Mass spectrometry - identification of protein allergen contaminants

Chris Hopley (christopher.hopley@lgcgroup.com)
28 February 2018

LGC

Science
for a safer world

Cumin & Paprika recalls – Referee Cases

LGC

1. • 31 Jan '15 FSA recall cumin - almond not listed on the label
2. • 10 Feb '15 FSA refer cumin sample to Government Chemist
3. • Elisa Testing
4. • March '15 supplier "…mahleb gives positive ELISA for almond"
5. • Protein by Mass Spectrometry
6. • 30 April '15 Canada rescinds recalls " … mahleb false positives"
7. • DNA Testing
8. • 26 June '15 Govt Chemist confirms cumin contains mahleb
9. • 29 June '15 FSA rescinds cumin recalls "mahleb present; not almond"
10. • 13 Aug '15 FSA refer sample of paprika to Government Chemist
11. • 9 Nov '15 Govt Chemist confirms paprika contains almond

Mass Spectrometry

- Gold standard for (Bio)analytical chemistry
- Able to both characterise and identify contaminants in matrix, be they small or large molecule
- Particularly suited for trace level work
- Our lab has 12 Mass Spectrometers and associated separation science available
- Also, in particular relevance to protein/peptide work, has a number of databases and search programmes available
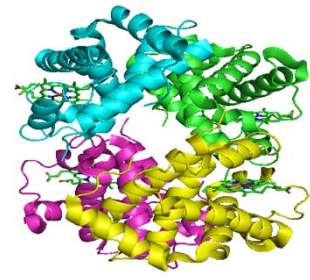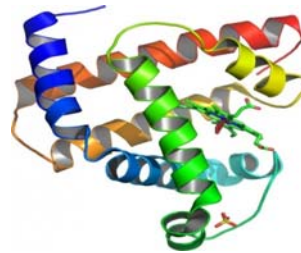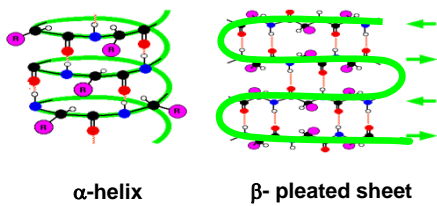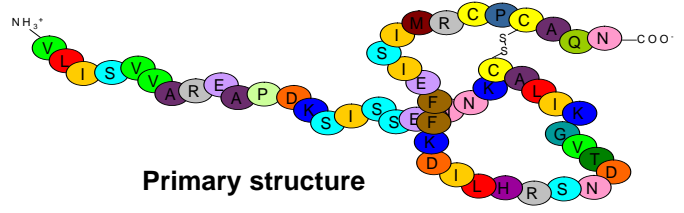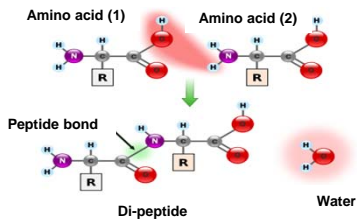
Food Allergen Identification

- Food Allergens are usually proteins
- Mass spectrometry is a highly specific and accurate platform for the identification of both known and unknown proteins
- Mass spectrometry can be used to both identify and quantify food allergens

Protein

Amino acid (1)   Amino acid (2)

Peptide bond

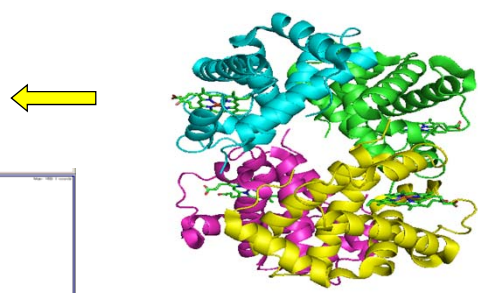Di-peptide   Water

Primary structure

α-helix   β- pleated sheet

Secondary structure

Tertiary structure

Quaternary structure

Identification of proteins by MS: top down

Lysozyme

Phosphorylase B
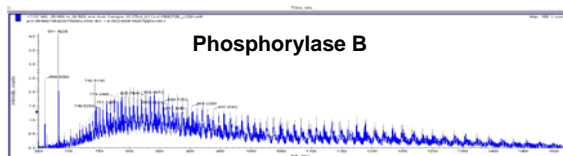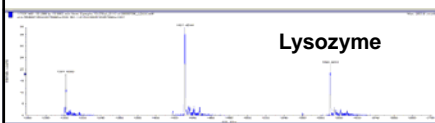
Limited sample prep

