March 2015

METABOLOMICS: PROSPECTS AND PITFALLS

Answers to Common Questions

LCGC

UBM

TOC
Table of contents

METABOLOMICS: PROSPECTS AND PITFALLS

March 2015

CONTENTS

LCGC UBM

METABOLOMICS: PROSPECTS AND PITFALLS

Answers to Common Questions

SPONSORED

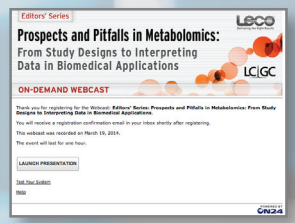
Broader Coverage of the Metabolome with Complementary LC-MS and GC-MS Techniques

Click to view PDF

FREE WEBINAR

Prospects and Pitfalls in Metabolomics

Click to launch webinar



Introduction

In metabolomics, one studies the small-molecule metabolites that both result from and influence cellular processes, as a way of learning about cellular biology, systems biology, and disease. Metabolite levels and fluxes are often seen as being at the receiving end of the information flow from the genotype to the phenotype, but such a view is too simplistic. Many metabolites provide feedback inhibition to protein expression, and thus regulate the flux of metabolites through the system, and others — signals or hormones — affect gene expression and activity. Thus, there is not a linear flow of information from the genotype to the phenotype; rather, a systems view will provide a more accurate understanding of the mechanisms of biology, physiology, and disease.

Given the complexity of biological systems, studying the metabolome is a challenging undertaking. In a recent LCGC web seminar, Prof. Oliver Fiehn, PhD, the director of the NIH West Coast Metabolomics Center at the University of California, Davis, provided advice for dealing with the challenges involved, by explaining the steps to follow in the metabolomic workflow and how to avoid the most serious pitfalls along the way. He covered considerations for defining a suitable hypothesis, designing a study, preparing biological specimens for comprehensive chemical analysis, choosing the most suitable analytical platforms, analyzing raw data and identifying metabolites, using quality controls and validations, and interpreting the data on the background of biomedical questions. Here in this e-book, Fiehn answers audience questions raised during the web seminar. To watch a recording of the seminar, register for free at the link provided in the box at left.

General Questions

Why hasn't metabolomics taken off in industry?

Metabolomics is being used heavily in industry, for example, in the agro-biotechnology business. Monsanto (U.S.) and Metanomics (Germany) have a large-scale agreement on use of intellectual property that was gained by industry-sized metabolomics screenings of hundreds of thousands of plants. Metabolomics is also used heavily in the biomedical industry, including pharma, again using a service provider (Metabolon, Durham, North Carolina). Metabolon has gained several rounds of investment and has a strong standing in the market. There are ranges of other metabolomics service providers, from Stemina (Madison, Wisconsin), to Biocrates (Austria) and multiple others. Within the pharma or agbiotech industry, it is simply cheaper to outsource services for large-scale operations than to implement all equipment and expertise in-house, although most companies have smaller in-house teams as well to validate metabolomics findings.

Where do you think the main bottlenecks occur in metabolomics?

Where should I begin? Of course, the biggest bottleneck is correct compound annotations. There are about 40,000,000 small molecules represented in PubChem, but all publicly available tandem mass spectrometry (MS-MS) libraries (Metlin, NIST14, and MassBank) combined only comprise 40,000 unique compounds. So currently we have at best 0.1% coverage. That is just a sad state of affairs, mostly a result of poor funding decisions by the scientific agencies. Some scientists may argue that there are only 15,000 compounds in biochemical databases like KEGG, but others correctly reply that we are entering the era of the "exposome," integrating data and knowledge of both endogenous metabolites and exogenous compounds such as food, cosmetics, drugs, and pesticides, all of which have enormous implications on health and disease.

PHOTOGRAPHER KARIN HIGGINS/UC DAVIS

PROFESSOR OLIVER FIEHN, PhD, is the director of the West Coast Metabolomics Center at the University of California, Davis, and an enthusiastic faculty member in the Genome Center, where he conducts research in metabolomics method development and biomedical applications. He oversees the operation of 25 staff using 15 mass spectrometers. He has pioneered developments and applications in metabolomics since 1998 with more than 130 publications to date, and is keen to improve these methods with the aim of further standardization and higher content in biochemical information through automatic annotation of metabolites and by using a range of chemical and biochemical database identifiers with every report his group generates. For the Metabolomics Society, he has chaired the efforts in standardizing metabolomics reports, and developed and serviced a diverse set of projects such as organizing monthly seminar series, workshops, and symposia. He is also the organizer of the Metabolomics 2015 conference.

SAMPLE PREPARATION AND SEPARATIONS

Sample Preparation

Recently we've seen publications on solvent-free sample preparation for metabolomics. What are your thoughts on the coverage and representativeness of such protocols?

Solid-phase microextraction (SPME) is an example of solvent-free sample preparation of the volatile metabolome in liquid or gaseous samples, used in combination with gas chromatography–mass spectrometry (GC–MS). It is a fantastic and validated tool, used in daily practice in many laboratories. In principle, other systems could adopt the philosophy of selectively binding certain classes of metabolites (such as using proteins bound to beads), and then using these solid supports directly with GC–MS or liquid chromatography–mass spectrometry (LC–MS). The coverage obtained by these methods strongly relies on the quality and selectivity of the solid support that is used.

How much do different extraction solvent systems affect metabolomics in terms of coverage and accuracy?

