An LC/MS Metabolomics Discovery Workflow for Malaria-Infected Red Blood Cells Using Mass Profiler Professional Software and LC-Triple Quadrupole MRM Confirmation

Application Note

Authors

Theodore Sana
Steve Fischer
Shane Tichy

Agilent Technologies, Inc.
Santa Clara, CA USA

Abstract

The utility of a combined un-targeted discovery-based LC/MS and targeted confirmation LC/MS/MS approach for mining metabolomics data was demonstrated using malaria-infected red blood cells as a model system. Several compounds, including arginine, citrulline, and ornithine, were identified as metabolites whose biosynthesis was altered in erythrocytes infected by the malaria parasite.

Agilent Technologies

Introduction

Malaria is a global health problem, and one of the most serious infectious diseases, with approximately 300 million clinical cases and over one million deaths annually. We have used a metabolomics approach to demonstrate the utility of a combined un-targeted (discovery) LC/MS and targeted (confirmation) LC/MS/MS workflow (**Figure 1**) to find metabolites that are present at different relative levels in cultured non-infected red blood cells (NRBC), versus malaria-infected red blood cells (IRBC). Un-targeted data mining of LC/Q-TOF MS data resulted in the detection of thousands of metabolites which were subsequently analyzed for statistical differences in Mass Profiler Professional (MPP) software to discover patterns of association between sample groups. Differential metabolites were queried against the METLIN Personal Compound Database for identification. As an example of the powerful analytical capabilities of the software, the Molecular Feature Extraction (MFE) algorithm detected m/z 175.1190, corresponding to the amino acid arginine, to be significantly suppressed or absent in IRBC relative to NRBC samples. The use of a pathway analysis tool in MPP highlighted the arginine metabolism pathway that also includes the physiological amino acids ornithine

A combined un-targeted (discovery) LC/MS and targeted (confirmation) LC/MS/MS metabolomics workflow

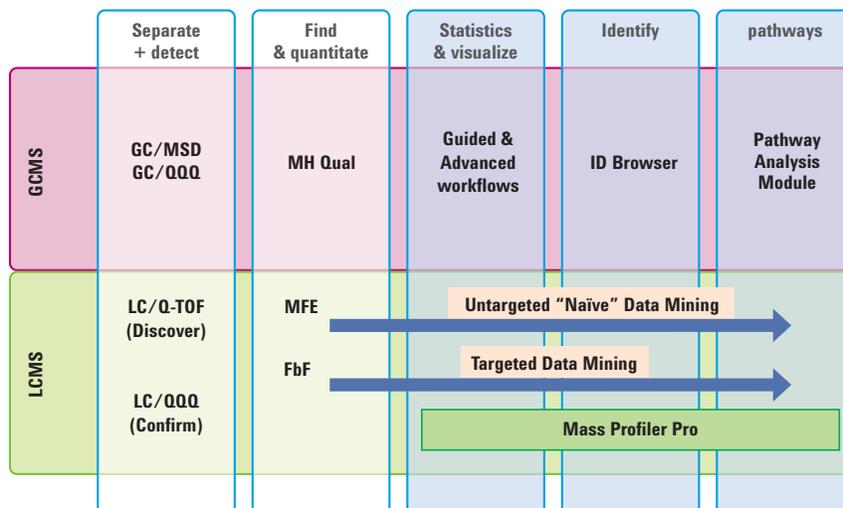|	Separate + detect	Find & quantitate	Statistics & visualize	Identify	pathways
GCMS	GC/MSD GC/QQQ	MH Qual	Guided & Advanced workflows	ID Browser	Pathway Analysis Module
LCMS	LC/Q-TOF (Discover)	MFE	Untagged "Naïve" Data Mining		
		FbF	Targeted Data Mining		
	LC/QQQ (Confirm)		Mass Profiler Pro		

Figure 1. A graphical representation of the analytical workflow used for the study of malaria infected erythrocytes. Agilent metabolomics workflow from separation to pathway analysis typically involves both GC/MS and LC/MS analyses. Molecular feature extraction (MFE) and Find by Formula (FbF) are two different algorithms used for finding peaks. All results files generated by the different analytical platforms can be imported into Mass Profiler Professional (MPP) for statistical analysis and visualization.

and citrulline. Subsequent "targeted" data mining of the original data using Extracted Ion Chromatograms (EICs) corresponding to these two compounds revealed that they were also present at low levels across all samples and that arginine metabolism was probably affected in IRBCs. This finding was subsequently tested by multiple reaction monitoring (MRM) assays targeting both samples and standards for arginine, citrulline, and ornithine on an Agilent 6460 Triple Quadrupole LC/MS with Jet Stream technology. Due to the increased sensitivity of this platform, we were able to confirm that arginine levels are significantly reduced, whereas citrulline and ornithine levels were elevated in IRBCs.

