Chapter 10

Metabolomics

Karl Burgess,[a,b] Naomi Rankin,[a,c] and Stefan Weidt[a,b]

[a]Glasgow Polyomics, University of Glasgow, UK, [b]Institute of Infection, Immunity and Inflammation, University of Glasgow, UK, [c]Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK

Chapter Outline

10.1 Overview 181
- 10.1.1 The Metabolome and Metabolomics 181
- 10.1.2 Biochemistry of the Metabolome 182
- 10.1.3 Applications of Metabolomics in Clinical and Biomedical Research 183
- 10.1.4 Experimental Design for Metabolomics 183

10.2 Metabolomics techniques and technologies 184
- 10.2.1 Fingerprinting 184
- 10.2.2 Metabolite Pool Analysis 185
- 10.2.3 Metabolite Profiling by MS 185
- 10.2.4 Metabolite Profiling Using NMR 185
- 10.2.5 Flux Analysis 187

10.3 Analytical platform 187
- 10.3.1 Nuclear Magnetic Resonance 187
- 10.3.2 Mass Spectrometry 192
- 10.3.3 Liquid Chromatography–Mass Spectrometry 194
- 10.3.4 Gas Chromatography–Mass Spectrometry 196

10.4 Analysis of metabolomic data 197
- 10.4.1 NMR-Based Metabolomics Data 197
- 10.4.2 MS-Based Metabolomics Data 199
- 10.4.3 Contextualization of Metabolomics Data 200
- 10.4.4 Annotation/Identification 200

10.5 Conclusion 200

References 201

10.1 OVERVIEW

The publication of the draft human genome on February 15, 2001, [1] was a significant milestone in the understanding of human disease and diagnosis. An enormous repository of information on the basic building blocks of the human organism had been made available to researchers, allowing groundbreaking research into the relationship between genes and inherited diseases and disease susceptibility. However, by exploring the linkage between genetic influence and the overall phenotype, it was demonstrated that sequencing the genome is a starting point in the analysis of the complexity of an organism.

Genetic events propagate from DNA (the genotype) through transcription and translation to ribonucleic acids and proteins [2]; these in turn have an effect on the small molecules that make up the bulk of a cell. The combination of these levels creates the phenotype [3]. Crucially, extracellular and extraorganismal signals propagate via small molecules, proteins, and nucleic acids, effecting transcription, translation, post-translational protein modification and chemical reactions with other small molecules.

The discipline of genomics was formed to exploit the plethora of genetic information available both on human disease, including single-nucleotide polymorphisms and epigenetic modifications within the human genome, and from genomic data from other human health–related organisms such as pathogens. It has already had significant impact on health research, via genome-wide association studies [4–6] and SNP analysis [7–9].

Following on from these efforts, post-genomic technologies have developed to exploit the enormous quantity of information available at the protein and metabolite levels, known as proteomics and metabolomics, respectively.

10.1.1 The Metabolome and Metabolomics

The metabolome was first described by Oliver and colleagues in 1998 [10], during their pioneering work on yeast metabolism, as the complete complement of small molecules in a biological system or fluid. It therefore follows that the discipline of metabolomics is the analysis of the metabolome: the small molecules in a biological sample. Metabolomics is also referred to as metabonomics or

metabolic profiling by some groups, but *metabolomics* is the term most commonly used and is therefore used here.

Use of gas chromatography–mass spectrometry for metabolite profiling of clinical urine samples is possibly the earliest application of what we would now describe as metabolomics, dating from the 1970s [11–13]. The application of nuclear magnetic resonance spectroscopy to profile clinical samples was pioneered by Nicholson and others in the 1980s [14–16]. Later, mass spectrometry coupled with both liquid chromatography [17] and capillary electrophoresis [18] came to be applied to metabolomics to round out the analytical techniques available.

10.1.2 Biochemistry of the Metabolome

Unlike the genome (comprising only 4 different nucleotides) and the proteome (comprising 20 different amino acids), which have a restricted set of building blocks (excluding epigenetic and post-translational modifications, which tremendously complicate these disciplines), metabolites, as individual molecules, include an enormous diversity of organic structures and functional groups that allow them to perform their particular functions. Indeed, these building blocks—amino acids, phosphosugars, nucleotides and their precursors, and degradation products—are all metabolites.

While many metabolites possess functions intrinsic to their structures, they do not exist in isolation *in vivo*. Metabolites are the substrates of enzymes and are chemically modified in anabolic and catabolic processes. Substrates may enter cells from the extracellular matrix and be transported out again as secreted molecules. They can also exist in chemical equilibrium with other molecules and convert readily between two states. For ease of reference, a series of chemical reactions mediated by enzymes in a cell are described as pathways. They include the most basic functions of any living cells, including

- Energy generation via glycolysis, the tricarboxylic acid cycle, the pentose phosphate pathway, and fatty acid oxidation, coupled with the electron transport chain; energy storage via gluconeogenesis; and fatty acid synthesis or glycogen synthesis.
- Synthesis of macromolecules via DNA replication, tRNA synthesis, and protein synthesis.

Understanding these pathways is critical to the interpretation of metabolomics data. The reader is encouraged to refer to biochemistry textbooks for help in interpreting the data [19–21].

Pathways themselves do not exist in isolation, but the inputs and outputs of each pathway are themselves the products of other pathways such that all pathways are interconnected in a complex network (see Figure 10.1) that constitutes an organism's complete metabolism [22]. This provides an enormous analytical challenge because, for example, an extremely sensitive method for the detection of organic acids may be unsuitable for the detection of amine compounds. As a consequence, the researcher must either focus her efforts on a single analytical platform

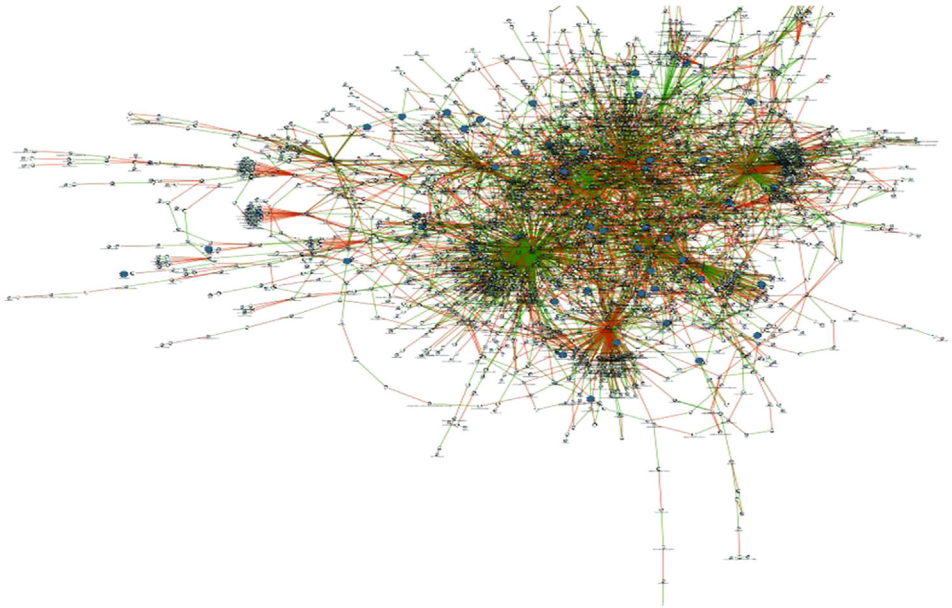

FIGURE 10.1 Complete network of plant metabolites. Metabolites are nodes; enzymatic reactions are edges. Some metabolites, notably water, are excluded from the network due to their ubiquity in reactions. Note the dense clusters surrounding some nodes, which are the consequence of highly promiscuous metabolites such as ATP. *Source: Image generated using MetExplore* [23].

that is capable of detecting a subset of the metabolome, or apply multiple methods to each sample to improve overall coverage.

