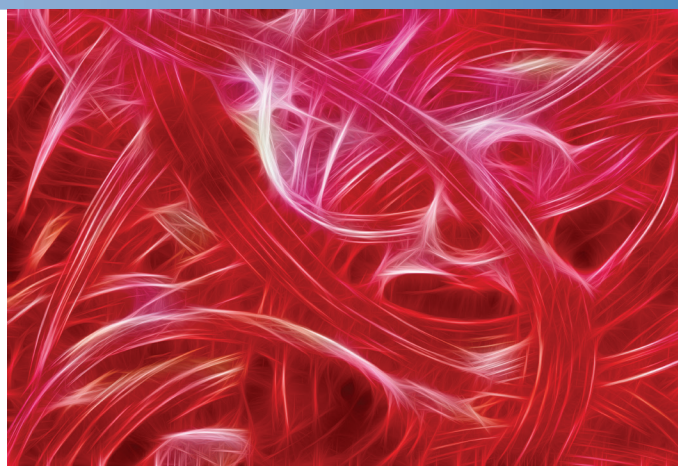
Untargeted Metabolomics Using Orbitrap-Based GC-MS

Stefan Weidt,[2] Bogusia Pesko,[2] Cristian Cojocariu,[1] Paul Silcock,[1] Richard J. Burchmore,[2] and Karl Burgess[2]
[1]Thermo Fisher Scientific, Runcorn, UK
[2]Glasgow Polyomics, University of Glasgow, Glasgow, UK

Application Note 10457

Key Words
Forensics, Metabolomics, Multivariate Statistics, Q Exactive GC System



Introduction

Metabolomics aims to characterize and quantify the complete small molecule complement, or metabolome, of a biological system. The metabolome consists of a diverse mixture of small molecules, including amino acids, sugars and phosphosugars, and biogenic amines and lipids. Untargeted metabolomics is exceptionally challenging due to the requirement to both identify and quantify hundreds of different compounds with limited a priori knowledge of the metabolites. It is, therefore, advantageous to use a detection system that is not only capable of sensitive detection of specific molecules in an untargeted way, but can also provide accurate mass information for confident confirmation and structural elucidation of unknowns.

Gas chromatography-mass spectrometry (GC-MS) is routinely used for metabolomics applications due to its inherent advantages, especially its chromatographic resolution, reproducibility, peak capacity, and convenient spectral libraries. GC provides excellent chromatographic separation capability for biomarker discovery using untargeted metabolomics, but has previously been hampered by the lack of high-end mass spectrometry support providing the dynamic range, accurate mass, and scan rate sufficient to analyze very complex samples, such as mammalian muscle tissue. The polar nature of the majority of central metabolites means that derivatization must be performed to allow effective volatilization and ensure good chromatography. High sample throughput and advanced automation is required for metabolomic analysis, especially for clinical metabolomics.

This work demonstrates the application of a complete untargeted metabolomics workflow using a novel Thermo Scientific™ Orbitrap™ MS-based GC to detect biomarkers for time of death in a rat model. Estimation of post-mortem interval (PMI) is one of the most critical, yet difficult, tasks in forensic investigation, particularly after the cadaver has equilibrated to the ambient environmental temperature. Current methods to determine PMI are inaccurate and primarily based on visual inspection of the body. A laboratory-based method, using a robust biomarker for PMI, would assist forensic investigation.

This GC-MS configuration using an Orbitrap-based detector enables ultra-high mass resolution, sub-ppm mass accuracy, a large dynamic range, and a scan rate commensurate with the efficient quantitative analysis of highly complex metabolomic samples. The high resolution, mass accuracy, and scan speed is critical for consistent data deconvolution to permit the detection of species from overlapping TIC peaks, allowing for an untargeted metabolomics pipeline. Accurate mass electron ionization (EI) fragment patterns are also suitable for matching against the widely available NIST and Wiley libraries for tentative compound identification, while providing accurate mass for more in-depth characterization.