Experimental

Study design

Two separate groups of RBC culture samples (total n=24): Control (NRBC) and infected (IRBC) cells at approximately 10% parasitemia were processed using a previously described sample extraction method for RBCs[1]. The extraction solvent pH was varied to either pH 2, 7 or 9 in a biphasic solution of methanol/water/chloroform to optimally drive biphasic separation in each of the quadruplicate samples per pH sub-group (condition), and to maximize metabolite recovery in the aqueous phase.

Cell culture

The 3D7 strain of *Plasmodium falciparum* parasite was grown in asynchronous erythrocytes, maintained in RPMI 1640 medium supplemented with 25 mM HEPES, 20 mM sodium bicarbonate, 370 µM hypoxanthine, and 0.5% AlbuMAX (Invitrogen) at 5% hematocrit. The cultures were incubated under 90% N_2, 5% O_2, and 5% CO_2 at 37°C. A 0.5 mL aliquot of erythrocyte culture was harvested for each sample 24 hours after parasite exposure, quenched with methanol, and metabolites extracted according to the protocol.

Sample filtering

In order to remove residual proteins that can cause significant detection interference in the LC/MS step, the aqueous phase extracts of the cell cultures were filtered through 0.2 µm Pall Nanosep MF filters (Pall Life Sciences, Ann Arbor, MI), followed by 10 kDa Pall Nanosep Pall centrifugal device filters, dried under vacuum, and re-suspended in 100 µL of 50:50 methanol:water, 0.2% acetic acid. One microliter was injected into each LC system.

Instruments

This study was performed in two parts: "un-targeted" discovery and "targeted" confirmation. Un-targeted discovery was carried out using reverse phase chromatography on an Agilent 1200 Series Rapid Resolution LC (RRLC) system equipped with a binary pump, autosampler with chiller and thermostatted column compartment, and coupled with an Agilent 6520 Accurate-Mass Q-TOF LC/MS with a dual ESI source (**Table 1**).

Confirmation of the discovery results was performed using Aqueous Normal Phase (ANP) chromatography on an Agilent 1200 Series RRLC system equipped with a binary pump, vacuum degasser, high performance thermostatted autosampler and a thermostatted column compartment, and connected to an Agilent 6460 Triple Quadrupole LC/MS System with Jet Stream technology (**Table 2**).

Table 1. LC and Q-TOF MS Conditions

LC Run Conditions	
Column	Guard column: ZORBAX SB-C8, 2.1 x 30 mm, 3.5 µm (P/N 873700-936) Analytical Column: ZORBAX SB-Aq, 2.1 x 50 mm, 1.8 µm (P/N 827700-914)
Column temperature	60°C
Injection volume	1 µL
Autosampler temperature	4°C
Needle wash	3 sec in wash port
Mobile phase	A = 0.2% acetic acid in water B = 0.2% acetic acid in methanol
Flow rate	0.6 mL/min
Linear gradient	• 2% methanol to 98% methanol in 13 min • 6 min hold at 98% methanol • Stop time: 19 min • Post time: 5 min

Q-TOF MS Conditions	
Ion mode	Both positive and negative
Drying gas temperature	250°C
Drying gas flow	10 L/min
Nebulizer pressure	45 psig
Capillary voltage	4000 V (positive mode) / 3500 V negative mode
Fragmentor	140 V
Reference Delivery	Agilent 1100 isocratic pump with 100:1 splitter (P/N G1607-60000)
Reference pump flow	1 mL/min for 10 µL/min to nebulizer
Reference ions	Positive mode: 121.050873 and 922.009798 Negative mode: 119.036320 and 966.000725
Instrument mass range	1700 Da
Acquisition rate	2 Hz
TOF spectra mass range	50 to 1000 *m/z*
Data storage	Centroid
Threshold	100
Instrument mode	Extended Dynamic Range

Table 2. LC and Triple Quadrupole MS Conditions

LC Run Conditions	
Column	Cogent Microsolv Diamond Hydride 100Å 4µm 2.1 x 150 mm
Column temperature	45°C
Injection volume	1 µL
Autosampler temperature	4°C
Needle wash	3 seconds in wash port
Mobile phase	A = 0.1% formic acid in water B = 0.1% form acid in acetonitrile
Flow rate	0.4 mL/min
Linear gradient	• 100% acetonitrile to 65% acetonitrile in 2.80 min • 2.2 min hold at 65% acetonitrile • Stop time: 5 min • Post time: 5 min

