Published in final edited form as:

Nat Protoc. 2013 March ; 8(3): 451–460. doi:10.1038/nprot.2013.004.

Liquid Chromatography Quadrupole Time-of-Flight Characterization of Metabolites Guided by the METLIN Database

Andrew W. Schultz*, Junhua Wang*, Zheng-Jiang Zhu, Caroline H. Johnson, Gary J. Patti\$, and Gary Siuzdak*

*Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037

\$Departments of Chemistry, Genetics, and Medicine, Washington University, 660 South Euclid Ave St. Louis, MO 63110

Abstract

Untargeted metabolomics provides a comprehensive platform to identify metabolites whose levels are altered between two or more populations. By using liquid chromatography quadrupole time-of-flight mass spectrometry (LC-Q-ToF-MS), hundreds to thousands of peaks with a unique m/z and retention time are routinely detected from most biological samples in an untargeted profiling experiment. Each peak, termed a metabolomic feature, can be characterized on the basis of its accurate mass, retention time, and tandem mass spectral fragmentation pattern. Here a 7-step protocol is suggested for such a characterization by using the METLIN metabolite database. The protocol starts from untargeted metabolomic LC-Q-ToF-MS data that has been analyzed with the bioinformatic program XCMS, and describes a strategy for selecting interesting features as well as performing subsequent targeted tandem mass spectrometry. The 7 steps described will require 2–4 hours to complete per feature, depending on the compound.

INTRODUCTION

Metabolomics has emerged as a powerful technique to understand the small-molecule basis of biological processes such as those associated with disease pathogenesis^{1, 2}, interactions of microbial communities³, microbial biochemistry^{4, 5}, plant physiology⁶, and drug mode of action⁷ and metabolism⁸. In general, there are two technological platforms used to perform metabolomics which involve either nuclear magnetic resonance (NMR) spectroscopy^{9, 10} or mass spectrometry (MS)^{11, 12}. Although NMR provides unique structural information about metabolites, it suffers from limitations in sensitivity and chemical resolution. In contrast, MS provides less conclusive structural information, but allows for the detection of many more species in a single experiment given its sensitivity and large dynamic range. Each of these technologies has been successfully applied to systematically study metabolites; however, MS methods are more commonly used for comprehensive investigations that are global in scope. The strength of MS-based metabolomics is best realized when coupled to a chromatographic technique such as capillary electrophoresis, gas chromatography (GC), or liquid chromatography (LC), the latter two being the most popular. GC/MS-based metabolomics is a robust, well-established technique^{13–15}. Because of the reproducibility of the chromatography, retention time can be paired with the electron impact (EI)-derived fragmentation spectra¹⁶ and compared against the NIST¹⁷ or Fiehn metabolomic¹⁸ databases

Corresponding authors: Gary Siuzdak (siuzdak@scripps.edu) phone (858) 784-9415 fax (858) 784-9496 Gary J. Patti (gjpatrick@wustl.edu) phone (314) 362-8358.

Author contributions statements: AWS and JW contributed equally to the work being described

to make identifications. However, the majority of metabolites must be derivatized to make them more volatile and more thermally stable, which introduces a source of error and complicates identification¹⁹.

In the last decade, LC/MS-based analysis has moved to the forefront because of its ability to analyze and identify underivatized and thermally labile metabolites. In contrast to EI, electrospray ionization (ESI)²⁰ (and to a lesser extent atmospheric pressure chemical ionization²⁰) is a soft mechanism to ionize molecules, leaving the molecular ion intact. There are two major approaches to LC/MS based metabolomic experiments, the targeted²¹⁻²³ and untargeted²⁴⁻²⁷ analysis. In untargeted metabolomics one tries to observe as many unknown and known metabolic peaks as possible, comparing the ion intensity between the same peaks present in two or more groups of samples. The disadvantage of this technique is that it is not optimized for a specific metabolite and is less quantitative. The advantage is that it provides an opportunity to observe a large number of known and unknown metabolites, which may provide novel insights into a biological system^{3, 5, 28}. Coupled to a high-resolution mass spectrometer²⁹, such as a time of flight (ToF)^{22, 30}, Orbitrap^{31, 32} or a Fourier transform ion cyclotron resonance (FT-ICR)³³ instrument, high mass accuracy can be obtained to make identifications of metabolites. The high mass accuracy greatly reduces the potential molecular formulas corresponding to one metabolic peak. However, the problem with this approach is that there may be several molecular formulas that are appropriate for the accurate mass data (depending on the resolution of the instrument), and numerous potential isomers for each molecular formula. Adding a fragmentation mechanism provides some structural information. When combined with the high-resolution precursor ion, provides the information needed to characterize a metabolite by liquid chromatography quadrupole time-of-flight (LC-Q-ToF) analysis to a single structure or a narrow set of structures (see limitations below). Therefore, matching the accurate mass and fragmentation data with standard MS/MS spectra in METLIN database³⁴ (metlin.scripps.edu) provides a more convincing identification than accurate mass alone. Retention times, relative to other metabolites of known identity and similar structural class, also support the structural determination. This protocol increases the power of LC-MS based untargeted metabolomics to provide new insights into biological processes by detailing methodologies that provide a more rigorous characterization of metabolites.

