A Q-TOF Generated, Metabolomics-Specific LC/MS/MS Library Facilitates Identification of Metabolites in Malaria Infected Erythrocytes

Application Note

Authors

Theodore R. Sana, PhD
Steven M. Fischer
Cindy Lai
Agilent Technologies, Inc.
Santa Clara, CA, USA

Dr. Sandra Chang
Professor of Tropical Medicine
John A. Burns School of Medicine,
Honolulu, HI

Abstract

As the premier vendor for metabolomics, Agilent Technologies has invested a great deal of effort to empower our customers with new metabolite identification tools. This study demonstrates the use of Agilent's new MS/MS spectral metabolomics library to elucidate the identification of several amino acids that are important in malaria metabolism.

Agilent Technologies

Introduction

Compound identification is the single most important challenge facing the metabolomics community today. A high level of domain knowledge and training is required for *de novo* MS/MS structure elucidation. Broader application of *de novo* MS/MS interpretation ultimately requires advances in the software used to interpret MS/MS spectra. The challenge is all the more pressing because an increasing number of scientists are realizing the potential for using the results of metabolomics studies in diverse applications, such as the analysis of biochemical pathways in functional genomics studies. Any conclusions that are made based on the observed phenotypic changes to specific compounds in biochemical pathways require the highest level of confidence in compound identity. Currently, there are only a few curated, publicly available accurate mass metabolite databases, and no comprehensive MS/MS libraries.

Research on the metabolism of the malaria parasite has led to a greater understanding of parasite biology and can inform the development of new drugs to control this disease. In our untargeted study, we performed global metabolomics analysis to provisionally identify metabolites in cultured *Plasmodium falciparum* infected (IRBC) and noninfected (NRBC) human red blood cell extracts. We also used the Agilent LC/MS/MS library of over 2,200 standards to confirm the identity of several provisionally identified amino acids.

Methods

Untargeted Discovery Analysis of Metabolites

Metabolomics workflows typically include sample preparation, separation, data collection, feature finding, statistical analysis, and compound identification. Table 1 shows the conditions used to collect data on an Agilent LC/MS/MS Quadrupole Time-of-flight (Q-TOF). Data was acquired in targeted MS/MS mode, with the collision cell energy (CE) set to 10 and 20 eV in positive ion mode.

For profiling experiments, data files were typically processed using a naïve or untargeted peak finding approach. To accomplish this in Agilent MassHunter Qualitative B.04 software, the **Find by Molecular Feature** (MFE) algorithm was used to process LC/MS data files very quickly, while finding most of the compounds in the data. A Compound Exchange File (*.cef*) of the results was subsequently generated and imported into Mass Profiler Professional (MPP) for differential and statistical analysis. Data filtering methods were applied to remove compounds that were not detected in at least one of the sample groups.

Table 1. LC and Q-TOF MS/MS Conditions.

LC 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	10 µL
Autosampler temperature	4 °C
Needle wash	3 s 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/MS Conditions	
Quad resolution	High resolution
Ion mode	Both positive and negative
Drying gas temperature	325 °C
Drying gas flow	9 L/min
Nebulizer pressure	45 psig
Capillary voltage	4000 V (positive mode) / 3500 V (negative mode)
Fragmentor	140 V
Skimmer	65 V
OCT1RFVpp	750 V
Isolation width	~ 1.3 m/z
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	3.35 spectra/s
TOF spectra mass range	25 to 1000 m/z
Collision energy (eV)	10 and 20
Data storage	Centroid
Threshold	100 (MS) and 5 (MS/MS)
Instrument mode	Extended Dynamic Range
MS/MS Library Match Settings	
Precursor ion	+/- 10 ppm +/- 2 mDa
Product ion	+/- 35 ppm +/- 2 mDa
Search mode	Reverse and forward

Results

Application of the Wilcoxan *t*-test to NRBC and IRBC groups in MPP, resulted in a list of statistically differential entities ($p < 0.05$). A profile plot of these features is shown in Figure 1, with Log_2 normalized abundance values plotted on the Y-axis. Next, the METLIN ID browser was launched in MPP so that masses from this list of significantly differential entities could be matched to compounds in the METLIN database for provisional identification using accurate mass. Compound annotations were automatically updated for all metabolites with provisional database matches. The molecular formula ($C_6H_{14}N_4O_2$), corresponding to arginine, was found to be present in NRBC but absent or barely detectable in IRBC samples.

