Mass Spectrometry Strategies in Metabolomics*

Published, JBC Papers in Press, June 1, 2011, DOI 10.1074/jbc.R111.238691

Zhentian Lei, David V. Huhman, and Lloyd W. Sumner
From the Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73402

MS has evolved as a critical component in metabolomics, which seeks to answer biological questions through large-scale qualitative and quantitative analyses of the metabolome. MS-based metabolomics techniques offer an excellent combination of sensitivity and selectivity, and they have become an indispensable platform in biology and metabolomics. In this minireview, various MS technologies used in metabolomics are briefly discussed, and future needs are suggested.

Metabolomics is idealized as the large-scale, qualitative, and quantitative study of all metabolites in a given biological system. Unlike transcripts and proteins, the molecular identity of metabolites cannot be deduced from genomic information. Thus, the identification and quantification of metabolites must rely on sophisticated instrumentation such as MS, NMR spectroscopy, and laser-induced fluorescence detection. Each of these technologies has its own unique advantages and disadvantages. Optimal selection of a particular technology depends on the goals of the study and is usually a compromise among sensitivity, selectivity, and speed.

NMR is highly selective and non-destructive and is generally accepted as the gold standard in metabolite structural elucidation, but it suffers from relatively lower sensitivity. Laser-induced fluorescence is one of the most sensitive techniques, but it lacks the chemical selectivity that is critical in structural identification. In contrast, MS offers a good combination of sensitivity and selectivity. Modern MS provides highly specific chemical information that is directly related to the chemical structure such as accurate mass, isotope distribution patterns for elemental formula determination, and characteristic fragment ions for structural elucidation or identification via spectral matching to authentic compound data. Moreover, the high sensitivity of MS allows detection and measurement of picomole to femtomole levels of many primary and secondary metabolites. These unique advantages make MS an important tool in metabolomics (1,2).

Modern MS offers an array of technologies that differ in operational principles and performance. Variations include ionization technique, mass analyzer technology, resolving power, and mass accuracy. The most common ionization techniques in metabolomics include electron ionization, electrospray ionization (ESI), and atmospheric pressure chemical ionization (APCI). Other ionization techniques such as chemical ionization, MALDI, and, more recently, desorption ESI (DESI) and extractive ESI (EESI) have also been used. Mass analyzers with different resolving powers have also been used in metabolomics. These include ultrahigh and high-resolution MS such as Fourier transform ion cyclotron resonance MS (FT-ICR-MS), orbitrap MS, and multipass TOF-MS. However, lower resolution instruments such as ion traps (both linear and three-dimensional quadrupoles) and single quadrupoles are utilized by many. Each of these mass analyzers has its own advantage and limitation. Selection of a specific MS platform for metabolomics depends on the goal of the metabolomics projects, throughput, and instrumental costs. In this minireview, we discuss MS strategies currently incorporated into metabolomics, including direct MS analysis and MS coupled to chromatography for the analysis of the chemically complex metabolome.

Direct MS Analysis

Direct MS analyses sample crude mixtures without chromatographic separations. This approach is the least informative but does provide a high throughput screening tool that is often the only practical choice for large sample numbers such as those encountered during clinical trials or screening of large mutant populations. For example, direct MS analyses enabled successful screening of thousands of yeast mutants to help elucidate the functions of mutated genes (5). Direct MS applicability in metabolomics is broadened by advanced instrumentation capable of high resolution, accurate mass measurements, and tandem MS (such as FT-ICR-MS and orbitrap MS).

FT-ICR-MS is an important and powerful tool in direct MS analyses due to its ultrahigh resolution (>1,000,000) and mass accuracy (<1 ppm). The high mass resolution is useful in empirical formula calculations and compound identification. For example, direct FT-ICR-MS analysis of a crude oil sample revealed >111,000 features in a singular mass spectrum, from which >8300 peaks could be assigned a unique elemental composition (6). Similarly, >1000 unambiguous chemical formulas were reportedly identified from the aerial parts of Arabidopsis using direct FT-ICR-MS (7). The disadvantage of FT-ICR-MS is its formidable cost, which prohibits its widespread availability and routine use in many metabolomics laboratories.