Unfortunately, there is not one single extraction solvent system that is optimal for all chemical classes, or even for all matrices. Different extraction systems always exert bias. For example, if you use acetonitrile as the extraction solvent, you will miss very lipophilic components. If you use hexane as the extraction solvent, you will miss sugars. Some extraction solvents may already alter the metabolome composition; for example, using methanol under acidic conditions may lead to artificial metabolite methylations. For that reason, my laboratory generally uses mixtures of solvents, such as acetonitrile–water–chloroform for plant samples, isopropanol–water–acetonitrile for primary metabolism in mammalian samples, or methyl *tert*-butylether–methanol–water for lipidomics studies. Just as you need more than metabolomics platform for data acquisition, you may find better coverage and accuracy if you use more than one solvent system for sample preparation.

What is the best method to extract cells, such as hepatocyte cell cultures?

Each extraction method introduces bias. It is not easy to objectively compare methods because you cannot just state optimal criteria such as "number of peaks" or "intensity of peaks" — these could be caused by metabolic degradation or enzymatic reactions during sample preparation. For that reason, microbiologists at least look at the "energy charge" by quantifying the levels of ATP, ADP, and AMP. I am not aware that a consensus for human cell cultures has been achieved, but in general, try to stop metabolic activity quickly when harvesting your cells, monitor potential cell leakage, and use target compounds such as cysteine and oxidized or reduced glutathione as a readout for the oxidative stress you incur during sample preparation. Have all your solvents carefully degassed, such as with nitrogen or argon, to remove residual oxygen from your solutions. Be wary of pH: Some compounds like acidic conditions, and others do not. There is not a single best recipe, I am afraid.

Chromatography

Given that targeted LC–MS can detect ~300–500 compounds, and GC–MS can quantify ~100, are the two approaches usually used as complementary or supplementary techniques in regard to quantification and compound identification?

Depending which GC–MS instrument you use, you get up to 900 detected and mass spectra–deconvoluted compounds in GC–MS, out of which we regularly identify some 120–200 compounds (depending on the sample, of course) after data curation (see above), plus some

200–300 unidentified compounds. The rest would be discarded as being too noisy and too low in abundance to be reliably reported. As stated above, with trimethylsilylation, these are mostly primary metabolites such as sugars, organic phosphates, hydroxyl acids, aromatics and sterols. In LC–MS with positive or negative electrospray ionization (ESI), such as in lipidomics, we get some 600–700 peaks detected, out of which we identify some 300 unique complex lipids, plus maybe some 200 unidentified signals, again, discarding very noisy and unreliable low-abundance peaks. So, these two platforms are very similar in total number of reported peaks (about 500 metabolites per platform), and also similar numbers of identified metabolites (150–250 per platform), and there would be almost no overlap between these platforms (except for such as free fatty acids and cholesterol). An alternative in-between platform is hydrophilic interaction chromatography–MS (HILIC–MS). Historically, HILIC–MS has been plagued by shifting and variable retention times, but this problem can be overcome nowadays. HILIC–MS can also nicely retain and quantify important compounds such as trimethylamine-*N*-oxide (TMAO) that would not be retained at all in reversed-phase LC–MS and that cannot be made volatile to be detected in GC–MS.

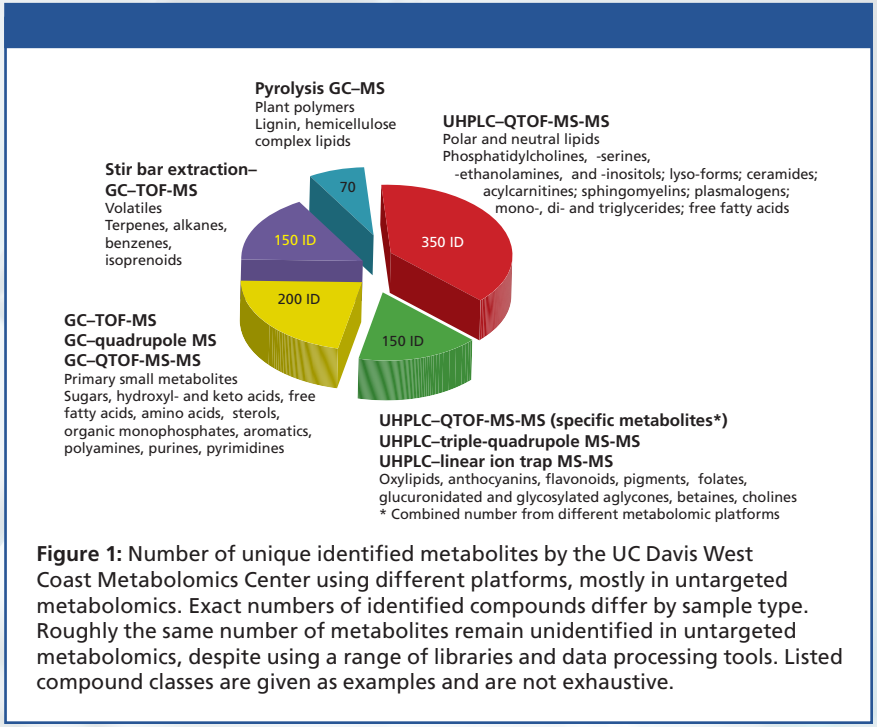


Figure 1: Number of unique identified metabolites by the UC Davis West Coast Metabolomics Center using different platforms, mostly in untargeted metabolomics. Exact numbers of identified compounds differ by sample type. Roughly the same number of metabolites remain unidentified in untargeted metabolomics, despite using a range of libraries and data processing tools. Listed compound classes are given as examples and are not exhaustive.

