lysozyme (Fig. 2B); however, for other proteins, such as vitellogenin-1 or vitellogenin-2 (Fig. 2C), an improvement was observed only for select peptides.

On the contrary, the food processing applied to the samples had a limited effect on the efficiency of the enzymatic digestion. The proteins that were resistant to trypsin digestion but were more efficiently digested with the tandem Lys-C/trypsin exhibited the same behavior independent of the food processing. The same observation was reported for the proteins without trypsin digestion resistance, such as lysozyme, or for the partly resistant proteins (Supplementary material 1).

According to the classic rules for trypsin and Lys-C proteolysis, trypsin cleaves at the carboxyl side of lysine and arginine whereas Lys-C cleaves only after lysine unless a proline immediately follows the cleavage site for both enzymes [37,38]. Additional factors are known to affect the digestion efficiency and the missed cleavage

probability of a particular cleavage site. Because of the goal of using a quantitative UHPLC-MS/MS method, the cleavage of the selected peptide biomarkers had to be as complete as possible. Indeed, if a selected peptide is subjected to a higher possibility of missed cleavages, its absolute quantification will be biased, and the egg protein concentration will be underestimated.

Assuming optimized digestion conditions (pH, temperature, duration, etc.), the amino acid sequence surrounding the cleavage site has been found to affect peptide bond hydrolysis efficiency [37–41]. This information has to be taken into account for the peptide selection, and, whenever possible, peptides containing or surrounded by a resistant sequence have to be avoided. These sequences are described, using the nomenclature formulated by Schechter and Berger [42], as P4-P3-P2-P1-P1'-P2'-P3', in which cleavage of the peptide bond occurs between P1 and P1'. As was previously mentioned, both proteases had a lower digestion rate for sequences with a proline at position P1'. Lysine (K) and arginine (R) also had a negative effect on cleavage efficiency when present at position P1'. In other words, R–R, R–K, K–R and K–K were prone to missed cleavage events. Finally, the negatively charged amino acid residues, aspartate (D) and glutamate (E) or phosphorylated serine (pS) and threonine (pT) had a negative influence when present close to the cleavage site. The amino acid residues affecting digestion efficiency are summarized in Table 2.

Whereas P, K, and R at the P1' positions can be easily avoided without too much interference of the overall sensitivity of the method, totally avoiding D and E in the surrounding of the cleavage site is more cumbersome given their relatively high natural occurrence. This criterion was taken into account for peptide selection but not as an exclusive rule, otherwise too many peptides would be rejected. The peptides under consideration (i.e., those containing D and/or E in the area surrounding the cleavage site) were treated on a case-by-case basis, with the importance of the potential resulting missed cleavage evaluated. A peptide was rejected when the MS signal intensity of the associated peptide containing a missed cleavage exceeded 10% of the MS signal of the fully tryptic peptides.

3.5. Selection criterion 4: robustness of peptides to food processing

Eggs are usually consumed in the form of processed products. It is known that food processing conditions and matrix interactions affect the recovery and/or the further detectability of target molecules [43,44]. Such adverse conditions include heat treatment, the pH of the matrix, its fat content, and the presence of tannins. To ensure the robustness and the accuracy of the method, the effect of food processing was taken into account for the peptide biomarker selection. The matrices and the food processing applied to the samples containing egg were selected to represent the conditions found in everyday food products. Ideally, the peptides selected for quantitative UHPLC-MS/MS should not be affected.

The food processing effect was evaluated by comparing the MS signal intensity of the peptides under the various test conditions. Examples of vitellogenin-1 and ovotransferrin, egg yolk and egg white proteins, respectively, are shown in Fig. 3. In the case of vitellogenin-1 (Fig. 3A), some peptides, such as ATAVS-LLEWQR or NVNFDGEILK, seemed resistant to the selected food processes. These peptides are thus ideal biomarkers for the quantitative UHPLC-MS/MS method. In contrast, other peptides are affected by the food processing like low pH for peptides ALLSEIR or YLLDLLPAAASHR. The situation is different for ovotransferrin (Fig. 3B). This protein seems to be unaffected by food processing investigated in this study. All of the peptides were detected at comparable intensities; thus, they could be potential biomarkers. In this case, the selected peptide biomarkers are those with the highest signal intensity to ensure an overall high sensitivity of the

