Journal of proteome reviews research

# From Exogenous to Endogenous: The Inevitable Imprint of Mass Spectrometry in Metabolomics

**Elizabeth J. Want, Anders Nordström, Hirotoshi Morita, and Gary Siuzdak***

*Department of Molecular Biology, The Scripps Center for Mass Spectrometry, 10550 North Torrey Pines Road, La Jolla, California 92037*

**Received September 27, 2006**

Mass spectrometry (MS) is an established technology in drug metabolite analysis and is now expanding into endogenous metabolite research. Its utility derives from its wide dynamic range, reproducible quantitative analysis, and the ability to analyze biofluids with extreme molecular complexity. The aims of developing mass spectrometry for metabolomics range from understanding basic biochemistry to biomarker discovery and the structural characterization of physiologically important metabolites. In this review, we will discuss the techniques involved in this exciting area and the current and future applications of this field.

**Keywords:** mass spectrometry • liquid chromatography • metabolomics • biomarker characterization • metabolite database

## Introduction

The application of modern mass spectrometry technology to endogenous metabolite research derives from its success in drug metabolite studies, both quantitative and structural.[1−11] Interest also originates from the ability to perform more comprehensive metabolite analyses with new liquid chromatography/mass spectrometry (LC/MS) technologies, such as nanoESI-LC/MS, and the desire to unravel basic biochemical events of cells and tissues, or to identify disease or pharmaceutical biomarkers.

One of the first metabolite profiling experiments was by Pauling and colleagues in 1971, who analyzed the metabolite content of human urine vapor and breath of subjects on a defined diet using gas chromatography (GC).[12] Approximately 250 substances were detected in a breath sample and 280 in a urine vapor sample. This group then went on to profile amino acids in urine, employing nonparametric statistical analysis for detecting profile differences related to gender and other variables.[13] This was the beginning of what we now call metabolomics, the aim of which is to provide a comprehensive profile of all the metabolites present in a biological sample.

From the 1970s, gas chromatography mass spectrometry (GC/MS) became popular for metabolite profiling and is still used for the detection of many metabolic disorders.[14] Advantages of GC/MS include high resolution and reproducibility, as well as the availability of EI spectral libraries for structural identification.[15] In addition, since the 1990s,[16] nuclear magnetic resonance (NMR) has also been applied to areas such as plant metabolism, Duchenne Muscular Dystrophy, neurological disorders, and hepatotoxicity and nephrotoxicity in rodents,[17−24] with advantages in both speed and accuracy. However, because of the limitations of NMR in terms of sensitivity, LC/MS has emerged as a powerful alternative technology for metabolomics. In this review, the role of mass spectrometry in metabolomics will be discussed, encompassing data acquisition, data analysis, metabolite characterization, and many exciting applications.

## 1. Data Acquisition

Due to the complex nature of biological samples, separation is often performed before mass spectrometric analysis to achieve the detection of as many metabolites as possible. Traditionally, GC was employed, as it is well-known for high resolution and reproducibility. However, disadvantages of GC include convoluted sample preparation (such as derivatization), lengthy analysis time, and the limitation on the size and type of molecule that can be analyzed (nonvolatile, polar macromolecules are unsuitable). However, GC-MS is still widely used in plant metabolomics due in part to the nature of the metabolites being investigated.[15,25−28]

Liquid chromatography electrospray ionization mass spectrometry (LC/ESI-MS) (Figure 1) is now a common metabolomics tool. Separation of the thousands of molecules present in biofluids using LC can reduce ion suppression[29−31] by decreasing the number of competing analytes entering the mass spectrometer ion source at any one time. This results in a selective approach that allows for both quantitation and structural information, where sensitivities in the pg/mL range can be achieved readily.[32] LC/MS techniques have replaced some of the traditional specialized clinical laboratory methods[33,34] that used immunological, fluorometric, and biological techniques.[35]

An important factor in LC metabolite separation is the choice of column. Many biofluids, particularly urine, contain a vast array of highly polar molecules that are not retained well on the more traditional reverse phase (RP) LC columns. Normal phase techniques, which result in the elution of less polar molecules first and thus the retention of more polar molecules,

10.1021/pr060505+ CCC: $37.00 



Published on Web 11/18/2006

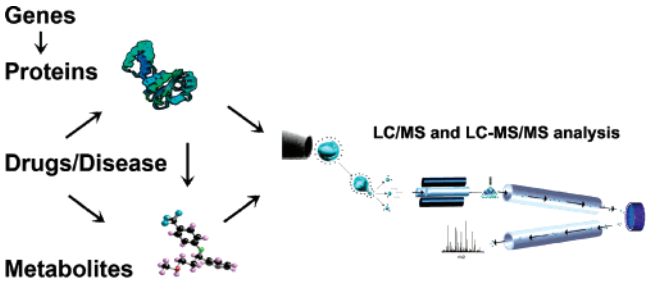


**Figure 1.** Metabolomics aims to measure all the metabolites in a biofluid or tissue. Common approaches include LC-MS and LC-MS/MS using electrospray ionization. Alterations in metabolite levels may reflect the activity of their corresponding enzymes. Additional factors affecting metabolite profiles include drug intake or the onset or progression of a disease. Metabolomics is complementary to proteomics and transcriptomics and the combination of data from all three approaches can provide important information regarding the status of a cell or organism.

require a different solvent system to that used by RP chromatography, typically containing no aqueous. A newer approach is hydrophilic interaction chromatography (HILIC), which can offer complementary information to that obtained using RP chromatography.[36] Here, water and acetonitrile can still be used, although starting from a high organic content and ending at high aqueous. HILIC approaches combined with ESI-MS techniques have already been applied to the analysis of dichloroacetic acid in rat blood and tissues,[37] plant metabolites such as oligosaccharides, glycosides, and sugar nucleotides,[38] and with APCI mass spectrometry for the determination of 5-fluorouracil in plasma and tissues.[39]

The ability of LC to separate complex mixtures prior to mass analysis comes at a cost of speed. An alternative to traditional reverse phase (RP) approaches is ultrahigh performance liquid chromatography (UPLC),[40] which utilizes columns with much smaller particle size packing material (1.4−1.7 $\mu$m) than traditional columns, thus allowing for improved separation and higher resolution (Figure 2). This technology permits pumping and injection of liquids at pressures exceeding 10 000 psi.[41] Using this approach, sample analysis times can be reduced to as little as 1 min,[42] resulting in much higher throughput. With UPLC, narrower chromatographic peaks can be achieved (peak widths at half-height <1 s), resulting in increased peak capacity, lower ion suppression and improved signal-to-noise ratio, and thus increased sensitivity (Figure 2). Recent studies comparing UPLC and HPLC for their application to metabolomics studies showed that UPLC can detect more components than HPLC,[32] with a 20% increase reported over the same chromatographic length.[43] This study also showed UPLC to display superior retention time reproducibility and signal-to-noise ratios over HPLC.

