This review presents an overview of the dynamically developing field of mass spectrometry-based metabolomics. Metabolomics aims at the comprehensive and quantitative analysis of wide arrays of metabolites in biological samples. These numerous analytes have very diverse physico-chemical properties and occur at different abundance levels. Consequently, comprehensive metabolomics investigations are primarily a challenge for analytical chemistry and specifically mass spectrometry has vast potential as a tool for this type of investigation. Metabolomics require special approaches for sample preparation, separation, and mass spectrometric analysis. Current examples of those approaches are described in this review. It primarily focuses on metabolic fingerprinting, a technique that analyzes all detectable analytes in a given sample with subsequent classification of samples and identification of differentially expressed metabolites, which define the sample classes. To perform this complex task, data analysis tools, metabolite libraries, and databases are required. Therefore, recent advances in metabolomics bioinformatics are also discussed. © 2006 Wiley Periodicals, Inc., Mass Spec Rev 26:51–78, 2007

Keywords: metabolomics; metabolic fingerprinting; metabolic profiling; lipidomics; mass spectrometry

I. INTRODUCTION

Dramatic technological advances in the biological sciences over the past few years have forged a new era of research including the emerging field of systems biology. Although the understanding of living organisms at the molecular system level is still in its infancy, it is evident that comprehensive investigations of the “omics cascade” with genomics, transcriptomics, proteomics, and metabolomics are important building blocks and will play a central role in this new science (see Fig. 1). The integrative analysis of an organism’s response to a perturbation on the transcriptome, proteome, and metabolome levels will lead to a better understanding of the biochemical and biological mechanisms in complex systems. However, whereas genomics, transcriptomics, and proteomics have made significant strides in technology development, the tools for the comprehensive examination of the metabolome are still emerging (Bino et al., 2004). Although metabolomics is the endpoint of the “omics cascade” and is the closest to phenotype, there is no single-instrument platform that currently can analyze all metabolites. Possibly, because there is at least the perception that the other “omic” approaches can be handled by a single platform, metabolomics has lagged behind the other technologies. This is illustrated in Figure 2, showing the bibliographic search containing the words metabolomics, metabonomics, and proteomics in Chemical Abstracts Plus (SciFinder Scholar). While in 1999 three articles containing the keywords metabolomics or metabonomics were published, the number increased to 147 articles in 2003 and 203 in 2004. Moreover, the journal Metabolomics (Springer) was recently launched, which is dedicated to publish research results related to metabolomics technology development, data analysis and storage, integrated studies with other “omics” techniques, and metabolomics applications. The rising number of publications in the field demonstrates that metabolomics is not just a new “omics” word but a valuable emerging tool to study phenotype and changes in phenotype caused by environmental influences, disease, or changes in genotype. The comprehensive investigation of the metabolome is being complicated by its enormous complexity and dynamics. Metabolite distributions are subjected to high temporal and spatial variability; for example, circadian fluctuations in mammalian organisms are well known. In addition, diet-dependent biological variability in mammalian systems can complicate the analysis (Vigneau-Callahan et al., 2001). A careful experimental design is therefore mandatory for the success of these types of investigations. The metabolome represents a vast number of components that belong to a wide variety of compound classes, such as amino acids, lipids, organic acids, nucleotides, etc. These compounds are very diverse in their physical and chemical properties and occur in a wide concentration range. For example, within lipids alone, not only high-abundance compounds, such as fatty acids, triglycerides, or phospholipids, are encountered, but also trace level components with important regulatory effects, such as eicosanoids derived from arachidonic acid. According to Beecher, 2,000 major metabolites seems to be a good estimate for humans (Beecher, 2006).
2003). This number can of course be vastly larger as one considers secondary metabolites. Some of these metabolites may be chemical mediators of great biological importance. Up to 200,000 metabolites can be encountered in the plant kingdom (Weckwerth, 2003). Consequently, studying the metabolome is a major challenge to analytical chemistry and a metabolomic analysis in its true sense, namely the quantitative analysis of all metabolites, cannot be achieved with the current analytical instrumentation.

Currently, two complementary approaches are used for metabolomic investigations: metabolic profiling and metabolic fingerprinting (see Fig. 3) (Dettmer & Hammock, 2004). A summary of metabolomics-related definitions is given in Table 1. Metabolic profiling focuses on the analysis of a group of metabolites either related to a specific metabolic pathway or a class of compounds. The quantitative analysis of fatty acids as fatty acid methyl esters by GC-FID (flame ionization detection) or the analysis of amino acids are examples for metabolic profiling. An even more directed approach is target analysis that aims at the measurement of selected analytes, such as biomarkers of disease or toxicant exposure, or substrates and products of enzymatic reactions (Fiehn, 2002). In most cases metabolic profiling is a hypothesis-driven approach rather than a hypothesis-generating one. Based on the questions asked, metabolites are selected for analysis and specific analytical methods are developed for their determination. The tremendous technology advances over the past few years allow a constant expansion of the number of analytes that are quantified simultaneously in a single analysis. Technologically, the analysis of single biomarker is often as complex as profiling all related key metabolites in a given biochemical pathway. However, the latter results will give a more complete and detailed description of metabolic perturbations than a single biomarker can provide. The results of metabolic profiling are quantitative and ideally independent of the technology used for data acquisition. Consequently, the data can be used to build databases that can be integrated with pathway maps or other “omics” data, which will enhance biological understanding. Although quantitative metabolite data from different model organisms are abundant in the literature, their integration in global databases has yet to be accomplished.
The disadvantage of metabolic profiling is that the system is not a global or true “omics” approach. However, numerous quantitative metabolic profiling methods analyzing different metabolite classes have already been developed and are routinely used. If these methods measuring key metabolites from different biochemical pathways are assembled as building blocks to study the metabolome, a powerful metabolomics approach will evolve.

The second approach towards metabolomics is metabolic fingerprinting. Initially in this approach the intention is not to identify each observed metabolite, but to compare patterns or “fingerprints” of metabolites that change in response to disease, toxin exposure, environmental or genetic alterations. A typical, but simplified workflow for a metabolic fingerprinting analysis is shown in Figure 4. Metabolic fingerprinting has been performed in a wide variety of biological matrices, such as urine, plasma or serum, saliva, and tissues or cells. In addition to metabolic fingerprinting of intracellular metabolites in cell culture systems, the analysis of extracellular metabolites excreted into the culture medium or taken up from the medium by cells can provide valuable information on their phenotype and physiological state. Pattern analysis of metabolites in conditioned cell culture media is called metabolic footprinting (Allen et al., 2003, 2004). Since metabolic fingerprinting can be simultaneously applied to a wide range of metabolites, it is a true “omics” approach. The

**TABLE 1.** Metabolomics-related definitions

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Small molecules that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolome</td>
<td>The complete set of metabolites in an organism.</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>Identification and quantification of all metabolites in a biological system.</td>
</tr>
<tr>
<td>Metabolic profiling</td>
<td>Quantitative analysis of set of metabolites in a selected biochemical pathway or a specific class of compounds. This includes target analysis, the analysis of a very limited number of metabolites, e.g. single analytes as precursors or products of biochemical reactions.</td>
</tr>
<tr>
<td>Metabolic fingerprinting</td>
<td>Unbiased, global screening approach to classify samples based on metabolite patterns or “fingerprints” that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites.</td>
</tr>
<tr>
<td>Metabolic footprinting</td>
<td>Fingerprinting analysis of extra-cellular metabolites in cell culture medium as a reflection of metabolite excretion or uptake by cells.</td>
</tr>
</tbody>
</table>

*This is in contrast to xenobiotic or foreign compound metabolites which may overshadow natural metabolites in an analytical procedure. However, they can be very valuable in evaluating the physiological status of the organism.*
1. Sample collection
Treatment, disease, etc. Control
tissue, plasma, urine, saliva, cells, etc.

2. Sample preparation
• Internal standard spike
• Extraction (LLE, SPE, etc)

3. Sample analysis
Derivatization-GC-MS, LC-MS, CE-MS

4. Data export

5. Data analysis
Multivariate analysis, e.g. PCA

6. Marker Identification
Libraries, MS/MS, NMR

Score Plot (Groups) Loading Plot (Markers)

FIGURE 4. Simplified workflow for a metabolic fingerprinting analysis.
implementation of nuclear magnetic resonance (NMR)-based metabolic fingerprinting has marked the beginning of a metabolomics approach as a tool in biochemistry and has proven to be extremely powerful in screening samples for a variety of signature patterns or clusters. Metabolic fingerprinting can be used as a diagnostic tool, for example, by evaluating a patient’s metabolic fingerprint in comparison to healthy and diseased subjects. In addition, the success of treatment strategies can be monitored by observing if the metabolic phenotype shifts back to the healthy state, or in other words if a sample after treatment falls in the cluster of healthy subjects. However, using metabolomics exclusively for fingerprinting without identifying the metabolites that cause clustering of experimental groups will only deliver a classification tool but not directly contribute to biochemical knowledge and understanding of underlying mechanisms of action. The real power of metabolomics is realized when qualitative and quantitative analyses are performed. The knowledge of metabolite identity and their quantitative perturbation as descriptors of differences in specific phenotypes will provide information that can be interpreted in the light of biochemical pathways. Therefore, metabolites causing group segregation in the fingerprinting approach need to be identified and quantitative methods for the analysis of these metabolites and related compounds should be developed, which will tie metabolic fingerprinting and profiling together. Consequently, the annotation of the metabolome is an important building block for successful metabolomics investigations.

Both metabolic fingerprinting and profiling can be used in the search for new biomarkers. The value of blood glucose and cholesterol tests in medical diagnostics illustrates the value of even simple biomarkers. Metabolomics can yield new biomarkers that can reach the clinic as tools to diagnose health status, disease, or outcome of pharmacological treatment. Metabolomics is not limited to individual biomarkers. It rather represents a new approach to diagnostics where large data sets can be employed in total to develop understanding. For example, evaluating related biochemical pathways in response to drug treatment will give a more complete description of feedback mechanisms or crosstalk than single biomarkers can deliver. Moreover, the concept of individualized health including nutrition but also tailored pharmacological treatment based on metabolic phenotype will rely strongly on metabolomics technology (Watkins & German, 2002). The promise of the technology in clinical medicine to move from milliliter to microliter samples and to move from a few to thousands of analytes is exciting, but the potential of the technology to generate biological understanding is possibly still more significant.