So, the physics of the different platforms make the different methods really complementary to achieve comprehensive metabolomics assessments.

You showed a pie chart (Figure 1) indicating the number of unique metabolites that your group has identified using various metabolomic platforms. Can you explain why you think that LC–MS analysis using a C18 column would not cover all the compounds, such as in primary metabolism?

C18 columns operate by separating compounds based on lipophilicity. Most primary metabolites, such as citrate, malate, glucose, phosphoglycerate, or ribulose-5-phosphate are simply not retained by C18 columns. There are some types of C18 phases that have a better capacity to retain polar compounds than others, but generally, capillary electrophoresis, HILIC or GC give you much easier, much more reliable, and much faster results than C18-based efforts.

For GC–MS, how much do the derivatization reagents affect the column, in terms of factors such as how often the column needs to be changed or replaced, septa need to be changed, autosampler needles need to be cleaned, and guard columns need to be replaced?

Our service recharge rate model estimates about $10 per sample for consumables, including GC–MS system maintenance and repairs, but not including service contracts (which we don't use because they are outrageously expensive) and also excluding instrument depreciation costs. On top of that $10 come labor costs for preparing samples, running the instruments, tuning and maintaining the instruments,

monitoring the quality controls, processing the data, and curating and reporting the data. You can imagine that these labor costs far outweigh the costs for consumables. For that reason, we use preventive maintenance: We change columns, liners, nuts, ferrules, filaments, syringes, seals, and septa on regular intervals before data quality is compromised. In addition, we use quality control samples and chemical mixtures to tightly monitor the system in-control situations.

Depending on the application and column, a column lasts for about 1000 injections. We use GC columns with integrated empty guard columns (10 m in length without film). From these guard columns we regularly cut off 15-cm-long pieces. Doing this will shorten absolute retention times, but not relative retention times (using the fatty acid methyl ester [FAME] retention index mixtures). We also use a robotic system called an *automatic liner exchange*. We exchange liners after every set of 10 samples with fresh liners, followed by blank injections and QC injections. We do that because metabolomics is a dirty business: Involatile components otherwise accumulate in the liners and cause quality problems, including matrix effects and compromised method blanks.

What are the advantages and disadvantages of using ion-pairing agents in LC?

The advantage of ion pairing is to be able to elute polar compounds from a reversed-phase LC column. The disadvantage is you will never lose these reagents from your system. That means you need to have a dedicated LC system for this application. A second disadvantage, of course, is that you will only be able to use one polarity in electrospray, but not the polarity used for your ion-pair agent.

Other Methods

What do you think about the role of capillary electrophoresis (CE) in metabolomics?

CE is great for cationic and anionic compounds. We believe that HILIC today is equally capable of detecting compounds for which CE is otherwise most suitable. The advantage of HILIC over CE is that HILIC is more scalable: You could, for example, use large-bore columns for fractionations, if you'd like to get larger quantities of a specific compound. CE is not scalable in this sense. Moreover, HILIC does not need specific equipment; you can use your existing LC pump and mass spectrometer without any modification. That's why CE–MS is not much used in U.S. or European metabolomics laboratories.

What do you think about using supercritical fluid chromatography (SFC)–MS in lipidomics?

SFC has been seen as an emerging, promising technology for a very long time, similar to other methods that never really made a breakthrough for use in mainstream applications. SFC–MS in lipidomics is a method in search for a question. LC–MS-based lipidomics works just fine, and is easier to operate. The open question in lipidomics is to separate positional isomers, from *sn*2/*sn*3 acyl-positions to the positions of double bonds. SFC has no role for that important task.

What do you think about ambient MS methods such as rapid evaporative ionization mass spectrometry (REIMS) and desorption electrospray ionization (DESI)-MS that are currently used to classify biomedical samples?

These methods are great new developments that extend the reach of mass spectrometry into clinical diagnostics, routine analysis of materials or suspicious goods, and to classify samples, as you correctly note. I am sure there might be many valuable applications, if use of the ambient MS tools can be validated with respect to specificity and sensitivity of their respective classifications.

Can a less complex detector like a flame ionization detector be used instead of MS to determine more complexity in lipids like phospholipids?

Flame ionization detection is exclusively used in samples of limited complexity in GC–MS. It cannot be combined with LC–MS. Phospholipids are always presenting in very complex mixtures. This fact requires high-resolution chromatography and highly selective detectors.

Separation Science

What Else is in Your Sample?

Missing something important in that sample? Our GC-TOFMS and GCxGC-TOFMS products allow you to see the complete picture by providing the separation, accuracy, resolving power, deconvolution, and speed to characterize your most complex samples. ChromaTOF® software provides the ability to identify significantly more analytes—so you can solve more analytical challenges than ever before.

PEGASUS® 4D

PEGASUS GC-HRT

See how Metabolomics experts use LECO GCMS solutions at
www.leco.com/metabolomics

LECO®

Delivering the Right Results

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

METHOD DEVELOPMENT

Do you have any tips for selecting an appropriate internal standard for quantitative metabolomics?