When coupled to separation techniques, MS analysis of biofluids can offer high sensitivity and specificity. However, despite LC/MS being the foremost technique for the analysis of known compounds,[44] as well as the determination of unknowns using MS/MS, one limitation is the inability of LC/MS alone to unequivocally distinguish between some coeluting stereoisomers.[45] However, the application of ion-mobility mass spectrometry to metabolomics might be a powerful strategy for addressing the problem of resolving isomers. Indeed, an LC approach has been combined with ion mobility/time-of-flight (TOF) mass spectrometry for the characterization of a combinatorial peptide library, enabling many peptide isomers with identical masses and retention times to be resolved.[46]

Furthermore, as LC-MS techniques for metabolomics can be affected by high noise levels, retention time shifts, and high variability in signal intensities, researchers are constantly investigating ways to reduce analysis time and sample preparation in metabolomics studies. There have been some recent explorations of chip-based mass spectrometry approaches for the delivery of the biological sample to the mass spectrometer with the aim of improving metabolite detection. One recent study using protein precipitation of plasma combined with chip-based nanospray infusion reported high reproducibility, sample throughput, and the observation of over 1800 different mass peaks up to 900Da.[47] Some of the samples were highly diluted to minimize ion suppression, and so although MS runs of 10 min were used, for MS/MS studies, runtimes of 60−90 min were needed to obtain meaningful data, offering no advantages over LC-MS/MS. In fact, using UPLC-MS/MS impressive fragmentation data can be collected in a run of 10 min.[48]

**Ionization Techniques.** Once the components of a biological sample have been separated, ions must be produced. In GC, samples are vaporized and then ionized by electron-impact (EI) or chemical ionization (CI). Extensive libraries of EI spectra, such as the NIST database, which contains over 100 000 compounds, are available to aid in the identification of molecules (http://www.nist.gov/srd/nist1a.htm). EI has the advantages of good sensitivity and unique fragmentation. However, the molecular ion is often not detected due to the extensive fragmentation, which may prove hinder the identification of unknown compounds. A disadvantage with EI is the limited mass range due to the thermal desorption requirement. As CI is much less energetic than electron ionization, it induces less fragmentation and in general, more stable ions, and so can be useful for identifying the molecular ion and thus determining the molecular weight of a compound. However, CI still requires thermal desorption. Negative CI is particularly sensitive for perfluorinated derivatives and proves a limited but powerful approach for certain derivatized molecules such as steroids.

However, for metabolomics studies, electrospray ionization (ESI) is most commonly used in conjunction with LC/MS. ESI offers soft ionization, excellent quantitative analysis and high sensitivity. With ESI, ions are generated directly from the liquid phase into the gas phase, establishing this technique as a convenient mass analysis platform for both liquid chromatography and automated sample analysis. In its simplest form, ESI can be quite effective even without separation, especially when combined with tandem mass spectrometry (MS/MS) where its direct application to metabolite screening is currently used for over 35 diseases.[49,50]

Three alternative solution-based ionization strategies to ESI are also being used for LC/MS-based metabolomics, namely nanoESI, atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI). NanoESI liquid chromatography, performed at low flow rates (~200 nL/min), has already proved useful in proteomics studies[51,52] where it significantly enhances sensitivity and dynamic range.[53−55] In nanoLC/nano-ESI-MS, ions are produced from small submicron sized droplets requiring less evaporation and a greater ability to focus the resulting ions into the analyzer, therefore increasing sensitivity and ultimately offering a greater dynamic range. APCI and APPI are widely used in the pharmaceutical industry[56−58] yet have had limited exposure to metabolomics

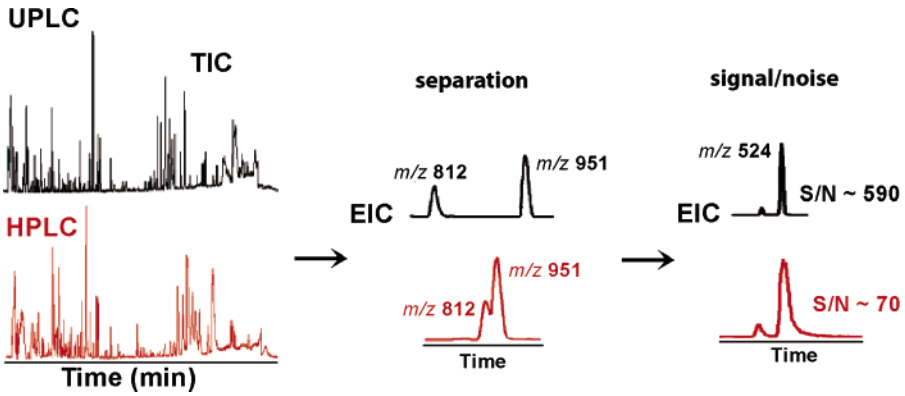


**Figure 2.** Ultrahigh performance liquid chromatography (UPLC) utilizes columns with smaller particle size packing material (1.4−1.7 $\mu$m) than traditional columns and can enhance several aspects of chromatography in a metabolomics context. (1) Separation of metabolites is improved, decreasing ion suppression and in turn improving data interpretability (2) Signal to Noise (S/N) is improved due to narrower peak widths allowing for increased peak capacity and improved accuracy and sensitivity. (3) Sample run time is decreased dramatically allowing for faster sample throughput.

studies. Analogous to the ESI interface, APCI and APPI typically induce little or no fragmentation and are considered robust and relatively tolerant of high buffer concentrations. It is now recognized that these approaches can be valuable for the analysis of nonpolar and thermally stable compounds such as lipids[59,60] with the apparent trend toward a "single" ionization source containing combinations of ESI and APCI or ESI and APPI.

There is a small but growing body of work using other ionization strategies for metabolomics. MALDI applications have been limited, due in part to matrix suppression issues for low molecular weight molecules. However, some researchers believe that there could be advantages to using this approach. Using a novel sample deposition approach, semiquantitation of amino acids from mammalian cells has been achieved using positive mode MALDI-TOF−MS.[61] Negative mode MALDI, rarely used due to the lack of suitable matrices, has been applied recently for the analysis of metabolites from Islets of Langerhans and *E. Coli*[62]. In all, over 100 metabolites were detected, although many could not be identified due to the lack of complete databases and the inability to distinguish isomers such as citrate and isocitrate. Recently, a matrix-suppressed laser desorption/ionization (MSLDI) strategy was evaluated. By decreasing the matrix/analyte ratio, less suppressed spectra were obtained, enabling the detection of lower abundance compounds.[63] However, it can be difficult to find a suitable matrix/analyte ratio without prior knowledge of metabolite concentrations.

Desorption ionization on porous silicon (DIOS) allows for the detection of small molecules in both positive and negative mode with little background interference[64] and has recently been applied to metabolomics studies.[65] Here, 26/30 known metabolites in a mixture were detected rapidly in positive mode and in negative mode, showing the potential of DIOS as a metabolomics approach.

Recently, the combination of desorption electrospray ionization mass spectrometry (DESI-MS) and NMR was investigated for its application to metabolomics.[66] This group studied urine without any sample preparation to differentiate between diseased and healthy mice. DESI is an ambient ionization direct analysis technique, providing high sensitivity and specificity with minimal sample preparation.[66] There is no sample separation and because the sample is placed on the surface rather than direct infusion, this affords a higher tolerance to salts.

However, some compounds do not ionize well using any of the common ionization techniques and so will not be detected using MS alone. The coupling of NMR and MS has been used in combination with liquid chromatography (LC-NMR−MS) and applied to metabolite studies, such as in the pharmaceutical drug discovery area.[44,67] This technique allows for the both MS and NMR data to be collected from a single LC run and the complementary information that can be provided makes this approach a powerful tool for the detection and identification of both known and unknown compounds.[44] Further, software is being developed to cope with the analysis of the complex data produced by these instruments, in particular, statistical heterospectroscopy (SHY), an approach to the integrated analysis of NMR and UPLC-MS data sets.[68]

**Mass Analyzers.** Along with advances in ionization sources, mass analyzers have improved with respect to speed, accuracy, and resolution. The most common mass analyzers are the quadrupole and time-of-flight (TOF) based analyzers. Other analyzers that can be used for metabolomics studies include ion traps, Fourier transform mass spectrometers (FTMS), and orbitraps, some of which will be discussed in this section.

Quadrupole mass analyzers can be coupled to many different ionization sources, with advantages including comparatively high pressure tolerance, good dynamic range, and excellent stability, all at a relatively low cost. To perform tandem mass analysis with a quadrupole instrument, three quadrupoles are placed in series. Each quadrupole has a separate function: the first quadrupole (Q1) scans across a preset $m/z$ range to select an ion of interest, which is then fragmented in the second quadrupole (Q2), the collision cell, using argon or helium as the collision gas. The third quadrupole (Q3) analyzes the fragment ions generated in the collision cell (Q2).