Low sensitivity

Complex interpretation

No real information on identity

Qualification of the highly ordered structure of the protein

Identification of proteins by MS: bottom up

Bottom up

Enzymatic digestion (trypsin)

peptides

FESNFNTQNTNR

F|E|S|N|F|N|T|Q|N|T|N|R
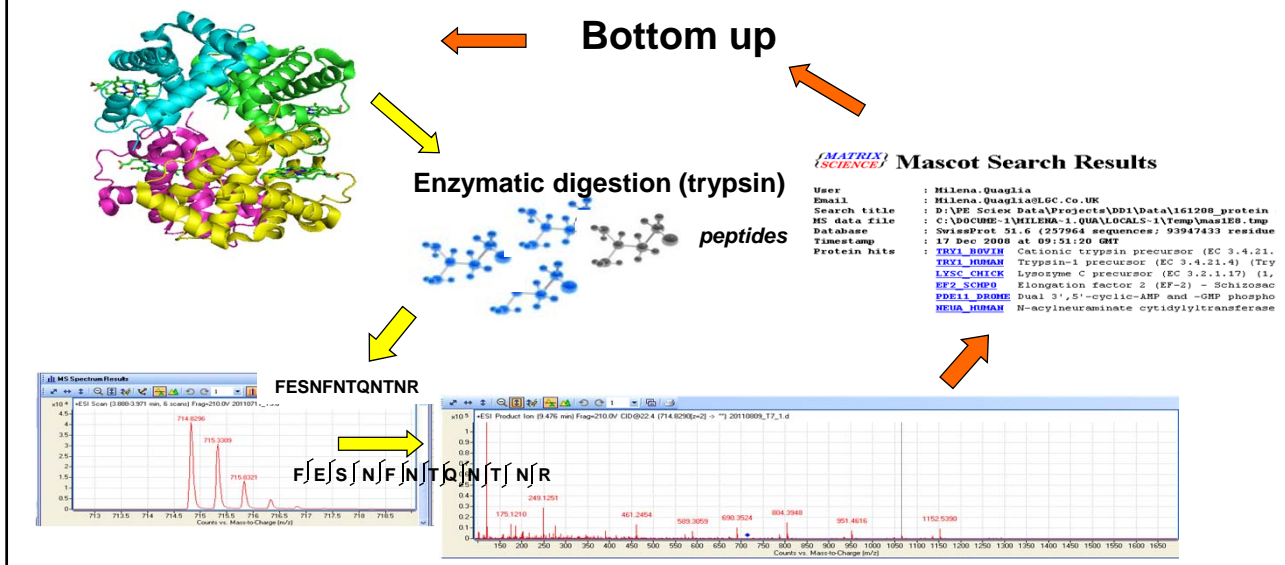
Databases and software

- Many databases are available
  - SwissProt, UniProt, NCBI
- And software packages
  - Peaks, SpectrumMill, Mascot, PLGS

- However, the majority of entries are either Human or Mammalian
- This causes issues for the detection of food allergen
  - database coverage is poor
  - sequence similarity can be quite high
  - Numerous entries are not experimentally derived, but derived from gene sequences