If you try using "one" appropriate internal standard for quantitative purposes, you have already lost. There is none. Unfortunately, calibration response curves in both GC–MS and LC–MS methods vary widely between classes of chemical compounds, and they can even vary widely within classes of chemical compounds, sometimes even for isomers (although that is more rare). In lipidomics, many scientists use one internal standard per lipid class, such as deuterated or odd-chain derivatives for phosphatidylcholines, for cholesteryl esters, and so on. However, your internal standards then may face matrix effects (ion suppression) in their local retention environment that are different to same-class compounds that are eluted a little farther away; hence, your internal standard may actually increase errors instead of fighting technical errors. Still, use of internal standards is considered a gold standard in analytical chemistry, and it is certainly wise to report data with these standards, even if the only aim is to give a good estimate of absolute molar concentrations and even if that value might be twofold off from a theoretical true value; at least you would still report the right order of magnitude for your metabolites.

But there is a secondary answer to your question. Internal standards can be used not only for absolute quantifications, but also for monitoring overall system suitability including absolute instrument response, retention times, and peak widths. They are very useful for laboratory staff, because they can look at these compounds every day in the same manner and spot problems right away. We use CUDA (InChI key HPTJABJPZMULFH-UHFFFAOYSA-N) because it is nicely retained both in reversed-phase LC and HILIC, and because it ionizes well both in positive and negative ESI MS. In GC–MS, we use our FAME internal standards for the same quality control purpose.

To hit the right concentration in GC–time-of-flight (TOF)-MS, would you recommend changing the dilution or the split ratio? What are the pros and cons of these options?

Assume you have a study with 100 samples. We strongly advise to perform test runs before running all the valuable 100 samples! Take either surrogate samples (such as commercial blood plasma, or some pooled samples) for testing and validating optimal extraction conditions and injection conditions before you run your 100 samples. However, I agree, sometimes the time pressure is so high, such as for a grant application submission, that one cannot perform such optimization. If you then shoot the first couple of samples and discover that many peaks are far too overloaded to yield actual quantifiable information, then there are three options to consider. The first is to talk to your collaborator. For example, maybe you analyze fruit samples, and you find glucose, fructose, sucrose, and malate to be very much overloaded, saturating the detector and compromising surrounding peak detections. However, in some samples, this is hardly avoidable (such as sucrose in grass leaf samples), and one can deliberately make the decision to ignore these overloaded peaks and focus on the lower abundance peaks. In fact, in all likelihood, your collaborator may not be really interested in these super-abundant peaks anyway. As a second option, you can shoot your sample twice, once in splitless injection and then with a split ratio of maybe 1:10. If your saturated peaks are still saturated, you could inject again at a 1:25 split ratio and provide two reports, one data set on the low-abundance peaks you saw in splitless mode, and one report only on the very abundant compounds. Do not multiply the abundant-metabolite values simply by the inverse of the split ratio; unfortunately, that would give very inaccurate estimates! The downside of split injections is that amino groups (such as in polyamines or amino acids) do not maintain their trimethylsilylation status very well. Splitless injection is far more advantageous for trimethylsilylated amino groups, depending of course on general

liner and injector conditions. Third, dilution does not work well for trimethylsilylated samples with respect to peak ratios and preserving amino-group derivatizations, and it is also costly if you think about using 1 mL of reagent for each sample instead of 100 µL!

In terms of data acquisition, how do you handle technical replicates? In what terms are they useful, such as replicate injections? And what is a reliable number of replicates?

Both LC–MS and GC–MS today are such reliable techniques that you do not need technical replicates if you use proper quality controls to maintain within-series and between-series reproducibility. As stated above, for every 10 biological samples we use one "pool quality control aliquot" and one technical chemical mixture QC sample. These QC samples are also used in large-scale normalizations, such as for clinical cohort samples for which we analyze thousands of patient samples.

It is more difficult to answer the following question: How many biological replicates does one need? Often, this question can only be answered after the study, when one knows the exact numbers for technical reproducibility of each peak (using the QC value, in %RSD) and the exact biological variance between the groups (such as healthy and disease groups) in comparison to the within-group variance (biological differences such as within a population of healthy people). In statistics, this is called *power analysis*. You can rightly assume that biological differences between humans are far larger than between cloned mice in a controlled cage, so you will need many more humans for statistical assessments of metabolome–phenotype relationships than you would need mice, or cell cultures. However, even cell cultures are not all the same, and scientists have noted that not all cells within a cell culture behave the same way! At the end, your statistical and your scientific questions define the number of biological replicates you need.

Is multiple-reaction monitoring (MRM) usually necessary when using mass spectrometry? With a single-quadrupole instrument, I tend not to get a stable signal, which makes determining peak heights or area difficult.

Single-quadrupole instruments in GC–MS are great instruments for entry-level metabolomics because they are used in conjunction with electron ionization, gaining reliable ion fragment spectra. Single-quadrupole instruments in LC–MS have limited use because in electrospray soft ionization, no fragmentation information is available. Selectivity of retention time and nominal mass information is not very high. This is the reason why metabolomicists mostly use either MRM for target compounds in LC–MS-MS — for example using triple-quadrupole or linear ion trap systems — or use retention time and high-resolution accurate mass spectrometry to gain selectivity by focusing on very small *m/z* windows for quantification.

What is your opinion of using "shotgun" approaches in metabolomics applications?

