

Development of a LC-MS method for the discrimination between trace level *Prunus* contaminants of spices

Sophie E. Inman*, Kate Groves, Bryan McCullough, Milena Quaglia, Chris Hopley

LGC, Queens Road, Teddington, Middlesex TW11 0LY, United Kingdom



A B S T R A C T

The need for an analytical procedure for the identification of allergens present at trace levels in foods was highlighted by conflicting results in a case of contamination of the spice cumin. The application of a bottom-up proteomics experiment was investigated to identify marker peptides for potential contaminant nuts which could then be monitored with high specificity and sensitivity by selective reaction monitoring experiments. The method developed allowed for the distinction between two closely related *Prunus* species, almond and mahaleb, in two different spices, cumin and paprika. The paprika sample was found to be contaminated with almond and the cumin sample, contaminated at a much lower level, was found to be contaminated with mahaleb. The method could be applied to any protein-dense food matrix allergen so long as suitable control and reference samples can be acquired.

1. Introduction

The issue of food allergens has significant impacts on sufferers (Muraro et al., 2014) and as a result, food labelling legislation has been introduced in the EU to protect consumers (European Parliament and the Council of the European Union, 2011). The EU legislation on food labelling requires declaration of any of 14 allergen groups used as ingredients in a food. However, contamination of the food chain with allergens can also occur through unexpected sources, for example, as a result of carryover from equipment used for the processing of multiple food stuffs.

The UK Government Chemist (<https://www.gov.uk/government/organisations/government-chemist>) is, in defined circumstances, responsible for the investigation of disputes in the food chain. Pursuant to these duties two cases of contaminated spices were referred to the Government Chemist for investigation as a result of apparent almond contamination. The first, in February 2015, was ground cumin (Food Standards Agency, 2015). The second, was ground paprika. These spices were analysed with multiple techniques (Burns et al., 2016; Government Chemist, 2015; Nixon, Hall, Wilkes, Walker, & Burns, 2016) to determine the *Prunus* contaminant species present. The presence of unrecognised almond in a food may present risks for people with almond allergy and there are significant economic impacts when products are withdrawn from the market hence testing procedures must be specific and robust.

The principal class of molecules that elicit an allergic reaction in food are proteins (Johnson et al., 2011). The use of immunochemical approaches such as ELISAs (enzyme-linked immunosorbent assays) facilitates high sensitivity for trace level protein contaminants. They also offer the advantage of being rapid and relatively inexpensive but the specificity of antibody recognition of target proteins may in some instances be compromised by peptide sequence homology (Walker, Burns, Elliott, Gowland, & Mills, 2016). The prunin protein, expressed in various proteoforms across the *Prunus* species as illustrated in Fig. 1, is an example of an allergenic protein that shows high sequence homology and may result in cross-reactivity.

The application of a polymerase chain reaction (PCR) experiment for the detection of an allergen facilitates high sensitivity for a specific species as it is able to amplify very low levels of DNA for detection. However, in this situation, the presence of DNA is used to infer the presence of the allergenic molecules, the proteins and although DNA is ubiquitous across all parts of the plant, the proteins may show localised expression profiles.

Mass spectrometry offers a number of advantages when considering the detection of protein allergens in food matrices. Firstly, it offers the stringent specificity required for confirmation of the presence of the analytes of interest especially when employing selected reaction monitoring (SRM) experiments on triple quadrupole instruments. Secondly, it offers a wide dynamic range when considering the abundance of the analyte and high sensitivity can be achieved if the correct sample

* Corresponding author.

E-mail addresses: Sophie.Inman@LGCgroup.com (S.E. Inman), Kate.Groves@lgcgroup.com (K. Groves), Bryan.McCullough@lgcgroup.com (B. McCullough), Milena.Quaglia@lgcgroup.com (M. Quaglia), Christopher.Hopley@lgcgroup.com (C. Hopley).

<http://dx.doi.org/10.1016/j.foodchem.2017.10.101>

Received 31 July 2017; Received in revised form 27 September 2017; Accepted 19 October 2017

Available online 21 October 2017

0308-8146/ © 2017 Elsevier Ltd. All rights reserved.

3EHK_A	Chain A, Crystal Structure Of Pru Du Amandin, An Allergenic Protein From <i>Prunus Dulcis</i>	
XP_007201716	Hypothetical protein PRUPE_ppa003759mg [<i>Prunus persica</i>]	
XP_008235678	PREDICTED: legumin type B-like [<i>Prunus mume</i>]	
1	-----ARQSQLSPQNQCQLNQLQAREPDNRIQAEAGQIETWNFNQGDQFCAGVAASRTITQRNGL	60
1	MAKASVFSLCLLLVFNGLA*****	80
1	MAKAEVFSLCLLLVFNGLA*****	80
61	HLPSYSSAPQLIYIVQGRGVLGAVFSGCPETFEESQQSSQQGRQQEQEQERQQQQQGEQGRQQGQQEQEQERQGRQQGRQ	140
81	*****SQQG*****GEQGRQQGQQE*****SQ	160
81	*****-----*****-----*****	143
141	QQEEGRQQEQQQGQQGRPQQQQQFRQLDRHQKTRRIREGDVVAIPAGVAYWSYNDGDQELVAVNLFHVSSDHNQLDQNP	220
161	QQEERR*****	240
144	-----*****	217
221	KFYLAGNPENEFNQGGSQPRQQGEQGRPGQHQQPFGRPRQQEQGGNGNVFSGFNTQLLAQALNVNEETARNLQGQNDN	300
241	*****EQGRPGQH*****	320
218	*****-----*****	289
301	RNQIIQVRGNLDFVQPPRGRQEREHEERQQEQQLQERQQQGEQIMANGLEETFCSLRLKENIGNPERADIFSPRAGRIST	380
321	*****	400
290	*****	369
381	LNSHNLPIRLFLRLSAERGFYRNGIYSPHWNVAHSVYVIRGNARQVNVNENGDAILDQEVQQGLFIVPQNHGVIQQ	460
401	*****	480
370	*****	449
461	AGNQGFYFAFKTEENAFINTLAGRTSFLRALPDEVLANAYQISREQARQLKYNRQETIALSSSQRRRAV	531
481	*****	551
450	*****	520