Thermo SCIENTIFIC

Instrument and Method Setup

Sample Preparation

Rat thigh muscle tissue sections were sampled from individual rats, post-mortem at increasing times of decomposition. Metabolite extractions were performed by homogenizing tissue sections with chloroform/methanol/water (1:3:1) and incubating for 1 hour over ice. Protein and DNA were pelleted by centrifugation. The supernatant was removed and stored at -80 °C until required.

Sample Derivatization

Of the extracted sample, 200 µL was transferred to a 9 mm screw cap borosilicate glass 1.5 mL vial. Samples were then dried in a Thermo Scientific™ Reacti-Vap™ evaporator, with a gentle nitrogen stream at 30 °C for 60 min. All following derivatization steps were performed using the Thermo Scientific™ TriPlus™ RSH autosampler. To each dried vial, 20 µL of 20 mg/mL (w/v) methoxyamine HCL in pyridine was added. The vials were vortexed for 10 seconds and incubated at 30 °C for 60 min. Following the methoximation step, 30 µL of MSTFA + 1% TMCS (N-Methyl-N-(trimethylsilyl) trifluoroacetamide + 1% trimethylchlorosilane) was added, followed by a further 30 sec of vortexing. Silylation was performed by incubating the vials at 45 °C for a further 60 min. Samples were cooled to room temperature and were then ready for injection.

GC-MS Analysis

All experiments used a Thermo Scientific™ Q Exactive™ GC hybrid quadrupole-Orbitrap mass spectrometer. Sample introduction was performed using a TriPlus RSH autosampler, and chromatographic separation was obtained using a Thermo Scientific™ TRACE™ 1310 GC and a Thermo Scientific™ TraceGOLD™ TG-5SilMS 15 m × 0.25 mm I.D. × 0.25 µm film capillary column (P/N: 26096-1301). Additional details of instrument parameters are shown in Tables 1 and 2.

Data Processing

The data analysis pipeline started with MSConvert (part of the ProteoWizard suite[1]) to convert the raw data files into MzXML files. Peak picking was performed using the XCMS package[2] and the centWave algorithm. Detected peaks were output in the PeakML format, and post-processing of the detected features (filtering for minimum detections per set of replicates, relative standard deviation, and correlation matching for EI fragment set grouping) was performed using the MzMatch.R package.[3] The resulting text output was processed using the IDEOM software[4] for univariate statistics and SIMCA™ 13.0.3[5] for multivariate statistical analysis. Peak clusters were identified using Thermo Scientific deconvolution software.

Results and Discussion

Eight rat cadavers were maintained at room temperature for four days. Muscle tissue was extracted on a daily basis to assay decomposition by change in metabolite concentration for a total of 16 samples. Samples were run in randomized order to ameliorate systematic errors. The complete optimized workflow for metabolomics is shown in Figure 1.

Table 1. GC Temperature program.

TRACE 1310 GC Parameters	
Injection Volume (µL)	1.0
Liner	Single taper (P/N 453A1345)
Inlet (°C)	250
Inlet Module and Mode	SSL, split 1:60
Carrier Gas (mL/min)	He, 1.2
Oven Temperature Program	
Temperature 1 (°C)	70
Hold Time (min)	2
Temperature 2 (°C)	325
Rate (°C/min)	10
Hold Time (min)	8.5

Table 2. Mass spectrometer parameters.

Q Exactive GC Mass Spectrometer Parameters	
Transfer Line (°C)	275
Ionization Type	EI
Ion Source (°C)	230
Electron Energy (eV)	70
Acquisition Mode	Full scan
Mass Range (*m/z*)	50–750
Mass Resolution (FWHM at *m/z* 200)	60,000
Lockmass (*m/z*)	207.03235

Figure 1. Workflow for the Q Exactive GC system metabolomics studies. Color-coding shows work package assignment: green for wet lab biologists, orange for lab technologist, purple for instrumentation, and blue for informatician.