The linear time-of-flight (TOF) mass analyzer is the simplest mass analyzer, with virtually unlimited mass range, whereas the TOF reflectron has mass range up to $m/z \sim$10 000. The TOF reflectron is now widely used with ESI and MALDI, and more recently for electron ionization in GC/MS applications. TOF instruments offer high resolution, fast scanning capabilities (ms), and accuracy on the order of 5 part per million (ppm). Quadrupole-TOF (Q-TOF) mass analyzers combine the stability of a quadrupole analyzer with the high efficiency, sensitivity, and accuracy of a time-of-flight reflectron mass analyzer, and are typically coupled to ESI sources. Q-TOF mass analyzers are an obvious choice for obtaining metabolite fragmentation data.

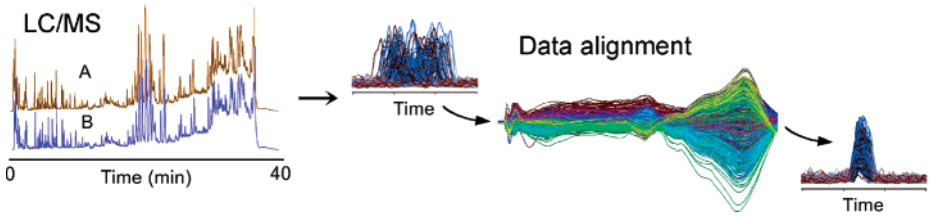


**Figure 3.** LC/MS metabolomics studies generate large, complex datasets, and there is the possibility of retention time drift between samples over the course of the study. Alignment of chromatographic data is fundamental for the production of comparable data sets. A data extraction tool like XCMS allows for nonlinear correction of retention time drift in the time domain. Other freely available data processing software includes MZmine and MET-IDEA.

The quadrupole can act as any simple quadrupole analyzer to scan across a specified $m/z$ range, but can also be used to selectively isolate a precursor ion and direct that ion into the collision cell. The resultant fragment ions are analyzed by the TOF reflectron mass analyzer. Q-TOF analyzers offer significantly higher sensitivity and accuracy over tandem quadrupole instruments when acquiring full fragment mass spectra.

The ion trap mass analyzer can be used for both MS scanning and MS/MS studies. It allows the isolation of one ion species by ejecting all others from the trap, whereby the isolated ions can subsequently be fragmented. However, a major limitation of the ion trap is its inability to perform high sensitivity triple quadrupole-type precursor ion scanning and neutral loss scanning experiments. Furthermore, the upper limit on the ratio between precursor $m/z$ and the lowest trapped fragment ion is ~0.3 (the "one-third rule"). The dynamic range is also limited due to space charge effects when too many ions are in the trap, which diminish the performance of the ion trap. Here, the linear ion trap has an advantage over the 3D trap, with a larger analyzer volume which lends itself to a greater dynamic range and an improved range of quantitative analysis. Quadrupole ion traps have $MS^n$ capabilities, allowing for multiple MS/MS experiments to be performed quickly without having multiple analyzers, such that real time LC-MS/MS is now routine. Other important advantages of quadrupole ion traps include their compact size, and their ability to trap and accumulate ions to provide a better ion signal.

Fourier transform mass spectrometry (FTMS) offers high resolution and the ability to perform multiple collision experiments (MSn). FTMS is capable of ejecting all but the ion of interest, fragmenting the selected ion and yielding high-accuracy fragment masses. Ultrahigh resolution FTMS can be coupled to MALDI, ESI, APCI, and EI, although the new quadrupole-FTMS and quadrupole linear ion trap-FTMS mass analyzers are typically coupled to electrospray ionization sources. Newer hybrid instrument designs are preferable over coupling FTMS/MS to separation techniques such as LC, as MS/MS experiments can be performed outside the magnet.[69] This presents some advantages because high resolution in FTMS is dependent on the presence of high vacuum. Performing MS/MS experiments outside the cell is thus faster because the ICR cell is not exposed to a pulse of gas to initiate dissociation and thus can be maintained at ultrahigh vacuum. Instruments, such as the LTQ-FT, combine the excellent performance and capabilities of the FT mass spectrometers with the well-established, tested, and validated features of quadrupoles and ion traps. Due to the robust, externally calibrated accurate mass determination for both parent and product ions, the LTQ-FT could be a very powerful analytical tool for metabolomics studies, allowing for the confirmation of known metabolites or to elucidate the structures of unknown metabolites.

## 2. Data Analysis

LC/MS based metabolomics studies generate large, complex datasets which require sophisticated software to enable interpretation. A current challenge is achieving the high-throughput conversion of these datasets into organized data matrices necessary for further statistical processing, as well as visualization. As two or more sample sets are often compared for changes in metabolite levels, metabolites must first be detected in all samples, matched between the samples and then their levels compared (Figure 3). Multiple adducts (such as sodium and potassium) can be formed using LC-ESI/MS, thus complicating the data produced by increasing the number of peaks detected. It is imperative that the same metabolites are identified correctly in all samples to enable this comparison. To this end, software has been produced in order to allow peak picking and evaluation. Many instrument manufacturers have produced their own software, which often works solely with data generated from a particular instrument. These include MarkerLynx (Waters), MassHunter (Agilent) and MarkerView (Applied Biosystems/MDS SCIEX). However, some researchers desire the freedom to modify many parameters and also to compare data from different instruments and so have developed their own software. Examples of these include MZmine,[70,71] XCMS,[72] and MET-IDEA[73] and are generally freely available for download and in some cases, user modification.

A huge challenge with metabolomic data analysis can often be the classification of data, either with or without prior knowledge that such classes exist. The development of new data analysis approaches[74,75] including multivariate statistical analysis for biomarker discovery[5,28,76,77] has facilitated the discovery of hidden structure in data. Therefore, with many of these software programs, data (as peak lists or similar) can be output in a suitable format to then be analyzed using multivariate statistics (Figure 4). These multivariate techniques can help to discern peaks with high discriminating power between the sample groups being analyzed, i.e., potential biomarkers.[78] Most companies and research groups involved in metabolite research supplement these available data mining techniques with in-house software to further enable compound identification and quantification.

A variety of multivariate statistics and pattern recognition methods are currently in use for metabolomics studies, which can be divided into two categories, unsupervised and supervised methods. In unsupervised methods, such as principal component analysis and hierarchical cluster analysis, the algorithm is not given a training set and so input data is

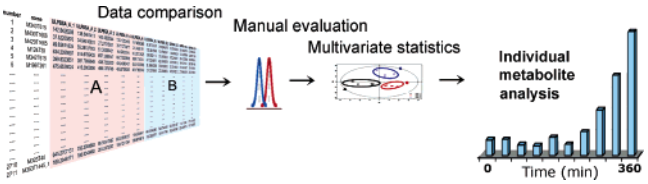


**Figure 4.** Once data has been aligned, the output can be manually inspected before data are further processed using multivariate statistics. These multivariate techniques can help to discern potential biomarkers, which then undergo targeted analysis to further validate their significance as biomarkers.

classified in an "unsupervised" manner. Conversely, with supervised methods, a classification system is given some input data together with the answers, known as the "training set", which can be used to build a model and estimate necessary parameters. These include discriminant analysis, such as projection to latent structures (also called partial least-squares) (PLS) and orthogonal projection to latent structures (O-PLS), artificial neural networks (ANNs), and evolutionary-based computational algorithms.