Fig. 1. The amino acid sequence alignments of the prunin 1 protein from three different *Prunus* species illustrating the sequence similarities (*).

preparation techniques are implemented. Thirdly, the sample preparation and method can be designed to target the particular class of molecule of interest. In the application of mass spectrometry for the detection of allergens, ideally, the target analyte would be the specific molecule giving rise to the allergenic response and a number of publications have reported the detection of peptides for the detection of allergens in foods (Sealey-Voyksner, Zweigenbaum, & Voyksner, 2016). All have taken different approaches to the sample preparation methods and use different criteria for the confirmation of the presence of particular foods.

The aim of this work was to develop a robust method for the detection of trace levels of *Prunus* nut protein contaminants, almond (*Prunus dulcis*) and mahaleb (*P. mahaleb*), in the spices cumin and paprika. Mahaleb was investigated as its presence had been found to give false-positive results in the detection for almond by the Canadian Food Inspection Agency at the same time as the contaminated cumin case was referred to the Government Chemist (Canadian Food Inspection Agency, 2015). To be able to differentiate between almond and mahaleb, required the identification of peptides unique to each species despite minimal information in the literature due to a lack of characterisation of the mahaleb proteome. Furthermore to improve the sensitivity of the method, the extraction procedure was maximised by applying an *in house* developed extraction procedure (Groves, Cryar, Walker, & Quaglia, in press). Once unique marker peptides had been identified, the spices were analysed for the presence or absence of the contaminants. The approach used here to identify and detect unique

peptide markers of food contaminants can be applied to different food matrices to improve food authenticity and allergen analysis.

2. Materials and methods

2.1. Samples

The samples referred to the Government Chemist for a definitive determination of which *Prunus* species was present were ground cumin (*Cuminum cyminum*) and ground paprika (*Capsicum annum*) hereafter referred to as 'referred cumin' and 'referred paprika' respectively.

Control materials were sourced from commercial retailers with traceable supply chains and processing procedures. Organic cumin seeds and almond nuts (*Prunus dulcis*) were finely ground *in house*. Kernels were taken from whole apricots (*Prunus armeniaca*) before grinding. The remaining samples, standard (blank) & smoked paprika and mahaleb (*Prunus mahaleb*), were acquired in ground form.

2.2. Chemicals

Water used for the sample preparation and mobile phases was from an ELGA purifier system (18.2 MΩ). Tris(hydroxymethyl)amino-methane (Tris-HCl), dithiothreitol (DTT), iodoacetamide (IAA) and TPCK-trypsin were from Sigma Aldrich (Poole, UK). Acetic acid and formic acid (FA) were from Fisher Scientific (Loughborough, UK). Acetonitrile (ACN) was from LGC Standards (Teddington, UK).

2.3. Extraction and digestion

Aliquots (0.15 g) of finely ground spice or nut were resuspended to 0.1 mg/mL in 100 mM Tris-HCl, pH 8. Disulfide bond reduction was achieved by the addition of 50 μ L of 200 mM DTT and incubation at 25 °C with shaking for 30 min. Reduced cysteine residues were then alkylated by the addition of 50 μ L of 400 mM IAA and incubation at 25 °C with shaking for 30 min. The alkylation reaction was quenched by the addition of a further 50 μ L of 200 mM DTT and incubation at 37 °C for 15 min.

A tryptic digestion cycle was performed. Thermal denaturation was achieved by heating the extract for 25 min in a 60 °C water bath with sonication. The samples were cooled to room temperature by incubation in a –20 °C freezer for five minutes. A 10 μ L aliquot of TPCK-trypsin, 1 μ g/ μ L in 1% acetic acid, was added to the cooled extracts before incubation at 37 °C with shaking for 30 min. This cycle was repeated until the fourth addition of trypsin at which point the digests were incubated at 37 °C overnight.

2.4. Peptide clean up

The overnight digests were centrifuged at approximately 12,000g (13,000 rpm) for 20 min and the supernatant removed. The supernatant was further clarified by centrifugation at 12,000g for an additional 20 min.

Solid phase extraction (SPE) was performed using SUPELLEAN™ LC-18 500 mg bed columns (Supelco, Sigma Aldrich). Columns were conditioned with 3 mL acetonitrile followed by equilibration with 3 mL 100 mM Tris-HCl, pH 8. Clarified digest supernatant (1 mL) was then loaded onto the column.

