LCGC®

FEBRUARY 2023

The Importance of GC/MS for Your Metabolomics Toolbox

- GC-TOFMS and GCxGC-TOFMS Metabolomics
- LC-MS and GC-MS Metabolomics Workflows
- Microbial Metabolomics Based on GC Analysis
- Metabolomics Using Pegasus BT GC-TOFMS

This custom ebook is sponsored by LECO and presented in partnership with *LCGC*.

LECO
EMPOWERING RESULTS

High Performance GC- and GC×GC-TOFMS Metabolomics-Based Approach for Discovery of Potential Biomarkers in Plasma

By David E. Alonso, Tomas Kovalczuk, Joe Binkley, Cristina DiPoto, and Habtom Ressom

GC-TOFMS and GCxGC-TOFMS techniques can be used to identify potential candidates for HCC biomarkers, such as mannitol, citric acid and more, that can be used to intervene and more effectively manage patients earlier in the course of disease.

INTRODUCTION

Hepatocellular carcinoma (HCC) is the leading cause of cancer-related death worldwide, with elevated death rates in the US and Asia, especially for people with co-morbidities such as hepatitis B and C, alcoholism, diabetes, obesity, and non-alcoholic fatty liver disease. There is a critical need for the discovery of early-stage HCC biomarkers that can trigger a rapid intervention and more effective medical treatment. This would help not only HCC, but practically all cancer cases. It was with this goal in mind that LECO Corporation collaborated with Georgetown University for a study to implement an untargeted analytical methodology for annotation of candidate cancer biomarkers in human plasma using GC-TOFMS and GCxGC-TOFMS techniques.

A metabolomics approach is ideal for the investigation of cancer physiology and discovery of biomarkers for disease diagnosis, but it is not without its challenges. A large number of metabolites can have very different concentrations in any given sample. A sample may contain very highly-concentrated and trace analytes at the same time, with a wide range of chemical and physical properties. This isn't a challenge unique to metabolomics, though, as food analysis, environmental, and pesticide analysis all deal with the same struggles.

The analytical strategy began with sample derivatization before analysis using GC-TOFMS and data processing with the ChromaTOF® brand software. The data was then loaded into the online XCMS tool where a principle component analysis (PCA) identified about 100 analytes with peaks of statistical significance, which were then selected as quantification targets for analysis using GCxGC-TOFMS. This comprehensive technique provided better sensitivity and spectral similarities over the 1D measurements. The GCxGC peak tables generated were processed using Metaboanalyst's advanced statistical tools, like PLSDA, random forests, and more, which allowed the team to begin sorting out the potential biomarkers.

RESOURCE PAGE

New Integrated Statistical Data Processing with ChromaTOF®

PART 1: METABOLOMICS STUDY USING GC-TOFMS

A plasma (100µl) metabolite extraction of 64 patient samples was carried out using a 1:1:1 acetonitrile, isopropanol, and water mixture. The extracts were each vortexed, centrifuged, and the supernatant transferred to 2 mL GC vials. The solvents were removed by placing the vials in a Speed Vac, followed by lyophilization for 20 minutes (Pressure < 0.1 mbar, Temperature -50 °C). The resulting residue was derivatized in two steps using methoxylamine hydrochloride in pyridine (60 °C, 1 hour), and followed by treatment with MSTFA (60 °C, 1 hour). The derivatized samples were then transferred to GC vials immediately prior to analysis

From the derivatized final extract, one microliter was injected into the GC and GCxGC TOFMS systems with a split ratio of 1 to 20. For the GCxGC TOF-MS analysis, a 200 spectra per second acquisition rate was used, since the GCxGC peaks are so narrow, about 100 milliseconds at the baseline. For proper sampling of such narrow peaks, high acquisition speeds need to be used.

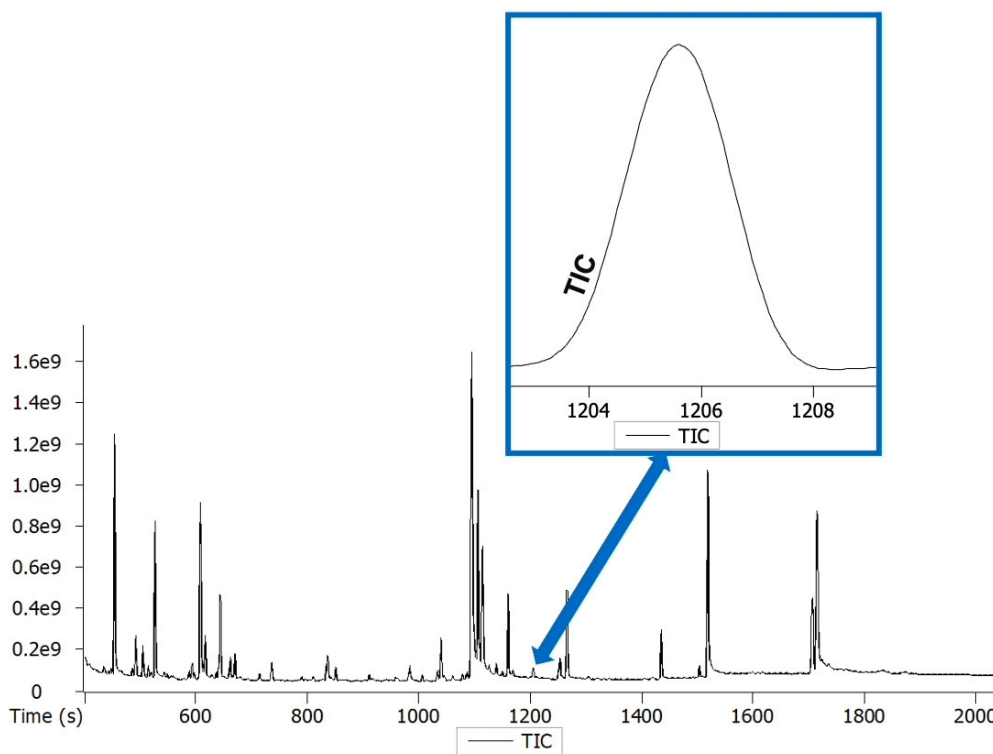
The acquired data were processed in *ChromaTOF* using the standard LECO peak find and non-targeted spectral deconvolution algorithms. The peak find is easy to use in *ChromaTOF*: set the minimum desired signal to noise ratio and from there, the software processes the data with no further user input needed. The resulting found peaks were then searched against the NIST 17 and Wiley 11 mass spectral libraries and further filtered

using a retention index criteria. For this study, a tolerance of 30 retention index units was used. Employing mass delta or mass accuracy measurements can narrow the selection of library hits even further. The libraries used contained the molecular formula from which the molecular ion can be calculated, and the comparison between the measured results and the technically expected results from the library can also help the identifications of the correct analytes, especially when using high-resolution measurements.

Data processing revealed many different metabolite classes such as acids, diacids, amino acids, sugars, fatty acids, monoacylglycerides, sterols, and more. There is a forest of peaks among the chromatogram. To make it simpler, let's focus on just one apparent peak. From the deconvoluted spectra, which is a spectrum clean of any interferences like solvent, column bleed, coeluted components, etc., the peak true can be compared with the library hit. At 89% similarity, this analyte was identified as myo-inositol.

On the zoomed-in view of this peak (FIGURE 1), the total ion current (TIC) shows what looks like a single peak, so a single peak might be expected. However, the *ChromaTOF*

FIGURE 1: Myo-inositol total ion current.

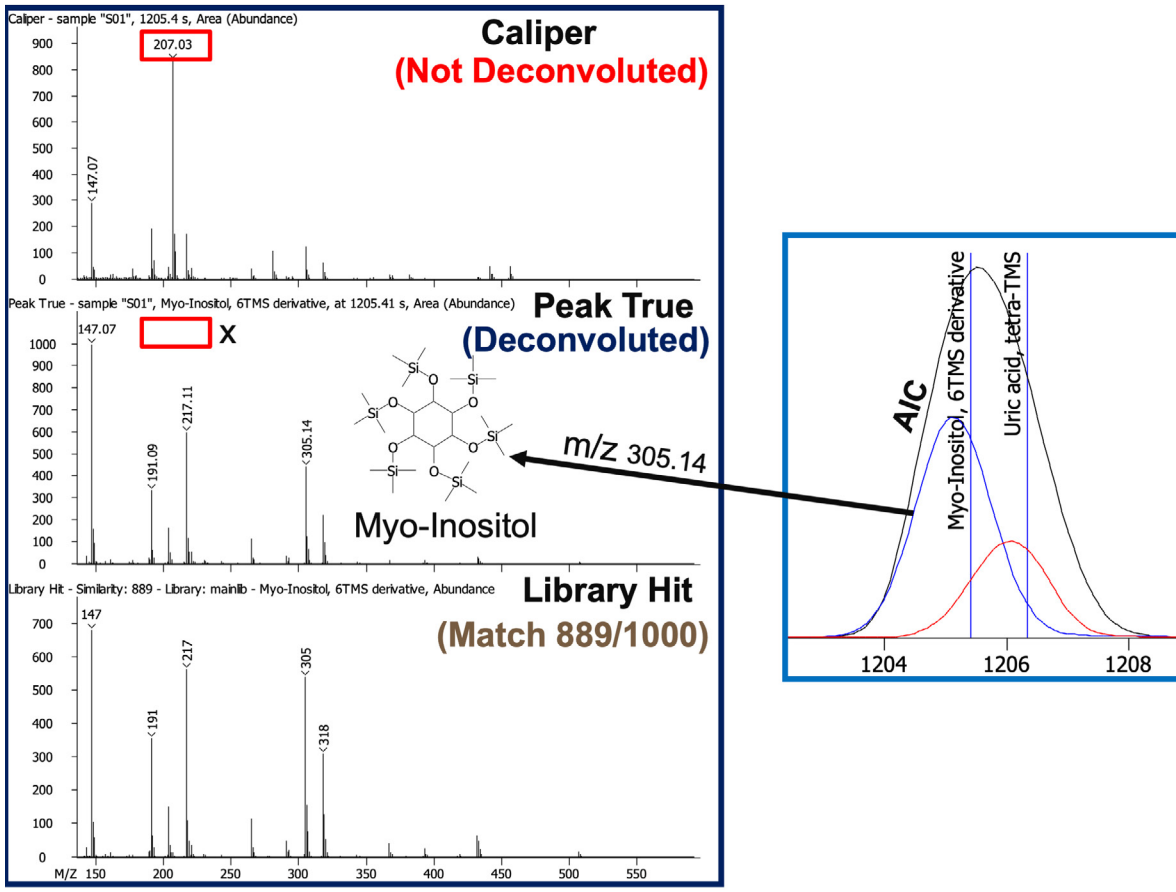


software actually identified two peaks beside each other.

Comparing the undeconvoluted spectra with the deconvoluted spectra (FIGURE 2), the mass at 207 was actually part of a siloxane interference, likely from a vial cap septum. Once that and other background ions were completely removed from the spectra, the cleaner deconvoluted spectra was a much closer match to the library hit. This was just one example of how deconvolution allowed for much better library similarity scores.

The identification of the smaller, co-eluting peak was also made possible by the deconvolution process which cleaned up the spectra, resulting in a similarity of 70% with the library hit for uric acid (FIGURE 3). As discussed earlier, mass accuracy or mass delta could be used to find the difference between

FIGURE 2: Comparison of undeconvoluted spectra and deconvoluted spectra.