Principal component analysis (PCA)[79] is often used for metabolomics.[80,81] PCA can be used in the reduction of data dimensionality, to investigate clustering tendency, such as with gene expression data,[82] to detect outliers, and to visualize data structure.[83,84] However, PCA gives a simplified representation of the information contained in the spectra and cannot generally use additional information about the data, such as class information. Therefore, PCA is often followed by a supervised analysis technique such as PLS-DA or O-PLS-DA. In fact, Lutz and colleagues showed by comparison of PCA with PLS-DA that there was a clear advantage in using a supervised model where class details are known.[78]

Hierarchical cluster analysis organizes information about variables in a data set, forming "clusters", where the degree of association is strong between samples within the same cluster and weak between those in different clusters. This approach may reveal associations and structure in data that were not previously evident. Hierarchical clustering can be represented as a tree, or dendrogram, where each step in the clustering process is illustrated by a join of the tree. The combination of proteomics and cluster analysis has been applied successfully to the classification of normal breast, benign breast and breast cancer tissues using just the protein expression profiles.[85]

Projection to latent structures, also called partial least-squares discriminant analysis (PLS-DA) is performed in order to enhance the separation between groups of observations, often by rotating PCA components to achieve maximum separation between classes, and to understand which variables are responsible for separating the classes. Orthogonal projection on latent structure discriminant analysis (O-PLS-DA), developed by Trygg and Wold, can be a powerful tool for the analysis of metabolomics data.[86,87] Like PLS-DA, O-PLS-DA is a supervised pattern recognition technique, but has improved predictive quality because the structured noise is modeled separately. O-PLS-DA has been used in conjunction with STOCSY (statistical total correlation spectroscopy) in the analysis of NMR metabolomics data.[88]

ANNs are powerful data modeling tools, capable of learning patterns and relations from input data, making good pattern recognition engines and robust classifiers. ANNs are being used effectively for problems including building nonlinear classification and regression models. Currently, ANNs are being developed which can predict patient responses to drugs, which would enable ideal dosing regimes to be established.[89]

A newer approach to the mining of highly complex metabolomics data is to apply evolutionary computational-based methods.[90] These are explanatory supervised learning techniques, including genetic algorithms, genetic programming, evolutionary programming and genomic computing, which could be ideal strategies for mining such high-dimensional data as that obtained from metabolomic studies.[90]

However, there appears to be no consensus on which multivariate statistics approach is truly superior, and so at present it seems that individual companies and research groups are employing their own combination of data analysis software and multivariate statistics to address their individual metabolomics challenges.

**Databases.** The collection of LC/MS data and subsequent comparative analysis is becoming more straightforward, yet a major challenge lies in characterizing the metabolites that have interesting biological properties and whose mass is initially identified. In contrast to the well-annotated gene and protein databases that can be searched easily, at present, no such comprehensive tools exist for metabolite researchers. However, current metabolite databases, although incomplete, offer a starting point for characterization. Among the databases currently available, the most widely used are the NIST database, which includes mass spectral data for some known metabolites (http://www.nist.gov/srd/nist1.htm), as well as the KEGG, HumanCyc, ARM, and METLIN databases. The KEGG database is a valuable resource for metabolomics researchers (http://www.genome.jp/kegg/ligand.html). HumanCyc (http://biocyc.org) includes known metabolites as well as those predicted by algorithms which project metabolic pathways from a genomic sequence. A database constructed as part of the Atomic Reconstruction of Metabolism (ARM) project, compiles metabolite structures together with molecular weight and MS fragmentation data (http://www.metabolome.jp). In addition, the University of Alberta hosts a mini-library of full mass spectra of newer drugs, metabolites and some breakdown products, (http://www.ualberta.ca/~gjones/mslib.htm). Other databases include the human metabolite database (http://www.hmdb.ca/), which acts as an electronic repository for identification of small molecule metabolites. The Spectral Database for Organic Compounds SDBS provides access to a wealth of spectra of organic compounds (NMR, MS, IR). Another metabolite database is the "tumor metabolome" database, established at the Justus-Liebig University Giessen in Germany (http://www.metabolic-database.com).

LIPID MAPS (http://www.lipidmaps.org/tools/index.html) and Lipid Search (http://lipidsearch.jp/LipidNavigator.htm) are useful databases to search lipid metabolites. Although phospholipids and some other lipids are important metabolites, numbers of the registered secondary metabolites are still limited.

The KNApSAcK database (http://kanaya.aist-nara.ac.jp/KNApSAcK/Manual/KNApSAcKManual.html) can also be used to pick up metabolites not registered in the above databases. This database is specific for secondary metabolites and MS-based data searches can also be performed. Some databases focus purely on electron impact mass spectrometry data, such as the Wiley Registry of Mass Spectral Data (http://www.wileyregistry.com), the largest commercially available reference library of mass spectra. The GOLM open access database at the Max-Planck Institute of Molecular Plant

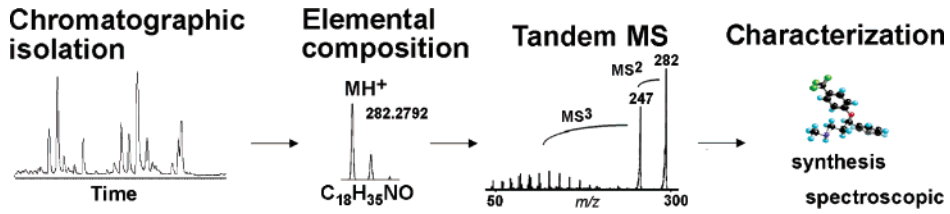


**Figure 5.** Identification of an unknown metabolite or confirmation of a known metabolite can be performed using mass spectrometry. The compound can be isolated using chromatography and characterization facilitated through accurate mass measurements to provide an elemental composition and tandem mass spectral data for structural information. Ultimately, synthetic standards are generated to validate (with perhaps other spectroscopic data), a structural hypothesis.

Physiology also focuses on electron ionization mass spectrometry data and is intended as a repository for experiments performed at this institute, as well as for data from collaborators.[91]

To support the identification of metabolites we have developed METLIN, a web-based data repository (http://metlin.scripps.edu/) on endogenous and exogenous metabolites. METLIN provides mass, elemental composition, CAS#, KEGG#, some MS/MS data, and a diverse collection of LC/MS and high-resolution Fourier transform mass spectrometry (FTMS) spectra, primarily from human biofluids and also some model organisms. The purpose of this data is to aid in metabolite identification through accurate mass measurement and isotopic pattern evaluation. METLIN also includes an annotated list of known metabolite structural information, both endogenous and drug metabolites can be easily cross-correlated with the LC/MS and FTMS data. Further, METLIN provides a number of data visualization tools including color 3D LC-MS plots and histograms.

A long-term aim in metabolomics is the establishment of data standards, to standardize experiment descriptions, particularly within publications. ArMet[92] (http://www.armet.org), is a data model to describe plant metabolomics experiments and their results.[93] Other groups have produced reporting requirements for metabolomics experiments,[15] to form a checklist of the information necessary for the publication of metabolomics data. A standard metabolic reporting structure policy document (SMRS Group, 2004) has been developed by a group from industry and academia.

**Metabolite Identification.** Once potential biomarkers have been selected, identification is required. Some metabolites observed in metabolomics studies may be well-known and characterized. Databases such as KEGG, human metabolite database, and METLIN can be used to search candidate molecules. If samples are analyzed using high-resolution mass spectrometry, then many candidates can be excluded. Once candidate molecules are obtained, co-chromatography and comparison of MS/MS data are necessary to confirm the identification of the molecule.

If the molecule is not known then the next task is identification, a significant challenge given the often limited sample amount and trace quantities of some metabolites. The overall procedure can be summarized in Figure 5 with the initial LC isolation of molecule of interest followed by tandem mass measurements on a Q-TOF for structural characterization and FTMS analysis for accurate mass measurements. Typical methods for obtaining elemental composition involve high-resolution ESI-FTMS and FTMS/MS technology for accurate mass determination, as well as the newer LTQ-FT technology. Orthogonal acceleration Q-TOF mass spectrometry[94] is also being used to obtain high accurate mass measurements. Furthermore, UPLC/MS$^E$, performed on a Q-TOF, has been presented recently as an approach for obtaining fragmentation data from LC/MS metabolomics studies.[95] This technique was applied to small molecules in complex mixtures and was achieved using simultaneous acquisition of exact mass at high and low collision energy, without reported loss of quality in the chromatographic data, offering an alternative approach to structural elucidation in complex mixture analysis problems.

However, despite the usefulness of this mass spectrometry data, the lack of comprehensive mass spectral libraries often precludes identification of molecules based on this data alone. Ultimately, the combination of many technologies will be required to identify unknown metabolites in biofluids including high sensitivity capillary NMR, which can provide metabolite structure characterization down to low microgram level,[96,97] chemical modification for functional group identification, and finally independent synthesis for verification.