Q-ToF based characterization of metabolites

In this protocol, a methodology is presented in which a LC-Q-ToF instrument in combination with the METLIN database³⁴ (metlin.scripps.edu) can be utilized to characterize metabolites (a flow chart is depicted in Figure 1). The Q-ToF provides the ability to collect both high-resolution precursor and fragmentation data, facilitating the characterization of metabolites. When used in conjunction with the METLIN database, which provides the user with the ability to search for the precursor ion, its fragments and neutral losses, the characterization of metabolites is highly augmented. In addition, METLIN is the largest curated database of high-resolution tandem mass spectra, covering over 5000 metabolites, thus identification of metabolites is greatly facilitated. The fragmentation spectra are essential for elucidation and confirmation of unknown metabolites. Characterization cannot be completed until this is performed. The characteristic fragments produced are equivocal to a metabolite fingerprint and when matched with the retention time and fragmentation of an authentic standard can confirm the identity. One of the advantages of tandem MS in the Q-ToF is that collision energies can be adjusted to enhance or decrease the degree of fragmentation therefore revealing more information about the metabolite. Some metabolites however do not fragment well or when an adduct is present. The adduct stabilizes the ion and can give limited fragmentation, but trying different ionization strategies or solvent mixtures can ameliorate this.

Untargeted metabolomics begins with an initial profiling experiment, often where two or more sample groups are profiled via LC-MS and statistically compared, with only the dysregulated metabolites being characterized^{2-4, 6, 28, 35}. There are a few exceptions in which only one sample group is analyzed in studies characterizing as many metabolites as possible in one biofluid^{36, 37}. Two excellent protocols are available for LC-MS profiling experiments in urine³⁸, and plasma and serum¹⁴. These protocols can be easily adjusted to other sample types. The key to obtaining good results is to carefully design the experiments so that there are enough biological replicates to make the results statistically significant, i.e. they must not be underpowered. Appropriate power calculations must be carried out first to determine the sample size which will have a statistically significant effect³⁹. There are a number of factors which can need to be considered such as biological variation, sample preparation and others discussed in more detail by Brown et al.⁴⁰. Depending on the biological variability of the system, we recommend the minimal numbers of each sample group are 4-6 for cell culture, 6-8 for animals, and 10 or more for humans, respectively. After analysis of the initial profiling data by using a peak alignment and statistical analysis package, such as XCMS⁴¹, a list of dysregulated metabolic peaks with a retention time and *m/z* will be generated. The protocol reported here starts to analyze and identify the list of dysregulated features one by one using the following 7-step procedure.

Limitations of this approach

Disclaimer—Many of the limitations listed below can be mitigated using specialized MS techniques and thus may not impose real challenges; however the following limitations discussed are important points to consider when carrying out general approaches to metabolomics before optimizing to specific species or problems.

Firstly, low-abundance ions can be hard to identify if the precursor ion intensity is low (generally below 5000 counts for an Agilent Q-ToF) it will be difficult to obtain the high-quality fragment spectra needed to support a structural assignment. But, overall there is not a problem with sensitivity when using LC-MS based approaches and examples of high sensitivity mass spectrometry-based metabolite identifications include 3.5 femtomol of DMS per mg of dorsal horn³⁵, or upper attomole range in the analysis of *Methylobacterium extroquens* AM1⁴².

Secondly, MS-based analysis provides little if any information about stereochemistry of the metabolites identified and is often insufficient to determine the positions of double bonds in acyl tails. However, some specialized techniques have been used involving ion mobility⁴³, addition of Li⁺ with multiple rounds of fragmentation⁴⁴ and ozone-induced dissociation⁴⁵. The location of these bonds may be important; for example, isobaric ω -3 or ω -6 isomers of a lipid can have significantly different biological roles⁴⁶.

Thirdly, isobaric species that co-elute will provide a convoluted mass spectrum making it difficult to characterize either species. MS is prone to ion suppression⁴⁷, therefore, co-eluted species also affect the quantification of molecules and reduce the ability to observe ions that are less capable of ionization in the presence of an interfering metabolite. Furthermore, isobaric and other species with very similar masses could be fragmented together if not well isolated, introducing contamination into the MS/MS spectra and hindering characterization, possibly leading to false negatives. Appropriate chromatographic methods can be developed, which can help resolve different species and reduce some issues with ion suppression. Ion mobility can also aid in the separation of isobaric species in gas phase, which alleviates contamination into the MS/MS spectra that hinders correct metabolite identification.

In addition, in-source fragmentation is sometimes observed for species containing a labile group. It can generate one or more abundant fragments that show a similar level of

dysregulation compared to other peaks at the same retention time⁴². If two or more dysregulated peaks coelute, one must exercise caution that the peaks are not fragments from the same molecule. In Supplemental Figure S1, an example of this is shown where two species (m/z of 339.2892 and m/z 480.3084) with the same retention time are observed to be dysregulated. The peak 480.3085 corresponds to a lysoPE(18:1/0:0) while 339.2892 is a major fragment of this lysoPE, a dehydrated oleoyl (18:1) glycerol. Without recognizing that the lysoPE is the dysregulated metabolite, one may falsely identify the in-source fragment, oleoyl glycerol, as a dysregulated metabolite.

