Utilization of Statistical Compare Software and Fisher Ratios Prior to Multivariate Analysis for Complex GCxGC-TOFMS Data in Order to Define Statistical Variation Between the Small Molecule Metabolite Profiles of Different Fish Species

John Heim • LECO Corporation; Saint Joseph, Michigan USA

Key Words: Fisher Ratio, Statistical Compare, Multivariate Analysis, GCxGC-TOFMS

1. Introduction

Metabolomics presents challenges for both the analytical methods used and the data reduction required to interpret the results. Metabolomic samples demand analytical solutions and instrumental methods that will identify the small molecule metabolite profile completely as well as discover significant key components of interest. Comprehensive multi-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) has emerged as an excellent instrumental option for the characterization of complex metabolite profiles.

This study investigates tissue extraction techniques to optimize a simple and fast procedure which provides characterization of the small metabolite profiles between two different fish species. The experimental research conducted explores the differences between the metabolite profiles of N-(t-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA) derivatized muscle tissue from different fish species analyzed by GCxGC-TOFMS. GCxGC-TOFMS provides increased peak capacity and resolution for the chromatographic separation while fast TOFMS acquires the mass spectral data density necessary to characterize complex biological samples. The combination of TOFMS data and deconvolution algorithms facilitate trace level analyte detection that would otherwise be hidden and coeluted with other compounds in the sample. Two-dimensional chromatographic plots of biological samples showing increased peak capacity and structural orientation not possible in one dimensional chromatography will be highlighted.

In addition, this research demonstrates a data mining strategy from GCxGC-TOFMS with results which focus on locating the metabolite differences between species. Processed sample data was loaded as separate classes into the Statistical Compare feature of ChromaTOF® software for peak table alignment, statistics generation, and Fisher Ratio calculations in order to define the unknown chemical variations between metabolites from different fish species, (Wild Canadian White Lake Perch, and Lake Michigan Lake Trout). Results from the statistical comparison were subsequently exported to a commercially available peripheral multivariate analysis software package, Miner3D, whereby analyte differences between the metabolite profiles of two fish species were examined.

2. Experimental Conditions

This research was designed to study MTBSTFA derivatized fish muscle tissue extractions for the small molecule metabolite profile intended to detect possible chemical variations between wild Canadian White Lake Perch and Lake Michigan Lake Trout. A series of 2.5 gram aliquots of muscle tissue from each species were finely ground and placed in 20 mL glass scintillation vials. Two drops of concentrated sulfuric acid were added to each sample to approximately pH 2. Samples were extracted with 5 mL of methylene chloride after sonication for 1 hour at 37°C. The methylene chloride supernatant was siphoned from the extraction mixture and placed in a new vial. The supernatant was centrifuged and solids were removed prior to evaporation under a stream of nitrogen. Before derivatization the samples were reconstituted in 500 µL of methylene chloride. Derivatization was carried out with N-(t-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA). A 200 µL aliquot of extractant was placed in a 2 mL amber glass autosampler vial containing 0.5 mg of sodium sulfate. To each sample vial 30 uL of pyridine and 100 µL MTBSTFA was added and sealed. The vials were heated at 60°C for 1 hour. Derivatized samples were then analyzed by GCxGC-TOFMS on the same day as prepared.

GCxGC-TOFMS results were generated with a LECO Pegasus® 4D time-of-flight mass spectrometer (TOFMS). The Pegasus 4D GC-TOFMS instrument was equipped with an Agilent 7890 gas chromatograph featuring a LECO two stage cryogenic modulator and secondary oven. LECO ChromaTOF software was used for all acquisition control, data processing, Statistical Compare, and Fisher Ratio calculations. A 30 m x 0.25 mm x 0.25 µm film thickness, Rxi-5ms, (Restek Corp., Bellefonte, PA) GC capillary column was used as the primary column for the GCxGC-TOFMS analysis. In the GCxGC configuration a second column 1.25 m x 0.10 mm id. x 0.10 µm film thickness, BPX-50, (SGE, Austin, TX) was placed inside the LECO secondary GC oven after the thermal modulator. Helium carrier gas flow rate was set to 1.0 mL/min at a corrected constant flow. The primary column was programmed with an initial temperature of 50°C for 2.0 minutes and ramped at 6°C/minute to 280°C for 12 minutes. The secondary column temperature program was set to an initial temperature of 55°C for 2.0 minutes and then ramped at 6°C/minute to 285°C with a 12 minute hold time. The thermal modulator was set to +20°C relative to the primary oven and a modulation time of 4.5 seconds was used. The total GCxGC-TOFMS run time was 53.17 minutes. The MS mass range was 40-750 m/z with an acquisition rate of 150 spectra/second. The ion source chamber was set to 230°C and the detector voltage was 1800V with an electron energy of -70eV.