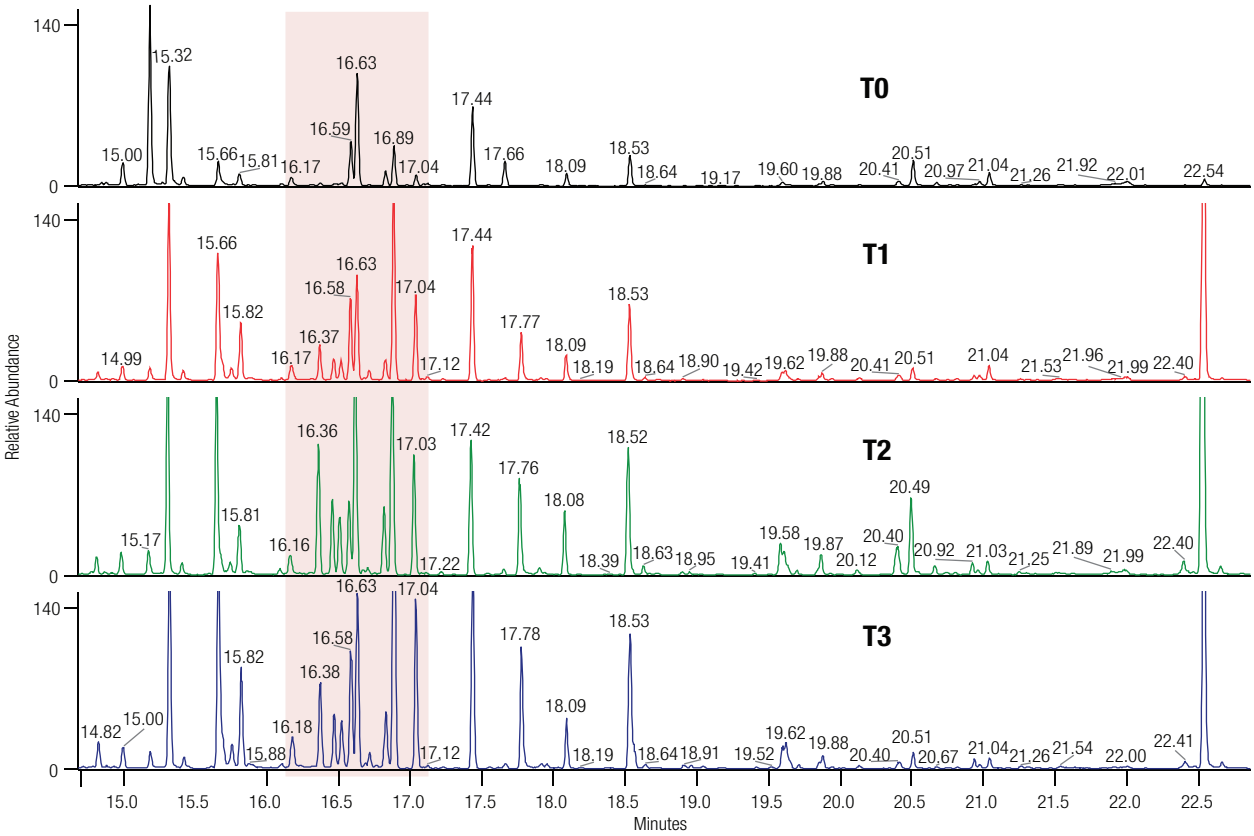


Figure 2. GC-MS chromatograms of a rat muscle tissue sample decayed for 0–3 days from top to bottom, with retention time on the X axis and intensity, fixed to 1E9 counts per second, on the Y axis. As an example, note the change in intensity over time for peaks in the highlighted region.