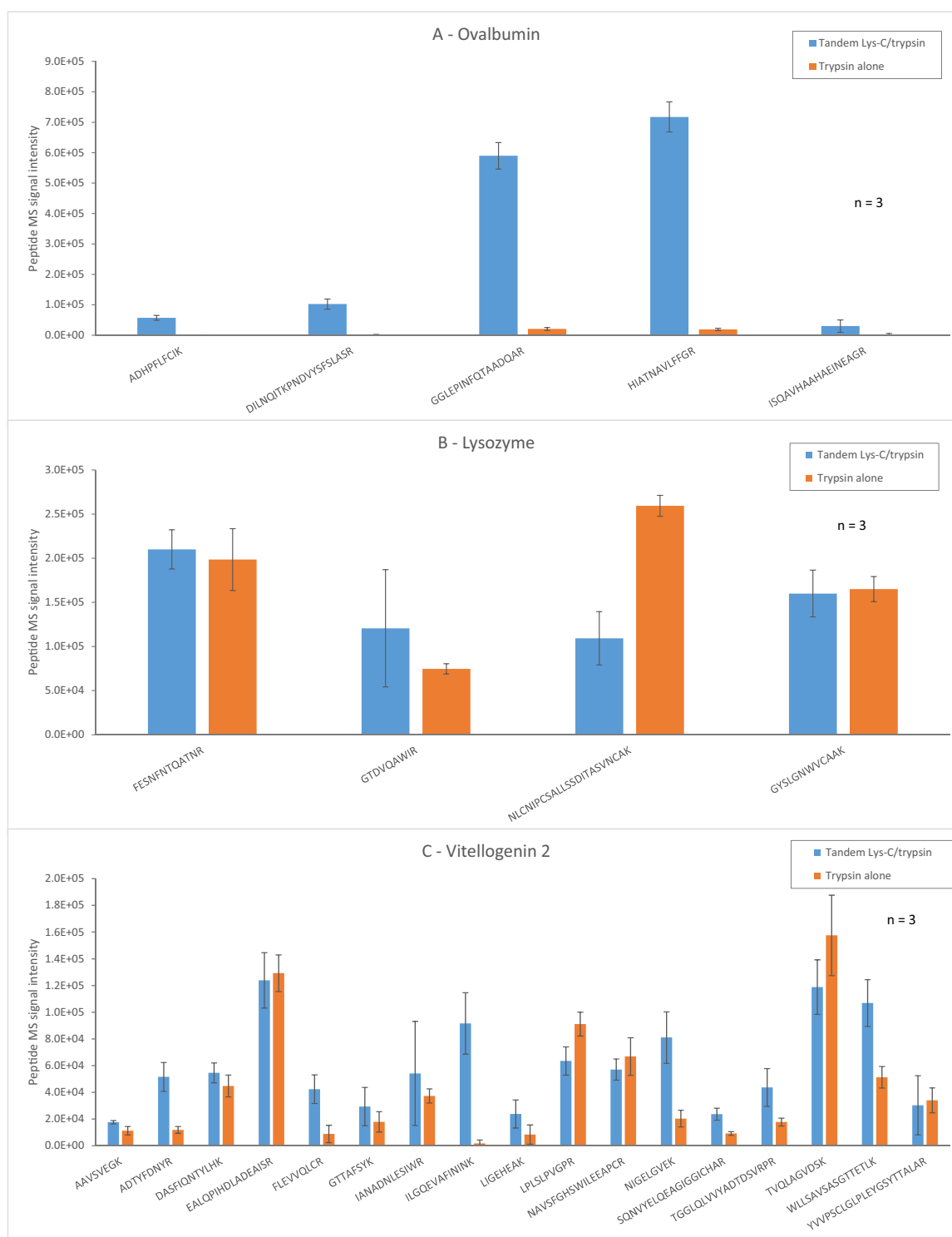


Fig. 2. Peptide MS signal intensity comparison between the two digestion protocols (tandem Lys-C/trypsin and trypsin alone) in unprocessed eggs for 3 proteins (A: ovalbumin; B: lysozyme; C: vitellogenin-2). Results are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 independent replicates.

Table 2

Summary of amino acids favoring missed cleavages by trypsin and Lys-C. The relative position of these amino acids to the cleavage site is shown using Schechter and Berger's nomenclature [46].

Relative position to cleavage site						
P4	P3	P2	P1	P1'	P2'	P3'
D E	D E	D E	–	D E K P R	D E	D E
pS pT	pS pT	pS pT		pS pT	pS pT	pS pT