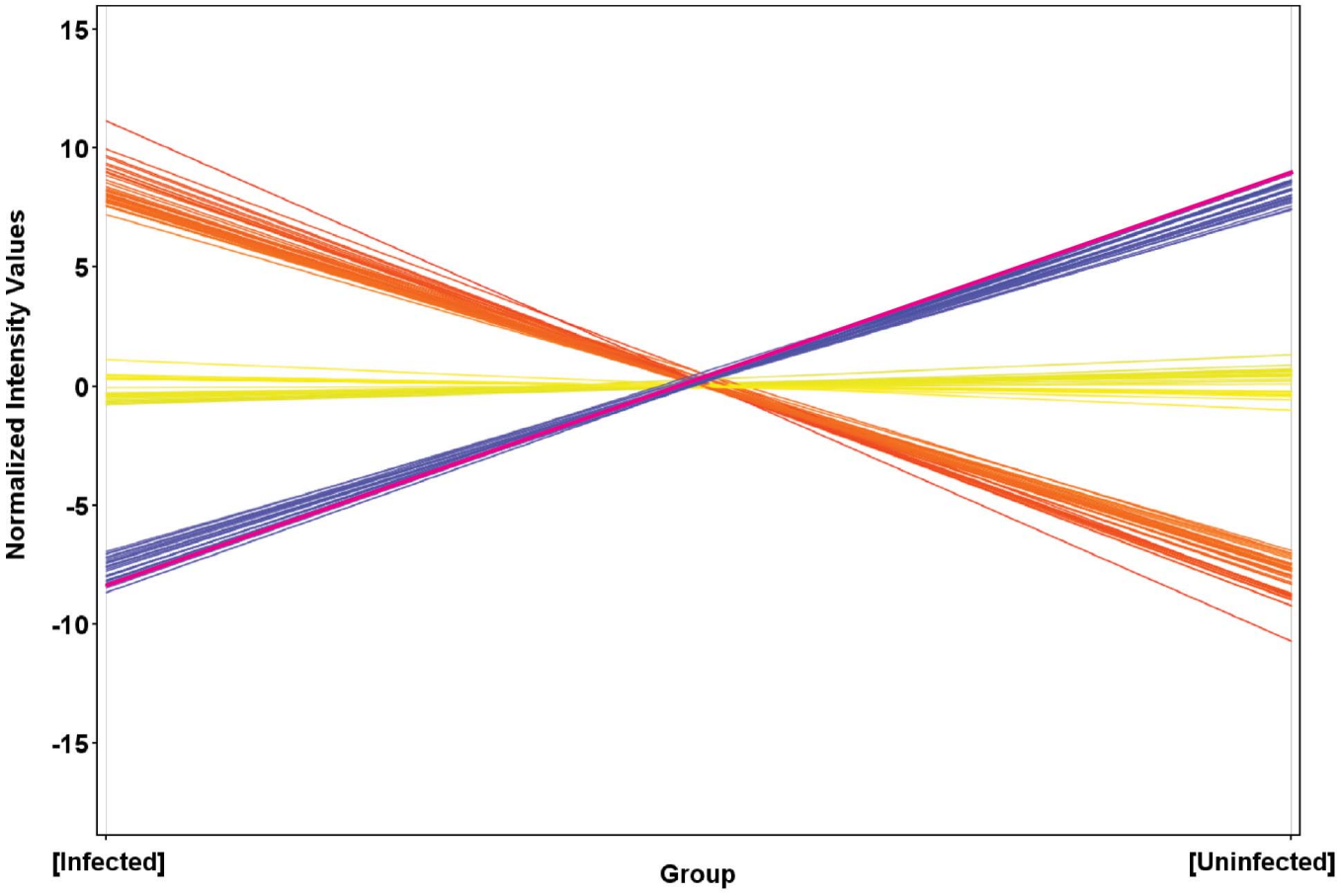


Figure 1. An MPP Profile plot after Wilcoxan *t*-test analysis. Significant differences were seen in normalized intensity values for many entities between infected and uninfected groups of samples. Entities are colored by normalized abundance: blue lines correspond to those having low abundance, yellow lines are intermediate in abundance, and orange lines represent high abundance. The slope of the line determines if the compound is higher or lower in the infected condition. The single pink line corresponds to arginine.