2.4.1. Cumin referee case conditions

The column was washed with 9 mL of 5% ACN, 0.1% (v/v) formic acid and eluted with 3 mL of 40% ACN, 0.1% (v/v) formic acid. The elution fractions were lyophilised using a vacuum centrifuge (AVC 2-22, Christ) at 6.0 mbar and –53 °C overnight. Prior to LC-MS/MS analysis, the lyophilised samples were resolubilised in 150 μ L of 0.1% formic acid in water with vigorous vortexing.

2.4.2. Optimised conditions

The column was washed with 9 mL of 10% ACN, 0.1% (v/v) formic acid and eluted with 1.5 mL of 30% ACN, 0.1% (v/v) formic acid. The elution fractions were lyophilised and resuspended as in the initial experiments.

2.5. Marker peptide selection

2.5.1. Literature peptides

An *in silico* digest was performed on the amino acid sequence of Pru du amandin (also known as almond major protein, AMP, or prunin, Fig. 1), the major allergenic protein in almond (Roux, Teuber, Robotham, & Sathé, 2001). For mahaleb, at the time of analysis, there were only 12 entries registered on UniProt. These translated to four proteins expected to be present in the leaves of the plant so simulated digests were not possible.

2.5.2. LC-DDA for marker peptide identification

The almond, mahaleb and apricot tryptic digests (prior to optimisation of the SPE procedure) were subjected to a data dependent analysis (DDA) experiment. Liquid chromatography (LC) was performed on an Agilent 1200 system coupled to an Agilent 6530 QToF operated in data dependent analysis (DDA) mode using Mass Hunter Data Acquisition (Agilent Technologies). A Waters BEH-C18 column (3.5 μ m, 2.1 $\times$ 150 mm) at 35 °C was used for separation with a 0.4 mL/min gradient from 5% A (0.1% FA in water) to 60% A over 20 min before a column wash with 100% B (0.1% FA in ACN) and re-equilibration to

starting conditions.

MS source conditions were: gas temperature 360 °C, drying gas 5 L/min, sheath gas temperature 400 °C, sheath gas flow 7.5 L/min, fragmentor 180 V, skimmer 70 V, capillary voltage 3500 V, nozzle voltage 1000 V. For the DDA experiments, MS survey scans were acquired for 250 ms from 250 to 2000 m/z and spectra were internally calibrated with a two-point lock mass calibration. The two most abundant peptides with a charge state greater than 2+ were selected for fragmentation with a collision energy gradient according to the equation: CE (V) = $(3.5 \times m/z)/(100 + (-1))$. The isolation window was set to medium ($\sim m/z$ 4) and fragmentation spectra were acquired for 500 ms.

The DDA data file was analysed in Mass Hunter Qualitative Analysis (Agilent Technologies) and exported as a MGF file for peptide and protein identification in MASCOT (Perkins, Pappin, Creasy, & Cottrell, 1999). An *in-house* Mascot server (version 2.2.06) was used. MASCOT search parameters included a 10 ppm peptide tolerance window and a 0.06 Da fragment ion tolerance window for sequences. There were no fixed modifications, but deamidation and methionine oxidation were allowed as variable modifications. The NCBI database (NCBI nr 100326, 10688764 sequences; 3647636407 residues) was used.

The 'ROUND' & 'MATCH' functions in Microsoft Excel were used to compile a list of precursor m/z values present in all the almond MGF files but none of the mahaleb MGF files and vice versa. The m/z values of the precursor ions of the identified peptides were used to generate extracted ion chromatograms from the MS survey scans and confirm their specificity to one species or the other. 'Unique' precursor m/z values were compared to the results of the MASCOT database search to determine which were assigned peptide sequences. The peptide sequences were further verified with targeted MS/MS on the QToF.

2.6. Skyline

The peptides selected as markers for almond or mahaleb were entered into Skyline (MacLean et al., 2010) to generate a list of transitions. The transitions and collision energies used were optimised for SRM using Skyline and the results manually verified. Each peptide was monitored by three transitions. Cysteine containing peptides were removed except for one of the peptides from the literature determined as unique to almond.

2.7. Marker peptide detection by SRM

High performance liquid chromatography (HPLC) was performed on an Waters Acquity H class system using the same Waters XBridge® Peptide BEH C18 column (130 Å, 3.5 μ m, 2.1 $\times$ 150 mm) as was used for the peptide identification experiments. Peptide separation was achieved with a gradient elution from aqueous (Buffer A, 0.1% FA in deionised water) to organic (Buffer B, 0.1% FA in ACN) mobile phases (0 min: 5% B, 2 min: 5% B, 4 min: 18% B, 14 min: 22% B, 16.5 min: 35% B, 17 min: 100% B, 19.5 min: 100% B, 19.7 min: reequilibrate at 5% B). The column eluent was diverted to waste for the first 5 and last 3.7 min of the gradient.

SRM of the peptides listed in Table 1 was performed on an ABSciex 4000 QTrap. Source parameters were optimised for peptide analysis: source temperature 650 °C, ion source gases 50 °C, IonSpray voltage 5500 V. The LC, divert valve and mass spectrometer were controlled by Analyst (AB Sciex). Some preliminary experiments were performed using the same LC conditions on an Agilent 1200 RR system coupled to an Agilent 6490 mass spectrometer, with source conditions optimised for the detection of peptides.

Table 1

The transitions monitored for the peptides of interest. The * symbol highlights the quantifier ions. The † symbol highlights the peptides used for confirming the presence of almond as found in the literature (Heick et al., 2011).