Finally, this approach does not provide an unequivocal identification of a metabolite. It does, however provide a higher level of confidence than high-resolution mass alone and is generally acceptable for publication. A scoring system is under development in our group to quantitatively evaluate the confidence of metabolite identification. For better confidence, standards should be acquired and run on the same instrument with the same instrument parameters. The retention time and fragmentation patterns must then match between the sample and standard to extend the Q-ToF based *characterization* to an *identification*, and if the retention time does not match, it implies the characterization is incorrect. For metabolites in which a higher level of confidence is needed, an orthogonal method should also be utilized to validate the metabolite structure. NMR for example has the benefit of structural identification and accurate characterization and furthermore when hyphenated to LC can be highly effective for metabolite elucidation⁴⁸. Metabolites lacking commercial standards should be chemically synthesized and compared as above for standards^{5, 49}. For some experiments, this level of rigor may be unnecessary depending on the scope of the biological question².

MATERIALS

REAGENTS

Acetonitrile with 0.1% formic acid (Honeywell B&J Brand LC-MS grade) ! AUTION
Acetonitrile is highly flammable

Water with 0.1% formic acid (Honeywell B&J Brand LC-MS grade)

Extracted samples from biofluids, yeast, cells, or animal tissues in autosampler vials

EQUIPMENT

LC-Q-ToF system: UPLC or LC system; Q-ToF mass spectrometer; Column (C18, HILIC, etc.) utilized in initial profiling experiment Instrument method from MS profiling experiment

A PC with an internet connection and a web browser

XCMS output spreadsheet from MS profiling experiment

Spectral files from original profiling experiment

Software for mass spectral analysis provided by instrument vendor (e.g. Agilent MassHunter; AB Sciex PeakView; Bruker Compass; and Waters Masslynx)

EQUIPMENT SETUP

LC-MS instrument setup—To insure a high level of mass accuracy, the instrument should be calibrated before running samples according to manufacturer guidelines. Ensure samples are properly mixed and thawed before placing in autosampler tray. Install mobile

phases, prime system pump and tubing. Install the column and ensure that it is properly equilibrated before injecting samples.

Note—This protocol is mainly based on utilizing an Agilent 1200 series high-performance liquid chromatography (HPLC) system coupled to an Agilent 6538 Q-ToF-MS (Agilent Technologies, Santa Clara, CA) with Agilent MassHunter (Version B.04.00) and XCMS software (current version 1.21.01). There are many other hardware and software combinations that can be used with METLIN, check the instrumentation and software documentation for assistance. For software analysis however it is recommended to use XCMS which can process and analyze data from AB SCIEX, Bruker, Thermo Fisher and Waters hardware. The file formats of these platforms can be seen at <https://xcmsonline.scripps.edu/docs/fileformats.html> with notes on how to convert the files into the appropriate formats.

PROCEDURE

1. DETERMINE ADDUCT AND CHARGE

The total ion chromatogram (TIC) and extracted ion chromatograms (EIC or XIC) should be retrieved from the spectral files from the original profiling experiment. This can be done through data analysis software provided by the instrument vendor. Each instrument vendor has its own software which has similar functions to retrieve the TIC and EICs. Here we used Agilent MassHunter as an example to demonstrate this step. Using MassHunter, open the spectral file for a sample and search for the peak of interest by retention time and accurate mass. The peaks of interest are picked from the XCMS output spreadsheet from the previous MS profiling experiment. In MassHunter, select File > Open Data File to select the data to analyze. The TIC should be displayed as in Figure 2A. Next, select Chromatograms > Extract Chromatograms. In Type: select EIC (extracted ion chromatogram). On the MS Chromatogram tab, set the MS level: to MS and m/z value(s): type in your value. On the Advanced tab, define the single m/z expansion to a symmetric (ppm) value. For this example 496.3409, ± 20 ppm was used. Click OK. The EIC should appear as in Figure 2B, and a peak with a RT appropriate for your peak of interest should be visible. The EIC will also display other species with very similar m/z , pointing out isobaric species that are present. With the Walk Chromatogram cursor selected, click on the EIC at the retention time of your peak of interest. The MS spectrum will appear. Using the Range Select cursor, zoom in on the MS spectra as in Figure 2C. Determine the adduct of your peak. In this case, 496.3409 is likely $[M+H]^+$ since a peak approximately 22 Da (518.3219) is present, which would correspond to the $[M+Na]^+$. Zoom in further on the MS spectra as in Figure 2D and determine the charge for the peak. Since there is a series of isotope peaks approximately 1 Da larger after the most intense peak, it is singly charged. Subtracting the proton provides the neutral mass for this species of 495.3336.