Triple Quadrupole MS Conditions	
Ion mode	Positive
Drying gas temperature	300°C
Drying gas flow	6 L/min
Sheath gas temperature	275°C
Sheath gas flow	11 L/min
Nebulizer pressure	45 psi
Capillary voltage	4000 V (Positive mode)
Nozzle voltage	500 V
Delta EMV	200 V
MRM	175.0 -> 70.1; 133.0 -> 166.1; 176.0 -> 159.1

Data acquisition and analysis

MassHunter Qual was used in a naïve, un-targeted data-mining approach to analyze the LC/MS-ESI TOF data in the discovery phase. The resulting files were then imported into Mass Profiler Professional for analysis. Agilent MassHunter Optimizer Software (B.02.00) was used with the triple quadrupole LC/MS system for identifying the most abundant multiple MRM transitions along with the associated fragmentor voltages and collision energies. MassHunter Workstation (version B.02.01) was utilized for data acquisition and MassHunter Quantitative Analysis Software (version B.04.00) was used for data processing.

Results and Discussion

Un-targeted data mining

Metabolomics workflows typically include sample preparation, separation, data collection, feature finding, statistical analysis and compound identification. The aqueous phase extracts for control (NRBC) and infected red blood cells (IRBC) were first analyzed by LC/MS-ESI TOF in both (+/-) ion modes. The data files were processed in MassHunter Qual using (1) a naïve, un-targeted data mining approach, and (2) a targeted data mining approach using a list of known compounds. To accomplish (1), the MFE algorithm was used which processes data files very quickly and is the method of choice for an un-targeted discovery workflow. Each of the thousands of extracted compound chromatograms was automatically drawn by the software and user-inspected before a Compound Exchange File (*.cef*) was generated for each sample file.

After data processing in MassHunter Qual, the *.cef* files were imported into MPP for analysis. The "Guided workflow" was selected from the menu, which is a simple way to follow in a step-by-step fashion a series of analytical steps including feature alignment, data file grouping, filtering, quality control, statistical analysis, and compound identification. Pathway analysis was performed as part of the "Advanced workflow".

Determining variability between various conditions for NRBC and IRBC groups

Data filtering:
The data was filtered to create entity lists based on the number of entities detected in at least one condition of the NRBC and IRBC sample replicates. The results are summarized in **Table 3**, which reveal the average number of entities per group in ESI+, ESI- and APCI+ modes, across pH 2, pH 7 and pH 9 extracted samples. For a given condition (i.e. pH), filtered entity lists were created based on the number of entities that passed the 100% or 75% sample replicate thresholds. The particular threshold to use depends on the number of replicates. Due to the limited number of biological replicates in this experiment, the entity list using stringent filter conditions (100%) was used as input for the subsequent step, "Filter by sample variability". In this case, entities displaying a coefficient of variation (CV) of the signal abundances of greater than 100% per condition were filtered away. This ensured that only reproducible data were analyzed in subsequent steps.

Condition	pH2_ESI neg	pH7_ESI neg	pH9_ESI neg	pH2_ESI pos	pH7_ESI pos	pH9_ESI pos	pH2_APCI pos	pH7_APCI pos	pH9_APCI pos
NRBC	1433	915	1267	2029	1401	2626	1128	463	500
IRBC	1327	1151	1232	2042	1542	2661	665	460	652
Rel Freq. 75%	1660	1282	1412	2175	1838	2822	916	502	679
Rel Freq. 100%	1259	917	1086	1550	1247	2086	486	342	483
CV < 100%	781	547	649	936	662	1083	284	235	271

Table 3. The number of features detected in each condition, NRBC and IRBC, is summarized for samples extracted under different solvent pHs, using both ESI and APCI sources and polarities

Principal component analysis (PCA):
PCA highlighted the difference between NRBC and IRBC samples extracted under the various pH conditions. **Figure 2** shows the distribution of the samples along the 3 major principal components. The most significant difference observed was the separation of samples based on solvent pH, confirming a previously published observation for RBCs[1]. Thus, these results suggest that one can maximize metabolite coverage by including additional aqueous solvent extractions at pH 2 and pH 9, as well as pH 7. Furthermore, a closer inspection of the PCA plot for pH 7 extracts, by rotation about the axes, highlighted clear separation between IRBC and NRBC samples.

