Comprehensive Food Profiling Combining High Resolution LC/MS and GC/MS Analyses

Application Note

Metabolomics

Authors

Zijuan Lai, Mine Palazoglu, and Oliver Fiehn
University of California at Davis

Sofia Nieto, Yuqin Dai, and Mark Sartain
Agilent Technologies, Inc.

Abstract

A comprehensive untargeted approach was applied to studying differences in food compositions from three distinct diets. To achieve a broad coverage of metabolite classes, this study used high-resolution accurate-mass LC/MS and GC/MS. These full spectrum mass spectrometers allowed for untargeted acquisition of widely varying metabolites. Data acquired by the different analytical techniques were processed using a common software platform, Agilent MassHunter, with multivariate analysis performed in Agilent Mass Profiler Professional (MPP). The results revealed distinct groups of metabolites characteristic of certain diets evaluated in this study.

Agilent Technologies

Introduction

Analysis of food components is essential to evaluate food nutritional quality and safety. This may also serve other purposes such as authentication and differentiation[1]. Although the effect of diet on human health is well recognized, the complete set of small molecules present in distinct diets is largely unknown[2].

The goal of the study was to compare three food plates referred to in this study as fast food (FF), pesco-vegetarian (PV), and eastern vegetarian (EV), that represent different diets with different chemical compositions. The differential analysis approach used in metabolomics is ideal for use in many areas of food science and nutrition research[3]. This comprehensive untargeted approach was applied to study the differences between these three plates, focusing on the detection of as many metabolites as possible[1,4].

To identify food components including small polar molecules, nonpolar molecules, and complex lipids, multiple analytical methods were used, and leveraged Agilent 7200 GC/Q-TOF, Agilent 6230 LC/TOF, and Agilent 6550 LC/Q-TOF systems. Agilent MassHunter software was used to efficiently process both GC/MS and LC/MS data with differential analysis performed using Agilent Mass Profiler Professional (MPP), a chemometrics software tool.

The multiple analytical techniques, in combination with the advanced differential software workflow, was crucial to perform comprehensive analysis and data interpretation.

Experimental

Materials and sample preparation

Food ingredients from each plate (Table 1) were mixed, homogenized, and freeze-dried. Lyophilized material was extracted with 80:20 (v/v) methanol/water, and centrifuged at 27,000 xg for 5 minutes at 4 °C to pellet proteins. For LC/MS samples, the supernatants were filtered through prewashed Pall Omega 3kDa Nanosep centrifugal devices. For GC/MS analysis, the extracts were dried and derivatized with 40 mg/mL methoxyamine hydrochloride in pyridine, followed by silylation with MSTFA + 1 % TMCS. Six technical replicates were used for each technique.

Instrumentation and analytical methods

GC/MS Metabolomics

Both electron ionization (EI) and chemical ionization (CI) spectra were acquired using an Agilent 7890B GC coupled to an Agilent 7200 series high-resolution accurate mass GC/Q-TOF. The data were acquired at 10 and 15 Hz to ensure best performance of the feature detection algorithm used for data processing. Further analytical methods and conditions for GC/Q-TOF were used as described elsewhere[5].

LC/MS Metabolomics

An Agilent 1290 Infinity LC, coupled to either an Agilent 6230 TOF, or an Agilent 6550 iFunnel Q-TOF, was used for data acquisition. The reversed-phase (RP) chromatographic separation was conducted on an Agilent ZORBAX SB-Aq column (2.1 × 50 mm, 1.8 µm, p/n 827700-914). The aqueous normal phase (ANP) chromatographic separation was performed on a Cogent Diamond Hydride HPLC column (2.1 × 150 mm, 4 µm). Data were acquired in both positive and negative ion electrospray ionization (ESI) modes. Further LC/MS conditions used in this study are described elsewhere[6,7].

LC/MS Lipidomics

A 1290 Infinity LC was coupled to a 6550 iFunnel Q-TOF equipped with a dual Agilent Jet Stream source. Data were acquired in both positive and negative ion modes. Reversed-phase chromatographic conditions and MS acquisition parameters were replicated as described elsewhere[8] with the changes shown in Table 2.

Table 1. Content of the food plates used in this study.

Fast food plate	Pesco-vegetarian plate	Eastern vegetarian plate
Hamburger • Bun • Beef • Cheese • Bacon • Lettuce • Tomatoes • Pickles • Ketchup	Salmon	White rice
French fries	Blended brown rice	Fried egg
Baked beans	Sliced almonds	Sesame seeds
Chocolate chip cookies	Lemon slices	Spicy sauce
Coke (regular)	Steamed vegetables • Carrots • Broccoli • Onions • Cabbage • Red bell pepper	Tofu
	Grapes	Spinach
	Yogurt	Bean sprouts
	Blueberries	Soy sprouts
	Green tea	Carrots
		Zucchini
		Radish
		Rice punch

Data analysis software

The Agilent MassHunter data analysis suite (B.07.00) and MPP (B.13.0) were used for differential analysis. Identification of the differential features was accomplished using the Agilent-Fiehn Metabolomics library, Agilent METLIN Personal Compound Database, NIST14 library, and SimLipid 4.40 (PREMIER Biosoft).

Data analysis workflow

Figure 1 summarizes the data analysis workflows.

Feature extraction

The initial data processing involved feature extraction for the subsequent differential analysis. LC/MS metabolomics and lipidomics data were analyzed in MassHunter Profinder in the MassHunter data analysis suite (Figure 2A)[9]. GC/MS data were processed using the Unknowns Analysis tool in the MassHunter data analysis suite (Figure 2B).

Differential analysis and initial compound identification/annotation

After feature extraction was performed, the features from GC/MS and LC/MS data were imported into MPP for differential analysis. ID Browser within MPP was used for either initial compound annotation (LC/MS), or identification (GC/MS). In the case of GC/Q-TOF EI spectra, both the NIST14 and the Agilent-Fiehn Metabolomics libraries were used to identify compounds. For LC/MS metabolomics datasets, compound annotation was performed by querying the Agilent METLIN PCD or other user-created databases. For lipidomics datasets, compound annotation was performed in SimLipid software, and then lipid annotations were imported into MPP.

Table 2. Chromatographic conditions for LC/MS lipidomics data acquisition.

Parameter	Value
LC flow	Held at 0 %B from 0–3 minutes, to 20 %B at 5 minutes, to 27.5 %B at 20 minutes, to 70 %B at 25 minutes, to 90 %B at 33 minutes, held for 1 minute, back to 0 %B at 35 minutes
Stop time	35 minutes
Post time	4 minutes
Drying gas	200 °C at 15 L/min
Sheath gas	300 °C at 11 L/min
Nebulizer gas	45 psi
Fragmentor	380 V
Auto MS/MS acquisition speed	4 spectra/s (MS) 4 spectra/s (MS/MS)
Auto MS/MS collision energy	20 and 35 eV

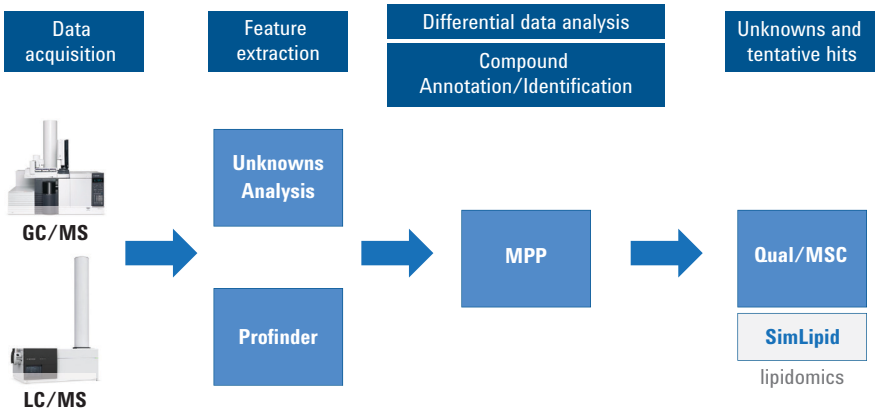

Figure 1. Data analysis workflows.