Numerous analytical platforms have been used for metabolomic applications, such as NMR (Nicholson & Wilson, 2003), Fourier transform-infrared spectroscopy (FT-IR) (Harrigan et al., 2004; Johnson et al., 2004) and mass spectrometry (MS) coupled to separation techniques, or using direct flow injection. The great advantages of NMR are the potential for high-throughput fingerprinting, minimal requirements for sample preparation, and the non-discriminating and non-destructive nature of the technique. However, only medium to high abundance metabolites will be detected with this approach and the identification of individual metabolites based on chemical shift signals that cause sample clustering in multivariate analysis is challenging in complex mixtures. Mass spectrometry-based metabolomics offers quantitative analyses with high selectivity and sensitivity and the potential to identify metabolites. Combination with a separation technique reduces the complexity of the mass spectra due to metabolite separation in a time dimension, provides isobar separation, and delivers additional information on the physicochemical properties of the metabolites. However, mass spectrometry-based techniques usually require a sample preparation step, which can cause metabolite losses, and based on the sample introduction system and the ionization technique used, specific metabolite classes may be discriminated. Therefore, parallel application of several techniques, for example, GC-MS and LC-MS is desired to study the metabolome comprehensively. Currently, mass spectrometry-based metabolomics is a dynamically emerging field with a number of annual publications exceeding published NMR-based investigations (see Fig. 2).

However, not only the choice of analytical techniques requires careful consideration, but the whole metabolomics experiment should be planned as an integrated unit, because the instrumental data are only as good as the experimental design and sample treatment. In this context Bino et al. (2004) proposed, the minimum information about a metabolomics (MIAMET) experiment, which should be reported with each study in order to facilitate the exchange of information and the establishment of databases. Similar recommendations regarding meta-data have been made for the other “omics” technologies.

In general, for every type of MS-based metabolomics experiment the following steps need to be addressed during method development and validation:

- Sampling,
- Sample preparation,
- Sample analysis including metabolite separation and MS detection,
- Data export,
- Data analysis.

II. SAMPLING AND QUENCHING

Sample acquisition is primarily driven by the experimental design and the experimental type. If possible, a power analysis should be performed to ensure that a sufficient number of samples are acquired and to reduce the influence of biological variability and obtain statistically validated data. In particular, when studying human samples, the influences of diet, gender, age, and genetic factors have to be considered. Therefore, in metabolic fingerprinting analysis a large number of samples are commonly analyzed to detect biologically relevant sample clustering. In addition, representative quality control samples, such as sample replicates, analytical replicates, and blanks including method blanks have to be analyzed.

While processing biological samples, special care must be taken to minimize the formation or degradation of metabolites after sampling due to remaining enzymatic activity or oxidation processes. Several techniques have been used to inhibit metabolism such as freezing in liquid nitrogen, freeze clamping, acid treatment (Theobald et al., 1993) or quenching in salt containing aqueous methanol at low temperatures (Maharjan &
Volatile metabolites can include alcohols, aldehydes, furans, ketones, pyroles, terpenes, and others (Mills & Walker, 2001). The analysis of volatile compounds is challenging by itself. Conventional sample preparation techniques, such as liquid extraction or SPE, are often not feasible, because the analytes might be incompletely extracted and losses will occur if the extract is concentrated. Moreover, the solvent can interfere with the analytes during gas chromatographic separation due to incomplete separation of the solvent peak from the analyte peaks. Specifically, in GC-MS the analytes will be cut off with the solvent delay. Therefore, solvent-free sample preparation techniques are often the method of choice for the analysis of volatile metabolites. Wahl et al. used headspace sampling with a gastight syringe in combination with cold trapping in a temperature-controlled cold injection system and GC-MS for the analysis of volatile metabolites in urine samples. Thirty-four compounds were identified using MS libraries and reference compounds (Wahl et al., 1999). Another rapid and solvent-less sample preparation technique for the determination of volatile and semi-volatile compounds is headspace-solid phase microextraction (HS-SPME) (Pawliszyn, 1997). SPME uses a silica fiber that is coated with a stationary phase for sampling. The fiber is housed in a syringe-type assembly. A number of different stationary phases comprising non-polar to polar materials are commercially available as fiber coatings. In the case of HS-SPME the sampling takes place in the headspace compartment above the sample, enabling an easy fractionation of volatile analytes and complex matrix. However, sample extraction is not exhaustive, but the amount of analytes enriched on the fiber depends on two equilibrium distribution steps, sample/gaseous phase and gaseous phase/fiber coating. Both distribution coefficients are temperature dependent, and extraction time and temperature need to be optimized to maximize sample enrichment. Moreover, the distribution of the analytes between sample and gaseous phase is influenced by analyte-matrix interactions and has to be considered for quantitative analysis. Desorption of the analytes from the fiber can be achieved by thermal energy in the hot injector of a GC or using solvent for coupling with HPLC. Mills and Walker (2000) reviewed the application of SPME for the GC analysis of volatile compounds in different matrices, such as urine, blood, faces, breast milk, hair, breath, and saliva. The authors also tested different fiber coatings for the profiling of volatile urinary metabolites by HS-SPME-GC-MS and found carboxen-polydimethylsiloxane fibers to be the best choice (Mills & Walker, 2001). In addition, SPME-GC-MS has been used to study VOCs in plants (Verdonk et al., 2003; Bino et al., 2005). Stashenko, Jaramillo, and Martinez (2004a, 2004b) used different extraction methods, such as hydrodistillation, simultaneous distillation–solvent extraction, microwave-assisted hydrodistillation, supercritical fluid extraction (SFE) with CO2, static headspace, simultaneous purge and trap in CH2Cl2, and HS-SPME for sampling volatile plant metabolites.

B. Solid-Phase Extraction

Solid-phase extraction (SPE) has become one of the most important sample preparation techniques to extract analytes from biological fluids. It is widely used in metabolic target and profiling analysis when a sufficient separation of analytes from
interfering matrix is needed. SPE describes the non-equilibrium, exhaustive removal of analytes from a flowing liquid sample by retention on a solid sorbent. The retained analytes are subsequently eluted from the sorbent using a solvent or solvent mixtures with sufficient elution strength (Mitra, 2003). Based on the analytes and the sorbent selected, compounds are retained by van der Waals interactions, dipole–dipole interactions, hydrogen bonding, or electrostatic forces. A variety of SPE sorbent materials are commercially available for the broad extraction of metabolites, such as silica, alkylated silica (e.g., C-18), carbon-based sorbents, ion exchange materials, polymer materials, and restricted access materials (RAM) used to remove macromolecules. While C-18 modified silica phases have gained the most widespread use, they are characterized by a number of limitations, including poor retention of polar analytes, instability at high pH, and irreproducible analyte losses when the sorbent runs dry after the equilibration step. In addition to silica-based sorbents, polymer materials have gained more analytic use over the past years as SPE sorbents. Specifically, functionalized polymeric resins, containing polar functional groups embedded in a non-polar polymeric backbone, have improved retention of polar analytes while retaining hydrophobic analytes. Polymeric SPE sorbents also show a better pH-stability and are less prone to analyte losses when running dry. Idborg-Bjorkman et al. (2003) used Oasis HLB (Waters, Inc., Milford, MA), a functionalized copolymer of divinylbenzene and N-vinylpyrrolidone, for the extraction of urine for LC-ESI-MS based metabolite screening (Idborg, Edlund, & Jacobsson, 2004). Wang et al. (2003) used a C-18 material to remove salts from serum after separating the serum metabolome from the proteome using a 5-kDa molecular cut-off filter.

A newer development within SPE-sorbent materials is mixed-mode sorbents, which have a high potential for metabolomics investigations. Conventional SPE sorbents rely on a primary mechanism for analyte retention. In contrast, mixed mode materials use multiple retention mechanisms caused by the incorporation of different ligands in one sorbent. Examples are polymer materials with weak/strong cation-exchange or anion-exchange sites in the same material (Mitra, 2003). For example, using a mixed mode cation exchange material, neutral metabolites, protonated acids, and bases can be extracted from pH-adjusted biofluids using the same solid phase. Neutrals and protonated acids are retained by non-specific hydrophobic interactions with the polymer backbone at low pH, while bases are retained by cation exchange. Consequently, elution with organic solvent will recover a fraction containing neutrals and protonated acids, and a second pH adjusted elution step will recover the basic compounds. Since a wide variety of metabolites can be extracted with these materials, they hold much promise for metabolic fingerprinting analyses. In addition to offline SPE extraction using individual cartridges, disks, or 96-well plates, online sample extraction using a pre-column coupled to LC-MS has been successfully used in our laboratory for metabolomic investigations.

C. Liquid-Liquid Extraction

Liquid-liquid extraction (LLE) for biofluids has a long history as a sample preparation technique. It is often the method of choice for tissue extraction. Upon collection, tissues are typically frozen in liquid nitrogen. In order to obtain an efficient tissue extraction, the tissue has to be homogenized first by grinding the frozen sample to a fine powder using mortar and pestle, mortar grinders (Weckwerth, Wenzel, & Fiehn, 2004), vibration mills (Jonsson et al., 2004), vortexing or mixing with beads made of an inert material (Colebatch et al., 2004), or homogenization of the tissue in a solvent using blenders (Roessner et al., 2000). Metabolite extraction from the tissue is commonly done by shaking or vortexing with organic solvents. Polar metabolites are extracted by isopropanol, ethanol, methanol, acidic methanol, acetonitrile (Aharoni et al., 2002), water, methanol:water (Roessner et al., 2000) or other mixtures of those solvents, while more lipophilic metabolites can be extracted by chloroform or ethyl acetate. The advantage of extracting samples with mixtures of water:methanol:chloroform or ethyl acetate is the generation of a biphasic sample and the fractionation of the metabolites into two regions: aqueous and lipophilic organic fractions, which can be analyzed separately (Fiehn et al., 2000; Colebatch et al., 2004; Gullberg et al., 2004) used a “design of experiment approach” to optimize the extraction protocol for plant tissue for subsequent GC-MS analysis. By evaluating the response of 66 metabolites, the authors found extraction in a vibration mill with chloroform (200 µL) followed by addition of a methanol/water mixture (800 µL, 75:25) to form a monophasic as the optimum extraction procedure.