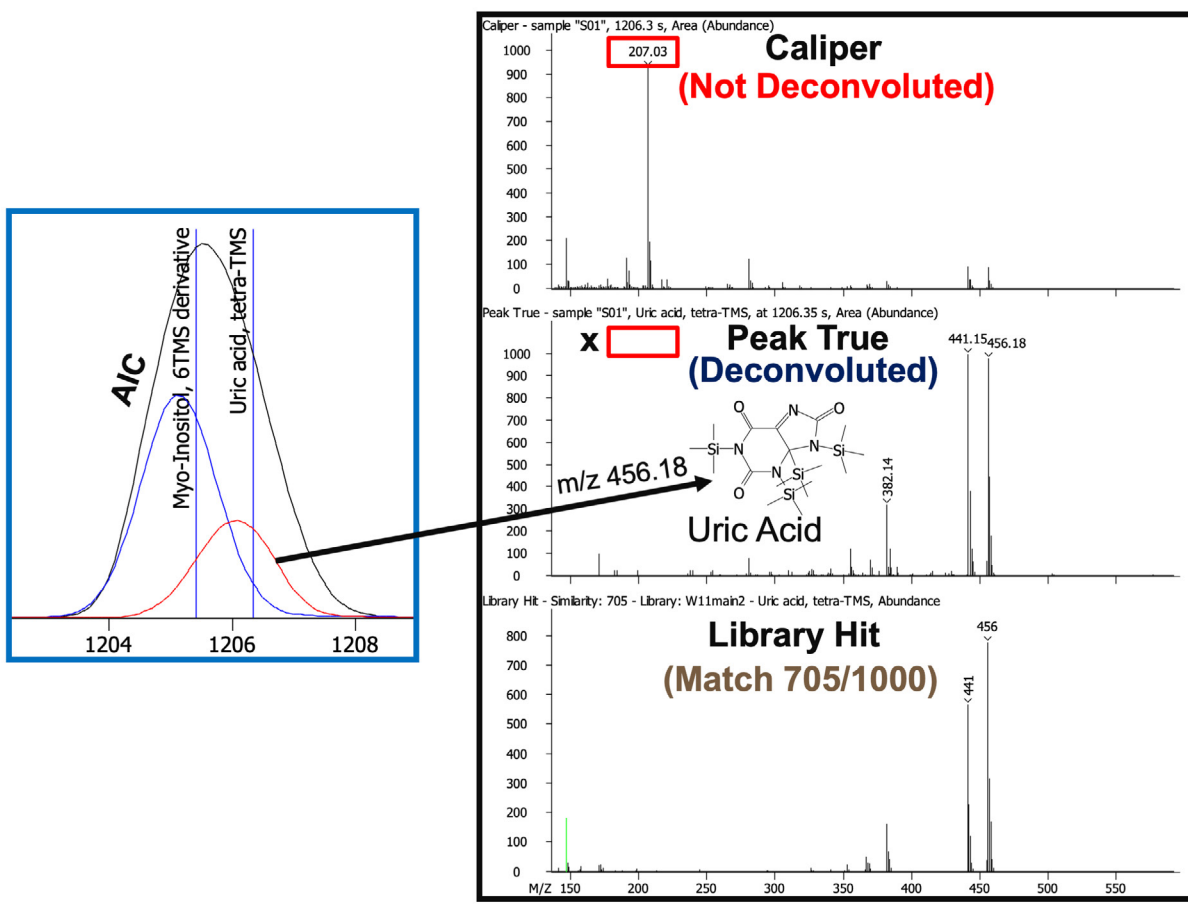


the calculated and observed mass. For this uric acid hit, the mass delta was 0.00 delta.

Other representative compounds had similarly excellent library hits. For some examples, l-leucine had a similarity score of 92%, while myristic acid had a similarity score of 86% and a mass delta of 0.00 again, and monopalmitin had a similarity score of 91%. In fact, out of about 100 selected peaks, the average similarity score was an incredibly high 86%.

Data alignment and advanced statistical filtering on this analysis was done using the XCMS online tool (xcmsonline.scripps.edu). From the significant analytes, about 100 provided by the PCA plots were then targetly searched in the GCxGC chromatograms and further processed. For example, mannitol had

FIGURE 3: Deconvolution resulting in identification of uric acid.

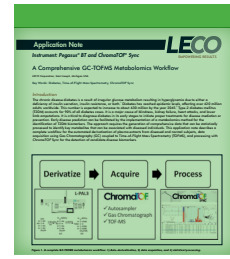


a p-value of 0.0025 with a 93% similarity match, while citric acid had a p-value of 0.028 and an 87% similarity match.

However, a closer look to the spectrum revealed that the molecular ion for citric acid was apparently missing (FIGURE 4). Analysis on LECO's high resolution MS system, the Pegasus® HRT, allowed for an additional analysis by a softer ionization technique, chemical ionization, which could reveal the hidden molecular ion. This was confirmed with a nice mass accuracy and corresponded with the citric acid.

PART 1: METABOLOMICS STUDY USING GCxGC-TOFMS

The next stage of the study was to run the samples in GCxGC-TOFMS mode. Using two columns with different separation mechanisms, the first nonpolar and the second more polar, connected through a dual stage quad jet modulator, two-dimensional contour plots were obtained.

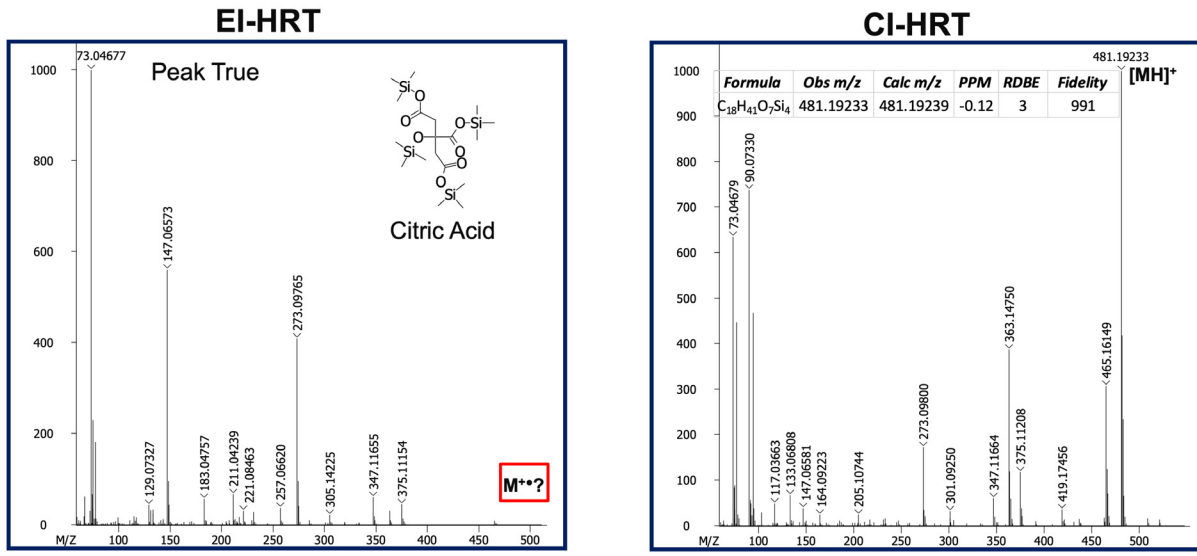


APP NOTE

A Comprehensive GC-TOFMS Metabolomics Workflow

As demonstrated in FIGURE 5, the GCxGC technique offered better chromatographic resolution due to the separation of the particular peak from interferences like matrix chemical noise and other target analytes. Two analytes that may elute in

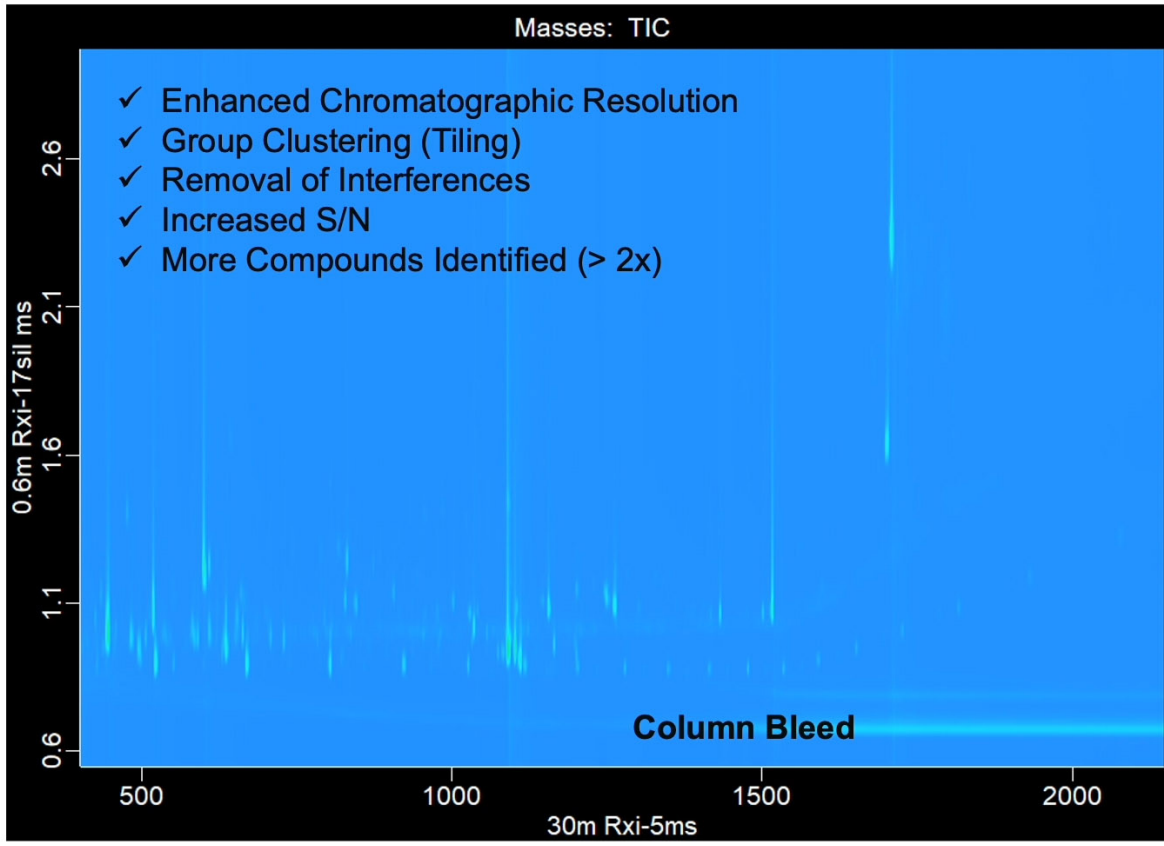
FIGURE 4: CI spectrum reveals a missing molecular ion for citric acid.



the same retention time on the first GC column can be clearly separated on the secondary column thanks to the different polarity. GCxGC chromatograms also show group clustering, or tiling, as groups of the same or similar class of compounds elute in bands. In addition, by narrowing the zones in the thermal modulator and therefore the separation, there is greater spectral clarity. All of this combined is why GCxGC was able to identify about twice as many analytes as the 1D analysis.

Looking at myo-inositol and uric acid again (FIGURE 6), it is clear to see how these two compounds, partially coeluted in the 1D analysis, are fully separated in the 2D analysis due to their different polarities. In addition, components that weren't even found in the GC-TOFMS analysis, such as naproxen, were able to be clearly identified in GCxGC with a similarity of 83%. This improvement of the spectral quality also meant that the similarity scores of known compounds, such as for myo-inositol and

FIGURE 5: Advantages of GCxGC anaylsis.



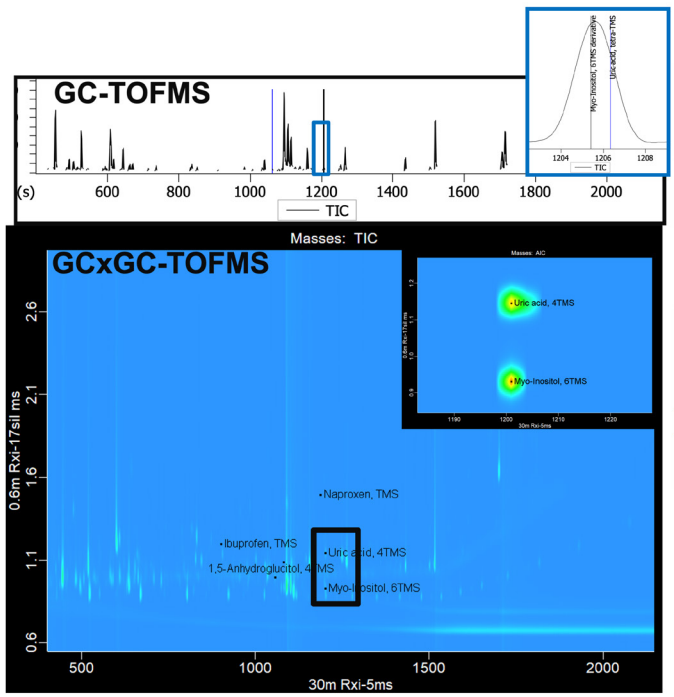
uric acid, went up. GCxGC enabled analyte identification to change from unknown to known and improved the confidence in already-identified analytes.

The statistically significant peaks were searched in GCxGC data using a target analyte finding approach. This strategy is simulating a single ion monitoring mode known from scanning instruments. However, in SIM mode, only the sought-after ions are acquired. With the GCxGC-TOFMS data, all of the ions have already been acquired, and the software simply has to search for the targeted masses. As a result, significantly faster data processing times are achieved. This targeted list and all of the locations of the analytes on the contour plot was the result (FIGURE 7).