2. INSPECT MS DATA TO DETERMINE IF PEAK IS REAL AND OF SUFFICIENT INTENSITY

Look for co-eluting ions within 1-2 m/z of the peak of interest in the MS spectra, since these may have convoluted the fragment spectra. In Figure 3A a group of peaks is observed in which the separation is insufficient. Several species, such as m/z 480.2805, m/z 480.3082, and m/z 482.2569 are not resolved, and will fragment together, creating convoluted MS/MS spectra, shown in Figure 3B. Once the species m/z 480.3082 are fully resolved by chromatography (Figure 3C), the generated MS/MS spectrum shows good spectral purity. In addition to achieving high quality MS/MS spectra, the feature of interest should have an intensity greater than 5000 (for an Agilent Q-ToF). The intensity requirement is very empirical. Other Q-ToF instruments from different vendors may have different empirical intensity requirements. The parent ion intensity is required to make sure the MS/MS spectra

have enough signal-to-noise ratios (S/N). If the peak is not pure (i.e. with co-eluting species within 1-2 m/z) or intense enough, it will be difficult to obtain good MS/MS spectra and thus a meaningful characterization. All examined features with good chromatographic resolution and peak intensities can be grouped for the MS/MS experiments in step 4.

3. SEARCH PRECURSOR IN METLIN

In your web browser, open METLIN (metlin.scripps.edu). Select Search: Simple. In the mass widow, input the accurate mass value of the parent ion (See Figure 4). Select the charge determined in step 1 or choose other ions as appropriate. The default and maximum tolerance of 30 ppm is generally acceptable for Q-ToF experiments; adjust as appropriate for your specific mass spectrometer. Generally, it is best to use a slightly wider window than the theoretical tolerance for an instrument. For example, 5 ppm would be an appropriate tolerance for an Orbitrap operating at 100,000 resolving mode. Click on Find Metabolites button.

4. PERFORM TARGETED MS/MS

Perform targeted MS/MS for the list of features with good chromatographic resolution and peak intensity as discussed in step 2. Various instruments have different ways to perform targeted MS/MS experiments. Here, we used the Agilent Q-ToF as an example. In the MassHunter, open the instrument method utilized to collect the original MS profiling data. Under the Q-ToF tab, click on tab for targeted MS/MS. Input the m/z value of the feature, set a RT window of at least 1 minute, and isolation to medium, unless co-eluting species dictate a narrower window. More than one feature may be programmed as needed. Save this method, and inject and analyze the sample with the new method. After the data has been collected, continue to step 5.

5. SEARCH MS/MS IN METLIN

Open the newly created data file in MassHunter. To examine the MS/MS spectra, select Chromatograms > Extract Chromatograms, select Type: TIC, and in the MS Chromatogram tab, select MS level: MS/MS and select the precursor ion of peak of interest. Using the Walk Chromatogram cursor to click on individual scans at and near your peak of interest. Inspect the individual MS/MS scans at and around this RT to assess spectral purity. Often a portion of the precursor ion will remain intact, making it easier to identify the spectrum of interest and assess spectral purity. Spectral purity is somewhat subjective, but generally if a similar fragmentation pattern is consistently seen across a few scans, and the MS spectrum lacks co-eluting species within a few m/z , then the spectra can be considered pure and sufficiently intense to identify the peak of interest. Scroll through the metabolites to find ones with MS/MS data (indicated by a View button) (Figure 5). Click on View. The spectrum will appear (Figure 6). Click on individual lines in the spectra table to select a specific precursor and voltage; the appropriate spectrum will appear. You can right click and drag a box to zoom in. Roll your cursor over a spectral peak and the exact mass for the fragment will be displayed along with a predicted structure for that fragment if available. Click Reset zoom in the upper left to zoom back out. Right click and hold move to move the spectral window around the page. To close, click on close in the upper right corner.

6. COMPARE EXPERIMENTAL MS/MS TO METLIN

Compare your experimental spectra with the spectra in METLIN by visual inspection. If the same fragment ions are present in the experimental spectra and the METLIN spectra with very similar intensity ratios, you have a perfect match as seen for phenylalanine Figure 7A, arachidonic acid Figure 7B, and hypoxanthine in positive and negative mode in Figure 7C and D. Hypoxanthine in positive mode (Figure 7C) is a good match since the major

experimental fragments are of similar intensity as the standard, although there is some low intensity contamination. If you found a good match, you can go to step 7. If several high intensity ions are missing or the ratios are significantly different, as seen in Figure 8B where the intensity ratios between the experimental spectra in black is significantly different than the standard spectra in red, you have not found a good match.

7. VERIFY CHARACTERIZATION WITH STANDARD

If you found an exact match between your experimental spectra at both the precursor and fragment levels, then you have characterized the metabolite. Depending on the level of confidence needed in your analysis, you should follow up with additional techniques to support your identification. Techniques such as FTICR-MS or NMR can give you an additional level of confidence, although metabolite concentrations often prevent the use of NMR to characterize metabolites. The highest level of confidence is obtained when standards are synthesized or purchased, and compared by LC-MS/MS to confirm retention time and MS/MS with the same parameters.