Weckwerth, Wenzel, and Fiehn (2004) developed a multiple solvent extraction protocol that sequentially extracts metabolites, proteins, and DNA from tissue samples of Arabidopsis thaliana. Different solvent systems were used for the extraction of cells, such as high-temperature extraction with ethanol or methanol (Nagels, Francois, & Renaud, 1997; Castrillo et al., 2003; Maharjan & Ferenci, 2003), 30% acetonitrile in water, 0.2% formic acid (50:50) (Vaidyanathan, Kell, & Goodacre, 2002), acid or alkaline treatment (Maharjan & Ferenci, 2003), cold methanol (Maharjan & Ferenci, 2003), chloroform-methanol extraction (Ruijter & Visser, 1996; Maharjan & Ferenci, 2003), or acidic methanol-dichloromethane-ethylacetate (Smetsgaard, 1997). Pilla-Boas et al. (2005) compared different extraction protocols, including chloroform:methanol:aqueous buffer, boiling ethanol, perchloric acid, potassium hydroxide, methanol:water, and pure methanol, for the extraction of polar intracellular yeast metabolites and found pure methanol the most and perchloric acid the least suitable for metabolite extraction.

D. Direct Injection

Direct injection of urine samples or diluted urine samples has been reported for LC-ESI-MS-based metabolic fingerprinting (Plumb et al., 2003b; Lenz et al., 2004a,b; Williams et al., 2005). The advantage of direct injection is that ideally no metabolites are lost during the sample preparation. However, high-salt content in urine can cause ion suppression and adduct formation in the electrospray process and can also cause a rapid deterioration of instrument performance due to contamination by non-volatile residues. The application of nanoelectrospray ionization (nanoESI) generally minimizes ion suppression and decreases source contamination (Scholz et al., 2004). If chromatographic
separation is used, the column flow can be diverted in the beginning of the run to reduce contamination of the ESI source (Lafaye et al., 2003). Direct injection of samples with high-protein content can cause poor chromatographic metabolite separation and a rapid increase of column backpressure, resulting in degradation of column performance (Pham-Tuan et al., 2003).

IV. METABOLIC PROFILING—AN ARRAY APPROACH TOWARDS METABOLOMICS

Metabolic profiling or the quantitative analysis of a selected number of metabolites involved in the same biochemical pathway is a widespread tool to study different aspects of metabolism. A thorough coverage of all published methods would greatly exceed the purpose of this review. However, selected contributions in the more recent literature describing the analysis of key metabolite classes will be highlighted below. While each method on its own does not represent a true metabolomic analysis, assembling a whole suite of quantitative methods that analyze key metabolites from different biochemical pathways will turn metabolic profiling into metabolomics. The advantage of this integrative array approach is that quantitative data are generated that will lead to an annotation of the metabolome and the subsequent generation of independent databases containing quantitative information that can be further mined. In addition, using individual assays to probe the metabolome allows the assessment of important low-abundance metabolites with regulatory functions, such as eicosanoids, which would not be detected by a general metabolic fingerprinting approach at the current level of technology development.

A. Mass Spectrometric Profiling of Metabolic Disorders

Mass spectrometric metabolic profiling plays an important role in the diagnosis of inborn errors of metabolism. The analysis of the urinary metabolome for diagnostics of metabolic disorders using GC-MS has recently been reviewed by Kuhara (2005). In addition, electrospray ionization with tandem mass spectrometry (MS/MS) has become an important tool in newborn screening for the assessment of inborn disorders of amino acid, fatty acid, and organic acid biosynthesis. Several reviews have been published recently, describing the methodology and biochemical interpretation of data in detail (Chace, Kalas, & Naylor, 2002, 2003; Schulze et al., 2003; Chace & Kalas, 2005). Standard protocols incorporate methanol extraction of disks of dried bloodspots on filter paper, addition of stable isotope-labeled internal standards, solvent evaporation, and derivatization with acidified butanol. Analysis of butyl esters is performed by flow injection analysis (FIA) with positive-mode ESI and MS/MS detection using a neutral loss scan of 102 amu (butyl formate) to selectively detect neutral amino acids, a neutral loss of 119 amu for ornithine and citrulline, and a loss of m/z 161 for arginine. Acylcarnitines are measured by a precursor ion scan of m/z 85 (Zytkovicz et al., 2001).

Piraud et al. (2003) developed a method based on LC-ESI-MS/MS for the analysis of 79 small molecules, mainly amino acids, related to known metabolic disorders. The goal of the study was to analyze underivatized amino acids and related analytes in a single run. Ionization behavior of the metabolites in positive and negative mode and the fragmentation in collision-induced dissociation (CID) were studied to select characteristic and signal-intense transitions for analyte detection. In order to avoid interferences between isobars and in-source fragmentation products, an LC separation was necessary.

B. Lipidomics

The analysis of lipids using mass spectrometry has become a valuable and mature tool for the assessment of changes in lipid metabolism and lipid-mediated signaling processes as a result of disease, toxicant exposure, genetic modifications, or drug therapy (Watkins et al., 2002). The lipidome, comprising lipid classes, subclasses, and lipid signaling molecules, is an important sub-compartment of the metabolome and has been well studied (German, Roberts, & Watkins, 2003). Lipidomics delivers detailed and quantitative information about the constitution of the cellular lipidome and provides insights in biochemical mechanisms of lipid metabolism, lipid–lipid and lipid–protein interactions.

Non-polar lipids have been routinely analyzed using gas chromatographic separation, electron impact ionization (EI), and MS. A derivatization step, such as methylation or silylation, is often needed to further expand the range of lipids accessible with this technique. However, GC-MS analysis is limited to thermally stable compounds with a sufficient vapor pressure for volatilization during injection. Using high-temperature gas chromatography, triglycerides have been analyzed, but the analysis of complex lipids, such as phospholipids, requires a pre-separation of lipid classes followed by hydrolysis, derivatization of free fatty acids and subsequent GC-EI-MS or GC-FID analysis. Although highly quantitative, this approach is very labor intensive and structural information regarding lipid composition is lost.

In addition to conventional hot split/splitless injection for sample introduction, analytical pyrolysis (Py) has been used to analyze the lipid composition in various sample matrices. Analytical pyrolysis is the thermal degradation of a non-volatile sample in an inert gas atmosphere before introduction into the GC. Pyrolysis, in combination with GC-MS, is a very powerful tool for the characterization of complex biological materials. Snyder et al. (1990) used Py-GC-Ion Trap-MS of whole, underivatized microorganism samples for the differentiation of bacterial species based on their lipid profiles. Basile et al. (1998) used in situ thermally assisted hydrolysis and methylation (THM) in a pyrolysis system for the differentiation of whole bacterial cells based on the profiles of fatty acid methyl esters formed during THM. Pyrolysis was coupled to GC-MS or used as a direct inlet system for a triple quadrupole mass spectrometer. In situ THM is performed using quaternary N-alkylammonium salts, such as tetramethylammonium hydroxide (TMAH), which forms tetramethylammonium carboxylates with organic acids and methylation takes place upon heating. Ester-bound fatty acids are saponified and trans-methylated. However, unsaturated fatty acids can undergo isomerization and recoveries are incomplete (Drechsel, Dettmer, & Engewald, 2003).
The introduction of solid-state ion sources, such as fast atom bombardment (FAB) and liquid-phase ion sources, such as ESI or atmospheric pressure chemical ionization (APCI) allowed the direct lipid analysis (Griffiths, 2003). Hermansson et al. (2005) developed a fully automated method for the analysis of the lipidome using LC-ESI-MS that delivers quantitative data for more than 100 polar lipids, originating primarily from phospholipids. The so-called shotgun lipidsome approach uses crude tissue extracts, commonly obtained by Folch or Bligh-Dyer extraction methods. It is based on direct sample injection of the extract, using both positive and negative mode ESI (Han & Gross, 2003, 2005). ESI in positive mode allows the detection of acylcarnitines, phosphatidylcholine, lysophosphatidylcholine, phosphatidyl ethanolamine, and sphingomyelins. Free fatty acids, phosphatidic acid, phosphatidyl serine, phosphatidyl inositol, phosphatidyl glycerol, and phosphatidylethanolamine are ionized by negative mode ESI. Neutral lipids, such as triacylglycerols are not readily ionized by ESI, however they can be detected as ammonium, lithium, or sodium adducts with ESI in positive mode (Duffin, Henion, & Shieh, 1991; Hsu & Turk, 1999; Han & Gross, 2005). MS/MS experiments are used to further identify lipid species or to screen for individual lipid classes. For example, phospholipids with specific head groups show characteristic fragmentation patterns in CID spectra that can be used for their identification (Griffiths et al., 2001; Pulfer & Murphy, 2003). A similar approach using characteristic CID fragmentation has been developed for sphingolipid analysis or sphingolipidomics (Merrill et al., 2005). Specifically, neutral-loss scans and precursor-ion scans can be used to identify lipids in samples and to determine the molecular masses of the compounds.

Using ESI in positive mode, precursor-ion scans for m/z 184 ([C₂H₁₄NPO₄]⁺) or neutral loss scans for 59 amu (N(CH₃)₃) can be used to identify the [M + Na]⁺ and [M + H]⁺ ions that correspond to phosphatidylcholine and lysophosphatidylcholine (see Fig. 5). Precursor-ion scans for m/z 164, ([NaHO₂POC₂H₄NH₃]⁺) can be used to identify the [M + Na]⁺ and [M + H]⁺ ions of phosphatidylethanolamine. Using ESI in negative mode, neutral loss scans for 153 amu ([HO₂POCH₂C(OH)CH₂]) can be used to identify precursors of phosphatidylglycerol (Pelizzi et al., 2002). In sphingolipidomics the same m/z 184 characteristic ion is used to identify sphingomyelin while m/z 264 and m/z 266 ions can be used for ceramide identification in precursor ion scans (Merrill et al., 2005).

Further MS/MS experiments can be performed for the compositional analysis of lipids. The regio-chemistry of phospholipids is determined based on the intensity ratio of the carboxylate anions originating from sn-1 and sn-2 position, because the intensity of the fragment ion derived from the sn-2 acyl substituent is greater than the product ion corresponding to the sn-1 acyl substituent (Fang & Barcelona, 1998). Branched saccharide moieties of higher order sphingolipids (such as gangliosides) requires MS/MS/MS experiments for structure elucidation (Merrill et al., 2005).

Because certain class of lipids may comprise 100 different species, their routine quantification represents a challenge. However, some lipids, such as phospholipids, differentiate only by their head groups and acyl groups (specifically their lengths and number of double bonds). Therefore, computational lipidomics approaches can be developed to automate spectra interpretation and quantification of multiple lipid species to use this similarity as an advantage. For example, automated detection of over 800 phospholipid species in direct injection-ESI-MS/MS analysis of cell extracts has been reported with semi-quantitative data output (lipid array) similar to DNA microarrays (Forrester et al., 2004).