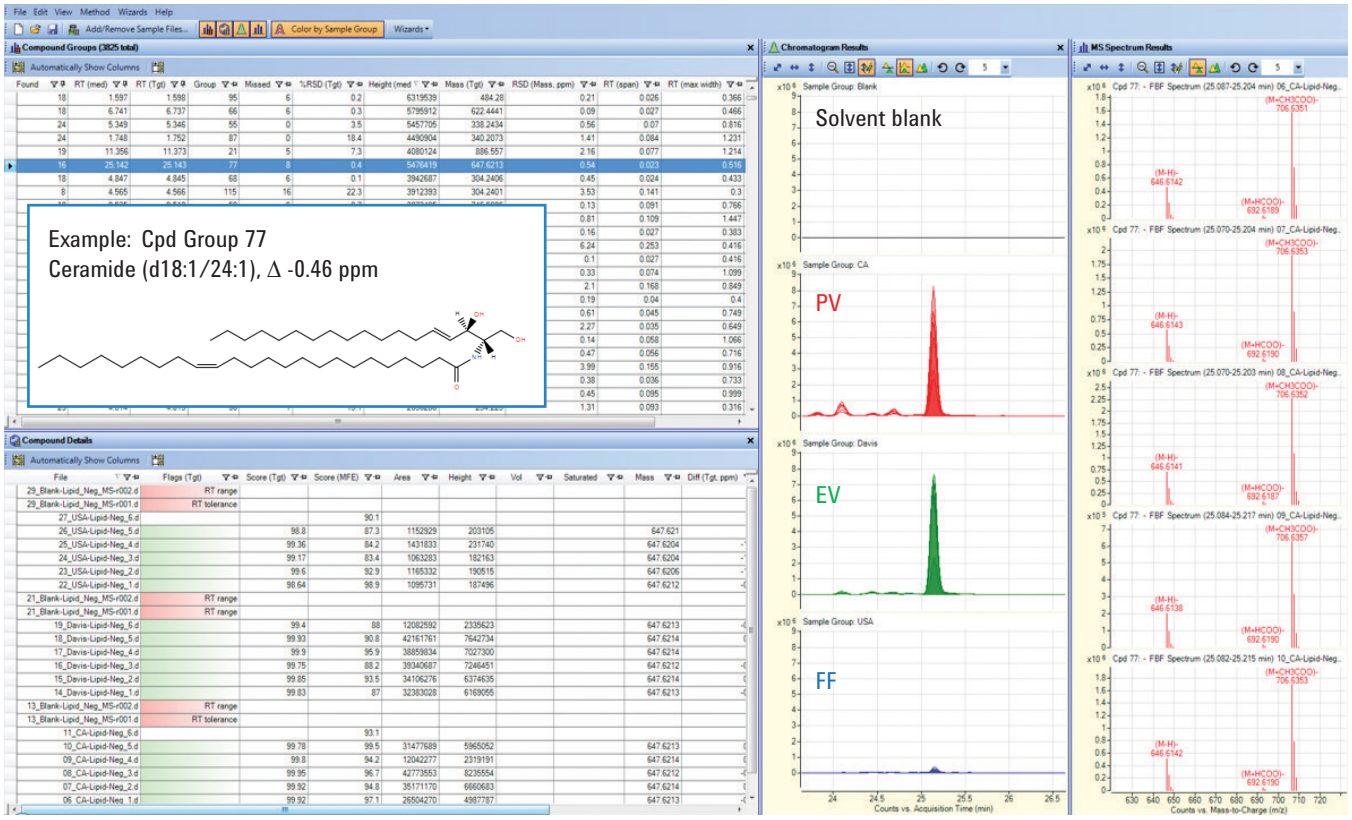

Figure 2A. Feature extraction: LC/MS data files were analyzed with Agilent MassHunter Profinder software using the recursive batch feature extraction algorithm. Shown is a representative extracted feature, later identified as ceramide (d18:1/24:1), and the relative compound abundance across all food plate samples in the lipid dataset.

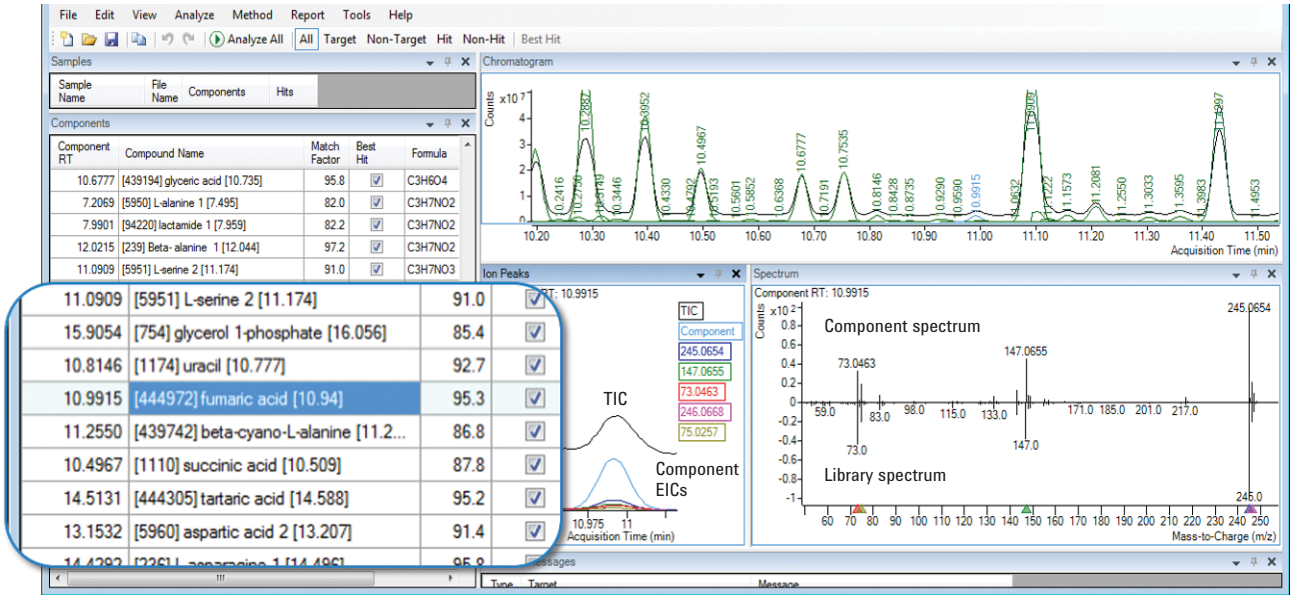

Figure 2B. GC/Q-TOF data files were processed using Agilent Unknowns Analysis for feature finding and library search.

Compound identification

For LC/MS/MS metabolomics data, MS/MS spectral library searching was performed against multiple Agilent PCDLs for compound identification.

When either no match, or a poor MS/MS library match was returned, Molecular Structure Correlator (MSC) was used to propose matches for annotated metabolites from both GC/MS/MS and LC/MS/MS data.

Lipid identifications of LC/MS/MS lipidomics datasets were performed with SimLipid, which match lipid *in silico* MS/MS fragments for structure-specific characteristic ions.

Results and Discussion

Maximizing the coverage of compound classes using multiple techniques, analytical methods, and polarities

For LC/MS, different analytical methods were used to cover the broad range of metabolites. In this study, optimized conditions were developed to provide the best separation and detection. Table 3 summarizes the thousands of features detected for the different LC/MS methods. To illustrate the importance of using multiple LC/MS methods, RP-ESI metabolomics data were compared for ESI positive or negative ion mode. Only 8 % of features were detected in both modes, demonstrating that many features would be missed if only a single polarity were used.

The data collected on the various Agilent systems showed a high level of reproducibility, a critical component for reliable differential analysis. Figure 3 shows an example from positive ion reversed-phase LC/TOF MS data; approximately 90 % of features were detected and extracted with CV of ≤30 %. The highly reproducible assay results provide high confidence for subsequent statistical analysis.

Table 3. Comparing the number of features found from food plates by different analytical approaches illustrated the importance of using multiple LC/MS methods and different polarities. Note that each feature represents multiple adducts and isotopes for a single compound.