The orbitrap is a relatively newer mass analyzer that uses an electrostatic field to trap ions (8). Orbitraps also have very high

* This work was supported by Oklahoma Center for the Advancement of Science and Technology Grant PSB10-027, National Science Foundation Grants MCB 1024976 and MCB 0520140, the LECO Corp., and Agilent Technologies. Financial personnel support was provided by the Samuel Roberts Noble Foundation. This is the third article in the Thematic Minireview Series on Biological Applications of Mass Spectrometry. This minireview will be reprinted in the 2011 Minireview Compendium, which will be available in January, 2012.

1 To whom correspondence should be addressed. E-mail: lwsumner@noble.org.

2 The abbreviations used are: ESI, electrospray ionization; APCI, atmospheric pressure chemical ionization; DESI, desorption ESI; EESI, extractive ESI; FT-ICR-MS, Fourier transform ion cyclotron resonance MS; CE, capillary electrophoresis; HILIC, hydrophilic interaction liquid chromatography; UHPLC, ultra-HPLC; CZE, capillary zone electrophoresis; CEC, capillary electrophrochromatography.
resolving power (typically 150,000) and excellent mass accuracy (1–5 ppm) and have been widely used in direct MS analyses of bovine lipids (9), yeast sphingolipids (10), and plant metabolites (11). Multiple pass TOF-MS instruments with high resolving power (40,000) and mass accuracy (<5 ppm) have also been recently developed and further facilitate the use of high resolution MS in metabolomics. Three-dimensional and linear ion traps have been used in direct MS analysis of metabolites as well. However, their relatively lower resolution and mass accuracy limit their roles in direct analysis of complex samples.

The development of several new ionization techniques such as DESI, direct analysis in real time, and EESI facilitates the use of direct MS in metabolomics. DESI uses a charged ESI aerosol focused on a separate surface to ionize the surface analytes, allowing not only liquid but also solid sample analyses directly by MS (12). For example, DESI-MS revealed heterogeneous distribution of antifungal compounds on the native surface of seaweed (13) and spatial accumulation of different lipids on rat spinal cords (14). Other techniques such as direct analysis in real-time MS (15) and EESI-MS (4) have been used also in direct MS analyses, including analyses of insect hormones (16); metabolic fingerprinting of human serum (17); screening of pesticides in produce (18); analyses of urine, milk, and polluted water (4); and fingerprinting of olive oil without sample preparation (19). EESI has also been used to directly analyze gaseous samples (20).

Overall, direct MS analyses have been used for a diversity of analytes; however, direct MS methods are very susceptible to ion suppression or enhancement. In addition, direct MS data interpretation can be challenging, as unique metabolite ions are difficult to distinguish from adduct and product ions. Another disadvantage is the inability to differentiate isomers. The majority of these disadvantages can be surmounted by coupling MS to chromatographic separations.

Chromatography Coupled to MS

Coupling chromatography to MS offers an excellent solution to complex mixture analyses and has been extensively used in metabolomics. Chromatographic separation of metabolites prior to MS analyses has several advantages: 1) reduces matrix effects and ionization suppression, 2) separates isomers, 3) provides additional and orthogonal data (i.e. retention time/factor/index) valuable for metabolite annotation, and 4) allows for more accurate quantification of individual metabolites. Currently, three predominant chromatographic techniques have been incorporated in MS-based metabolomics, i.e. GC, LC, and capillary electrophoresis (CE). Multidimensional separation techniques such as two-dimensional GC (GC×GC) and LC (LC×LC) have further enabled the separation of even more complex biological mixtures but are less widely employed. Here, we review the applications of chromatography coupled to MS for metabolite profiling.