Metabolic profiling of eicosanoids, a class of lipid signaling molecules, was recently termed eicosanomics (Balazy, 2004). Eicosanoids are metabolites of arachidonic acid produced by

---

**FIGURE 5.** Product ion spectra of dipentadecyl-phosphatidylcholine using ESI in positive [M + H]⁺ (upper spectrum) and negative mode [M + HCOO]⁻ (lower spectrum).
cyclooxygenases, lipoxigenases, cytochrome P450’s and non-enzymatic oxidation and are one of the most important pharmacological targets. Because many eicosanoids are structural isomers, LC or GC separation is required prior to MS detection. The profiling of eicosanoids is performed by LC-ESI-MS/MS in negative mode (Margalit, Duffin, & Isakson, 1996; Newman, Watanabe, & Hammock, 2002; Schmelzer et al., 2005) or LC-Electron Capture-APCI-MS/MS which requires derivatization of carboxylic groups into pentafluorobenzyl esters (Lee et al., 2003).

V. METABOLIC FINGERPRINTING

Metabolic fingerprinting describes the unbiased analysis of the metabolome by examination of metabolite patterns in different experimental groups and subsequent classification of the patterns. Samples can be classified if the metabolite fingerprints change between groups, resulting in sample clustering. In mass spectrometry-based investigation, metabolite fingerprints are described by m/z values and corresponding intensities of detected ions. If a separation step is performed, retention times are also used to index metabolites. Thus, m/z values, retention times, and intensities represent the metabolic fingerprint of the analyzed sample and are exported for sample classification using multivariate data analysis. The chemical structure of the detected metabolites typically remains unknown. In this review, the term metabolic fingerprinting refers not to investigations that aim at understanding the global plant response to nutritional stress. Instead, it is used as an integrated transcriptomics and metabolomics study to interpret the flight-tube geometry and instrument tuning, TOF-MS instruments provide mass resolution of 6,000–17,000 with mass accuracy in the range of 3–5 ppm. Allen et al. (2003, 2004) used TOF-MS with ESI in positive mode for metabolic footprinting analysis in yeast cell media samples.

Hybrid instruments, such as quadrupole-TOF-MS (Q-TOF-MS), have also been used for the rapid screening analysis of cell or plant extracts using the TOF as mass analyzer, which provides accurate mass measurement with high-mass resolution (Vaidyanathan, Kell, & Goodacre, 2002; Scholz et al., 2004). Structural elucidation of biomarkers can be performed by MS/MS experiments without switching to another instrument. Using the quadrupole as a scanning device, precursor ions can be selected for fragmentation in the collision cell, and product ion spectra with high-mass resolution for the fragment ions are produced. Accurate mass measurement of precursor and fragment ions allows clarification of the fragmentation process, thereby significantly facilitating the interpretation of spectra.

FT-ICR-MS offers very high resolution (100,000–1,000,000) and highest currently available mass accuracy (0.1–1 ppm), very low detection limits in the attomole to femtomole range and MS² capabilities (Marshall, Hendrickson, & Jackson, 1998), which makes it an ideal tool for metabolic fingerprinting investigations (Brown, Kruppa, & Dasseux, 2005). However, high-instrument costs have prevented their application from gaining wider distribution in metabolomic research. Aharoni et al. (2002) used FT-ICR-MS to study metabolic changes in ripening strawberry fruit by direct infusion of crude fruit extracts (methanol:water, 0.1% formic acid (50:50), or 100% acetonitrile) and ESI or APCI in positive and negative mode. In the study, a total of 5,844 unique ¹²C single charged masses were obtained by screening both extracts using four different ionization modes, and a single empirical formula could be assigned to over half of them with an error of ±1 ppm. Hirai et al. (2004) used FT-ICR-MS for metabolic fingerprinting of Arabidopsis thaliana extracts in an integrated transcriptomics and metabolomics study to understand the global plant response to nutritional stress.

Direct injection experiments for ESI are performed by continuous flow injection or by loop injection of 10–20 μL sample, using a solvent flow rate of 5–50 μL/min. The application of nanoESI has also been described. High-throughput fingerprinting can be achieved using automated nanoESI systems, such as the Nanomate system (Advion BioSciences, Inc., Ithaca, NY), which comprises a liquid sampler and an ESI chip consisting of an array of nanoelectrospray nozzles (Scholz et al., 2004).

Although direct injection-MS is a high-throughput metabolic fingerprinting technique, it has several disadvantages. For example, chemical isomers cannot be distinguished by this rapid screening technique because they have the same exact mass and therefore would require chromatographic separation (see Fig. 6). In addition, the classification of in-source fragmentation products or adduct formation is complicated. However, in the case of ESI the most important concern is ion suppression. Since all sample components are introduced simultaneously into the ionization source, signal suppression or enhancement can occur. This can impede a meaningful data analysis, especially if the matrix content varies between sample groups, for example, urine samples.
FIGURE 6. Combined MS spectra from 2.5 to 35 min obtained after online-SPE extraction of 10 µL human urine, reversed phase HPLC separation, and TOF-MS detection (LCT, Micromass, UK) with ESI in positive mode. Expanded section of the spectrum illustrates the need for high-mass resolution to distinguish isobaric peaks, which originate from different metabolites as shown by the extracted-ion chromatogram. The third extracted-ion chromatogram shows that m/z 202.129 originates from two different metabolites that can be chromatographically separated. Direct infusion would not readily distinguish these two metabolites.
An example for ion suppression is given in Figure 7, which shows the direct injection of a rat urine sample spiked with internal standards after online SPE compared to the analysis of internal standards spiked into a saline solution. Although salts are removed by the online-SPE extraction, severe signal suppression is observed for the internal standard signals in the urine sample relative to the saline solution.

**FIGURE 7.** Online SPE extraction of (A) 10 μL saline solution and (B) 10 μL rat urine spiked with internal standards coupled to direct infusion analysis using ESI in positive mode and TOF-MS detection (LCT, Micromass). Internal standards used: 1. Indole-3-acetic acid-d2, 2. 11-Aminoundecanoic acid, 3. Melatonin-d4. Rat urine spectrum (B) shows significant ion suppression for the internal standards in the range of 60%.

An example for ion suppression is given in Figure 7, which shows the direct injection of a rat urine sample spiked with internal standards after online SPE compared to the analysis of internal standards spiked into a saline solution. Although salts are removed by the online-SPE extraction, severe signal suppression is observed for the internal standard signals in the urine sample relative to the saline solution.

**B. LC-MS**

In order to overcome the drawbacks of directly injecting complex samples, liquid chromatography can be interfaced with the MS detector. Liquid chromatography can reduce ion suppression caused by coeluting compounds, isobaric interferences in the case of low-resolving mass analyzers, and often can separate isomers. In addition, a good analytical separation will result in better detection limits and MS data quality due to reduced background noise. Reversed phase liquid chromatography using C-18 narrow bore columns with particle sizes of 3–5 μm is widely used for metabolomic investigations, but the application of reversed phase capillary column separation is expected to increase in the future. In addition, the application of a monolithic-silica based C-18 capillary column for plant metabolomics has been reported (Tolstikov et al., 2003). However, conventional RP-HPLC separation is often insufficient for the separation of complex biological samples, resulting in poor metabolite resolution (see Fig. 8). One approach to increase chromatographic resolution and peak capacity is to use small particle sizes, such as the 1.7-μm particles used in ultra performance liquid chromatography (UPLC). Sub-2 μm particles provide narrow chromatographic peaks, which results not only in a better resolution but also in lower detection limits. Analysis of mouse urine using a 2.1 × 100 mm, 1.7 μm particle size C-18 column, and gradient elution produced chromatographic peaks with a peak width in the order of 1.8 sec at the base, resulting in a peak capacity of ca. 250 for a 10 min separation (Wilson et al., 2005a,b). The downside of this approach is the high pressure (10,000–15,000 psi) needed to operate these columns, and thus special UPLC systems are required.

Reversed phase chromatography is a standard tool for the separation of medium polar and non-polar analytes. However,
very polar metabolites are not retained on classical reversed phase stationary phases and elute with the void volume. An interesting alternative has been presented by Tolstikov and Fiehn (2002), who used hydrophilic interaction chromatography (HILIC) for the analysis of highly polar compounds in plant extracts. HILIC is orthogonal to reversed phase chromatography and uses either silica or derivatized silica, including amino, diol, amide, polysulfoethyl aspartamide, or polyhydroxyethyl aspartamide groups, with low-aqueous/high-organic solvent systems. Although similar to normal phase chromatography, HILIC requires a substantial amount of water in the mobile phase, which forms a stagnant water layer on the stationary phase surface into which polar analytes can be partitioned and retained. Contrary to normal phase chromatography, HILIC uses water-miscible polar organic solvents, such as acetonitrile or methanol. The application of HILIC for bioanalytical LC-MS/MS has been reviewed by Naidong, who also discusses the retention mechanisms involved in HILIC separations (Naidong, 2003). LC-MS based metabolic fingerprinting investigations have been performed using different mass analyzers including ion trap instruments (Tolstikov & Fiehn, 2002; Lafaye et al., 2003; Tolstikov et al., 2003; Wang et al., 2003), triple quadrupole instruments (Idborg-Bjorkman et al., 2003), TOF (Wang et al., 2003), and Q-TOF instruments (Plumb et al., 2002, 2003a,b; Lenz et al., 2004a,b; Williams et al., 2005). Triple quadrupole or Q-TOF analyzers provide the ability to perform MS/MS experiments for the structural elucidation of biomarkers. In the case of Q-TOF hybrid instruments, accurate mass measurement can be performed for precursor and product ions. Ion trap instruments have the advantage of MS^n capabilities. MS/MS experiments can also be used to probe the metabolome for specific compound classes by screening for characteristic ions and neutral losses, for example, the neutral loss of 44 amu (CO$_2$) observed in negative mode CID spectra is characteristic for carboxylic acids (Lafaye et al., 2003). In addition, phase II metabolites often show characteristic fragmentation patterns that can be used for rapid screening. Levensen et al. (2005) gives an excellent overview on the CID behavior of major Phase II metabolite classes, including glutathione conjugates, glucuronides, glucosides, malonylglucosides, sulfates, acetates, methyl and glycine conjugates.