CONCLUSION

In conclusion, all of the technical features from these instruments helped to obtain great chromatographic and mass spectrometric data that could be further processed. Excellent features in particular included: great spectral similarities achieved by the deconvolution on 1D and GCxGC, great mass accuracy, the ability to select the right library hit together with the retention index filtering, and the improvement in

FIGURE 6: Comparison of myo-inositol and uric acid with GCxGC.



Name	R.T.(s)	Peak S/N	Similarity
Ibuprofen, TMS	Not Detected		
1,5-Anhydroglucitol, 4TMS	1061.64	2070	871
4-Hydroxyphenyllactic acid, 3TMS	Not Detected		
Naproxen, TMS	Not Detected		
Myo-Inositol, 6TMS	1205.41	2436	889
Uric acid, 4TMS	1206.35	375	705

Name	R.T. (s)	Peak S/N	Similarity
Ibuprofen, TMS	901 s, 1.199 s	137	828
1,5-Anhydroglucitol, 4TMS	1057 s, 1.000 s	1830	926
4-Hydroxyphenyllactic acid, 3TMS	1081 s, 1.090 s	873	874
Naproxen, TMS	1186 s, 1.495 s	97	828
Myo-Inositol, 6TMS	1201 s, 0.931 s	4642	951
Uric acid, 4TMS	1201 s, 1.147 s	4864	863

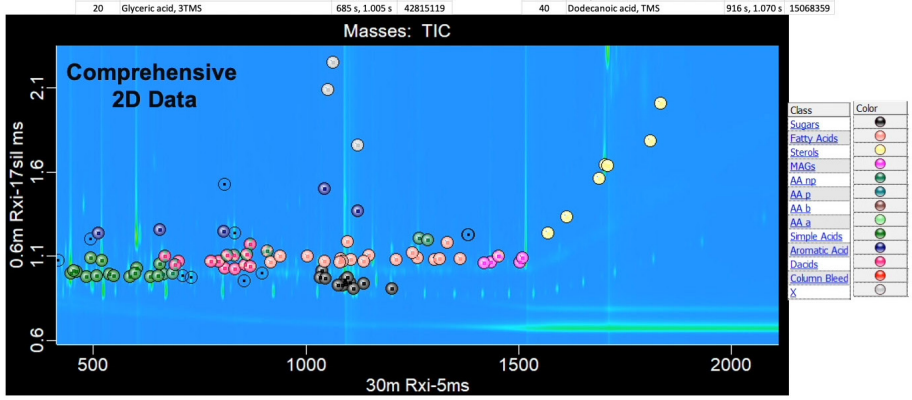
Unknowns → Knowns
More Compounds Identified

FIGURE 7: TAF results – GCxGC-TOFMS data.

Peak #	Name	R.T. (s)	Area
1	Lactic Acid, 2TMS	448 s, 1.005 s	1530743063
2	Hexanoic acid, TMS	454 s, 1.020 s	13162262
3	Glycolic acid, 2TMS	460 s, 1.015 s	80842344
4	L-Alanine, 2TMS	484 s, 0.985 s	482966060
5	Butanoic acid, 2-(methoxyimino)-3-methyl-, TMS	493 s, 1.095 s	13622033
6	2-Hydroxybutyric acid, 2TMS	508 s, 0.990 s	180287953
7	Levulinic acid, TMS	511 s, 1.240 s	428778
8	Oxalic acid, 2TMS	520 s, 1.080 s	784133934
9	3-Hydroxybutyric acid, 2TMS	538 s, 1.000 s	88630520
10	2-Aminobutanoic acid, 2TMS	547 s, 0.990 s	79490922
11	L-Valine, 2TMS	586 s, 0.985 s	199449790
12	2-Ethyl-3-hydroxypropionic acid, 2TMS	598 s, 1.005 s	300073
13	4-Hydroxybutanoic acid, 2TMS	601 s, 1.035 s	8510894
14	L-Leucine, 2TMS	634 s, 0.985 s	235656845
15	L-Isoleucine, 2TMS	652 s, 0.990 s	83917335
16	L-Proline, 2TMS	655 s, 1.060 s	328573200
17	Benzeneacetic acid, TMS	655 s, 1.265 s	537811
18	Glycine, 3TMS	664 s, 1.000 s	286915045
19	Succinic acid, 2TMS	667 s, 1.105 s	31819297
20	Glyceric acid, 3TMS	685 s, 1.005 s	42815119

Peak #	Name	R.T. (s)	Area
21	Fumaric acid, bis-TMS	691 s, 1.055 s	8891312
22	Nonanoic acid, TMS	700 s, 1.075 s	19826030
23	Serine, 3TMS	709 s, 0.990 s	109659036
24	L-Threonine, 3TMS	730 s, 0.980 s	158957563
25	Decanoic acid, TMS	775 s, 1.075 s	3660100
26	2-Aminomalonic acid, 3TMS	793 s, 1.075 s	5616482
27	benzoic acid, 3,5-dimethyl-, TMS	805 s, 1.255 s	6077820
28	Malic acid, 3TMS	808 s, 1.035 s	13730253
29	Adipic acid, 2TMS	814 s, 1.110 s	2264692
30	L-Methionine, 2TMS	829 s, 1.110 s	116146521
31	L-Aspartic acid, 3TMS	832 s, 1.030 s	3356604
32	L-5-Oxoproline, , 2TMS	832 s, 1.245 s	481524860
33	Erythronic acid, 4TMS	853 s, 0.960 s	1000486
34	L-Cysteine, 3TMS	856 s, 1.055 s	12124921
35	Creatinine enol, 3TMS	859 s, 1.115 s	9280518
36	Pentanedioic acid, 2-hydroxy, 3TMS	868 s, 1.045 s	940599
37	Pentanedioic acid, 2-oxo,2TMS MOX	868 s, 1.175 s	8610200
38	L-Ornithine, 3TMS	895 s, 1.005 s	10006017
39	Phenylalanine, 2TMS	907 s, 1.135 s	108523725
40	Dodecanoic acid, TMS	916 s, 1.070 s	15068359

Peak #	Name	R.T. (s)	Area
41	Asparagine, 3TMS	937 s, 1.105 s	12024101
42	L-Glutamine, 3TMS	1003 s, 1.105 s	103589292
43	L-Ornithine, 4TMS	1033 s, 0.980 s	55448880
44	Citric acid, 4TMS	1036 s, 1.020 s	221423396
45	Myristic acid, TMS	1042 s, 1.075 s	23015134
46	Hippuric acid, TMS	1042 s, 1.505 s	882709
47	D-Fructose 1, 5TMS MOX	1045 s, 0.975 s	502297
48	Caffeine	1051 s, 2.090 s	12225781
49	Theobromine	1063 s, 2.255 s	441308
50	D-Fructose 2, 5TMS MOX	1075 s, 0.935 s	116178769
51	Pentadecanoic acid, TMS	1078 s, 1.070 s	379262
52	4-Hydroxyphenyllactic acid, 3TMS	1081 s, 1.090 s	2936130
53	D-Galactose (E), 5TMS MOX	1084 s, 0.935 s	39729792
54	Pentadecanoic acid, TMS	1084 s, 1.075 s	758140
55	D-Galactose (Z), 5TMS MOX	1090 s, 0.955 s	194365532
56	Sedoheptulose, 6TMS MOX	1090 s, 0.965 s	207254222
57	D-Glucose 1, 5TMS MOX	1093 s, 0.940 s	495733088
58	L-Lysine, 4TMS	1096 s, 0.980 s	64860588
59	Histidine, tri-TMS	1096 s, 1.190 s	13386968
60	D-Glucose 2, 5TMS MOX	1102 s, 0.950 s	258338306
61	L-Tyrosine, 3TMS	1105 s, 1.085 s	173794020
62	D-Mannitol, 6TMS	1111 s, 0.915 s	386279467
63	3-Indoleacetic acid, 2TMS	1120 s, 1.375 s	512142
64	Theobromine, TMS	1120 s, 1.765 s	1647960
65	Glucopyranose, 5TMS	1135 s, 0.945 s	27670238
66	Palmitic Acid, TMS	1135 s, 1.075 s	1217644
67	Palmitelaidic acid, TMS	1147 s, 1.110 s	17556506
68	Myo-Inositol, 6TMS	1201 s, 0.915 s	112559586
69	Margarate, TMS	1210 s, 1.085 s	5001025
70	Oleic Acid, (Z)-, TMS	1249 s, 1.125 s	116038276
71	Stearic acid, TMS	1261 s, 1.095 s	446806072
72	Tryptophan, 3TMS	1264 s, 1.210 s	24687640
73	Lanthionine, 4TMS	1285 s, 1.200 s	2553020
74	L-Cystine, 4TMS	1300 s, 1.085 s	20578488
75	Nonadecanoic acid, TMS	1312 s, 1.090 s	511285
76	Arachidonic acid, TMS	1330 s, 1.185 s	6171938
77	Arachidic acid, TMS	1360 s, 1.090 s	10312922
78	Uridine, 3TMS	1378 s, 1.230 s	1786668
79	2-Palmitoylglycerol, 2TMS	1417 s, 1.065 s	33950465
80	1-Monopalmitin, 2TMS	1432 s, 1.070 s	155285362
81	Behenic acid, TMS	1450 s, 1.105 s	1364489
82	2-Monostearin, 2TMS	1501 s, 1.070 s	48790109
83	1-Monooleoylglycerol, 2TMS	1507 s, 1.095 s	6823066
84	δ-Tocopherol, TMS	1567 s, 1.245 s	396733
85	γ-Tocopherol, TMS	1612 s, 1.340 s	1084349
86	α-Tocopherol, TMS	1687 s, 1.565 s	5280690
87	Cholesterol, TMS	1702 s, 1.645 s	240037185
88	Cholestan-3-ol, TMS	1708 s, 1.640 s	1852286
89	Deoxycholic Acid, 2TMS	1807 s, 1.790 s	5444685
90	β-Sitosterol, TMS	1831 s, 2.010 s	3138035



quality of generated spectra between 1D and 2D, which helps to improve the number of identified peaks. The initial non-targeted approach of the study was followed by a targeted, relative quantitation approach using target analyte finding which was performed quickly despite the large datasets. All of this later resulted in identification of potential candidates for HCC biomarkers such as mannitol, citric acid and more.

View the full webinar presentation here.

David E. Alonso
LECO Corporation

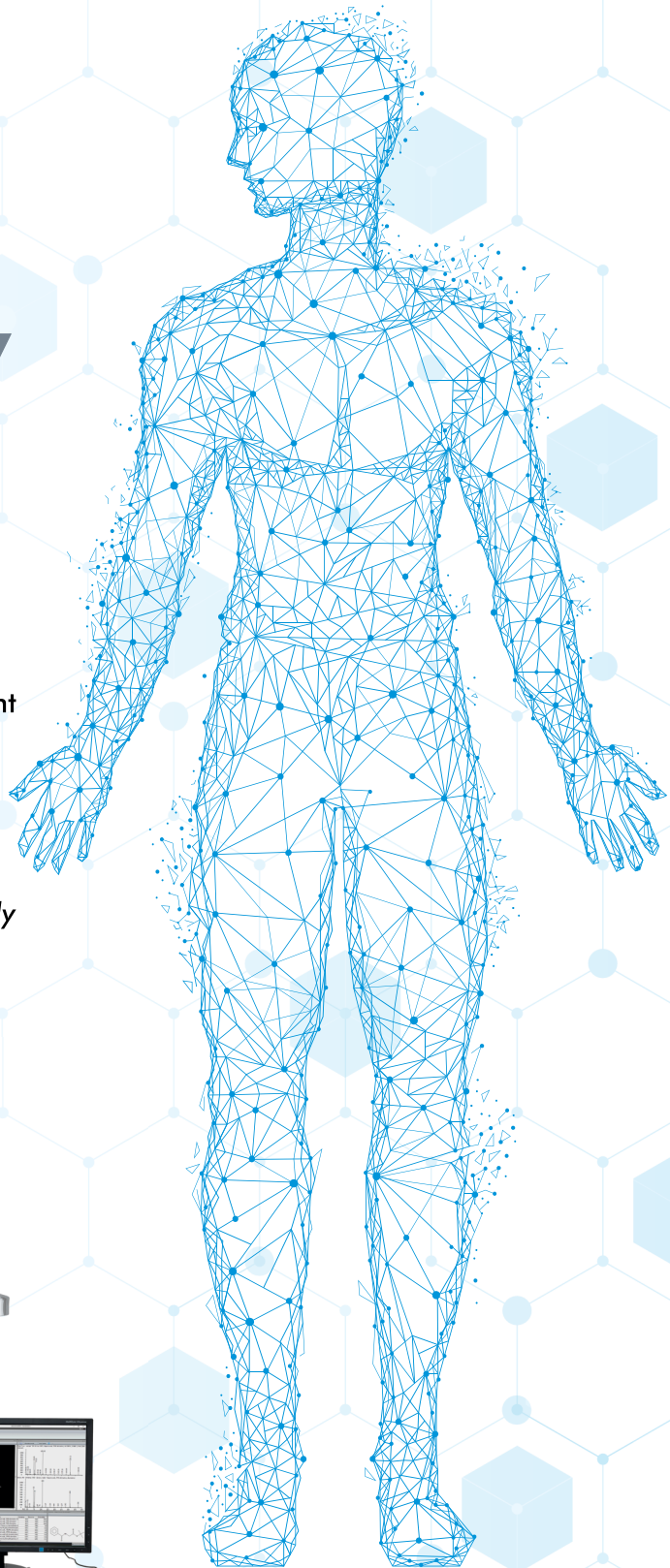
Tomas Kovalczuk
LECO Corporation

Joe Binkley
LECO Corporation

Cristina DiPoto
Georgetown University

Habtom Ressom
Georgetown University

Studying the Small Has Never Been a Bigger Priority



The diversity of the metabolome requires a diversity of analytical tools to understand it. Discover how to complete your instrument suite with GC-TOFMS. LECO's all-in-one metabolomics solution takes your samples from start to finish.