**GC-MS**—GC-MS is ideally suited for the analyses of both volatile and nonvolatile compounds following derivatization. The high resolution and reproducible chromatographic separations offered by modern capillary GC make it an excellent tool for complex metabolic mixture analyses. In addition, the standardized MS electron ionization energy of 70 eV leads to reproducible mass spectra and highly transferable electron ionization MS spectral libraries that allow compound identification through mass spectral library matching. The highly reproducible retention indices can also be used for orthogonal confirmation of compound identification such as in identification of stereoisomers that often produce similar mass spectra but distinctly separate in the chromatographic domain. Several spectral libraries have incorporated retention indices such as the National Institute of Standards and Technologies (21) and Metabolomics FiehnLib (22).

Volatile analytes are a specialized class of metabolites that contribute to vegetable and fruit aromas and plant defense responses. Current metabolomics technologies used to study volatiles center on GC-MS coupled to headspace solid-phase microextraction (HS-SPME) or other sorbent-based sampling techniques. HS-SPME is a sensitive and robust technique. The effectiveness of various commercial HS-SPME fibers has recently been evaluated for the analysis of fruit volatiles and led to the identification of 14 novel volatiles (23). GC-MS-based analyses have also been used to identify volatile tomato repellents against whitefly (24) and to profile volatiles from various plant species, including tomato (25) and grape (26).

GC requires volatile and thermally stable analytes such as those described above. However, relatively few compounds meet this requirement in their native state (e.g. short chain alcohols, acid, esters, and hydrocarbons). Many other compounds can be analyzed by GC only following derivatization, i.e. alkylation and silylation (27, 28). An in-liner derivatization method for ultra-small sample volumes (2 µl down to 10 nl) was recently developed and used to profile the intracellular content of frog oocytes (29). Although derivatization is often necessary in GC analyses, it does introduce variability and produce derivatization artifacts.

GC has often been coupled to single quadrupole MS detectors, which have the advantages of high sensitivity and good dynamic range but suffer from slower scan rates and lower mass accuracy relative to TOF-MS detectors. However, the availability, reliability, effectiveness, and affordable cost of GC–quadrupole MS analyzers have made them a popular and robust metabolomics platform. Other mass analyzers such as TOF-MS and triple quadrupoles have also been interfaced to GC. GC–triple quadrupole MS/MS is capable of multiple reaction monitoring of analytes, which can overcome the challenging identification and quantification problems associated with co-eluting analytes in complex matrices. It has been employed to detect multiple pesticide classes in various fruits and vegetables (30–32); to profile sugar in olive fruits, leaves, and stems (33); to reveal responses of cell cultures to external stresses (27); and to detect fatty acid amides in human plasma (34).

GC-TOF-MS technology offers high mass resolution, high mass accuracy, and fast scan speeds. The relatively faster scan rates associated with TOF-MS are extremely useful for the accurate deconvolution of overlapping high resolution or ultrafast GC peaks such as those encountered during complex metabolic mixture analyses. Recent application of GC-TOF-MS in metabolomics includes large-scale metabolite profiling of human serum (28) and plant samples (35–37).
A unique innovation was the development of GC×GC, which offers dramatically increased separation efficiencies and peak capacities (38). In GC×GC, two capillary columns of different stationary-phase selectivity are coupled in series through a flow modulator. Effluents from the first column (usually a long non-polar column) are captured and transferred by the modulator onto the second column. The second column is normally a short polar or semi-polar column that quickly separates the effluent within seconds before the next effluent enters the column. The sharp and narrow peaks generated in fast GC or GC×GC require the use of fast scanning analyzers such as TOF-MS (i.e. approximately >100 Hz) or “semi-fast” scan quadrupoles (i.e. ~20 Hz) (39). Current GC×GC-TOF-MS analyzers can operate with very high acquisition rates, typically up to 500 Hz, and offer higher resolution and sensitivity. Recent metabolomics applications of GC×GC-MS include animals (40), plants (41–44), microorganisms (45), and other samples such as human serum and tissues (46–48). Fig. 1 (reprinted from Ref. 41 with permission) shows a recent application of GC×GC-TOF-MS to resolve plant terpenoids.