The ionization technique of choice for LC-MS based metabolic fingerprinting is ESI. In order to obtain a broad coverage of the metabolome, ionization must be performed in positive and negative mode. This is illustrated in Figure 8, which shows the total ion chromatograms obtained in positive and negative ESI of a human urine sample after online SPE and HPLC separation. The figure emphasizes that ionization in positive and negative mode leads to the detection of two sets of analytes that can differ significantly. Depending on the mass analyzer used, detection in positive and negative mode can be performed simultaneously in a single run (Idborg-Bjorkman...
et al., 2003) thus, reducing the time needed for analysis and eliminating possible variability due to injection errors.

The application of LC-MS based metabolic fingerprinting includes the analysis of urine samples in toxicity studies (Idborg-Bjorkman et al., 2003; Lafaye et al., 2003; Lenz et al., 2004a,b; Williams et al., 2005) or genetic studies (Plumb et al., 2003a), plasma samples (Wang et al., 2003), plant extracts (Tolstikov & Fiehn, 2002; Tolstikov et al., 2003) and the analysis of fungal metabolites (Nielsen & Smedsgaard, 2003). Saghatelian et al. (2004) identified novel substrates of fatty acid amidase hydrolase (FAAH) by performing metabolic fingerprinting in brain and spinal cord extracts using RP LC-MS.

C. GC-MS

The combination of gas chromatography with electron impact ionization MS (EI-MS) provides high-chromatographic metabolite resolution, analyte-specific detection, and quantification of metabolites, as well as the capability to identify unknowns. However, a major prerequisite for GC-MS analysis is a sufficient vapor pressure and thermally stability of the analytes. The analysis of polar metabolites usually requires derivatization at the functional group to reduce polarity and increase thermal stability and volatility. A summary of common derivatization reagents is given in Table 2. Active hydrogens in functional groups, such

<table>
<thead>
<tr>
<th>Derivatization reagent</th>
<th>Product</th>
<th>Reagent leaving group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Silylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Example: Organic acid</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSTFA</td>
<td>Trimethylsilylester</td>
<td>N-Methyl-trifluoroacetamide</td>
</tr>
<tr>
<td>N-Methyl-N-trimethylsilyl-trifluoroacetamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTBSTFA</td>
<td>tert-Butyldimethylsilylester</td>
<td>N-Methyl-trifluoroacetamide</td>
</tr>
<tr>
<td>N-Methyl-N-tert-butyl-dimethylsilyl-trifluoroacetamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSTFA</td>
<td>Trimethylsilylester</td>
<td>N-trimethylsilyl trifluoroacetamide</td>
</tr>
<tr>
<td>N,N-Bis(trimethylsilyl)-trifluoroacetamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oximation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Example: Aldehyde</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O-Methylhydroxylamine hydrochloride</td>
<td>Methoxime</td>
<td>HCl, H2O</td>
</tr>
<tr>
<td>H2C=O—NH2*HCl</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2.** Common derivatization reactions used for GC-MS based metabolic fingerprinting investigations
as –COOH, –OH, –NH, and –SH can be derivatized by alkylation, acylation or silylation. In silylation, the active hydrogen is replaced by an alkylsilyl group, primarily trimethylsilyl (TMS), with formation of TMS ethers, TMS esters, TMS sulfides or TMS amines. Silyl derivatives show a better thermal stability and higher volatility and can produce more distinct MS spectra than their underivatized precursors. A number of different reagents for trimethylsilylation are available, including BSTFA (N,O-bis-(trimethylsilyl)-trifluoroacetamide) and MSTFA (N-methyl-trimethylsilyl trifluoroacetamide), which are being predominantly used in metabolomic investigations. BSTFA is highly volatile and produces volatile by-products, for example, the leaving group is trifluoroacetamide, which causes little interference with early eluting peaks. Comparable in silyl donor strength, MSTFA is the most volatile of the TMS acetamides and is therefore very suitable for GC-MS analysis. Often 1% trimethylchlorosilane (TMCS) is added to the reagents as a catalyst. A disadvantage of silylation is the sensitivity of the derivatives to moisture. Sample extracts must be thoroughly dried before derivatization and care has to be taken when handling the derivatized samples to avoid contact with moisture, resulting in derivative degradation. It has also to be noted that silylation can cause conversion reactions, for example, it has been described that arginine is converted into ornithine by reaction with BSTFA or MSTFA (Halket et al., 2005). Moreover, artifacts can be formed, complicating data interpretation (Little, 1999). In addition to trimethylsilylation, tert-butylidemethylsilylation (tert-BDMS) is being used for derivatization in metabolic fingerprinting. The EI-MS spectra of tert-BDMS derivatives often show characteristic fragmentation with a fragmentation of [M-57]⁺ caused by σ-bond cleavage at the Si atom and loss of the tert-butyl moiety, which can be helpful for the identification of unknown metabolites (Fiehn et al., 2000). Moreover, tert-BDMS derivatives have been reported to be stable with 2% residual water in the samples (Birkemeyer, Kolasa, & Kopka, 2003).

In addition to the derivatization of exchangeable protons, carbonyl groups are often transformed into the corresponding oximes with hydroxylamine or alkoxyamines, which primarily stabilizes α-ketoacids and locks sugars in open-ring conformation. Oxime derivatives can be formed as syn and anti isomers depending on the orientation at the carbon-nitrogen double bond. The isomers can be chromatographically resolved, leading to two GC peaks for each compound. Fiehn et al. (2000) found alkoxyamines superior to hydroxylamines for identification purposes. In order to obtain broad coverage of the metabolome by GC-MS, a two-step derivatization procedure including methoximation followed by silylation is often used for metabolic fingerprinting by GC-MS (Roessler et al., 2000; Colebatch et al., 2004; Jonsson et al., 2004). Gullberg et al. used a “design of experiment approach” to optimize factors influencing the extraction of plant tissue and subsequent derivatization for GC-TOF-MS analysis. The authors used methoximation and studied the influence of oximation temperature and time followed by silylation with MSTFA with 1% trimethylchlorosilane (TMCS) and variation of silylation temperature, amount of reagent, and addition of cosolvent (hexane, acetonitrile) (Gullberg et al., 2004). Using 12 reference compounds from different metabolite classes to study derivatization efficiency, the authors found methoximation with methoxyamine in pyridine for 17 hr at room temperature, followed by silylation with MSTFA for 1 hr at room temperature to be the optimum derivatization procedure.

Sample preparation for the GC-MS analysis of biofluids often includes only a lyophilization step. However, in this case, highly concentrated matrix components can interfere with the analysis. For example, urea in urine samples can cause a significant column and detector overload. Fu et al. (2000) describe a simple urease treatment to remove excessive urea for urine samples prior to silylation and GC-MS analysis.

Metabolomic investigations using GC-MS rely primarily on quadrupole instruments (Roessler et al., 2001; Roessner, Willmitzer, & Fernie, 2001) or TOF mass analyzers. Here, TOF-MS is mainly used for fast analyte detection, rather than for high-mass resolution. For example, the commonly used Pegasus TOF-MS instruments (LECO Instruments) provide high-data acquisition rates with nominal mass resolution. Nevertheless, high-mass resolution TOF-MS will be important for the identification of unknowns. Interestingly, GC-MS with ion trap instruments has not been adapted for general metabolomic investigations.

The combination of GC with TOF-MS provides rapid metabolite detection and has the potential for significant reduction of analysis time by implementing fast GC methods, thereby increasing laboratory throughput. Fast GC methods will reduce conventional GC run times from 40–60 min to less than 10 min by using short GC columns with small inner diameters. Fast GC results in very narrow peaks (peak width 0.5–1 sec) which cannot be sampled by conventional scanning quadrupole mass analyzers.

Although standard capillary GC columns provide a high-peak capacity, in the case of very complex biological samples the separation power of one column is often not enough. As an example, Figure 9 shows the GC chromatogram of a human urine (1 mL) extract after silylation with MSTFA. Deconvolution of the chromatogram using AMDIS resulted in 1,582 components, many of them overlapping.

One way to increase metabolite separation is the use of comprehensive two-dimensional chromatography (GC × GC). In GC × GC, each peak from a first usually non-polar column is transferred to a second (polar) column for further separation. In this way, three-dimensional (3D) chromatograms are obtained, with the retention times on the two columns forming two dimensions and peak intensity representing the third dimension. With MS detection a fourth dimension, the mass spectra, is introduced. The two columns are connected by a so-called modulator. The analytes are transferred from the first to the second column by dividing the eluent from the first column into small segments and focusing them in the modulator, for example, by cooling. In these cryogenic modulators, each individual fraction is cryo-trapped using CO₂ or N₂ cooled jets, focused and then, after rapid heating transferred into the second column as a narrow band. The focusing step reduces the peak width of the elution peaks significantly and therefore decreases the detection limit. Even without a second column dimension, an increased peak capacity can be obtained by the focusing step alone. The second short column separates the individual fractions within a few seconds. The whole process must be carried out very quickly because trapping, release and separation of a fraction must be finished before the next segment reaches the modulator. The resulting peaks are very narrow with baseline peak widths.
ranging from 80 to 600 msec (Adahchour et al., 2005). The narrow peaks require a fast scanning rate, and for identification purposes a TOF mass spectrometer is commonly used. The introduction of rapid-scanning quadrupoles that have scan speeds of up to 10,000 amu/sec provides a less expensive alternative for detection (Adahchour et al., 2005).

D. CE-MS

Capillary electrophoresis (CE) is a powerful technique for the separation of charged metabolites, offering high-analyte resolution. The combination with mass spectrometry makes CE-MS an ideal tool for the analysis of the metabolome. However, only a few applications have been published to date. Soga et al. (2003) developed a methodology using three CE separations for the analysis of anionic metabolites, cationic metabolites, and nucleotides and CoA compounds, to achieve a comprehensive coverage of the metabolome. Since the authors used a single quadruple mass analyzer, they choose a scan range of 30 amu to maximize signal intensity for the detection of low-abundance metabolites. In order to cover a range of \( m/z \) 70–\( m/z \) 1,027, 33 runs were necessary, resulting in a total analysis time of 16 hr for one sample. Using this technique, 1,692 metabolites in a bacterial extract were analyzed. Sato et al. used a similar approach to analyze major metabolites in rice leaves using CE-MS and CE-diode array detection (DAD) (Sato et al., 2004).