- Automated, just-in-time derivatization
- Specialized metabolite library for confident identification
- Unparalleled speed, sensitivity, and ruggedness
- GC and GCxGC options

No matrix is too complex when you're properly equipped. Turn your data into discovery with LECO's turn-key metabolomics solution.

*View our **Metabolomics Analysis Resources***

PEGASUS® BT 4D

Phone: 1-800-292-6141 | info@leco.com
www.leco.com | © 2023 LECO Corporation

LECO
EMPOWERING RESULTS

Integrating Best Practices into LC-MS and GC-MS Metabolomics Workflows

Experts Warwick (Rick) Dunn, BSc, PhD, FHEA, FRSC; Jane E. Hill, PhD; and Oliver Fiehn, PhD, comment on roles and opportunities using LC-MS and GC-MS for the study of metabolomics.

The study of metabolomics can be a powerful tool to researchers, as it allows for better characterization of chronic diseases and the discovery of important biomarkers that can be used to detect and, in some instances, improve understanding of pathophysiological processes related to diseases and their treatments.

Understanding how to properly assess metabolites in the laboratory is critical, as interest increases with the expansion of precision medicine for various fields. Experts who study metabolomics share their insight and provide recommendations for establishing LC-MS and GC-MS workflows into routine practice below.

Q. What are the key elements of metabolomics workflows?

Hill: It all starts with excellent study design, which starts with reliable and reproducible sample collection. If your research question has a clinical focus, such as a research question with a clinical focus around biomarkers for a particular disease, your study design will need to include study subject controls (i.e., subjects with the same symptoms as the disease, but who do not have the diagnosis you are interested in). These sorts of controls, as well as the usual environmental controls and analytical controls form the basis of a good metabolic workflow.

Thereafter, high QA/QC regarding sample analysis is essential. Following this, your data analytic pipeline will need to be thorough and should include checks for batch effects and biases in algorithms. An evaluation of statistics and biomarkers for biological relevance is also important. In the context of biomarkers of diseases, one must ensure the differences are measurable and place them within the context of a subject or a group of subjects' natural biological variation.

Dunn: The metabolomics workflow includes multiple elements, starting with robust experimental design, through sample collection and metabolite extraction, to data acquisition, raw data processing, and data analysis and biological interpretation. All elements of the metabolomics workflow are important and, when operated appropriately, provide confidence in results and biological conclusions. If one element is not applied appropriately, this can influence the quality and confidence of your results. Designing both the biological and metabolomics studies together and determining what will be applied for each element of the workflow should ensure data will be reported, which can generate robust biological conclusions.

RESOURCE

Metabolomics Research (Curated Compilation of Articles)

Q. What aspects of these workflows are important in terms of quality, best practices, and data confidence?

Fiehn: Study design is most important. Errors in the biomedical or biology study design cannot be healed by perfect chemistry or perfect data treatment. Unfortunately, too many studies still ask too easy questions. For example, some investigators only look at a knockout mutant and a comparator wildtype, but not other genotypes to be compared to. Or studies would use far too few samples to yield valid, statistical results. For example, even cell cultures have inherent variance in metabolome levels. If you only use n=3 cell cultures per genotype or per treatment, you will only find very large metabolic differences (like absence/presence of compounds), not smaller ones.

Dunn: In what is sometimes called "untargeted metabolomics," the quality of data should be assessed and reported for all studies. A number of QC samples can be applied in this process, including system suitability samples, pooled QC samples, process blanks, and standard reference materials. NIST SRM 1950 - Metabolites in Human Plasma is most commonly applied.

mQACC [Metabolomics Quality Assurance & Quality Control Consortium] is an international consortium working on the development of best practices for the application and reporting of data quality. I have been involved with mQACC from its inception, in 2017. I was also delighted to be involved in the development of early best practices for data quality and standardized protocols for sample extraction, data acquisition, and data quality reporting.

Confidence in the data reported is important for biological discoveries. Data quality and its reporting is one important aspect of this.

One other area is the confidence of reporting metabolite chemical structures in untargeted studies. The data collected is applied to identify a metabolite's structure (compared to targeted assays where data is applied to confirm a metabolite's structure). Using complementary data (e.g., RI and EI-MS) and matching to libraries constructed with data from chemical standards provides the highest confidence.

Another area is in quantification. Most studies use peak areas as surrogates for concentration. Reporting on how data is normalized is important. Even for targeted assays, the need to report which internal standards were used and which chemical standard was applied for quantification, as well as whether the calibration was a single-point or multiple-point calibration is important.

Q. Can you tell us more about the roles of liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS) in metabolomics' workflows?

Hill: Both are incredibly important. I think LC-MS is ahead of GC-MS in many ways, based on the high throughput systems that you see in, for example, hospital clinical chemistry systems. However, there is immense potential in the use of GC-MS systems for metabolomics.

Dunn: The collection of metabolites in biological samples (the metabolome) are chemically complex. No single analytical method can detect/report all metabolites. In untargeted metabolomics that are discovery-based, where the goal is to detect as many metabolites as is possible, the use of complementary analytical platforms to provide as broad a coverage of the

metabolome is required. GC-MS and LC-MS are two complementary platforms that both provide detection of hundreds of metabolites at (sub)-micromolar concentrations in a reproducible and medium throughput style.

Fiehn: GC-MS is perfect for primary metabolites, including glycolysis, TCA cycle, pentose phosphate pathway, and isotope label tracing experiments. Obviously, GC-MS is also indispensable for profiling volatile organic compounds. LC-MS cannot do that. While GC-MS is limited to compounds that are thermostable, some thermolabile compounds can be reproducibly quantified by degradation products. Do not forget, LC-MS also suffers from heat. This challenge occurs, for example, during electrospray ionization, leading to in-source neutral losses. Electrospray is not as soft as some researchers may believe.

Yet, of course complex lipids such as phospholipids and triacylglycerides can only be analyzed by LC-MS. For targeted analyses, LC-triple quadrupole MS/MS adds more sensitivity and selectivity than GC-triple quadrupole MS/MS.

Q. How significant is the role of GC-MS in metabolomics? Should all lab workflows employ it alongside LC?

Fiehn: Interestingly, more than 30% of all metabolomics papers use GC-MS, a stable ratio over the past 10 years, whereas the use of NMR is steadily declining in favor of LC-MS. So, by and large, the community continues to see the benefit of this technology.

GC-MS is much easier to operate and maintain, and is generally more robust compared to high resolution LC-MS/MS. An advantage of GC-MS is electron ionization, meaning there is no need for "data dependent" MS/MS or sophisticated schemas to convince the instrument to perform fragmentations. Everything is fragmented, all the time, meaning peak deconvolution by proper software is much easier than in LC-MS.

GC-MS is also generally cheaper than LC-MS, with respect to operations and maintenance. GC-MS data are also easier to interpret, because the identified metabolites nicely map on classic metabolic pathways. So, yes, I have always wondered why more labs haven't employed GC-MS?

Hill: I am biased. I think GC-MS represents an untapped resource. It is not as easy to quantify in these systems, and untargeted analysis is probably more challenging. However, GC-MS has a lot of potential and I hope it ultimately becomes a regular part of metabolomic workflows.

Q. There has also been some interesting work using hydrophilic interaction liquid chromatography (HILIC). Why is HILIC useful in clinical metabolomics studies? Can GC-MS help with this workflow?

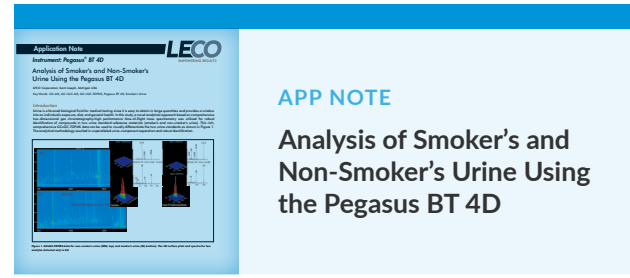
Dunn: The use of HILIC chromatography in metabolomic studies has increased over the last 5-10 years. However, multiple HILIC stationary phases can be applied, and, typically, two different assays are required for positively and negatively charged metabolites.

The mobile phase and associated modifiers need to be carefully developed to ensure good peak shapes and good sensitivity. And so, significant method development is required. The classes of metabolites detected by HILIC assays can generally also be detected by GC-MS, and in a single analytical run.

Fiehn: In my lab, we usually employ GC-TOF MS, reversed phase, and hydrophilic interaction LC-MS/MS for all samples, unless a project has a more limited metabolic scope or hypothesis. In HILIC-MS/MS, we get a much better overview on drug metabolism compared to GC-MS or reversed phase LC-MS/MS, in addition to short peptides that convey information about protein turnover, cachexia, and other phenotypes. The epimetabolome that investigates methylated, hydroxylated, acetylated, or otherwise modified canonical metabolites is also more comprehensive in HILIC-MS/MS. However, HILIC-MS/MS is far worse in separating sugar alcohols, sugar acids, and sugar phosphates. Hence, GC-MS is a necessary complement for these compound classes to HILIC-MS/MS. We published a couple of papers on clinical cases for these sugars, including in hepatic diseases and liver cancer.

Q. What is your view on deploying GC-MS metabolomics assays in routine settings? Will it be possible to use the technique in the future for mass screening or other purposes?

Hill: I use GCxGC-MS all the time. For me, this is routine practice. Regarding the mass use of GC-MS, sure, why not? GC-MS is already used for many things beyond metabolomics, so why not use the method for metabolomics screenings and assays?



Dunn: GC-MS allows for detection of a wide range of metabolites, and these systems have been proven to be robust in large-scale studies. Therefore, yes, GC-MS can be applied in routine settings, with appropriate QA and QC processes, to collect robust and comparable data over long periods of time. For clinical applications, it would be expected that targeted metabolite assays reporting accurate concentrations would be applied, and these are already present. The use of automation in sample preparation (e.g., chemical derivatization), data processing, and data interpretation are key components to include in the workflow to allow 24/7 operation in a clinical setting.

Fiehn: Most studies do not have more than 50-100 samples. However, indeed, in epidemiology cohort studies we routinely use 4000-14,000 samples, including in GC-TOF/MS. The consortium on type 1 diabetes in the young (TEDDY) recently identified interesting statistical results in such cohorts from our GC-TOF/MS data. We have also completed cohort studies

on type 2 diabetes in 4000 American Indians currently undergoing statistical tests, and on thousands of subjects in our Longevity consortium cohorts. So, yes, we have data that convincingly demonstrate that this technique can be used on a large-scale basis.

Q. What are your current research activities? What are the main analytical challenges in this research, how are you trying to overcome them, and what have been your findings so far?

Dunn: The Analytical and Clinical Metabolomics Group focus on two areas of research. The first focus is analytical chemistry development to enhance the application of metabolomics. Three significant areas are metabolite annotation, data quality, and (semi)-targeted assays. We are also focusing on human biofluid sampling away from the clinic (e.g., dried blood spots). The second focus is applying metabolomics in the study of human health, disease, and ageing. We work in several disease areas, including cancer, liver transplantation, steroid-related diseases, and diet. The research focus here is precision medicine, where metabolites can be applied to determine prognosis or identify responders/non-responders at the individual level and not at the population level.