LC/MS Method	Compounds targeted	No. of features positive ion mode (+)	No. of features negative ion mode (–)
ANP-ESI	Metabolites (polar)	2,957	2,005
RP-ESI	Metabolites (intermediate)	3,279	2,135
Lipid-RP-AJS	Lipids	2,689	3,816

ANP-ESI: Aqueous-normal phase chromatography coupled to the electrospray source
RP-ESI: Reversed-phase chromatography coupled to the electrospray source
Lipid-RP-AJS: Reversed-phase chromatography coupled to the Agilent Jet Stream source

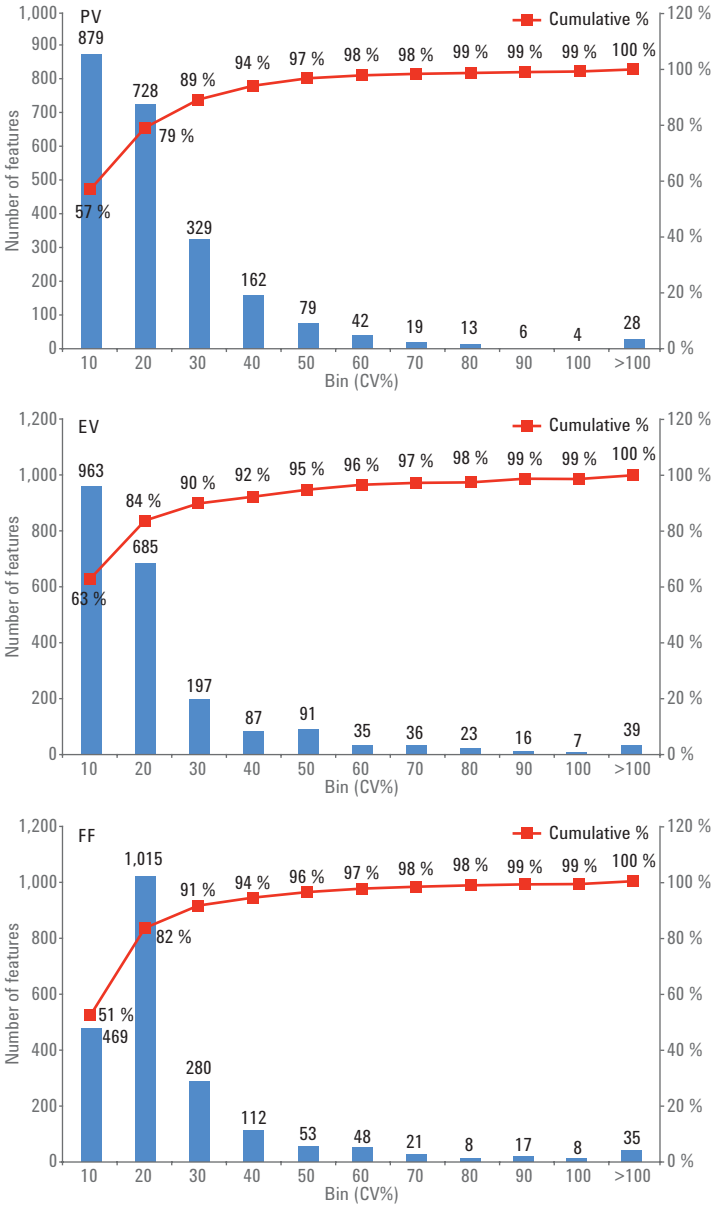

Figure 3. Histograms from RP-positive LC/TOF MS data demonstrated highly reproducible assay results.

Statistical analysis and annotation using MPP

Statistical analysis of untargeted profiling data presents substantial challenges because of the complexity of the compound profiles, and the size and variability of the data sets. MPP, a sophisticated chemometrics software with a variety of statistical tools, was used in this study for differential analysis. Principal Component Analysis (PCA) is one of the most common unsupervised data analysis methods used to detect data patterns, and facilitate the evaluation of data quality. In this study, the PCA plot shows tight clustering between the sample replicates in each food plate for all analytical techniques indicating good repeatability, and demonstrating significant differences in the composition of each plate (Figure 4).

Venn diagrams were used to compare the composition of the different plates in this study. Figure 5 shows that a large number of compounds was shared between food plates. The PV plate had the highest number of unique metabolites by all techniques. To further investigate if the shared features (Figure 5B) showed statistically meaningful differences between food plates, a two-way ANOVA was performed. We found that 1,131 out of 1,362 features were different ($P \leq 0.05$). More than 200 of the shared features had a ≥5-fold difference in abundance when comparing PV with EV, PV with FF, and EV with FF plates.

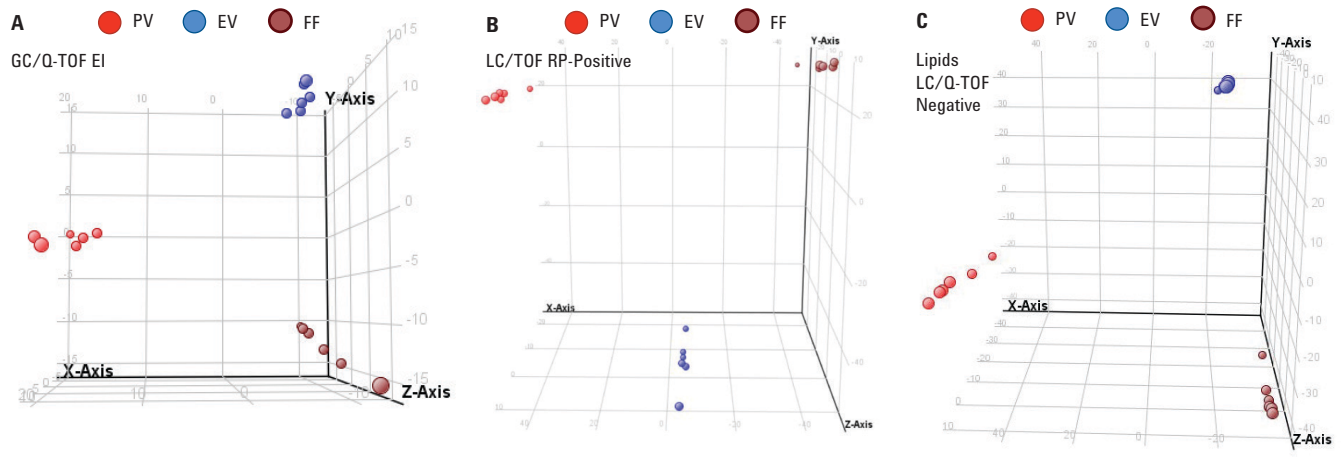

Figure 4. The PCA plot of the food plate data demonstrates excellent repeatability across all three techniques, as well as highlighting the distinct nature of the samples.

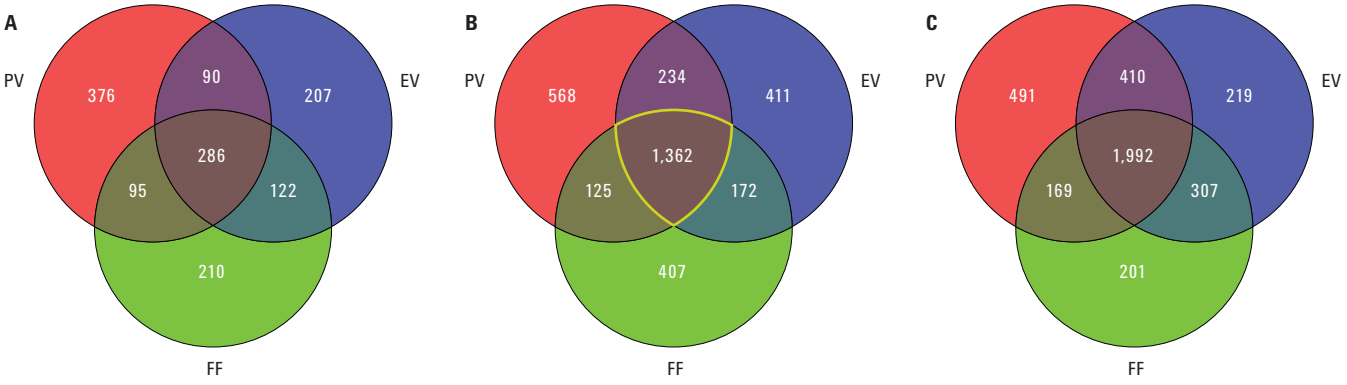

Figure 5. Number of compounds found by each technique in each plate represented by Venn diagrams. A) GC/Q-TOF, B) LC/TOF, RP-positive mode, C) LC/MS lipids. The highest number of unique components was found in the pesco-vegetarian (PV) plate.

Correlation analysis provides another useful tool to help reveal or confirm relationships between samples in an experiment. Figure 6 demonstrates the repeatability between sample replicates in each food plate, confirming the PCA assessment.