Specificity	Peptide	Precursor Ion / <i>m/z</i>		Fragment Ion / <i>m/z</i>		Decustering Potential /V	Collision Energy /V
Mahaleb	SGGQILPIR	2+	470.7824	y ₆ ¹⁺	739.4825	65.4	21.0
				y ₅ ¹⁺	611.4239	65.4	21.0
				*y ₄ ¹⁺	498.3398	65.4	26.0
	DFVSSPFR	2+	477.7376	y ₅ ¹⁺	593.3042	50.0	22.0
				y ₄ ¹⁺	506.2722	50.0	22.0
				*b ₂ ¹⁺	263.1026	50.0	23.0
	VPTPVPPRVSSPR	2+	694.9041	y ₁₀ ¹⁺	1091.6320	81.8	30.0
				y ₈ ¹⁺	895.5108	81.8	30.0
				*y ₂ ¹⁺	272.171	81.8	35.3
Almond	ISTLNSHNPILR	3+	493.2877	y ₁₁ ²⁺	639.3699	67.1	22.4
				b ₂ ¹⁺	201.1234	67.1	22.4
				*y ₁₂ ²⁺	682.8859	67.1	22.4
	GVLGAVFGSCPETFEESSQQSSQGR [†]	3+	895.7452	y ₇ ¹⁺	790.3802	96.4	34.9
				y ₃ ¹⁺	360.1990	96.4	34.9
				*y ₆ ¹⁺	662.3216	96.4	34.9
	TEENAFINTLAGR	2+	718.3624	y ₆ ¹⁺	631.3522	83.5	36.7
				y ₇ ¹⁺	744.4363	83.5	36.7
				*y ₈ ¹⁺	891.5047	83.5	36.7
Prunus	FVSPAYR	2+	420.2241	b ₂ ¹⁺	247.1441	45.0	19.7
				y ₄ ¹⁺	506.2722	45.0	19.7
				*y ₅ ¹⁺	593.3042	45.0	19.7
	GNLDFVQPPR [†]	2+	571.8013	y ₇ ¹⁺	858.4468	55.0	27.0
				y ₄ ¹⁺	497.2831	55.0	28.3
				*y ₃ ¹⁺	369.2245	55.0	25.0
	VQGQLDFVSPFRS	2+	740.3832	y ₅ ¹⁺	593.3042	85.1	37.9
				y ₈ ¹⁺	954.4680	85.1	37.9
				*y ₁₁ ¹⁺	1252.6321	85.1	34.0
	ALPDEVLANAYQISR [†]	2+	830.4387	y ₉ ¹⁺	1035.5582	91.7	44.0
y ₇ ¹⁺				851.4370	91.7	43.1	
*y ₈ ¹⁺				922.4741	91.7	44.0	
Literature	NGLHLPSYSNAPQLIYIVQGR [†]	3+	780.8	y ₆ ¹⁺	735.7	N/A	N/A
				b ₁₁ ¹⁺	1154.7	N/A	N/A

3. Results & discussion

3.1. Initial experiments

Initial experiments were performed on five neat “nut” samples (almond kernel, almond shell, inner and outer apricot kernel and mahaleb) and the control and referred cumin samples by replicating a literature method that extracted proteins from the matrix prior to tryptic digestion (Heick, Fischer, & Pöpping, 2011). Briefly, extraction was performed on 2 g of ground material by incubation in Tris buffer. Extracted proteins were subjected to a size exclusion step (using BioRad spin columns) for sample clean up prior to concentration by ultra-filtration (Amicon Ultra). Heick et al. reported four peptides from the prunin protein as indicative of the presence of almond. When analysed by LC-MS/MS on the Agilent 6490 QqQ, the peptides could be detected in the neat “nut” samples but not in the referred cumin. Furthermore, the peptides were found not to be specific to the almond sample.

To improve recovery of the nut peptides from the spice matrix, an *in house* thermal cycling digestion protocol (Groves et al., *in press*) was applied in which extraction and digestion should occur simultaneously and the repeated cycles expose more of the solid matrix for extraction. Although this did improve the nut protein recovery and the peptides were detected in the referred cumin, the lack of specificity of the peptides meant the contaminating species could not be unequivocally

assigned as almond.

3.2. Novel peptide identification

Further almond, apricot and mahaleb samples were subjected to the *in house* thermal cycling digestion protocol and an SPE clean-up procedure (5% ACN wash, 40% ACN elution). A layer of fat was clearly visible after the centrifugation step. The SPE eluents were pooled before analysis by LC-MS/MS using DDA experiments and triplicate injections. The fragmentation data was exported as MGF files to allow for peptide identification with a MASCOT powered database search. Initially the Swissprot database was used for peptide ID however, due to the lack of verified protein sequences from the *Prunus* species, a poor number of protein hits were returned and the larger NCBI database was used.