The total number of metabolites in a biological system is currently the subject of considerable debate, with estimates of between a few hundred for small eukaryotes [24] to thousands in higher mammals [25], to hundreds of thousands in plants [26], with their complex array of secondary metabolites. In addition, while drugs and xenobiotics are generally not considered to be metabolites per se, analytical techniques used to detect endogenous metabolites also detect drugs because of their similar physicochemical properties; therefore, metabolomics is a powerful tool for drug mode-of-action studies.

10.1.3 Applications of Metabolomics in Clinical and Biomedical Research

While analysis of the entire human metabolome, even at the low estimate of a few thousand metabolites, is a goal that metabolomics is far away from reaching, the capability to analyze hundreds of individual small molecules or profile the characteristic signatures of the entire metabolome makes metabolomics an extremely powerful tool that can be applied to a wide variety of biological questions. In cancer science, the Warburg effect [27] of altered energy metabolism in tumor cells has recently been an important target of metabolomic research, with several groundbreaking papers on the energy metabolism of cancer [28,29]. Moderate-scale clinical metabolomics studies in the cardiovascular field, including preeclampsia [30], diabetes [31], and stroke [32], have provided clues as to common disease mechanisms. More specific studies on organs and tissues associated with significant issues in human health have been performed (e.g., liver function [33], pancreatic cancer [34], and lung function [35]).

Parasite (e.g., Barrett et al. [36], Vincent et al. [37], and Vincent et al. [38]) and bacterial (e.g., see references [39–41]) metabolomes that respond to drug action and host interactions have been extensively studied. Indeed, there is a growing interest in the microbiome and metabolomics [42,43] because, while genomics can provide descriptive analyses of the species of microorganisms and their microenvironments (e.g., the gut), the majority of metabolites are cross-species, which allows researchers to study the complete metabolome produced by a community of microbes, including its unique molecules, toxins, and other compounds.

Gas-chromatography mass spectrometry (GC-MS) is particularly powerful in the profiling of samples that contain molecules with structural similarity, such as sugars [44–48], drugs of abuse [49–51], and steroids [52–55]. While the majority of applications are targeted methods for quantitation, GC-MS has also been used for global metabolomics profiling [56–60].

Clinical samples analyzed by nuclear magnetic resonance (NMR) metabolomics include whole blood [61], reconstituted blood spots [62], plasma [63], serum [63], urine [63], saliva [64], cerebrospinal fluid [65], and digested tissue biopsies [63]. NMR-based metabolomics has also been used in lipoprotein analysis for prediction of cardiovascular risk [66,67], early identification of sepsis in critically ill patients [68,69], identification of novel inborn errors of metabolism [70], and drug toxicity [68,71].

Many organisms that model human disease have also been analyzed by metabolomics. For example *Drosophila* studies have been related to kidney function [72–74] and spleen function has been studied in mice [75]. Finally, interdisciplinary projects spanning bioengineering and orthopedic surgery have used metabolomics to optimize surfaces for the maintenance and differentiation of stem cells [76–78].

10.1.4 Experimental Design for Metabolomics

Of course, with such a diversity of applications, good experimental design is essential to achieving high-quality results. It is important to decide whether the experiment will be an unbiased study for hypothesis generation or a specific study chosen to test a hypothesis [79]. The former is the canonical biomarker discovery experiment, while the latter exemplifies validation of biomarkers in a larger cohort.

As in any experiment, a sufficiently large number of controls must be carefully chosen to eliminate bias. For health/disease studies, it is important to use positive and negative controls to separate general effects of ill health from specific disease-associated markers. Samples should also be appropriately randomized and quality control samples must be rigorously acted on if the data obtained fail to meet appropriate standards to eliminate or compensate for inter- and intrabatch variation [80]. Internal standards should also be selected to ensure effective extraction [81].

An appropriate number of biological replicates should be chosen to provide the appropriate statistical power [82]. More specifically, the larger the coefficient of variation in metabolite levels in the samples, the higher the number of replicates and the larger the minimal fold change required for statistical significance. In general, for clinical samples from small sample groups only very large variations from control levels have statistical power. Less than twofold changes in metabolite concentration in highly variable clinical datasets require very large numbers of samples (hundreds or thousands, dependent on the variation observed). While well-controlled cultured cells have low variation, a smaller number of replicates provides similar power. Technical variation in metabolomics experiments is predominantly lower than biological variation [83]. A typical experimental design for a simple pairwise comparison is shown in Figure 10.2.

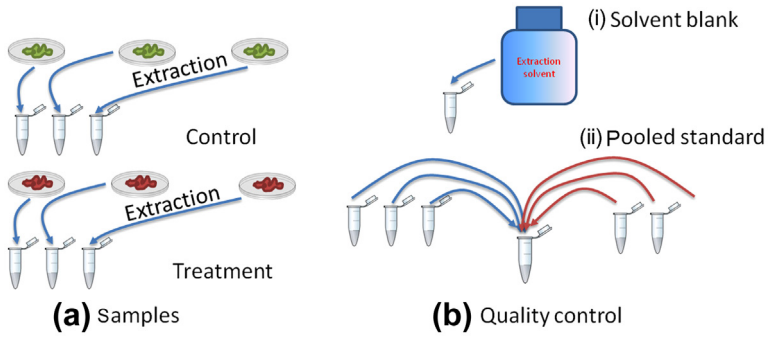

FIGURE 10.2 Simple experimental design of a pairwise metabolomics comparison. (a) Minimum three biological replicates from each state to be analyzed (replicate numbers should be carefully chosen to optimize detection of the desired fold change). An aliquot of extracted fresh medium could be added to these samples, as well as the spent medium from each biological replicate, to observe depletion of metabolites. (b) Quality control samples, including a solvent blank (i) to eliminate spurious compounds introduced in the extraction solvents and (ii) a pooled standard (i.e., a mixture of a small aliquot of the other samples) to enable tracking of compound degradation, batch effects, and instrument drift.

Once the experiment has been carefully planned, it is important to select an analytical method that will have the best chance of producing the desired outcome.

10.2 METABOLOMICS TECHNIQUES AND TECHNOLOGIES

While all "omics" technologies are capable of providing information on the state of an organism, metabolomics can also provide information on food or medium intakes as well as outputs in terms of urine, feces, spent media, and so forth. This application of metabolomics to the inputs and outputs of a system, coupled with analysis of intracellular metabolites, provides additional information. Thus it is possible to determine by comparison between spent and fresh medium that cells have depleted a compound from their environment. A commonly described example of this is the consumption of glucose from the medium. Fresh medium may contain large quantities of glucose; spent medium may not. If there is a very small pool of glucose in the intracellular metabolome, it can be inferred that the glucose has been taken up and fully utilized. Conversely, if glucose is low intracellularly and unchanged between fresh and spent medium, this may indicate that the glucose is not being used by these cells.

Metabolomics uniquely comprises a growing number of competing platforms. The primary tools for metabolomics analysis are liquid or gas chromatography hyphenated to mass spectrometry (LC-MS or GC-MS) and nuclear magnetic resonance (NMR). Capillary electrophoresis, as pioneered by Soga [18], is becoming more popular as a separation method for charged metabolites prior to mass spectral analysis. High-throughput spectroscopic profiling techniques relying on FT-IR [84,85] and Raman spectroscopy [86,87] have also become popular. Each methodology has strengths and weaknesses for particular experimental types, which should be borne in mind when designing a particular type of metabolomics experiment.

In this section, we follow the four major classifications of metabolomics experiment (fingerprinting, metabolite pool analysis, flux analysis, and metabolite profiling) that are dealt with in Birkemeyer et al.'s excellent review on metabolomic experimental paradigms [88].

10.2.1 Fingerprinting

Metabolite fingerprinting is essentially a sample stratification and classification strategy. In this type of experiment, the identification of individual compounds is not the focus, as the principle is to obtain diagnostic patterns from the data that can be used, for example, to distinguish disease status from healthy status. This strategy was pioneered in the NMR, FT-IR, and Raman spectroscopy fields, and it has been applied to several major health issues such as coronary heart disease [66] and diabetes mellitus [89].