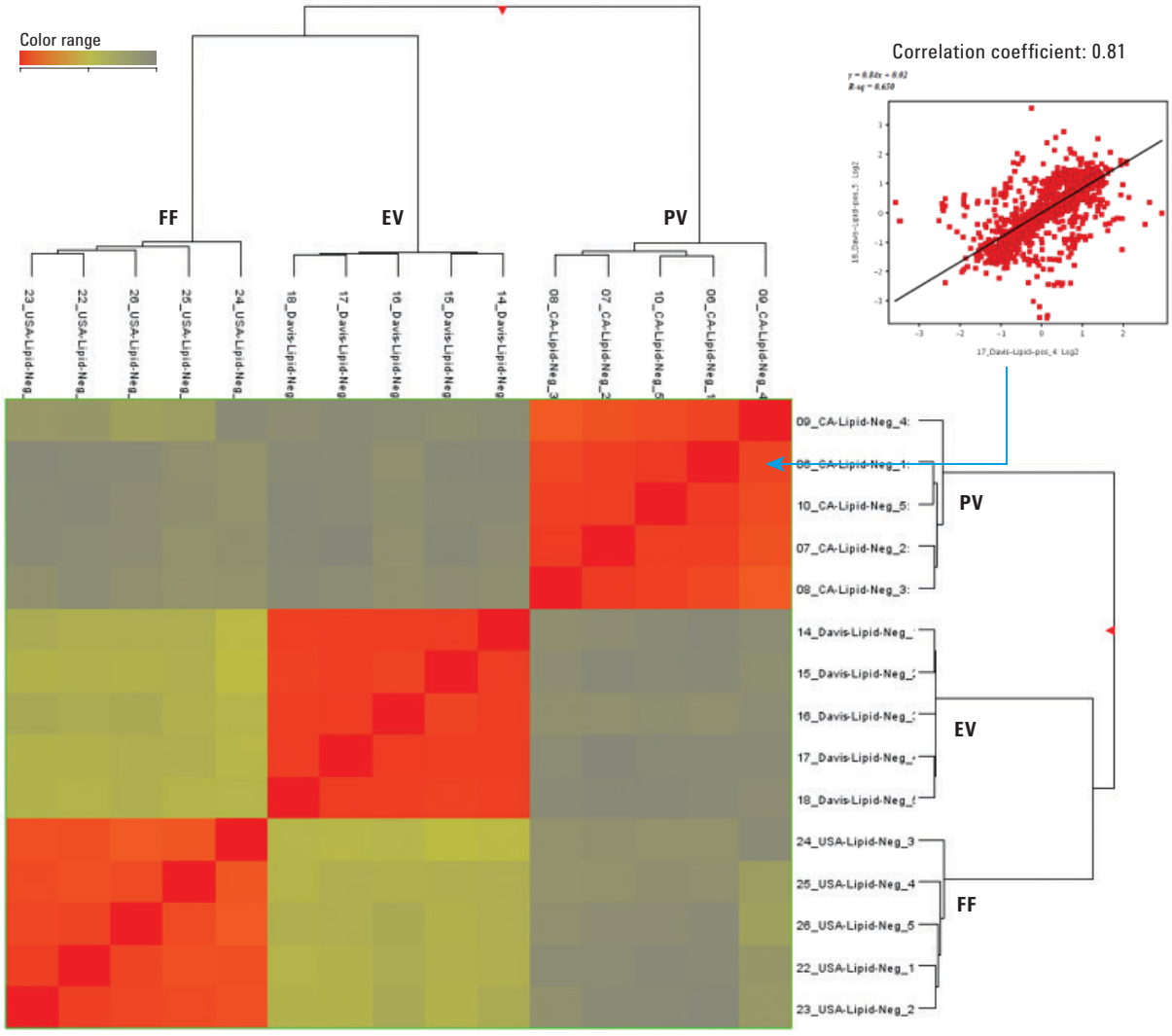

Figure 6. Example of sample correlation analysis for negative-ion LC/MS lipid analyses. The scatterplot shows the abundance correlation between two sample replicates from the PV plate.

MPP also incorporates ID Browser, a tool that allows users to annotate features based on accurate mass information and database matching using Agilent METLIN PCD or any customer-created database using Agilent PCDL Manager software.

Figure 7A shows the compound annotation results using MPP ID Browser and database matching. Using this approach, compound classes can be categorized (Figure 7B) by searching different types of databases.

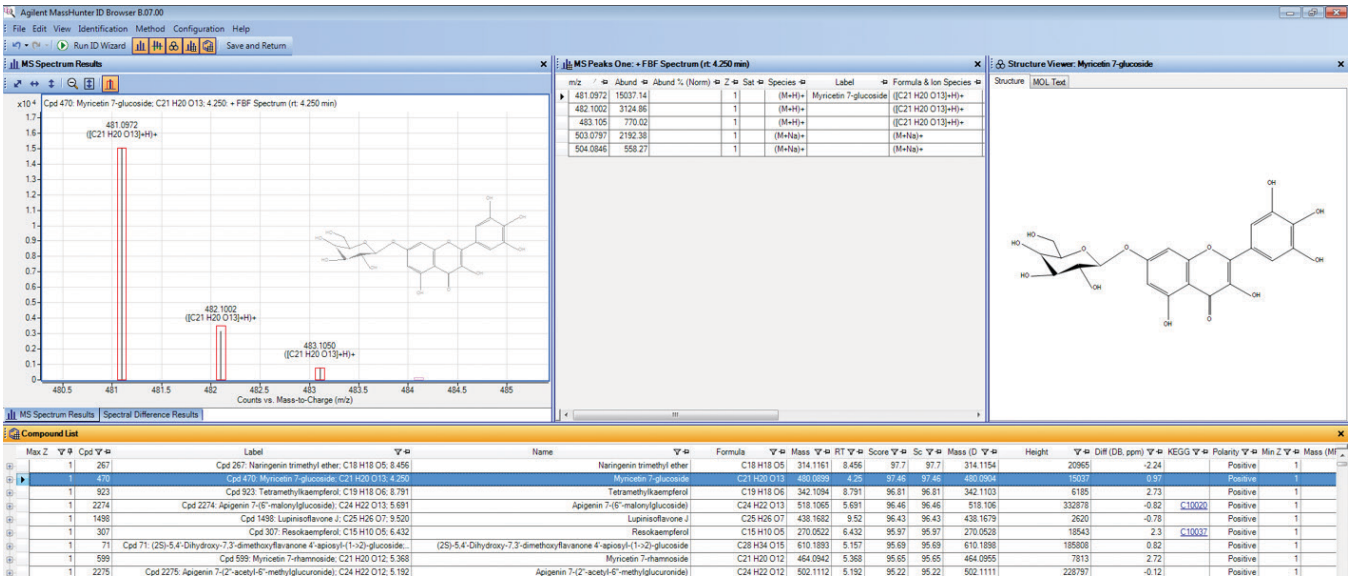

Figure 7A. Compound annotation of features found in the food plates using ID Browser and database matching.

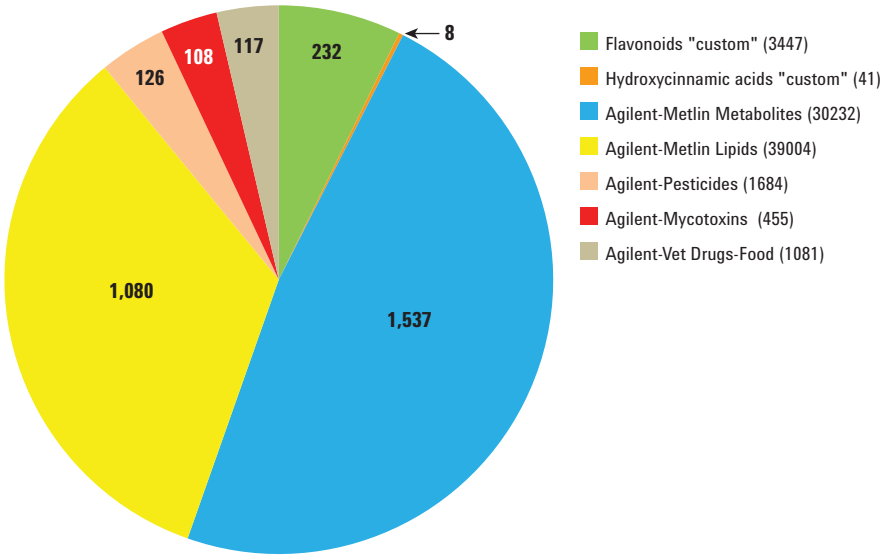

Figure 7B. Annotation results from the LC/TOF data acquired using the reversed-phase positive ion method. The numbers in parenthesis represent the size of the database, and the numbers in the pie chart represent the number of compounds matched to the database.