There are many papers claiming high-throughput operation, but few papers showing successful analyses of thousands of samples. The usual logic behind shotgun approaches is to sacrifice data accuracy for speed of sample throughput. Unless authors show that this logic is being actually used — for example, for a quick-and-dirty shotgun screening of 5000 samples, followed by finding the 50 most interesting samples for which detailed coverage of metabolomics is achieved — I think most shotgun approaches do not produce interesting data. In fact, current high-quality LC–MS-based lipidomics with ultrahigh-pressure liquid chromatography (UHPLC) systems and 1.7-µm particle columns yield better, more comprehensive, and more quantitative data with 15-min sample-to-sample cycle times than direct infusion nanoelectrospray-MS-MS shotgun investigations that actually require the same time, when performed in a comprehensive sweep fashion. Many scientists overestimate the selectivity of MRM screens, unfortunately. Orthogonal methods like chromatography really help for getting reliable data.

How do you balance the tradeoff between data points per peak vs. the stability of signal intensity? What criteria do you use for finding that balance?

This can only be answered on an instrument-to-instrument basis, and method-to-method application basis. It is not only the stability of the signal intensity, but really a question of optimal signal-to-noise ratios for quantification purposes, the number of coeluted compounds or interfering signals, and the mass spectral deconvolution software that is being used, and of course, the linear dynamic range of both the column capacity and the MS detector capacity.

Win the Fight Against Coelution

Your coelution problem is one of the past when you have the leader in GCMS metabolomics on your side. LECO products deliver the separation, accuracy, resolving power, deconvolution, and speed to characterize the most complex biological systems.

Do you know what else is in your sample? With our solutions, you'll have the all advantages you need to see the complete picture.

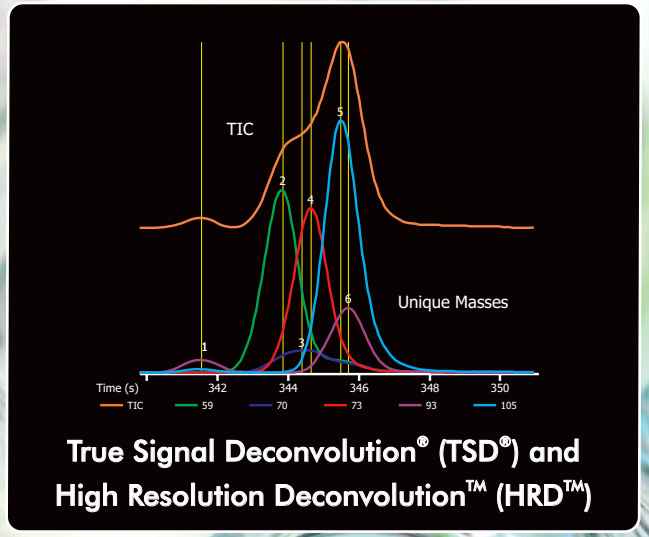
Deconvolution | Sensitivity

Productivity | Reproducibility

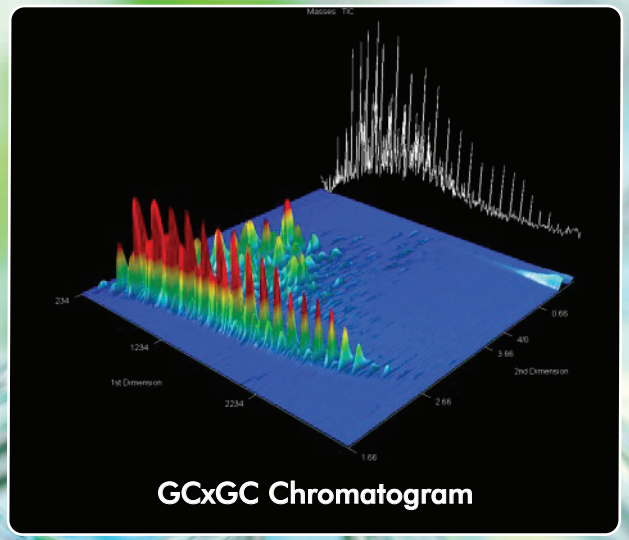
GCxGC

Accurate Mass | Structural Interpretation

Don't believe us? See how metabolomics experts use LECO GCMS solutions at *www.leco.com/metabolomics*



True Signal Deconvolution® (TSD®) and High Resolution Deconvolution™ (HRD™)



GCxGC Chromatogram

Separation Science

Delivering the Right Results

Phone: 1-800-292-6141 | info@leco.com
www.leco.com | © 2015 LECO Corporation

PEAK IDENTIFICATION

If several metabolites are eluted at the same time in an LC–MS run, how we can identify them?

In metabolomics, coelution of compounds is unavoidable. This leads to numerous problems, ranging from ionization suppression of one compound by its coeluted neighbors, to peak picking and deciding which ions are actually valid molecular adducts (such as M+H or M+Na) and which ions might be in-source fragmentation ions. A simple answer is, the higher your data acquisition rate, the better you will be able to distinguish small but measurable retention time differences to tell two coeluted peaks apart. By rule of thumb and the best of my experience, you need to have at least a retention time difference of two MS scans to tell two peaks apart. That means, if you run your instrument at 4 spectra per second, your peaks must be at least 0.5 s apart, even under otherwise optimal conditions (that is, Gaussian peak shapes, similar peak intensities, and so on). Obviously, faster data acquisition rates usually come at the price of lower absolute peak heights, and potentially, lower signal-to-noise ratios. One option, of course, is to improve the chromatography, but that is not always easily achieved. Another option is to improve the MS resolving power, but again, many laboratories do not have the option of using very expensive high-resolution equipment.