Cumin and Paprika

Overall MS Approach

Bottom-up Proteomics
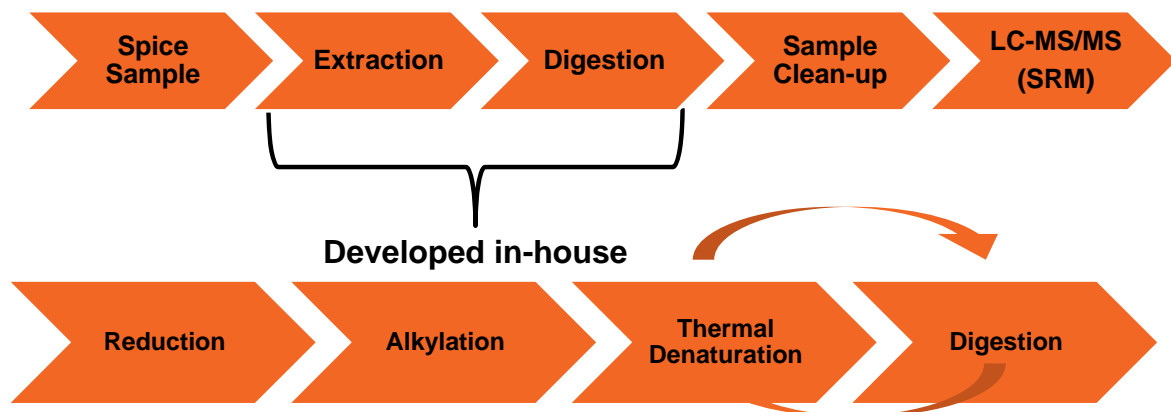

Literature

- Targeted to allergenic proteins, digested to indicator peptides

Peptide 1	GNLDFVQPPR
Peptide 2	GVLGAVFSGCPETFEESQQSSQQGR
Peptide 3	ALPDEVLANAYQISR
Peptide 4	NGLHLPSYSNAPQLIYIVQGR

In-house Approach

| Spice Sample | Extraction | Digestion | Sample Clean-up | LC-MS/MS (SRM) |

Developed in-house

| Reduction | Alkylation | Thermal Denaturation | Digestion |

Literature approach

- Several peptides previously identified as indicative of presence of Almond
- Method build for the detection of peptides by targeted LC-MSMS
- Tested on other *prunus* species
- Specificity of literature peptides are called into question