In a similar manner, lipidomics datasets within MPP were annotated to categorize lipid classes based on accurate mass information using SimLipid. From MPP results, the summed relative abundances of the annotated lipid classes were used to compare lipid profiles from the different food plates (Figure 8). The analyses revealed interesting differences in lipid class abundance (Figure 8A), fatty acid profiles (Figure 8B), and lipid saturation levels (Figure 8C). One notable observation was that the PV plate contained higher levels of the polyunsaturated lipids 22:6 (DHA) and 20:5 (EPA), while the FF plate contained the highest level of saturated lipids.

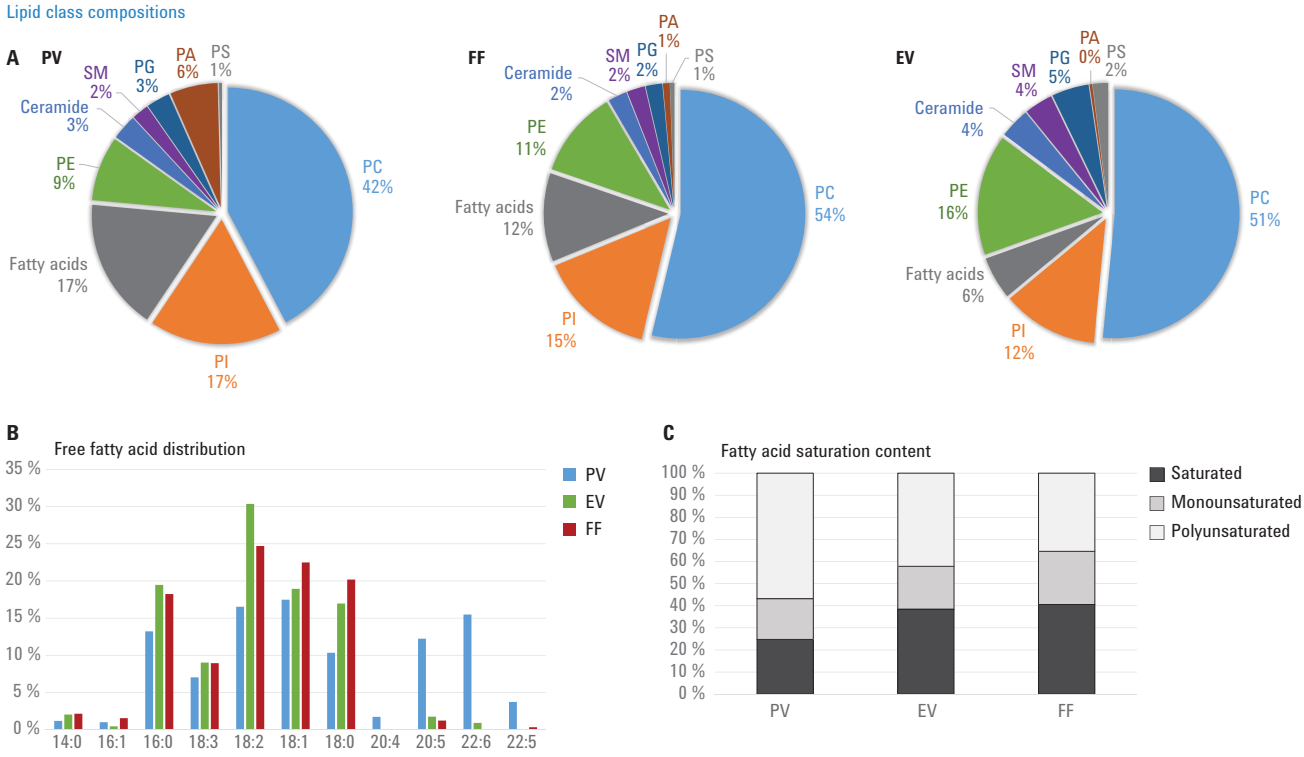

Figure 8. Relative comparisons of lipid compositions. Negative-ion LC/MS lipid data were annotated with Agilent-METLIN PCD, and compound abundances were used to compare relative differences in A) lipid class composition, B) free fatty acid profiles, and C) lipid saturation content.

A volcano plot is another tool in MPP that allows a pair-wise comparison between different groups (food plates in this study) where the results can be visualized by fold change and significance. A volcano plot for GC/Q-TOF data is shown in Figure 9, demonstrating the difference in the metabolite content between the EV and PV plates.

While many antioxidants, such as sinapinic acid, quinic acid, and gallic acid, among others, were present at higher levels in the PV plate, some of the other characteristic metabolites (components of the specific ingredients present in the EV plate) were present at higher levels in the EV plate compared to the PV plate. Examples of these metabolites are raffinose, present at high levels in beans, and daidzein, typically found in soy (Figure 9).

Compound identification with MS/MS data

Accurate mass MS/MS information provided by high-resolution Q-TOF systems is valuable for compound identification and confirmation of tentative hits, as it allows more accurate assignment of formulas for fragments derived from selected precursors.

For GC/MS metabolite analysis, EI spectra from GC/Q-TOF provide significant structural information, but the relatively low abundance of the molecular ion can complicate identification of unknowns. To identify and confirm the molecular ion, CI spectra from GC/Q-TOF were also acquired.

The compound identification and confirmation workflow for GC/Q-TOF using CI MS and EI MS/MS can be demonstrated with caffeine, which was tentatively identified by searching the NIST14 library. Caffeine eluted in an extremely congested part of the chromatogram (Figure 10A), and its identification in this type of extract represents a challenge for any compound-finding algorithm. In CI mode, an abundant molecular ion, as well as isotopes, were observed with the correct mass and expected isotope ratios corresponding to caffeine (Figure 10B). Since an authentic MS/MS spectrum for caffeine was not available, further confirmation of the match was done with MSC software (Figure 10C).

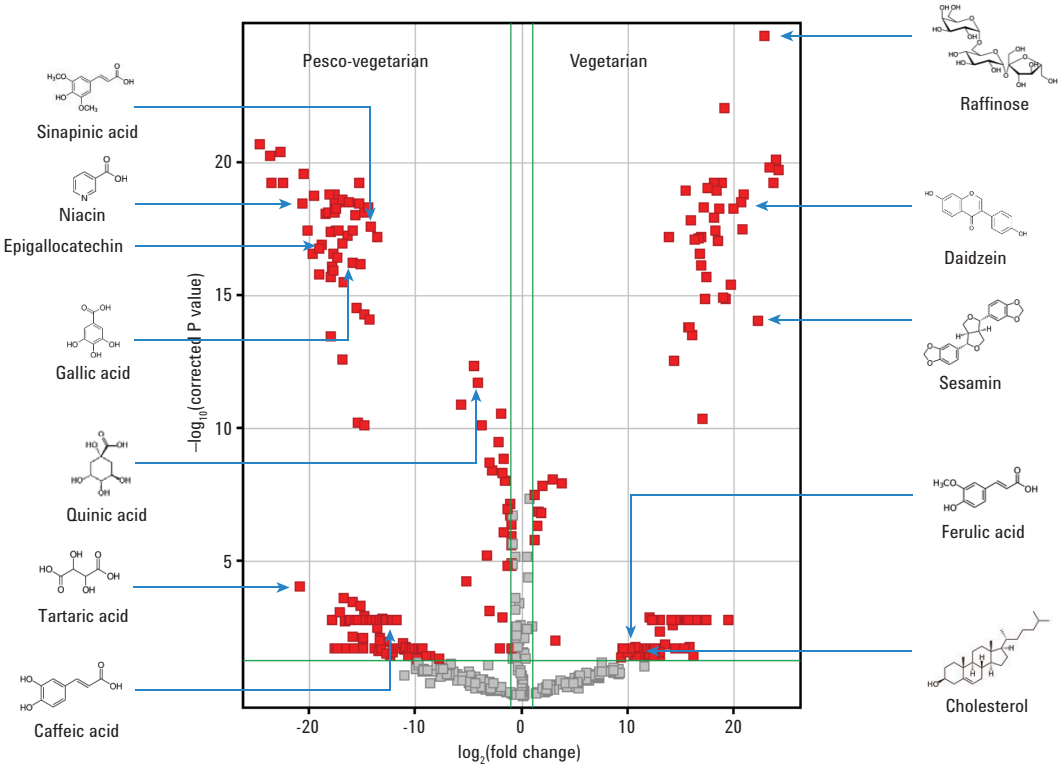

Figure 9. Example of a volcano plot in MPP shown for GC/Q-TOF data. A few selected differential compounds are labeled.