3. Results and Discussion

Results of the metabolite profile study between perch and lake trout subjects are shown in Figures 1 and 2 by the total ion chromatograms depicted as contour plots. These chromatographic examples visually illustrate peak differences between sample types as well as highlight the benefits GCxGC-TOFMS offers which include increased peak capacity, improved analyte detectability, and enhanced resolution. On average over 1900 peaks were found per sample with a S/N ratio of greater than 200 for this study. The red cross hatched area in each contour plot is an unprocessed region developed in the Classifications feature of ChromaTOF software which eliminates unwanted background peaks.

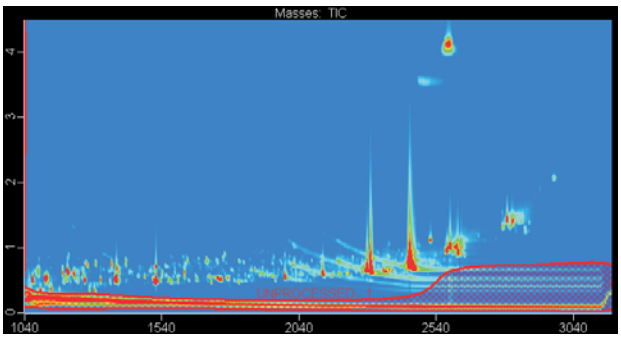

Figure 1. Lake Michigan Lake Trout: Contour Plot Total Ion Chromatogram of MTBSTFA-derivatized fish extraction showing the small molecule metabolite profile. Over 1900 peaks were found at a S/N of 200.

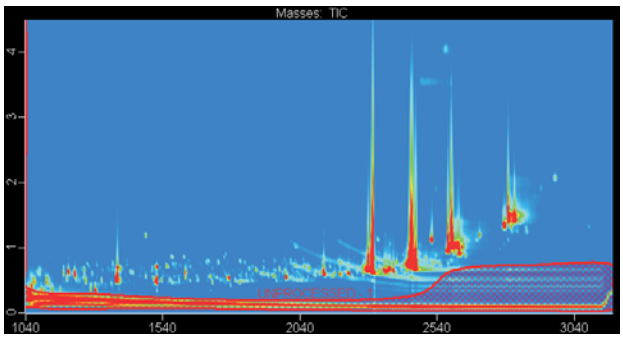

Figure 2. Wild Canadian White Lake Perch: Contour Plot Total Ion chromatogram of MTBSTFA-derivatized fish tissue extraction showing the small molecule metabolite profile. Over 1950 peaks were found with a S/N of 200.

Data Mining Strategy Using the Statistical Compare Feature

The "Statistical Compare" option available in ChromaTOF software allows the user to view statistical comparisons as a data processing step for groups of samples. The groups of samples are divided into different classes. ChromaTOF software aligns the data for the specified group of samples from the data processed peak tables. Upon completion of peak alignment, statistical information on various peak properties (such as peak height, peak area, retention time, etc.) can be viewed in the Compound Table generated by Statistical Compare. Additionally, "Fisher Ratios" can be calculated from the Compound Table for each analyte. The Fisher Ratio is a statistical calculation that can be used to discover the unknown chemical differences among known classes of complex samples. In addition Statistical Compare results can be exported as a .csv file and applied to third party software programs for supplemental data reduction such as multivariate analysis.