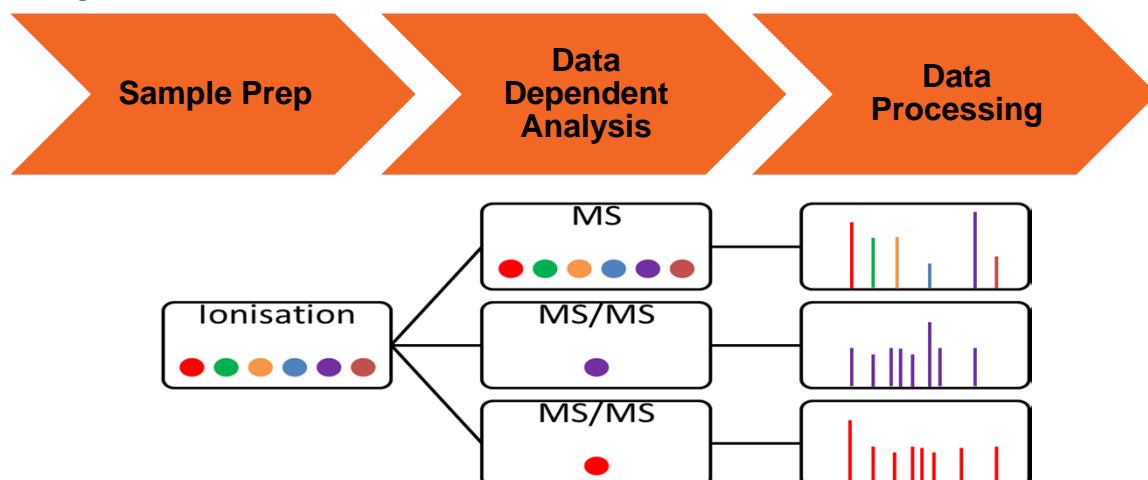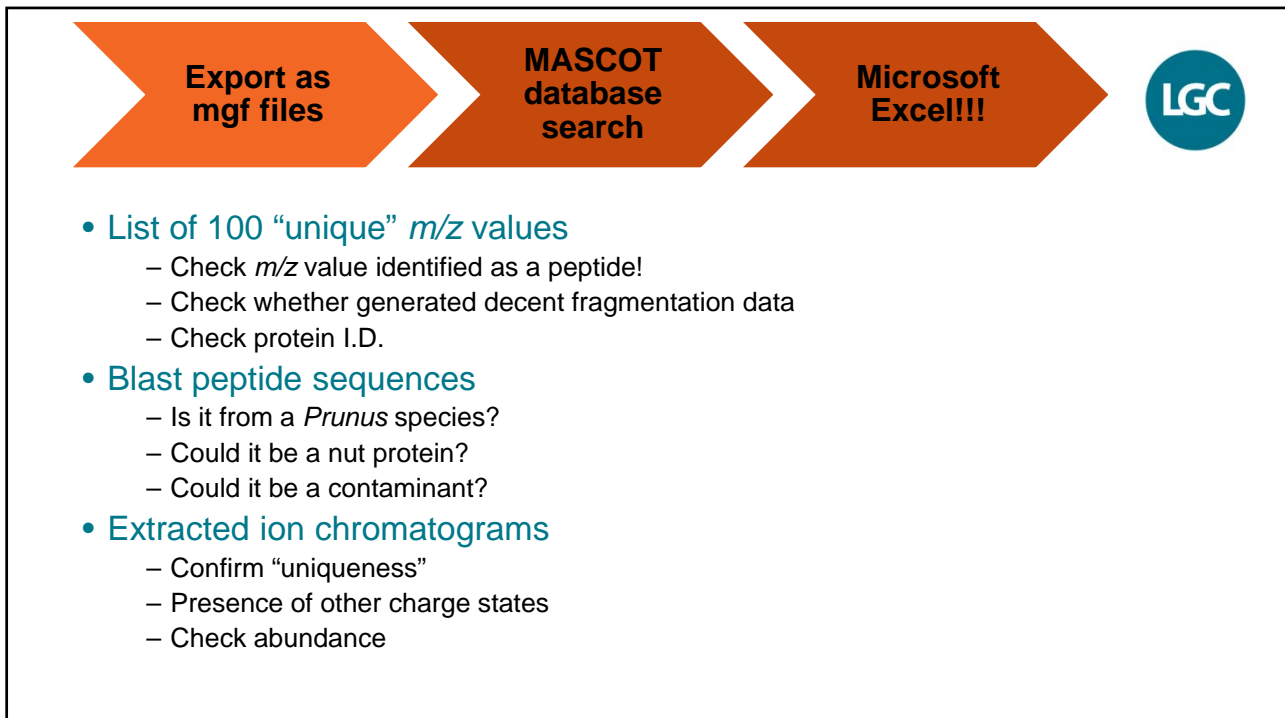| Precursor ion m/z	Charge state	Retention time min	Peptide sequence (Assigned using Mascot)	Almond Kernel	Apricot Kernel	Mahleb	Peach Kernel
571.8	2+	7.7	GNLDFVQPPR	Y	Y	Y	Y
876.74	3+	10.0	GVLGAVFSGCPETFEESQQSSQQGR	Y	Y	N	Y
830.44	2+	9.9	ALPDEVLANAYQISR	Y	Y	Y	Y
1170.63	2+	10.7	NGLHLPSYSNAPQLIYIVQGR	Y	Y	N	Y

12

New peptides are required

- The Literature peptides for detection of Almond allergen were not found to be specific enough
- The databases were found to be lacking in information regarding proteins from *Prunus*
- A new approach was required
    - Authentic samples of relevant prunus species were obtained
    - Samples were ground, then extracted/digested
    - Samples analysed by LC-QTOF-MS to determine distinctive peptides

# Identifying New Distinctive Peptides

- Agilent 6530 QToF

Export as mgf files ▸ **MASCOT database search** ▸ **Microsoft Excel!!!** LGC

- Round all precursor ion *m/z* values to 2 decimal places

$$-\frac{0.01}{500\ m/z}\ x\ 10^6 = 20\ ppm\ error$$

- Compare lists using the "MATCH" function

Precursor Ions ▸ **Remove *m/z* duplicates** ▸ **Match within replicates** ▸ **Unique *m/z* values**