When MS/MS is performed by selecting a molecular ion as a precursor, the compound information from MassHunter Qualitative Analysis can be imported to MSC for structural elucidation. Based on the tentative molecular ion formula(s) of the unknown or tentatively identified compound, MSC searches a user-specified database for any possible structures corresponding to the proposed formula or observed mass. It then evaluates the observed MS/MS fragmentation versus the possible fragments for the proposed match. The resulting MSC score accounts for mass accuracy for the precursor and fragment ions and how well the observed fragmentation is explained by the proposed match. In this example, caffeine had the highest number of literature references, a mass error of 0.1 ppm, and an overall score of 89.7 (Figure 10C).

Figure 10D summarizes the GC/Q-TOF compound identification results. The highest amounts of antioxidants and flavonoids were observed in the PV plate, followed by the eastern vegetarian EV plate. The FF plate was characterized by a high number of the pyrolysis products that results from the heat-processing of foods, the highest number of sugars and saturated fats, and a lower number of unique amino acids.

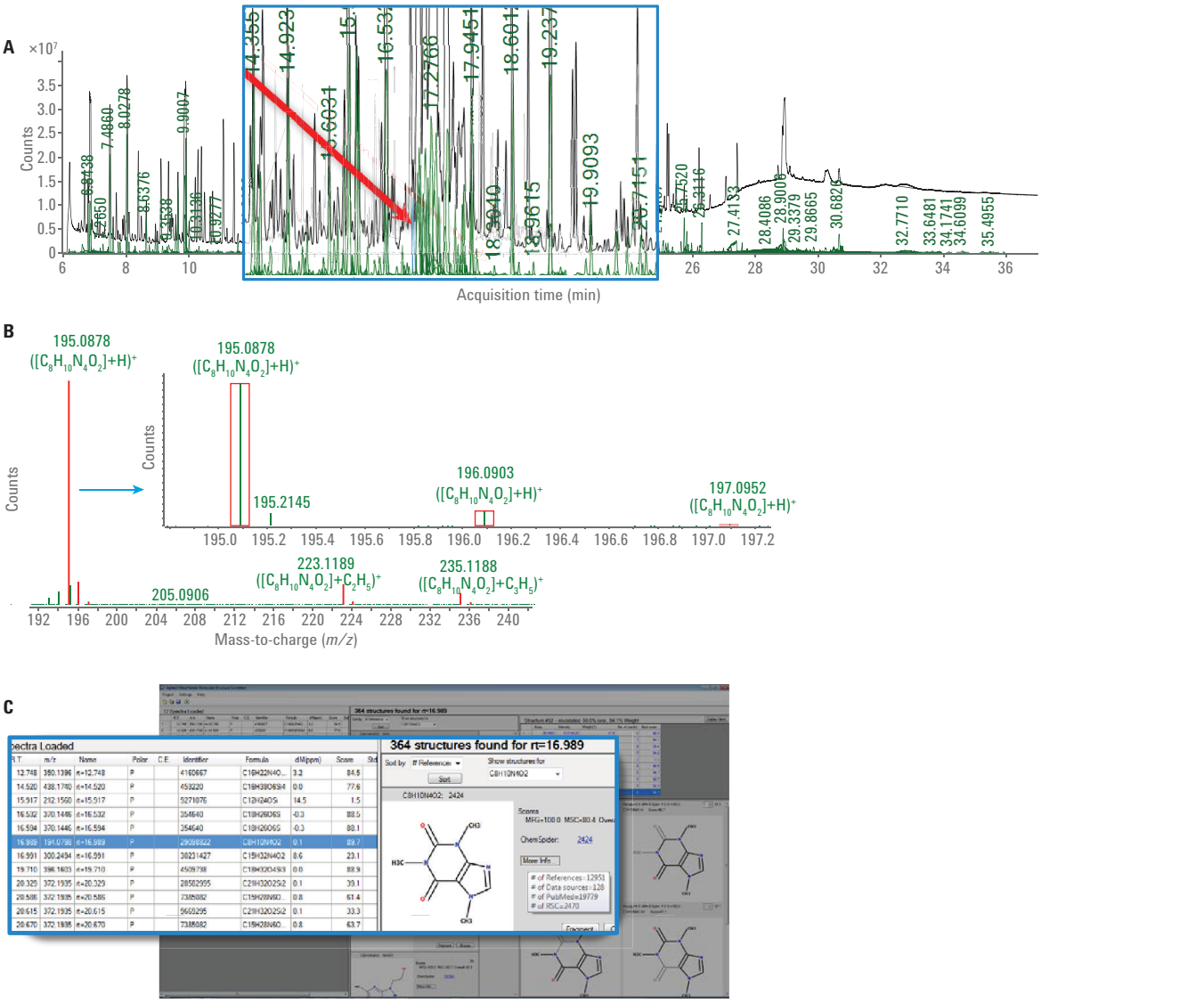

Figure 10. Compound identification workflows and results for GC/Q-TOF. A) EI TIC overlaid with components EIC from Agilent Unknowns Analysis. B) CI spectrum, showing methane adducts of the molecular ion with overlaid theoretical isotope pattern automatically assigned by Molecular Formula Generator (MFG) in Agilent MassHunter Qualitative software. C) ID confirmation using MS/MS and Molecular Structure Correlator (MSC).

To validate the initial annotations from the LC/MS metabolomics data, targeted MS/MS experiments were performed for those metabolites of interest using a 6550 iFunnel Q-TOF LC/MS. Compound identification and structure elucidation were achieved using either MS/MS library matching or structure correlation using Molecular Structure Correlator (MSC)[6,10]. The MS/MS spectra were searched against Agilent METLIN PCDL, and the spectral comparison is shown, as demonstrated for the identification of L-glutamate (Figure 11A). The mirror image plot (middle) showed excellent matches between the acquired sample spectrum (top) and the library standard spectrum (bottom) for both *m/z* values and abundances of the precursor and fragment ions with a reverse match score of 96. Using this approach, many food components were identified. Betaine and L-phenylalanine were found in all food plates, daidzein was found in both PV and EV plates, quercetin 3-galactoside and neohesperidin were found in only the PV plate, and genistein was only found in the EV plate. A mycotoxin metabolite, mycrosporin/physcion, was identified in the PV and EV food plates.

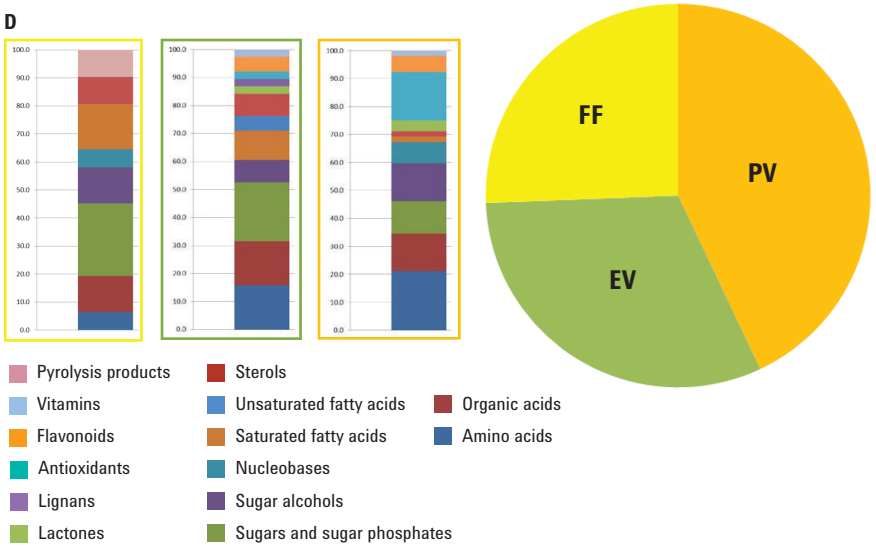

Figure 10D. GC/Q-TOF compound identification results summary.