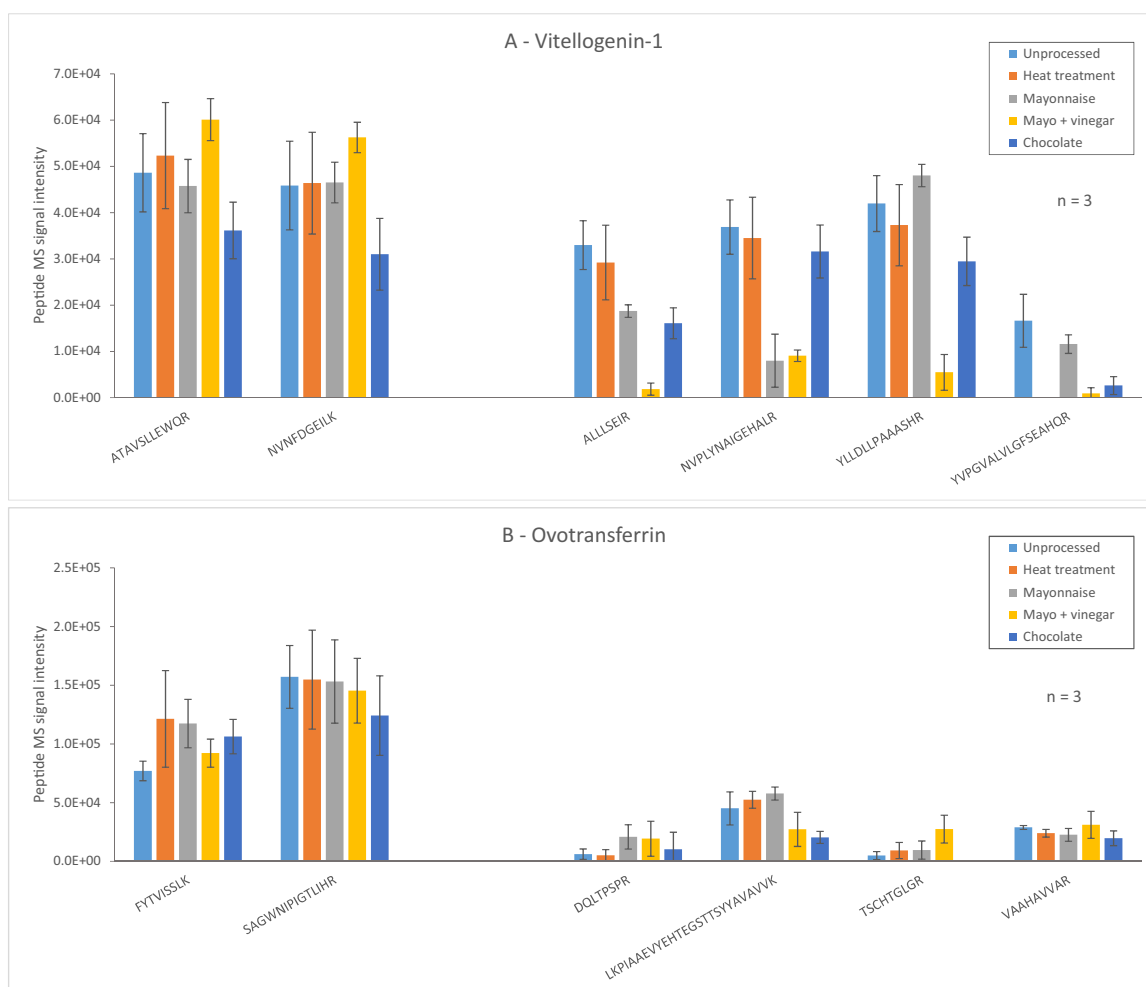


Fig. 3. Effect of food processing on a protein from egg yolk (A - Vitellogenin-1) and a protein from egg white (B - Ovotransferrin). Five matrices containing egg are compared. From left to right: no process, heated egg, fatty-rich environment (mayonnaise), acidic environment (mayo + vinegar), and the presence of tannins (chocolate). The comparison is based on the MS signal intensity. The ideal peptide biomarkers are located on the left of the figure. The results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 independent replicates.

UHPLC-MS/MS method. This should therefore allow the detection of low(er) amounts of egg proteins in processed food products.

Finally, the signal intensities of the peptides identified in the chocolate matrix were observed to be globally lower than those identified in the other matrices. This might be explained by matrix effects, such as the presence of tannins, which can interfere with sample preparation and HRMS analysis [9,45] or, simply, by the presence of cacao proteins. Indeed, proteins were extracted from the samples and digested, and 3 μ g of the obtained peptides were analyzed by HRMS without consideration of their origin. Because the chocolate matrix was the only matrix containing endogenous proteins, the signals associated with the egg peptides were lower. This assumption was strengthened by the identification of peptides from cacao proteins, such as Vicilin-A and the 21 kDa seed protein.

3.6. Final selection of egg peptide biomarkers

Starting from more than 600 egg peptides identified by HRMS, the selection criteria discussed in the previous sections were systematically applied to identifying the ideal biomarkers. A list of 16 peptides from 6 proteins (3 from egg white and 3 from egg yolk) was obtained (Table 3). These peptides are potential biomarkers and will be used to develop a UHPLC-MS/MS method to quantify egg in processed food products.