Export as mgf files ▸ **MASCOT database search** ▸ **Microsoft Excel!!!** LGC

- List of 100 "unique" *m/z* values
  - Check *m/z* value identified as a peptide!
  - Check whether generated decent fragmentation data
  - Check protein I.D.
- Blast peptide sequences
  - Is it from a *Prunus* species?
  - Could it be a nut protein?
  - Could it be a contaminant?
- Extracted ion chromatograms
  - Confirm "uniqueness"
  - Presence of other charge states
  - Check abundance

Final Peptides

Peptide	Precursor Ion /m/z		Almond	Mahaleb
FVSPAYR	2+	420.2241		
SGGQILPIR	2+	470.7824		
DFVSSPFR	2+	477.7376		
VPTPVPPRVSSPR	2+	694.9041		
ALPDEVLANAYQISR	2+	830.4387		
VQGQLDFVSPFRS	2+	740.3832		
TEENAFINTLAGR	2+	718.3624		
ISTLNSHNLPILR	3+	493.2877		
GNLDFVQPPR	2+	571.8013		
GVLGAVFSGCPETFEESQQSSQQGR	3+	895.7452		

Identifying the Cumin Contamination

- Once the peptides have been identified, method moved to a LC-QQQ-MS and further refined
- Most sensitive instrument for trace analysis
- Use fixed transitions (minimum of 2) and retention time to unequivocally identify peptides
- Samples run in comparison to blanks, spikes and standards to identify the contamination
- Preliminary results suggest Cumin contamination identified to be Mahaleb
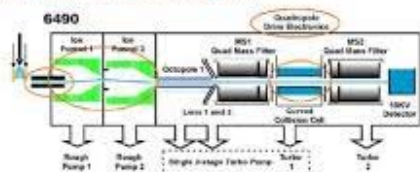
What about paprika?

- Paprika screened by ELISA found to contain almond
- Sample run using method for Cumin
  - Preliminary analysis indicates it is almond
  - 400 ppm, much higher than Cumin ~ 1ppm

- Incurred material
  - Confirm identify of the contaminant
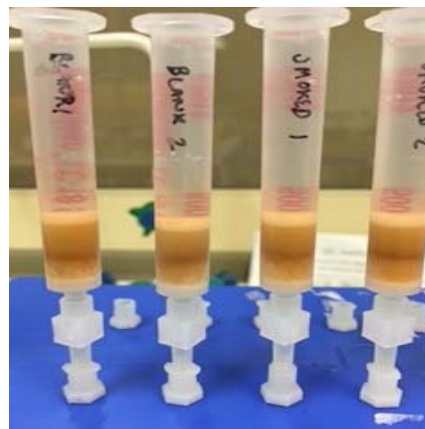 - Optimise the sample preparation methods

Considering Reverse Phase SPE
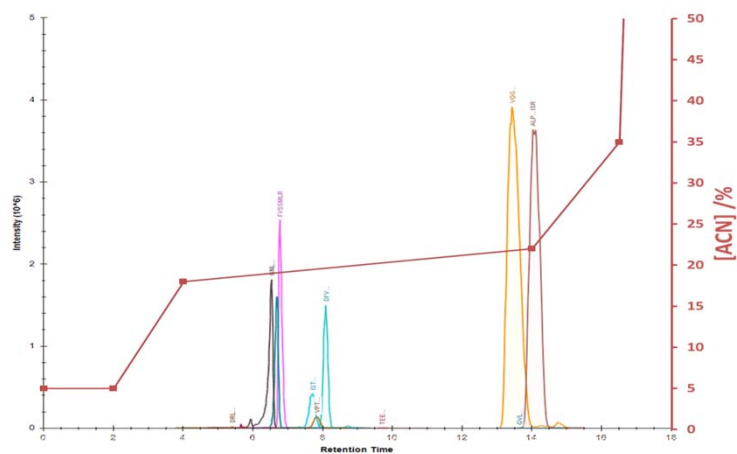
Optimising the SPE