Discovery Phase

A complex chromatogram was obtained for each sample (Figure 2). The Q Exactive GC system makes these chromatograms possible by offering a wide dynamic range to capture metabolites at various concentration levels without the loss of accurate mass information. Automated peak picking (using a combination of XCMS and MzMatch.R) was necessary to extract each EI peak cluster. During this process, 1193 distinct peak clusters were detected and quantified across the dataset, with a conservative intensity threshold of 100,000 counts. An example of a deconvoluted peak cluster is shown in Figure 3.

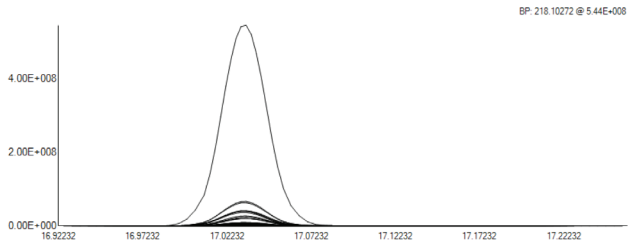


Figure 3. Deconvoluted peak cluster putatively identified as tyrosine.

Quantitation Phase

Results were initially analyzed using univariate statistical analysis. Student's t-tests were applied to compare each time point to time zero. Mean T0 intensity of each metabolite was set to 1, and fold changes were displayed in relation to 1 to allow easy comparison of metabolites with significantly different intensities (Figures 4 and 5). Mean intensities of 272 significantly (P value <0.05) changing metabolites were detected and an example of a quantitation matrix containing detected peak clusters is shown in Table 3.

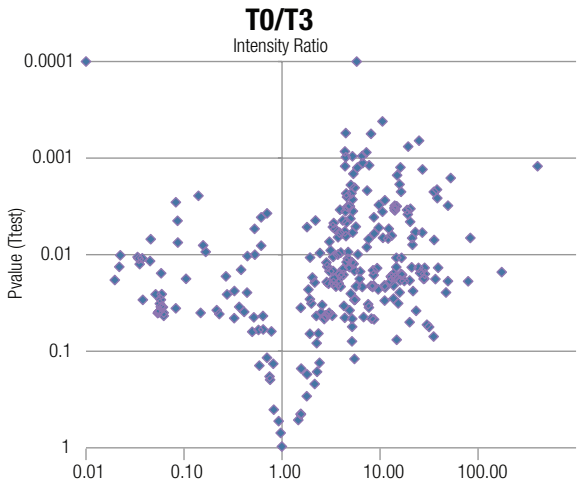
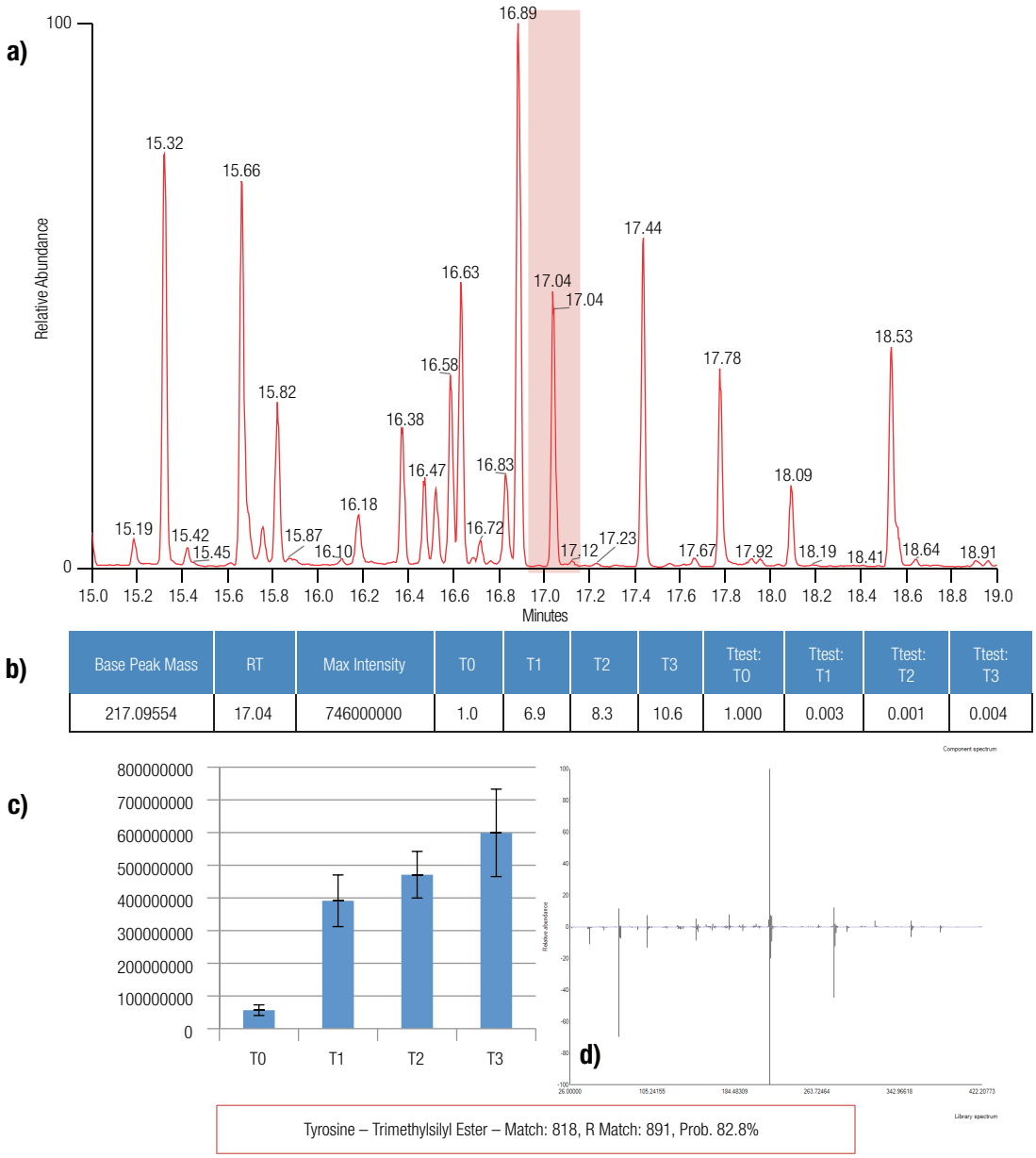


