

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### Comparative study of concatemer efficiency as an isotope-labelled internal standard for allergen quantification

Gavage, Maxime; Van Vlierberghe, Kaatje; Van Poucke, C.; De Loose, M.; Gevaert, K.; Dieu, Marc; Renard, Patricia; Arnould, Thierry; Filee, Patrice; Gillard, Nathalie

*Published in:*  
Food Chemistry

*DOI:*  
[10.1016/j.foodchem.2020.127413](https://doi.org/10.1016/j.foodchem.2020.127413)

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

#### [Link to publication](#)

#### *Citation for pulished version (HARVARD):*

Gavage, M, Van Vlierberghe, K, Van Poucke, C, De Loose, M, Gevaert, K, Dieu, M, Renard, P, Arnould, T, Filee, P & Gillard, N 2020, 'Comparative study of concatemer efficiency as an isotope-labelled internal standard for allergen quantification', *Food Chemistry*, vol. 332, 127413, pp. 1-10.  
<https://doi.org/10.1016/j.foodchem.2020.127413>

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Comparative study of concatemer efficiency as an isotope-labelled internal standard for allergen**  
2 **quantification**

3 Maxime GAVAGE<sup>a,b</sup>, Kaatje VAN VLIERBERGHE<sup>c</sup>, Christof VAN POUCKE<sup>c</sup>, Marc DE LOOSE<sup>c</sup>, Kris  
4 GEVAERT<sup>d,e</sup>, Marc DIEU<sup>b,f</sup>, Patsy RENARD<sup>b, f</sup>, Thierry ARNOULD<sup>b</sup>, Patrice FILEE<sup>a</sup>, Nathalie  
5 GILLARD<sup>a,\*</sup>

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

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

9 c. Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Brusselsesteenweg 370,  
10 9090 Melle, Belgium

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

12 e. Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, B9000 Ghent,  
13 Belgium

14 f. MaSUN, Mass Spectrometry Facility, University of Namur, 61, Rue de Bruxelles, 5000 Namur,  
15 Belgium

16

17 \* Corresponding author; telephone: +32 (0) 84 31 00 90; e-mail: [n.gillard@cergroupe.be](mailto:n.gillard@cergroupe.be)

18 CER Groupe, rue du Point du Jour, 8, 6900 Marloie, Belgium

19 **Abstract**

20 Mass spectrometry-based methods coupled with stable isotope dilution have become effective and  
21 widely used methods for the detection and quantification of food allergens. Current methods target  
22 signature peptides resulting from proteolytic digestion of proteins of the allergenic ingredient. The choice  
23 of appropriate stable isotope-labelled internal standard is crucial, given the diversity of encountered food  
24 matrices which can affect sample preparation and analysis. We propose the use of concatemer, an  
25 artificial and stable isotope-labelled protein composed of several concatenated signature peptides as  
26 internal standard. With a comparative analysis of three matrices contaminated with four allergens (egg,  
27 milk, peanut, and hazelnut), the concatemer approach was found to offer advantages associated with  
28 the use of labelled proteins, ideal but unaffordable, and circumvent certain limitations of traditionally  
29 used synthetic peptides as internal standards. Although used in the proteomic field for more than a  
30 decade, concatemer strategy has not yet been applied for food analysis.

31

32 **Keywords**

33 Food allergen analysis, mass spectrometry, isotope dilution, isotope-labelled internal standard, isotope-  
34 labelled concatemer

35

## 36 1. Introduction

37 Food allergy is defined as an adverse health effect arising from a specific reproducible immune  
38 response that occurs on exposure to a given food (Boyce et al., 2011). Several studies indicate an  
39 increase in the prevalence of food allergy with nearly 5% of adults and 8% of children being affected  
40 (Sicherer & Sampson, 2014). Given the absence of accepted treatment, the current solution for allergic  
41 patients relies on allergen avoidance to circumvent allergic reactions. However, this essentially requires  
42 correct food labelling and efficient risk management from food business operators to reduce the risk of  
43 contamination by allergens to acceptable levels. European legislation (Regulation [EU] No 1169/2011)  
44 requires the labelling of 14 allergenic ingredients when they are part of a foodstuff recipe. However, this  
45 legislation does not cover the presence of hidden allergens that are due to cross-contamination during  
46 food processing. Even if strongly requested by food producers and control laboratories, no harmonized  
47 regulatory framework for managing hidden allergens or action thresholds have been enacted in Europe.  
48 Some countries have set legal thresholds but with a high disparity among allergens and among countries  
49 (Planque et al., 2019). A quantitative risk assessment was also developed by VITAL® (Voluntary  
50 Incidental Trace Allergen Labelling) combining reference doses and exposure (Allen et al., 2014). The  
51 thresholds for allergenic proteins in food are based on clinical data and are indicators of the action levels,  
52 expressed as the total protein amount of the allergenic food (mg), below which only the most sensitive  
53 allergic subjects might react (1% of allergic patients or 5% of them for the less common foods). These  
54 values are often used by laboratories as a targeted limit of quantification (LOQ) in the absence of legal  
55 thresholds.

56 The development of a quantitative allergen risk assessment requires quantitative allergen analysis.  
57 During the last decade, mass spectrometry became the method of choice for allergen analysis (Ahsan,  
58 Rao, Gruppuso, Ramratnam, & Salomon, 2016). Allergen analysis by mass spectrometry is  
59 predominantly performed by specific analysis of peptides obtained by an enzymatic digestion of the  
60 proteins of the sample, including the proteins of the allergenic ingredients. One of the advantages of  
61 mass spectrometry-based methods is the possibility to simultaneously detect multiple peptides from  
62 multiple allergens, thus enabling time- and money-saving multiplexed analysis. Such a targeted  
63 approach, named multiple reaction monitoring (MRM), offers high sensitivity and specificity. Targeted  
64 proteomics is often used for absolute peptide quantification in combination with isotope dilution, a  
65 technique based on the use of an internal standard corresponding to the stable isotope-labelled version  
66 of the analyte (Monaci, Losito, De Angelis, Pilolli, & Visconti, 2013; Nitride et al., 2019; Planque et al.,

67 2019). The introduction of this isotope-labelled internal standard corrects for variability and various  
68 matrix effects during the actual analysis. Notably, ion suppression effects and, depending on the type  
69 of internal standard, matrix effects and analyte loss during sample preparation may be corrected by the  
70 use of isotope-labelled internal standards.

71 Peptides specific for allergen proteins are the analytes in mass spectrometry analysis of food  
72 allergens; however, the initial analytes are proteins. Stable isotope-labelled internal standards can  
73 therefore adopt different forms. In theory, a stable isotope-labelled protein is the ideal internal standard  
74 as, when added to the food that needs to be analyzed, it can correct for sample losses during all the  
75 steps of the sample preparation procedure (including protein extraction and digestion), as well as for  
76 matrix effects during mass spectrometry analysis. Such an approach was proposed by Newsome and  
77 Scholl (Newsome & Scholl, 2013) for the quantification of bovine milk  $\alpha_{S1}$ -casein in baked goods. The  
78 main limitation of this approach, besides technical issues for protein production, is its cost. When one  
79 aims at multiplexed analysis, this necessitates the use of multiple isotope-labelled proteins, which is  
80 unrealistic for laboratories performing routine analyses (Planque, Arnould, & Gillard, 2017). Therefore,  
81 most laboratories rely on stable isotope-labelled synthetic peptides (Boo, Parker, & Jackson, 2018;  
82 Henrottin et al., 2019; Planque et al., 2019). However, in food allergen analysis, the initial analytes are  
83 proteins. Peptide internal standard and protein analytes can exhibit different behaviors during the  
84 extraction, leading to different extraction yield. Moreover, the peptides do not undergo the enzymatic  
85 digestion step which is known to be highly affected by the matrix effects (Korte, Oberleitner, &  
86 Brockmeyer, 2019).