10.2.1.1 Metabolite Fingerprinting via NMR Spectral Binning

For metabolome-wide association studies (MWAS), multivariate statistical methods are commonly used [68,90]. High-throughput NMR is performed on the samples. The data are binned and commonly, orthogonal PLS multivariate analysis is performed on the raw spectra, with correction for false discovery rate [68,90]. Once peaks that distinguish cases from controls (or similar) are identified as being important, the unknown metabolite is then identified. This is essential for biological interpretation.

Metabolic fingerprinting can be used to great practical effect—for example, in the iKnife, developed by Balog et al. [91] for real-time tissue identification during tumor surgery.

10.2.2 Metabolite Pool Analysis

The canonical targeted metabolomics experiment is essentially one where the pool of metabolite abundances is assessed (for an early example, see reference [92]). When applied to both LC-MS (e.g., see reference [93]) and GC-MS [94], absolute levels of a particular metabolite can be obtained with the addition of a 13C-, 15N-, or deuterium-labeled internal standard to the sample, accompanied by preparation of a standard curve. This obviously renders the experiment a targeted one, where only the metabolites of interest are obtained. Some efforts to produce uniformly labeled organisms have resulted in a larger number of available labeled standards [95,96], but this requires that the organism in question be culturable in bulk and have identical metabolites to those of interest, and that standard curves for abundance are still prepared for each metabolite.

Metabolite pool analyses are static—they provide no information about rates of change or movement of metabolites. They simply tell you that one or more metabolites change or do not change in abundance as a result of a stress or treatment. For drug mode-of-action studies or analysis of a knockout mutant, if the drug is an enzyme inhibitor, the metabolite prior to the enzyme is likely to be significantly upregulated, and the metabolites downstream will be downregulated or not present post-inhibition. Due to the presence of metabolic shunts that can route around biochemical pathway blockages, this is not an infallible method for analyzing the phenotype of a mutant or drug treatment, but in many cases it can produce valuable data.

Metabolite pool analyses are the standard clinical measures of metabolites. Biomarkers such as glucose for diabetes; bilirubin, used to screen for or monitor disorders of the liver; urea and creatinine used to monitor kidney function; and lactate to detect hypoxia are measured routinely as they provide an indication of dysfunction of biochemical pathways.

10.2.3 Metabolite Profiling by MS

Metabolite profiling is somewhat halfway between observing the overall phenotype by using a metabolomic measure and applying a targeted approach. Metabolite profiling focuses on identification of the largest number of compounds possible, sacrificing absolute quantitation of the sample in exchange for relative quantitation of a far greater number of compounds.

Mass spectrometers are most commonly used for this purpose, as their sensitivity to individual compounds is far greater than any other technique. It is possible to produce relative quantitation between groups of samples, such as wild type/mutant or healthy/diseased for a large number of compounds [80,97–99]. Further details on MS profiling can be found in later sections.

10.2.4 Metabolite Profiling Using NMR

NMR metabolomics can also restrict analysis to a smaller set of well-characterized metabolites. For example, the Ala-Korpela group's method measures over 200 metabolite measures: >100 directly and >100 derived. Directly measured metabolites include >20 amino acids and other clinically relevant metabolites such as glucose, creatinine, and 3-hydroxybutyrate. Also directly quantified by NMR are approximately 80 measures of lipoprotein subclass and lipoprotein content. The 99 derived measures include apollipoproteins, total branched-chain amino acids, Fischer ratio for liver fibrosis, and lipid ratios [100].

It is common in NMR metabolite profiling experiments to apply multiple pulse programs or differential extraction of samples to increase the number and type of compounds detected. The exemplar method uses three different "windows": the LIPO window for lipoproteins and the low-molecular-weight metabolites (LMWM) window, which are both performed on native serum, and the LIPID window, which is performed on extracted lipids [101].

The LIPO window is a standard 1D-^{1}H-NOESY pulse program (to be discussed later). The spectrum obtained contains highly overlapping spectra. The baseline is automatically zeroed and the peaks are aligned. The albumin background must be subtracted [103]. Regression modeling is used to quantify VLDL, LDL, IDL, HDL, and triglycerides. These models are cross-validated to NMR-independent lipid and lipoprotein analysis [101,103].

The LMWM window is a standard CPMG pulse sequence [101]. As described in the Section 10.3.2. Common 1D NMR Pulse Programs, the FIDs are acquired after a longer period, once the larger proteins and lipoproteins have stopped spinning. There is therefore much less spectral overlap. The LIPID window is obtained from extracted lipids reconstituted in deuterated chloroform containing TMS as the internal standard [101]. (See Figure 10.3.)

The results of all the different windows are brought together, and the 99 derived measures are calculated. Self-organizing maps, based on unsupervised pattern recognition of the spectra, are then used to group patients [103]. Other clinical measures are then overlaid on these maps as a visual representation of the correlation between the spectra and the variable of interest; p-values are also calculated [104].

The huge benefit of this method is that it is high-throughput, at almost 200 samples per day, which makes it ideal for large clinical cohorts. It has already been shown to be important for risk assessment for cardiovascular disease [103], prediction of diabetic complications [104], development of cognitive impairment and Alzheimer's disease [105]. Moreover, it has been used to observe the effect of birth weight, diet, physical activity, and timing of puberty on the NMR metabolites detected.

