Streamlining the Use of High Resolution Mass Spectrometry Data to Fingerprint Adulterated Honey using Multivariate Data Analysis to Facilitate Food Product Quality Control

Christopher Buck, Waters Australia, 3/38-46 South St, Rydalmere, NSW 2116. Email: christopher_buck@waters.com

The adulteration of food products with materials of lesser value or even potentially unsafe origins has been observed worldwide in foods diverse as olive oil, basmati rice, meats and honey [1]. At first glance approaching suspect products with high resolution mass spectrometry can seem intimidating due to the high complexity of the data, as it is not unusual to find evidence of several thousand components. Development for many years has focused on addressing these large mass spectrometric data sets with statistical analysis to rapidly produce results easy to interpret [2,3]. With the Progenesis QI software package Waters has sought to provide an easy to use, highly visual guided workflow for the statistical analysis of mass spectrometry data. In Progenesis QI the steps to the workflow are in a logical order and include everything from importing data, detection of peaks, normalisation of samples, multivariate statistical analysis to determine statistically relevant components of interest that differentiate sample groups, and tools for identifying components of interest. The Progenesis software has been successfully utilised for the statistical analysis of metabolites [4,5], as metabolomics research has pioneered the comparison of groups for finding key differences. In this study the goal was to demonstrate the utility of using high resolution mass spectrometer data derived from honey and two possible adulterants to show that mass fingerprints inclusive of hundreds of compounds, can quickly differentiate these foods and combinations of the foods.

Data Acquisition Methods

To facilitate the comparison of several samples methods must be selected carefully and maintained while collecting data from all samples. The data set for statistical analysis is composed of Exact Mass Retention Time (EMRT) information for all components detected in the samples. Considerations are appropriate chromatography and mass spectrometer methods. Data was acquired using a Waters Acquity UPLC (classic model) and a Waters Xevo G2 QTof mass spectrometer (Melbourne, Australia).

Methods were carried out using both positive and negative ion modes, with LC columns, gradients and buffers selected for appropriate screening with the specified ion modes. For positive ion mode data acquisition reverse phase chromatography was selected using a Waters UPLC BEH C18 1.7 µm particle, 2.1 mm x 100 mm column. Mobile phase A was water containing 0.1% (v/v) formic acid, mobile phase B was acetonitrile containing 0.1% (v/v) formic acid. Mobile phase B was held at 2% (v/v) for 15 seconds followed by a gradient to 99% (v/v) B over 12 minutes. The flow rate was maintained at 0.450 mL/min and column temperature was set at 45°C. For negative ion mode data acquisition HILIC chromatography was performed using a Waters BEH Amide 1.7 µm particle, 2.1 mm x 100 mm column, as this chromatography and ion mode are complimentary for simple sugars and some polar metabolites (5). Mobile phase A was 100% acetonitrile, mobile phase B was water containing 10 mM ammonium formate pH 8. Mobile phase B was held at 2% for 15 seconds followed by a gradient to 90% B over 12 minutes. The flow rate was maintained at 0.450 mL/min and column temperature was set at 60°C. Total acquisition times were 15 minutes. The mass spectrometry methods for both ion modes were MSE - a simple, patented method of unbiased data acquisition that comprehensively catalogues complex samples in a single analysis [6,7]. A mass range of 50 to 1200 m/z was scanned, with alternating low and elevated collision energy scans of 0.3 seconds. For elevated energy scans (the high energy fragment channel) the collision energy was ramped from 10 to 45 eV. Leucine-enkephalin was used as a lockspray [6], acquired for 0.5 seconds at 20 seconds intervals. The mass spectra of the lockspray are recorded in a separate channel that is used to recalibrate acquired data improving mass accuracy to within 5 ppm.

Samples used were honey (a major domestic Australian brand), a supermarket brand golden syrup (major ingredient was cane sugar syrup, referred to as 'golden'), and a brand name maple flavoured syrup (major ingredient was wheat glucose syrup, referred to as 'wheat'). As corn syrup based products were unavailable they were not included, but would make sense to include in some markets where it is a likely adulterant. Additional samples analysed were 1:1 mixtures of honey and the same golden syrup, and also of honey and the wheat glucose syrup. A QC mixture was prepared that was equal parts of each product. Each product was diluted 20x with water, mixed until homogeneous and passed through 0.2 micron microcentrifuge filters. A negative control of extraction solvent was the same dilution water passed through an identical filter. An injection volume of 5 µL was used for each sample.