87 Here, we implemented an alternative method based on the synthesis of a concatemer used as a  
88 stable isotope-labelled internal standard for allergen quantification. This strategy has been well adopted  
89 by proteomics researchers, and the concatemers are known as QconCAT (Pratt et al., 2006), but, as far  
90 as we know, these molecules have not yet been explored for food analysis. Concatemers are artificial  
91 proteins composed of concatenated, proteotypic peptides originating from different proteins of interest.  
92 The peptides themselves are typically first identified following mass spectrometry or are predicted from  
93 theoretical peptide sequences. Concatemers are typically recombinantly produced in an environment  
94 that allows labelling with stable isotopes (e.g.,  $^{13}\text{C}$  or  $^{15}\text{N}$ ). In contrast to synthetic peptides, concatemers  
95 need to be proteolytically digested to release their peptides, and thus, this peptide release is also  
96 affected by the interference caused by the matrix during the digestion step, in a manner similar to the  
97 analyte of interest. Another advantage of concatemers is their potential for multiplexing. A single

98 concatemer can be composed of numerous proteotypic peptides and can therefore be used for  
99 multiplexed allergen analysis. The limitation of this approach is fixed by the protein size reachable with  
100 recombinant protein expression, which is more than 100 kDa (Chambers, Austen, Fulghum, & Kim,  
101 2004). This approach can be cost-effective when compared with using synthetic peptides for multiplexed  
102 analysis. For our study, we developed, produced, and purified a <sup>15</sup>N isotopically labelled concatemer  
103 composed of 19 proteotypic peptides, allowing for the analysis of 4 allergenic ingredients (egg, milk,  
104 peanut, and hazelnut). We evaluated the performance of this concatemer by the analysis of three  
105 uncontaminated food matrices spiked with increasing and defined concentrations (2.5 ppm to 50 ppm,  
106 where ppm corresponded to mg of total allergen protein per kg of matrix) of the selected allergen  
107 extracts. In addition, we compared the use of the concatemer with that of five synthetic peptides  
108 corresponding to tryptic peptides from the four considered allergens and with β-lactoglobulin, a bovine  
109 milk protein that was <sup>15</sup>N isotopically labelled.  
110

## 111 2. Material and methods

### 112 2.1. Reagent and materials

113 Gene synthesis and cloning were ordered from GeneCust (Boynes, France). Acetic acid, ammonium  
114 bicarbonate, ampicillin sodium salt, chloramphenicol, dimethyl sulfoxide (DMSO), DL-dithiothreitol  
115 (DTT), expression plasmid pET17b(+) Novagen, HiLoad® 26/600 Superdex® 200 pg, imidazole  
116 hydrochloride, iodoacetamide (IAA), kanamycin monosulfate, Lennox broth (LB), Ni Sepharose® 6 fast  
117 flow GE Healthcare, Origami™ B(DE3) pLysS competent cells Novagen, phenylmethanesulfonyl  
118 fluoride (PMSF), Q Sepharose® Fast Flow, select agar, sodium chloride, sodium phosphate dibasic,  
119 sodium phosphate monobasic, tetracycline hydrochloride, tetraethylammonium bicarbonate (TEAB),  
120 trypsin from bovine pancreas, tris(hydroxymethyl)aminomethane (Tris) and urea were obtained from  
121 Sigma-Aldrich (Bornem, Belgium). One Shot™ BL21(DE3) chemically competent *Escherichia coli*,  
122 isopropyl β-D-thiogalactopyranoside (IPTG), SnakeSkin™ dialysis tubing, 3.5K MWCO, 22 mm were  
123 purchased from Thermo Fisher Scientific (Waltham, MA, USA). Bioexpress cell growth media (U-15N,  
124 98%) (10x concentrate) was obtained from Buchem B.V. (Apeldoorn, the Netherlands), Trypsin Gold,  
125 Mass Spectrometry Grade from Promega (Madison, WI, USA), 4–20 Mini-PROTEAN® TGX™ precast  
126 protein gels from Bio-Rad (Hercules, CA, USA), Sep-Pak C18 6 cc Vac solid-phase extraction (SPE)  
127 cartridges from Waters (Milford, MA, USA), and 0.2 μm acrodisc syringe filters with supor membrane  
128 from Pall Corporation (Port Washington, NY, USA). Water, acetonitrile (ACN), and formic acid (FA) were  
129 obtained from Biosolve (Valkenswaard, the Netherlands). Labelled synthetic peptides  
130 ADIYTEQV[<sup>13</sup>C<sub>5</sub><sup>15</sup>N]GR, FFVAPFPEVFGK[<sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub>], GGLEPINF[Ring-D<sub>5</sub>]QTAADQAR, LSF[Ring-  
131 D<sub>5</sub>]NPTQLEEQCHI, TANELNLLIL[<sup>13</sup>C<sub>6</sub><sup>15</sup>N]R were ordered from Eurogentec (Seraing, Belgium).

132 Food samples were analyzed by ultra-high performance liquid chromatography–tandem mass-  
133 spectrometry (UHPLC MS/MS) using an Acquity liquid chromatograph equipped with a C18 Acquity  
134 BEH130 column (2.1 x 150 mm; 1.7 μm) and coupled with a Xevo TQ-S micro triple quadrupole system  
135 (Waters, Milford, MA, USA). Characterization of <sup>15</sup>N isotopically labelled concatemer and β-lactoglobulin  
136 was performed by ultra-high performance liquid chromatography–high resolution mass spectrometry  
137 (UHPLC-HRMS) using an Acquity liquid chromatograph equipped with a C18 Acquity BEH130 column  
138 (2.1 x 150 mm; 1.7 μm) and coupled to a Xevo G2-XS QToF quadrupole time-of-flight system (Waters,  
139 Milford, MA, USA).

### 140 2.2. <sup>15</sup>N isotopically labelled concatemer production and purification

141 Design and production of the concatemer were adapted from the method of Pratt (Pratt et al.,  
142 2006).The first step focused on concatemer design and the selection of the concatenated peptides.  
143 Here, we considered 19 peptides (Table 1) originating from seven proteins of four allergenic ingredients  
144 ( $\alpha$ <sub>s1</sub>-casein and  $\beta$ -lactoglobulin from cow milk; ovalbumin, ovotransferrin, and vitellogenin-1 from hen's  
145 egg; Cor a 9 allergen from hazelnut; and Ara h 1 allergen from peanut). These 19 peptides were selected  
146 from a set of relevant peptide biomarkers identified by an empirical approach based on UHPLC-HRMS  
147 analysis of incurred and processed samples. The applied food processing steps, sample preparation,  
148 and selection criteria have already been detailed in our previous studies (Gavage et al., 2019, 2020;  
149 Van Vlierberghe et al., 2020). The peptides were then *in silico* concatenated, and the resulting  
150 polypeptide was flanked with an N-terminus initiator sequence including a methionine start and a  
151 C-terminus hexahistidine purification tag (His-tag). Hydrophobicity of each of the 19 peptides was  
152 evaluated based on their grand average of hydropathy (GRAVY) parameter. Hydrophobic and  
153 hydrophilic peptides were alternated in the concatemer sequence to avoid the formation of high  
154 hydrophobic clusters that can interfere with solvent accessibility of concatenated peptides and thus with  
155 their subsequent proteolysis during the sample preparation. Translation-associated aspects such as  
156 tRNA-mediated codon usage bias and mRNA secondary structure, known to impact the translation  
157 process (Gorochoowski, Ignatova, Bovenberg, & Roubos, 2015), were also considered. Visual Gene  
158 Developer (University of California-Davis, Davis, CA, USA) was used to predict and optimize the mRNA  
159 secondary structure. The *in silico* designed DNA construct was finally chemically synthesized and cloned  
160 into the pET17b(+) expression vector using *NdeI* and *XhoI* restriction sites to give the pET17b(+)-  
161 concat1.