TIMING—This protocol should take 2-4 hours depending on the metabolite

TROUBLESHOOTING

Step 1

If it is determined that your metabolic peak of interest is an isotope peak, one must be cautious that this may be a false positive. If your peak is an adduct other than M+H or M-H, one should look back at the original profiling experiment to see if the monoisotopic peak or M+H or M-H is also dysregulated. If it is, complete this protocol with the M+H or M-H ion. If it is not dysregulated, do another simple search in METLIN with the correct adduct selected. As we discussed above (Supplemental Figure S1) in-source fragments should also be checked. These in-source fragments always co-eluted with their parent ions. If the in-source fragment ion is identified, one should look for the parent ion at the same retention time. If the parent ion is also dysregulated, complete this protocol with the parent ion.

Step 2

If a co-eluting metabolic peak is within 1-2 m/z of your ion of interest, it may provide a convoluted spectrum. If you suspect this is the case, you should re-fragment this species with a narrower isolation window. If it is within 1 Da, this may not be sufficient to isolate the species, and you may need to use another approach to identify this peak. If two ions are co-eluting, different chromatographic conditions may allow these two species to be separated as in Figure 3.

Step 2

If you cannot identify the precursor ion, you may want to rerun the sample on the LC-MS, fragmenting at a lower energy. If the precursor is identified, but there is insufficient fragmentation, you may want to rerun the sample fragmenting at a higher energy.

Step 3

If no metabolites are returned from the search, you can increase the tolerance value, or add additional adducts if appropriate. For the ionic metabolites, when doing the search with METLIN database, the “neutral” should be chosen for the “charge” setting. In addition, the isotopic pattern distribution also helps to predict the empirical formula of unknown compounds. Most data analysis tools provided by instrument vendors have this function.

Step 3

As seen in Figure 5, a search for m/z 137.0450 returns back seven hits. The first five hits such as threonate are organic acids and the remaining two hits (allopurinol and hypoxanthine) are more basic metabolites. The two types of metabolites could be differentiated by their retention time and ionization efficiency using positive mode ESI. This helps to narrow down the candidates before comparing MS/MS spectrum. However, to further differentiate allopurinol and hypoxanthine, MS/MS matching is necessary.

Step 4

Recently, new tandem MS techniques such as MS^E (from Waters) and SWATH (from AB Sciex) have emerged. Unfortunately, we have not fully tested MS/MS data acquired from MS^E and SWATH techniques with the METLIN database. Thus, currently we do not suggest to the use of MS^E and SWATH data for METLIN MS/MS spectral comparison.

ANTICIPATED RESULTS

Following this protocol allows one to characterize a peak of interest in an untargeted metabolomic experiment if it is a metabolite found in METLIN, or is an analog of a metabolite in METLIN. Metabolites that are not analogs of known metabolites are difficult to identify with this technique, although this protocol will provide information that would be valuable when used in combination with other analytical techniques. Some cases which have proved challenging when attempting to identify unknown metabolites are discussed below, they include examples of metabolites which have no exact match in METLIN and metabolites that co-elute with other metabolic peaks of similar m/z .

For our first example, the metabolic peak of interest has a m/z of 496.3409 and a RT of 24.5. The ion spectrum is extracted (Figure 2C) from the TIC and upon inspection of the spectrum at m/z 496.3409, another peak is observed at m/z 518.3219, which is 21.981 amu larger. This is characteristic of the $[M+Na]^+$ peak and supports that m/z 496.3409 is the $[M+H]^+$ peak ($Na^+-H^+ = 21.9820$). Also noted in (Figure 2D), two isotope peaks for the m/z 496.3409 peak can be seen, m/z 497.3440 and m/z 498.3455. Since these peaks are approximately +1 and +2 from the $[M+H]^+$ peak, it adds validation that this is a singly charged ion and that m/z 496.3409 is indeed the protonated monoisotopic mass of the molecule.

In order to determine the structure of the species at 480.3082 in Figure 3, caution must be taken due to potential contamination from the species at m/z 479.7786 $[M+2H]^{2+}$, m/z 480.2805, and m/z 482.2569 $[M+2H]^{2+}$. Indeed, when m/z 480.3082 is isolated and fragmented, the spectrum in Figure 3B is obtained which contains both m/z 480.2805 (isotope of m/z 479.7786) and m/z 482.2567 species. In this situation m/z 480.3082 cannot be identified since the MS/MS spectrum is suppressed and contaminated. If chromatography is used to separate these species like shown in Figure 3C, a pure MS/MS spectrum can be obtained for m/z 480.3084 (Figure 3D), which is characterized as LysoPE (18:1/0:0). Use of a narrow isolation window may also be useful to prevent contamination by other species if the mass difference of two species is sufficient.

The characterization of three metabolites, phenylalanine, arachidonic acid, and hypoxanthine are depicted in Figure 7. The simple fragmentation of the experimental phenylalanine (Figure 7A) and the more complex arachidonic acid (Figure 7B) match the standard METLIN spectra in both intensity ratio and accurate mass of the fragments, supporting their identification. The experimental spectrum for hypoxanthine in negative mode (Figure 7D) matches well with the METLIN spectrum, although there is significantly more contamination in the experimental sample than observed in positive mode (Figure 7C).