An example of the isolation and characterization of completely novel metabolites was recently shown with the discovery of a family of taurine-conjugated fatty acids.[11] The challenge in identifying these metabolites was addressed in a three-step approach, (1) ultrahigh accuracy FTMS mass measurements, (2) high accuracy tandem mass analysis using a Q-TOF, and (3) chemical synthesis of potential candidates using the results and structural information gained from experiments (1) and (2).

## 3. Applications of Mass Spectrometry in Metabolomics

As metabolomics techniques become more robust and sophisticated, their applications become more widespread (Table 1). Combined with proteomics and genomics, metabolomics can help gain insight into systems biology, by studying the metabolite alterations and their relationships to changes in gene expression, protein expression and enzyme activity.[98−100] Despite the obvious challenges facing mass spectrometry in metabolomics, including the confirmation of known metabolites and the identification of unknown metabolites, many studies are underway employing these techniques. One particularly successful application of metabolomics has been in understanding gene function in model organisms such as yeast, plants and mice.[25,101−103] Notably, metabolomics has been applied to mouse models of Huntingtons Disease,[104] cardiac disease,[103] and Duchenne muscular dystrophy.[20]

However, despite a great need and potential, there are currently very few metabolomic studies in cancer therapeutics. Metabolomics can be applied to the study of cancer by monitoring tumor growth and regression, and has already been used to study the function of hypoxia-inducible factor 1$\beta$ in tumors.[105] By combining functional genomics with metabolomics, features of neuroendocrine cancers associated with a poor outcome have been identified.[99]

**Table 1.** Summary of Some of the Applications of Metabolomics, Together with Relevant References.

| application of metabolomics | description | reference |
|---|---|---|
| systems biology | study of metabolite alterations and their relationship to changes in gene expression, protein expression, and enzyme activity | 79−81 |
| mouse models | to study Huntington's disease | 86 |
| | to study cardiac disease | 85 |
| | to study Duchenne muscular dystrophy | 18 |
| cancer | study of the function of hypoxia-inducible factor 1ss in tumors | 87 |
| | in combination with functional genomics to identify features of neuroendocrine cancers associated with a poor outcome | 80 |
| pharmaceutical arena | investigate drug efficacy and toxicity, diagnose or predict disease states, and classify patient groups based on their specific metabolism | 88, 89, 15, 21 |
| plant biotechnology | phenotyping of plants and the assessment of the natural variance in metabolite profiles between plants to improve compositional quality | 84, 90−99 |

Another important application of metabolomics is in the pharmaceutical arena, where it can be used to investigate drug efficacy and toxicity, to diagnose or predict disease states, or to classify patient groups based on their specific metabolism. By measuring alterations in biofluid metabolite concentrations after administration of a therapeutic agent, and applying multivariate statistical analysis techniques to highlight any differences between dosed and control samples, the effect of a potential drug can be studied.[106] Alterations in specific metabolites, such as succinate, glycine, and dimethylamine in the blood indicate kidney damage.[17] In addition, the nephrotoxin gentamicin, when administered to male Wistar-derived rats, has been shown to increase *N*-acetyl-beta-D-glucosaminidase (NAG) activity significantly, accompanied by kidney damage. Using a combination of NMR and HPLC-TOF−MS/MS, raised glucose and reduced trimethylamine *N*-oxide (TMAO), as well as reduced xanthurenic acid and kynurenic acid were observed in the urine of treated animals.[23] Furthermore, bromobenzene treatment to rats induces the formation of the novel biomarker, 5-oxoproline, in liver tissue, blood plasma, and urine.[107] These studies could be eventually expanded to humans, where metabolomics techniques may be able to highlight the responses of different groups of patients to a given drug. In this way, metabolomics may dramatically reduce the costs of drug development, by eliminating the progression of compounds destined to fail due to toxicity. Additionally, in the drug development phase, metabolomics could also aid in the discovery of new preclinical and clinical safety and efficacy biomarkers. The timing of the appearance of small molecule markers in the particular biofluids may also be of importance.

The value of metabolomics in plant biotechnology has increased significantly, and despite the convoluted nature of plant metabolism, the interpretation of metabolomics data is becoming easier, in part due to more sophisticated data analysis approaches. Metabolomics can be used for the phenotyping of plants, and has been used in part to assess the natural variance in metabolite profiles between plants, with the potential to improve compositional quality.[26] Already, metabolomic techniques have been applied to a vast array of plant species, such as potato,[27] tomato,[108,109] wheat,[110] rice,[111] *Arabidopsis*,[25] aspen,[112] cucumber,[113] strawberry,[114] and lettuce.[115] Over 1000 small molecules have been quantitated in a single leaf extract, as well as more than 500 compounds from potato tubers.[25,27]

**Summary and Outlook.** The area of metabolomics is expanding rapidly and applications for this science range from basic biochemistry to clinical biomarker discovery. The primary challenge in metabolomics is in the generation of comprehensive, quantitative profiles of the thousands of components present in biofluids, an issue that is largely being addressed with LC/MS technology. Data analysis is becoming more mature, due to the development of sophisticated bioinformatics software packages that will ultimately drive the discovery process. However, probably the greatest challenge in metabolomics is in structurally characterizing physiologically important molecules. The application of high-accuracy instruments and advancements in the generation of fragmentation data, along with the growing numbers of databases available, are gradually making this task possible. As these challenges are being met, it is encouraging that new potential biomarkers for diseases such as myocardial ischemia,[116] atherosclerosis,[117] muscular dystrophy,[118] influenza-associated encephalopathy,[119] and various cancers[66,99,120] are being identified. As metabolomics data is complementary to transcriptomics and proteomics, the data from all three approaches can be meshed to provide a more complete picture of cells and even whole organisms. Ultimately it is the discovery of novel metabolites[10,11] as well as correlating the changes of multiple metabolites with physiological events that make this area alluring and challenging.

**Acknowledgment.** This work was supported by NIH Grant MH062261 and DOE Grant DE-AC02-05CH11231. A.N. is supported by a postdoctoral fellowship from the Swedish Research Council (V.R.).

## References

(1) Dear, G. J.; Ayrton, J.; Plumb, R.; Fraser, I. J. The rapid identification of drug metabolites using capillary liquid chromatography coupled to an ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **1999**, *13* (5), 456−463.
(2) Zhang, N. Y.; Fountain, S. T.; Bi, H. G.; Rossi, D. T. Quantification and rapid metabolite identification in drug discovery using API time-of-flight LC/MS. *Anal. Chem.* **2000**, *72* (4), 800−806.
(3) Shockcor, J. P.; Holmes, E. Metabonomic applications in toxicity screening and disease diagnosis. *Curr. Top. Med. Chem.* **2002**, *2* (1), 35−51.
(4) Tiller, P. R.; Romanyshyn, L. A. Liquid chromatography/tandem mass spectrometric quantification with metabolite screening as a strategy to enhance the early drug discovery process. *Rapid Commun. Mass Spectrom.* **2002**, *16* (12), 1225−1231.
(5) Plumb, R. S.; Stumpf, C. L.; Granger, J. H.; Castro-Perez, J.; Haselden, J. N.; Dear, G. J. Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun. Mass Spectrom.* **2003**, *17* (23), 2632−2638.