Fiehn: Precision of quantitative data on a metabolome-wide scale in GC-TOF/MS was a bit harder than for HILIC- or reversed-phase LC-MS/MS. For LC, we had developed a normalization method called "systematic error removal by random forest" (SERRF). Surprisingly, for GC-TOF/MS data, this method did not work as well as we had thought. Here, we recently completed and validated a new algorithm using denoising autoencoders that brought the residual technical error in GC-MS evaluations to the same level as LC-MS/MS data.

Hill: We are currently focused on studying human and animal patients with acute or chronic infections as an effort to develop next-generation devices for breath and blood projects. Through the discovery and validation of certain biomarkers, we can better understand their impact on cellular activity and translate them to portable devices for patients.

> ***In terms of analytical challenges, we need to overcome batch effects between instruments and address alignment issues with hundreds to thousands of chromatograms.***

Q. What else is your group focusing on at the moment?

Dunn: Two research grants that are currently operating are focused on enhancing metabolite annotation in metabolomics data repositories and the re-use of data deposited by research groups globally. Surprisingly, the number of metabolites reported in deposited studies is small compared to the number of metabolites detected. These grants will increase the metabolite information in studies and make these studies more

desirable to be reused in the future (for example, in meta-analyses).

Hill: We are also focusing efforts on the biodiscovery of certain plant species. This includes studying cannabis.

Fiehn: We recently submitted a grant proposal on volatile organic compounds and brain receptors in honeybees to the HFSP program. GC-MS has many more applications than just blood, blood, blood!

Q. How do you think metabolomics analysis will evolve in the future?

Hill: Really, I think the question is, "What *won't* metabolomics do in the future?" I see a world that is multi 'omic, with metabolomics playing a critical role, probably in terms of better understanding the human condition as well as via diagnostics and other assays that employ metabolomics to help us manage human disease.

Fiehn: Metabolomics has matured. However, we need to be able to compare results better quantitatively across studies and between centers.

Ring trials and round-robin tests must validate our results to enable data meta-analysis and eventual adoption of techniques by oversight boards like the American Heart Association.

Dunn: The number of research groups applying metabolomics into practice is increasing rapidly, as is the number of published papers. This means there are research groups who have no metabolomics expertise and require training in the metabolomics workflow, analytical platform operations and maintenance, and in software for data processing and analysis. Dr Cate Winder and I direct the Liverpool Centre for Metabolomics Training, which provides face-to-face and online training courses for different aspects of the metabolomics workflow. This helps to develop the next generation of metabolomics experts.

Metabolomics is and will continue to evolve from untargeted studies (where the metabolites to be detected are not known before data acquisition) to semi-targeted studies (where the metabolite structure is known before data collection). With these advances, the process from data collection to biological interpretation is easier and somewhat automated. This still allows for discovery studies, but avoids some of the hurdles related to derivation of metabolite structures from metabolomics data. The next step will be semi-targeted assays, for which hundreds of metabolites are quantified via comparison to a calibration curve, similar to a clinical chemistry readout but for many more analytes.

Warwick (Rick) Dunn, BSc, PhD, FHEA, FRSC, holds a chair in Analytical and Clinical Metabolomics at the University of Liverpool. He obtained a lectureship in 2011 at the University of Manchester and moved to a lectureship at the University of Birmingham in 2013. His research is focused on two areas: The development of new analytical tools and methods to enhance data quality, efficiency of metabolite annotation, coverage of detectable metabolites, and sample collection strategies; and the application of untargeted and targeted metabolomics to the study of metabolism across the life course in humans, including pre-birth, ageing, endocrinology, inflammatory and immune diseases, and cancers, with a focus on precision medicine. His career goals are to make metabolomics a standard resource that can be applied in biological research, and to train the next generation of metabolomics researchers.

Jane E. Hill, PhD, is the Canada Research Chair of Breath Science and Technologies at the University of British Columbia. Her team specializes in the discovery and clinical validation of breath biomarkers for respiratory diseases, especially those with an infectious etiology. She leads the Human Breath Atlas initiative.

Oliver Fiehn, PhD, has pioneered developments and applications in metabolomics with over 420 publications to date, with an h-index of 112 and i10-index 367. In 2000 he began as group leader at the Max-Planck Institute for Molecular Plant Physiology in Potsdam, Germany. He joined University of California, Davis, as a professor in 2004. He oversees his research laboratory and the service core laboratory in metabolomics research. Since 2012, he has served as Director of the UC Davis West Coast Metabolomics Center, supervising 35 staff operating 18 mass spectrometers and coordinating a range of outreach activities, including a yearly symposium, monthly webinars, and a range of professional courses in metabolomics. Professor Fiehn's laboratory members develop and implement new informatics approaches and analytical chemistry methods for covering the metabolome. He applies metabolomics to metabolic questions in a range of human and animal models. He also studies fundamental biochemical questions from metabolite damage repair to the new concept of epimetabolites, the chemical transformation of primary metabolites that gain regulatory functions in cells. Professor Fiehn has received a range of awards, including the 2014 Molecular and Cellular Proteomics Lecture Award and the 2014 Metabolomics Society Lifetime Achievement Award. He served on the Board of Directors of the Metabolomics Society until 2015. In 2017, he co-founded the Metabolomics Association of North America (MANA), which he serves on.

The Clock is Ticking

Don't let uncertainty waste your time. Learn more about how LECO's turn-key solution to metabolomics works with your data to save you time.

- Full mass spectra every time, with every run
- Identification Grading System™ (IGS™) lets you quickly see, justify, have confidence in, and report on the chemicals in your sample
- Statistical analysis tools like ChromaTOF® Tile to highlight significant differences between your samples

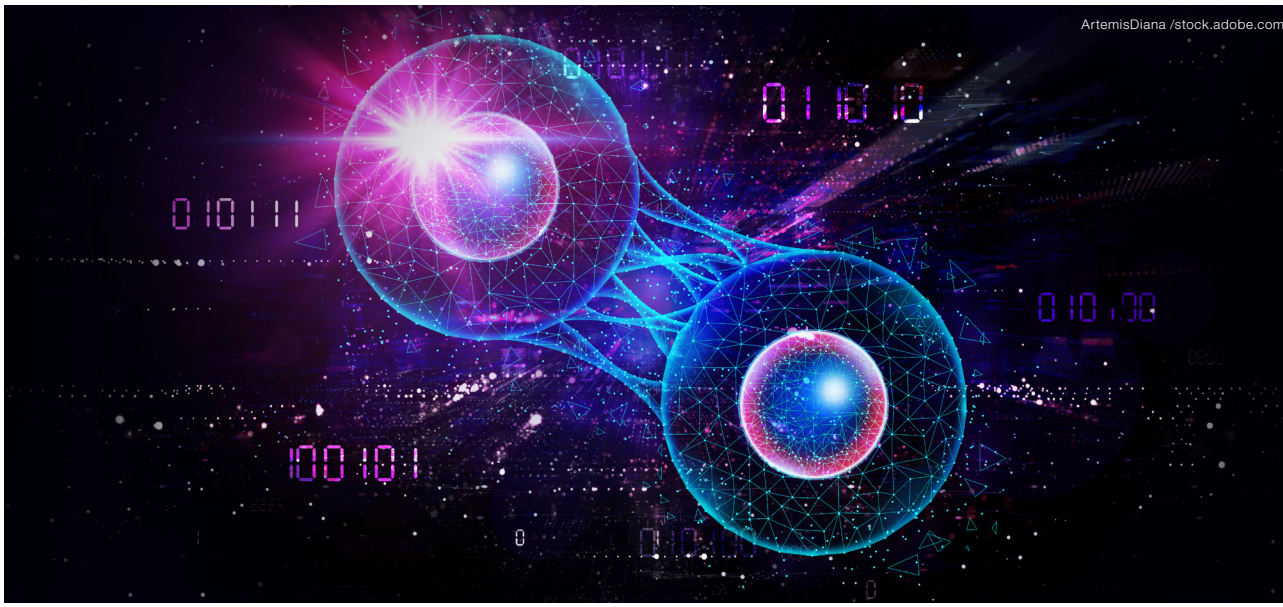
When you don't have to spend your time finding what's important, you can spend your time finding out why it is.

*View our **Metabolomics Analysis Resources***

PEGASUS® HRT+ 4D

Phone: 1-800-292-6141 | info@leco.com
www.leco.com |

LECO
EMPOWERING RESULTS



Microbial Metabolomics Based on GC Analysis as the New Avenue in Understanding the Complex Systems of Microorganisms: *Current Challenges on Sample Preparation*

By Sílvia M. Rocha

Microbial metabolomics includes a set of analytical tools that have been gaining enormous relevance because of the amount of data generated and the information that has been made available. Because of the wide coverage of microbial metabolomics, increased focus has emerged regarding sample preparation of microbial metabolomics before chromatographic analysis. One challenging issue on constructing microbial metabolomics workflows and several case-studies were used to illustrate the main challenges and new achievements in this field of study.

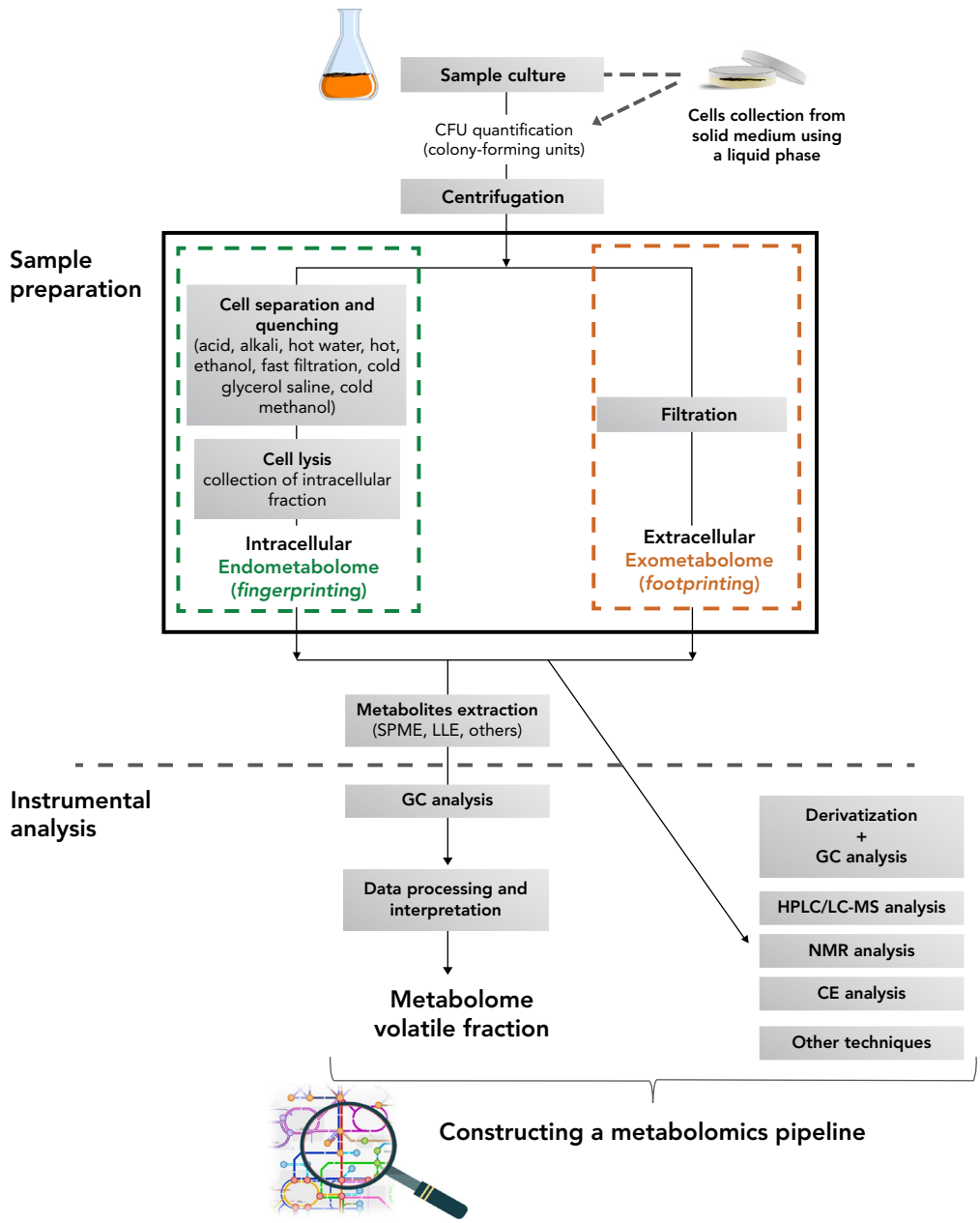
Microorganisms are found everywhere in the environment and play an important role in various processes. Pathogenic microorganisms can cause diseases in hosts during interactions or provoke damage and off-flavors in a wide variety of foods (1). The virulence factors aggravate the defense mechanisms of the host to establish infections, and the nature and type of virulence factors determine the pathogenicity of microorganisms. For instance, Staphylococcal

food poisoning is a disease that results in significant health and economic losses and is caused by *Staphylococcus aureus (S. aureus)* strains, which are able to produce enterotoxins. Recent studies revealed that *S. aureus* enterotoxic and non-enterotoxic strains may be distinguished based on their volatile metabolites' profiles (1). Otherwise, microorganisms can therefore be seen as suppliers of so-called ecosystem services, which are fundamental to the environment, human life, and industrial activities (2–5). By studying the complete set of metabolites within a microorganism and monitoring the global outcome of interactions between its development processes and the environment, it might be possible to provide a more accurate snapshot of the physiological state of the cell and the phenomena that modulate its metabolic pathways (6).