Although both normal-phase (NP) and reversed-phase (RP) columns have been employed in metabolomics, RP columns such as C\(_{18}\) and C\(_{8}\) are by far the most utilized. However, NP separations provide complementary views of the metabolome, as demonstrated in metabolic profiling of urine using hydrophilic interaction liquid chromatography (HILIC)-MS and RP-HPLC-MS (49). HILIC is ideal for highly polar and ionic compounds and therefore suitable for samples that contain predominantly polar metabolites such as urine. LC-MS using conventional C\(_{18}\) columns with particle sizes of 3–5 \(\mu\)m has been widely used in metabolomics elucidation of plant secondary metabolism (27, 50–52). Many established lipidomics programs also rely upon LC-MS for the large-scale study of cellular lipids (53, 54), which has improved our understanding of lipid metabolism, signaling, and neurodegenerative disorders such as Alzheimer disease (55–57).

The development of fast and more efficient ultra-HPLC (UHPLC), which utilizes higher pressures (12,000–15,000 p.s.i. compared with ~6000 p.s.i. for HPLC) and sub-2-\(\mu\)m packing particles, has substantially increased chromatographic resolution and peak capacity compared with HPLC. Fig. 2 shows a UHPLC-QTOF-MS base peak chromatogram of a highly complex metabolic plant mixture. The superiority of UHPLC has also been demonstrated by Nordström et al. (58), who reported that UPLC-MS resulted in a >20% increase in detectable components compared with a similar HPLC-MS-based approach.
More recently, HILIC-UHPLC-MS has been introduced and used in urinary metabolic profiling (59).

LC×LC utilizing two columns of different selectivity offers an effective platform for separating both polar and non-polar compounds simultaneously. The peak capacity of two-dimensional HPLC is much higher and is the product of the two independent dimensional peak capacities given that the first and second separations are truly orthogonal. The higher peak capacities offer greater metabolome coverage. Fig. 3 (reprinted from Ref. 60 with permission) illustrates NP/HPLC-RP and RP/HPLC-RP two-dimensional HPLC separations, with the NP/HPLC system achieving an overall peak capacity of 1095 when applied to the analysis of a lemon oil extract (60). LC×LC is typically superior to one-dimensional HPLC even if the columns used in two-dimensional LC are of similar chemistry (61). LC×LC-MS has been employed for the analyses of carotenoids in different orange juices (62), drug metabolism (63), and triacylglycerols in vegetable oil (64). LC×LC-MS has also been tailored for target analyses. For example, Aturki et al. (65) used an RP/C_{18} column in the first dimension to separate flavanone-7-O-glycosides from a complex sample and then used a carboxymethylated β-cyclodextrin-based column in the second dimension to resolve the individual flavanone-7-O-glycoside stereoisomer. The disadvantages of LC×LC are its relatively complex setup and the loss of sensitivity due to a sample dilution effect in the second dimension (66).

The ionization technique selected for LC-MS-based metabolomics can also have a substantial impact on metabolite profiles. Generally, ESI is ideal for semi-polar and polar compounds, whereas APCI is more suitable for neutral or less polar compounds. These two ionization techniques provide complementary data; both are desirable in large-scale non-targeted metabolomics. For example, complementary ESI-APCI analyses resulted in an ~20% increase in the number of detected ions in a human blood serum extract (67). Multi-ionization mode or dual ESI-APCI ion sources are now commercially available from some instrument vendors, allowing for simultaneous acquisition of ESI and APCI data.

Many modern MS instruments are now capable of fast polarity switching during data acquisition and have been exploited in simultaneous acquisition of both positive and negative ion mode data (68, 69). The use of both positive and negative ionization LC-MS offers more comprehensive metabolome coverage than the use of a single polarity (50, 67). Fig. 4 (reprinted from Ref. 50 with permission) shows the positive and negative ion mode HPLC-MS chromatograms of *Medicago truncatula* cell extracts. Several analytes were detected only in the negative ion mode, whereas others were observed only in the positive ion mode. Similarly, Nordström et al. (67) noted that >90% of the human blood plasma ions observed in the positive ion ESI mode were not found in the negative ion mode and vice versa.
It is expected that LC-MS will continue to play an important role in MS-based metabolomics, and with the continuous advancement of LC and MS technologies, both the sensitivity and depth of coverage of LC-MS-based metabolomics will continue to improve. As demonstrated in a recent human serum metabolome study, higher metabolome coverage can best be achieved by using multiple metabolomics technologies (70).