Another promising separation technique that has high potential for metabolomics investigations is capillary electrophoresis (CEC). CEC can be described as a hybrid of liquid chromatography and capillary electrophoresis. It commonly uses capillary columns that are packed with LC-stationary phases or have monolithic stationary phases. The mobile phase flow is caused by the electroosmotic flow, resulting in the plug-like flow profile typical of CE, which in comparison to the parabolic flow profile of LC, leads to narrow peaks, and high resolution. Pressure-assisted CEC has also been performed. The hyphenation of CEC with MS and the interfaces used has recently been reviewed by Klampfl (2004). The author also discusses different bio-analytical applications of CEC-MS, such as the analysis of proteins, peptides, amino acids, and saccharides.

VI. FLUX ANALYSIS

Metabolism is a dynamic phenomenon. In living organisms there is a continuous flux of matter mediated by fast enzymatic transformation of metabolites. Thus, measurements of intermediates of metabolic pathways represent measurements of
steady-state metabolite concentrations or snapshots of metabolite concentrations at the given point in time the samples were taken. In some instances, the steady-state concentration of metabolites may remain constant while flux changes dramatically. Therefore, steady-state concentrations may not reflect the underlying biological processes and do not describe the phenotype completely (Fig. 1). Flux analysis provides a true dynamic picture of the phenotype because it captures the metabolome in its functional interactions with the environment and the genome. Thus, metabolism-wide measurements of fluxes are a basis of phenomics, “phenotypic analysis of genomic information or entire mutant collections with the goal of understanding the relationship between genes and higher levels of organization in the cell” (Sauer, 2004). In addition, flux analysis is important for studying the kinetics and dynamics of xenobiotic biotransformation (e.g., pharmaceuticals or environmental pollutants).

The analysis of metabolite fluxes is performed using isotope-labeled precursors of metabolic pathways. Most of the known biochemical pathways were initially deciphered using radioactive isotope labels. However, detection limits of decay counters are not low enough to perform flux measurements in vivo using safe radioisotope doses. Accelerator mass spectrometry (AMS) can precisely quantify atoms in the attomole range and, therefore, be used to monitor how trace amounts of 13C-labeled compounds and other isotopes below the levels of legally regulated radioactivity move through metabolic pathways (White & Brown, 2004). AMS was applied to study in vivo the metabolic flux of human folate (Lin et al., 2004) and betacarotene (Dueker et al., 2000; Lemke et al., 2003). The metabolic fate of xenobiotics, such as the herbicide atrazine, was studied by HPLC-AMS in humans as well (Buchholz et al., 1999). However, AMS cannot perform structural elucidation of unknowns to confirm identity of metabolites. Therefore, additional techniques are required for identity confirmation (e.g., chromatographic coelution of the labeled compounds with a reference standard). Thus, modern AMS can hardly be used for global flux analysis on the level of the metabolome but it can be an excellent tool to study flux in a specific pathway. The high cost of AMS instruments precludes their widespread application, although the cost is dropping dramatically.

Metabolites labeled with 13C or other stable isotopes can be readily detected with standard NMR and MS instruments. Although their detection limits are significantly higher than the detection limits of AMS, standard instrumentation is commonly available, can confirm metabolite identity and provides structural information about labeled atom positions (positional isotopomers) and the number of labeled atoms in the molecule (mass isotopomers). Lower detection limits and high throughput make LC-ESI-MS/MS the preferential technique over NMR for flux analysis (Wittmann, 2002; Sauer, 2004). Because MS detects both natural and artificially added isotopes (e.g., 12C and 13C) and a large number of isotopomers can be derived from an initial precursor label, special data analysis methods are required to interpret the data.

An overview of mass spectrometry application in flux analysis was given recently by Wittmann (2002). Flux analysis is usually performed by GC-MS because EI fragmentation spectra provide information about isotopomer distribution. This technique was used to study metabolic fluxes, for example, in microorganisms Phaffia rhodozyma (Cannizzaro et al., 2004), Escherichia coli (Fischer & Sauer, 2003), in potato (Roessner-Tunali et al., 2004) and in humans (Hellman et al., 1997). To analyze non-volatile but ionic metabolites or to avoid derivatization steps, LC-ESI-MS/MS can be used as an alternative technique (Mashego et al., 2004; van Winden et al., 2005), In addition other techniques such as MALDI-TOF (Wittmann & Heinzle, 2001) were applied for flux analysis. Overall, MS-based integrated studies of all metabolite fluxes or MS-based fluxomics is just beginning to emerge (Hellerstein, 2004; Sauer, 2004).

VII. DATA EXPORT AND ANALYSIS

The ultimate goal of data export is to convert metabolic fingerprinting data into a standard and uniform format. This includes data pretreatment to reduce noise and background, alignment of mass spectra or mass chromatograms and automated picking and annotation of mass spectrometric or total-ion/extracted-ion chromatographic peaks of metabolites. Currently, several public and proprietary software tools exist to implement data export and subsequent data analysis. Since metabolomics is essentially the analysis of complex mixtures, any software that facilitates such analysis is referred to as metabolomics software in this review.

A. MS Data Format

With a variety of commercial and custom-built mass spectrometers a number of data file formats exist. This impedes the development of universal metabolomics software that does not depend on an analytical platform, especially the development of General Public License (GNU) software that has undergone tremendous growth in the analysis of genomics and proteomics data. Several solutions to this problem have been implemented. First, instrument manufacturers simply create their own software for the export of metabolomics data (e.g., LECO, Micromass). Second, independent software developers create programs, which are able to access a variety of commercial MS file formats (e.g., Advanced Chemistry Development, Inc.). Finally, open-access file formats such as ASCII text or binary netCDF exist, and instrument manufacturers often provide tools to convert original data files into these formats. Thus, platform-independent software can be created based on such standards.

While ASCII text is a good entry point file format for keeping low resolution or low volume MS data, the size of ASCII data file increases tremendously for high resolution or chromatography-MS data. In addition, there is no widely accepted text markup standard for MS data. Thus, a standard extensible markup language (XML) format for MS metabolomics data is yet to be developed. Currently, a universal mzXML format (Pedrioli et al., 2004) is becoming more widely accepted, however it is not design to contain biological descriptive information. XML format for metabolomics will not only provide standardization of quantitative data output but also a thorough sample annotation, which is required for the biological interpretation of the data.
Consequently, data become noisy and even false positive results two bins, the same centroided peak from different data files may the accurate mass of a metabolite is close to the margin between drifts in high-resolution mass measurement can occur. Thus, if assigning them zero intensities. However, for high-resolution masses to integer values, inserting missing mass values, and amu. In this case, the binning procedure is simply a rounding of with nominal mass resolution spectra because no information is son, & Tranter, 2001; Viant, 2003). Binning is highly compatible approach widely used in NMR metabolomics (Holmes, Nichol- transform mass spectral data into vectors of uniform length (see Fig. 10A). The most trivial way to bring the mass-intensity vectors to uniform length is to bin them in a fashion similar to the Fig. 10B). Although the information gained by separation is sacrificed, in the case of high-resolution mass spectrometry (Jonsson et al., 2005). If the bin size is kept low to represent instrument resolution, then very large data sets are created with only very few bins containing information (sparse matrices). In addition, there is a high probability of signal splitting between adjacent bins, which complicates subsequent data analysis. Another possibility to minimize the mass scale drift problem is to bin original continuous spectra, however, this makes subsequent analysis cumbersome.

Metabolites usually consist of carbon, hydrogen, nitrogen, and oxygen to a lower extent sulfur and phosphorus atoms. While hydrogen and nitrogen increase the decimal part of the accurate mass, oxygen, sulfur, and phosphorus atoms decrease it. Therefore, the accurate mass of some metabolites rich in oxygen and phosphorus (e.g., fructose 1,6-bisphosphate \( \text{C}_6\text{H}_{12}\text{O}_{7}\text{P}_2 \)) will be slightly less than the integer nominal mass. In addition, since metabolites are considered small molecules, the decimal part of their accurate mass will rarely exceed 0.4 amu for the majority of metabolites with molecular mass below ~500 amu. Therefore, the optimum set of low-resolution bins for majority of primary metabolites except heavy lipids are \( m_{\text{nom}} = 0.1 \text{ amu; } m_{\text{nom}} + 0.4 \text{ amu} \) and \( m_{\text{nom}} + 0.4 \text{ amu; } m_{\text{nom}} + 0.9 \text{ amu} \) where \( m_{\text{nom}} \) is a nominal mass. An example is given in Figure 11, which shows a histogram of metabolite decimal place distribution as detected by MarkerLynx (Micromass, UK) in metabolic fingerprints of cell culture media (m/z 100–1,200). Consequently, all ions representing metabolites will be added into the first bin, while the second bin ideally will contain multi-charged species (Smedsgaard & Nielsen, 2005).

An elegant approach to extract masses from high-resolution mass spectrometry is to build a density function of masses weighted by their intensities by combining all acquired spectra (Scholz et al., 2004). Thus, masses at high-intensity density are determined as the maximum of the density function. Then, each data file is reexamined and mass intensities, which coincide with the determined maxima within the given instrument precision, are extracted.

When the mass intensity vectors are transformed to uniform length, the entire dataset obtained by direct injection analyses (a single scan) can be represented as a 2D matrix \( (m \times f) \), where \( m \) is the length of the uniform mass vector and \( f \) is the number of files. In contrast, the entire dataset obtained with a separation step, which involves multiple scans, can be characterized by a 3D matrix \( (m \times s \times f) \), where the number of scans \( s \) is added, representing the retention time. Samples characterized by high-resolution mass spectrometry can be further classified using multivariate data analysis methods, such as principal component analysis (PCA) or parallel factor (PARAFAC) analysis for 3D matrices. Because the statistical analysis of 3D matrices is computationally difficult, it is also possible to sum mass spectra of specific chromatogram segments and convert them into 2D matrices (Plumb et al., 2002) (see also Fig. 10B).
spectrometry, only peaks of separated isomers and metabolites with small mass differences in the low-mamu range lose their identity (see Fig. 6).

The approaches described above are suitable for the classification of samples or the discovery of abundant metabolite peaks. However, the detection of low-abundance peaks in 3D data sets, for example, LC-MS or GC-MS runs, may be required for exact sample classification as well as for the discovery of biomarkers which are expressed at low abundances. In this computationally more intense approach, all detected peaks in the 3D data sets can be represented as a 2D matrix \( \binom{n}{C, f} \), where \( n \) is the number of individual peaks detected in all samples and \( f \) is the number of files in the datasets. Detection of metabolite peaks at nominal resolution can be performed using available peak detection algorithms (see Fig. 10C) (Windig, Phalp, & Payne, 1996; Koradi et al., 1998; Hastings, Norton, & Roy, 2002). For the analysis of high-resolution mass spectrometry data, signal density functions as those described above can be applied (Scholz et al., 2004) (see also Fig. 10D).