Microbial metabolomics is a comprehensive analysis of metabolites within a microorganism that may help to better understand complex biological systems and supports a new approach in microbial research. The microorganisms produce a wide range of metabolic products, which are the final products of biochemical processes and a result of environmental and genetic interactions, such as volatile metabolites, that can be used as unique volatile metabolic fingerprints of each species, and possibly of strains. Metabolomic profiling may provide additional information about the processes affected by specific microorganisms or microbiota alterations. Metabolomic profiling may also provide information about how the environment affects organisms that are susceptible to environmental changes and stress conditions. This data is particularly relevant because it takes into account the current climatic changes and the drought-related hydric stress (3). Also, because microbial communities represent huge complex systems, a global approach is necessary to understand the formation of consortia, communication between members, and functional interaction in a dynamic setting (6–9).

In recent years, there has been much progress in studying the metabolomics of several species of microorganisms, essentially taking advantage of using advanced equipment and highly efficient extraction techniques. To extract as much information as possible, a workflow should be defined to construct a microbial metabolome signature, as shown in FIGURE 1. A microbial metabolomics pipeline includes the sample preparation and the instrumental analysis steps, and a wide set of techniques that may be used to map the metabolome as completely as possible. For targeted and untargeted analysis of small volatile metabolites,

FIGURE 1: Schematics showing a workflow that may be used for the construction of a microbial metabolomics platform, including the sample preparation step and the instrumental analysis component, in which gas chromatography is highlighted for targeted and untargeted analysis of volatile metabolites. Also, data processing and interpretation should be considered to have a complete metabolomics pipeline.



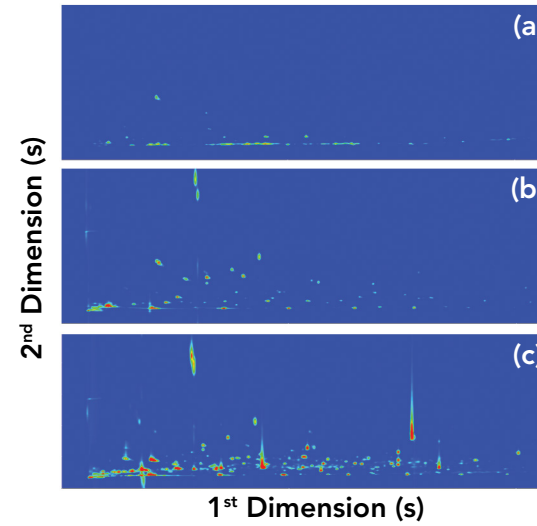
gas chromatography–mass spectrometry (GC–MS) has been the most popular technique. However, comprehensive two-dimensional gas chromatography (GCxGC) has emerged as a technique with high potential for constructing a more detailed picture of the microbial metabolome (1–5,7,10).

GC×GC advantages are based on the principle of the orthogonality mechanism that separates the constituents of the sample within a single analysis based on using two GC columns coated with different stationary phases. The interface samples small portions of the first dimension (^{1}D) eluate, in general, by cryofocusing, and re-injects them into the second column (^{2}D). Each ^{1}D peak is modulated several times, largely preserving the ^{1}D separation (11). Thus, compounds that are co-eluted from ^{1}D undergo additional separation on ^{2}D, and sensitivity and limits of detection are improved because they focus on the peak in the modulator and separate analytes from their chemical background. Also, the combination of the GC×GC instrument and a mass spectrometer with time-of-flight (ToF) analyzer allows the detection and quantification of analytes in picograms (pg), which is in line with the microbial metabolomics sensitivity requirements. The narrow peaks produced by GC×GC (peak width at half height of 0.1 s or less) require a detector with high data acquisition speed (hundred full-mass-range spectra per second), such as ToF-MS, thus providing sufficient data density. Moreover, ToF-MS allows the acquisition of full mass spectra at trace levels and mass spectral continuity, providing a reliable spectra deconvolution of overlapping peaks (10,11).

In constructing the microbial metabolome workflow, there are still several challenges to overcome, namely in the sample preparation stage. These challenges are related to a) microbial culturing in representative conditions; b) metabolic quenching approaches since strict control of the quenching procedure that arrests the cellular metabolism and enzymatic reactions should be done to reduce data variability; c) cell lysis approaches, depending on if the study is focused on the extracellular or intracellular metabolites, or both; and d) the extraction methodology that allows the determination of metabolites in representative conditions, and avoids the formation of artifacts or damage of metabolites since the amount of metabolites is very small. These challenges, alongside the technical difficulties to identify and quantify trace metabolites within complex matrices and the inherent problems related to data processing, are partially responsible for the paucity of information on the full volatile metabolome of common microbial pathogens or species with important biotechnological applications.

The definition of conditions for the culture of microorganisms is one of the first steps considered, where the selection of the medium is particularly important. Despite providing the appropriate nutrients into the culture medium, it is also necessary to attend the biophysical requirements during microorganism growth, such as oxygen levels, temperature, and the growth time. On the other hand, the volatile composition of the media must be minimal to avoid overlapping and interference with the metabolites produced by the microorganisms. FIGURE 2 illustrates the GC×GC total ion chromatogram contour plots of three culture media commonly used, which clearly shows the differences between the media regarding their volatile profiles, which is related to the composition of each medium as well as compounds that can be formed in the initial stage of sterilization. The volatile organic compounds (VOCs) screening medium exhibits a higher complexity of volatile composition compared with the others, which may be explained by the presence of the yeast extract (1% m/v). Thus, to confirm the origin of all detected analytes from microorganism cultures, excluding the medium effect, it is recommended to determine the amount of the analytes in the cultures and respective medium used as background control, analyzing both under the same conditions. To see if the VOCs do not exhibit significant differences between medium used as background control and microbes' cultures, statistical significance should be verified by variance analysis.

FIGURE 2: GC×GC total ion chromatogram contour plots of three culture media: (a) synthetic defined (SD) medium is based upon a yeast nitrogen base without amino acids supplemented with amino acids 0.67% m/v and glucose 2% m/v; (b) synthetic complete (SC) medium is based upon a yeast nitrogen base supplemented by a mixture of amino acids and vitamins; and (c) volatile organic compound (VOC) screening medium (yeast extract 1% m/v, diammonium phosphate ((NH4)2HPO4) 1% m/v, and glucose 2% m/v).



In general, the analytes not detected in the control medium or determined in the medium in levels at least three standard deviations lower than in the microbe's culture for similar conditions, may be attributed as having their origin associated with the microbe's metabolism (10). For instance, *S. aureus* strains were cultured in nonbuffered brain heart infusion (BHI) broth,

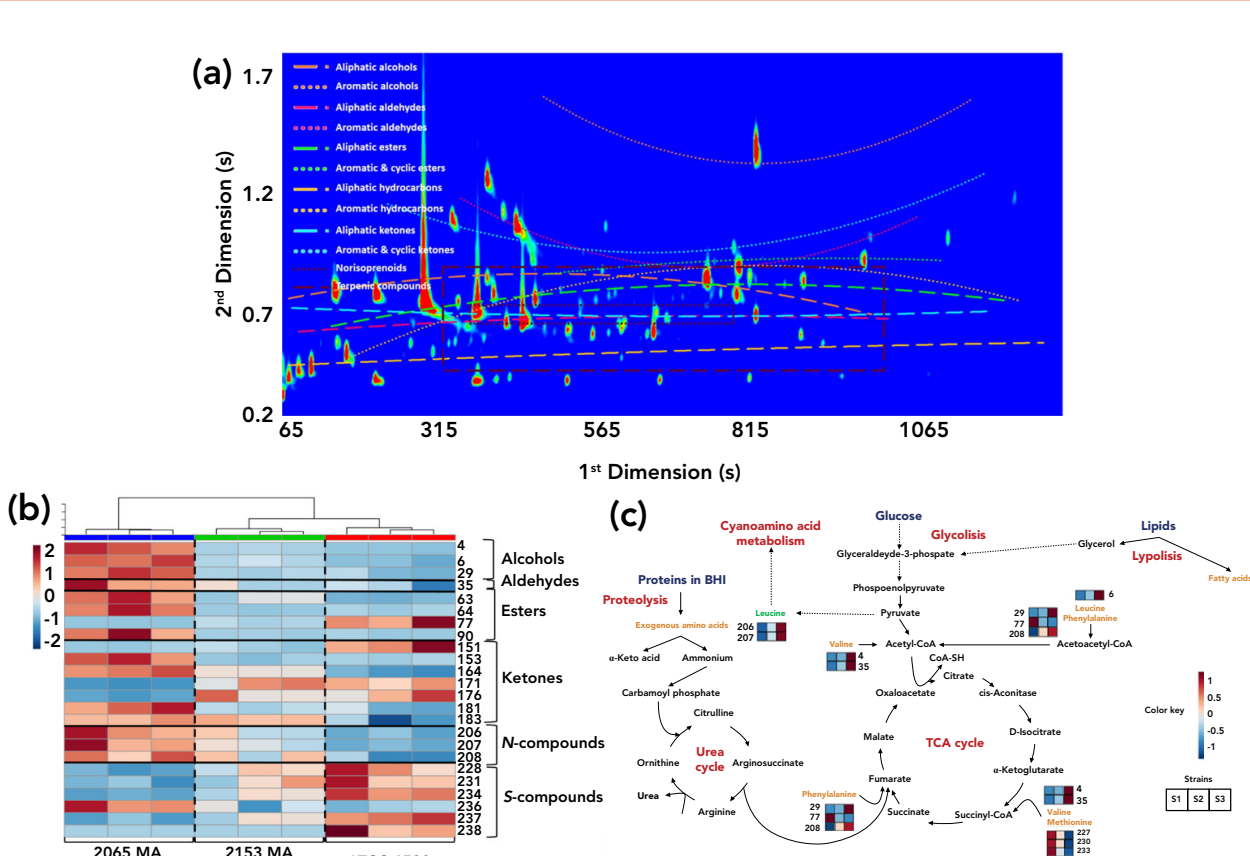
and the analysis of variance tests indicated that 75 VOCs (most of them belonging to the chemical families of pyrroles, pyridines, furan-like compounds, pyrazines, Strecker aldehydes, and thiazoles) should not be considered as a microbe's metabolites, and most of them are considered likely Maillard reaction products (1).

The methodology based on headspace solid-phase microextraction (HS-SPME)–GC×GC–ToF-MS provides novel data on the *S. aureus* volatile exometabolome (FIGURES 3A AND 3B), showing higher complexity (240 volatiles) than previously reported (1). For untargeted profiling, the divinylbenzene/carboxen/poly(dimethylsiloxane) (DVB/CAR/PDMS) sorbent coating was selected as it is recommended for the extraction of a wide range of chemical species. This sorbent coating is produced using three different polymers, which gives it a synergistic effect between adsorption and absorption. This mutually synergetic effect promotes a higher retention capacity and, consequently, a higher sensitivity, which is very convenient to in-depth analyze complex samples such as the metabolites released by microorganisms. The metabolites released from *S. aureus* strains were mainly by-products of branched-chain amino acids and methionine degradation, pyruvate metabolism, diacetyl pathway, oxidative stress, and carotenoid cleavage (FIGURE 3C). The HS-SPME/GC×GC-ToFMS tandem with clustering analysis unveiled that the metabolites released by the first two pathways were produced in higher contents by the enterotoxic strains, which distinguished strains of *S. aureus* by the number of produced enterotoxins (1). This information is especially important from the food safety point of view.