- Simple reverse phase cartridges
  - Supelco LC-18
  - 500 mg bed
  - 3 mL cartridge
- Variables
  - Wash buffer
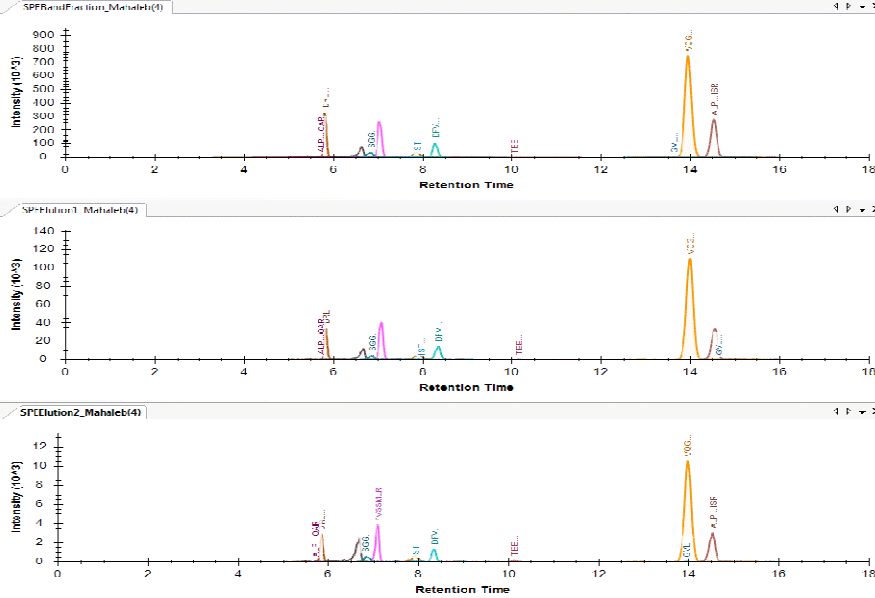 - Elution buffer
  - Volumes

Optimising the SPE?

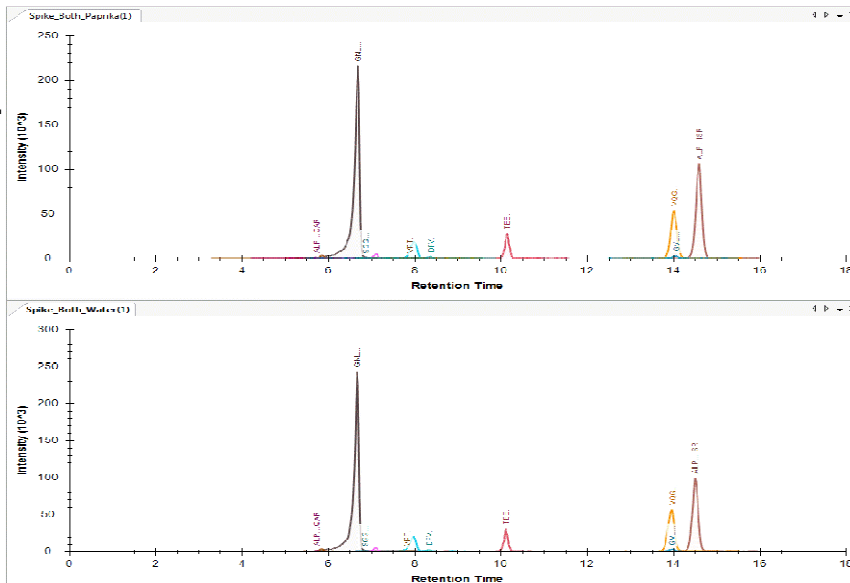| SPE Method	Conditioning Buffer	Wash Buffer	Elution Buffer	Resuspension Buffer
1	0.1 % FA	5 % ACN	2x1ml 40% ACN	? CH
2	0.1 % FA	15% ACN, 0.1% FA	2x3ml 35% ACN 0.1% FA	0.1 % FA
3	0.1 % FA	15% ACN, 0.1% FA 20% ACN, 0.1% FA 30% ACN, 0.1% FA	40% ACN, 0.1% FA	0.1% FA
4	100 mM Tris pH 8	15% ACN, 0.1% FA	45% ACN, 0.1% FA	0.1% FA
5	100 mM Tris pH 8	10% ACN, 0.1% FA	45% ACN, 0.1% FA	5% ACN, 0.1% FA
6	100 mM Tris pH 8	10% ACN, 0.1% FA 1.2 ml 30% ACN, 0.1% FA	45% ACN, 0.1% FA	5% ACN, 0.1% FA