The MGF files were further analysed manually using Excel as described in the Materials & Methods. Each stage of the data processing protocol reduced the number of peptides: approximately 800 *m/z* values selected for fragmentation were filtered down to approximately 30 *m/z* values for manual investigation. Extracted ion chromatograms were generated for these 30 *m/z* values of the precursor ions (± 10 ppm) to confirm their uniqueness to almond or mahaleb, their retention times and signal intensity. If they were found to be unique by high resolution MS, the result of the Mascot database search was used to give a peptide sequence. The peptide sequence was verified with a targeted MS/MS

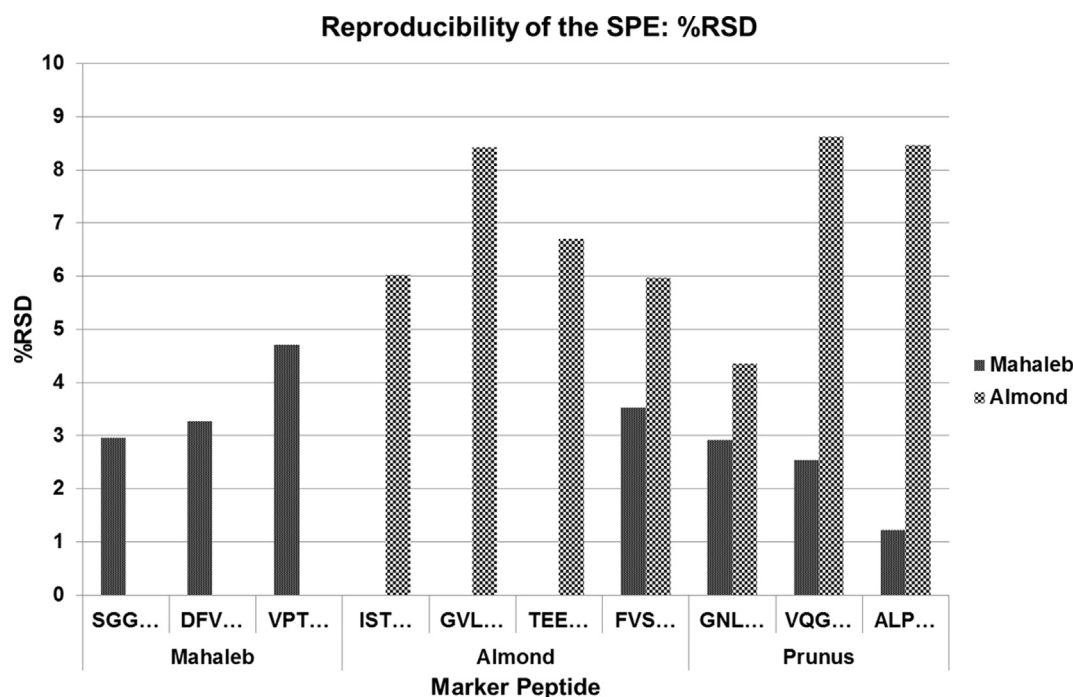


Fig. 2. The RSD of the SPE method was assessed from the mean and standard deviation of the summed peak area of the three transitions for each peptide from $n = 3$ experimental replicates of a pooled digest sample.

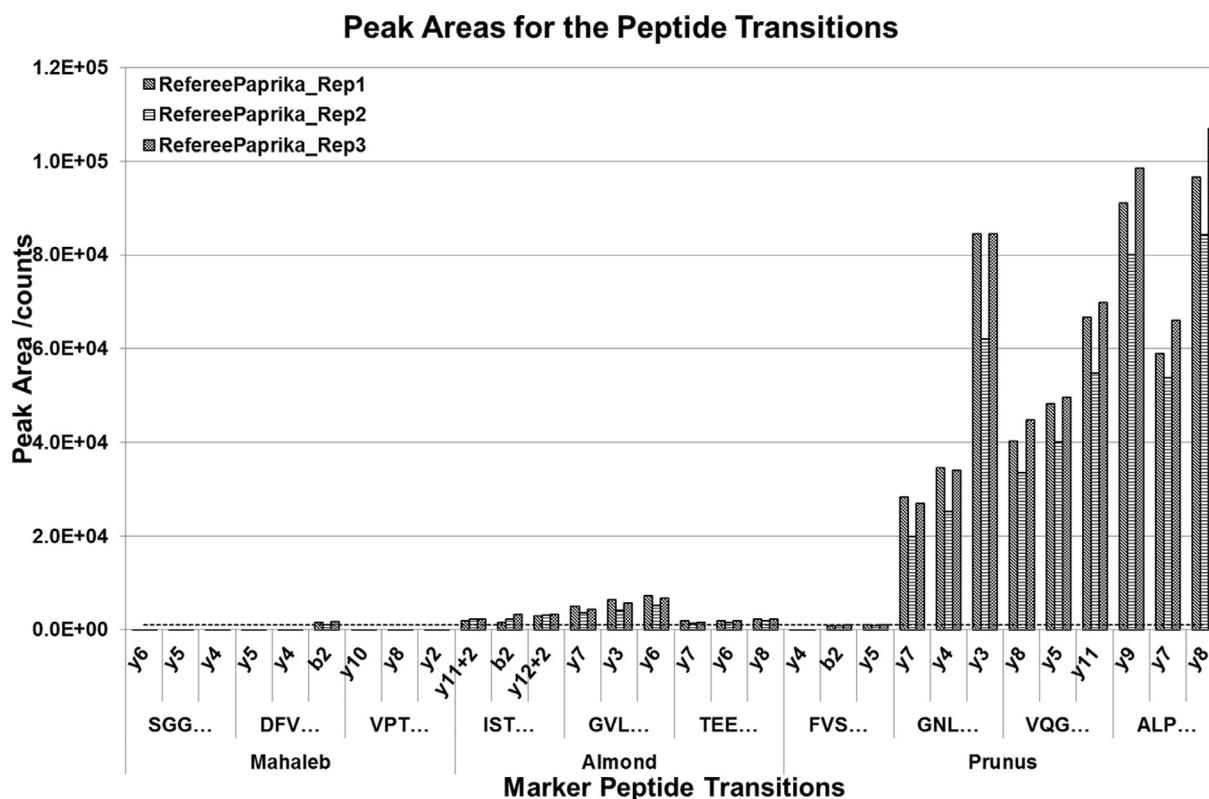


Fig. 3. Integrating the peak areas for the three transitions monitored for each peptide of interest. The dotted lines shows the 1000 count threshold used for determining whether the peak was genuine although manual confirmation was also performed.

experiment.