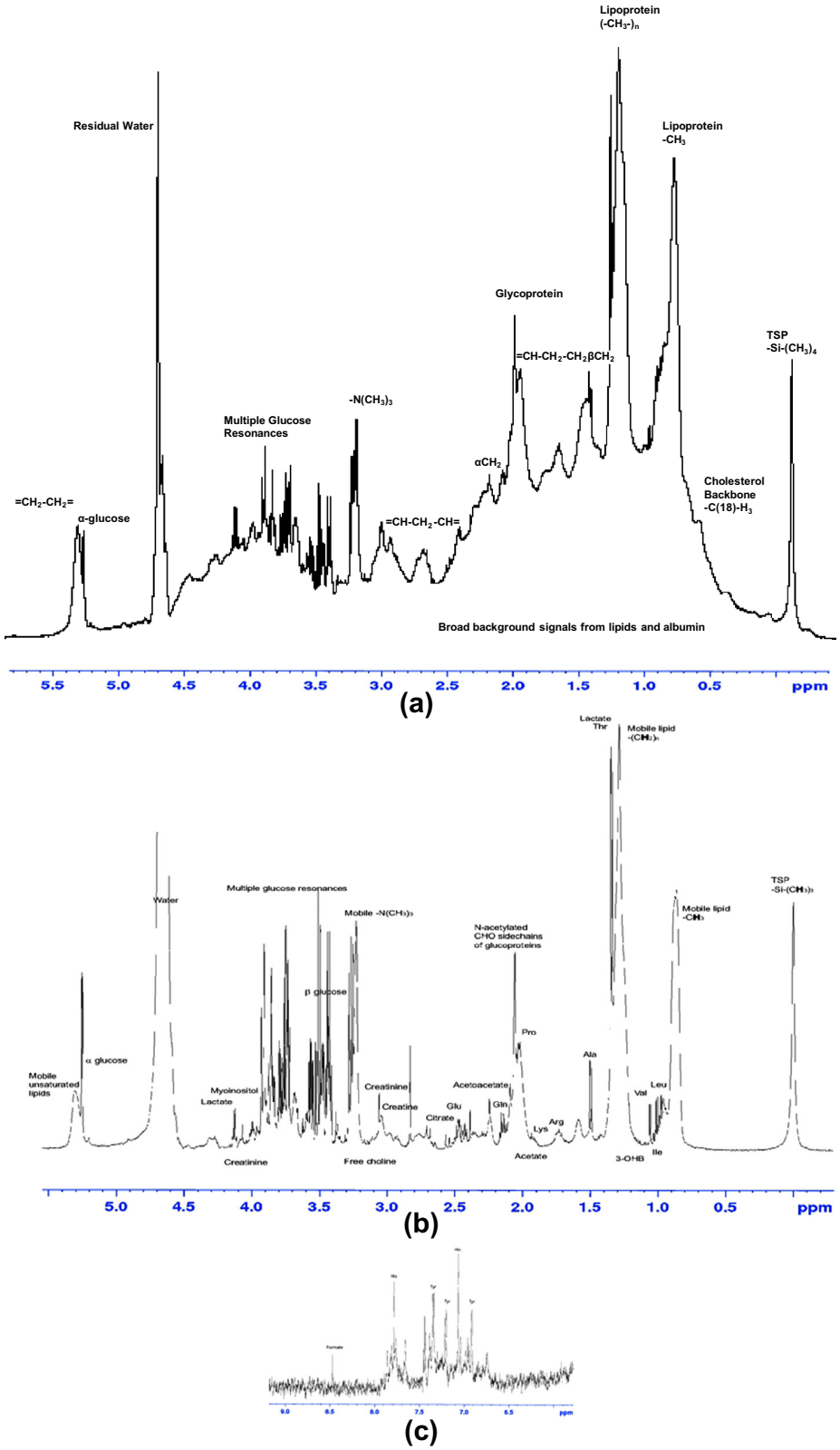

FIGURE 10.3 Example ^{1}H NMR spectra of serum obtained by the Ala-Korpela method. Samples were analyzed on a 14.1 T (600 MHz) NMR with Avance III HD console. (a) ^{1}H NOESY (LIPO window). (b) ^{1}H CPMG (LMWM window). (c) Aromatic region of LMWM window. Note that the large diffuse background in the LIPO window has been suppressed in the LMWM window. The resonances in the LIPO window can be used to determine quantitative lipoprotein profiles. The LMWM window is used to identify and quantify an array of small molecules. The inset shows the aromatic amino acid resonances. *Source: Annotation based on Soininen et al.* [102].

As the method is high-throughput and can identify and quantify multiple metabolites, it combines well with genome-wide association studies [106,107]. The more detailed the phenotyping is, the greater the ability to detect genotype-phenotype associations. This high-throughput NMR method has been used to identify genes associated with heritability of the above metabolites, especially lipids [100]. The method has also been used to identify novel genes associated with atherosclerosis [108].

10.2.5 Flux Analysis

Metabolic flux is the passage of a metabolite through a reaction system over time, and flux analysis is the combination of time-course methodologies in metabolomics and computational modeling of pathways. It is one of the most effective methods for analyzing pathways and determining the fate of a molecular substrate (see reviews in references [109,110]) and is used in fields as diverse as synthetic biology [111], antiviral therapy [112], and cancer [113].

Essentially, in stable isotope labeling an organism is provided with a food, usually medium, although some success in stable isotope labeling in higher organisms has been obtained which contains a heavy labeled version of a particular metabolite. The most common example in metabolomics is the application of ^{13}C glucose to a cell culture to trace the fate [114]. Labeled substrate is normally added in a 50% load with the unlabeled compound. Time courses investigating the rate of incorporation of a labeled substrate can be applied to inform flux balance analysis of a specific biochemical pathway with the use of appropriate software such as MzMatch.ISO [115] or MAVEN [116]. Metabolic flux analysis in the era of metabolomics can be performed using either mass spectrometry or NMR platforms, with characteristic gaps between mass peaks denoting labeled and unlabeled compounds in the case of the former.

^{13}C signals can also be detected by NMR, one of the advantages of which is that the chemical shifts are spread over approximately 200 ppm [71] (see the relevant section). Natural abundance ^{13}C (1.1%) can be difficult to detect due to the low abundance and reduced sensitivity of ^{13}C NMR. ^{13}C labeled metabolites can be made, which can help with the few metabolites that are ^{1}H NMR invisible, such as uric acid. This is a result of either a lack of protons (azide) or a very rapid exchange of protons on the metabolite with matrix protons (uric acid) [117]. It is also useful for tracking metabolite pathways.

10.3 ANALYTICAL PLATFORM

Once the study is designed and the type of experiment to be performed has been selected, it is important to determine the analytical platform. When selecting a platform for metabolomic analysis, it is important to consider its benefits and limitations prior to designing the study.

10.3.1 Nuclear Magnetic Resonance

The advantages of using nuclear magnetic resonance (NMR) for metabolomics are as follows:

- It is nondestructive and selective.
- Sample preparation is simple.
- It can be used for absolute quantitation.
- It provides information on structure
- It is independent of analyte polarity.
- It can be high-throughput and automated (associated with cost reduction).
- It can be considered an almost universal detector [118,119].
- The spectra are highly reproducible for most sample types; urine is more variable.

One-dimensional NMR is very fast, and no chromatographic preseparation or sample derivatization is required.

NMR metabolomics is also highly suitable for large-scale cohort and similar studies. Perhaps the most unique and important feature of NMR is that is can be used to identify novel compounds (generally accurate mass MS to obtain the molecular weight and 2D NMR methods are required) [120]. NMR spectroscopy is a technique that exploits the magnetic properties of atoms in order to obtain information about the structure of a molecule, and hence its identity. With its powerful quantitation capabilities [121], it is highly suited to metabolomics.

NMR detects nuclei that spin when placed in a strong magnetic field. The spin is either aligned with the magnetic field (spin-up) or opposed to it (spin-down). The nuclei in the spin-up state are in the low-energy state; those opposed to the magnetic field are in the high-energy state, as energy is required to oppose the magnetic field. Electromagnetic radiation, in the form of radiofrequency (RF) pulses, is used to push some of the nuclei in the low-energy state into the high-energy state. Once the pulse has been stopped the nuclei relax back to equilibrium. The energy is returned in the form of free induction decay (FID). These signals result from oscillation of the nuclei, which decays with time. The NMR instrument acts as both a radio transmitter and a radio receiver. The FID can be Fourier-transformed in order to convert the time domain into the frequency domain. This allows the data to be interpreted [122] (see Figure 10.3 for an example of a typical NMR spectrum of plasma using a CPMG pulse program, discussed later).

The most common nucleus observed in NMR experiments is the proton (^{1}H). Proton NMR allows absolute quantitation. It can detect almost anything with ^{1}H nuclei, including amino acids, carbohydrates, carboxylic acids, and so forth, provided it is in a high enough concentration.

There is a high abundance of hydrogen in metabolites, and NMR is very sensitive to proton signals [119]. Other commonly observed nuclei are carbons (^{13}C) and nitrogens (^{15}C) [121].

The RF must be exactly equal to the energy gap between the spin-up and spin-down states. Therefore, the higher the magnetic field strength (Tesla), the larger the difference in energy (MHz) between the two spin states (up versus down), resulting in better resolution. Sensitivity also increases with increased magnetic field strength [118]. For example, an increase from a 14.1 T/600 MHz magnet ($14.1^{3/2} = 52.9$) to a 16.4 T/700 MHz magnet ($16.4^{3/2} = 66.4$) results in an increase in sensitivity of 1.25 (66.4/52.9).

The first step of an NMR experiment is to design the pulse program. There are numerous predesigned sequences available with NMR software. These can be chosen from the catalog depending on the type of experiment you wish to perform and can be adapted to suit your sample type, solvent type, research question, and so forth.

A detailed guide to NMR pulses is outside the scope of this chapter. The basic pulse program begins with a delay (as this pulse will be repeated multiple times). A 90 pulse is then applied to transfer the magnetization from the z-axis to the x-y axis, where it can be observed, and to push the nuclei from the low- to the high-energy state. This is followed by the acquisition period, where the signal emitted by the precessing nuclei is recorded [121].

10.3.1.1 NMR Instrumentation

A typical NMR spectrometer setup consists of a superconducting magnet, a probe, a console, and a computer. The superconducting magnet is often a solenoid made from nobium titanium alloy. A large current flows around the loop, creating the strong continuous magnetic field required to align the nuclei. The magnet must be kept extremely cold, 4.3 K (−268.85 °C), to maintain minimal resistance (i.e., to continue acting as a superconducting magnet). This is achieved by maintaining the magnet under liquid helium in a vacuum jacket. As liquid helium is expensive and stocks are diminishing, an outer jacket containing liquid nitrogen is generally used in addition.