(6) Deng, Y. Z.; Wu, J. T.; Zhang, H. W.; Olah, T. V. Quantitation of drug metabolites in the absence of pure metabolite standards by high-performance liquid chromatography coupled with a chemiluminescence nitrogen detector and mass spectrometer. *Rapid Commun. Mass Spectrom.* **2004**, *18* (15), 1681−1685.
(7) Maurer, H. H. Advances in analytical toxicology: the current role of liquid chromatography-mass spectrometry in drug quantification in blood and oral fluid. *Anal. Bioanal. Chem.* **2005**, *381* (1), 110−118.
(8) Staack, R. F.; Varesio, E.; Hopfgartner, G. The combination of liquid chromatography/tandem mass spectrometry and chip-based infusion for improved screening and characterization of drug metabolites. *Rapid Commun. Mass Spectrom.* **2005**, *19* (5), 618−626.
(9) Liu, D. Q.; Hop, C. E. C. A. Strategies for characterization of drug metabolites using liquid chromatography-tandem mass spectrometry in conjunction with chemical derivatization and on-line H/D exchange approaches. *J. Pharm. Biomed. Anal.* **2005**, *37* (1), 1−18.
(10) Cravatt, B. F.; Prosperogarcia, O.; Siuzdak, G.; Gilula, N. B.; Henriksen, S. J.; Boger, D. L.; Lerner, R. A. Chemical Characterization of a Family of Brain Lipids That Induce Sleep. *Science* **1995**, *268* (5216), 1506−1509.
(11) Saghatelian, A.; Trauger, S. A.; Want, E. J.; Hawkins, E. G.; Siuzdak, G.; Cravatt, B. F. Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* **2004**, *43* (45), 14332−14339.
(12) Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68* (10), 2374−2376.
(13) Dirren, H.; Robinson, A. B.; Pauling, L. Sex-related patterns in the profiles of human urinary amino acids. *Clin. Chem.* **1975**, *21* (13), 1970−5.
(14) Chace, D. H. Mass spectrometry in the clinical laboratory. *Chem. Rev.* **2001**, *101* (2), 445−477.
(15) Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **2004**, *9* (9), 418−425.
(16) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Antti, H.; Bollard, M. E.; Keun, H.; Beckonert, O.; Ebbels, T. M.; Reilly, M. D.; Robertson, D.; Stevens, G. J.; Luke, P.; Breau, A. P.; Cantor, G. H.; Bible, R. H.; Niederhauser, U.; Senn, H.; Schlotterbeck, G.; Sidelmann, U. G.; Laursen, S. M.; Tymiak, A.; Car, B. D.; Lehman-McKeeman, L.; Colet, J. M.; Loukaci, A.; Thomas, C. Contemporary issues in toxicology - The role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* **2003**, *187* (3), 137−146.
(17) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discovery* **2002**, *1* (2), 153−161.
(18) Bligny, R.; Douce, R. NMR and plant metabolism. *Curr. Opin. Plant Biol.* **2001**, *4* (3), 191−196.
(19) Ratcliffe, R. G.; Shachar-Hill, Y. Probing plant metabolism with NMR. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2001**, *52*, 499−526.
(20) Griffin, J. L.; Williams, H. J.; Sang, E.; Clarke, K.; Rae, C.; Nicholson, J. K. Metabolic profiling of genetic disorders: A multitissue H-1 nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue. *Anal. Biochem.* **2001**, *293* (1), 16−21.
(21) Holmes, E.; Tsang, T. M.; Tabrizi, S. J. The application of NMR-based metabonomics in neurological disorders. *NeuroRx* **2006**, *3* (3), 358−372.
(22) Bollard, M. E.; Keun, H. C.; Beckonert, O.; Ebbels, T. M. D.; Antti, H.; Nicholls, A. W.; Shockcor, J. P.; Cantor, G. H.; Stevens, G.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. *Toxicol. Appl. Pharmacol.* **2005**, *204* (2), 135−151.
(23) Lenz, E. M.; Bright, J.; Knight, R.; Westwood, F. R.; Davies, D.; Major, H.; Wilson, I. D. Metabonomics with H-1-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. *Biomarkers* **2005**, *10* (2−3), 173−187.
(24) Craig, A.; Sidaway, J.; Holmes, E.; Orton, T.; Jackson, D.; Rowlinson, R.; Nickson, J.; Tonge, R.; Wilson, I.; Nicholson, J. Systems toxicology: Integrated genomic, proteomic and metabonomic analysis of methapyrilene induced hepatotoxicity in the rat. *J. Proteome Res.* **2006**, *5* (7), 1586−1601.
(25) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18* (11), 1157−1161.
(26) Schauer, N.; Fernie, A. R. Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci.* **2006**, in press.
(27) Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* **2000**, *23* (1), 131−142.
(28) Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.* **2004**, *76* (6), 1738−1745.
(29) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. Matrix effect in quantitative LC/MS/MS analyses of biological fluids: a method for determination of finasteride in human plasma at picogram per milliliter concentrations. *Anal. Chem.* **1998**, *70* (5), 882−889.
(30) Gangl, E. T.; Annan, M. M.; Spooner, N.; Vouros, P. Reduction of signal suppression effects in ESI-MS using a nanosplitting device. *Anal. Chem.* **2001**, *73* (23), 5635−5644.
(31) Gustavsson, S. A.; Samskog, J.; Markides, K. E.; Langstrom, B. Studies of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents. *J. Chromatogr. A* **2001**, *937* (1−2), 41−47.
(32) Plumb, R.; Castro-Perez, J.; Granger, J.; Beattie, I.; Joncour, K.; Wright, A. Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004**, *18* (19), 2331−2337.
(33) Bremer, H. J. *Disturbances of Amino Acid Metabolism*; Urban & Schwarzenberg: Baltimore, MD, 1981.
(34) Hommes, F. A. *Techniques in Diagnostic Human Biochemical Genetics: A Laboratory Manual;* Wiley-Liss: New York, 1991.
(35) Niwa, T. Procedures for MS analysis of clinically relevant compounds. *Clin. Chim. Acta* **1995**, *241−242*, 75−152.
(36) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J. Chromatogr., B: Analyt. Technol. Biomed. Life Sci.* **2005**, *828* (1−2), 9−13.
(37) Delinsky, A. D.; Delinsky, D. C.; Muralidhara, S.; Fisher, J. W.; Bruckner, J. V.; Bartlett, M. G. Analysis of dichloroacetic acid in rat blood and tissues by hydrophilic interaction liquid chromatography with tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19* (8), 1075−1083.
(38) Tolstikov, V. V.; Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **2002**, *301* (2), 298−307.
(39) Pisano, R.; Breda, M.; Grassi, S.; James, C. A. Hydrophilic interaction liquid chromatography-APCI-mass spectrometry determination of 5-fluorouracil in plasma and tissues. *J. Pharm. Biomed. Anal.* **2005**, *38* (4), 738−745.
(40) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* **2005**, *4* (2), 591−598.
(41) Swartz, M. E.; Murphy, B. J. Ultr performance liquid chromatography: tomorrow's HPLC technology today. *Labplus Int.* **2004**, *18* (3), 6−9.
(42) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M. HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr., B: Analyt. Technol. Biomed. Life Sci.* **2005**, *817* (1), 67−76.
(43) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal. Chem.* **2006**, *78* (10), 3289−3295.
(44) Yang, Z. Online hyphenated liquid chromatography-nuclear magnetic resonance spectroscopy-mass spectrometry for drug metabolite and nature product analysis. *J. Pharm. Biomed. Anal.* **2006**, *40* (3), 516−527.