Using a combination of platforms, including LC-MS/MS, GC-MS, TLC-GC-flame ionization detection, direct infusion MS/MS, and NMR, 4229 compounds were tentatively identified from human serum, with each platform identifying a subset of unique compounds (70).

**CE-MS**—CE separates analytes based on charge and size, and it is particularly suitable for the analysis of highly polar and...
MINIREVIEW: MS in Metabolomics

Ionic metabolites. Separation of neutral compounds can be achieved using micellar electrokinetic chromatography, which employs charged surfactants such as SDS to form charged micelles containing the analyte. CE is a fast, relatively inexpensive, and highly efficient separation technique. Capillary zone electrophoresis (CZE) is the most utilized separation technique in CE-MS-based metabolomics because many compounds can be separated readily by CZE. CE can be interfaced with various MS analyzers; however, TOF-MS is the most commonly used CE-MS analyzer due to its fast acquisition rates, which are necessary to statistically sample the narrow CZE peaks. ESI is the ionization technique of choice for CE-MS.

CE-MS has been used in both targeted and non-targeted metabolomics. CE-TOF-MS was used for global profiling of endogenous metabolites in tumor and normal tissues, and the results revealed elevated glycolysis in tumor tissues evidenced by extremely low glucose, high lactate, and high glycolytic intermediates in tumor tissues (71). CE-TOF-MS metabolic profiling of *Illicium anisatum* seed, pulp, stem, and leaf tissues detected >1000 tentative polar metabolites in 40 min and revealed spatial distributions of numerous metabolites (72). Use of capillary electrophoretic chromatography (CEC) coupled to MS for metabolomics has been reported as well. CEC achieves separations using an electrostatic field imposed on a packed particle or a monolithic column, which allows for the high separation efficiency of CE with the high selectivity of the stationary phase of LC columns. For example, CEC-ESI-MS methods have been developed and used to analyze the metabolome of a human hepatocellular carcinoma cell line (73) and drug abuse through urine analysis (74).

Unfortunately, CE-MS has inherent limitations. These are mainly low sensitivity, poor reproducibility, and electrochemical reactions of metabolites. Recently, the performances of GC-MS, LC-MS, and CE-MS were compared in quantitative metabolomics, and it was concluded that CE lacked the necessary robustness and was the least suitable platform for analyzing complex biological samples (75).

**Progressive MS-based Metabolomics Applications and Quantifications**

The above noted MS technologies form the core analytical platforms used by most in metabolomics. However, metabolomics applications and techniques continue to expand. Several highlighted areas are discussed below, including spatially resolved metabolomics, fluxomics, integrated metabolomics, personalized medicine, and computational methods for metabolite annotation.

Advanced sampling and MS technologies have made it possible to perform spatially resolved metabolomics, which can be achieved through the profiling of laser-dissected tissues (76), single cell sampling using microcapillaries (77), or metabolite imaging MS. Microdissection and single cell sampling using microcapillaries provide information on differential metabolite accumulation in specific tissue or cell types, whereas imaging MS enables the spatial visualization of metabolites and their relative abundances across various tissues. Spatial metabolomics can be used to decipher the functional roles of the metabolites based upon their localization and/or co-localization with other metabolites/proteins/transcripts.

Flux analyses or fluxomics is another important application of metabolomics (78) and has been used for mechanistic studies (79), integrated metabolomics for gene discovery (80) and personalized medicine (81). For example, analysis of flux through the TCA cycle in a human cancer cell line confirmed glutaminolysis and reductive carboxylation as the major cancer cell pathways that provide nitrogen for amino acid and nucleotide (adenine) syntheses (78). Differential regulation of the same metabolic pathways in response to different elicitors in plant cells was identified through large-scale metabolic profiling (79). Combined with transcriptomics, metabolomics was successfully used to identify two *Arabidopsis* Myb transcription factor genes that regulate aliphatic glucosinolate biosynthesis (80). The emerging MS-based omics technologies are providing a systems approach to disease and are transforming medicine from reactive to proactive (i.e. predictive, personalized, preventive, and participatory or “4P”) (82).