Observing that hypoxanthine is dysregulated in both positive and negative mode also validates the characterization of this peak. In addition to the MS/MS pattern, the retention time is another important parameter to consider. For example, the precursor ion m/z 300.2889 is appropriate for both sphingosine C-18 and palmitoylethanolamide (PEA), which have the same formula of $C_{18}H_{37}NO_2$ (Figure 8A). These molecules are indistinguishable by accurate mass alone. If these molecules were not resolved by chromatography, both species would be selected to fragment at the same time, generating a convoluted spectrum that would hinder identification of either species. When resolved, the individual species can be analyzed and structures assigned to each peak, as represented by peaks 1 and 3 in Figure 8A. The relative retention time can support a structural assignment. In Figure 8, two additional peaks, 2 and 4, can be seen which are analogs of 1 and 3 but are an additional two carbon units long. In general, on C18 based columns, increasing chain number and increasing saturation increases the retention time for a group of molecules with the same functional group. Observing a later retention time for sphingosine C-20 over sphingosine C-18 and stearyl ethanolamide over PEA is consistent with their characterization.

In a recent study⁵⁰, an unknown endogenous compound was identified utilizing this process, which illustrates the power of LC-Q-ToF based characterization of metabolites. A species with a m/z of 328.3213 was observed and was up-regulated in a rat model of neuropathic pain. At that time, searching m/z 328.3213 in METLIN returned two structural isomers, stearyl ethanolamide and sphingosine C-20. The 20 V fragment experimental spectrum was compared against the stearyl ethanolamide and sphingosine C-20 spectra in METLIN. Stearyl ethanolamide was quickly ruled out since the experimental spectra lacked the m/z 62.060 ion characteristic of ethanolamides. Comparison to sphingosine C-20 revealed several fragments in common in the 250-310 m/z range, although their ratios varied significantly. In the range of 40-120 m/z , there were some low intensity ions that did not match well. Fragmentation at 40 V provided more intense signals in the 40-120 m/z range, which did not match well to sphingosine C-20, ruling out this metabolite. Additionally, a peak with a m/z and fragmentation pattern appropriate for sphingosine C-20 eluted a few minutes earlier than the unknown m/z 328.3213 species. Since matching to METLIN was exhausted, other databases were searched with the m/z , which returned seven results within 0.01 Da. Searching again with the formula $C_{20}H_{41}NO_2$, which is the same molecular formula for stearyl ethanolamide and sphingosine C-20, and the most reasonable formula calculated from the accurate mass of 328.3213, stearylethanolamide and N, N,-dimethylsphingosine (DMS) was found. DMS was purchased, and LC-Q-ToF analysis of the sample and standard provided matching retention times and MS/MS spectra as reported⁵⁰. This strongly supported the identification of the species as DMS and provided the first characterization and quantitation of DMS as a naturally occurring metabolite. Since this analysis was completed DMS has been added to METLIN. Investigators that have access to pure standards of compounds not currently characterized on METLIN can first metlin@scripps.edu.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Wang Z, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011; 472:57–63. [PubMed: 21475195]
2. Wikoff WR, Gangoiti JA, Barshop BA, Siuzdak G. Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clin. Chem*. 2007; 53:2169. [PubMed: 17951291]

3. Wikoff WR, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:3698. [PubMed: 19234110]
4. Vinayavekhin N, Saghatelian A. Regulation of Alkyl-dihydrothiazole-carboxylates (ATCs) by Iron and the Pyochelin Gene Cluster in *Pseudomonas aeruginosa*. *ACS Chem. Biol.* 2009; 4:617–623. [PubMed: 19621937]
5. Kalisiak J, et al. Identification of a New Endogenous Metabolite and the Characterization of Its Protein Interactions through an Immobilization Approach. *J. Am. Chem. Soc.* 2008; 131:378–386. [PubMed: 19055353]
6. Leiss KA, Maltese F, Choi YH, Verpoorte R, Klinkhamer PGL. Identification of Chlorogenic Acid as a Resistance Factor for Thrips in *Chrysanthemum*. *Plant Physiol.* 2009; 150:1567–1575. [PubMed: 19448039]
7. Allen J, et al. Discrimination of Modes of Action of Antifungal Substances by Use of Metabolic Footprinting. *Appl. Environ. Microbiol.* 2004; 70:6157–6165. [PubMed: 15466562]
8. Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:14728–14733. [PubMed: 19667173]
9. Ludwig C, Viant MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem. Anal.* 2010; 21:22–32. [PubMed: 19904730]
10. Powers R. NMR metabolomics and drug discovery. *Magn. Reson. Chem.* 2009; 47:S2–S11. [PubMed: 19504464]
11. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* 2007; 26:51. [PubMed: 16921475]
12. Lei Z, Huhman D, Sumner LW. Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry.* 2011; 286:25435–25442. [PubMed: 21632543]
13. Smart KF, Aggio RBM, Van Houtte JR, Villas-Boas SG. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat. Protoc.* 2010; 5:1709–1729. [PubMed: 20885382]
14. Dunn WB, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2011; 6:1060–1083. [PubMed: 21720319]
15. Chan ECY, Pasikanti KK, Nicholson JK. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* 2011; 6:1483–1499. [PubMed: 21959233]
16. Fiehn O, et al. Metabolite profiling for plant functional genomics. *Nat Biotechnol.* 2000; 18:1157–61. [PubMed: 11062433]
17. Babushok VI, et al. Development of a database of gas chromatographic retention properties of organic compounds. *J. Chromatogr. A.* 2007; 1157:414–421. [PubMed: 17543315]
18. Kind T, et al. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Anal. Chem.* 2009; 81:10038–10048. [PubMed: 19928838]
19. Xu F, Zou L, Ong CN. Multiorigin of Chromatographic Peaks in Derivatized GC/MS Metabolomics: A Confounder That Influences Metabolic Pathway Interpretation. *J. Proteome Res.* 2009; 8:5657–5665. [PubMed: 19835423]
20. Nordstrom A, Want E, Northen T, Lehtio J, Siuzdak G. Multiple Ionization Mass Spectrometry Strategy Used To Reveal the Complexity of Metabolomics. *Anal. Chem.* 2007; 80:421–429. [PubMed: 18085752]
21. Wishart DS, et al. The human cerebrospinal fluid metabolome. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2008; 871:164–73.
22. Lu W, Bennett BD, Rabinowitz JD. Analytical strategies for LC–MS-based targeted metabolomics. *J. Chromatogr. B.* 2008; 871:236–242.
23. Kaddurah-Daouk R, et al. Lipidomic analysis of variation in response to simvastatin in the Cholesterol and Pharmacogenetics Study. *Metabolomics.* 2010; 6:191–201. [PubMed: 20445760]
24. Vinayavekhin, N.; Saghatelian, A. *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.; 2001.