Confirming Identification Results with the New Agilent METLIN Metabolomics LC/MS/MS Library

We have previously reported LC-QQQ results for several compounds in the arginine biosynthetic pathway.[1] Ornithine and citrulline also participate in arginine metabolism, and were identified separately using a targeted Find by Formula approach. All three compounds were also tentatively identified as present in the sample by matching their retention times to the accurate mass and retention time database (AMRT).

However, since these compounds elute near the void volume of the reverse phase method, the retention time value is informative but not sufficiently conclusive for identification. Therefore, the availability of MS/MS library spectra for compound matching as an orthogonal variable provides very strong evidence of compound identity. Fortunately, introduction of the Agilent METLIN Personal Compound Database and Library (PCDL) now enables the user to perform compound identification via MS/MS library matching. The METLIN PCDL contains MS/MS spectra for over 2,200 compounds that were collected in both ESI positive and negative ion modes. MS/MS spectra for all 2,200 standards, including arginine, ornithine, and citrulline were produced from the isolated mono-isotopic ion. Fragmentation data was collected at three collision energies: 10, 20, and 40 eV. In addition, all MS/MS spectra were curated for quality prior to loading into the library: the fragment ions were confirmed and mass corrected, the noise ions were removed, and the spectra were manually reviewed. To confirm the identities of arginine, citrulline, and ornithine in NRBC and IRBC samples, an inclusion list (Figure 2) was generated for analysis on the Agilent 6520 Q-TOF with targeted MS/MS mode analysis.

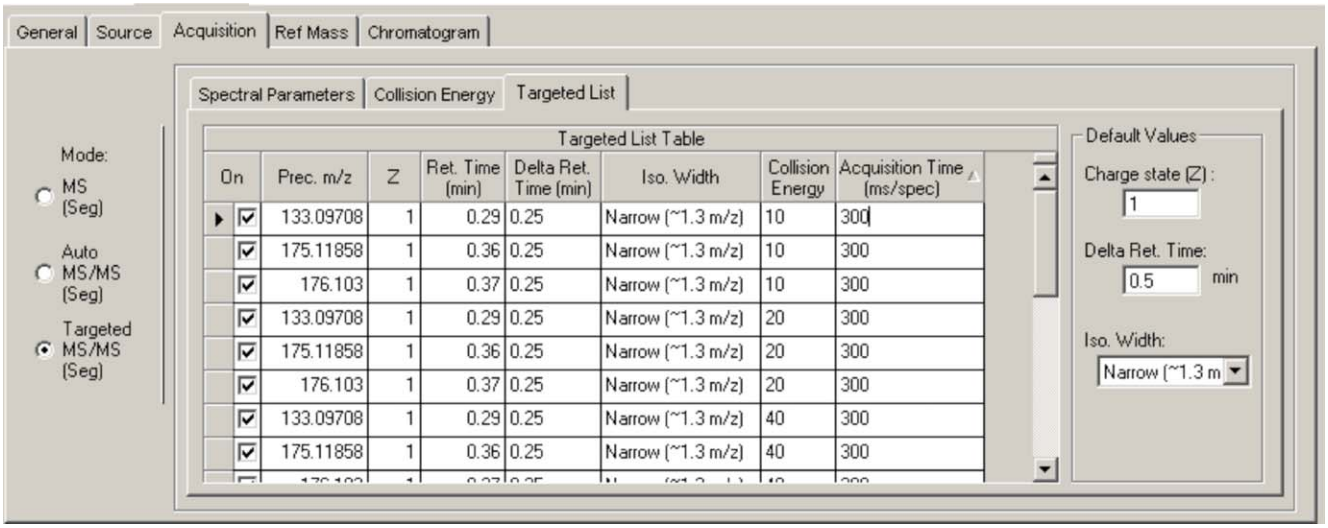


Figure 2. Q-TOF acquisition inclusion list.