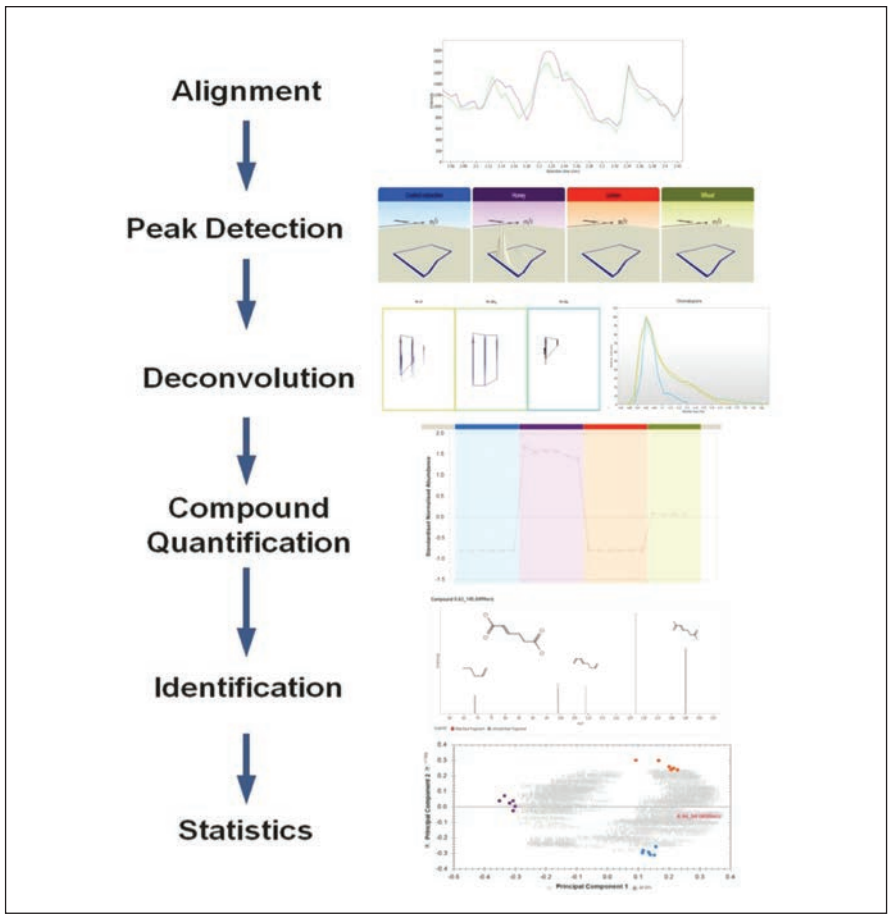

Figure 1. General workflow for Progenesis software. Chromatograms at the top are imported, and components are detected, compared, and analysed statistically. An example of peak picking across sample groups is pictured from the 3D peak view. Deconvolution identifies peaks that are the same compound but forming several different adducts, such as the protonated, ammonia and sodium adducts shown here. Deconvolution is illustrated here with 2D images of the isotopic patterns and the overlaid chromatograms for the different adducts observed. This can assist accurate quantification and avoid misidentifications. Compound quantification, identification and statistics are detailed further in the following figures.