In general, a significant proportion of profiled metabolites remain unannotated, but several groups are creating large MS/MS spectral libraries as part of an advanced scheme to characterize and cross-correlate both known and unknown metabolites. For example, Matsuda et al. (83) have generated an MS2T library that contains nearly 1.6 million MS2T spectra to facilitate peak annotation in non-targeted plant metabolomics studies. Annotation of metabolites through MS or MS/MS spectra typically involves spectral matching against spectral libraries compiled with authentic standards. Identification of metabolites whose MS/MS spectra are not present in the spectral libraries remains very challenging and is being addressed through both empirical and computational MS. MetFrag, a computer program for metabolite identification based upon *in silico* spectral fragmentation matching, was recently described (84). It is expected that continued development of computational methods such as MetFrag and other fragment tree-related software (85) will further assist metabolite identification in metabolomics.

Quantification in metabolomics is critical in understanding biological processes, and it is generally performed in two manners, i.e. relative or absolute quantification. Relative quantification, which normalizes the metabolite signal intensity to that of an internal standard or another relative metabolite, is typically used in large-scale non-targeted profiling. Absolute quantification uses external standards or internal isotopically labeled standards to determine the absolute metabolite quantity and is used mostly in targeted metabolomics. The major obstacle in metabolite quantification is that the metabolite’s signal intensity is dependent not only on its concentration but also on its chemical structure and matrix. Ion suppression and enhancement caused by matrix effects can result in inaccurate quantification of the metabolites. Stable isotope dilution is one solution for absolute quantification (i.e. stable isotope-labeled standards are added to the samples to account for sample processing variation and matrix effects encountered during MS analysis). Absolute quantification in metabolic profiling of obese and lean humans clearly linked branched chain amino acid-related metabolites to the development of obesity-associ-
ated insulin resistance (86). Quantitative MS intermediary metabolite profiling and 13C NMR-based flux analyses identified a critical link between pyruvate and the TCA cycle mediated by pyruvate carboxylase, which plays an important role in regulating glucose-stimulated insulin secretion (87, 88). Quantitative lipidomics revealed that abnormal lipid profiles occur even at the very earliest stages of diabetes (89). Stable isotope dilution GC-MS of cholesterol in human plasma revealed that oxysterols are sensitive and specific biomarkers for Niemann-Pick type C1 disease (90).

It is clear that stable isotope standards are critical for absolute quantification in metabolomics. However, the availability of commercial isotope-labeled standards is limited, and the costs can be prohibitive to large-scale use. Several groups have begun to chemically synthesize a diverse range of stable isotope-labeled compounds for absolute quantification. These include fatty acids (91), human steroids (92), and plant hormones (93), but this approach is laborious and requires synthetic expertise not available in every metabolomics group. In addition, preparing multiple isotope-labeled standards for the entire metabolome is challenging given the diverse and complex structures of the metabolites. One way to circumvent this problem is to introduce a stable isotope tag to the metabolites through chemical labeling. A dimethylation method to label amine-containing metabolites using commercially available and inexpensive 13C- or deuterium-labeled formaldehyde has been reported (94). Standards of known concentrations are labeled in the same manner and used to generate calibration curves for absolute quantification.

Conclusions and Future Challenges

MS has become an indispensable productive tool in metabolomics due to its high sensitivity and selectivity. However, there are still many metabolomics challenges, including limited dynamic range, lack of comprehensive coverage, and limited metabolite annotations. Currently, the best dynamic range of modern MS is ~106, which is significantly lower than the estimated concentration range of cellular metabolites of 1012 or more. In addition, the estimated number of metabolites within a given plant species can be 10,000 or more, and the current metabolome depth of coverage is approximately <20%. Thus, there is substantial need for improvement. Advancements and solutions to the above limitations will continue to expand the scope and propel the utility of MS in metabolomics.