The resulting acquired MS/MS sample spectra were matched to the library using reverse and forward library searches in MassHunter Qualitative software. A reverse search uses only the ions present in the library spectra to test for the presence of the library ions in the observed spectrum. It is used to detect the presence of a compound in a MS/MS spectrum. In contrast, a forward search uses all ions above a threshold score in the acquired data to search against the library spectrum, testing for the purity of the acquired spectrum as well as its fit to the MS/MS library spectrum. It is used to highlight a noisy spectrum, since that may imply the observed spectrum is impure and that the reverse search match is questionable.

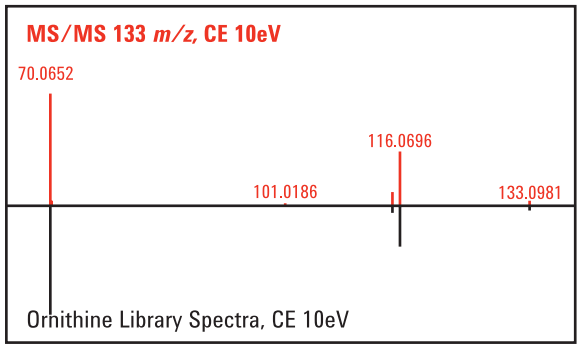
Accurate mass matching of the precursor ion filters out most compounds in the database, while the MS/MS spectral match differentiates the few remaining compounds that are possible identifications. Observed ions that fall within the user-defined tolerance are matched to the corresponding library ion. A spectral match score is determined using a dot product calculation that compares the relative intensity of the observed ions to the library spectrum. The search results are ranked based on a match quality score which ranges from 0 to 100. A reverse search score above 80 is considered a good match. A forward search score below 50 is considered uncertain and would necessitate a review of the reverse match.

Targeted MS/MS acquisition was performed for each ion (133, 175, and 176 *m/z*) at 10 and 20 eV collision energy (Figure 3). Compound identification was based on the match results for both collision energies. The match results (Figure 3A) for 133 *m/z* were relatively easy to interpret because the signal strengths were high. The strong signal in the acquired sample resulted in a rich fragment spectrum that was easily matched to the MS/MS library (reverse score = 100). The forward score (97 and 99) was also very high, indicating the acquired spectra is pure. Based on the library scores, there was a good compound library match to ornithine.

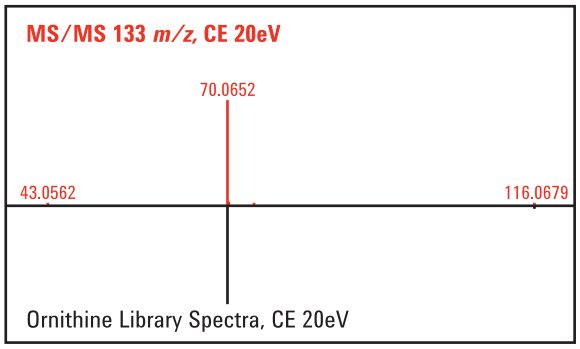
Figure 3B shows that the signal strength for 175 *m/z* was much lower than ornithine. However, a good reverse match was reported for both spectra. Given that the major library ions were matched in both acquired spectra this is a reasonably good match to arginine.

The signal strength for 176 *m/z*, like arginine, is also weak (Figure 3C). It is a good example of the benefit of having acquired spectra at two different collision energies. With four common ions in the 10 eV spectrum and one ion in the 20 eV spectrum, it shows a good match to the library despite lower reverse scores. Furthermore, when one also considers the mass accuracy settings for MS and MS/MS, 10 ppm +/- 2 mDa and 25 ppm +/- 5 mDa respectively, the results provide a good degree of confidence in confirming the presence of citrulline in the samples.