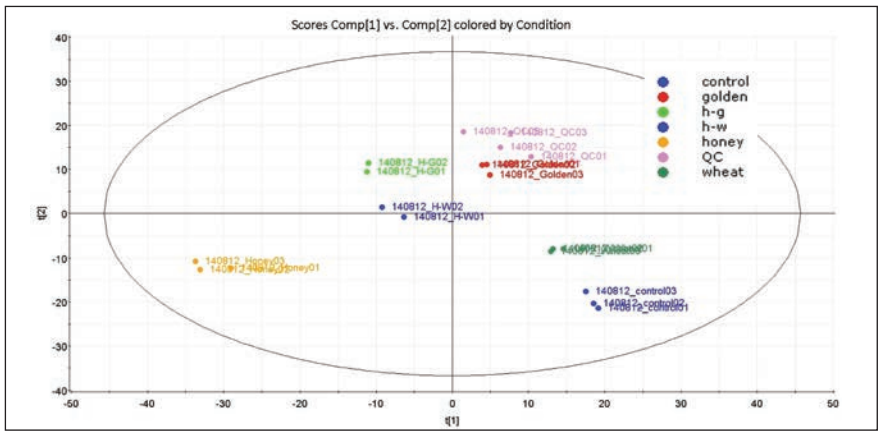

Figure 2. Principal Component Analysis (PCA) of samples analysed. Repeat injections of each sample clustered together and away from the different products, mixtures clustered together midway between their constituent sources. Control = extraction solvent negative control, golden = golden syrup, wheat = artificial maple syrup, h-g = honey/golden syrup mixture, h-w = honey/ artificial maple syrup mixture, QC = mixture of all three products, and honey is the domestic honey.

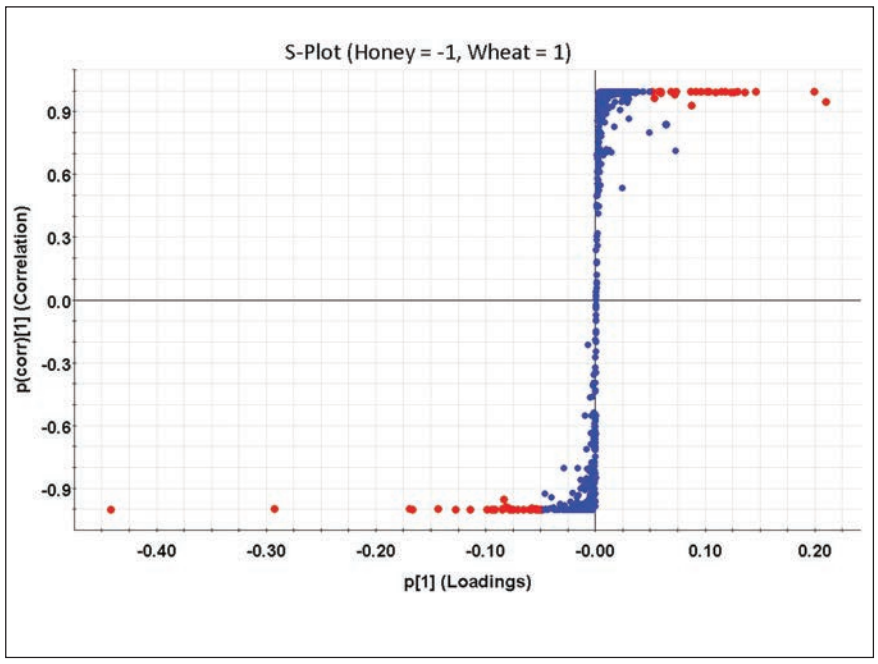

Figure 3. S-plot produced in the EZinfo software package. The S-plot is a loadings plot using Pareto scaling to emphasise components which are most responsible for variation between the sample groups being compared. Components with a high correlation migrates away from the origin on the y-axis (are consistent within a group), and those with high intensity difference between the groups migrate away from the origin on the x-axis. Individual components highlighted are of the greatest interest as individual markers, and some are further explored in the abundance profiles of Figure 5.