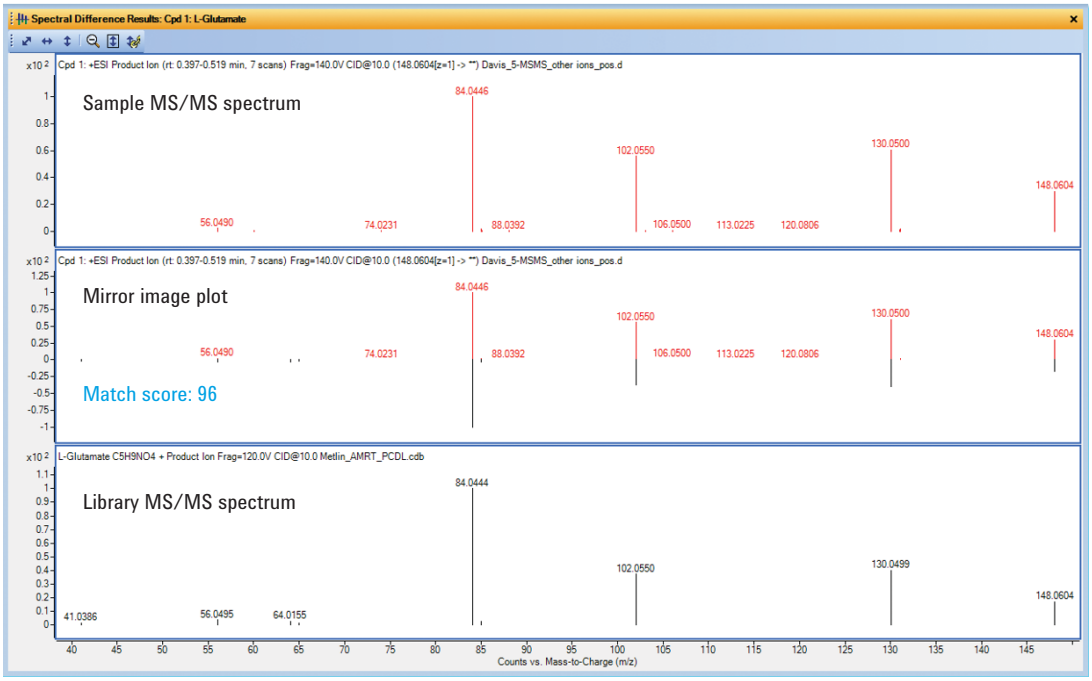

Figure 11A. Compound identification using LC/Q-TOF MS/MS spectra matching with a reverse score of 96 for L-glutamate in positive ion ESI mode. The reverse score reflects ion matches for those peaks from the library standards that are reflected in the spectrum acquired from the sample. The collision energy (CE) of 10 eV was used to acquire the sample MS/MS spectrum.

When an acquired LC/MS/MS spectrum had poor matches or no matches to the Agilent MS/MS library, MSC was used for putative compound identification. Based on the tentative molecular ion formula(s) of the unknown or tentatively identified compound, MSC searches a user-specified database for any possible structures corresponding to the proposed formula or observed mass, then evaluates the observed MS/MS fragmentation versus the possible fragments for the proposed match. Figure 11B shows an example of an MSC search result where a total of 22 possible structure matches were found for the unknown highlighted with blue. The top match, quercetin 3-rhamnosyl-(1→6)glucosyl-(1→6)-galactoside, showed a high MSC score, and accounted for a high percentage of the observed fragment ions, suggesting that this is a likely match for the unknown.

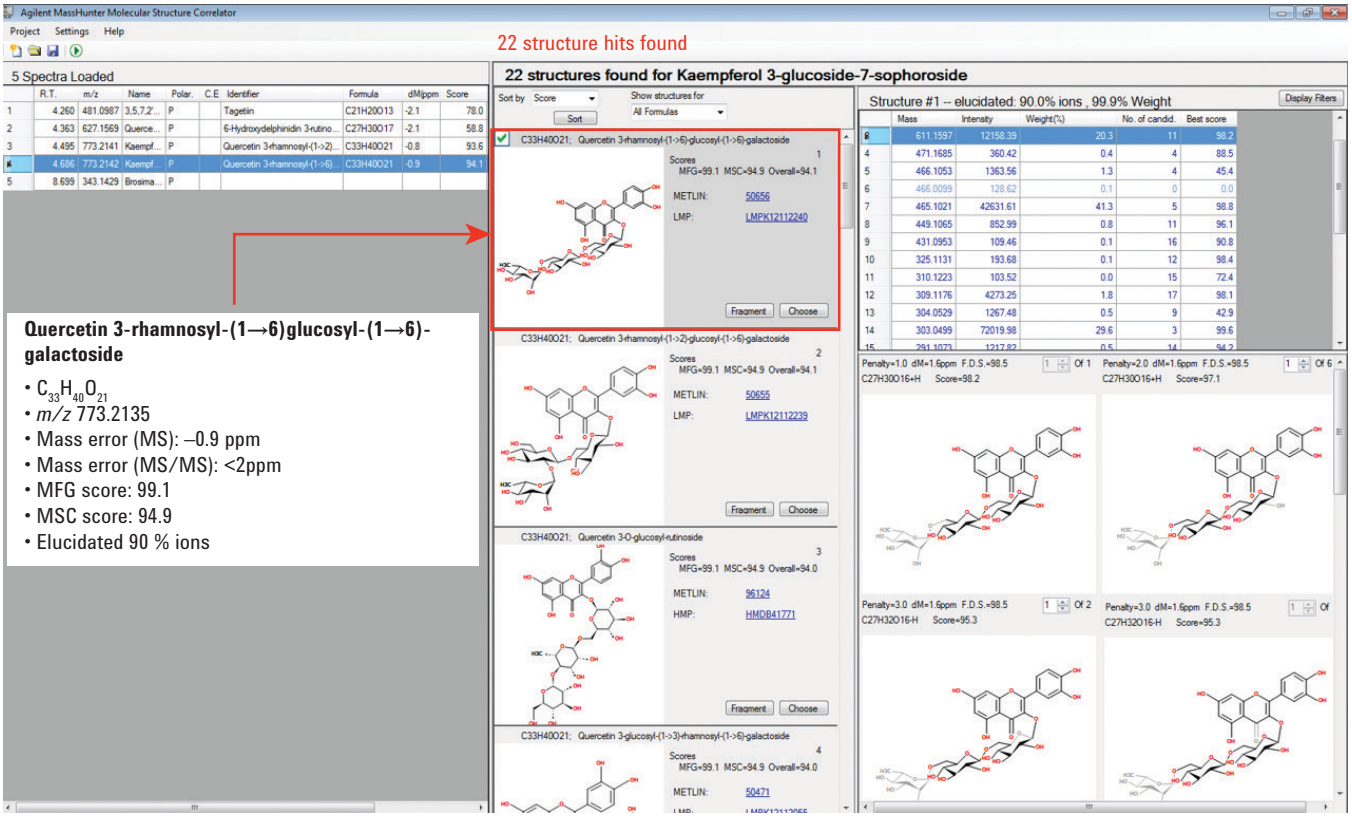

Figure 11B. MSC results obtained by searching Agilent PCDL.

LC/MS/MS lipidomics identifications were accomplished by searching extracted MS/MS spectra with SimLipid software. Figure 12 shows an example where two isomers of phosphatidylethanolamine (PE) were identified, and the MS/MS data afforded additional structural information. This high-confidence lipid identification approach can be used to construct alternative workflows to profile lipid classes in more detail (Figure 13). A custom subset lipid PCDL can be curated with retention time values from identified lipids, and this PCDL can be leveraged in Profinder for a recursive targeted workflow (Figure 13A). In this manner, the PE lipid class was profiled in depth, and significant differences were observed across the three food plates, even across PE isomers (Figure 13B).