Concluding Remarks

Despite the enormous interest in knowing the metabolome of microorganisms, information about subject, even for the most common species, is still scarce or unknown. The advances observed in recent years have essentially taken advantage of using sophisticated equipment such as GC×GC–ToF-MS instrumentation. Sample preparation is a multistep phase that continues to present many challenges, but the use of SPME is a more environmentally friendly alternative because it does not use solvents and is a simple and user-friendly technique. In addition, SPME allows the mapping of metabolites present in headspace, which have a central role in several biological phenomena of communication between species of microorganisms and microorganism-plants. The challenges related to data processing and interpretation, especially the difficulty to implement a computerized combined analysis of instrumental signals and metabolic pathway data bases (when available), are also noteworthy. Currently, methodologies for in situ analysis, such as for human biofluids, food, or numerous biotechnology applications, represent huge challenges in the field of microbial metabolomics. In these cases, the simultaneous presence of several species of microorganisms raises additional questions

FIGURE 3: (a) GC×GC–ToF-MS total ion chromatogram contour plot of the *Staphylococcus aureus* ATCC 6538 culture headspace volatile components. Volatiles chemical families used for statistical analysis are represented by the lines and clusters. The increase in volatility (low $^{1}t_{R}$) is related to the decrease in the number of carbons through the first dimension. On the other hand, an increase in the $^{2}t_{R}$ correlates to an increase in polarity through the second dimension. (b) Hierarchical clustered heatmap visualization of the volatiles with variable importance in projection (VIP) values higher than 1.5 from the three strains of *Staphylococcus aureus* cultures headspace volatiles and organized by chemical families. The chromatographic area of each metabolite was normalized by CFU/mL followed by autoscaling. Each line corresponds to one metabolite and each column corresponds to each independent assay. (c) Metabolic pathways related with the VIP volatiles and their relative content in three strains of *Stapylococcus aureus* under study: S1 - ATCC 6538; S2 - 2153 MA; and S3 - 2065 MA. The relative content of the metabolite is illustrated on a red (high) to blue (low) chromatic scale. Adapted from reference (1).



that can only be overcome with the use of equipment and sample preparation strategies of high efficiency, sensitivity, or specificity, as those previously reported.

Acknowledgments

Special thanks goes to the FCT/MEC for their financial support of LAQV-REQUIMTE (UIDB/50006/2020) Research Unit through national funds and where applicable co-financed by the FEDER, within the PT2020 Partnership Agreement.

References

1. I. Baptista, M. Santos, A. Rudnitskaya, J.A. Saraiva, A. Almeida, and S.M. Rocha, *Int. J. Biochem. Cell Biol.* **108**, 40–50 (2019). doi:10.1016/j.biocel.2019.01.007
2. P. Cardoso, M. Santos, R. Freitas, S.M. Rocha, and E. Figueira, *Environ. Pollut.* **231**, 802–811 (2017). doi: 10.1016/j.envpol.2017.08.067
3. D.R. Matos, C.F. Sá, P.J. Cardoso, A. Pires, S.M. Rocha, and E. Figueira, *Ecotoxicol. Environ. Saf.* **186**, 109759 (2019). doi:10.1016/j.ecoenv.2019.109759
4. Z. Alves, A. Melo, A.R. Figueiredo, M.A. Coimbra, A.C. Gomes, and S.M. Rocha, *PLoS One* **10**(11), 1–16 (2015). doi: 10.1371/journal.pone.0143641
5. C. Martins, T. Brandão, A. Almeida, and S.M. Rocha, *J. Sep. Sci.* **40**, 2228–2237 (2017). doi:10.1002/jssc.201601296.
6. J. Tang, *Microb. Genom.* **12**, 391–403 (2011). doi: 10.2174/138920211797248619
7. C.P. Costa, A.R. Bezerra, A. Almeida, and S.M. Rocha, *Microorganisms* **8**, 1911 (2020). DOI:10.3390/microorganisms8121911.
8. D.S. Childers and D. Kadosh, *PLoS One* **10**, 1–15 (2015). doi: 10.1007/978-1-4939-8757-3_2
9. M. Martins, M. Henriques, J. Azeredo, S.M. Rocha, M.A. Coimbra, and R. Oliveira,*J. Basic Microbiol.* **50**, 1–9 (2010). doi: 10.1002/jobm.200900442
10. C.P. Costa, D.G. Silva, A. Rudnitskaya, A. Almeida, and S.M. Rocha, *Sci. Rep.* **6**(27441), 1–13 (2016). doi:10.1038/srep27441.
11. S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, and R.E. Synovec, *Anal. Chem.* **90**, 505–532 (2018). doi.org/10.1021/acs.analchem.7b04226

Sílvia M. Rocha is with the Chemistry Department and LAQV/REQUIMTE at the University of Aveiro, in Aveiro, Portugal. Direct correspondence to: smrocha@ua.pt.

We're Not the Only Ones Talking About It

Professor Antonio Randazzo
University of Naples

"Pegasus BT is the perfect choice for identifying well-known metabolites at low concentration level in our sample, and also discovering new compounds."

Alex Dickens
Turku Metabolomics Facility

"The StayClean™ ion source allows us to run hundreds of samples in a batch and be confident that the ion source will remain clean even with dirty samples such as plasma."

Dott. Marcello Manfredi
Unversita Degli Studi Del Piemonte Orientale

"A combination of very good sensitivity, quality to price ratio and high flexibility make the BT 4D system perfect for analysis of every sample."

Learn why these labs chose the Pegasus series and never looked back.

View our Metabolomics Analysis Resources

PEGASUS® BT 4D

LECO
EMPOWERING RESULTS

Phone: 1-800-292-6141 | info@leco.com
www.leco.com |

Replacing LECO *Pegasus*® HT with *Pegasus* BT GC-TOFMS

By Jan Hazebroek, Ph.D.

The impacts and benefits of Pegasus BT in a laboratory's high-throughput metabolomics research.

INTRODUCTION

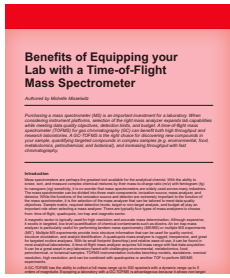
It's easy to fall into comfortable routines in the laboratory. Over the years, chemists learn how to work with their instruments and software, knowing exactly what they need to provide and how they need to translate the results to get useful data. A change to that rhythm can be daunting, but sometimes, it's necessary. Sometimes an instrument needs to be upgraded, or software needs to change. When an old workhorse retires, it's important to know the actual workflows in place to get the replacement up and running as painlessly as possible. Such was the case at Jan Hazebroek's laboratory at DuPont Pioneer.

AUTOMATED PLANT METABOLOMICS RESEARCH

Hazebroek's laboratory used the *Pegasus*® HT for high-throughput metabolomics, but the HTs were only growing older. In order to upgrade, he needed a good reasoning, and that reasoning started with the reasons why his lab did metabolomics research in the first place.

Lab- or field-based assays were costly and time-consuming. There had to be a way to predict desired phenotypes without needing to go through them. In the plant breeding field, where Hazebroek worked, it would take growing plants over multiple years in multiple locations to get a good read on the phenotypes due to the variability of environments year by year. A greenhouse assay that could predict the field performance would be much simpler and lower-cost.

Metabolomics research also allowed Hazebroek to identify gene discovery leads and confirm shifts in intended biochemical pathways. In addition, it augmented other profiling approaches with metabolite data that was inherently closer to the physical phenotype his company sold while still being cost effective enough to enable extensive plant-to-plant replication.


Benefits of Equipping your Lab with a Time-of-Flight Mass Spectrometer

WHITE PAPER

Benefits of Equipping your Lab with a Time-of-Flight Mass Spectrometer

That low-cost, cheaper replicability kept coming up as a reason why Hazebroek worked on plant metabolomics research in the first place. Because the major source of variability in his metabolomics experiments was the biology of the sample, high replication and thorough statistics on the results was essential. High replication costs money, though, so savings needed to be found in other places to keep the overall costs down. The major cost in the majority of labs isn't the instruments themselves, but actually the labor used to run them. To reduce the cost of labor, labs like Hazebroek's needed to automate.

Hazebroek's laboratory had already automated the sample weighing, sample extraction, high-throughput GC-MS, and high-throughput, high-resolution LC-MS. Combined with robust data management, they had a pipeline that went from the experimental design, carrying the metadata through the field to greenhouse experiments, through the analytics, through the pre-processing of data, all the way to the final reporting.

This lab worked with many different species and plant tissues, but the lion's share of their work is with corn leaf. Hazebroek has preparation down to a science, using methods established for over a decade. First, samples are harvested by taking a leaf puncher with an attached tube and collecting four leaf punches into the tube. The tube is then immediately, within three to five seconds, dipped in liquid nitrogen (or dry ice if the liquid nitrogen isn't available)

to stop the metabolomic process, while the rest of the plant is just fine to continue to grow and be sampled later, if necessary. This immediate freezing means that the biology of the plant itself is what is measured, and not the biology or chemistry that happens once the plant has been harvested.

The samples are stabilized by lyophilization, followed by grinding, weighing and extracting them in a chloroform methanol water (2:5:2) solution containing internal standards. After centrifugation, the samples endure a very traditional methoxyamine derivitaization, followed by trimethylsilylation with MSTFA before the sample is introduced to the GC-TOFMS instruments with a short column designed for speed. An aliquot of the sample can go to the LC-MS if necessary.

Sample weighing is just as automated. 2D barcoded tubes are used for the samples, which are loaded onto a robot. The robot weighs the empty tubes to the nearest hundredth of a milligram, and then the tubes are sent out to the clients to receive samples. When the samples return, the robot weighs them again, and the difference between the two weights is the sample weight. All of this data is recorded into the Laboratory Information Management System (LIMS). Control tubes with known weights are included to insure accuracy, and the LIMS is smart enough to flag weights that are negative or close to zero as missing samples. By keeping track of all of this information with the barcoded tubes, this also means that samples can be rearranged across the plate or plates, or even within a plate, so metadata doesn't get confounded with run order or run batch. With metabolomics experiments, it's very important to avoid these kinds of artifacts.

After fifteen years of metabolomics research, Hazebroek was convinced that consistency of samples is more important than accuracy, and so he used liquid handlers to do most of the pipetting. This had the added advantage of removing the chance for ergonomic injuries that can occur with high-throughput pipetting, especially of larger samples.

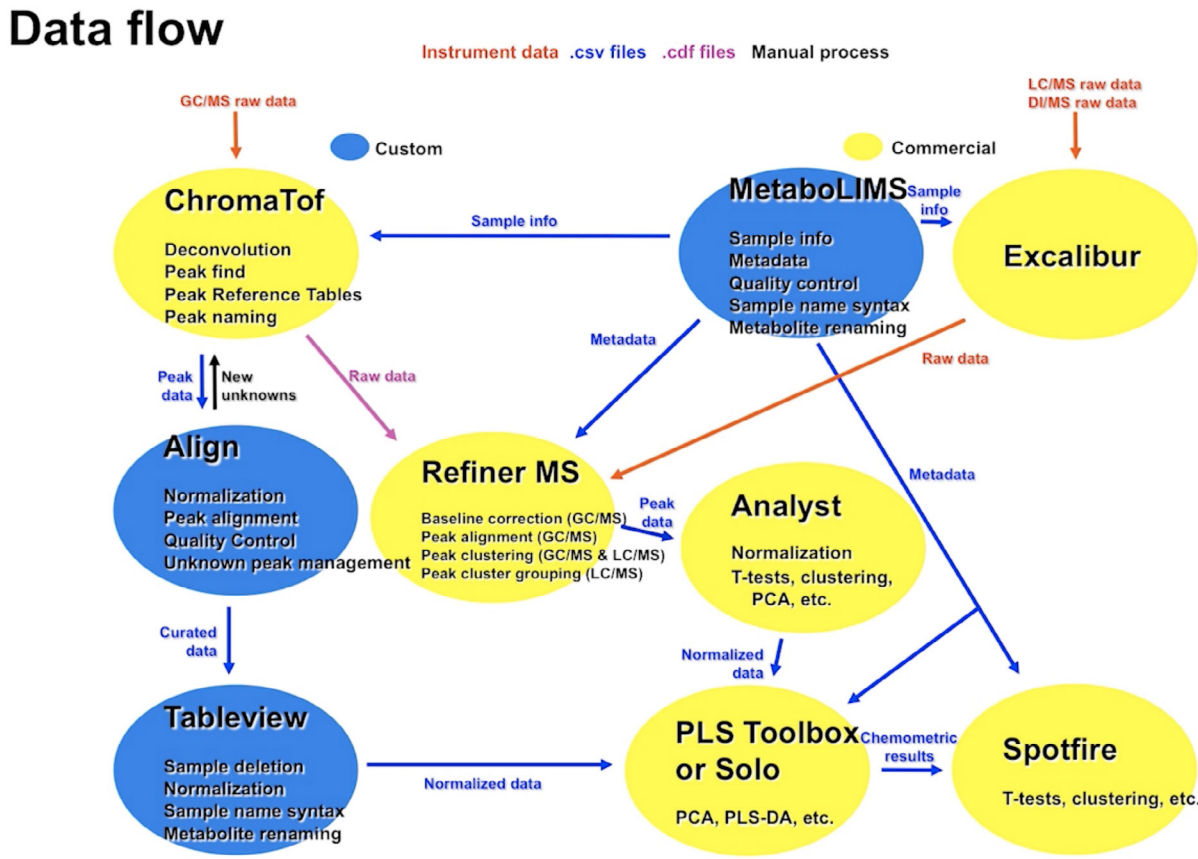
Fifteen years ago, the biggest challenge Hazebroek faced wasn't ergonomic injuries, but rather software. The hardware was impressive among most of the vendors, like LECO, but no vendor had an A-to-Z path for software that satisfied his needs. So in Hazebroek's laboratory, a combination of custom and commercial software was stitched together to create the solution he was after.