The sample is placed in an NMR tube and dropped into the magnet on a cushion of compressed air. The strong magnetic field causes the nuclei to become aligned or opposed to it.

The pulse program parameters are sent to the console. From here the required RF pulses are transferred to the probe. The probe is positioned in the middle of the magnetic field and contains the RF coils that go around the sample. It transmits the required RF pulses to excite the sample. It also acquires the weak FID from the nuclei in the excited sample. This is amplified and converted from an analog to a digital signal before the data are sent to the computer. Here the data can be manipulated (Fourier-transformed, phased, etc).

10.3.1.2 Experimental Techniques

Preanalytical Factors

To minimize artifacts and bias due to variation in sample collection, handling, and storage, standard operating procedures should be followed [90,123]. Care should be taken to avoid prolonged venostasis or contamination with ethanol or other substances used to clean the skin. For patients on IV drips or total parenteral nutrition, the arm with the drip in it should be avoided [19]. For toxicological metabolomics studies, consideration should be given to when the drug was administered. Fasting samples are generally recommended, as dietary intake can cause significant variation in metabolites. Time of day of sample collection should be consistent, as many metabolites, including cortisol, are subject to diurnal variation [19,123]). Where possible, multiple measures from the same patient should be used to give an average baseline.

Tube types should also be standardized. A comparison study [123] showed that the differences in NMR metabolite profile of serum compared to plasma (collected in heparin tubes) was minimal. Plasma samples collected in EDTA tubes results in huge contaminating peaks in ^{1}H NMR spectra, corresponding to free EDTA, magnesium EDTA, and calcium EDTA complexes. These peaks can potentially obscure the resonance signals of other metabolites. However, the influence of EDTA contamination was investigated and it was found that metabolites with resonances at these chemical shifts have other chemical shifts in areas not affected by EDTA and so they can still be identified and quantified [124].

For urine samples, 24 h timed urine collections are better as outputs can be normalized to time. The 24 h urine volume should be recorded and assessed for complete collection (a low urine volume may suggest an incomplete collection).

The time before centrifugation (serum-clot contact time) should be standardized. It has been found that this was the preanalytical factor, after interindividual variation, that caused the most variation in results [123]. Clotting on ice reduces this effect and is recommended [123]. There must be enough time (approximately 30 min) to allow the sample to fully clot but not so much that changes accumulate due to continued metabolism of glucose.

As much information as possible should be recorded with the sample: time of collection, time of centrifugation, temperature, time before freezing, and so forth [90,123]. Freeze–thaw cycles should be minimized by aliquoting the sample into multiple tubes. One freeze–thaw cycle has only a small affect, but multiple freeze–thaw cycles can significantly affect the metabolite profile [90,123].

Sample Preparation

For serum samples, proteins and lipoproteins can interfere with metabolite quantification; they can be removed by

protein precipitation, solid-phase extraction, liquid-liquid extraction, or spectral filtering [118,125]. If not removed, they can result in variation in analytes and in analyte loss [118]. The alternative, spectral subtraction, where the sample background is subtracted, can also cause signal loss and distort the spectrum [118].

Deuterium oxide (D_2O) for locking the signal, pH buffer to minimize shifts due to pH, imidazole as a pH indicator, and azide as a biological and internal standard for referencing the chemical shift and as a quantification standard [70] are often added to each sample prior to analysis. A deuterated solvent is generally used in NMR. The magnetic field can then be locked to the ^2H in the deuterated solvent so that any changes in chemical shift of the metabolites are due to the environment of the protons within the molecule and not simply due to changes in magnetic field strength [121].

While large numbers of samples are waiting to be analyzed by NMR, the temperature should be controlled, as differences in the amount of time at room temperature can result in systematic bias [102,123]. Temperature-controlled sample changers are available for NMR.

10.3.1.3 Common 1D NMR Pulse Programs

One of the most commonly used pulses is the 1D-^1H-Nuclear Overhauser Effect Spectroscopy (NOESY) pulse. This is because it is a simple sequence, with few optimization requirements or hardware restrictions and good-quality water suppression is generally easily obtained [126].

The Carr-Purcell-Meiboom-Gill (CPMG) pulse program is used to identify small molecules in the presence of large proteins and lipoproteins [127]. An initial 90° pulse is applied. During the first delay, the spins begin to dephase (the nuclear spins get out of synch over only a couple of seconds). This is due to spin−spin relaxation (T2) caused by the interaction of spinning nuclei with other spinning nuclei. Their spins are different due to differences in the Larmor frequency, differences in the nuclear environment, and tiny inhomogeneities in the magnetic field. A 180° pulse is applied to rephase the spins (fast ones to the back and slow ones to the front). The FID is then acquired. This allows a longer delay between initial pulse and acquisition. Large molecules such as proteins will have stopped spinning by the time the acquisition phase is begun, so only the FIDs of the small molecules, which can spin for longer, are recorded. This pulse program, a form of spectral filtering, is very useful for serum samples, which are full of lipids and lipoproteins.

10.3.1.4 Quantitative NMR

Although NMR is described as a quantative method, errors in quantitation can be caused inadvertently if careful consideration is not given to the acquisition and processing of the samples.

Preanalytical considerations are as above (e.g., freeze−thaw). For absolute quantitation (quantitative NMR, or qNMR) a signal:noise (S:N) ratio of >100:1 is required, whereas only 10:1 is required for relative quantitiation [118]. The number of scans can be increased to increase the S:N. The increase is proportional to the square root of the number of scans added (4 scans = 2*; 16 more scans = 4*; 64 more scans = 8*); therefore, only small improvements are achieved with large numbers of extra scans [119]. A compromise must be made between analysis time and S:N achieved.

Another essential consideration for qNMR is spin-lattice relaxation time (T1). This is caused by the interaction of the nuclei with the surroundings (the lattice). It determines the time taken for the net magnetization (vector sum of the magnetic moments of the nuclei) to return to equilibrium (completely relaxed). Different protons in different metabolites will have different T1s, as they have different rates of relaxation due to differences in the nuclear environment [90]. The temperature and the solvent, because of their effects on viscosity, also affect T1. Carbon T1s are typically much longer than proton T1s. T1s can be measured by the inversion recovery method [128]. After one T1, approximately 63% of nuclei have returned to equilibrium. After five T1s, >99% have returned to equilibrium. For quantitative NMR, the pulse delay (time between pulses) must be at least five times T1 to ensure that the relative peak areas are not underestimated [118].

If this is not possible, incomplete relaxation can be corrected for, providing the T1 times have been measured in the same sample matrix [118].

10.3.1.5 Two-Dimensional NMR

Two-dimensional NMR (2D NMR) is a technique used to increase the resolution of NMR. The data are plotted against two frequency domains: the direct domain (F1), based on FT of the direct time domain (t1), and the indirect domain (F2), based on the indirect time domain (t2) obtained by increasing the delay between multiple 1D spectra. There are many methods for performing 2D NMR such as Correlaction SpectrsocopY (COSY), TOtal Correlation SpectryoscopY (TOCSY), and Nuclear Overhauser Effect SpectroscopY (NOESY). These methods are very important in metabolomics because they help determine which peaks are connected, part of the same spin system on the same molecule, when multiple metabolites are present. It is useful when there is significant spectral overlap in 1D NMR, as it increases signal dispersion. It also allows structural determination of metabolites that are too complex to work out with 1D NMR alone [63,129].

To perform a basic 2D NMR experiment, the 1D spectrum must first be optimized. Hundreds of 1D spectra are recorded, and the delay in the pulse sequence is increased with each acquisition. Resulting spectra are stacked, with

the lowest t1 value at the bottom. FIDs of each acquisition are Fourier-transformed as normal to obtain the F2 (direct) frequency spectra. A new FID is formed across the stacks—in the indirect time domain (t2)—which is then Fourier-transformed to obtain the orthogonal F1 (indirect) spectrum. The indirect time domain is a result of the FIDs being sampled at discrete time intervals. F1 and F2 may be plotted against each other to observe cross-peaks [129], which are formed when the frequency of two nuclei overlap. The intensity of each cross-peak is shown on the z-axis in a topographic (contour) plot. It is determined by the efficiency of the transfer of magnetization from an excited nucleus to another, nearby nucleus. The full 2D spectrum represents all the interactions resulting from magnetization transfer: if there are a large number of paired nuclei, there will be a correspondingly large number of cross-peaks. These can then be assigned to specific positions on the metabolite [121]. Quantitation of 2D-NMR spectra is challenging [130].