Figure 2. Principal Component Analysis of metabolite data from biological replicate samples highlights variability in abundance profiles between NRBC and IRBC for pH 7 extracted samples.
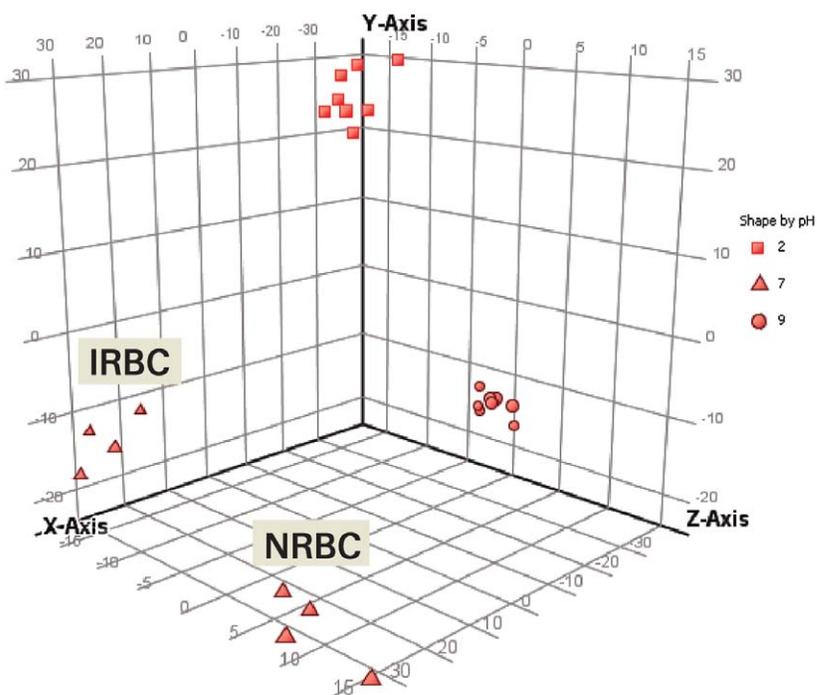
Identify variability between groups with Principal Component Analysis

Determining significant abundance differences of compounds across groups

Identification

MPP includes a convenient, integrated ID Browser that enables accurate mass compound matching of the list of differential metabolites to the METLIN database. For masses that did not match a compound in the database, an Agilent empirical formula calculator, called Molecular Formula Generator (MFG), was used to calculate the formula with the 'best' match based on the mass accuracy of the isotope masses and their ratios. The compound list displayed results for database matches < 5 ppm. In addition to an MS spectrum and structure, the base peak abundance and the abundances for all the isotopes, including any specified adducts, are summarized

(**Figure 3**). One of the compounds, *m/z* 175.1190, matched the amino acid arginine. After identifying compounds using the ID browser, MPP automatically annotated the entity list and projected the compound names onto the graphical plots and pathways for visualization.

Statistical analysis

Since PCA had shown group differences based on extracted metabolite abundance profiles when samples were extracted at different pH values, the results of each pH extract were analyzed independently to find entities with statistically significant abundance differences between NRBC and IRBC. A useful tool for visualizing differential abundance between any two different conditions is the Volcano plot, which provides a visual summary of *p*-values and fold-change values,

adding statistical significance to the interpretation of the results. As an example, for the ESI+, pH 7 extracted samples, the results for a differential abundance analysis of NRBC and IRBC are summarized in **Figure 4**, where the Volcano scatter plot is shown using Fold Change (FC) and *p*-value on each axis. A summary table for all entities revealed the number of differential entities based on FC and *p*-value. For example, there were 69 entities that had a $p < 0.05$ and FC > 2. Inspection of a separate table provided details of the actual FCs, *p*-values and the direction of regulation for each significant entity, based on the results of an unpaired t-test between NRBC and IRBC samples. From this table we were able to see that the result for one of the 'identified' compounds, arginine, was highly significant, with an absolute FC=16 and down regulation. This means that using MFE, its presence was detected in NRBC samples and not detected in IRBC samples.

ID Browser in Mass Profiler Professional facilitates identification of significant abundance differences
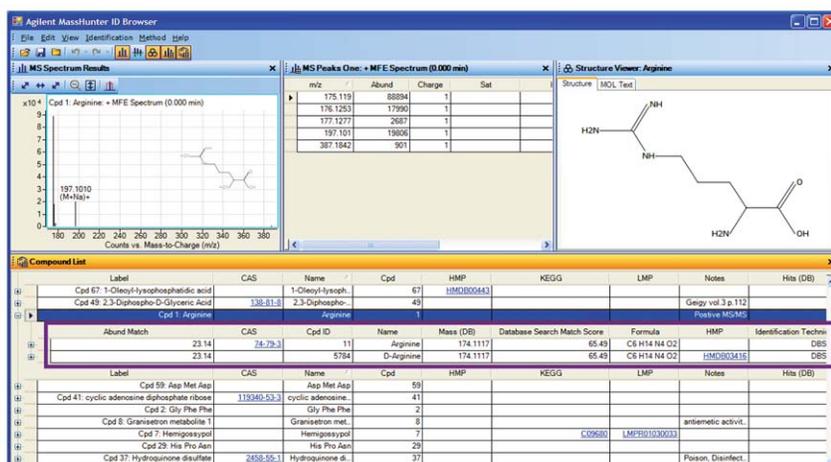