162 The *E. coli* BL21(DE3)/pET17b(+)-concat1 strain was inoculated in a 30 ml starter culture of <sup>15</sup>N  
163 labelled media (Bioexpress cell growth media [U-15N, 98%] with 100 µg/mL ampicillin) and grown  
164 overnight at 37 °C under 300 rpm orbital shaking. Cells were harvested by centrifugation (4000 x g,  
165 5 min) and the pellet was resuspended in 1 mL of <sup>15</sup>N labelled media. Next, a volume of 660 µl of this  
166 bacterial suspension was used to inoculate a 1L <sup>15</sup>N labelled main culture. This culture was grown at 37  
167 °C under 300 rpm orbital shaking until the optical density at 600 nm reached 0.6–0.8. Concatemer  
168 expression was next induced with 1 mM IPTG and cells were cultured overnight at 25 °C under 300 rpm  
169 orbital shaking. Cells were harvested by centrifugation (5000 x g, 15 min) and stored at –80 °C until  
170 concatemer purification.

171 The cell pellet of the 1 L culture was resuspended in 40 mL of lysis buffer (50 mM Tris - 10 mM  
172 imidazole - pH 8) with 1 mM PMSF. Cells were disrupted using a Vibra-Cell™ (Sonics, Newtown, CN,  
173 USA) ultrasonic probe. The cell lysate was centrifuged twice (40000 x g, 20 min) and filtered through  
174 0.2 µm syringe filters before to be submitted to metal affinity chromatography purification. The protein  
175 solution was loaded on a 8 ml Ni Sepharose 6 Fast Flow column equilibrated with lysis buffer. An  
176 intermediate washing step was performed in the presence of 20 mM imidazole and the His-tag labelled  
177 concatemer was finally eluted by using a linear imidazole gradient from 20 mM to 250 mM. The elution  
178 fractions were analyzed on SDS-PAGE (Supplementary data 1). The positive fractions were pooled and  
179 dialyzed against the storage buffer (50 mM Tris - pH 8) to eliminate imidazole.

180 Total protein concentration was measured by absorbance at 280nm. A SDS-PAGE/densitometry  
181 method based on ImageJ software was used to estimate concatemer purity. A total of 84.5 mg of <sup>15</sup>N  
182 isotopically labelled concatemer were produced and purified with an estimated purity higher than 90%.  
183 Protein sequences, concentration calculations, and purity estimation are detailed in Supplementary  
184 data 3.

### 185 **2.3. <sup>15</sup>N isotopically labelled β-lactoglobulin production and purification**

186 The production of β-lactoglobulin, a cow milk protein, was adapted from the work of Loch and  
187 collaborators (Loch et al., 2016) who implemented a method leading to the cytoplasmic accumulation of  
188 correctly folded disulfide bond-dependent proteins. Briefly, two mutations (L2A/I3S) were introduced in  
189 the β-lactoglobulin to facilitate *in vivo* cleavage of the N-terminal methionine allowing for correct protein  
190 folding. ) Further, the *E. coli* Origami B (DE3) pLysS strain, a glutathione reductase (gor) and thioredoxin  
191 reductase (trxB) mutated strain, was used for conducting the cytoplasmic co-expression of the protein  
192 of interest with DsbC, an *E. coli* cytoplasmic disulfide bond isomerase. The co-expression was achieved  
193 with the same expression vector (pET17b(+)-DsbC-BLg) in which the two genes were transcribed from  
194 individual T7 IPTG-inducible promoters.

195 To achieve the production of <sup>15</sup>N labelled β-lactoglobulin, expression (starter culture, main culture,  
196 and IPTG induction) conditions were similar as used for concatemer production. The antibiotics that  
197 were used were tailored to 200 µg/mL ampicillin, 34 µg/mL chloramphenicol, 15 µg/mL kanamycin, and  
198 12.5 µg/mL tetracycline, and the IPTG concentration was 0.5 mM. Harvested cells were resuspended  
199 in 50 mM phosphate buffer, pH 6.5, with 1 mM PMSF and prepared for protein purification using the  
200 same procedure as for the concatemer. The purification of <sup>15</sup>N labelled β-lactoglobulin was performed

201 according to the procedure described by Loch and collaborator (Loch et al., 2016). Briefly, this protocol  
202 combines anion-exchange chromatography (Q Sepharose® Fast Flow) with a NaCl linear elution  
203 gradient (up to 2 M) followed by size-exclusion chromatography (HiLoad® 26/600 Superdex® 200 pg)  
204 in initial conditions (50 mM phosphate buffer, pH 6.5). Eluates of these two purification steps were  
205 collected in 1 mL fractions and analyzed on SDS-PAGE (Supplementary data 2).

206 Total protein concentration was measured by absorbance at 280nm. A SDS-PAGE/densitometry  
207 method based on ImageJ software was used to estimate protein purity. Using this approach, a total of  
208 2.4 mg of <sup>15</sup>N isotopically labelled β-lactoglobulin were produced and purified with an estimated purity  
209 higher than 70%. Protein sequences, concentration calculations, and purity estimation are detailed in  
210 Supplementary data 3.

#### 211 **2.4. Characterization of produced <sup>15</sup>N isotopically labelled proteins**

212 Protein <sup>15</sup>N stable isotope enrichment was evaluated by UHPLC-HRMS analysis of its constitutive  
213 tryptic peptides. In separated containers, concatemer and β-lactoglobulin were diluted to 0.1 mg/mL with  
214 50 mM TEAB, pH 9.2, to a final volume of 20 μl. Disulfide bridges of β-lactoglobulin were successively  
215 reduced and alkylated with DTT (10 mM final concentration, 45 min incubation at 37 °C under 300 rpm  
216 orbital agitation) and IAA (40 mM final concentration, 45 min incubation in the dark at 37 °C under 300  
217 rpm orbital agitation). Concatemer and β-lactoglobulin were then proteolytically digested by adding 0.1  
218 μg of trypsin gold (protein:trypsin ratio of 1:20). Digestion was conducted for 1 h at 37 °C under 300 rpm  
219 orbital agitation and stopped by the addition of 1% (final concentration) of FA followed by centrifugation  
220 (20000 x g, 5 min). Samples were ten-fold diluted with 5% ACN before UHPLC-HRMS analysis.  
221 Peptides (5 μl of sample was injected) were first separated by reverse-phase liquid chromatography  
222 using a 20 min water/ACN + 0.1% FA linear gradient from 5% to 40% of ACN. Data was acquired in  
223 MS<sup>E</sup> mode with 0.3 s scan time within the 50 to 2000 m/z mass range. The data were processed using  
224 UNIFI software (Waters, Milford, MA, USA) and peptide mapping analysis type with traditional tryptic  
225 cleavage rules and setting cysteine carbamidomethylation and <sup>15</sup>N isotope labelling as a fixed  
226 modifications.