A typical 2D NMR experiment consists of four stages: preparation, evolution (t1), mixing, and detection. In the preparation phase, a pulse is used to excite the nuclei. An evolution period (t1) is the delay; this is increased by a fixed amount for each pulse. During the mixing time the magnetization is transferred from the excited nuclei to other coupled nuclei within the same spin system via J-coupling or Nuclear Overhauser Effects (NOE). In the acquisition phase, the FIDs produced by the precessing nuclei are detected [129].

There are many different types of 2D NMR experiment. These mainly differ in the pulses and delays used during the mixing time to transfer the magnetization. There are two types of 2D NMR: homonuclear, where magnetization is transferred from one nucleus to another nucleus of the same type (usually ^{1}H to ^{1}H), and heteronuclear, where magnetization is transferred from one nucleus to another nucleus of a different type (usually ^{1}H to ^{13}C) [121].

Two-dimensional COrrelation SpectroscopY (COSY) [121] is a method for performing structural analysis by tracing atomic relationships through closely coupled bonds. Magnetization is transferred via J-coupling. Coupling distances can be 2-bond (geminal), 3-bond (vicinal), or (rarely) 4- or 5-bond (long range).

2D NOESY allows structural determination via relationships between atoms through space. The z-magnetization is transferred via NOE from the excited nucleus to all other nuclei in the same spin system via multiple transfers. Cross-peaks are observed if a proton is less than 5 angstroms away from another proton in the same spin system [121].

Heteronuclear single-quantum coherence (HSQC) [131] correlates proton resonances with 13C or 15N resonances that are directly coupled. This method is improved by derivatized metabolites [118]. For example, ^{13}C-glucose, can be used to perform metabolic flux analysis to identify which metabolic pathways are using the metabolite [118].

J-resolved NMR is one of the simplest 2D NMR experiments [119]. It uses a double spin echo sequence to supress strong coupling artifacts [132], so the peak splitting caused by J-coupling is reduced. This allows spectral overlap to be minimized, making the spectra easier to interpret. The lack of correlation makes J-resolved NMR less useful for structural determination; however, if the metabolites at the chemical shifts of interest are known, this is not a problem. A 2D NMR spectrum is obtained and used to identify J-coupling. The split peaks are removed and the 2D NMR is projected back as a cleaned-up version of the original 1D NMR spectrum, with only singlets [119]. The spectra are optimized and aligned, as in 1D NMR. This method has been shown to aid the identification of small molecules that would otherwise be hidden by the overlap caused by J-coupling in typical 1D NOESY or CPMG pulse programs [119]. Signals from lipoproteins and proteins are also decreased. Metabolite quantification can be more challenging in J-resolved spectra due to alterations in line shape [119,133].

A recently developed method is ultrafast 2D NMR (UF-NMR). It virtually splits the sample into multiple slices and performs a different excitation pulse on each slice. This is termed spatial encoding. Acquisition is performed on the spatial gradients within the sample. It can therefore provide 2D data in a single scan, drastically improving analysis time. Ultrafast 2D NMR has been described in detail [129]. Its benefits include increased throughput or timesaving and minimization of spectrometer instabilities that decrease the reproducibility of typical 2D NMR spectra [130]. This method has been used to look at the metabolite changes (in terms of absolute quantitation) in breast cancer cell extracts [130].

2D NMR is difficult to interpret manually, especially for complex mixtures. 2D NMR spectra may be analyzed using a variety of softwar, for example, Amix (Bruker Biospin, Germany), MetaboMiner [134], or the CCPN Metabolomics Project [135].

10.3.1.6 Other Techniques

Statistical TOtal Correlation SpectroscopY (STOCSY) [136] is used to identify correlations of spectroscopic signals with other signals. The 1D spectra of many biological replicates are compared. Peaks from the same metabolite are statistically correlated—even across multiple spin systems. Therefore, if individual peaks increase and decrease by the same magnitude as other peaks in the same sample in each replicate, it can be inferred that they are from the same metabolite, or from the molecules regulated by the same metabolic pathway [63,68].

Statistical HeterospectroscopY (SHY) [137] is used to correlate peaks from MS and NMR (or any two or more independent spectropscopic methods). Again, multiple biological replicates are analyzed by MS and NMR. If the increase or decrease of a peak is statistically correlated across the two methods, it can be inferred that they are from

the same metabolite or from metabolites belonging to the same biochemical pathway [63,68].

NMR Hyphenation

Although one of the advantages of NMR is that sample preseparation is not required, separation can help in cases where a low-abundance metabolite in a sample otherwise full of high-abundance metabolites is of special interest. Generally, NMR is hyphenated to an LC system, LC-NMR. This is a technically demanding technique, as it is very difficult to shim the sample, which can result in poor spectra. Another method is simply collecting fractions from an LC using a fraction collector and running individual fractions on the NMR. Solid-phase extraction is often required before NMR spectroscopy of the fractions, as the ionic strength or pH can affect the results.

The samples can also be split so that one part is analyzed by MS and the other by NMR [63]. Depending on the MS method used, this allows determination of the exact mass, which can be used to deduce the molecular formula. Information from the retention time can also be used to putatively identify the compound. The NMR is then used to confirm the structure of the compound using the molecular formula and spectra obtained. This can be very important when cohort MS detects a metabolite that is consistently up- or downregulated in cases versus controls, but cannot be identified confidently with MS alone.

Water Suppression

The majority of samples and solvents used in metabolomics contain a large amount of water. For example, serum and urine are approximately 95% water [125]. The water signal must be suppressed or it will swamp all the other signals and can cause radiation damping [63]. There are a number of water suppression methods available that allow higher receiver gains to be used, resulting in greater sensitivity [118]. Presaturation of the water signal, the most common, is generally achieved using a long low-powered pulse at the frequency of the solvent [118]. Water suppression Enhanced by T1 (WET) and WATER suppression by GrAdient-Tailored Excitation (WATERGATE) and excitation sculpting use gradient pulses to suppress water [118,138].

Care must be taken in the quantitation of peaks close to solvent peaks, as their resonances may have been suppressed along with the solvent suppression or there may be baseline distortions resulting from the suppression [118].

10.3.1.7 Optimization of Detection

Acquisition time should be optimized to be short enough to prevent excess noise and long enough that the FID is not truncated [118]. The digital resolution (number of data points per Hz) should be high enough that there are at least five data points for each peak of interest [118]. Zero filling (adding zeros to the end of the FID) can be used to increase the number of data points without increasing noise [118,121]. This can allow fine coupling to be detected but can result in artifacts in the spectra. The baseline should be flat, corrected to zero, and the spectra must be properly phased in order to obtain accurate integration [118,121]. Alignment of the spectra may also be required to compensate for drift due to matrix effects [90].

10.3.1.8 Limitations of NMR

In comparison to mass spectrometry, NMR is less sensitive. The amount of pure material required in order to obtain a decent S:N ratio depends on the type of probe and the magnetic field strength. The probes used in metabolomics are generally cryoprobes, where the detector and preamplifier are cooled to approximately 20 K [71]. Cryoprobes have up to four times the S:N compared to room-temperature probes because the detection sensitivity of the RF coil is increased and thermal noise in the electronics is reduced [63,71,118]. Microcoil probes can also be used with small-diameter NMR tubes to increase sensitivity and are especially good for limited sample amounts [63,71,118]. The pulse is repeated multiple times on the sample in the magnet to increase the S:N ratio. For good peaks that can be used for quantitation, the sample must be carefully shimmed. A series of electromagnets are used to adjust the magnetic field in order to maximize the homogeneity of the magnetic field. This is essential for obtaining sharp peaks. If the magnetic field is poorly shimmed, the chemical shift of the nuclei will vary across the sample due to local changes in the magnetic field, resulting in broad peaks that could overlap with other peaks and be difficult to identify or quantify [121].