Figure 3. ID Browser results for a list of differential compounds reveals METLIN database matches for several compounds, including arginine. MassHunter ID Browser was used for identification (METLIN database annotation) of compounds from an entity list generated in MPP. The result for mass 174.1117 matched (< 5 ppm) a single compound, arginine, in the METLIN database. In addition to the formula, database match score, Name, KEGG and CAS ID, the figure for the mass spectrum, the isotope abundances associated with the base peak, and a compound structure viewer based on the MOL file are included.

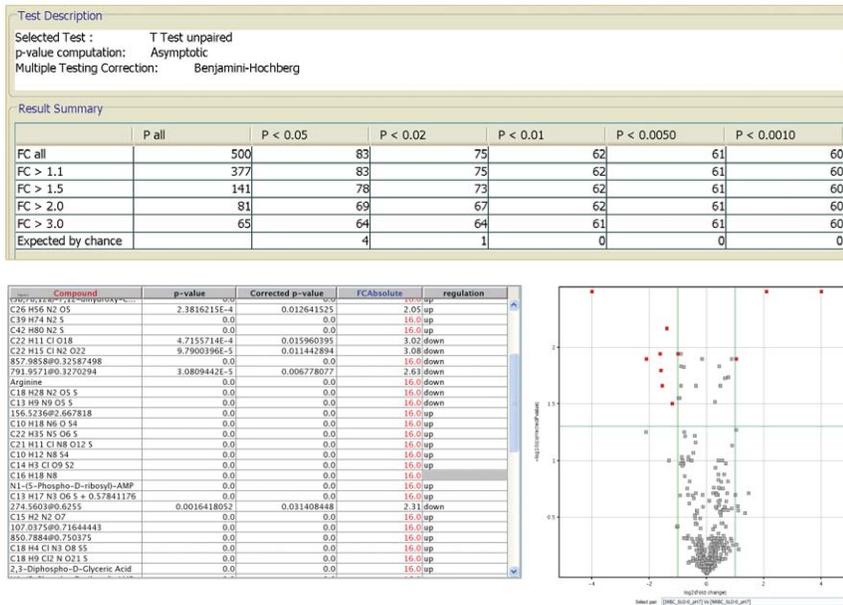**Volcano plots help visualize differential abundance**

Figure 4. The results of a t-test between NRBC and IRBC samples is displayed as a Volcano Plot, revealing the fold-change in abundance as well as p-value of significance for each mass entity in a list of compounds that were 'identified' using ID Browser. The p-value, fold change, and directionality of regulation are also summarized in a table. For example, arginine levels appear to be significantly reduced or absent in IRBC based on the result of $\log_2$ transformation of the abundances, expressed as a ratio.

Hierarchical cluster analysis of entities and conditions

Hierarchical cluster analysis was performed to organize the compounds or entities and replicate samples in the two conditions into clusters based on the similarity of their abundance profiles. This resulted in a hierarchical tree for the identified entities and a hierarchical tree for the conditions, organized into a heat map. The heat map confirmed the separation of the sample replicates for the two conditions. **Figure 5** includes the differential abundance pattern for arginine between the IRBC and NRBC groups in the tree, with the branches colored red and blue, respectively.

Hierarchical cluster analysis organizes entities based on similarity of abundance profiles