227 For each identified tryptic peptide, the most intense charge state was considered to define the <sup>15</sup>N  
228 stable isotopic enrichment. The isotopic enrichment or isotope incorporation rate was evaluated for each  
229 peptide by comparing the intensity (in counts) of the peak corresponding to the fully <sup>15</sup>N labelled (U-<sup>15</sup>N)  
230 peptide with other peaks corresponding to partially <sup>15</sup>N labelled peptides. For practicality, we considered

231 a  $^{13}\text{C}$  natural abundance of 1.1% and neglected hydrogen and oxygen isotopic distributions in our  
232 calculations. Furthermore, only peaks corresponding to peptide with 1  $^{14}\text{N}$  isotope were considered in  
233 our calculation. The proportion of U- $^{15}\text{N}$  peptide was then obtained after comparing the intensity of the  
234 peak corresponding to the (U- $^{15}\text{N}$  & U- $^{12}\text{C}$ ) peptide with the peak corresponding to the [(U-1)- $^{15}\text{N}$  & U-  
235  $^{12}\text{C}$ ] peptide. Protein isotopic enrichment was evaluated with the exponential trend given by the  
236 proportion of the U- $^{15}\text{N}$  version of each peptide considering its nitrogen content.

## 237 **2.5. Food matrices preparation**

238 Three blank food matrices – thus, not contaminated with the considered allergenic ingredients –  
239 were prepared to assess the variability due to the food sample used in our study. These blank matrices  
240 were baked cookies, chocolate, and freeze-dried cookie dough.

241 Cookie dough was produced in batches of 3 kg by mixing (Kenwood Major Titanium, Stainless Steel  
242 Dough Hook, 15 min, max speed) the following ingredients purchased from a local supermarket in the  
243 respective weight proportions as follows: wheat flour (Carrefour type 55)/water (Milli-Q)/olive oil (Bertoli  
244 Classico)/salt (sodium chloride ACS,  $\geq 99\%$ , Thermo Scientific™)/baking powder (Dr. Oetker  
245 Baking)/Sugar (Grand Pont Crystal Sugar): 57%/18%/10%/0.2%/0.8%/14%. The dough was  
246 subsequently rolled out to a thickness of 0.5 mm, and cookies with a diameter of 8 cm were pressed out  
247 of the dough (weight =  $25 \pm 2$  g). Cookies were baked for 25 min with the following program: 1–10 min:  
248 180 °C heat from above and 180 °C heat from below; 11–25 min: 180 °C heat from above and 160 °C  
249 heat from below. This was done to ensure that the warming of the baking plate would not result in uneven  
250 cookie baking. Cookies were left at ambient temperatures to cool down, and subsequently milled and  
251 sieved (Retsch® ZM 200 ultra-centrifugal mill [Retsch GmbH, Haan, Germany] with a 0.75 mm pore  
252 size sieve, 14000 rpm). Cookie powder was stored at 4 °C in the dark until further use.

253 Cookie dough was produced as described above, rolled out to a thickness of 1 cm, stored at  $-20$   
254 °C, and subsequently freeze-dried. Freeze-dried cookie dough was then milled and sieved (Retch® ZM  
255 200 ultra-centrifugal mill with a 0.75 mm pore size sieve, 14000 rpm). The freeze-dried cookie dough  
256 powder was stored at 4 °C in the dark until further use.

257 Chocolate was made by warming chocolate walsenpowder (90%; Callebaut, Belgium) and cacao  
258 butter (10%, Callebaut, Belgium) in a water bath at 40 °C (maximum temperature). The mixture was  
259 stirred for 15 min, after which 2% ammonium phosphatide (kindly provided by Palsgaard, Julesminde,  
260 Denmark) was added. This mixture was again stirred for 15 min and subsequently poured into chocolate

261 molds, resulting in chocolate chips of around 5 g each. The chocolate was left to cool down and solidify  
262 at 4 °C for 2 h, and the chocolate chips were packed under vacuum and stored at 4 °C in the dark until  
263 further use.

## 264 **2.6. Sample preparation for UHPLC-MS/MS analysis**

265 Two series of samples were prepared and analyzed to be able to cover the three internal standards.  
266 Concatemer and  $\beta$ -lactoglobulin were isotopically labelled with the same strategy ( $^{15}\text{N}$  uniform labelling)  
267 and share common tryptic peptides, which cannot be distinguished after enzymatic digestion. Two series  
268 of samples were prepared. Labelled peptides and  $\beta$ -lactoglobulin were spiked in the first series (only  
269 one shared peptide LSFNPTQLEEQCHI) and labelled concatemer in the second one. For each series,  
270 the three blank matrices (baked cookie, chocolate, and lyophilized unbaked cookie dough) were spiked,  
271 before extraction, with the appropriate internal standard and with increasing amounts of a standard  
272 extract of the four allergens (milk, egg, peanut, and hazelnut). These allergen amounts corresponded to  
273 0, 2.5, 5, 10, 25 and 50 ppm level points expressed in total allergen protein per matrix kg. For each  
274 series, each blank matrix and each point of the allergen curve, three biological sample replicates were  
275 prepared and analyzed. Stock solutions containing the four allergen standards at 20 mg/mL were  
276 prepared using a similar extraction protocol as that used for the samples (extraction, sonication, and  
277 centrifugation; see below). These stock solutions were then combined and diluted in appropriate ratios  
278 to spike samples at different contamination levels with a 100  $\mu\text{l}$  volume. Combination and dilution were  
279 calculated based on theoretical protein content of standards assuming 100% extraction yield. Each  
280 internal standard was spiked at the similar molar level (0.25 nmol) with a 100  $\mu\text{l}$  volume. Then, 1 mg/mL  
281 stock solutions of the five considered labelled peptides (ADITYTEQV[ $^{13}\text{C}_5^{15}\text{N}$ ]GR,  
282 FFVAPFPEVFGK[ $^{13}\text{C}_6^{15}\text{N}_2$ ], GGLEPINF[Ring-D<sub>5</sub>]QTAADQAR, LSF[Ring-D<sub>5</sub>]NPTQLEEQCHI, and  
283 TANELNLLIL[ $^{13}\text{C}_6^{15}\text{N}$ ]R) were combined and diluted at the appropriate concentration with 0.1% FA.  
284 Concatemer and  $\beta$ -lactoglobulin solutions were also diluted to be spiked at 0.25 nmol level with a 100  
285  $\mu\text{l}$  volume. This level, converted in equivalent allergen ppm, ranged from 10 ppm for abundant proteins,  
286 such as  $\alpha_{\text{S1}}$ -casein, to more than 300 ppm for less abundant proteins, such as vitellogenin-1. This  
287 estimate was based on the natural abundance of each considered protein in the corresponding  
288 allergenic ingredient. Allergen standards and internal standard were added to blank matrices before  
289 extraction.

290 Samples were prepared as previously described (Planque et al., 2016). Briefly, protein from 2 g  
291 samples was extracted in 50 mL conical tubes with 20 mL of 200 mM Tris, pH 9.2, 2 M urea by shaking

292 at 20 °C for 30 min (Agitelec, J. Toulemonde, Paris, France) prior to ultrasonic treatment at 4 °C for 15  
293 min. After centrifugation (4660 x g, 10 min), 10 mL of supernatant were diluted in digestion buffer (200  
294 mM ammonium bicarbonate, pH 8.2). Protein disulfide bridges were successively reduced and alkylated  
295 with 45 min incubation steps at room temperature with the addition of 1 mL of 200 mM DTT and 1 mL  
296 of 400 mM IAA (in the dark). Protein was then enzymatically digested with the addition of 1 mL of trypsin  
297 solution (trypsin from bovine pancreas, 1 mg/mL in 50 mM acetic acid, pH 2.8) and incubation for 1 h at  
298 37 °C. The digestion reaction was stopped by adding 300 µl of 20% FA to the samples, which were then  
299 centrifuged (4660 x g, 5 min). Obtained peptides were then purified and concentrated using C18 SPE  
300 cartridges, which were first conditioned with 18 mL of ACN followed by 18 mL of 0.1% FA before loading  
301 of 20 mL of the centrifuged sample. The cartridges were washed with 18 mL of 0.1% FA and eluted in  
302 15 mL conical tubes with 6 mL of 80% ACN and 0.1% FA. A volume of 30 µl of DMSO was added to the  
303 sample before evaporation (40 °C under nitrogen flow) to avoid dryness. The pellet was finally dissolved  
304 in 600 µl of 5% ACN with 0.1% FA and centrifuged twice (4660 x g, 5 min in conical tube and 20 000 x  
305 g, 5 min in 1.5 mL microtube, keeping the supernatant) before UHPLC-MS/MS analysis.