Figure 4. Volcano plot of significantly changing metabolites. Individual metabolites are represented by dark blue diamonds. Intensity ratio (the X axis) is the fold-change of the metabolite at T3 comparison to T0, while the Y axis shows the P-value of the metabolite. Therefore, metabolites in the top right and left corners are those with the largest fold change and highest statistical significance.



Base Peak Mass	RT	Max Intensity	T0	T1	T2	T3	Ttest: T0	Ttest: T1	Ttest: T2	Ttest: T3
217.09554	17.04	746000000	1.0	6.9	8.3	10.6	1.000	0.003	0.001	0.004

Figure 5. Summary of data analysis pipeline. Peaks are detected and deconvoluted from the background (a), then displayed as basepeaks in the IDEOM metabolomics software (b) along with quantitative information and statistical measures of quantitation. IDEOM also provides graphical information on the quantitation (c), where the graph displays the mean intensity and standard deviation for each condition, in this case consisting of four biological replicates each. Interesting peaks identities are assigned using the NIST libraries, the match quality is demonstrated in (d) with the score for the assigned derivatized amino acid underneath.

Table 3. Quantitation matrix of detected peak clusters. Base peaks (the most intense peak from a peak cluster), retention time (RT) and maximum detected intensity are shown in columns 1–3. Intensities are normalized to 1 for T0 and other time points are compared to T0 intensities with color-coding as appropriate in columns 4–7. Columns 8–11 contain T-test P-values for each comparison.