(45) Dachtler, M.; Glaser, T.; Kohler, K.; Albert, K. Combined HPLC-MS and HPLC-NMR on-line coupling for the separation and determination of lutein and zeaxanthin stereoisomers in spinach and in retina. *Anal. Chem.* **2001**, *73* (3), 667−674.
(46) Srebalus, B.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. Resolving isomeric peptide mixtures: a combined HPLC/ion mobility-TOFMS analysis of a 4000-component combinatorial library. *Anal. Chem.* **2002**, *74* (1), 26−36.
(47) Boernsen, K. O.; Gatzek, S.; Imbert, G. Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal. Chem.* **2005**, *77* (22), 7255−7264.
(48) Wang, G.; Hsieh, Y.; Cui, X.; Cheng, K. C.; Korfmacher, W. A. Ultra-performance liquid chromatography/tandem mass spectrometric determination of testosterone and its metabolites in in vitro samples. *Rapid Commun. Mass Spectrom.* **2006**, *20* (14), 2215−2221.
(49) Chace, D. H.; Millington, D. S.; Terada, N.; Kahler, S. G.; Roe, C. R.; Hofman, L. F. Rapid diagnosis of phenylketonuria by quantitative analysis for phenylalanine and tyrosine in neonatal blood spots by tandem mass spectrometry. *Clin. Chem.* **1993**, *39* (1), 66−71.
(50) Levy, H. L. Newborn screening by tandem mass spectrometry: a new era. *Clin. Chem.* **1998**, *44* (12), 2401−2402.
(51) Chelius, D.; Zhang, T.; Wang, G.; Shen, R. F. Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. *Anal. Chem.* **2003**, *75* (23), 6658−6665.
(52) Nagele, E.; Vollmer, M.; Horth, P. Two-dimensional nano-liquid chromatography-mass spectrometry system for applications in proteomics. *J. Chromatogr. A* **2003**, *1009* (1−2), 197−205.
(53) Chatman, K.; Hollenbeck, T.; Hagey, L.; Vallee, M.; Purdy, R.; Weiss, F.; Siuzdak, G. Nanoelectrospray mass spectrometry and precursor ion monitoring for quantitative steroid analysis and attomole sensitivity. *Anal. Chem.* **1999**, *71* (13), 2358−2363.
(54) Griffiths, W. J.; Liu, S.; Yang, Y.; Purdy, R. H.; Sjovall, J. Nanoelectrospray tandem mass spectrometry for the analysis of neurosteroid sulphates. *Rapid Commun. Mass Spectrom.* **1999**, *13* (15), 1595−1610.
(55) Abian, J.; Oosterkamp, A. J.; Gelpí, E. Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: practical considerations. *J. Mass Spectrom.* **1999**, *34* (4), 244−254.
(56) Keski-Hynnila, H.; Kurkela, M.; Elovaara, E.; Antonio, L.; Magdalou, J.; Luukkanen, L.; Taskinen, J.; Kostiainen, R. Comparison of electrospray, atmospheric pressure chemical ionization, and atmospheric pressure photoionization in the identification of apomorphine, dobutamine, and entacapone phase II metabolites in biological samples. *Anal. Chem.* **2002**, *74* (14), 3449−3457.
(57) Raffaelli, A.; Saba, A. Atmospheric pressure photoionization mass spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (5), 318−331.
(58) Kratzsch, C.; Tenberken, O.; Peters, F. T.; Weber, A. A.; Kraemer, T.; Maurer, H. H. Screening, library-assisted identification and validated quantification of 23 benzodiazepines, flumazenil, zaleplone, zolpidem and zopiclone in plasma by liquid chromatography/mass spectrometry with atmospheric pressure chemical ionization. *J. Mass Spectrom.* **2004**, *39* (8), 856−872.
(59) Byrdwell, W. C. Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids. *Lipids* **2001**, *36* (4), 327−346.
(60) De Marchi, N.; De Petrocellis, L.; Orlando, P.; Daniele, F.; Fezza, F.; Di Marzo, V. Endocannabinoid signalling in the blood of patients with schizophrenia. *Lipids Health Dis.* **2003**, *2*, 5.
(61) Dally, J. E.; Gorniak, J.; Bowie, R.; Bentzley, C. M. Quantitation of underivatized free amino acids in mammalian cell culture media using matrix assisted laser desorption ionization time-of-flight mass spectrometry. *Anal. Chem.* **2003**, *75* (19), 5046−5053.
(62) Edwards, J. L.; Kennedy, R. T. Metabolomic analysis of eukaryotic tissue and prokaryotes using negative mode MALDI time-of-flight mass spectrometry. *Anal. Chem.* **2005**, *77* (7), 2201−2209.
(63) Vaidyanathan, S.; Gaskell, S.; Goodacre, R. Matrix-suppressed laser desorption/ionisation mass spectrometry and its suitability for metabolome analyses. *Rapid Commun. Mass Spectrom.* **2006**, *20* (8), 1192−1198.
(64) Shen, Z.; Thomas, J. J.; Averbuj, C.; Broo, K. M.; Engelhard, M.; Crowell, J. E.; Finn, M. G.; Siuzdak, G. Porous silicon as a versatile platform for laser desorption/ionization mass spectrometry. *Anal. Chem.* **2001**, *73* (3), 612−619.
(65) Vaidyanathan, S.; Jones, D.; Broadhurst, D. I.; Ellis, J.; Jenkins, T.; Dunn, W. B.; Hayes, A.; Burton, N.; Oliver, S. G.; Kell, D. B.; Goodacre, R. A laser desorption ionisation mass spectrometry approach for high throughput metabolomics. *Metabolomics* **2005**, *1* (3), 243−250.
(66) Chen, H.; Pan, Z.; Talaty, N.; Raftery, D.; Cooks, R. G. Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (10), 1577−1584.
(67) Corcoran, O.; Spraul, M. LC-NMR-MS in drug discovery. *Drug Discovery Today* **2003**, *8* (14), 624−631.
(68) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal. Chem.* **2006**, *78* (2), 363−371.
(69) Patrie, S. M.; Charlebois, J. P.; Whipple, D.; Kelleher, N. L.; Hendrickson, C. L.; Quinn, J. P.; Marshall, A. G.; Mukhopadhyay, B. Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (7), 1099−1108.
(70) Katajamaa, M.; Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* **2005**, *6*, 179.
(71) Katajamaa, M.; Miettinen, J.; Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **2006**, *22* (5), 634−636.
(72) Smith, C. A.; Want, E. J.; 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* (3), 779−787.
(73) Broeckling, C. D.; Reddy, I. R.; Duran, A. L.; Zhao, X.; Sumner, L. W. MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Anal. Chem.* **2006**, *78* (13), 4334−4341.
(74) Hastings, C. A.; Norton, S. M.; Roy, S. New algorithms for processing and peak detection in liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom.* **2002**, *16* (5), 462−467.
(75) Floter, A.; Nicolas, J.; Schaub, T.; Selbig, J. Threshold extraction in metabolite concentration data. *Bioinformatics* **2004**, *20* (10), 1491−1494.
(76) Norton, S. M.; Huyn, P.; Hastings, C. A.; Heller, J. C. Data mining of spectroscopic data for biomarker discovery. *Curr. Opin. Drug Discov. Devel.* **2001**, *4* (3), 325−331.
(77) Idborg, H.; Edlund, P. O.; Jacobsson, S. P. Multivariate approaches for efficient detection of potential metabolites from liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom.* **2004**, *18* (9), 944−954.
(78) Lutz, U.; Lutz, R. W.; Lutz, W. K. Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Anal. Chem.* **2006**, *78* (13), 4564−4571.
(79) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002.
(80) Taylor, J.; King, R. D.; Altmann, T.; Fiehn, O. Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics* **2002**, *18* Suppl 2, S241−248.
(81) Choi, H. K.; Choi, Y. H.; Verberne, M.; Lefeber, A. W.; Erkelens, C.; Verpoorte, R. Metabolic fingerprinting of wild type and transgenic tobacco plants by 1H NMR and multivariate analysis technique. *Phytochemistry* **2004**, *65* (7), 857−864.
(82) Yeung, K. Y.; Ruzzo, W. L. Principal component analysis for clustering gene expression data. *Bioinformatics* **2001**, *17* (9), 763−774.
(83) Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley & Sons Inc.: New York, 1989.
(84) Malinowski, E. R. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1991.
(85) Dwek, M. V.; Alaiya, A. A. Proteome analysis enables separate clustering of normal breast, benign breast and breast cancer tissues. *Br. J. Cancer* **2003**, *89* (2), 305−307.
(86) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16* (3), 119−128.
(87) Trygg, J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. *J. Chemom.* **2002**, *16* (6), 283−293.