## 306 **2.7. UHPLC-MS/MS analysis and data analysis**

307 The peptides were separated by reverse-phase chromatography on-line connected to a triple  
308 quadrupole mass spectrometer. The following 26 min solvent gradient (solvent A, 0.1% FA and solvent  
309 B, ACN and 0.1% FA) was applied to the 20 µl injected sample volume: 0–3 min: 92% solvent A; 3–18  
310 min: linear gradient from 92% to 58% solvent A; 18–22.5 min: 15% solvent A; and 22.5–26 min: 92%  
311 solvent A, always at constant 0.2 mL/min flow rate. Eluted peptides were ionized using the positive  
312 electrospray source and analyzed in MRM mode. The source gas flow was set at 50 L/h and the source  
313 voltage at 2.5 kV for the capillary and 30 V for the cone. The source temperature was set at 150 °C and  
314 the desolvation temperature at 400 °C with a gas flow at 1200 L/h. Targeted transitions are summarized  
315 in Table 1. For each peptide, three transitions were analyzed, as well as the corresponding transitions  
316 for the related isotopically labelled internal standard(s) (peptides, concatemer, and  $\beta$ -lactoglobulin). The  
317 transitions were selected beforehand using criteria that included the MS signal intensity and the absence  
318 of interference for the three considered matrices. The MS/MS acquisition method was generated using  
319 the open source Skyline software (MacLean et al., 2010). The most intense transition was used for  
320 internal standard comparison calculation and the two others as confirmatory transitions. Internal  
321 standards were compared using the peak area ratio (for the most intense transition) between the peptide

322 from the allergenic ingredient and its corresponding isotopically labelled version from the internal  
323 standard.

### 324 3. Results and discussion

#### 325 3.1. Choice of the isotope labelling strategy

326 Stable isotope internal standard labelling and associated isotopic enrichment are key elements in  
327 the design of quantitative mass spectrometry-based methods. The isotopic enrichment and mass shift  
328 combination has to be sufficient to avoid any potential risk of false positive introduction. The resolution  
329 of quadrupole analyzers is typically around 1 atomic mass unit (Georgiou & Danezis, 2015). Taking into  
330 account that most of the peptide ion precursors carry multiple charges and that peptides contain tens of  
331 carbons, which lead to widespread isotopic distribution (see Fig. 1), the mass shift introduced by the  
332 stable isotopes has to be sufficient to be able to totally distinguish the natural analyte from its internal  
333 standard. Considering these aspects, a mass shift of  $m/z \geq 3$  is necessary. Furthermore, attention has  
334 to be paid to the actual isotope enrichment. Depending on the labelling strategy, an insufficient isotope  
335 enrichment may lead to the introduction of the unlabeled form of the internal standard, thus  
336 corresponding to the natural analyte itself and contaminating the quantitative analysis.

337 Several strategies have been developed to produce isotopically labelled proteins, including selective  
338 labelling using auxotrophic *E. coli* strains and growth medium supplemented with isotopically labelled  
339 amino acids (Mondal, Shet, Prasanna, & Atreya, 2013) or post-translational protein deuteration (Galan  
340 et al., 2018). In this study, we decided to use a rich bacterial cell growth medium specifically designed  
341 for  $^{15}\text{N}$  labeling protein using *E. coli* as a host cell for recombinant protein expression. This original  
342 medium is an algal hydrolysate that contains the same level of amino acids as LB medium. This strategy  
343 allowed for stable and protein sequence independent labelling (as each amino acid contains at least  
344 one nitrogen) with a high isotopic enrichment. As one of the peptide biomarkers selection criteria  
345 concerned the actual peptide length (peptides should have at least 8 amino acids),  $m/z \geq 3$  mass shift  
346 precaution is respected for triply charged precursor. Indeed, selected peptide biomarkers are tryptic  
347 peptides, with a lysine or an arginine in C-terminal position, holding two and four nitrogen atoms,  
348 respectively.

#### 349 3.2. Characterization of $^{15}\text{N}$ isotopically labelled proteins

350 The isotopic enrichment in the concatemer and  $\beta$ -lactoglobulin was evaluated following analysis of  
351 their constitutive tryptic peptides by UHPLC-HRMS. The proportion of the fully  $^{15}\text{N}$  labelled version of  
352 each tryptic peptide was estimated by comparing the intensities of the monoisotopic peak ( $\text{U-}^{15}\text{N}$  &  $\text{U-}$   
353  $^{12}\text{C}$ ) and those of its isotope containing one  $^{14}\text{N}$  isotope ( $[\text{U-1}]^{15}\text{N}$  &  $\text{U-}^{12}\text{C}$ ). As shown in Fig. 1, the

354 intensities of the peaks from peptides with more than one  $^{14}\text{N}$  isotope were found to be negligible  
355 (relative peak intensity <1% compared to the  $[\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}]$  peak). Given the resolution of the MS  
356 system (40000), carbon and nitrogen isotopes could not be distinguished. As a result, the monoisotopic  
357 peak ( $\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) was combined with the peak corresponding to the peptide with one  $^{14}\text{N}$  and one  
358  $^{13}\text{C}$  isotope ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ ^{13}\text{C}_1$ ). The proportion of fully  $^{15}\text{N}$  labelled peptide was evaluated by comparing  
359 ( $\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) and ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) peak intensities. The part of the peak intensity corresponding  
360 to the ( $\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) isotope therefore had to first be discriminated from the combined ( $\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ )  
361 and ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ ^{13}\text{C}_1$ ) isotopes' peak intensity. Since isotopes with more than one  $^{14}\text{N}$  were found to be  
362 negligible, we assumed that the ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) isotope peak would only correspond to this  
363 combination of isotopes. The peak intensity of the ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ ^{13}\text{C}_1$ ) isotope could therefore be predicted  
364 from the ( $[\text{U-1}]\text{-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) isotope peak intensity assuming a 1.1% natural abundance of  $^{13}\text{C}$  isotopes  
365 and knowing the number of carbon atoms in the peptide. With this prediction, the ( $\text{U-}^{15}\text{N} \ \& \ \text{U-}^{12}\text{C}$ ) isotope  
366 peak intensity could be deduced from the combined isotopes' peak intensity.

367       The proportion of fully  $^{15}\text{N}$  labelled peptide was evaluated for all the 19 concatenated tryptic peptides  
368 of the concatemer and for all identified tryptic peptides from  $\beta$ -lactoglobulin. As shown in Fig. 2, the  
369 relation between the labelling proportion and the number of nitrogen atoms in the peptides follows an  
370 exponential decay. The associated exponential decay constant corresponds to the natural logarithm of  
371 the isotopic enrichment. Indeed, for a given isotopic enrichment ( $\varphi$ ), the proportion of fully  $^{15}\text{N}$  labelled  
372 peptide with  $n$  nitrogens is given by  $\varphi^n$ , which can be transformed into  $e^{\ln(\varphi)*n}$ . Isotopic enrichment is  
373 deduced from this mathematical transformation by equating  $\ln(\varphi)$  to experimentally obtained  
374 exponential arguments ( $-0.00446$  for the concatemer and  $-0.00411$  for  $\beta$ -lactoglobulin). These results  
375 give an isotopic enrichment of 99.5% for the concatemer and 99.6% for  $\beta$ -lactoglobulin, and are in  
376 agreement with the >98% isotopic enrichment of the growth medium.