10.3.1.9 Spectral Overlap

Proton NMR is limited by the small chemical shift scale (0–10 ppm for the majority of compounds). 1D 1H NMR of mixtures can be extremely difficult. Each metabolite will have multiple signals for each proton in a different environment. Multiple signals are also obtained for stereoisomers of slowly interconverting conformations. Due the multiplicity of signals from many metabolites, NMR of mixtures can result in significant signal overlap. To overcome this problem, 2D methods can be used, but they take much longer to perform (hours compared to minutes for 1D NMR) because of the number of scans required. J-resolved NMR has the benefit of minimizing spectral overlap in a much shorter time than other 2D NMR methods [119].

10.3.1.10 Matrix Effects

Although NMR does not suffer from the effects of ion enhancement or suppression, as in MS, differences between sample matrices will still affect the spectra. Subtle differences in pH, ionic strength and protein content cause differences in peak position and line width. Each metabolite is

differentially sensitive to these effects. In many compounds, NMR resonances of each nuclei are affected independently—making global correlation extremely difficult.

Such effects can be minimized by careful consideration of sample preparation. For example, mixing the sample with a good buffer will minimize differences in pH between samples, although for urine samples that have a wide pH range, shifts can still be seen even after buffering [63]. Note that the ionic strength of the buffer must not be so great as to adversely affect tuning and matching of the probe [63]. Also, a known concentration of imidazole can be added to act as a pH reference for the sample, as its ppm can be used to determine pH [139]. The NMR should also be regularly temperature-calibrated, for example, with deuterated methanol, to eliminate skew caused by temperature effects [140].

10.3.2 Mass Spectrometry

Mass spectrometry is an analytical technique that essentially "weighs" each molecule in a sample and plots the detected masses on a graph (Figure 10.4).

Mass spectrometers commonly used in the life sciences range from low-resolution quadrupole mass filter [141] and quadrupolar ion trap [141] instruments to high-resolution time-of-flight [142,143], orbitrap [144], and ion cyclotron resonance [145] instruments. High-resolution instruments are also those with the highest mass accuracy, but both high- and low-resolution instruments are commonly used in metabolomics applications. High resolution allows us to distinguish between, for example, tyrosine ($C_9H_{11}NO_3$, with a protonated monoisotopic mass of 182.081) and methionine sulphone ($C_5H_{11}NO_4S$, with a monoisotopic mass of 182.048) because they appear as two separate peaks (Figure 10.5). On a low-resolution quadrupole instrument, these peaks overlap and appear to be a single peak. High mass accuracy, on the other hand, provides the capability to calculate the empirical formula of a compound, based on its mass defect—that is, its deviation from unit mass resulting from the presence and number of heteroatoms in the structure, often supplemented by the intensity ratio of detected isotopes [146,147]. While this is not infallible, especially when including a wider variety of possible atoms than C, H,

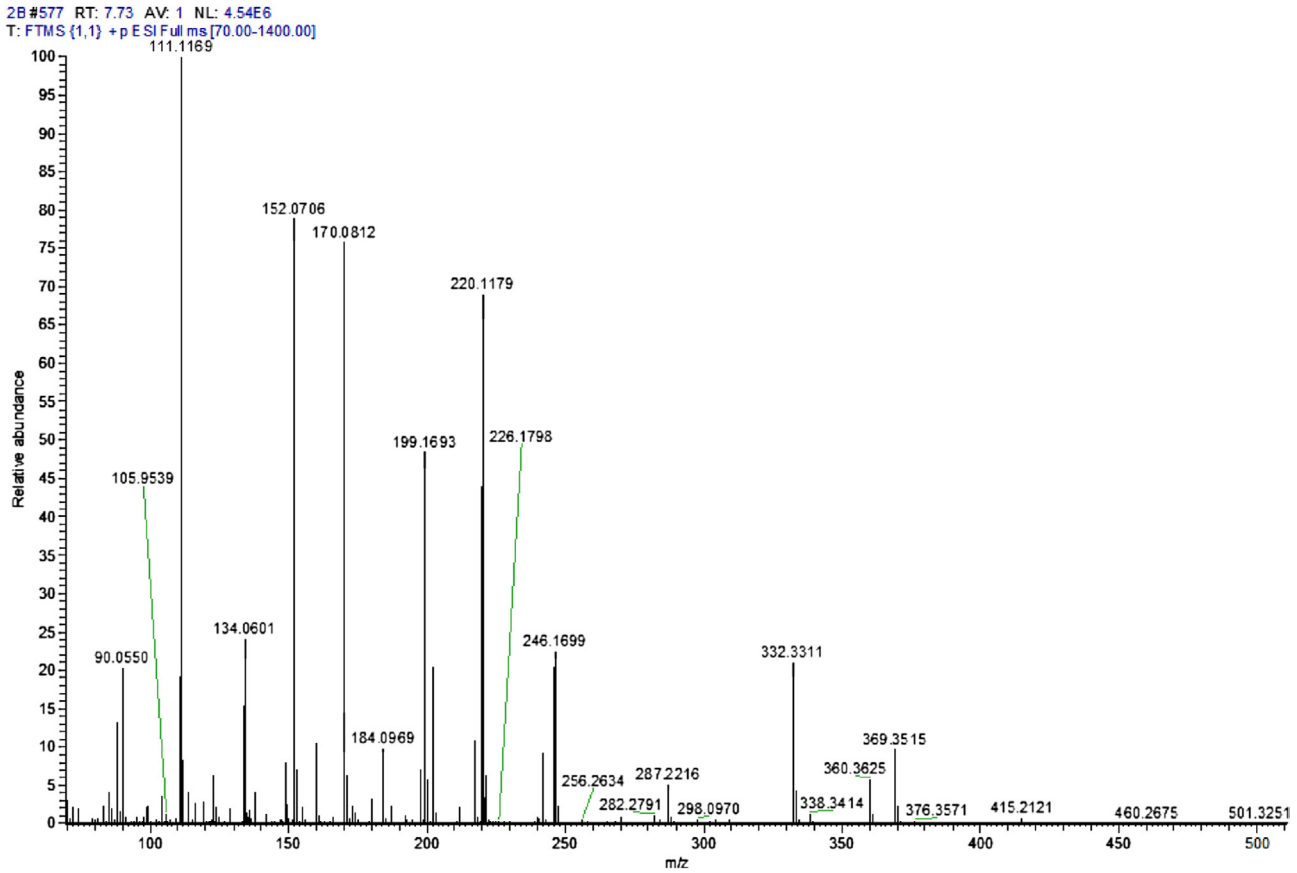

FIGURE 10.4 Typical mass spectrum from a metabolomics experiment. Each fine vertical peak is a detected metabolite. Mass:charge ratio is displayed along the x-axis, and relative abundance is plotted on the y-axis. Peaks are labeled with their constituent masses. The spectrum, despite being a single data point from a section of an LCMS analysis, is very complex.