3.7. Peptide specificity

Egg is one of the most important components in the human diet. Usually, it is hen eggs that are consumed; however, 7% of the entire egg production is from other poultry species [46], such as goose, duck, turkey, quail, and guinea fowl. Serologic and clinical cross-reactivity between hen eggs and other poultry eggs have been reported [47,48]. Thus, patients with hypersensitivity to hen eggs may also react allergically to other poultry eggs. The main goal of the method used in this study was to detect hen eggs, but there is also a need to detect eggs from other farm birds. Indeed, the European legislation regarding labelling clearly states that “egg” refers to eggs from all farm birds (Commission Notice of 13 July 2017 relating to the provision of information on substances or products causing allergies or intolerances as listed in Annex II to Regulation [EU] No 1169/2011 of the European Parliament and of the Council on the provision of food information to consumers). On the contrary, the selected peptides may not be identified in the proteins of other food ingredients otherwise, false positive results could be obtained. The sequences have to be specific to hen eggs, but they could possibly be extended to the eggs of other farm birds.

Given the empirical approach used in this study, it can be asserted that the selected peptide sequences are found in hen egg proteins. Indeed, the selection of egg peptides was based on the hen’s proteome. The specificity of these sequences to the eggs

Table 3

List of egg peptide biomarkers that fulfil the above-mentioned criteria. The peptides are linked to their respective proteins (with their positions in the protein sequences) and egg part (white or yolk). The protein accession numbers in the UniProt database have been placed in brackets.

Allergenic ingredient	Protein	Selected egg peptide biomarkers
Egg White	Lysozyme (P00698)	52 FESNFTQATNR ₆₃
	Ovalbumin (P01012)	128 GGLEPINFQTAADQAR ₁₄₃ 371 HIATNAVLFFGR ₃₈₂ 141 SAGWNIPIGTLIHR ₁₅₄
	Ovotransferrin (P02789)	155 GAIEWEGIESGSVEQAVAK ₁₇₃ 279 VEDIWSFLSK ₂₈₈ 680 FYTVISLKLK ₆₈₈
	Apolipoprotein B (F1NV02)	257 HFLPSSYK ₂₆₅ 432 ASFYGLSHAVTK ₄₄₃ 725 ALFDYFGYSHDGK ₇₃₇ 840 FALSGIVTPGAK ₈₅₁
		942 TEEIPPLIENR ₉₅₂ 75 NFLINETAR ₈₃
Egg Yolk	Apovitellenin 1 (P02659)	1745 TVIVEAPIHGLK ₁₇₅₆ 1757 NVNFDGEILK ₁₇₆₆ 1874 ATAVSLEWQQR ₁₈₈₄
	Vitellogenin 1 (P87498)	

Table 4

Poultry inter-species homology of selected egg peptide biomarkers using BLAST analysis with the UniProtKB database from UniProt and the non-redundant protein sequence database from NCBI. "X" indicates that the peptide was sequenced in the species, and X* that the sequence was predicted in the species.

	Hen (<i>Gallus gallus</i>)	Goose (<i>Anser cygnoides domesticus</i>)	Duck (<i>Anas platyrhynchos</i>)	Turkey (<i>Meleagris gallopavo</i>)	Quail (<i>Coturnix japonica</i>)	Guinea fowl (<i>Numida meleagris</i>)
GGLEPINFQTAADQAR	X					
HIATNAVLFFGR	X					
FYTVISLKLK	X					
GAIEWEGIESGSVEQAVAK	X					
SAGWNIPIGTLIHR	X	X*	X			X
VEDIWSFLSK	X					
FESNFTQATNR	X				X	
ATAVSLEWQQR	X	X	X	X*	X*	
NVNFDGEILK	X					X
TVIVEAPIHGLK	X					
ALFDYFGYSHDGK	X			X	X*	
ASFYGLSHAVTK	X			X	X*	X
FALSGIVTPGAK	X			X*		X
HFLPSSYK	X	X*	X	X*	X*	X
TEEIPPLIENR	X	X*	X	X*	X*	
NFLINETAR	X					

of farmed birds and the identification of inter-poultry species homologies were assessed using Basic Local Alignment Search Tool (BLAST) analyses using the UniProt and NCBI protein databases. The results obtained for the poultry species are shown in Table 4. The GGLEPINFQTAADQAR peptide from the ovalbumin was specific for hen egg whereas the ATAVSLEWQQR peptide from vitellogenin 1 was found in several poultry species. This information allows for the development of methods specific to hen eggs or methods for differentiating the eggs of poultry species. Concerning the specificity to poultry egg, none of the selected peptides was found in the sequence of proteins of other food ingredients. For example, Peptide HFLPSSYK has been identified in the protein of turtle, alligator, and snake eggs; however, because these eggs are not usually consumed, this has not been a problem. Thus, the procedure in this study ensures specificity, with regard to the databases considered. Indeed, all of the proteins in potential food ingredients have not yet been sequenced.