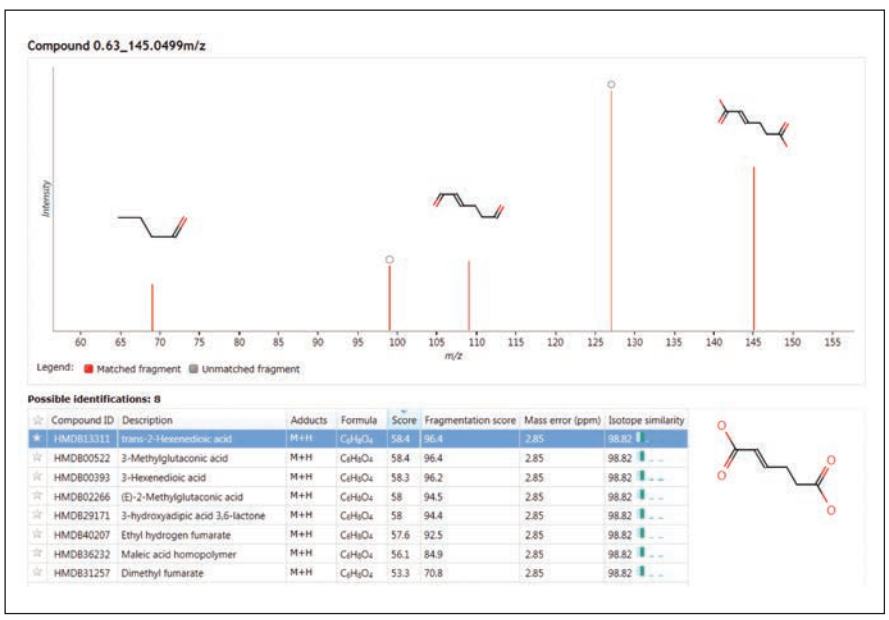

Figure 4. A database in the Structure Data Format (SDF) was searched. As an SDF database contains the structural information of the compounds it contains, in the MOL structure format, it was possible to automatically search the high energy fragmentation data for theoretical molecular fragments which support the exact mass identification of the parent ion. As can be seen in the results of this search several isomers are potential hits for the observed ion at m/z 145.0499, and the supporting fragment information is displayed for trans-2-Hexenedioic acid.

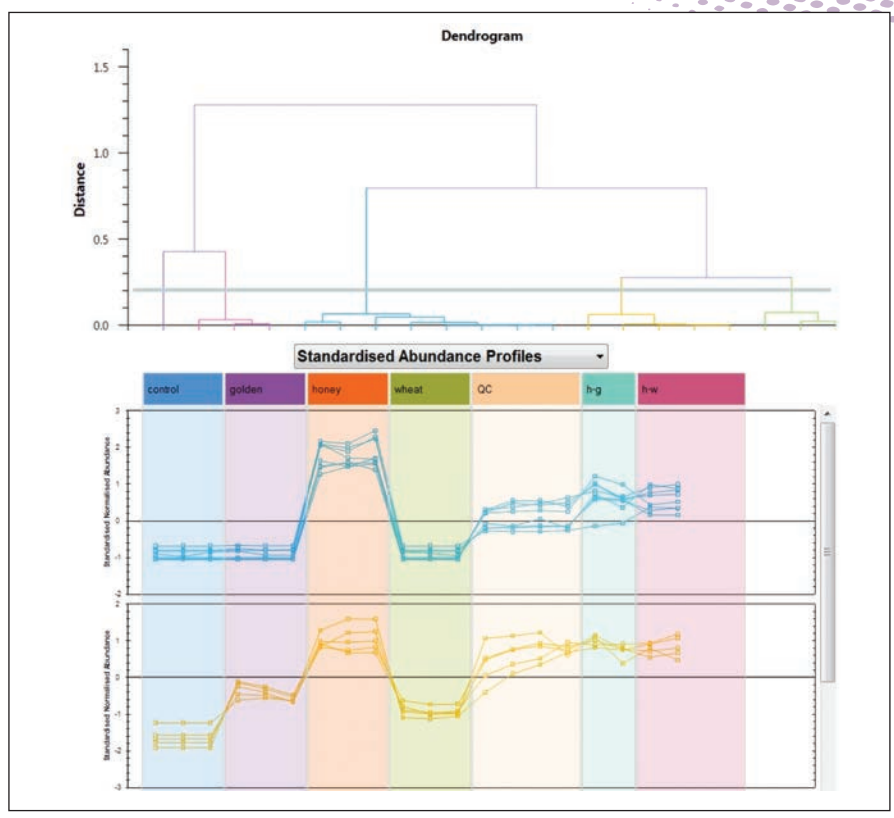

Figure 5. Dendrogram and Standardised Abundance Profiles. Progenesis QI generates PCA and loadings plots, and also dendrograms and abundance profiles such as for these selected components. Components with the same abundance profiles cluster together on the dendrogram, and the blue and yellow clusters from the dendrogram are displayed at the bottom showing their relative abundance in the different sample groups. In this case twenty components of particular interest have been filtered from a data set of over a thousand components detected.