A further part of answering your question is likely geared toward obtaining clean MS-MS spectra. In data-dependent MS-MS spectra acquisition, the mass inclusion window is usually about 3 Da, even if you try to pick only one precursor ion. That means your MS-MS spectrum may contain fragments from coeluted precursor ions. It also means that you may not obtain MS-MS data on low-abundance peaks, because the instrument starts with the most abundant precursor ions and may not be fast enough to get to the low-abundance compounds. A novel way to get around this problem is using data-independent fragmentation information, by fragmenting large ranges of *m/z* values in a very fast manner, and assembling all fragment ions and their corresponding precursor ions using mass spectral deconvolution algorithms. This concept has been successfully used in GC–MS for 16 years now, and is currently tested in LC–MS software as well.

If the same *m/z* value of metabolite is eluted several times in an LC–MS run, how can we be sure which peak is major?

By definition, if you have several compounds with different physicochemical properties but identical elemental compositions — that is, identical accurate masses — these are isomers. Some researchers have indeed defined metabolomics as the science of isomers! Important examples are, for example, hexoses such as glucose and fructose, or vitamin isomers such as tocopherols. But we also regularly see many natural product isomers in food products such as strawberries. Often, but not always, isomers may yield indistinguishable mass spectra. In this case, alternative structure annotation methods may be needed, such as ion mobility, MS^n fragmentations and, of course, fraction collection and in-depth nuclear magnetic resonance (NMR) investigations.

How do you identify peaks by retention time using libraries if laboratories use different mobile phases in different gradient elution programs?

In GC–MS, the use of internal standards for a standardized retention index was established decades ago by Kováts. In GC–MS, the dominant retention factor is the boiling point, and the interaction with the column film is only a secondary factor. That is why GC–MS is much more useful for standardizing retention times compared to LC–MS. In liquid chromatography, there are multiple factors determining relative retention times, ranging from buffers to pH to highly different effects of solvents to particle and pore sizes, packings, the density

and nature of silica-based polarity modifiers, and other effects. In short, as you indicated in your question, you cannot compare retention times when laboratories use different mobile phases and different gradients, and presumably, different types of columns. There are literally hundreds of different C18 reversed-phase columns on the market, and they all do not behave in the same manner. However, it is still important to report retention time for two reasons: First, a laboratory could duplicate the entire method to validate the exact reports from another laboratory or paper. This is at the very heart and essence of the scientific approach itself. Second, if there are many reports on the use of identical LC conditions, or a very large database with retention time information with many compounds run under the same conditions, researchers can build retention time prediction models using quantitative structure–property relationship (QSPR) methods. The Metabolomics Society Standards Initiative (MSI) demands that there be at least two independent properties for metabolite annotations, such as retention time and full mass spectrum in GC–MS, or accurate mass and MS-MS in LC–MS, or other types of supporting data. This is of course a minimal requirement, while in fact, one may want to use retention time, accurate mass, isotope ratio, and MS-MS information for compound annotations. The term *compound identification* is reserved for when you have an actual authentic standard to verify the annotation.

How confident can you be in the identification of metabolites using a virtual MS-MS database?

Any prediction of mass spectra, or prediction of other compound properties such as retention times, requires extensive validation steps. Without showing validation data, including true positive/false positive ratios and true negative/false negative identifications, computer algorithms and MS-MS spectra predictions are not very valuable. Among those validations need to be the identification of compounds that were not used in generating the virtual MS-MS spectra (that is, true positive data), using decoy data of MS-MS spectra that do not represent the class of investigated compounds (that is, true negative data) and very clear statements about the coverage and extensions of the tools. For example, our initial virtual LipidBlast MS-MS library did not contain betaine lipids but focused on 29 other lipid classes. We have now extended this library to also be used for betaine lipids (which are prominently found in algae) and extended ranges of double bond numbers and carbon numbers in acyl groups. As long as these restrictions (or extensions) are clearly stated and understood, virtual MS-MS libraries can be a highly important tool in metabolomics.

Since we are comparing among samples, are there any advantages of using single *m/z*–retention time values rather than combining them (such as Na and NH_4)?

I very much endorse the concept of combining different adduct ions, and potential in-source fragment ions, to give a valid quantitative signature of a metabolite in LC–MS analyses. Counting beans in *m/z*–retention time pairs greatly overstates the number of independent signals, confuses readers about the difference of *m/z*–retention time pairs and genuine metabolites, and yields wrong quantitative information. For example, you cannot control the ratio of M+Na and M+NH_4 ions in the LC–MS analysis of triglycerides. If you combine both ions, you get valid quantitative information. If you do not combine these ions, which is what occurs in many vendor-released and freeware software solutions, you simply get wrong data. In addition, you need to consider the statistical problem of "multiple testing": A 95% confidence level, or a *p*-value of $p < 0.05$, means that just by random bad luck, 5 peaks out of 100 will appear significantly different even if they are not. But you do not know which ones! If you do not combine *m/z*–retention time pairs, this problem gets bigger, then you may have 50 peaks out of a total of 1000 that randomly give false significance levels. For that reason, false discovery rate (FDR) corrections are used, meaning that even your true positive biomarkers have to get over a higher bar! It is simply better analytical chemistry, and better statistics, to use fewer, but valid, signals, rather than more inaccurate and noisy signals.

DATA ANALYSIS

In a profiling experiment, what is the threshold for missing data points for compounds to make them liable to be rejected (as opposed to not being present at all due to effects of biological treatment)? For example, is it acceptable to report metabolites when 30% of the data points are missing?