The final list of peptides of interest was entered into Skyline for prediction of optimal transitions and collision energies although the targeted MS/MS data was also used to decide which transitions to use. The almond, mahaleb and apricot samples were further analysed with a targeted SRM method on the Agilent 6490 QqQ instrument to confirm

the peptides were distinctive for almond and mahaleb. This removed a number of peptides from the analysis since, as with the literature peptides, they were not specific to almond or mahaleb. The sensitivity towards the mahaleb peptides was poor and thus a third class of peptides was included: 'Prunus'. This contained peptides that were markers of both almond and mahaleb. Given a number of the 'Prunus' peptides

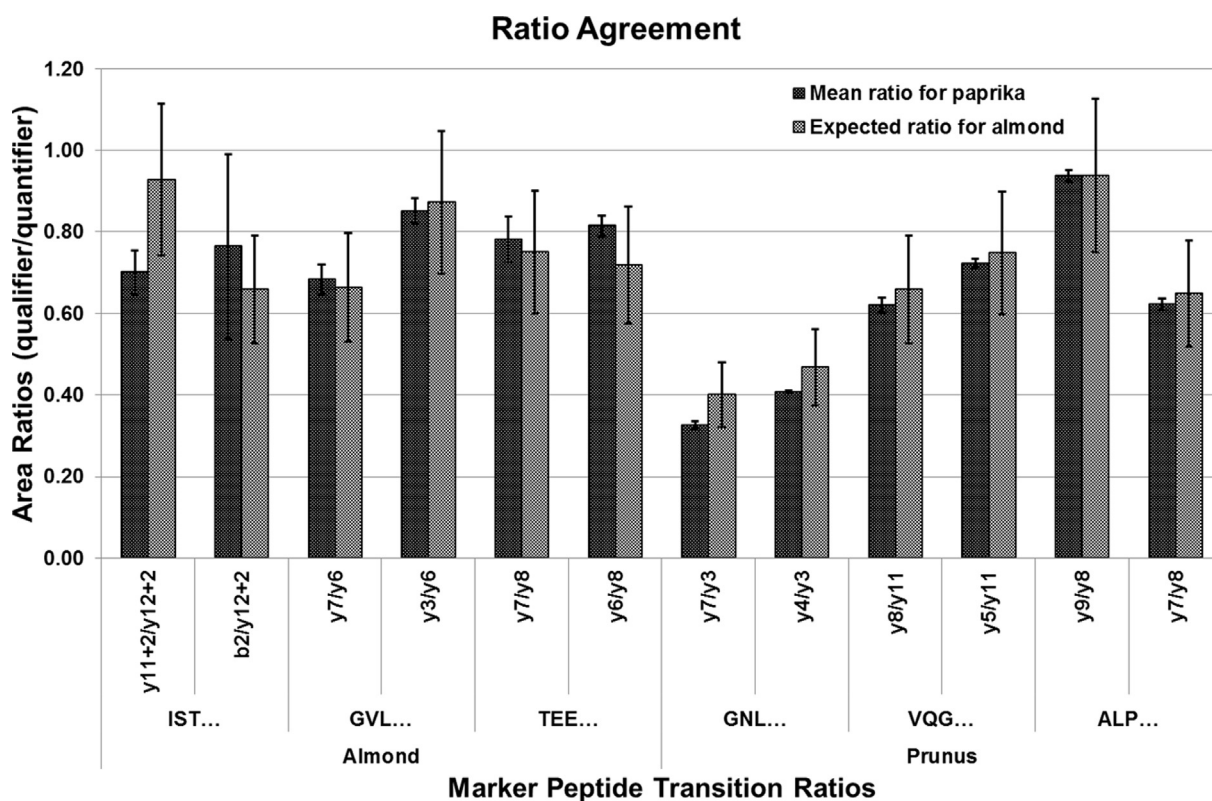


Fig. 4. Taking the ratio of the qualifier ions to the quantifier ions gives another layer of specificity to the data as they can be characteristic of the analyte of interest. The error bars on the mean ratio show one standard deviation. The error bars on the expected ratio show the boundaries of a 20% variation.