Base Peak Mass	RT	Max Intensity	T0	T1	T2	T3	Ttest: T0:T0	Ttest: T1:T0	Ttest: T2:T0	Ttest: T3:T0
219.1100	10.71	507930367	1.0	2.7	3.4	4.5	1.000	0.008	0.000	0.003
232.1184	11.78	291673570	1.0	1.5	1.6	2.1	1.000	0.043	0.055	0.030
232.1184	12.41	941914122	1.0	5.9	8.5	11.0	1.000	0.013	0.001	0.002
156.0840	15.32	1163630714	1.0	2.1	1.9	2.0	1.000	0.019	0.008	0.021
174.1128	16.18	203636192	1.0	2.5	3.2	5.0	1.000	0.015	0.001	0.026
156.1203	16.88	1169362271	1.0	3.6	3.9	5.1	1.000	0.000	0.000	0.007

Multivariate statistical analysis was also performed, using the SIMCA software.[5] Data was log transformed, Y categories were set as the individual time points and a partial least squares discriminant analysis (PLS-DA) model of the data was generated. Scores and loadings plots were generated and showed distinct clustering and separation of each time point along the principle component (Figure 6). From this analysis, it is clear that the samples taken immediately post-mortem (RAT_T0) cluster together and are significantly different from decomposing rat samples (T1 to T3). Group clustering and a continuum of decomposition can be observed from the T1–T3 samples. The shift on the X or Y axis denotes the contribution a metabolite makes to the separation between the sample clusters shown in the scores plot. In this case, the X axis separates T0 from T1–3, and the Y axis separates T1, T2, and T3 (Figure 6). Each blue point on the loadings plot denotes a detected metabolite consisting of a cluster of EI fragment ions (Figure 7).