25. Johnson CH, et al. Radiation metabolomics. 4. UPLC-ESI-QTOFMS-Based metabolomics for urinary biomarker discovery in gamma-irradiated rats. *Radiat Res.* 2011; 175:473–84. [PubMed: 21309707]
26. Trupp M, et al. Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment. *Plos One.* 2012; 7:e38386. [PubMed: 22808006]
27. Wikoff WR, Kalisak E, Trauger S, Manchester M, Siuzdak G. Response and recovery in the plasma metabolome tracks the acute LCMV-induced immune response. *Journal of Proteome Research.* 2009; 8:3578–87. [PubMed: 19496611]
28. Yanes O, et al. Metabolic oxidation regulates embryonic stem cell differentiation. *Nat. Chem. Biol.* 2010; 6:411–417. [PubMed: 20436487]
29. Marshall AG, Hendrickson CL. High-Resolution Mass Spectrometers. *Annu. Rev. Anal. Chem.* 2008; 1:579–599.
30. Verhoeven, H.; Ric de Vos, C.; Bino, R.; Hall, R. Saito, K.; Dixon, RA.; Willmitzer, L., editors. Springer Berlin Heidelberg; 2006. p. 33–48.
31. Kamleh A, et al. Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. *Rapid Commun. Mass Spectrom.* 2008; 22:1912–1918. [PubMed: 18470888]
32. Breitling R, Pitt AR, Barrett MP. Precision mapping of the metabolome. *Trends Biotechnol.* 2006; 24:543–548. [PubMed: 17064801]
33. Brown SC, Kruppa G, Dasseux J-L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* 2005; 24:223–231. [PubMed: 15389859]
34. Smith CA, et al. METLIN: A Metabolite Mass Spectral Database. *Ther. Drug Monit.* 2005; 27:747–751. [PubMed: 16404815]
35. Patti GJ, et al. Untargeted Metabolomics Implicants Dysregulated Sphingomyelin Metabolism in the Central Nervous System During Neuropathic Pain. *Nat. Chem. Biol.* 2011 In Press.
36. Psychogios N, et al. The Human Serum Metabolome. *PLoS ONE.* 2011; 6:e16957. [PubMed: 21359215]
37. Chen L, Zhou L, Chan ECY, Neo J, Beuerman RW. Characterization of The Human Tear Metabolome by LC-MS/MS. *J. Proteome Res.* 2011; 10:4876–4882. [PubMed: 21800913]
38. Want EJ, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* 2010; 5:1005–1018. [PubMed: 20448546]
39. Nebert DW, Zhang G, Vesell ES. From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab Rev.* 2008; 40:187–224. [PubMed: 18464043]
40. Brown M, et al. A metabolome pipeline: from concept to data to knowledge. *Metabolomics.* 2005; 1:39–51.
41. Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* 2006; 78:779–787. [PubMed: 16448051]
42. Kiefer P, Delmotte N.I. Vorholt JA. Nanoscale Ion-Pair Reversed-Phase HPLC–MS for Sensitive Metabolome Analysis. *Anal. Chem.* 2010; 83:850–855. [PubMed: 21166460]
43. Castro-Perez J, et al. Localization of Fatty Acyl and Double Bond Positions in Phosphatidylcholines Using a Dual Stage CID Fragmentation Coupled with Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2011; 22:1552–1567. [PubMed: 21953258]
44. Hsu F-F, Turk J. Elucidation of the double-bond position of long-chain unsaturated fatty acids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J. Am. Soc. Mass Spectrom.* 2008; 19:1673–1680. [PubMed: 18692406]
45. Thomas MC, et al. Ozone-induced dissociation: Elucidation of double bond position within mass-selected lipid ions. *Anal. Chem.* 2007; 80:303–311. [PubMed: 18062677]
46. Gian Luigi R. Dietary *n*-6 and *n*-3 polyunsaturated fatty acids: From biochemistry to clinical implications in cardiovascular prevention. *Biochem. Pharmacol.* 2009; 77:937–946. [PubMed: 19022225]