Creating a Statistical Compare Method

The figures below illustrate several key windows used in the process of building the Statistical Compare method for this research. A Statistical Compare method is created in the database tree as a new method. Classes are then created in the Class Table and samples are added to the appropriate class. The Sample Table can be viewed and checked for accuracy in ChromaTOF. A Compound Table is generated from the Class and Sample Table. From the Compound Table, analyte peak alignment is completed and various statistics calculated for all of the analytes. After analyte alignment is completed, a Fisher Ratio can be calculated for each analyte found in the Compound Table. Data from the Compound Table was then exported as a .csv file for multivariate analysis in the peripheral software program Miner3D. Multivariate analysis including PCA and K-means clustering was calculated for the top 340 analytes according to their Fisher Ratio values. The graphical results are shown in Figure 7.

Sample Table-071509 PERCH VS...

Sample	Name	Class
1*	2nd DP_W W L PERCH 3:1	PERCH
2	2nd DP_W W L PERCH 4:1	PERCH
3	2nd DP_W W L PERCH 1:2	PERCH
4	W L LAKE PERCH:1	PERCH
5	W L LAKE PERCH 2:1	PERCH
6	W L LAKE PERCH 3:1	PERCH
7	W L LAKE PERCH 4:1	PERCH
8	2nd DP_L M LAKE TROUT 2:1	LAKE TROUT
9	2nd DP_L M LAKE TROUT 3:1	LAKE TROUT
10	2nd DP_L M LAKE TROUT 4:1	LAKE TROUT
11	LAKE TROUT 1:1	LAKE TROUT
12	LAKE TROUT 2	LAKE TROUT
13	LAKE TROUT 3:1	LAKE TROUT
14	LAKE TROUT 4:1	LAKE TROUT

Figure 3. The figure above shows the Sample Table of seven Perch and seven Lake Trout samples data processed from the GCxGC-TOFMS analysis.

The partial Compound Table shown below was generated by the processed data files from seven perch and seven lake trout samples which were derivatized with MTBSTFA and analyzed by GCxGC-TOFMS.

Compound Table-071509 PERCH...

Analyte	Name	Area	Fisher Ratio
		Average	
1219*	Arachidonic acid, trimethylsilyl ester	14901908.6	16916
487	1-Nonadecene	7413366.54	13533
471	8-Heptadecene	1987485.77	8972.3
289	Propanoic acid, 2-[(tert-butyldimethylsilyl)oxy]-, tert-b	76637.020	7438.3
1265	4,7,10,13,16,19-Docosahexaenoic acid, methyl ester	64237880.5	4970.2
15	3-BUTEN-2-ON, 3-TRIMETHYLSILYLOXY-	42337537.1	4256.1
901	Linolenic acid, trimethylsilyl ester	963155.609	4242.7
300	PENTADECANE	13470009.3	3826.3
133	8-METHYLENE-3-OXA-TRICYCLO[5.2.0.0 2,4]NONAI	920829.900	3662.8
771	Tetradecanal	1193339.90	3653.2
210	1-ETHYLHEXYL TRIMETHYLSILYL ETHER	1454699.69	3542.2
642	Dodecanoic acid, tert-butyldimethylsilyl ester	2445712.67	3435.4
983	Heptadecanoic acid, tert-butyldimethylsilyl ester	11607435.3	3433.1
493	PENTADECANE, 2,6,10,14-TETRAMETHYL-	73894.882	3429.7
101	3-PYRIDINYL TRIMETHYLSILYL ETHER	7263933.48	3349.8
151	2-Ketoisocaproic acid, trimethylsilyl ester	2128436.73	3341.9
1053	Octadecanoic acid, tert-butyldimethylsilyl ester	14186110.6	3139.4
1264	Squalene	23850817.3	3121.6
158	Acetamide, 2,2,2-trifluoro-N-methyl-N-(trimethylsilyl)-	1673543.50	2860.6

Figure 4. A partial Sample Group Compound Table is illustrated above showing the top 19 compounds with the highest Fisher Ratio values. These compounds represent the most variation between the lake trout and perch analytes.