Table 1. Statistically relevant honey components. Components from the sample set which were observed to particularly favour the honey samples, and therefore correlate with the presence of honey, are listed here. While further extensive work may be required to positively identify these compounds and verify that they are consistently found in honey, they illustrate the challenge someone counterfeiting a product would face in fooling a multivariate statistical analysis of high resolution mass spectrometer data. The analysis of variance (ANOVA) is used to compare differences of means between groups. ANOVA compares the extent of variation between groups with the amount of variation within groups. A low ANOVA p score could be considered good, as it indicates the observed component is producing a very consistent peak area within the replicate samples of each group, and the difference in composition between the groups is unlikely to arise from chance variability. The max fold changes indicate how much larger the mean peak response was found in one group when compared to the other. If no evidence of a component was found in one group, the fold change is infinity (more accurately the component is below our limit of detection in one group). Identifications were produced by comparisons of the observed masses of components observed to those of the molecules in the searched database. As can be seen in Figure 4 additional information in the form of fragments was found for some compounds, but identifications should be treated as very tentative until further verified using a standard injected using the same chromatography method. Ideally we would like to match a standards exact mass, retention time, and observe the same fragments when applying collision energy.

Compound	Ion Mode	m/z	Retention time (min)	Anova (p)	Max Fold Change	Identifi-cations
1.64_310.1288m/z	Positive	310.1288	1.64	<1.1E-16	Infinity	4
0.63_145.0499m/z	Positive	145.0499	0.63	<1.1E-16	Infinity	8
1.59_166.0867m/z	Positive	166.0867	1.59	<1.1E-16	Infinity	15
0.71_117.0550m/z	Positive	117.055	0.71	<1.1E-16	Infinity	7
0.63_127.0393m/z	Positive	127.0393	0.63	<1.1E-16	Infinity	6
2.70_536.1746n	Positive	559.1645	2.70	<1.1E-16	Infinity	0
0.66_99.0447m/z	Positive	99.04467	0.66	<1.1E-16	Infinity	0
0.60_116.0711m/z	Positive	116.0711	0.60	<1.1E-16	Infinity	5
0.93_185.0427m/z	Positive	185.0427	0.93	6.27E-14	Infinity	8
0.59_260.1090n	Positive	278.1236	0.59	1.64E-13	Infinity	0
0.55_378.1762m/z	Positive	378.1762	0.55	7.22E-13	Infinity	5
0.59_144.1025m/z	Positive	144.1025	0.59	4.17E-12	Infinity	13
6.05_274.2745m/z	Positive	274.2745	6.05	1.76E-10	11	10
5.28_357.5718m/z	Negative	357.5718	5.28	<1.1E-16	Infinity	0
3.98_195.4374m/z	Negative	195.4374	3.98	<1.1E-16	Infinity	0
5.54_549.6863m/z	Negative	549.6863	5.54	<1.1E-16	Infinity	0
5.76_503.6714m/z	Negative	503.6714	5.76	<1.1E-16	Infinity	0
5.65_1007.8788m/z	Negative	1007.879	5.65	<1.1E-16	Infinity	0
3.97_195.5112m/z	Negative	195.5112	3.97	<1.1E-16	Infinity	0
2.51_133.3495m/z	Negative	133.3495	2.51	<1.1E-16	Infinity	0
5.02_179.4306m/z	Negative	179.4306	5.02	1.33E-15	40456	0
3.38_179.4307m/z	Negative	179.4307	3.38	8.89E-14	148	0
3.60_215.4323m/z	Negative	215.4323	3.60	8.45E-13	40	0
3.54_377.5605m/z	Negative	377.5605	3.54	2.69E-11	13093	0
4.26_161.4059m/z	Negative	161.4059	4.26	5.97E-09	15	0
3.56_217.4311m/z	Negative	217.4311	3.56	7.91E-08	91	0
5.24_683.7634m/z	Negative	683.7634	5.24	3.73E-07	1117	0

Processing of Data with Progenesis QI

The sample data is imported directly from within the Progenesis QI software. Progenesis QI is compatible with a wide variety of instruments, not just those supplied by Waters. A summary of the workflow can be seen in *Figure 1*. For the analysis of proteomics sample sets there is a separate software package which differs in some ways (quantification and identification), but has a parallel workflow. Once data is imported, Progenesis QI uses a co-detection workflow that begins with chromatographic peak matching, followed by peak picking and normalisation.