Each collected peak is annotated by a retention time and mass. Hard ionization techniques, such as EI, or post-source fragmentation in MS/MS and MS\(^n\) experiments add another layer of complexity because each metabolite produces several fragment ions. Thus, spectra deconvolution is required, which also reduces the complexity of the data. However, fragmentation also provides additional structural information. Therefore, each peak may be annotated not only by retention time and ion masses but also by a match from the library search (see Section VII.C). Deconvolution can also be used to reduce the complexity of chromatograms obtained with soft ionization techniques by filtering multiple charged species, clusters, and adducts (Jonsson et al., 2005). To match the peaks extracted from \( f \) files, all mass

---

**FIGURE 10.** Diagram showing different approaches for metabolomics data analysis. A: Mass-intensity vectors in each scan of the raw LC-MS or GC-MS data are binned to form mass-aligned vectors of uniform length. B: Scans across chromatogram area of interest are combined. C: Components are detected in binned data by peak detection algorithms in all scans belonging to one bin (selected ion chromatograms) or several bins (peak deconvolution). D: Peaks are detected in raw data using peak picking algorithms.
spectra and scans (chromatograms) have to be aligned across the entire dataset and/or matching criteria should be set, usually the mass and retention time windows. For the alignment, internal reference compounds for retention time and mass measurements can be added to the samples, or major peaks present in all chromatograms are used as references. Externally added non-endogenous compounds are preferred as references because they can also be used as retention index references, which further facilitate the identification of unknowns (see Section VII.C).

A few software products are currently available that meet the described requirements for data export. The Micromass MarkerLynx Applications Manager (Micromass, Manchester, UK) was used to detect drug metabolites in urine samples analyzed by LC-ESI-Q-TOF-MS (Plumb et al., 2003a). The program detects peaks in LC-MS chromatograms at high-mass resolution, performs automated peak integration and alignment using a single retention time reference standard, matches detected peaks between samples and outputs data as the described \((n \times f)\) matrix in a table which can be analyzed further by multivariate data analysis tools. In general, thousands of peaks (markers) can be extracted from a 1-hr gradient run of a complex biological sample. However, the software is not transparent and does not allow assessing the quality of peak extraction, for example, by visual examination of extracted ion chromatogram integration. Thus, a reasonable noise threshold level should be set to identify significant markers among all peaks.

A test alpha version of metAlign software (Plant Research International, Wageningen, Netherlands) was used for the extraction of plant metabolomics data (Tolstikov et al., 2003). The program performs noise prefiltering, baseline correction, and chromatogram alignment. Preprocessed chromatograms are summed into sample groups and summed chromatograms are compared pairwise. The output of the program provides significantly different characteristic peaks, which differentiate the two chromatograms. Integration of detected peaks is performed manually using instrument software. Similar pairwise comparison can be also performed by the COMPARELC function of the MS Processor program (ACD/Labs, Toronto, Canada) (Tolstikov, UC Davis; personal communication). The program applies an advanced component detection (CODA) algorithm to detect peaks. It is transparent and allows reexamination of each extracted-ion chromatogram of a detected peak.

MS Resolver and ReOrder (Pattern Recognition Systems, Bergen, Norway) were used to process data acquired by LC-ESI-MS at nominal resolution (Idborg-Bjorkman et al., 2003; Idborg,

**FIGURE 11.** A histogram of decimal places of metabolite accurate masses which demonstrates the optimum bin margins. Accurate masses of 1,008 metabolites were used which were discovered by MarkerLynx (MicroMass, UK) in a typical LC-MS fingerprint of cell culture media samples.
C. Data Normalization, Transformation, and Analysis

A detailed description of statistical methods for data analysis is beyond the scope of this review. A thorough review of currently available statistical methods was published for related NMR metabonomics (metabolomics) applications (Lindon, Holmes, & Nicholson, 2001). The statistical tools for metabolomics data analysis should be selected according to the aim of the study. If the aim is sample classification and prior information about the sample identity is unknown (e.g., in identification of silent mutation phenotype in progeny), unsupervised methods such as hierarchical clustering analysis (HCA), principal component analysis (PCA), or independent component analysis (ICA) are used. On the other hand, sample identity is often known and the aim of the study is to discover characteristic biomarkers (e.g., search for biomarkers of a disease comparing samples from healthy and diseased subjects). In this case, supervised methods such as principal least square (PLS) or soft-independent method of class analogy (SIMCA) can be used as well. The supervised methods use prior information about sample class and perform better in biomarker discovery (Jonsson et al., 2005). The supervised methods are also the preferred option for diagnostics studies (especially SIMCA), when classes of some samples are known and identities of others need to be determined (e.g., cancer diagnosis by comparing a patient’s metabolic fingerprint with the fingerprints of healthy and diseased subjects). The most popular data analysis tool in metabolomics is PCA which is available in many statistical software packages. However, if the aim of the study is a mere biomarker discovery in samples of known classes, common statistical methods such as ANOVA or the recently described ASCA (ANOVA-simultaneous component analysis) can be used (Smilde et al., 2005).

When significantly different markers are discovered, PCA can be used for data visualization. Strictly speaking, PCA is rather a data visualization tool that projects most variable components of a multidimensional data on a 2D or 3D plot, which can be comprehended by human mind in contrast to the multidimensional space. Thus, since variability of the data usually increases with absolute values, data transformation should be performed along with PCA. Otherwise, PCA will discover insignificant but variable components with high intensities and neglect significant but low-abundance biomarkers. The impact of variable components is actually one of the major reasons why high-intersample reproducibility is desired in metabolomics experiments. To minimize the impact of variability of high-intensity peaks, data normalization, and data transformation are used. Each sample is usually normalized to the sum of its intensities or to Euclidean norm. Transformation is intended to reduce high-intensity values but keep low-intensity values. Commonly, a logarithm function is used for the transformation. Because log \( 0 \rightarrow \infty \), an infinitesimal constant (Idborg-Bjorkman et al., 2003) or unity are added to all intensity values. An improvement of log transformation is a general logarithm (glog) transformation that was successfully applied to NMR metabolomics data (Purohit et al., 2004). The glog transformation takes advantage of the fact that the variance does not depend on mean values at low-signal intensities. Thus, more precise cluster analysis can be performed for signals close to the limit of detection.

Another perspective method to apply in metabolomics data analysis is ICA. ICA is similar to PCA but is less strict to the choice of components. It was shown that ICA is able to detect intersample variability caused by instrument drift as a separate component of the data while PCA failed to separate it (Scholz et al., 2004). As a result, improved separation between the samples was achieved because intersample variability was filtered out.

VIII. IDENTIFICATION OF METABOLITES

A. De Novo Identification of Metabolites

An important advantage of mass spectrometry tools in metabolomics investigation is achieving low-detection limits, which enables the detection of low-abundance metabolites. However,
identification of those compounds can become a problem if other spectroscopic techniques are needed as alternative tools for structure elucidation, as these tools often lack the necessary sensitivity. In this case, multiple fraction collection steps and sample pre-concentration are needed, which is not only labor-intensive but also bears the risk of metabolite decomposition during sample workup. For high-abundance metabolites, proton or other types of NMR is the technique of choice for structure elucidation. Often multiple metabolites can be produced by a single enzyme. Hence, changes in enzyme expression may change the expression of multiple but related metabolites that can be structurally similar due to the substrate specificity of the enzyme. Thus, structure determination of the most abundant differentially expressed metabolite by NMR may provide leads for the structure identification of less-abundant differentially expressed metabolites, which subsequently can be confirmed by mass spectrometry.

Experimental approaches for the structure elucidation by mass spectrometry are extensively described in the literature. Valuable information can be obtained by FT-ICR-MS because of its ultra-high resolution and mass accuracy, which allows determination of elemental composition and its capability for high-resolution MS* experiments, which provides detailed information about molecular structure units. A less expensive alternative to FT-ICR-MS for MS* experiments is a 3D ion trap mass analyzer. Finally, high-resolution Q-TOF instruments can be used for MS/MS experiments providing high-resolution data on fragment ions.

Often a combination of different mass spectrometric techniques is required for the structural elucidation of unknowns. For example, the structure of N-acyl taurines, representing a novel class of metabolites, was identified by FT-ICR-MS accurate mass measurements and triple quadrupole MS fragmentation studies (Saghatelian et al., 2004). From accurate mass measurements, the elemental composition was determined, while MS/MS spectra provided evidence of taurine residue in the molecules.

Moreover, derivatization can be used to identify functional groups and number of functional groups based on the observed mass shift. A hydrogen/deuterium exchange assay can be used to determine the number of exchangeable protons as a tool to screen for functional groups such as alcohols, amines, carboxylic acids, etc.

Metabolites can be also identified by library search. However, library search results should be confirmed by comparison with a reference spectrum of the standard and desirably by chromatographic retention of the standard because spectral differences for structural isomers can be minimal or not present at all. Thus, the use of retention time indices provides more confidence in assigning metabolite structure. Retention indices, or more precisely the Kovats indices, are routinely used in GC with the homologous series of n-alkanes used as references compounds. However, a retention index system for LC is yet to be established.

B. GC-MS Libraries

EI-MS spectra have been studied for several decades and, therefore, many EI-MS spectra libraries exist containing thousands of spectra. The creation of EI-MS libraries was facilitated by high-interinstrument reproducibility and the standardized electron energy of 70 eV commonly used for ionization. The largest spectra libraries are available from Wiley (~400,000 spectra) and the National Institute of Standards and Technology (NIST) (~200,000 spectra) as well as a combination of both with some additional spectra from Palisade (~600,000 spectra).

Large commercial libraries contain many spectra of synthetic compounds, which are unlikely to appear in biological samples. Although their spectra may provide leads for the structure identification of unknown metabolites, searching through thousands of spectra can be time consuming if hundreds of unknown metabolites need to be identified. In addition, less common metabolites are often not represented in large commercial libraries. However, there are metabolite libraries available, which contain mass spectra of metabolites found in specific biological samples. For example, a set of publicly available libraries of quadrupole and TOF spectra of plant metabolites was recently created as a part of the Golm Metabolome Database (GMD) (Schauer et al., 2005). These MS libraries also contain retention indices (MSRI libraries), which tremendously improve the identification of spectrally similar compounds. One of the tools performing spectra deconvolution and application of retention indices is the Automated Mass Spectral Deconvolution and Identification System (AMDIS) freely available from NIST (http://chemdata.nist.gov/mass-spc/amdis/) (Stein, 1999). The program was used for profiling of carboxylic acids in human urine (Halket et al., 1999) and for the identification of metabolites in phloems of Curucbita maxima (Fiehn, 2003).