The Fisher Ratio plot shown below illustrates the analytes of highest variance from the Compound Table developed in Statistical Compare. The most intense peaks shown in

Figure 5 by their Fisher Ratio show the compounds that exhibit the highest class to class variation. The application of the Fisher Ratio to the Statistical Compare "Compound Table" provides a data reduction technique to find metabolites of interest in a large data set.

Results from the Compound Table can be conveniently exported as a .csv file to be used as an excel spreadsheet in external data analysis software packages. The partial spreadsheet shown below represents the analytes with the highest Fisher Ratios and shows the columns used for the multivariate analysis conducted for the fish metabolite study. The columns represent the variables Peak Compound Name, the Class (perch or lake trout) and the peak area for each analyte in a sample.

Multivariate Analysis

Multivariate analysis is based on multivariate statistics, which involves observation and analysis of more than one statistical variable at a time. The technique is used to perform studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. This study applied ChromaTOF's Statistical Compare and Fisher Ratios to a data set of fourteen samples from Wild Canadian White Lake Perch and Lake Michigan Lake Trout that determined the analytes with highest variation across the entire sample population.

The Statistical Compare results generated a Compound Table which was exported as a .csv file in Excel format and applied to the peripheral multivariate analysis platform, Miner3D. PCA analysis was conducted on the variables of analyte identification (Compound Name), class, (Perch or Lake Trout), and analyte (Peak Area). Following PCA analysis, K-means clustering was applied for five clusters.

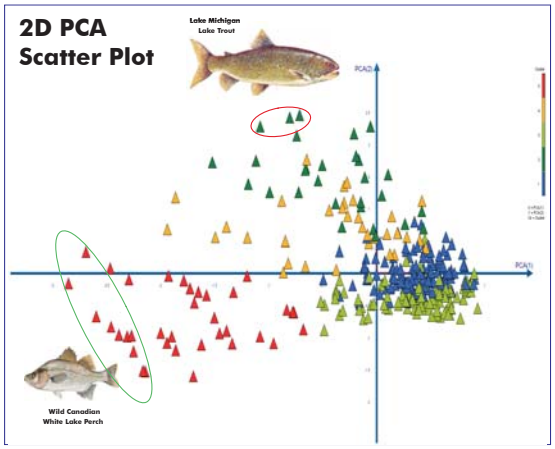

Figure 7. The figure above shows the 2D graph of PCA 1 on the x-axis and PCA 2 on the y-axis, as well as the cluster analysis depicted for 5 different colors developed in the third party software program Miner3D. The goal for this GCxGC-TOFMS analysis was to find metabolites that were unique to either Lake Trout or Perch species. Ten analytes illustrated inside the green oval were found to be unique to the Wild Canadian White Lake Perch, while three analytes shown inside the red oval in the graph above were found to be unique to the Lake Michigan Lake Trout.

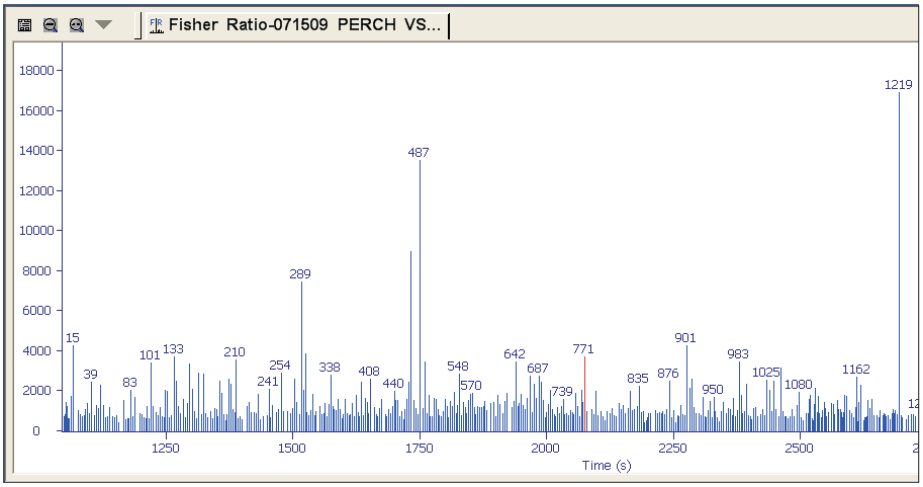

Figure 5. The partial Fisher Ratio plot shown above graphically represents unknown chemical differences between the Lake Michigan Lake Trout and the Wild Canadian White Lake Perch. The greatest Fisher Ratio value (analyte 1219) is shown as the highest intensity value in the graph.