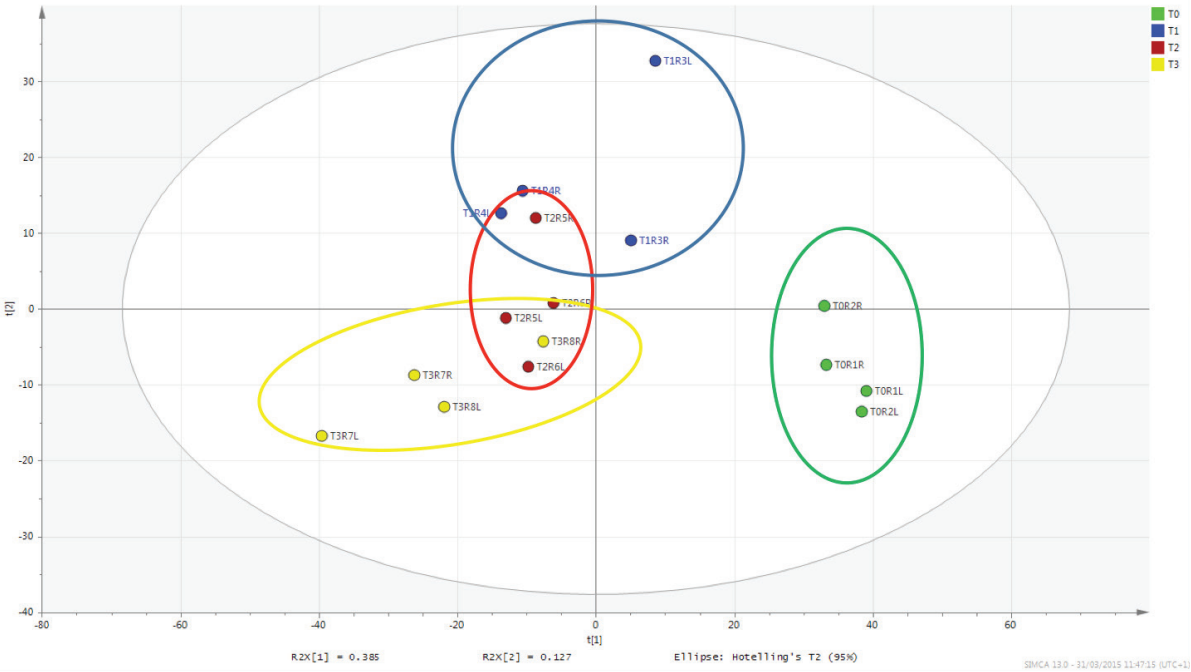


Figure 6. PLS-DA model of the decomposition data. A supervised multivariate analysis that collapses high-dimensional data (e.g. a large number of metabolites with varying intensities) to principal components that encompass the majority of variance in the dataset. In this case the X axis is principal component 1 and the Y axis is principal component 2. Note that the samples cluster appropriately—each group clusters together and T0 is distinctly separated from the other groups.

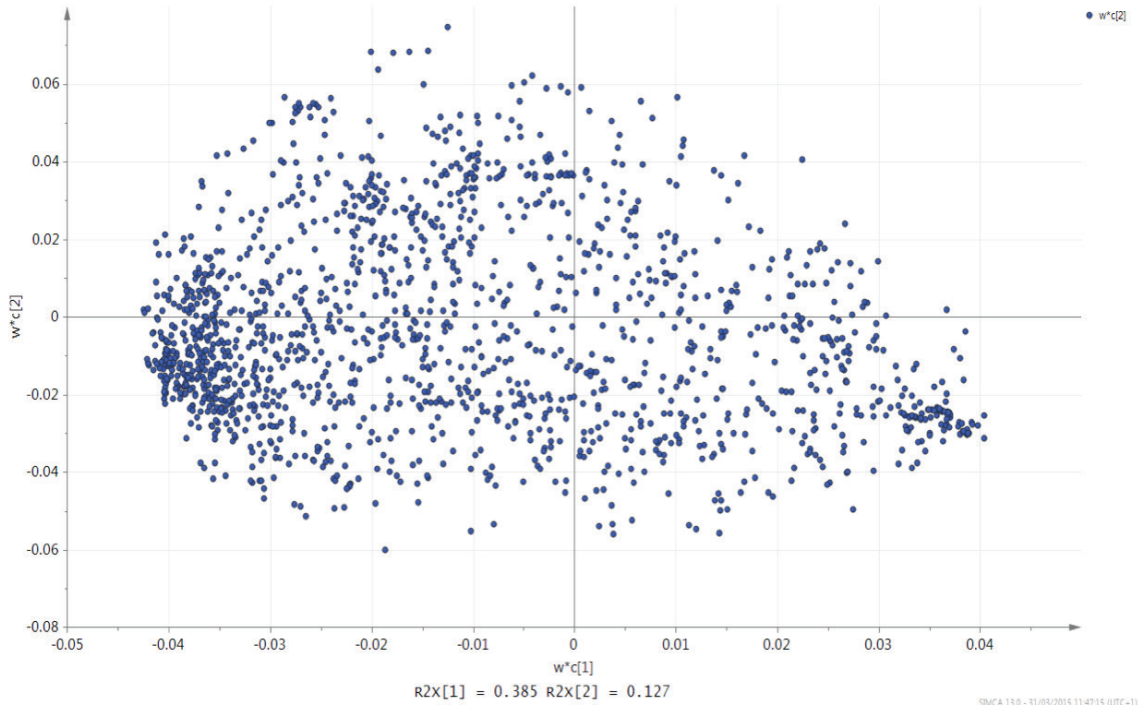


Figure 7. Loadings plot for the PLS-DA model. Metabolites (denoted with a blue dot) cluster according to their contribution to separation of the groups shown in Figure 6. Thus, for example, metabolites on the far right contribute considerably to the definition of T0 samples.

Table 4. List of putative IDs of metabolites increasing during decomposition. Compound confirmation can be performed using accurate mass and formula prediction. Putrescine was detectable only due to the high resolution and mass accuracy allowing deconvolution from background ions.