As demonstrated in **FIGURE 1**, LECO's *ChromaTOF*® software collected the data from the GC-MS instruments. The resulting data was then exported as netCDF files to a program called Refiner MS, from a company called Genedata. This did the alignment on a retention index, m/z value, and fingerprint level. It handled baseline corrections, de-noising, peak alignment, and clustering to create peaks, as well as some annotation.

This peak data then passed to another Genedata module called Analyst, where the data was normalized for sample weight and internal standard signal.

At this point, simple T-tests, multivariate, and univariate tests could be done on the data. Most of it was done by a software tool called PLS Toolbox, from a company called Eigenvector research. This entire time, the LIMS was managing the data and metadata, including any LCMS data. This data could be further worked with the *ChromaTOF* software, taking advantage of additional tools. A defined peak list, for example, could be annotated and deconvoluted, which Hazebroek liked a lot. These deconvoluted lists can be further annotated and combed through for more detailed peak reports. Unfortunately, each sample generated its own CSV file, and statistics couldn't be done on multiple CSV files.

FIGURE 1: Data flow using LECO's ChromaTOF® software collected data from GC-MS instruments.

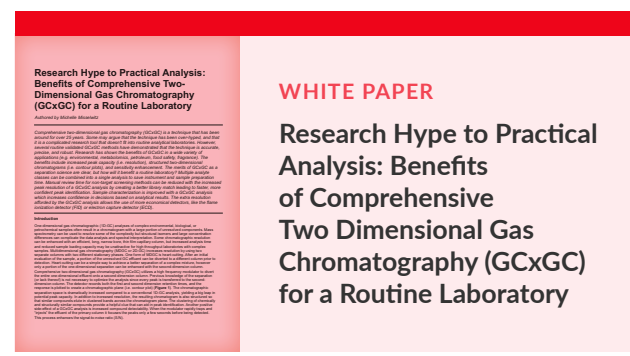


Hazebroek's solution was a self-written program called Align (**FIGURE 1**) that aligned metabolites based on their name, if known, or their retention time if unknown, and mass spectral similarity. Align created one file with a bunch of lists, a bunch of metabolites, that also wouldn't work with statistics, so the data went to another custom piece of software called Tableview (**FIGURE 1**) that would make a 2D array of these lists with samples on one dimension and m/z values on another, and finally, Hazebroek had the data that could be explored with statistics.

Though it sounded convoluted and looked incredibly messy, Hazebroek found a strategic advantage in this style of solution. As one module was updated or replaced, his lab didn't have to worry about databases communicating with each other. Instead, the only thing they needed to keep straight was the file syntax, which was much easier to deal with. This type of workflow meant that when change happened, as it inevitably would, the workflow itself would survive without stumbling.

The Genedata software that handled the alignments could visualize some of the data, such as a parallel plot for internal standards. With a bar chart or whisker plot for internal standards, outliers or irregularities in the data could be identified even before statistics were involved. This helped build the incredible capacity of Hazebroek's lab. Since 2003, his lab has run hundreds of thousands of samples, with a GC-MS capacity of 1,800 samples per week and 150 samples per week for LC-MS capacity, as they have fewer LC-MS instruments and do both positive and negative electrospray.

With this robust workflow in place, the immediate trepidation of a major change, like switching instruments, was already alleviated. The next question was: What could the *Pegasus* BT do that the *Pegasus* HT could not?



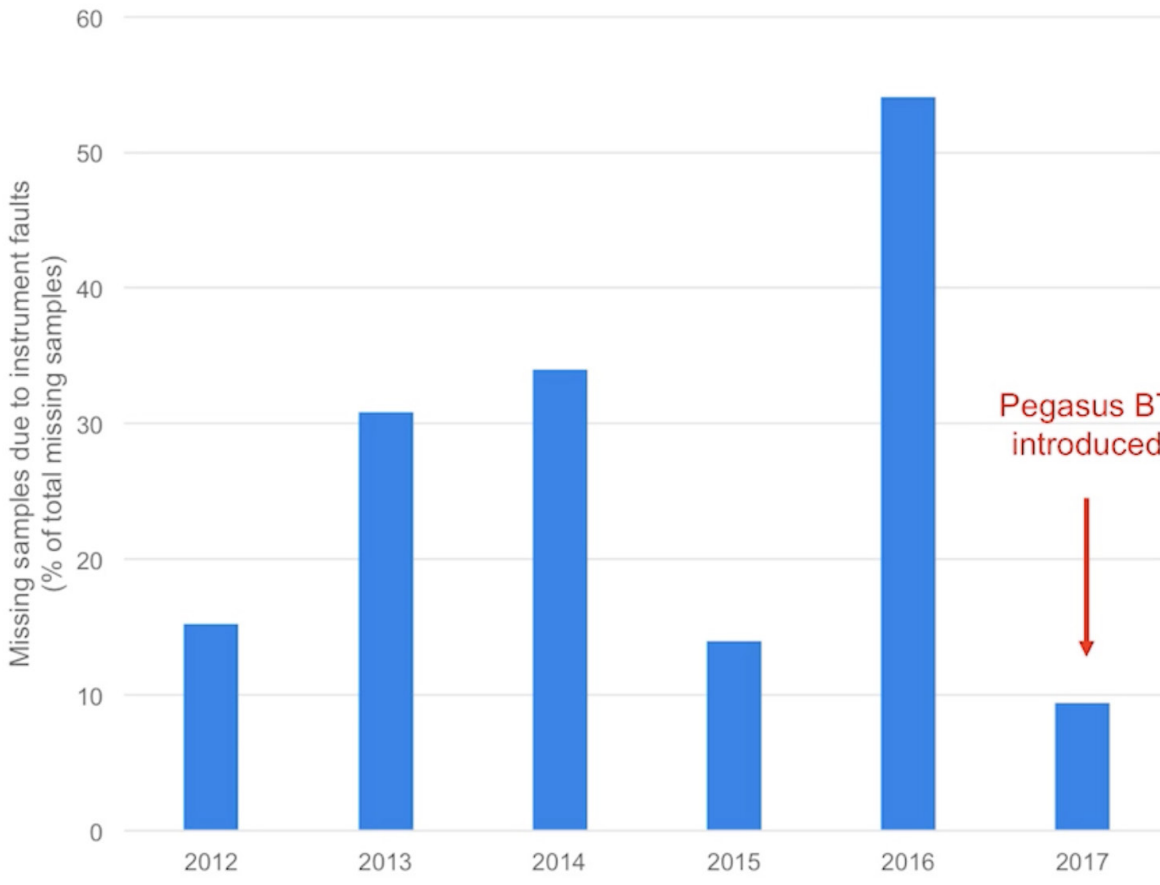
The initial point in the BT's favor was simply the fact that they were new instruments. The LIMS had tracked the reason why samples were missing. From 2012 to 2016, an instrument fault, whether it was the autosampler, the GC, or the MS, was an increasing reason for missing samples (except for 2015). In 2016, over half of the missing samples were due to the aging HTs. After the introduction of the BTs (**FIGURE 2**), less than 10% of missing samples were related to the instruments, and well into 2018, the StayClean™ Ion Source still had not needed cleaning.

The *Pegasus* BT was also much more sensitive than the HT. Because the samples

Hazebroek worked with were required to be so small to preserve the life of the plant, only 400 microliters of the extraction solvent could be generated, of which 300 microliters was required for the HT. If they wanted to run LC-MS on the same sample, they needed to take a second sample, which doubled the cost and complexity of an experiment. With the BT, however, only 50 microliters of the extraction solvent was necessary for the same signal strength, which left plenty for LC-MS analysis, as well as some extra in case something gets confounded in the process or they want to try a different test on that sample.

The BT also had a dynamic range of at least four orders of magnitude, which eliminated the need to run sample dilutions. Big peaks, such as sucrose, glucose, and fructose in plant tissues, would overload the columns. However, these are important metabolites

FIGURE 2: Introduction of *Pegasus* BT resulted in fewer instrument faults.



for plant biochemistry, so those samples would need to be diluted in order to measure them with the HT. As expected, diluting samples doubles the analytics and increases experimental complexity. With the BT, the columns were not overloaded and small peaks were still visible alongside the larger ones. The BTs also had higher mass resolution capabilities than the HTs, but as Hazebroek's lab had separate high-resolution instruments, this wasn't as important in his decision making.

LECO's autosampler, the L-PAL3, was used to do just-in-time derivatization. Trimethylsiylation derivatization is not a particularly stable reaction. Some metabolites derivatized quickly, some derivatized slowly, and some derivatized and start breaking down immediately, such as amino acids. If the trimethylsiylation derivatization was done at the start of the day, the chemistry of the first sample was going to be different than the chemistry of the last sample if it sat on the autosampler for 20 hours before going into the instrument. Just-in-time derivatization alleviated this problem by having the autosampler transfer the sample into an agitator 30 minutes prior to injection, deliver an aliquot of MSTFA, shake it for 30 minutes, and then inject the sample into the GC-MS. This reduced run order effects in the analytics.

Just-in-time derivatization required a syringe exchange tool, because the syringe required to deliver the proper volume of MSTFA was bigger than the syringe necessary for the injection into the GC. It also required an agitator and a LECO script embedded within *ChromaTOF*. Hazebroek's samples required the script to be able to handle two samples, because the incubation period was thirty minutes but the runtime was only 15 minutes. In addition, he wanted to continue to use the VT96 sample tray, because so much of their automation was built around this particular tray. While the L-PAL3's tray holder did not fit the VT96, LECO designed and built a tray holder adaptor that did fit over the L-PAL3's tray holder slots to accommodate the VT96, which allowed Hazebroek's workflow to continue unimpeded. Other minor modifications included little clamps to hold the solvent wash bottles in place and solvent wash bottle venting to keep the nasty fumes out of the lab. These were all little things that needed to be considered when upgrading instruments.

The next worry about upgrading instruments was the integration of the software, but as Hazebroek's workflow dealt with file type transfers instead of databases talking, all he needed to do was make sure the BT could import sequence files from the LIMS as a CSV file. There was no difference here between the HT and the BT, so no changes were required. The same was true for exporting instrument information out of *ChromaTOF* and into the LIMS, such as data acquisition rate, detector voltage, active filament, and injection time/date. This was important so sample differences

in a PCA plot could be accurately attributed to something like a filament that blew in the middle of a run. Instead of misleading the research, tracking this information allows for differences to be appropriately categorized when they aren't actually differences in the sample itself. Custom scripts written to convert retention times from the netCDF files to retention indexes also worked exactly the same with the BT as they had with the HT.

The main challenge with the BT's netCDF files was actually the size, itself. The initial version of the *ChromaTOF* software produced huge netCDF files. A *ChromaTOF* patch has since addressed that issue, but in the interim, Hazebroek would dummy the data down to normal nominal mass only. He was resistant at first, because he didn't want to throw away data, but his fifteen years of experience and success, and all their libraries, were based on nominal mass, so as long as they kept that, he figured they would be okay. Fortunately, the Genedata software they already used had a tool to convert the netCDF files to normal nominal mass, which reduced the file size by 90% and solved their memory issues.

CONCLUSION

By taking the time to future-proof their workflows, Hazebroek was able to easily change his laboratory from the *Pegasus* HT to the *Pegasus* BT for an increase in productivity. Working with the manufacturer, he had identified and solved problems with the physical workflow before they could slow anyone down. Upgrading an instrument doesn't need to be a challenge. If done right, it will only bring advantages.

View full webinar here.

Jan Hazebroek, Ph.D.
Corteva Agriscience