3.8. Comparison of selected peptide biomarkers with literature

The 16 egg peptides selected with our empirical approach were finally compared with egg peptide biomarkers described in the literature. In the table placed in Supplementary material 4, our potential biomarkers are highlighted in bold. The matrices used by authors are also referenced in this table.

A total of 54 egg peptides were identified in the literature, 44 from egg white proteins and 10 from egg yolk proteins. This tendency for peptides from egg white proteins is partially due to methods detecting egg white used for some winemaking techniques. But, as explained above, peptide biomarkers have to be selected from both egg white and egg yolk proteins to be able to detect egg in any situation.

Out of these 54 peptides from the literature, 44 were identified with our HRMS based method. Many of them would have been rejected by our selection criteria. For example, ovalbumin peptide LTEWTSSNVMEER is considered as egg biomarker in 11 publications. But, since it contains a potentially modified methionine, it was rejected for our future quantitative method. In contrast, 9 peptides of our final biomarkers selection aren't considered in other publications. The majority of them (8/9) are coming from egg yolk proteins.

4. Conclusions

Food allergy is a growing health problem; thus, there is an urgent need for accurate detection and quantification methods of the allergenic ingredients in processed food. Mass spectrometry-based methods are promising. They can overcome the limitations of classic detection methods, such as ELISA or PCR. The proteins of allergenic ingredients are detected and quantified through their

constitutive peptides, and the selection of these peptides is the first, and crucial, step in method development.

This study presented an empirical approach for identifying potential peptide biomarkers to detect and to quantify egg in processed food. The approach was based on the analysis of processed egg matrices (raw egg, heated egg, fatty rich environment, and low pH matrix) with HRMS. The identified peptides were filtered using a series of criteria to ensure method specificity, trueness, and robustness. Thus, peptides must be specific for the allergenic ingredient, be robust to food processing, have originated from abundant proteins, be size-compatible with QqQ, and not be prone to amino acid modifications or missed cleavages by the protease used. The use of a tandem Lys-C/trypsin enzymatic digestion improved the digestion of some of the proteins as compared to digestion with trypsin alone.

Considering these criteria, a list of 16 robust potential peptide biomarkers was obtained, allowing for the detection of the egg yolk and egg white proteins for eggs from hens and other farm birds (goose, duck, turkey, quail, and guinea fowl). In future research, these peptides will be used as targets in a quantitative UHPLC-MS/MS method for detecting food allergens.

Acknowledgments

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS.

Appendix A. Supplementary data

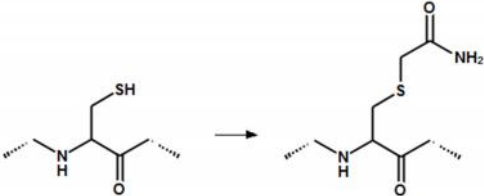
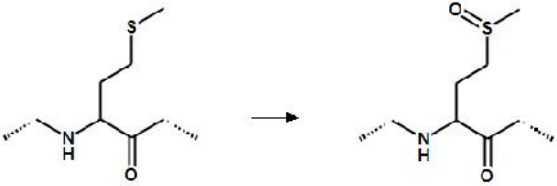
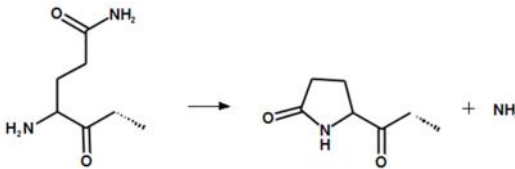
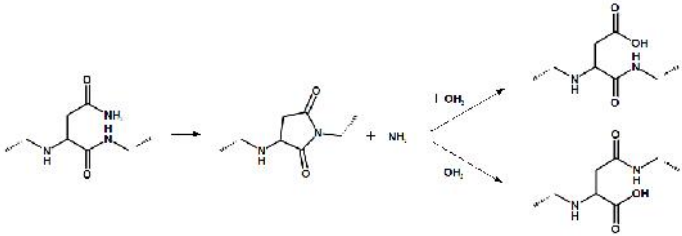
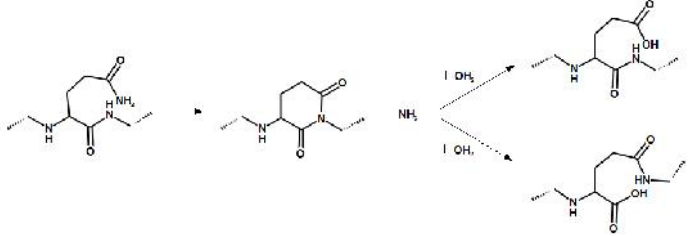
Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.chroma.2018.11.036>.