There is an easy answer and a more difficult one. The easy one: The mere fact that a software program gives you missing data although true positive peaks are present means that the software is not operating well. The more difficult answer, of course, is that any software algorithm tries to optimize true positive/false positive peak finding ratios (and false positive/false negative ratios, of course). Each software algorithm is trained on specific assumptions about peak shapes, number of coeluted peaks, data acquisition rates, and finally, example chromatograms. That's why no single software program can work for all conditions, and all samples, equally well.

That said, I would like to mention that good software also employs algorithms that backfill missing data, and may highlight such "replaced values." It is simply not acceptable to not get a value — all GC–MS and all LC–MS ion traces have noise levels. Exceptions to this statement about noise are very specific cases such as selected-reaction monitoring [SRM] target scanning, or selected-ion monitoring [SIM] scanning for very-high-resolution MS instruments. In such cases, indeed chemical noise levels might be zero, except perhaps for electronic noise.

What we do in my own lab is to define true peak findings based on the experimental designs. For example, if you have 10 mutant samples and 10 wild-type samples, the mutants may lack specific peaks completely. If you then have an experiment with 10 different mutants all under one wild-type genetic background, maybe a peak is only present in one of the mutants, but lacking in all others and also lacking in the wild type. So, then you'd have 90% missing data, but still very important information! That's why we have thresholds for entering novel peaks into our BinBase database. That threshold is to have at most 20% missing-true-positive peak detections for any specific experimental class, and a minimum of six positive peak detections. Basically, we tell all our collaborators to give us a minimum of six independent biological replicate samples. If they only give three samples (as microbiologists tend to do), novel peaks simply do not enter our database.

How much biological variation is observed or acceptable for any compound in spite of having a similar starting material, method, and protocol? Is variation only a minor problem, or should a CV or RSD cut-off level be applied before proceeding to normalization?

This is another very good question, and one that we also ask ourselves in daily practice. The answer is that you need to use quality control (QC) samples that have identical biological origin (that is, a pooled sample), and then aliquot, extract, and prepare such pooled QC aliquots in the same manner as your real samples. You can then intersperse these QC samples into your analytical sequence to get your real within-series reproducibility. We then compare the %RSD values of the QC samples to the %RSD values for the real samples for the same compound. It's then a matter of taste where to put the threshold and deleted peaks from the table, before selecting the data sets for statistical investigations. Some scientists would choose a threshold ratio of twofold higher RSD values in the biological samples compared to the QC samples. Others are OK with lower bars, and yet others would say that these not-so-variable compounds would not appear as statistically differently regulated compounds anyway in either multivariate or univariate statistics, so it is not important to trim the data beforehand. A counter argument would be that if you want to use your method for clinical diagnosis, you

GETTY IMAGES/JODIE GRIGGS

need to use FDR adjustments, and the more signals you detect, the harsher your FDR adjustment must become.

Would you please comment on statistical tools, such as PCA, PLSDA, or OPLS-DA? How should one choose a statistical tool?

We use principal component analysis (PCA) mainly to gain an initial overview of a total data set. It is a so-called unsupervised technique; that means you get an unbiased view of your samples with respect the most important set of data variance (vector 1), the second most important set of data variance (vector 2), and so on. Hopefully, you will then see that a PCA separation by vector 1 and vector 2 already highlights the groups of samples that you were hoping to investigate, such as your mutant and wild-type control samples. However, you may also find that there was a big bias in your data set, such as that variance vector 1 in your data was separated by the days you operated your instrument, or other types of analytical error. So, using PCA really helps you ensure that your data set was not totally biased by effects that you did not want to investigate.

On the other hand, so-called supervised techniques like partial least squares (PLS) or support-vector machines (SVM) or decision trees or K-nearest neighbor methods (KNN) all factor in your experimental design. You will need to tell these software algorithms which are the groups of samples you are interested in, for example, which are your wild-type controls and which are your mutants. The algorithms will then focus only on variance in your data set that can best tell these groups apart. When you then investigate which metabolites had the biggest impact on generating these classification vectors, you may get important information about underlying biochemistry. However, be careful! Usually, metabolomics studies are underpowered, meaning you have many more variables (metabolites) than cases (samples). That means you have a positive bias to finding differences where in fact, there are none. This phenomenon is called *data overfitting*. The best way to control for overfitting is to develop a data model (called *discriminating vectors*, *biomarkers*, or *panels of biomarkers*), and test that model with an independent set of samples that you acquired at another time. Then you could be sure that your biomarkers are real and not just a statistical artifact. This process is called *validation*. Once again, be careful! Software algorithms will allow you to perform "leave-one-out" cross validation or Monte-Carlo simulation validation options. These solutions are often invalid, and will not find bias in your samples.

What is difference between the InChi Key and the CAS number? And why should we use the InChi key instead of the CAS number?

The Chemical Abstracts System (CAS) number is an arbitrary and proprietary number owned by the American Chemical Society, a private organization. You are not allowed to store or disseminate more than 5000 CAS numbers in any public domain. CAS numbers are used for individual queries on databases owned by the American Chemical Society, such as SciFinder, where information found in chemical literature is collated.

The International Chemical Identifier code (InChI) is a public and freeware code that denotes chemical structures as such, including stereochemical information. The InChI has been developed by the U.S. National Institute of Science and Technology (NIST) in cooperation with the International Union of Pure and Applied Chemists (IUPAC). InChI codes can be used for a large variety of cheminformatic services, and all major databases such as ChemSpider or PubChem entirely rely on InChI codes. Abbreviated InChI codes, called InChI keys, are the best codes to be used in your chemical reports and manuscripts, even if found in your supplementary material, because PubMed and Google will use these short strings when web queries try to find information about specific chemicals.