Figure 5. Hierarchical cluster analysis organized the replicate samples into two distinct clusters based on the similarity of entities' abundance profiles for the NRBC and IRBC conditions. Clustering analysis by entity profiles across sample groups allows for the grouping together of compounds that have similar abundance profiles. This may reveal interesting relationships, as entities that exhibit similar behavior across a set of experimental conditions may share similar reaction pathways.
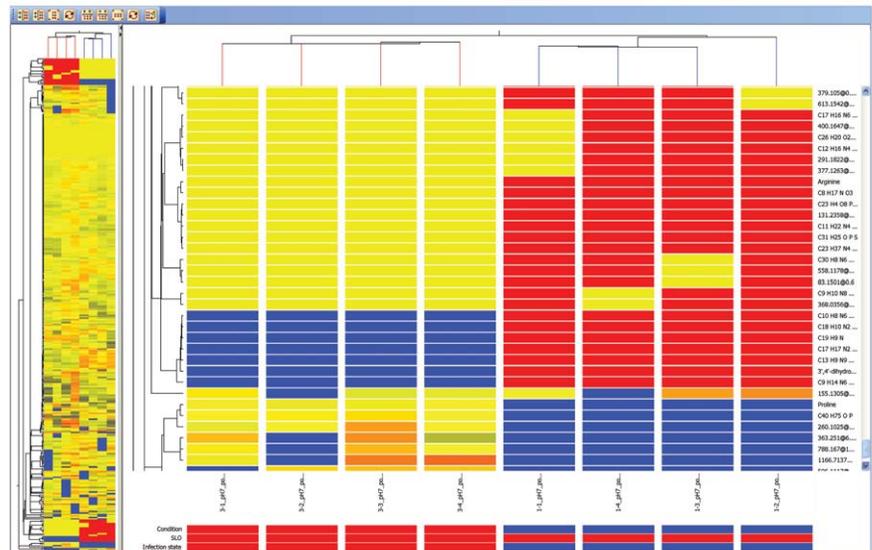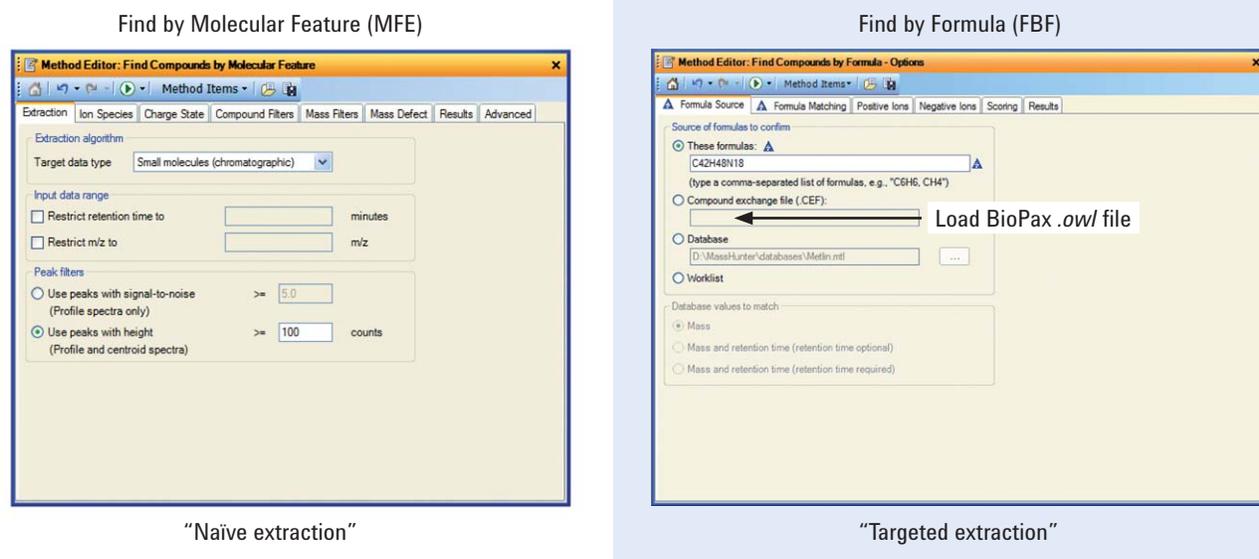
Targeted data mining of TOF data

Feature extraction algorithms that are un-targeted or naïve can sometimes compromise detection sensitivity. Therefore, we also analyzed the data with the more sensitive EIC-based, "Find by Formula" or FbF algorithm for targeted data mining. **Figure 6** shows two different method editors in MassHunter Qual that are used for un-targeted or naïve data mining (left panel) and targeted data mining (right panel). Targeted data mining of TOF data was achieved in a couple of different ways: (1) a list of compound formulas was created based on literature associated with malaria. The formulas were entered into a new METLIN personal compound database (PCD) for malaria. The database was then queried from within MassHunter Qual against all the sample files to find any matches. The 'best' EIC was integrated and *.cef* files were generated for analysis in MPP; (2) A pathway-specific approach was used in which, for example, the arginine biosynthesis pathway was downloaded in Biopax format (*.owl*) from Pathway Commons: http://www.pathwaycommons.org/. The file was loaded into MassHunter Qual and each formula in the pathway was automatically queried against each data file. The masses corresponding to the formulas for both citrulline and ornithine, which participate in the arginine biosynthetic pathway, were found in IRBC and NRBC data by targeted analysis.