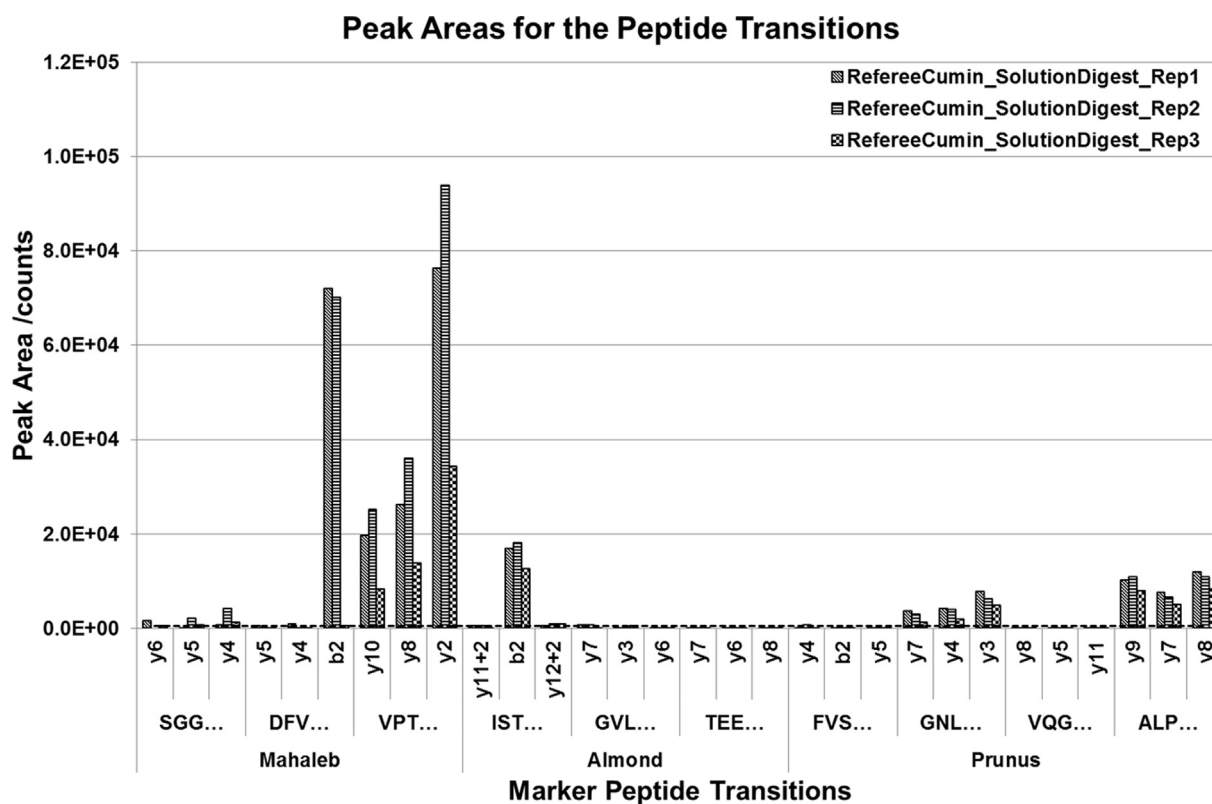


Fig. 5. Integrating the peak areas for the three transitions monitored for each peptide of interest. The dotted lines shows the 500 count threshold used for determining whether the peak was genuine although manual confirmation was also performed.

were derived from the prunin protein, they were present at high abundance.

3.3. Method optimisation

Initial targeted SRM experiments were performed on the blank and referred cumin samples. Although these experiments confirmed mahaleb was the contaminating species, they indicated further method development would be desirable to improve the sensitivity. Although it would have been possible to spike the blank cumin with almond or mahaleb, at the trace levels required, it would have been difficult to ensure the blend was homogeneous. The referred paprika sample was deemed an appropriate material for further optimisation of the sample preparation procedures as it was an incurred material with the contamination homogeneously distributed across the sample and ELISA results had previously suggested contamination at a level of approximately 500 mg of almond per kg⁻¹ (Nixon et al., 2016).

The initial experiments revealed the marker peptides eluted within a narrow concentration range of the LC gradient and thus optimal separation was achieved with a shallow elution gradient from 18 to 22% ACN over 10 min. As a result, the SPE elution conditions were investigated. For optimal clean-up of the marker peptides, the concentration of the ACN in the wash and elution solutions was investigated. A concentration of 10% (v/v) ACN, 0.1% FA was found optimal for the wash solution and a 1.5 ml elution of 30% (v/v) ACN, 0.1% FA gave the optimal conditions for mahaleb and almond peptide detection. The variability of this method was investigated by pooling three replicate extraction-digestion supernatants and separating them into three SPE load aliquots. After analysis by LC-MS, the %RSD in the integrated peak areas was less than 10% in all instances, Fig. 2, when investigating neat mahaleb and almond digests indicating the method was reproducible.

To ensure the thermal cycling digestion protocol resulted in reproducible digestion efficiency, the variability in the protocol was investigated with three replicate digestion experiments of the referred paprika. From the mean and standard deviation of the summed peak areas for marker peptides in the referred paprika sample, the %RSD was calculated to be between 9 and 18%, (data not shown). This degree of uncertainty is to be expected for a qualitative method in development and in the absence of an internal standard.

3.4. Paprika contaminant

The referred paprika sample was subjected to the thermal cycling digestion protocol in triplicate before sample clean-up with the optimised SPE method and analysis by SRM.

A number of criteria were set for the specificity of the peptides. Firstly, peptide retention times within 30 s of those from the neat almond or mahaleb samples. Secondly, all three transitions for each peptide present. Thirdly, after integration, peak area > 1000 counts, with manual investigation if any of the three transitions < 1000 counts.

As a result of investigating the peak areas in the referred paprika, Fig. 3, the lack of mahaleb peptides identified and abundance of almond peptides present indicated the presence of almond. One of the *Prunus* peptides (FVS...) was removed from the analysis as the peak areas did not exceed the 1000 counts threshold. The peptides observed were confirmed as undetected in the blank paprika (data not shown).

As a final specificity filter the mean of the ratio between each qualifier ion and the quantifier ion for each peptide from the referred paprika was required to be within 20% of those observed in the replicate analyses of neat almond.