SPE Method 6

SPE Method 6: Matrix Effects

Final Results – Cumin/Paprika

Peak Areas for the Peptide Transitions

25

Conclusions

- Analytical method to determine low level allergen contaminants in spices developed
- Method able to differentiate between different *Prunus* species
- Able to detect contaminants at low level ~ 1ppm
- High degree of confidence due to multiple peptides used to distinguish the samples
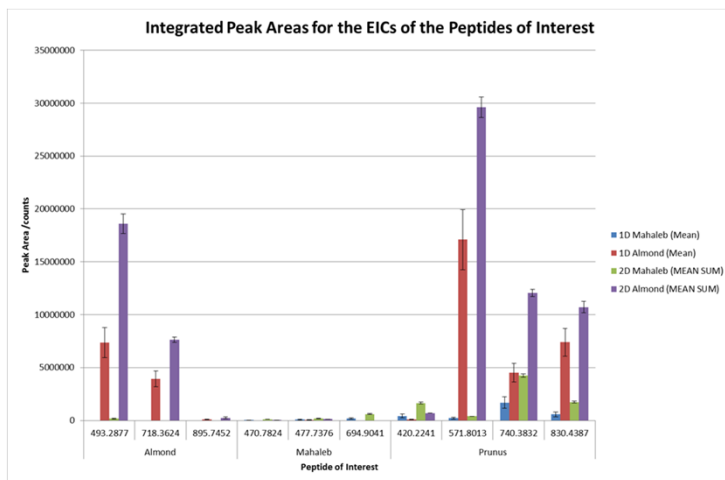
New developments

- New technologies in LC-MS constantly being developed
- We have recently been loaned a 2D-LC system
- Applied the system to the spices contamination issue
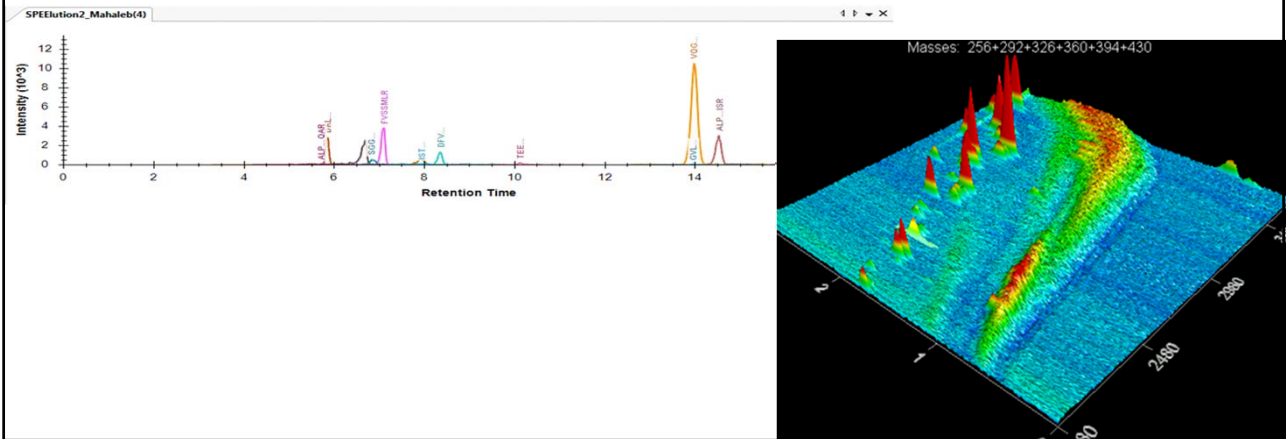
2D LC applied to Almond/Mahaleb

2D LC results In larger peak areas, therefore better sensitivity

Acknowledgements

- The work described in this presentation was funded by the Department of Business, Energy & Industrial Strategy, and the Food Standards Agency

- Sophie Thurlow, Bryan McCullough, Kate Groves, Kirstin Gray, Malvinder Singh and Milena Quaglia