(88) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem.* **2005**, *77* (5), 1282−1289.
(89) Gaweda, A. E.; Jacobs, A. A.; Brier, M. E.; Zurada, J. M. Pharmacodynamic population analysis in chronic renal failure using artificial neural networks−a comparative study. *Neural Netw.* **2003**, *16* (5−6), 841−845.
(90) Goodacre, R. Making sense of the metabolome using evolutionary computation: seeing the wood with the trees. *J. Exp. Bot.* **2005**, *56* (410), 245−254.
(91) Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **2005**, *21* (8), 1635−1638.
(92) Jenkins, H.; Hardy, N.; Beckmann, M.; Draper, J.; Smith, A. R.; Taylor, J.; Fiehn, O.; Goodacre, R.; Bino, R. J.; Hall, R.; Kopka, J.; Lane, G. A.; Lange, B. M.; Liu, J. R.; Mendes, P.; Nikolau, B. J.; Oliver, S. G.; Paton, N. W.; Rhee, S.; Roessner-Tunali, U.; Saito, K.; Smedsgaard, J.; Sumner, L. W.; Wang, T.; Walsh, S.; Wurtele, E. S.; Kell, D. B. A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* **2004**, *22* (12), 1601−1606.
(93) Jenkins, H.; Johnson, H.; Kular, B.; Wang, T.; Hardy, N. Toward supportive data collection tools for plant metabolomics. *Plant Physiol.* **2005**, *138* (1), 67−77.
(94) Wolff, J. C.; Eckers, C.; Sage, A. B.; Giles, K.; Bateman, R. Accurate mass liquid chromatography/mass spectrometry on quadrupole orthogonal acceleration time-of-flight mass analyzers using switching between separate sample and reference sprays. 2. Applications using the dual-electrospray ion source. *Anal. Chem.* **2001**, *73* (11), 2605−2612.
(95) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (13), 1989−1994.
(96) Schlotterbeck, G.; Ross, A.; Hochstrasser, R.; Senn, H.; Kuhn, T.; Marek, D.; Schett, O. High-resolution capillary tube NMR. A miniaturized 5-microL high-sensitivity TXI probe for mass-limited samples, off-line LC NMR, and HT NMR. *Anal. Chem.* **2002**, *74* (17), 4464−4471.
(97) Olson, D. L.; Norcross, J. A.; O'Neil-Johnson, M.; Molitor, P. F.; Detlefsen, D. J.; Wilson, A. G.; Peck, T. L. Microflow NMR: concepts and capabilities. *Anal. Chem.* **2004**, *76* (10), 2966−2974.
(98) Hirai, M. Y.; Yano, M.; Goodenowe, D. B.; Kanaya, S.; Kimura, T.; Awazuhara, M.; Arita, M.; Fujiwara, T.; Saito, K. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (27), 10205−10210.
(99) Ippolito, J. E.; Xu, J.; Jain, S.; Moulder, K.; Mennerick, S.; Crowley, J. R.; Townsend, R. R.; Gordon, J. I. An integrated functional genomics and metabolomics approach for defining poor prognosis in human neuroendocrine cancers. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (28), 9901−9906.
(100) Clish, C. B.; Davidov, E.; Oresic, M.; Plasterer, T. N.; Lavine, G.; Londo, T.; Meys, M.; Snell, P.; Stochaj, W.; Adourian, A.; Zhang, X.; Morel, N.; Neumann, E.; Verheij, E.; Vogels, J. T.; Havekes, L. M.; Afeyan, N.; Regnier, F.; van der Greef, J.; Naylor, S. Integrative biological analysis of the APOE*3-leiden transgenic mouse. *Omics* **2004**, *8* (1), 3−13.
(101) Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **2001**, *19* (1), 45−50.
(102) Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* **2003**, *21* (6), 692−696.
(103) Jones, G. L.; Sang, E.; Goddard, C.; Mortishire-Smith, R. J.; Sweatman, B. C.; Haselden, J. N.; Davies, K.; Grace, A. A.; Clarke, K.; Griffin, J. L. A functional analysis of mouse models of cardiac disease through metabolic profiling. *J. Biol. Chem.* **2005**, *280* (9), 7530−9.
(104) Griffin, J. L.; Cemal, C. K.; Pook, M. A. Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3. *Physiol. Genomics* **2004**, *16* (3), 334−340.
(105) Griffin, J. L.; Shockcor, J. P. Metabolic profiles of cancer cells. *Nat. Rev. Cancer* **2004**, *4* (7), 551−561.
(106) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. *Rapid Commun. Mass Spectrom.* **2002**, *16* (20), 1991−1996.
(107) Waters, N. J.; Waterfield, C. J.; Farrant, R. D.; Holmes, E.; Nicholson, J. K. Integrated metabonomic analysis of bromobenzene-induced hepatotoxicity: novel induction of 5-oxoprolinosis. *J. Proteome Res.* **2006**, *5* (6), 1448−1459.
(108) Schauer, N.; Zamir, D.; Fernie, A. R. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicum complex. *J. Exp. Bot.* **2005**, *56* (410), 297−307.
(109) Moco, S.; Bino, R. J.; Vorst, O.; Verhoeven, H. A.; de Groot, J.; van Beek, T. A.; Vervoort, J.; de Vos, C. H. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol.* **2006**, *141* (4), 1205−1218.
(110) Hamzehzarghani, H.; Kushalappa, A. C.; Dion, Y.; Rioux, S.; Comeau, A.; Yaylayan, V.; Marshall, W. D.; Mather, D. E. Metabolic profiling and factor analysis to discriminate quantitative resistance in wheat cultivars against fusarium head blight. *Physiol. Mol. Plant Pathol.* **2005**, *66* (4), 119−133.
(111) Sato, S.; Soga, T.; Nishioka, T.; Tomita, M. Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant J.* **2004**, *40* (1), 151−163.
(112) Jonsson, P.; Johansson, E. S.; Wuolikainen, A.; Lindberg, J.; Schuppe-Koistinen, I.; Kusano, M.; Sjostrom, M.; Trygg, J.; Moritz, T.; Antti, H. Predictive metabolite profiling applying hierarchical multivariate curve resolution to GC-MS data−a potential tool for multi-parametric diagnosis. *J. Proteome Res.* **2006**, *5* (6), 1407−1414.
(113) Tagashira, N.; Plader, W.; Filipecki, M.; Yin, Z.; Wisniewska, A.; Gaj, P.; Szwacka, M.; Fiehn, O.; Hoshi, Y.; Kondo, K.; Malepszy, S. The metabolic profiles of transgenic cucumber lines vary with different chromosomal locations of the transgene. *Cell. Mol. Biol. Lett.* **2005**, *10* (4), 697−710.
(114) Aharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *Omics* **2002**, *6* (3), 217−234.
(115) Garratt, L. C.; Linforth, R.; Taylor, A. J.; Lowe, K. C.; Power, J. B.; Davey, M. R. Metabolite fingerprinting in transgenic lettuce. *Plant Biotechnol. J.* **2005**, *3* (2), 165−174.
(116) Sabatine, M. S.; Liu, E.; Morrow, D. A.; Heller, E.; McCarroll, R.; Wiegand, R.; Berriz, G. F.; Roth, F. P.; Gerszten, R. E. Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* **2005**, *112* (25), 3868−3875.
(117) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethell, H. W.; Clarke, S.; Schofield, P. M.; McKilligin, E.; Mosedale, D. E.; Grainger, D. J. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nat. Med.* **2002**, *8* (12), 1439−1444.
(118) Griffin, J. L.; Sang, E.; Evens, T.; Davies, K.; Clarke, K. Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy. *FEBS Lett* **2002**, *530* (1−3), 109−116.
(119) Kawashima, H.; Oguchi, M.; Ioi, H.; Amaha, M.; Yamanaka, G.; Kashiwagi, Y.; Takekuma, K.; Yamazaki, Y.; Hoshika, A.; Watanabe, Y. Primary biomarkers in cerebral spinal fluid obtained from patients with influenza-associated encephalopathy analyzed by metabolomics. *Int J. Neurosci.* **2006**, *116* (8), 927−936.
(120) Yang, J.; Xu, G.; Zheng, Y.; Kong, H.; Pang, T.; Lv, S.; Yang, Q. Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *J. Chromatogr., B: Analyt. Technol. Biomed. Life Sci.* **2004**, *813* (1−2), 59−65.

PR060505+