As a result of this ratio comparison, Fig. 4, one of the peptides characteristic of almond (IST...) was removed from the analysis leaving two distinctive almond peptides and three *Prunus* peptides observed in the referred paprika.

3.5. Cumin contaminant

The cumin samples were subjected to the improved sample preparation process. The low level of *Prunus* contamination meant a 40 µL injection volume was required.

The above selectivity filter approach was applied. The retention time criterion was satisfied for all peaks. At a peak area threshold of 500 counts (Fig. 5) all three transitions for one of the mahaleb peptides were detected, as well as three transitions for two of the '*Prunus*' peptides however one of the *Prunus* peptides (GNL...) failed the ion ratio criterion in one transition and was removed. Thus the presence of mahaleb, rather than almond, in the referred cumin sample was confirmed. The weaker signal for the *Prunus* peptides is consistent with the lower level of contamination in the referred cumin as compared to the referred paprika.

4. Conclusions

The work presented describes a method for the identification of novel marker peptides in food matrices and the development of a robust sample preparation method for screening for marker peptides at trace levels. The methods were applied in the identification of part per million contaminants of spices with nuts from different, but closely related, *Prunus* species. Although the protocol described is qualitative, the inclusion of isotopically labelled standards could facilitate quantification of the marker peptides.

Acknowledgements

The authors wish to thank Kirstin Grey and Malvinder Singh for sourcing suitable control samples and Michael Walker for constructive discussions on multidisciplinary approaches for the tackling of the problem.

The authors gratefully acknowledge funding through the UK Department for Business, Energy & Industrial Strategy as part of the Government Chemist Programme and funding assistance from the UK Food Standards Agency.

References

- Burns, M., Walker, M., Wilkes, T., Hall, L., Gray, K., & Nixon, G. (2016). Development of a real-time PCR approach for the specific detection of *Prunus* mahaleb. *Food and Nutrition Sciences*, 7, 703–710.
- Canadian Food Inspection Agency. (2015). Canadian Food Inspection Agency Food Recall Warnings. <http://www.inspection.gc.ca/about-the-cfia/newsroom/food-recall-warnings/complete-listing/2015-03-20b/eng/1426911229505/1426911232888>.
- European Parliament and the Council of the European Union (2011). Regulation (EU) No. 1169/2011 of the European Parliament & of the Council of 25 October 2011 on the Provision of Food Information to Consumers. *Official Journal of the European Union*, 304, 18–63.
- Food Standards Agency. (2015). Bart Ground Cumin Recalled. <https://www.food.gov.uk/news-updates/news/2015/13512/bart-ground-cumin-recalled>.
- Government Chemist. (2015). Paprika Referee Sample: Further Testing Identifies Almond Present. <https://www.gov.uk/government/news/paprika-referee-sample-further-testing-identifies-almond-present>.
- Groves, K., Cryar, A., Walker, M., & Quaglia, M. (in press). Assessment of recovery of milk protein allergens from processed food for mass spectrometry quantification. *Journal of AOAC International*.
- Heick, J., Fischer, M., & Pöpping, B. (2011). First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry. *Journal of Chromatography A*, 1218(7), 938–943. <http://dx.doi.org/10.1016/j.chroma.2010.12.067>.
- Johnson, P. E., Baumgartner, S., Aldick, T., Bessant, C., Giosafatto, V., Heick, J., ... Mills, E. N. C. (2011). Current perspectives and recommendations for the development of mass spectrometry methods for the determination of allergens in foods. *Journal of AOAC International*, 94(4), 1026–1033.
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., & MacCoss, M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26(7), 966–968. <http://dx.doi.org/10.1093/bioinformatics/btq054>.
- Muraro, A., Werfel, T., Hoffmann-Sommergruber, K., Roberts, G., Beyer, K., Bindslev-Jensen, C., ... Akdis, C. A. (2014). EAACI Food Allergy and Anaphylaxis Guidelines: Diagnosis and management of food allergy. *Allergy: European Journal of Allergy and*

- Clinical Immunology*, 69(8), 1008–1025. <http://dx.doi.org/10.1111/all.12429>.
- Nixon, G., Hall, L., Wilkes, T., Walker, M., & Burns, M. (2016). Novel Approach to the Rapid Differentiation of Common Prunus Allergen Species by PCR Product Melt Analysis. *Food and Nutrition Sciences*, 7, 920–926.
- Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(August), 3551–3567. [http://dx.doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20](http://dx.doi.org/10.1002/(SICI)1522-2683(19991201)20).
- Roux, K. H., Teuber, S. S., Robotham, J. M., & Sathe, S. K. (2001). Detection and stability of the major almond allergen in foods. *Journal of Agricultural and Food Chemistry*, 49(5), 2131–2136. <http://dx.doi.org/10.1021/jf001307k>.
- Sealey-Voyksner, J., Zweigenbaum, J., & Voyksner, R. (2016). Discovery of highly conserved unique peanut and tree nut peptides by LC-MS/MS for multi-allergen detection. *Food Chemistry*, 194, 201–211. <http://dx.doi.org/10.1016/j.foodchem.2015.07.043>.
- Walker, M. J., Burns, D. T., Elliott, C. T., Gowland, M. H., & Mills, E. N. C. (2016). Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs. *The Analyst*, 141(1), 24–35. <http://dx.doi.org/10.1039/C5AN01457C>.