C. LC-MS Libraries

LC systems are usually interfaced with APCI or ESI ion sources that produce even-electron pseudo-molecular ions ([M + H]⁺ or [M - H]⁻) usually without fragmentation. Therefore, additional fragmentation techniques are required such as in-source CID or conventional CID in triple quadrupole instruments or ion traps, which are more selective than in-source techniques and will be discussed below. Fragmentation spectra vary between different types of mass analyzers (triple quadrupole instruments and ion traps are usually compared) and even between instruments of the same mass analyzer type but different brands (Bristow et al., 2004; Jansen, Latchatre, & Marquet, 2005). Thus, the creation of universal libraries for LC-MS is impeded.

To minimize instrument-dependent variability, calibration ions can be used to tune fragmentation conditions according to conditions used to acquire the library spectra (Lemire & Busch, 1996). Ion-trap spectra were shown to be less dependent on collision energy parameters than triple quadrupole spectra, as observed for fragmentation spectra of steroids (Josephs & Sanders, 2004). To reduce the dependence of triple quadrupole instrument libraries on collision energy parameters, the library spectra can be generated by summing several spectra acquired at different collision energies (Josephs & Sanders, 2004). If the criteria for spectrum matching are lowered, for example, as a percentage fit of ≥60%, satisfactory inter-instrumental
identification through a library search can be achieved (Bristow et al., 2004). However, it is unclear if this criterion will be specific enough for the analysis of complex mixtures of metabolites. For metabolites that readily ionize in both ionization modes, similarity of both positive and negative ionization spectra can be a strong-fit criterion, even if a low-fit criterion is used for individual spectra.

Despite the described difficulties, MS/MS spectral libraries are being created. Recently NIST has published a new update of their library (NIST 2005) containing 5,191 MS/MS spectra of 1,943 different ions (1,671 positive and 341 negative ions). The spectra were acquired with different triple quadrupole and ion trap instruments. In addition, several user libraries exist, which are usually built using the same instrument model. To date 79 MS/MS spectra of common metabolites and drugs acquired with a Q-TOF instrument are available at the METLIN database (http://metlin.scripps.edu/) (Smith et al., 2005). To increase the specificity of library search, retention time is also included in the library. This approach was used for the identification of drug metabolites in urine samples (Pelander et al., 2003). A more detailed overview on currently available LC-MS and GC-MS libraries was recently published (Halket et al., 2005).

IX. DATABASES

Although a number of genomics and proteomics databases are available now, metabolomics databases are still limited. Biochemical databases can be used to identify unknown metabolites, for example, to identify structure from known elemental composition, or to determine the biological function of the identified metabolite. Information about biochemical pathways and metabolites involved in those pathways are available in the KEGG (http://www.genome.jp/kegg/) and BioCyc (http://biocyc.org/) databases. More specific lipidomics databases exist, such as LipidMaps (http://www.lipidmaps.org/data/index.html), SphinGOMAP (http://sphingomap.org/), and Lipid Bank (http://lipidbank.jp/index00.shtml), which contain structural and nomenclatural information as well as standard analytical protocols. General information about physico-chemical properties of metabolites can be obtained by searching general chemical databases such as PubChem or CAS. A general database search can also be useful for identifying the chemical structure of unknowns by starting with the elemental composition, which can be obtained from accurate mass measurements. For example, a search in the Merck Index was used to identify unknown water pollutants (Ibanez et al., 2005). The development of specific metabolomics mass spectrometry databases is also in progress. The Golm metabolome database (GMD) (http://csdbd.mpimp-golm.mpg.de/csbdb/gmd/gmd.html) contains publicly available mass spectra and retention index metabolite (MSRI) libraries and GC-MS metabolic profiles of plant samples (Kopka et al., 2005). The METLIN database (http://metlin.scripps.edu/) catalogues metabolites, MS/MS spectra, and LC-MS profiles of human plasma and urine samples (Smith et al., 2005). To date, existing metabolomics databases aim primarily at the structural identification of metabolites in various biological samples. However, once a better annotation of the metabolome in various organisms is achieved, the generation of databases containing quantitative metabolite data can be expected. An example for this type of database is the human metabolite database that contains more than 1,400 metabolites found in the human body. Each metabolite is described by a MetaboCard designed to contain chemical data, clinical data and molecular biology/biochemistry data (http://www.hmdb.ca/).

X. CONCLUSIONS

Technological advances in NMR and mass spectrometry have opened a new chapter in biochemistry by introducing metabolomics as an approach to study metabolism and its regulation in relation to disease, genetic, and environmental factors. Since the metabolome is the most predictive of phenotype, metabolomics holds the promise to extensively contribute to the understanding of phenotypic changes as an organism’s answer to disease, genetic changes, and nutritional, toxicological, environmental, and pharmacological influences. With regards to human health alone, multiple benefits of metabolomics investigations can be envisioned. It can deliver new tools to diagnose disease or monitor the success of nutritional and pharmaceutical interventions, for example, by using an unbiased metabolic fingerprinting approach. Based on research studies, metabolomics can provide new biomarkers to assess human health, and over time a powerful list of diagnostic markers will evolve, which can be measured using high-throughput assays. However, in order to proceed from the single biomarker concept to the global metabolome evaluation outside a research environment, the technology has yet to be developed to provide the clinician with the tools to assess entire wide classes of metabolites in biofluids and automatically process the data to evaluate the biochemical status of an individual.

The high selectivity of mass spectrometers in combination with low-detection limits, as well as their compatibility with separation techniques and their ability to deliver quantitative data makes mass spectrometry an ideal tool for metabolomic applications. So far, several excellent scientific contributions have been published demonstrating the enormous power of mass spectrometry applied to metabolomics. However, the field is rapidly developing and exciting new techniques and applications are anticipated in the near future, as we have seen in the rapidly growing field of proteomics.

Nevertheless, many technical and methodological issues have to be addressed to create analytical platforms that readily answer biological questions efficiently.

A major obstacle in this type of investigation is high diversity and variability encountered in the metabolome. The physico-chemical properties of metabolites are so diverse that none of the currently available mass spectrometric techniques can analyze all metabolites simultaneously. In general, an efficient sample introduction and ionization of all metabolite classes is of high importance. While EI and ESI ion sources are very efficient for certain types of analytes they have their own limitations. Among liquid MS interfaces, a promising ionization technique is the combination of ESI/APCI and atmospheric pressure photo ionization (APPI) or laser-based ionization that
enhances ionization efficiency for non-polar species. In addition, a more efficient separation of metabolites using chromatographic or electrophoretic techniques will decrease interferences and can provide lower detection limits. Multidimensional chromatographic techniques, such as GC × GC, are expected to gain wider distribution. A major goal is of course to further decrease the detection limits of the techniques used, not only to expand metabolomics studies towards the detection of low-abundance metabolites, but also to move to small sample volumes. This is not only important in the area of medicine but also opens up the exciting field of single-cell metabolomics. The trend towards miniaturization is already in progress with the evolving lab-on-the-chip technologies. Moreover, automated high-throughput sample preparation tools that are interfaced with the analytical platform are a vital technological development for metabolomics. A major obstacle in global metabolomics investigations using mass spectrometry is the quantification of the metabolites encountered, because signal intensity is not only function of concentration or mass but also depends on the chemical structure of the analytes and can be influenced by matrix interferences as observed in ESI. Stable isotope-labeled internal standards are therefore commonly used in metabolic profiling techniques. However, it seems unrealistic at this point that isotopic-labeled standards will become available, for example, for all plasma metabolites. Nevertheless, a centralized effort to design standard sets of isotopic-labeled reference compounds covering a wide range of metabolites would tremendously accelerate the work in metabolomics and would aid in the inter-laboratory comparison of semi-quantitative metabolic fingerprinting data.

Besides the technological aspects, data standardization, export, and analysis are essential components of metabolomics (Mendes, 2002). While high-throughput and semi-quantitative proteomics, with automated identification of proteins by computer algorithms, is available now (e.g., SEQUEST algorithm search), data analysis and visualization tools, libraries, and databases for metabolomics have yet to be developed.

ACKNOWLEDGMENTS

This work was supported by NIEHS Grants R37 ES002710, R01 ES013933, P30 ES05707, NIEHS Superfund Basic Research Program Grant P42 ES04699, and NIEHS Center for Children’s Environmental Health & Disease Prevention Grant P01 ES11269, the Cure Autism Now Foundation Frank del Olmo Memorial Award and in part by the German Research Foundation. We thank D. Hanna for the discussion of the manuscript.

REFERENCES


Katja Dettmer, Ph.D., studied Chemistry at the University of Leipzig and received her Ph.D. in 2001 from the Institute of Analytical Chemistry, University of Leipzig. She spent 4 years as a postdoctoral fellow and later as an assistant research scientist at the University of California, Davis, USA, where she started to work in the field of metabolomics. Currently, she is a research scientist at the Institute of Functional Genomics at the University of Regensburg, Germany. She is conducting research in the field of metabolomics, focusing on the development of mass spectrometry-based tools for metabolic fingerprinting in biofluids as well as metabolic-profiling methods.

Pavel A. Aronov studied Environmental Chemistry at the Novosibirsk State University and graduated with honors in 2001. His interest in Analytical and Biological Chemistry brought him to the laboratory of Bruce Hammock at the University of California, Davis, where he started to work in the field of metabolomics. Currently he works on his Ph.D. focused on metabolic fingerprinting, metabolic profiling of biologically active lipids, and chemical derivatization for electrospray ionization.

Bruce D. Hammock received his Ph.D. in Entomology and Toxicology at U.C. Berkeley in 1973 and has subsequently taught at Northwestern University, UC Riverside and since 1980 at UC Davis where he is a distinguished professor in Entomology and the Cancer Research Center. He has been PI of a NIEHS Superfund Basic Research Program for over 20 years, is the director of a NIH Biotechnology Training Grant and is a member of the US National Academy of Sciences. The UCD-NIEHS analytical laboratory that he directs has had a metabolomic emphasis for 10 years. His laboratory works in aspects of analytical chemistry, biosensor development, mammalian pharmacology, and agricultural-medical pest control.