So in summary, a CAS number only helps scientists who buy CAS licenses. The InChI code is free chemical information. You choose!

When we analyze metabolites, do we have to also analyze MS-MS data?

It depends on the hypothesis that you are investigating in your study. When you only aim at discriminating groups or classes of samples, low-resolution ("shotgun") methods may help, including DESI-MS or similar methods. However, if you want to identify the compounds that distinguish these classes, you will need to use different compound annotation data, including fragment ion spectra investigations.

Can LipidBlast be used on GC–MS data?

LipidBlast has been developed on collision-induced MS-MS fragment spectra of complex lipids and cannot be used for electron-ionization MS fragment spectra. The physics of the mass spectrometers and (by the way) the compounds that are amenable to GC–MS and LC–MS are just too different. There are no validated GC–MS virtual libraries available at this point, although there are certainly tools and papers to predict GC–electron ionization MS spectra directly from structures.

If you use a FAME-based retention index (RI), do you use an equation to transfer it to an alkane RI for comparison in a library?

Yes, that is very easy to do. Just run the Kováts alkane mixture along with your FAME mixture, and you then have the data for converting alkanes to FAME retention indices (or whatever other logical sequence of internal standards you'd like to use). We use FAMEs because they do not occur in nature, except for one marine microorganism; under electron ionization conditions, you see the M^+ ion, which is very important in automatic data processing in algorithms like our BinBase database; and we use RI units that originally equated to absolute retention times for a specific column and instrument setup, so there is a chromatographic reality behind these numbers.

In fact, we used all spectra in the NIST14 library, and used the NIST algorithm, corrected for trimethylsilylations, to predict retention indices for our conditions and our BinBase database.

What platforms are used in your lab for alignment of GCxGC data?

After the last Riva del Garda 2014 chromatography conference, two interesting debates were discussed in the on-line forum on GCxGC on LinkedIn. At this point, my laboratory does not utilize GCxGC because we think the software is not correctly handling the data processing side. We had been rejected for NIH funding to write appropriate software, so until someone gets it done, we will stick to one-dimensional GC–MS.

SPECIFIC TYPES OF METABOLOMICS STUDIES

You did not talk about analysis of metabolic fluxes. What do you think is the best way to look for the contributions in glucose and glutamine to metabolic pathways in cancer cells?

For primary metabolism, GC–MS gives the best overview, for the three reasons. First, GC gives great theoretical plate numbers, with fantastic overall peak capacities. It's the tool of choice for sugars up to trisaccharides (such as with trimethylsilylation), but also works great for amino acids, hydroxyl acids, monophosphates and keto acids. Second, for keto acids, hydroxyl acids, and amino acids, *tert*-butyldimethylsilylation gives abundant M-*tert*-butyl fragments, formerly called *pseudomolecular ions*, even under electron ionization conditions. The natural abundance of silicon isotope ratios (^{29}Si is 4.7%; ^{30}Si is 3.1%) gives good detectable $(M\text{-}tBu)^+$ and $(M\text{-}tBu)^{2+}$ ion traces even for the "unlabeled" controls in flux experiments. That's good if you want to see increases in ion ratios in your chromatograms! Third, electron ionization with derivatization is a great tool because in electron ionization you see a lot of fragments, unlike too many MS–MS fragmentation spectra in LC–MS. With that information, you can determine the position where isotopes are actually labeled, which is very important to know when you consider alternative pathways! Only for complex lipids (such as palmitate-incorporation into phospholipids) you really need to use LC–MS tools.

In cell-based studies, the energy charge is very important. Can you analyze ATP, ADP, and AMP along with NADPH/NADP and NADH/NAD ratios?

Good question. The answer depends on the cell type and the extraction conditions. It's certainly not easy, but there are services offered by the industry to help determine such ratios. See also the next question.

What happens if the system under study is a dynamic system, such as in the case of in vivo tissue sampling? What else should be considered in those situations?

Metabolism is dynamic. Any sampling must ensure that metabolic activity is quenched as soon as possible. Unfortunately, most professors, myself included, still teach undergraduate students concepts of metabolic homeostasis or steady-state metabolite levels. Such concepts are inadequate at best, but should rather be viewed as borderline to erroneous. That being said, there is a huge variability in metabolic fluxes and enzyme turnover rates, so it really depends which metabolic pathways you are interested in, and to answer which hypotheses. Methods for metabolic quenching are different between bacteria cells, eukaryotes, or human cell cultures. Some tissue samples can be snap-frozen within seconds, whereas others, such as those being handled in clinical surgery, need to be screened by pathologists before they can be frozen. Metabolic energy charge is defined as E = (ATP + 0.5*ADP)/(ATP + ADP + AMP) with a very high turnover of ATP toward ADP and ultimately AMP during sample harvesting. So if your metabolomics analyses show that your cells or tissue samples had energy charge states of around 0.9, you will know that you were very fast and efficient in quenching. However, even if your ATP levels have dropped significantly, you can still use metabolomics data for many other pathways. There are also other types of tissue preservation techniques such as laser-based ultrafast heating, which may be an alternative to snap-freezing, especially when tissues are to be applied to metabolomic imaging. There are some reports on use of formalin-fixed paraffin-embedded tissues in metabolomics (FFPE slides) used in cancer pathology, but these techniques yet need to be validated to be reproducible.