References

- [1] B.I. Nwaru, L. Hickstein, S.S. Panesar, A. Muraro, T. Werfel, V. Cardona, A.E.J. Dubois, S. Halken, K. Hoffmann-Sommergruber, L.K. Poulsen, G. Roberts, R. Van Ree, B.J. Vlieg-Boerstra, A. Sheikh, The epidemiology of food allergy in Europe: a systematic review and meta-analysis, *Allergy Eur. J. Allergy Clin. Immunol.* 69 (2014) 62–75, <http://dx.doi.org/10.1111/all.12305>.
- [2] A. Muraro, T. Werfel, K. Hoffmann-Sommergruber, G. Roberts, C. Beyer, C. Bindslev-Jensen, V. Cardona, A. Dubois, G. duToit, P. Eigenmann, M. Fernandez Rivas, S. Halken, L. Hickstein, A. Høst, E. Knol, G. Lack, M.J. Marchisotto, B. Niggemann, B.I. Nwaru, N.G. Papadopoulos, L.K. Poulsen, A.F. Santos, I. Skypala, A. Schoepfer, R. Van Ree, C. Venter, M. Worm, B. Vlieg-Boerstra, S. Panesar, D. de Silva, K. Soares-Weiser, A. Sheikh, B.K. Ballmer-Weber, C. Nilsson, N.W. de Jong, C.A. Akdis, EAACI Food Allergy and Anaphylaxis Guidelines Group, EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy, *Allergy* 69 (2014) 1008–1025, <http://dx.doi.org/10.1111/all.12429>.
- [3] S.H. Sicherer, H.A. Sampson, Food allergy: epidemiology, pathogenesis, diagnosis, and treatment, *J. Allergy Clin. Immunol.* 133 (2014), <http://dx.doi.org/10.1016/j.jaci.2013.11.020>.
- [4] B. Popping, C. Diaz-Amigo, K. Hoenicke, *Molecular Biological and Immunological Techniques and Applications for Food Chemists*, 2010, <http://dx.doi.org/10.1002/9780470637685>.
- [5] A.J. Van Hengel, Food allergen detection methods and the challenge to protect food-allergic consumers, *Anal. Bioanal. Chem.* 389 (2007) 111–118, <http://dx.doi.org/10.1007/s00216-007-1353-5>.
- [6] Nomenclature for automated and mechanised analysis (Recommendations 1989), *Pure Appl. Chem.* 61 (1989) 1657, <http://dx.doi.org/10.1351/pac198961091657>.
- [7] D.C. Liebler, L.J. Zimmerman, Targeted quantitation of proteins by mass spectrometry, *Biochemistry* 52 (2013) 3797–3806, <http://dx.doi.org/10.1021/bi400110b>.
- [8] J. Heick, M. Fischer, B. Popping, First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry, *J. Chromatogr. A* 1218 (2011) 938–943, <http://dx.doi.org/10.1016/j.chroma.2010.12.067>.
- [9] M. Planque, T. Arnould, M. Dieu, P. Delahaut, P. Renard, N. Gillard, Liquid chromatography coupled to tandem mass spectrometry for detecting ten allergens in complex and incurred foodstuffs, *J. Chromatogr. A* 1530 (2017) 138–151, <http://dx.doi.org/10.1016/j.chroma.2017.11.039>.
- [10] M. Planque, T. Arnould, M. Dieu, P. Delahaut, P. Renard, N. Gillard, Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs, *J. Chromatogr. A* 1464 (2016) 115–123, <http://dx.doi.org/10.1016/j.chroma.2016.08.033>.
- [11] A.I. Sancho, E.N.C. Mills, Proteomic approaches for qualitative and quantitative characterisation of food allergens, *Regul. Toxicol. Pharmacol.* 58 (2010) S42–S46, <http://dx.doi.org/10.1016/j.yrtph.2010.08.026>.
- [12] M. Rauh, LC-MS/MS for protein and peptide quantification in clinical chemistry, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 883–884 (2012) 59–67, <http://dx.doi.org/10.1016/j.jchromb.2011.09.030>.
- [13] N.R. Kitteringham, R.E. Jenkins, C.S. Lane, V.L. Elliott, B.K. Park, Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 877 (2009) 1229–1239, <http://dx.doi.org/10.1016/j.jchromb.2008.11.013>.
- [14] S.H. Sicherer, H.A. Sampson, 9. Food allergy, *J. Allergy Clin. Immunol.* 117 (2006), <http://dx.doi.org/10.1016/j.jaci.2005.05.048>.
- [15] M. Eggesbø, G. Botten, R. Halvorsen, P. Magnus, The prevalence of allergy to egg: a population-based study in young children, *Allergy Eur. J. Allergy Clin. Immunol.* 56 (2001) 403–411, <http://dx.doi.org/10.1034/j.1398-9995.2001.056005403.x>.
- [16] USDA, Code of Federal Regulations, Section 590.570 – Pasteurization of liquid eggs, 2012 <https://www.gpo.gov/fdsys/pkg/CFR-2012-title9-vol2/xml/CFR-2012-title9-vol2-sec590-570.xml>.