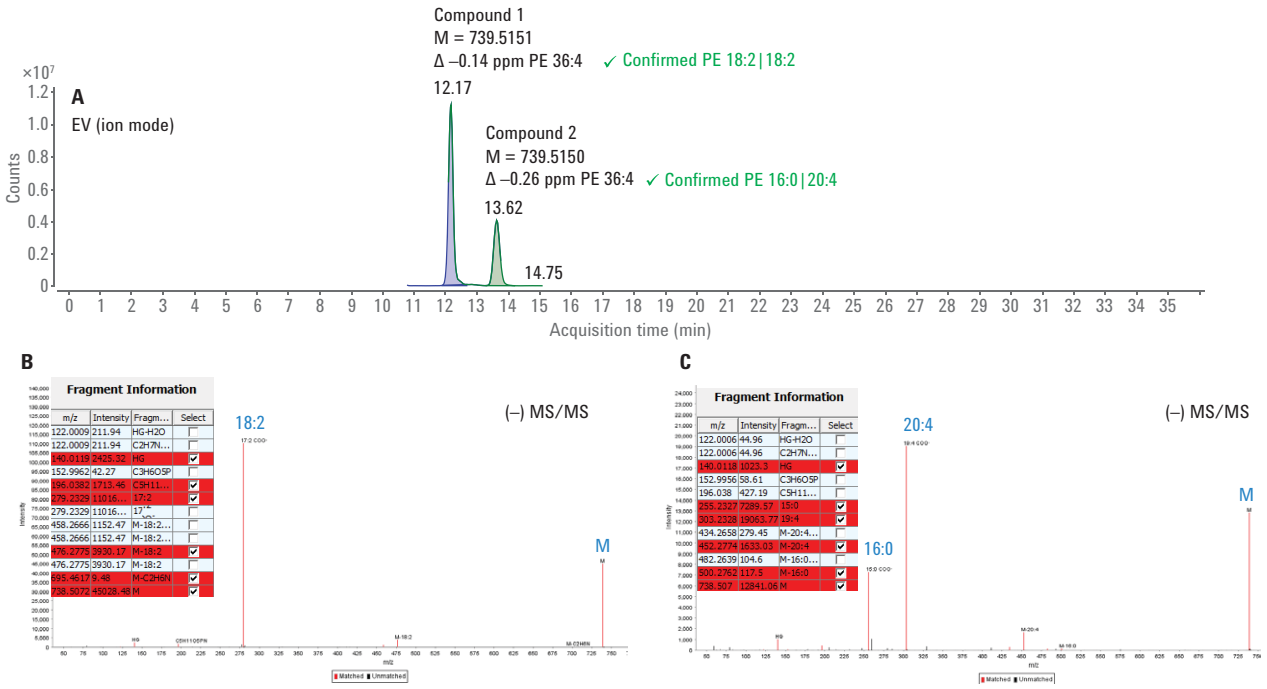

Figure 12. Lipid MS/MS Annotation with SimLipid. Two isobaric lipids were resolved by retention time, and automated identification with SimLipid confirmed these compounds as distinct PE 36:4 isomers with different fatty acyl compositions (PE 18:2|18:2 and PE 16:0|20:4).

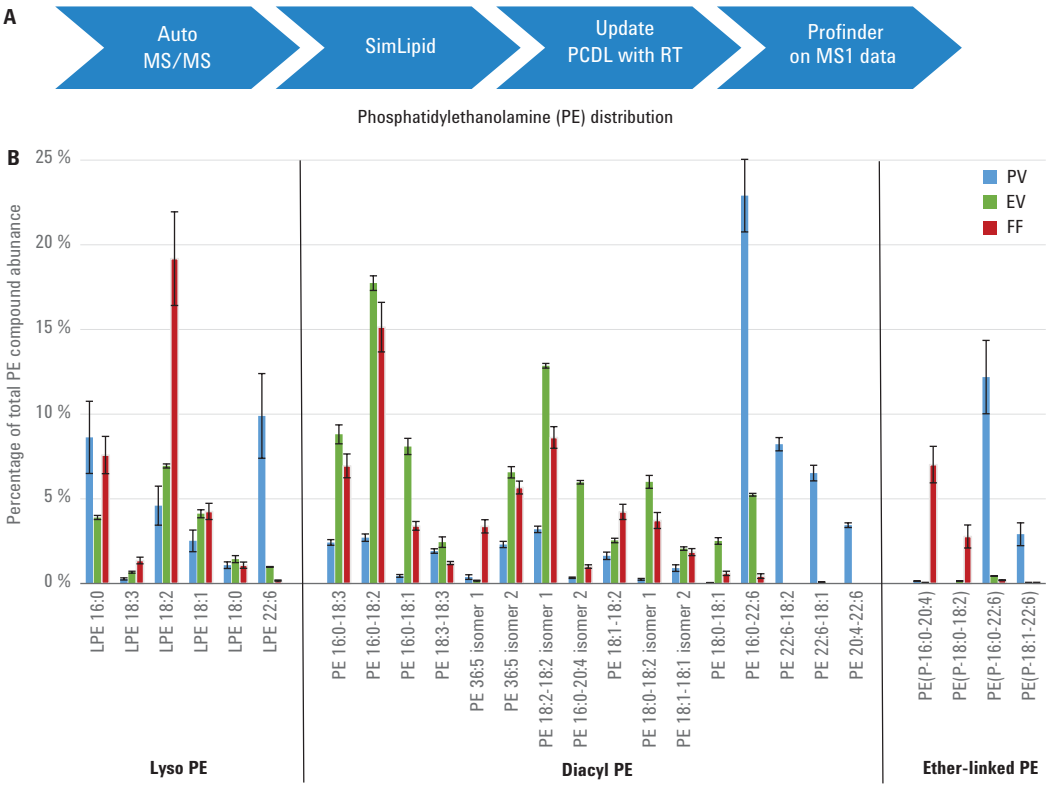

Figure 13. Targeted workflows for in-depth lipid profiling. A) PE lipids were identified with high confidence from MS/MS data acquired from pooled plate replicates. Observed retention times were used to curate a subset PE PCDL, which was used as the source for Profinder Batch Targeted Feature Recursion for individual plate replicates. B) Results showing PE distribution showing the relative abundances of lyso-PEs, Diacyl PEs, and ether-linked PEs profiled in depth across the three food plates. Error bars = ±1 SD.

Conclusion

Because of the complex nature of food, no single extraction, separation, or ionization technique can cover all compounds. The Agilent portfolio of both LC and GC high-resolution accurate mass TOF-based mass spectrometers enabled the comprehensive analysis of metabolites and lipids with a wide range of physiochemical properties, using a common, powerful data analysis platform. Multiple analytical techniques, in combination with the advanced differential analysis software workflow, were crucial to successfully analyzing and interpreting the food profiling results, confirming the diverse nature of the three food plates.

References

1. Cevallos-Cevallos, J. M.; *et al.* Metabolomic analysis in food science: a review. *Trends in Food Science & Technology* **2009**, 20, 557-566.
2. Johanningsmeier, S. D.; Harris, G. K.; Klevorn, C. M. Metabolomic technologies for improving the quality of food: practice and promise. *Annu. Rev. Food Sci. Technol.* **2016**, *7*, 413-438.
3. Wishart, D. S. Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology* **2008**, *19(9)*, 482-493.
4. Fiehn, O. Metabolomics – The link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155-171.
5. Wu, M.; *et al.* Metabolomics of opiate-induced changes in murine brain, *Agilent Technologies Application Note*, publication number 5991-2481EN.
6. Dai, Y.; Fischer, S. M. Metabolomics batch data analysis workflow to characterize differential metabolites in bacteria, *Agilent Technologies Application Note*, publication number 5991-5706EN.
7. Jenkins, S.; Fischer, S. M.; Sana T. R. Mass Profiler Professional and Personal Compound Database and Library Software facilitate compound identification for profiling of the yeast metabolome, *Agilent Technologies Application Note*, publication number 5990-9858EN.
8. Sartain, M.; Sana, T. Impact of Chromatography on Lipid Profiling of Liver Tissue Extracts, *Agilent Technologies Application Note*, publication number 5991-5494.
9. MassHunter Profinder: Batch processing software for high quality feature extraction of mass spectrometry data, *Agilent Technologies Technical Overview*, publication number 5991-3947EN.
10. S. Joseph, Y. Dai, Pharmaceutical Impurity Identification and Profiling Using Agilent Q-TOF LC/MS Combined with Advanced MassHunter Data Processing Software, *Agilent Technologies Application Note*, publication number 5991-1375EN.

www.agilent.com/chem

Published in the USA, August 31, 2017
5991-8312EN

Agilent Technologies