377       By using a method for efficient isotopic labelling of recombinant protein, we demonstrated that the  
378 purified  $^{15}\text{N}$  isotopically labelled concatemer and  $\beta$ -lactoglobulin internal standards fulfilled the required  
379 criteria regarding isotopic enrichment and the introduced mass shift. With this  $^{15}\text{N}$  uniform labelling  
380 strategy, the introduced mass shift was sufficient to distinguish the internal standard from the natural  
381 analyte using the quadrupole analyzer. The lowest mass shift corresponded to the double charged  
382 FYTVISSLK peptide (from egg white ovotransferrin), one of the 19 concatenated peptides, which  
383 contained 10 nitrogen atoms and an associated mass shift of a  $m/z$  of 5. Such a mass shift and obtained  
384 isotopic enrichment combination prevented the risk of false positive introduction.

### 385           **3.3. Comparison of isotopically labelled internal standards**

386           Performance of the three types of isotopically labelled internal standards (peptides, concatemer,  
387 and protein) were evaluated following analysis of three food matrices (baked cookie [cookie], chocolate,  
388 and lyophilized unbaked cookie dough [dough]). In theory, a perfect internal standard would have the  
389 same exact behavior as its corresponding analyte during sample preparation and analysis. Hence, any  
390 analyte loss or matrix effect (during sample preparation or UHPLC-MS/MS analysis) which affects the  
391 analyte should equally affect the internal standard. Consequently, for a given natural analyte  
392 concentration and internal standard spike level, the signal ratio between a natural analyte and the  
393 internal standard would remain constant, independent of analyte losses and matrix effects. The three  
394 internal standards considered in this study were compared based on this correlation.

395           Similar matrix-matched calibration curves were prepared for the three matrices. These curves  
396 included a blank and five allergen concentrations ranging from 2.5 to 50 ppm (expressed in mg total  
397 allergen protein per kg of matrix), with each sample prepared in triplicate. For each combination of matrix  
398 and allergen contamination level, the appropriate internal standard(s) (isotopically labelled peptides and  
399 U-<sup>15</sup>N β-lactoglobulin for the first sample series, and U-<sup>15</sup>N concatemer for the second one) was spiked  
400 at the same concentration. Results are presented separately for each targeted peptide and its  
401 corresponding internal standard (five synthetic peptides, 19 allergenic tryptic peptides from U-<sup>15</sup>N  
402 concatemer digestion, and four tryptic peptides from U-<sup>15</sup>N β-lactoglobulin digestion). Representative  
403 peptides of each internal standard are shown in Fig. 3, and complete results are shown in  
404 Supplementary data 4. Performance of the different internal standards were evaluated by comparing  
405 the peak area ratio for the most intense transition (highlighted in Table 1) between the analyte and its  
406 corresponding internal standard for the three considered matrices. As shown, for a given analyte and  
407 internal standard concentration, the signal ratio remained constant when the internal standard was  
408 effective. The overlay of the generated linear regression lines was therefore used to evaluate internal  
409 standards performance. Overlapping regression lines indicated, for each allergen contamination level,  
410 a constant peak area ratio among matrices and thus, an effective internal standard, compensating for  
411 matrix effects. In addition to visual evaluation, overlapping regression lines were evaluated using the  
412 coefficient of variation (CV) between the slopes of the linear regression lines.

413           Overall, the best results were obtained for the isotopically labelled protein, U-<sup>15</sup>N β-lactoglobulin.  
414 Assuming that recombinant protein folding was similar to the native protein and that the introduced N-  
415 terminal mutations had no significant impact, as previously demonstrated (Loch et al., 2016), this

416 approach seemed to be the one best suited one for quantifying allergen proteins. Aside from their mass  
417 (given the mass shift introduced by isotope labelling), both the analyte protein and the internal standard  
418 protein must have had the same properties. This was confirmed by the analysis of four constitutive  
419 tryptic peptides from  $\beta$ -lactoglobulin. Regression lines overlapped with all CV values below 15%. This  
420 confirmed that the internal standard had efficiently balanced matrix effects during sample preparation  
421 and UHPLC-MS/MS analysis, further supported by the fact that the analyte absolute peak area varied  
422 by a factor of up to 10 among the three considered matrices, depending on the peptide (data not shown)  
423 while the analyte/internal standard peak area ratio remained constant. However, the labelled protein  
424 was spiked into the different samples after food processing, which is known to impact peptide  
425 detectability and quantification (Korte et al., 2019; Parker et al., 2015). Peptide biomarker selection is  
426 therefore a crucial preliminary step in the development of a quantitative method, and selected peptides  
427 have to be robust to the food process.

428 The results obtained with isotopically labelled peptides and the concatemer were less  
429 straightforward to interpret. For some targeted peptides, such as LSFNPTQLEEQCHI with labelled  
430 peptides, or TNDNAQISPLAGR with the U-<sup>15</sup>N concatemer, the internal standard efficiently  
431 compensated for matrix effects with observed CV values below 15%. However, for some other targeted  
432 peptides, such as GGLEPINFQTAADQAR with both U-<sup>15</sup>N concatemer and labelled peptides, the  
433 analyte and internal standard signal ratio was highly matrix-dependent. In these cases, internal  
434 standards did not correctly balance matrix effects, potentially leading to biased allergen quantification.  
435 These results are consistent with those reported by Planque and co-workers (Planque et al., 2019). No  
436 significant difference was observed for the three peptides which were common to synthetic peptides and  
437 concatemer used as internal standards.

438 Isotopically labelled peptides are not subject to one of the crucial steps during sample preparation,  
439 this being the proteolytic digestion with trypsin. The composition of the food matrix directly impairs the  
440 efficiency of enzymatic digestion at least in two different ways. First, different matrices have different  
441 protein concentrations, directly affecting the protein/enzyme ratio. Labelled peptides do not balance for  
442 this aspect. Second, some other sample components, such as polyphenols and tannins, may also affect  
443 the efficiency of trypsin digestion (Gonçalves, Mateus, Pianet, Laguerre, & De Freitas, 2011), which  
444 might help to explain why the chocolate matrix gave lower signals for most of targeted peptides. Contrary  
445 to the labelled peptides, the U-<sup>15</sup>N concatemer needed to be digested by trypsin to yield peptides that  
446 could be detected upon UHPLC-MS/MS analysis. Therefore, factors such as the sample protein content

447 or the presence of tannins should be balanced when using such an internal standard. However, our  
448 results indicated that the performance of the concatemer was peptide-dependent. For some peptides,  
449 such as TNDNAQISPLAGR from hazelnut Cor a 9 allergen and FFVAPFPEVFGK from milk  $\alpha$ <sub>S1</sub>-casein,  
450 matrix effects were efficiently balanced with linear regression lines CV below 15% between the matrices.  
451 However, for other peptides, such as NVNFDGEILK from egg vitellogenin-1 and TPEVDDEALEK from  
452 milk  $\beta$ -lactoglobulin, the associated CVs were much higher (>30%).

453 Matrix effects can also affect analytes by other means. Robustness to food processing was one of  
454 the criteria for peptide biomarkers selection (Gavage et al., 2019, 2020; Van Vlierberghe et al., 2020)  
455 and can therefore be excluded. Variation in protein extraction can also be excluded as, for all selected  
456 proteins, multiple peptides were included in the U-<sup>15</sup>N concatemer, and no general trend of the matrix  
457 effect was observed for all the peptides of a given protein. Indeed, if protein extraction of the analyte  
458 and/or the internal standard was affected by the matrix, all peptides from a given protein should be  
459 equally impacted, which was not observed.