Microsoft Excel - 72009_fish metabolites test 2B.csv

File Edit View Insert Format Tools Data Window Help Adobe PDF

A39: 7-(3-{[TERT-BUTYL(DIMETHYL)SILYL]OXY}PROPYL)BICYCLO[2.2.1]HEPTAN-2-ONE

	A	B	C	D	E	F	G
1		2nd DP_W W L PERCH 3.1		2nd DP_W W L PERCH 4.1		2nd DP_W W L PERCH	
2	Peak	Class	Area	Class	Area	Class	Area
3	Arachidonic acid, trimethylsilyl ester	PERCH	13339599	PERCH	12042077	PERCH	23065516
4	1-Nonadecene	PERCH	389854.4	PERCH	119268.4	PERCH	23089.24
5	8-Heptadecene	PERCH	511340.5	PERCH	497075.4	PERCH	515367.2
6	Propanoic acid, 2-[(tert-butyldimethylsilyl)oxy]-, tert-butyldimethylsilyl ester	PERCH	300323.2	PERCH	410405.8		
7	4,7,10,13,16,19-Docosahexaenoic acid, methyl ester, (all-Z)-	PERCH	34409406	PERCH	4740981	PERCH	8593795
8	3-BUTEN-2-ON, 3-TRIMETHYLSILYLOXY-	PERCH	21001424	PERCH	7355273	PERCH	8805697
9	Linolenic acid, trimethylsilyl ester	PERCH	289976.6	PERCH	93608.98	PERCH	150942
10	PENTADECANE	PERCH	3277606	PERCH	1920414	PERCH	1797864
11	8-METHYLENE-3-OXA-TRICYCLO[5.2.0.0.2,4]NONANE	PERCH	342466	PERCH	67859.72	PERCH	100577.5
12	Tetradecanal	PERCH	1643411	PERCH	772183.9	PERCH	1674763
13	1-ETHYLHEXYL TRIMETHYLSILYL ETHER	PERCH	2338437	PERCH	383203.4	PERCH	1294280
14	Dodecanoic acid, tert-butyldimethylsilyl ester	PERCH	3139935	PERCH	2567842	PERCH	2830517
15	Heptadecanoic acid, tert-butyldimethylsilyl ester	PERCH	12947175	PERCH	8830823	PERCH	14527758
16	PENTADECANE, 2,6,10,14-TETRAMETHYL-			PERCH	581.895		

Figure 6. Shown above is a partial section of the Microsoft Excel spreadsheet exported as a .csv file from ChromaTOF which is then imported into a peripheral multivariate analysis software program.

Life Science and Chemical Analysis Solutions

The two-dimensional graph shown in Figure 7 was developed in the commercially available Miner3D software. The graph shows clear differences as well as similarities in the small molecule metabolites found in both fish species analyzed by GCxGC-TOFMS.

Using Statistical Compare and Fisher Ratios to Find Unknown Data Patterns

The Statistical Compare feature can be utilized to find significant differences between analytes of different classes. The example shown in Figures 8 and 9 illustrates the use of the *Fisher Ratio* and the *Compound Statistics by Class Table* to find an analyte with dietary benefits for cardiovascular health. This example shows that significantly higher concentrations of the essential fatty acid, Linolenic acid, can be found in the Perch species.

Compound Stats By Class-0715...