This process eliminates missing values and enables more efficient application of uni-variate and multi-variate statistics. After peak detection the results of Principal Component Analysis (PCA) can be seen in *Figure 2*. The PCA is unsupervised, meaning it does not know which samples are in the same group. Samples which are most similar will cluster together, and will separate from samples which are very different. Waters can provide the EZinfo software package (Umetrics, Sweden) for additional analysis such as the S-plot pictured in *Figure 3*. The S-plot is one of several tools that can aid in the graphical visualisation of which components detected in the samples are responsible for sample groups being different from each other. We can select components in the S-plot and then view their chromatographic information and abundance profiles in Progenesis QI. A list of components determined to be more abundant in honey samples is listed in *Table 1*. Samples in negative ion mode did not return hits from a database search, however several tentative identifications were returned from the positive ion mode data. One such example identification can be seen in *Figure 4*. The simplification of a very large complex set of data (from hundreds to possibly thousands of components) to isolating the components of greatest interest is further illustrated in the abundance profiles of *Figure 5*. These abundance profiles show components unique to or highly elevated in concentration in the honey. Attempting to pick such a list of compounds with abundance specific to particular samples by manual processes is tedious and much less likely to be successful.

Acknowledgments

I would like to thank Sebastian Barone and Saman Buddhadasa for access to the instrument system at the Australian National Measurement Institute facility in Melbourne, and John Shockcor of the Imperial College, London, for insights into the application of multivariate statistical analysis.

References

1. Castle S, Carvajal D. Counterfeit Food More Widespread Than Suspected. The New York Times article, June 27, 2013, on page B1 of the New York edition.

2. Silcock P, and Uria D. Characterization and Detection of Olive Oil Adulterations Using Chemometrics. (2008) Waters application note 720002786en.

3. Cleland G, Ladak A, Lai S, and Burgess J. The Use of HRMS and Statistical Analysis in the Investigation of Basmati Rice Authenticity and Potential Food Fraud. (2014) Waters application note 720005218en.

4. Sedic M, Gethings LA, Vissers JP, Shockcor JP, McDonald S, Vasieva O, Lemac M, Langridge JI, Batinic D, and Pavelic SK. Label-free mass spectrometric profiling of urinary proteins and metabolites from paediatric idiopathic nephrotic syndrome. (2014) Biochem Biophys Res Commun 452:21-26

5. Paglia G, Langridge J, and Astarita G. Development of a Metabolomic Assay for the Analysis of Polar Metabolites Using HILIC UPLC/QTof MS. (2013) Waters application note 720004612EN

6. Bateman RH, Carruthers R, Hoyes JB, Jones C, Langridge JI, Millar A, Vissers JP. A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. (2002) J Am Soc Mass Spectrom. 13 (7),792-803.

7. Silva JC, Gorenstein MV, Li GZ, Vissers JPC, and Geromanos SJ. Absolute quantification of proteins by LC MSE a virtue of parallel MS acquisition. (2006) Molecular & Cellular Proteomics 5 (1), 144-156

Conclusions

The samples of honey, its potential adulterants, and honey adulterated with the other foods could be clearly differentiated using the full set of data from the high resolution mass spectrometer. Potential markers either common to the honey or found at higher levels in the adulterants were identified and could be used to develop targeted analysis via other instruments, such as Multiple Reaction Monitoring (MRM) on tandem quad mass spectrometers. Targeted analysis can be fraught with peril for fraud as purity mimicry is possible if the nature of an assay becomes common knowledge, and MRM analysis is blind to uncharacterised components that may be a new adulterant not seen before. Further considerations are that natural products such as honey can change from year to year or season to season, but data acquired over time can in fact build an ever larger library of these natural variations. Progenesis QI can be used to nurture a growing set of data to characterise all the constants and variables typical of a food product (especially with primary producer's cooperation). This library of acceptable samples can then be used to make any adulterated product stand out from the crowd based upon unusual chemical components, without any preconceived idea what the tell-tale components will be. For liquid chromatography high resolution mass spectrometry data there are two programs available, Progenesis QI and Progenesis QI for Proteomics. Progenesis QI was used for this study, and is applicable to all small molecules (sugars, vitamins, pesticides, lipids, etc.) The parallel workflow for proteomics is similar, but protein database searching is a very different process. Both applications are described in detail at www.nonlinear.com.