A

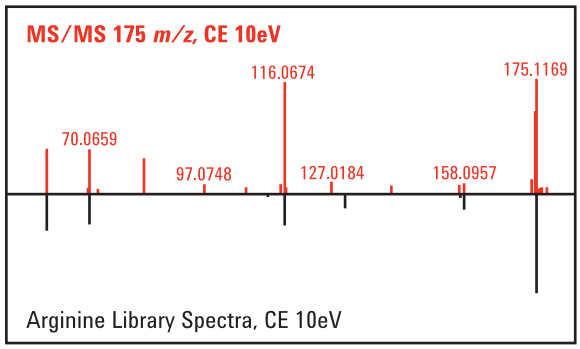


Maximum MS/MS Signal: 168333 counts
- Number of library ions: 4
- Reverse score: 99.4
- Forward score: 99.4

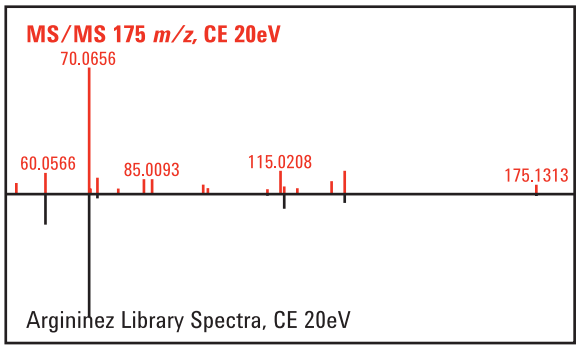


Maximum MS/MS Signal: 44566 counts
- Number of library ions: 3
- Reverse score: 100
- Forward score: 96.8

B

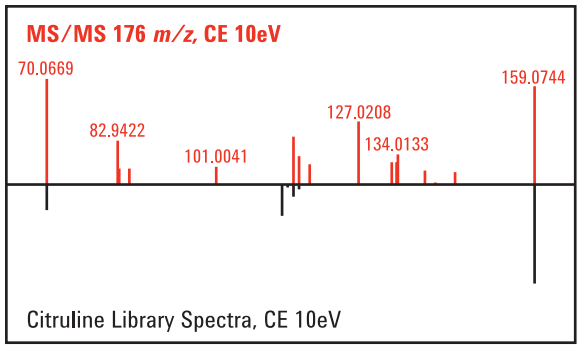


Maximum MS/MS Signal: 526 counts
- Number of library ions: 11
- Reverse score: 83.2
- Forward score: 41.8

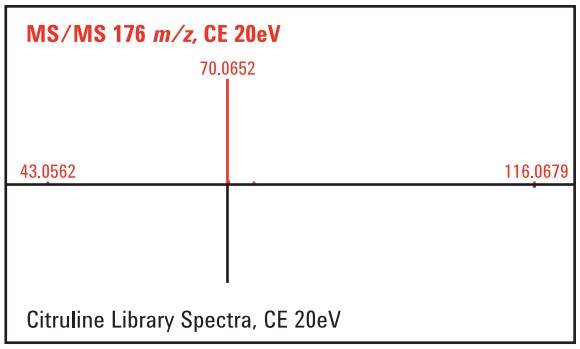


Maximum MS/MS Signal: 101 counts
- Number of library ions: 9
- Reverse score: 99.98
- Forward score: 96.75

C



Maximum MS/MS Signal: 209 counts
- Number of library ions: 8
- Reverse score: 67.9
- Forward score: 21.3



Maximum MS/MS Signal: 456 counts
- Number of library ions: 9
- Reverse score: 84.8
- Forward score: 31.0

Figure 3. Mirror image MS/MS matches between acquired (top trace) and library spectra (bottom trace) for ornithine (A), arginine (B), and citrulline (C) at 10 and 20 eV, as well as forward and reverse match scores for each compound.

Conclusions

A targeted LC-QTOF MS/MS analysis was performed on arginine, citrulline, and ornithine that had been provisionally identified by our global untargeted analyses using LC/MS data mining experiments. This study demonstrates the powerful capabilities of the METLIN PCD/PCDL to aid in confirmation of provisional identifications through accurate mass matching on both MS and MS/MS levels. MassHunter Qualitative software provides the user with all the necessary tools to quickly query the database and library for high confidence matches to metabolites that were found to be significantly different between NRBC and IRBC samples on the basis of both accurate mass and retention time. The METLIN PCDL search results provide strong confirmation for the presence of these amino acids in the sample.

References

1. Fischer, S., and Sana, T., An LC/MS Metabolomics Discovery Workflow for Malaria-Infected Red Blood Cells Using Mass Profiler Professional Software and LC-Triple Quadrupole MRM Confirmation. *Agilent Application Note 5990-6790EN.*

www.agilent.com/chem/metabolomics

Published in USA, September 30, 2011
5990-9132EN

Agilent Technologies