460 Proteolytic digestion of extracted proteins is a key step in sample preparation and could be a source  
461 of the observed variability. Even if the concatemer internal standard needs to be digested to release its  
462 constitutive peptides, multiple factors could influence the digestion kinetics. For instance, amino acids  
463 surrounding trypsin recognition sites are known to influence the efficiency of peptide bond hydrolysis  
464 (Siepen, Keevil, Knight, & Hubbard, 2007). Cleavage sites are described using the nomenclature  
465 formulated by Schechter and Berger (Schechter & Berger, 1967), as P4-P3-P2-P1-P1'-P2'-P3', in which  
466 cleavage of the peptide bond occurs between P1 and P1'. Arginine, lysine, and proline in position P1'  
467 have, for instance, a negative effect on the digestion efficiency. The acidic amino acids aspartate and  
468 glutamate also negatively influence digestion when they are present near the cleavage site. These  
469 aspects were taken into account during peptide biomarkers selection, and sequences known to  
470 negatively affect trypsin digestion were rejected. However, peptide biomarkers were synthetically  
471 stitched together in the concatemer. Considering a given peptide in the concatemer, its cleavage site is  
472 surrounded at the N-terminal side (P4 to P1) by amino acids from this peptide but also by amino acids  
473 from its neighboring peptide at the C-terminal side (P1' to P3'). Consequently, at a local scale, enzymatic  
474 digestion of the concatemer only partially reflects digestion of the natural proteins. This difference  
475 between natural analytes and concatemers might lead to differences in enzymatic digestion kinetics and  
476 could have been a source of the observed variations. A relatively simple solution to overcome this would  
477 be the introduction of amino acids between each targeted peptide of the concatemer. Such introduced

478 amino acids could be the flanking amino acids in the corresponding natural protein sequence, a solution  
479 known as a peptide-concatenated standard (PCS) (Kito, Ota, Fujita, & Ito, 2007). However, amino acids  
480 surrounding the cleavage site in the three-dimensional structure of the protein might also affect trypsin  
481 digestion. Hence, cleavage sites surrounded by acidic amino acids, characterized by a greater average  
482 exposed area, are more subject to missed-cleavages.

483 Besides flanking amino acids, structural parameters also interfere with enzymatic digestion of a  
484 protein. According to the work of Hamady and co-workers (Hamady, Cheung, Tufo, & Knight, 2005),  
485 secondary protein structures affect trypsin digestion efficiency. Cleavage sites within unstructured  
486 domains are more prone to be cleaved incorrectly, whereas cleavage sites in alpha-helices are more  
487 favorable. The structures of proteins targeted by the UHPLC-MS/MS method, when available, were  
488 analyzed to define whether observed variability among peptides could be linked to findings of Hamady  
489 and co-workers or not (Hamady et al., 2005). No general trend emerged from our data, limited to the 19  
490 concatenated peptides. However, three-dimensional and structural aspects could be included in a future  
491 peptide biomarker selection, in addition to all other criteria already considered in this study.

#### 492 **4. Conclusions**

493 Mass spectrometry-based detection and quantification of food allergens in processed food products  
494 remains challenging. Currently, no threshold values for allergen trace-level contamination have been  
495 established in European legislation, but these are highly expected by all stakeholders involved in the  
496 food chain, from producers to control laboratories, and will require quantitative analysis methods.  
497 Quantitative methods based on stable dilution techniques need isotopically labelled internal standards.

498 Here, we presented and compared the performances of three different types of isotopically labelled  
499 internal standards for allergen analysis in processed food products: synthetic peptides, concatemer, and  
500 protein. These internal standards were compared through the analysis of three matrix-matched  
501 calibration curves (cookie, chocolate, and unbaked lyophilized cookie dough) for four targeted allergens  
502 (egg, milk, peanut, and hazelnut). An effective internal standard needs to behave similar to the natural  
503 analyte and is therefore identically impacted by matrix effects during sample preparation and UHPLC-  
504 MS/MS analysis. As expected from a theoretical point of view, the isotopically labelled protein that was  
505 used as an internal standard gave the best results. A constant signal ratio between the analyte and the  
506 internal standard peak areas was observed in all matrices tested for the four tryptic peptides generated  
507 from the studied protein. However, we need to emphasize that these results only come from one  
508 investigated protein,  $\beta$ -lactoglobulin from milk.

509 Results from our studies using peptides and the concatemer were more equivocal and seemed to  
510 be peptide-dependent. For some synthetic peptides or some tryptic peptides from the concatemer,  
511 matrix effects during sample preparation and UHPLC-MS/MS analysis could be efficiently countered by  
512 the applied internal standards, whereas for other peptides, significant matrix effects were observed.  
513 However, the non-inferiority of the results obtained for the tryptic peptides from the concatemer was  
514 established, when compared to synthetic peptides. Moreover, the addition of any synthetic peptide in a  
515 method represent an additional cost, limiting therefore the number of targeted peptides for routine  
516 laboratories. The concatemer production costs are relatively independent of the number of concatenated  
517 tryptic peptides. From a rough estimate of ten peptides, the use of a concatemer as internal standard is  
518 financially advantageous and supersedes synthetic peptides.

519 Even though isotopically labelled synthetic peptides are currently the most commonly used internal  
520 standard for allergen analysis, they do not exactly reflect the natural situation as they do not need to be  
521 subjected to proteolytic digestion, while part of the variability observed in our study could have come  
522 from proteolytic digestion. Concatemers clearly need to be digested to release their constituting

523 peptides. However, our data seem to indicate that the digestion of the concatemer could be improved  
524 to more efficiently represent analyte protein digestion. In this respect, introducing flanking amino acids  
525 between each individual peptide (i.e. the PCS strategy) could be a future asset. Moreover, for our  
526 concatemer construct, peptide biomarker selection was mainly focused on robustness to food  
527 processing and local sequences, but additional criteria, such as protein structure and the local digestion  
528 site environment, could be included in the peptide selection process. Such possible future improvements  
529 strongly suggest that isotopically labelled concatemers could represent relevant internal standards, as  
530 they overcome limitations of the use of synthetic peptides, while combining advantages of the use of  
531 labelled proteins and, further, allowing for multiple allergen quantification by mass spectrometry.

532 **Acknowledgments**

533 The research that yielded these results was funded by the Belgian Federal Public Service of Health,  
534 Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS. The authors also  
535 want to gratefully acknowledge the contributions of the University of Namur Mass  
536 Spectrometry platform (MaSUN).