Putative Compound ID	RT (min)	NIST Forward Match	Fold Increase Compared to T0	Base Peak Fragment Elemental Composition	ppm Accuracy (Base Peak)	ppm Accuracy (Molecular Ion)
L-Threonine, 3TMS	10.71	795	2.8	$C_9H_{24}ONSi_2$	0.27	0.13
L-Aspartate, 3TMS	11.78	707	7.0	$C_9H_{22}NO_2Si_2$	0.18	0.34
L-Methionine, 2TMS	12.40	749	15.0	$C_7H_{18}NSSi$	0.24	0.04
L-Glutamine-3TMS	15.32	815	2.0	$C_7H_{14}NOSi$	0.53	0.21
Putrescine, 4TMS	16.18	870	2.0	$C_7H_{20}NSi_2$	0.05	N/A
Lysine, 4TMS	16.88	732	5.1	$C_8H_{18}NSi$	0.19	0.05

Identification Phase

The most significant changes were identified against existing commercial libraries (NIST) and matched primarily to amino acids and compounds associated with decomposition (Table 4). A summary of the complete workflow based around univariate statistics, where peaks are deconvoluted, quantified, and identified, is shown in Figure 5. High scores (>700) were obtained for all the selected compounds and library matches were enhanced with the application of accurate mass to fragment matches (Table 3).

Conclusion

The Q Exactive GC system workflow described here runs from sample preparation, automated derivatization, GC separation and mass spectrometry detection, to data analysis and reporting of the results. This comprehensive workflow makes the Q Exactive GC system is a unique analytical tool that can be used for metabolomics analyses of both volatile and non-volatile compounds following derivatization.

Excellent chromatographic resolution and reproducible chromatographic separation together with fast data acquisition make the Q Exactive GC system an ideal platform for complex metabolomics analysis.

Routine ultra-high resolution and consistent sub-ppm, accurate-mass measurements offer reliable and selective analysis of a variety of metabolites present in the complex biological decomposition matrix.

The wide dynamic range allows for sensitive and consistent detection of the metabolites in the samples analyzed, without any compromise in mass accuracy, while providing accurate, relative quantitation of detected metabolites.

The EI data obtained can be used for tentative compound identification against existing commercial libraries, enabling researchers to make value judgments about the results, while the accurate mass available allows compounds of interest to be further confirmed by fragmentation analysis (e.g. Mass Frontier) or pure standards. In this case, time-dependant evolution of amino acid signals provides the potential for a facile biochemical forensic assay for post mortem interval.

Acknowledgements

We would like to acknowledge Dr. Mark McLaughlin for providing us with the rats used in the study.

References

1. ProteoWizard Page. http://proteowizard.sourceforge.net/ (accessed Apr. 22, 2015).
2. C.A. Smith, E.J. Want, G.C. Tong, R. Abagyan, and G. Siuzdak. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78 (3)*, 779–787.
3. Richard A. Scheltema, Andris Jankevics, Ritsert C. Jansen, Morris A. Swertz, and Rainer Breitling. PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis. *Anal. Chem.* **2011** *83 (7)*, 2786–2793.
4. Creek, D.J., Jankevics, A., Burgess, K.E.V., Breitling, R., and Barrett, M.P. IDEOM: an excel interface for analysis of LC-MS based metabolomics data. *Bioinformatics*. **2012**. *28 (7)*, 1048–1049.
5. SIMCA Page, Umetrics Website. http://www.umetrics.com/products/simca (accessed Apr. 22, 2015).

www.thermoscientific.com/QExactiveGC

Thermo Fisher Scientific, Sunnyvale, CA USA is ISO 9001 Certified.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Brazil +55 11 3731 5140
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 6 6885 1213
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
Taiwan +886 2 8751 6655
UK/Ireland +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand

AN10457-EN 0415S