47. Ding J, et al. Capillary LC Coupled with High-Mass Measurement Accuracy Mass Spectrometry for Metabolic Profiling. *Anal. Chem.* 2007; 79:6081–6093. [PubMed: 17636878]
48. Lindon JC, Nicholson JK. Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics. *Annual Review of Analytical Chemistry.* 2008; 1:45–69.
49. Cravatt B, et al. Chemical characterization of a family of brain lipids that induce sleep. *Science.* 1995; 268:1506–1509. [PubMed: 7770779]
50. Patti GJ, et al. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nat Chem Biol.* 2012; 8:232–4. [PubMed: 22267119]

Figure 1.
Flow chart for LC-Q-ToF characterization of metabolites utilizing the METLIN database.

Figure 2.

Determination of monoisotopic peak, charge state, and adduct of the precursor ion. (A) The total ion chromatogram (TIC) for a representative sample. (B) The extracted ion chromatogram, showing one peak for 496.3409. (C) The mass spectrum at 24.5 minutes, scaled to highlight the peaks at 496.3409 and 518.3219 m/z, which represents the protonated ($[M+H]^+$) and sodiated ($[M+Na]^+$) species for the same metabolic feature. (D) Zooming in further on the peak at 496.3409 reveals a series of isotope peaks of M+1 and M+2.

Figure 3.

Insufficient chromatographic resolution of a species can lead to overlapping peaks that produce convoluted MS/MS species. (A) Insufficient resolution of the species m/z 480.3082 from other components in a sample provides several overlapping peaks. (B) Fragmentation of the unresolved species (m/z 480.3082) from panel (A) results in a convoluted spectrum containing at least two species, m/z 480.3084 and m/z 482.2567. (C) Chromatographic resolution of the species m/z 480.3084. (D) Fragmentation of a resolved spectrum of m/z 480.3084 from panel (C) allows for the characterization of lysoPE(18:1/0:0).

METLIN: Metabolite Search
Simple

[Simple](#) | [Advanced](#) | [Batch](#) | [Fragment](#) | [Multiple Fragment](#) | [Neutral Loss](#) | [Unknowns](#)

Mass:	<input type="text" value="137.045"/>	Tolerance (±)	<input type="text" value="30"/>	<input type="text" value="ppm"/>
Charge:	<div><div>Neutral Positive Negative</div><div>M+H M+NH4 M+Na M+H-2H2O M+H-H2O M+K M+ACN+H M+ACN+Na M+2Na-H M+2H M+3H M+H+Na M+2H+Na M+2Na M+2Na+H M+Li M+CH3OH+H</div></div>			

• To select multiple Adducts:
⌨ - Hit Ctrl + Adducts
⌨ - Hit Command + Adducts
Select: **all** | none

Figure 4. Metabolite search in METLIN. The simple metabolite search panel, with 137.045 inputted and M+H selected as the adduct.

METLIN
Metabolites

Mass 137.045 with 30 ppm mass accuracy

[Change Query](#)

Total: 7 Metabolites

METLIN ID	MASS	Δ ppm	NAME	MS/MS	STRUCTURE
4244	[M+H] ⁺ m/z 137.0445 M 136.0372	4	Threonate Formula: C ₄ H ₈ O ₅ CAS: 70753-61-6	View	
35473	[M+H] ⁺ m/z 137.0444 M 136.0372	4	D-threonic acid Formula: C ₄ H ₈ O ₅ CAS:	NO	
45859	[M+H] ⁺ m/z 137.0444 M 136.0372	4	Threonic acid Formula: C ₄ H ₈ O ₅ CAS:	NO	
45855	[M+H] ⁺ m/z 137.0444 M 136.0372	4	Erythronic acid Formula: C ₄ H ₈ O ₅ CAS:	NO	
35474	[M+H] ⁺ m/z 137.0444 M 136.0372	4	DL-erythronic acid Formula: C ₄ H ₈ O ₅ CAS:	NO	
865	[M+H] ⁺ m/z 137.0458 M 136.0385	5	allopurinol Formula: C ₅ H ₄ N ₄ O CAS: 315-30-0	View	
83	[M+H] ⁺ m/z 137.0458 M 136.0385	5	Hypoxanthine Formula: C ₅ H ₄ N ₄ O CAS: 68-94-0	View	

Figure 5.
The returned metabolites from the search for 137.045, with structural and mass spectral information.

Figure 6.

The spectra of hypoxanthine. The 20 V fragmentation spectrum is displayed here. Clicking on the other voltages in the black bounded box displays the appropriate spectra. Hovering over a fragment peak, such as 119 depicted here, reveals a predicted fragment structure and details about the exact mass and relative intensity of that fragment.

Figure 7.

A comparison of experimental (black) and METLIN standard (red) spectra for three metabolites. (A) Phenylalanine, (B) arachidonic acid, and hypoxanthine in (C) positive and (D) negative mode.