ID	Name	1st Dimensic	2nd Dimensi	Area	Count
		Average	Average	Average	
901-1	PERCH	2277.3	0.757114	1727224.19	7
901-1-1	2nd DP_W W L PERCH 3:1	2276.02	0.7194	289976.550	
901-1-2	2nd DP_W W L PERCH 4:1	2280.51	0.726	93608.977	
901-1-3	2nd DP_W W L PERCH 1:2	2280.51	0.726	150941.952	
901-1-4	W L LAKE PERCH:1	2271.52	0.7788	5359489.46	
901-1-5	W L LAKE PERCH 2:1	2280.51	0.7788	491028.745	
901-1-6	W L LAKE PERCH 3:1	2280.51	0.7788	566264.311	
901-1-7	W L LAKE PERCH 4:1	2271.52	0.792	5139259.33	
901-2*	LAKE TROUT	2275.27	0.7194	71742.266	6
901-2-8	2nd DP_L M LAKE TROUT 2:1	2276.02	0.7194	50129.910	
901-2-9	2nd DP_L M LAKE TROUT 3:1	2271.52	0.7326	17486.627	
901-2-10	2nd DP_L M LAKE TROUT 4:1	2276.02	0.7194	104612.387	
901-2-11	LAKE TROUT 1:1	2271.52	0.7722	131371.228	
901-2-12	LAKE TROUT 2	2280.51	0.7656	126831.585	
901-2-13	LAKE TROUT 3:1	2276.02	0.6072	21.860	
901-2-14	LAKE TROUT 4:1				
901	Total	2276.36	0.739708	963155.609	13

Figure 8. Shows the Compound Statistics by Class table for Linolenic acid-(1TMS). The Compound Stats by Class table reveals an unknown pattern between the different concentrations of Linolenic acid found in perch and lake trout for this study. The results indicate that the average peak area of Linolenic acid found in perch is nearly 100 times greater than the average peak area of Linolenic acid found in lake trout. The statistical inference implies that Wild Canadian White Lake Perch is a better source of dietary Linolenic acid than Lake Michigan Lake Trout.

4. Conclusions

This research presents a step by step data mining approach using Wild Canadian White Lake Perch versus Lake Michigan Lake Trout GCxGC-TOFMS metabolomics data to demonstrate the value of the LECO ChromaTOFStatistical Compare feature. In addition, the capability to export useful results tables to peripheral multivariate software providing supplemental data interpretation was shown. A comprehensive GCxGC-TOFMS analysis accompanied by statistical comparison targeting high variance data through Fisher Ratios along with multivariate PCA and Clustering analysis was illustrated. This exploratory research presents an optimized GCxGC-TOFMS analysis with a data mining strategy using preliminary statistical methods prior to multivariate analysis that establishes a viable experimental approach which identifies significant metabolite variation between complex biological samples. The results presented illustrate that significantly increased analytical performance is achieved by utilizing GCxGC-TOFMS for the characterization of small molecule metabolite profiles. Several new ChromaTOF software features were applied which allow the analyst to find significant unknown chemical differences among known classes of complex samples. The new features available in LECO's ChromaTOF software were used to align a large set of data and define the highest variance for analytes between two different fish species. Furthermore, it was demonstrated that results from Statistical Compare and Fisher Ratio calculations can be exported quite simply into multivariate analysis programs whereby PCA and Clustering analysis can be applied. This research illustrates the capability of the Statistical Compare software feature to facilitate data reduction, and define metabolite variance thereby increasing the overall experimental results between complex biological sample classes.

Compound Table-071509 PERCH...

Analyte	Name	R.T. (s)	Area	Fisher Ratio
		Average	Average	
901	Linolenic acid, trimethylsilyl ester	2276.36,0.7	963155.609	4242.7

Figure 9. Shows TMS derivatized Linolenic acid in the Compound Table generated by Statistical Compare. A relatively large Fisher Ratio was calculated indicating high variation between the lake trout and perch classes. Linolenic acid is an essential fatty acid which has been assessed for its dietary benefits in cardiovascular health.

LECO Corporation • 3000 Lakeview Avenue • St. Joseph, MI 49085 • Phone: 800-292-6141 • Fax: 269-982-8977
info@leco.com • www.leco.com • ISO-9001:2000 • No. FM 24045 • *LECO is a registered trademark of LECO Corporation.*

Form No. 203-821-377 3/10-REV1