- [17] C.M. Stoscheck, Quantitation of protein, in: M.P.B.T.-M. E. Deutscher (Eds.), *Guid. to Protein Purif.*, Academic Press, 1990, pp. 50–68, [http://dx.doi.org/10.1016/0076-6879\(90\)82008-P](http://dx.doi.org/10.1016/0076-6879(90)82008-P).
- [18] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem.* 75 (2003) 4646–4658, <http://dx.doi.org/10.1021/ac0341261>.
- [19] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, *Anal. Chem.* 74 (2002) 5383–5392, <http://dx.doi.org/10.1021/ac025747h>.
- [20] M.J. Walker, D.T. Burns, C.T. Elliott, M.H. Gowland, E.N.C. Mills, Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs, *Analyst* 141 (2016) 24–35, <http://dx.doi.org/10.1039/c5an01457c>.
- [21] W. Daengprok, W. Garnjanagoonchorn, Y. Mine, Fermented pork sausage fortified with commercial or hen eggshell calcium lactate, *Meat Sci.* 62 (2002) 199–204, [http://dx.doi.org/10.1016/S0309-1740\(01\)00247-9](http://dx.doi.org/10.1016/S0309-1740(01)00247-9).
- [22] E.C.Y. Li-Chan, H.-O. Kim, Structure and chemical compositions of eggs, in: *Egg Biosci. Biotechnol.*, John Wiley & Sons, Inc., 2008, pp. 1–95, <http://dx.doi.org/10.1002/9780470181249.ch1>.
- [23] M.T. Hincke, J. Gautron, M. Panheleux, J. Garcia-Ruiz, M.D. McKee, Y. Nys, Identification and localization of lysozyme as a component of eggshell membranes and eggshell matrix, *Matrix Biol.* 19 (2000) 443–453, [http://dx.doi.org/10.1016/S0945-053X\(00\)00095-0](http://dx.doi.org/10.1016/S0945-053X(00)00095-0).
- [24] M. Wong, M.J.C. Hendrix, K. von der Mark, C. Little, R. Stern, Collagen in the egg shell membranes of the hen, *Dev. Biol.* 104 (1984) 28–36, [http://dx.doi.org/10.1016/0012-1606\(84\)90033-2](http://dx.doi.org/10.1016/0012-1606(84)90033-2).
- [25] M. Ebisawa, H. Tachimoto, K. Ikematsu, C. Sugizaki, Y. Masuda, M. Kimura, Allergenic activity of non-calcinated egg shell calcium, *Arerugi* 54 (2005) 471–477.
- [26] B.J. Walsh, D. Barnett, R.W. Burley, C. Elliott, D.J. Hill, M.E.H. Howden, New allergens from hen's egg white and egg yolk, *Int. Arch. Allergy Immunol.* 87 (1988) 81–86, <http://dx.doi.org/10.1159/000234653>.
- [27] J.S. Cottrell, Protein identification using MS/MS data, *J. Proteom.* 74 (2011) 1842–1851, <http://dx.doi.org/10.1016/j.jprot.2011.05.014>.
- [28] T.P. Coultaire, R.S. of C. (Great Britain), *Food: The Chemistry of its Components*, Royal Society of Chemistry, 2009 <https://books.google.be/books?id=KF2A8Cz7B-cC>.
- [29] N.E. Robinson, A.B. Robinson, Molecular clocks, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 944–949, <http://dx.doi.org/10.1073/pnas.98.3.944>.
- [30] B. Li, R.L. Schowen, E.M. Topp, R.T. Borchardt, Effect of N-1 and N-2 residues on peptide deamidation rate in solution and solid state, *AAPS J.* 8 (2006) E166–E173, <http://dx.doi.org/10.1208/aaps080120>.
- [31] A.A. Kosky, V. Dharmavaram, G. Ratnaswamy, M.C. Manning, Multivariate analysis of the sequence dependence of asparagine deamidation rates in peptides, *Pharm. Res.* 26 (2009) 2417–2428, <http://dx.doi.org/10.1007/s11095-009-9953-8>.
- [32] N.E. Robinson, A.B. Robinson, Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects, *J. Pept. Res.* 63 (2004) 437–448, <http://dx.doi.org/10.1111/j.1399-3011.2004.00148.x>.
- [33] J. Bunkenborg, G. Espadas, H. Molina, Cutting edge proteomics: benchmarking of six commercial trypsins, *J. Proteome Res.* 12 (2013) 3631–3641, <http://dx.doi.org/10.1021/pr4001465>.
- [34] L. Tsiaitsiani, A.J.R. Heck, Proteomics beyond trypsin, *FEBS J.* 282 (2015) 2612–2626, <http://dx.doi.org/10.1111/febs.13287>.
- [35] P. Giansanti, L. Tsiaitsiani, T.Y. ew Low, A.J.R. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin, *Nat. Protoc.* 11 (2016) 993–1006, <http://dx.doi.org/10.1038/nprot.2016.057>.
- [36] T. Glatter, C. Ludwig, E. Ahnér, R. Aebersold, A.J.R. Heck, A. Schmidt, Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin

- proteolysis over trypsin digestion, *J. Proteome Res.* 11 (2012) 5145–5156, <http://dx.doi.org/10.1021/pr300273g>.
- [37] J. Kay, Specificity of proteolysis, *FEBS Lett.* 331 (1993) 201, [http://dx.doi.org/10.1016/0014-5793\(93\)80331-N](http://dx.doi.org/10.1016/0014-5793(93)80331-N).
- [38] R. Raijmakers, P. Neerincx, S. Mohammed, A.J.R. Heck, Cleavage specificities of the brother and sister proteases Lys-C and Lys-N, *Chem. Commun.* 46 (2010) 8827, <http://dx.doi.org/10.1039/c0cc02523b>.
- [39] J.A. Siepen, E.J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, *J. Proteome Res.* 6 (2007) 399–408, <http://dx.doi.org/10.1021/pr060507u>.
- [40] J. Rodriguez, Does trypsin cut before proline? *J. Proteome Res.* 7 (2008) 300–305, <http://dx.doi.org/10.1021/pr0705035>.
- [41] P.D. Gershon, Cleaved and missed sites for trypsin, Lys-C, and Lys-N can be predicted with high confidence on the basis of sequence context, *J. Proteome Res.* 13 (2014) 702–709, <http://dx.doi.org/10.1021/pr400802z>.
- [42] E. Harper, A. Berger, On the size of the active site in proteases: pronase, *Biochem. Biophys. Res. Commun.* 46 (1972) 1956–1960, [http://dx.doi.org/10.1016/0006-291X\(72\)90076-9](http://dx.doi.org/10.1016/0006-291X(72)90076-9).
- [43] S.E. Khuda, L.S. Jackson, T.J. Fu, K.M. Williams, Effects of processing on the recovery of food allergens from a model dark chocolate matrix, *Food Chem.* 168 (2015) 580–587, <http://dx.doi.org/10.1016/j.foodchem.2014.07.084>.
- [44] C.H. Parker, S.E. Khuda, M. Pereira, M.M. Ross, T.J. Fu, X. Fan, Y. Wu, K.M. Williams, J. DeVries, B. Pulvermacher, B. Bedford, X. Zhang, L.S. Jackson, Multi-allergen quantitation and the impact of thermal treatment in industry-processed baked goods by ELISA and liquid chromatography–tandem mass spectrometry, *J. Agric. Food Chem.* 63 (2015) 10669–10680, <http://dx.doi.org/10.1021/acs.jafc.5b04287>.
- [45] K.J. Shefcheck, J.H. Callahan, S.M. Musser, Confirmation of peanut protein using peptide markers in dark chocolate using liquid chromatography–tandem mass spectrometry (LC-MS/MS), *J. Agric. Food Chem.* 54 (2006) 7953–7959, <http://dx.doi.org/10.1021/jf060714e>.
- [46] E.S. Polat, O.B. Citil, M. Garip, Fatty acid composition of yolk of nine poultry species kept in their natural environment, *Anim. Sci. Pap. Rep.* 31 (2013) 363–368.
- [47] T. Langeland, A clinical and immunological study of allergy to hen's egg white: VI. Occurrence of proteins cross-reacting with allergens in hen's egg white as studied in egg white from Turkey, duck, goose, seagull, and in hen egg yolk, and hen and chicken sera and fles, *Allergy* 38 (1983) 399–412, <http://dx.doi.org/10.1111/j.1398-9995.1983.tb05083.x>.
- [48] C. Alessandri, M. Calvani, L. Rosengart, C. Madella, Anaphylaxis to quail egg, *Allergy Eur. J. Allergy Clin. Immunol.* 60 (2005) 128–129, <http://dx.doi.org/10.1111/j.1398-9995.2004.00449.x>.

Supplementary material 2: Identified amino acid modifications based on literature review and experimental results. For each modification, the structural formula is developed, and the corresponding mass shift is given.

Modification	Chemical reaction	Mass shift (Da)	Ref.
Cysteine carbamido-methylation		+57.021	[26–28]
Methionine oxidation		+15.995	[28–30]
N-terminal glutamine to pyroglutamate conversion		−17.027	[28,29]
Asparagine deamidation		+0.984	[31–34]
Glutamine deamidation		+0.984	[31,32, 34]