Two algorithms for data mining increase the breadth of peak finding

Find by Molecular Feature (MFE)

"Naïve extraction"

Find by Formula (FBF)

Load BioPax *.owl* file

"Targeted extraction"

Figure 6. The algorithms used in MassHunter Qual for finding or extracting the peaks in data acquired in un-targeted mode are shown. The peaks were extracted in either a "naïve" or "targeted" fashion using Find by Molecular Feature (MFE) and Find by Formula (FbF), respectively. The use of two independent methods increases the comprehensiveness of peak finding. The software allows one to change variables for each algorithm located under the different tabs. Furthermore, for targeted analyses the formulas from 'Pathway Commons' for a specific set of compounds were also loaded in an .owl file format, belonging to the citrulline biosynthetic pathway (which includes arginine).

Pathway analysis

Compound annotation using the integrated ID browser in MPP enabled the use of integrated pathway software (Pathway Module) to mine interaction databases provided by Agilent that are created using natural language processing of published literature. It was used in conjunction with the imported Biopax formatted arginine biosynthesis pathway. The pathways were searched using an entity list of differential entities to determine which pathways might be relevant. An interactive pathway viewer revealed the average compound abundances for arginine by sample group (**Figure 7**).

MRM assays by triple quadrupole LC/MS/MS for pH 7 extracted metabolites

Finally, once several members of the arginine biosynthesis pathway were identified, MRM assays, using the parameters shown in **Table 4**, were performed on the triple quadrupole LC/MS platform for arginine, ornithine and citrulline to confirm the discovery LC/Q-TOF findings, and to provide quantitative information for the levels of several amino acids in this biosynthetic pathway. Since the reverse phase LC method used for the initial discovery analysis does not retain highly polar metabolites such as arginine, the MRM assay used a hydrophilic method, based on a silica hydride column operated in the Aqueous Normal Phase (ANP) mode.

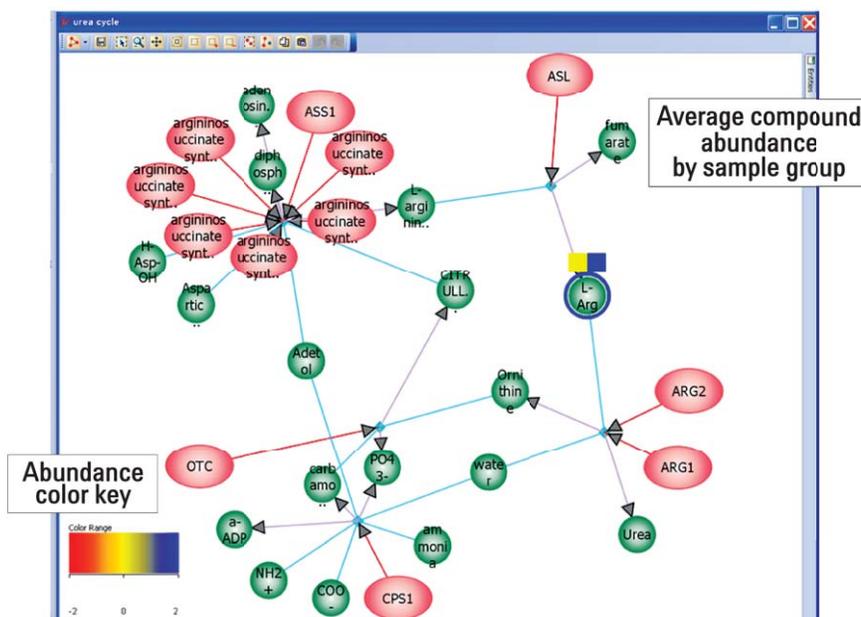**Pathway software facilitates pathway analysis**

Figure 7. Pathway Architect was used to evaluate MS data in biological context by direct navigation between biological pathways and their associated compounds. The average abundance for arginine (Arg) based on the sample group is summarized using an abundance color key in a yellow (IRBC) or blue (NRBC) block.