537

538 **5. Bibliography**

- 539 Ahsan, N., Rao, R. S. P., Gruppuso, P. A., Ramratnam, B., & Salomon, A. R. (2016). Targeted  
540 proteomics: Current status and future perspectives for quantification of food allergens. *Journal of*  
541 *Proteomics*. <https://doi.org/10.1016/j.jprot.2016.04.018>
- 542 Allen, K. J., Remington, B. C., Baumert, J. L., Crevel, R. W. R., Houben, G. F., Brooke-Taylor, S., ...  
543 Taylor, S. L. (2014). Allergen reference doses for precautionary labeling (VITAL 2.0): Clinical  
544 implications. *Journal of Allergy and Clinical Immunology*.  
545 <https://doi.org/10.1016/j.jaci.2013.06.042>
- 546 Boo, C. C., Parker, C. H., & Jackson, L. S. (2018). A targeted LC-MS/MS method for the simultaneous  
547 detection and quantitation of egg, milk, and peanut allergens in sugar cookies. *Journal of AOAC*  
548 *International*. <https://doi.org/10.5740/jaoacint.17-0400>
- 549 Boyce, J. A., Assa'ad, A., Burks, A. W., Jones, S. M., Sampson, H. A., Wood, R. A., ... Schwaninger,  
550 J. M. (2011). Guidelines for the Diagnosis and Management of Food Allergy in the United States:  
551 Summary of the NIAID-Sponsored Expert Panel Report. *Journal of the American Academy of*  
552 *Dermatology*. <https://doi.org/10.1016/j.jaad.2010.11.020>
- 553 Chambers, S. P., Austen, D. A., Fulghum, J. R., & Kim, W. M. (2004). High-throughput screening for  
554 soluble recombinant expressed kinases in Escherichia coli and insect cells. *Protein Expression*  
555 *and Purification*. <https://doi.org/10.1016/j.pep.2004.03.003>
- 556 Galan, S. R. G., Wickens, J. R., Dadova, J., Ng, W. L., Zhang, X., Simion, R. A., ... Davis, B. G.  
557 (2018). Post-translational site-selective protein backbone  $\alpha$ -deuteration. *Nature Chemical*  
558 *Biology*. <https://doi.org/10.1038/s41589-018-0128-y>
- 559 Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N.  
560 (2019). Selection of egg peptide biomarkers in processed food products by high resolution mass  
561 spectrometry. *Journal of Chromatography A*, 1584, 115–125.  
562 <https://doi.org/10.1016/J.CHROMA.2018.11.036>
- 563 Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N.  
564 (2020). High-resolution mass spectrometry-based selection of peanut peptide biomarkers  
565 considering food processing and market type variation. *Food Chemistry*.  
566 <https://doi.org/10.1016/j.foodchem.2019.125428>
- 567 Georgiou, C. A., & Danezis, G. P. (2015). Elemental and Isotopic Mass Spectrometry. *Comprehensive*  
568 *Analytical Chemistry*, 68, 131–243. <https://doi.org/10.1016/B978-0-444-63340-8.00003-0>

569 Gonçalves, R., Mateus, N., Pianet, I., Laguerre, M., & De Freitas, V. (2011). Mechanisms of tannin-  
570 induced trypsin inhibition: A molecular approach. *Langmuir*. <https://doi.org/10.1021/la202280c>

571 Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L., & Roubos, J. A. (2015). Trade-offs between  
572 tRNA abundance and mRNA secondary structure support smoothing of translation elongation  
573 rate. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkv199>

574 Hamady, M., Cheung, T. H. T., Tufo, H., & Knight, R. (2005). Does protein structure influence trypsin  
575 miscleavage? *IEEE Engineering in Medicine and Biology Magazine*, 24(3), 58–66.  
576 <https://doi.org/10.1109/MEMB.2005.1436461>

577 Henrotin, J., Planque, M., Huet, A.-C., Marega, R., Lamote, A., & Gillard, N. (2019). Gluten analysis in  
578 processed foodstuffs by a multi-allergens and grain-specific UHPLC-MS/MS method: One  
579 method to detect them all. *Journal of AOAC International*. <https://doi.org/10.5740/jaoacint.19->  
580 0057

581 Kito, K., Ota, K., Fujita, T., & Ito, T. (2007). A synthetic protein approach toward accurate mass  
582 spectrometric quantification of component stoichiometry of multiprotein complexes. *Journal of*  
583 *Proteome Research*. <https://doi.org/10.1021/pr060447s>

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

588 Loch, J. I., Bonarek, P., Tworzydło, M., Polit, A., Hawro, B., Łach, A., ... Lewiński, K. (2016).  
589 Engineered  $\beta$ -Lactoglobulin Produced in *E. coli*: Purification, Biophysical and Structural  
590 Characterisation. *Molecular Biotechnology*. <https://doi.org/10.1007/s12033-016-9960-z>

591 MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., ... MacCoss,  
592 M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted  
593 proteomics experiments. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btq054>

594 Monaci, L., Losito, I., De Angelis, E., Pilolli, R., & Visconti, A. (2013). Multi-allergen quantification of  
595 fining-related egg and milk proteins in white wines by high-resolution mass spectrometry. *Rapid*  
596 *Communications in Mass Spectrometry*, 27(17), 2009–2018. <https://doi.org/10.1002/rcm.6662>

597 Mondal, S., Shet, D., Prasanna, C., & Atreya, H. S. (2013). High yield expression of proteins in *E. coli*  
598 for NMR studies. *Advances in Bioscience and Biotechnology*.  
599 <https://doi.org/10.4236/abb.2013.46099>

600 Newsome, G. A., & Scholl, P. F. (2013). Quantification of allergenic bovine milk  $\alpha$ 1-casein in baked  
601 goods using an intact  $^{15}\text{N}$ -labeled protein internal standard. *Journal of Agricultural and Food*  
602 *Chemistry*. <https://doi.org/10.1021/jf3015238>

603 Nitride, C., Nørgaard, J., Omar, J., Emons, H., Estes, M. J. M., & O'Connor, G. (2019). An  
604 assessment of the impact of extraction and digestion protocols on multiplexed targeted protein  
605 quantification by mass spectrometry for egg and milk allergens. *Analytical and Bioanalytical*  
606 *Chemistry*. <https://doi.org/10.1007/s00216-019-01816-z>

607 Parker, C. H., Khuda, S. E., Pereira, M., Ross, M. M., Fu, T. J., Fan, X., ... Jackson, L. S. (2015).  
608 Multi-allergen Quantitation and the Impact of Thermal Treatment in Industry-Processed Baked  
609 Goods by ELISA and Liquid Chromatography-Tandem Mass Spectrometry. *Journal of*  
610 *Agricultural and Food Chemistry*. <https://doi.org/10.1021/acs.jafc.5b04287>

611 Planque, Arnould, Delahaut, Renard, Dieu, & Gillard. (2019). Development of a strategy for the  
612 quantification of food allergens in several food products by mass spectrometry in a routine  
613 laboratory. *Food Chemistry*. <https://doi.org/10.1016/j.foodchem.2018.08.095>

614 Planque, Arnould, & Gillard. (2017). Food Allergen Analysis: Detection, Quantification and Validation  
615 by Mass Spectrometry. In *Allergen*. <https://doi.org/10.5772/intechopen.69361>

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

620 Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., & Beynon, R. J. (2006).  
621 Multiplexed absolute quantification for proteomics using concatenated signature peptides  
622 encoded by QconCAT genes. *Nature Protocols*. <https://doi.org/10.1038/nprot.2006.129>

623 Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical*  
624 *and Biophysical Research Communications*, 27(2), 157–162. [https://doi.org/10.1016/S0006-](https://doi.org/10.1016/S0006-291X(67)80055-X)  
625 [291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X)

626 Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and  
627 treatment. *Journal of Allergy and Clinical Immunology*. <https://doi.org/10.1016/j.jaci.2013.11.020>

628 Siepen, J. A., Keevil, E. J., Knight, D., & Hubbard, S. J. (2007). Prediction of missed cleavage sites in  
629 tryptic peptides aids protein identification in proteomics. *Journal of Proteome Research*, 6(1),  
630 399–408. <https://doi.org/10.1021/pr060507u>

631 Van Vlierberghe, K., Gavage, M., Dieu, M., Renard, P., Arnould, T., Gillard, N., ... Van Poucke, C.  
632 (2020). Selection of universal peptide biomarkers for the detection of the allergen hazelnut in  
633 food through a comprehensive, high resolution mass spectrometric (HRMS) based approach.  
634 *Food Chemistry*, 125679. <https://doi.org/10.1016/J.FOODCHEM.2019.125679>  
635