Table 4. MRM assay parameters

Compound	Precursor Ion (m/z)	Product Ion (m/z)	Fragmentor (V)	Collision Energy (V)	Retention Time (min)
Citrulline	176	159	75	6	1.5
Arginine	175	70.1	75	16	2.4
Ornithine	133	116	75	6	2.6

Calibration curves were constructed for arginine, citrulline, and ornithine, using concentrations of 0.1, 1.0 10, 100 and 1000 ng/mL. The **Figure 8** inset shows the calibration curve for arginine, and illustrates the excellent linearity obtained with R^2 values very close to 1.000. Calibration curves for each compound were generated over five orders of magnitude. The three calibration curves were generated using a linear fit, no inclusion through the origin, and a 1/x weighting. **Figure 8** also illustrates the differential quantitative results between NRBC and IRBC based on MRM transitions for arginine (175.0 →70.1), ornithine (133.0 →116.1), and citrulline (176.0 →159.1). A summary table (**Table 5**) lists the relative amounts (pg) of each compound determined from the calibration curves based on a 1 µL injection, the associated standard deviation (SD) for the biological replicates, and a calculated amount of each compound contained in 1 mL of RBC suspension. It is clear that the amount of arginine drops significantly in IRBC relative to NRBC (almost 8-fold). Conversely, the relative amounts of ornithine, and particularly citrulline, were increased in IRBC, indicating a reduction in arginine biosynthesis.

MRM assays on a triple quadrupole LC/MS confirm the abundance findings
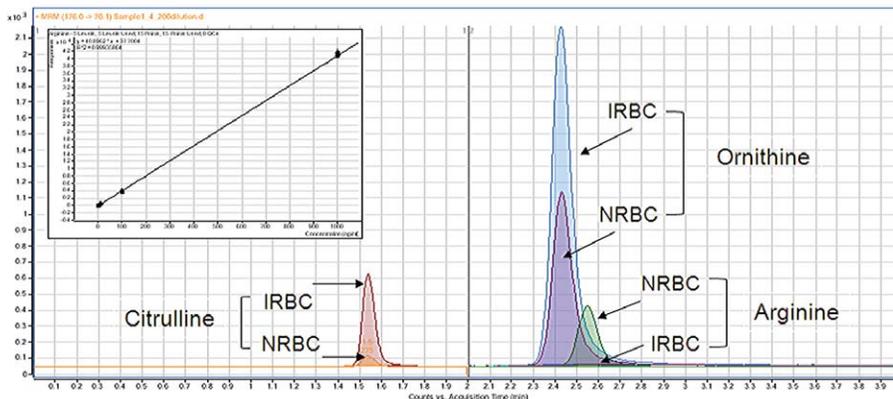
Figure 8. LC/MS/MS analysis of pH 7 extracted metabolite samples on a polar silica hydride column confirms the LC/TOF results. The MRM transitions for arginine, citrulline and ornithine are displayed. Standard curves (inset) were built across a wide dynamic range (five logs), which enabled the estimation of the concentrations for each compound in each RBC sample.

Compound	NRBC			IRBC			Fold Change
	pg	SD	pg/mL	pg	SD	pg/mL	(NRBC/IRBC)
Citrulline	2.3	1.6	187.1	26.2	2.4	2093.1	0.1
Ornithine	48.4	6.9	3873.8	75.5	9.0	6041.5	0.6
Arginine	35.6	19.8	2846.9	4.5	6.5	356.5	8.0

Table 5. Relative amounts of each compound extrapolated from the calibration curves

Conclusion

A global metabolomics workflow was used to initially identify several compounds, including arginine, citrulline and ornithine, as metabolites whose biosynthesis was altered in *in-vitro* cultured erythrocytes infected by the malaria parasite. This example demonstrates that separate algorithms in MassHunter Qualitative data analysis software can be used for finding compounds acquired in an un-targeted or global (MS-only) mode. The data were extracted in an "un-targeted" (i.e. no a priori knowledge of what compounds are present in the sample) fashion, and in a "targeted" fashion using a list of known compounds that were expected to be present/absent in the samples. Initially, the ID Browser in Mass Profiler Professional was used to match mass 174.1117 to arginine in the METLIN database. This compound was present in NRBC samples only. MPP's Pathway Module was then used to determine biological relevance. Insights gained from the pathway analysis led to recursive interrogation of the data in a targeted fashion. Once a preliminary conclusion was reached from the discovery (global) metabolomics experiment, a targeted analysis was designed for the triple quadrupole LC/MS system to quantify the metabolites identified in the initial data mining experiment: arginine, citrulline and ornithine. This workflow demonstrates how a discovery based approach for mining data can be subsequently confirmed in a quantitative fashion. To validate the results on a larger scale, the data could also be imported into MPP to verify the biological conclusions on a larger population set.

References

1. T. R. Sana, K. Waddell, S. M. Fischer, "A sample extraction and chromatographic strategy for increasing LC/MS detection coverage of the erythrocyte metabolome," *J Chromatogr B*, 871:314–321, **2008**.

Agilent Technologies