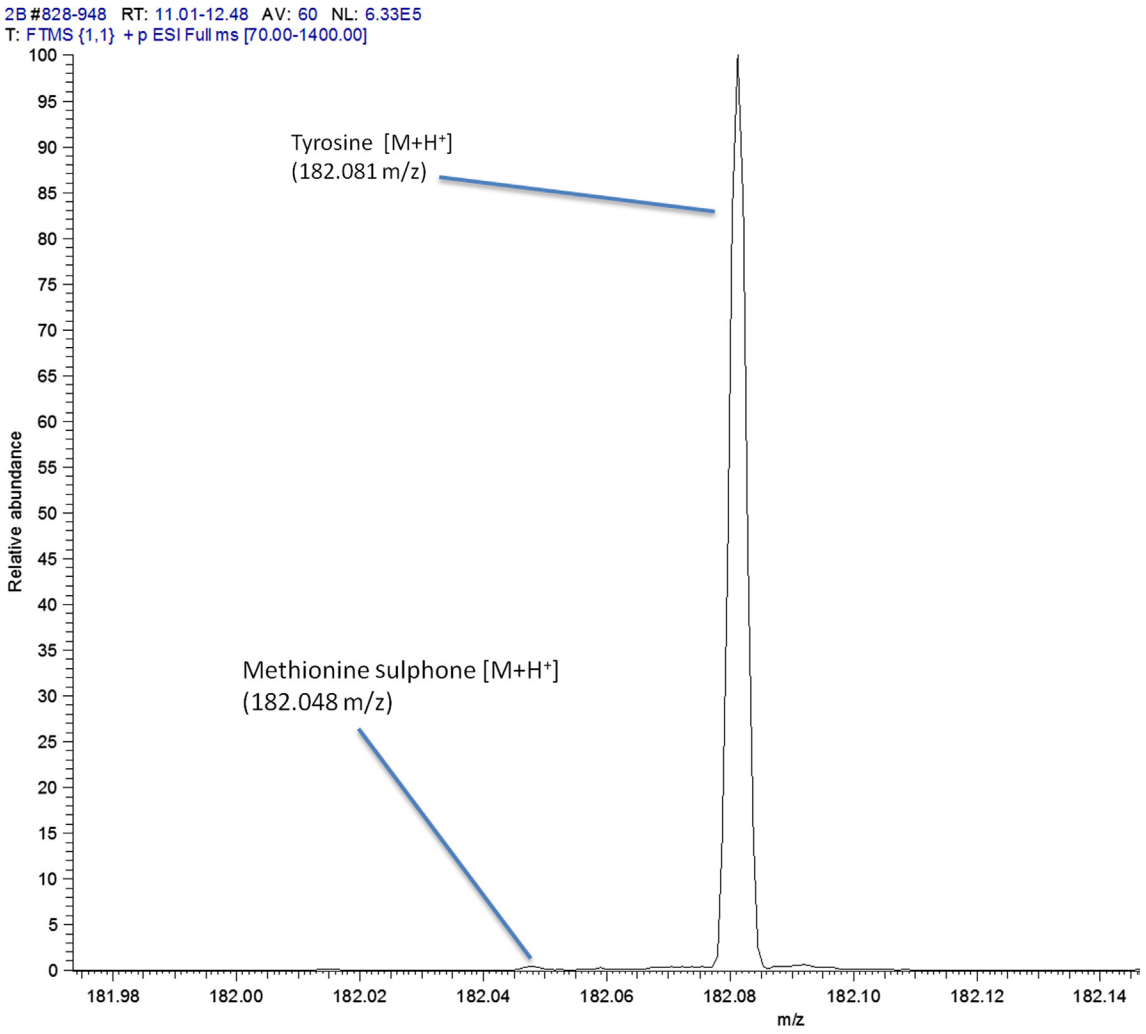

FIGURE 10.5 **Compounds with similar mass.** Tyrosine ($C_9H_{11}NO_3$, with a protonated monoisotopic mass of 182.081) and methionine sulphone ($C_5H_{11}NO_4S$, with a monoisotopic mass of 182.048). While methionine sulphone is very low intensity compared to tyrosine, it is still a distinct peak. A resolving power of 50,000 is amply sufficient to distinguish these compounds. Unit mass resolution quadrupole instruments would display a single peak in this instance.

O, and N [148], it is a powerful method for narrowing down the search space of possible metabolite identifications. Due to their high resolution and mass accuracy, these instruments are normally used in an untargeted manner, where an effort is made to detect, annotate, and quantify every available peak in a sample.

Unit mass resolution instruments, such as quadrupole-based mass spectrometers, often apply fragmentation to generate orthogonal information that can help them approach the diagnostic capabilities of accurate mass instruments. Of course, these fragmentation techniques are also available for high-resolution instruments for confirmation of identification. In the most common use of this type of experiment, a multiple reaction monitoring (MRM) [149,150] experiment is carried out on a triple-quadrupole instrument [151].

MRMs are based on the application of quadrupolar mass filters to exclude any ion that does not possess the appropriate unit mass. Once an ion of particular mass passes the first quadrupole, it is fragmented in a second quadrupole before the resulting fragment ions are filtered in a third quadrupole (thus, the "triple-quad" instrument type). Only ions passing through both filters strike the detector and are analyzed (see Figure 10.6). For this reason, MRM experiments are extremely sensitive and, if unique fragment ions for a particular structure can be obtained, can also be very specific. The significant drawback to this technique is that it can only be used in a targeted manner: if the experiment does not

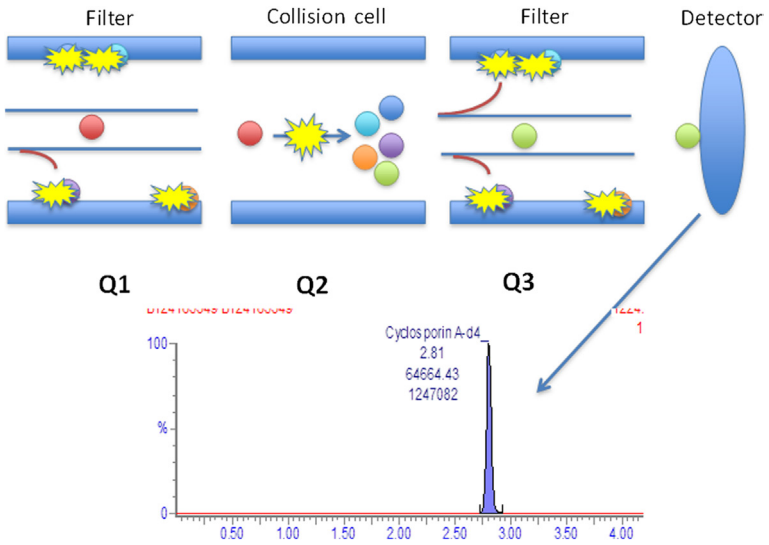

FIGURE 10.6 **Multiple reaction monitoring experiment using a triple-quadrupole mass spectrometer.** It was used for selective and sensitive detection of specific compounds. In this instance, the intact mass of cyclosporin is selected for quadrupole 1 (Q1), then fragmented in quadrupole 2 (Q2). The resulting ions are filtered to exclude nondiagnostic fragments in quadrupole 3 (Q3). Diagnostic fragments are allowed through the filter to strike the detector, producing a peak at the appropriate retention time in the chromatogram.

apply a particular MRM designed for a specific molecule, the molecule will not be detected.

10.3.3 Liquid Chromatography–Mass Spectrometry

Two significant drawbacks to any mass spectrometry method are isomers and ion suppression. Isomers will be discussed in more detail in Section 10.3.3.2; they are compounds with identical chemical formulas (and therefore mass) but different structures. Ion suppression (whose implications in clinical analysis are reviewed in reference [152]) occurs due to a variety of factors, predominantly modification of the solvation characteristics of the sample by an eluting compound. For this reason, it is considerably advantageous to apply a separation upstream of the mass spectrometer and usually directly coupled to it. Liquid and gas chromatography are the most commonly used methods for separating compounds prior to their introduction to the mass spectrometer. Both are based on the transient interactions of analytes with a surface (the stationary phase) while carried by a mobile phase (either a liquid or a gas). Analytes that interact strongly with the stationary phase will elute more slowly than analytes that interact weakly or not at all.

Reversed phase chromatography is the most common separation technique for LC-MS metabolomics [80,153]. It separates on the basis of hydrophobicity, so nonpolar compounds are eluted more slowly than polar compounds. Newer, but gaining more acceptance, are HILIC separations, which separate compounds on the basis of hydrophilicity and are more selective for polar compounds [72,73,154].

10.3.3.1 Sample Preparation for Liquid Chromatography–Mass Spectrometry

A significant consideration with any analytical technique is the requirement for sample preparation. Mass spectrometers are sensitive to contamination and interfering compounds, and liquid chromatography can be fouled by the use of incorrect buffers. There are, essentially, three methods for the extraction of metabolite samples: extraction of the metabolites followed by introduction of the extract to the chromatography system; extraction of the metabolites followed by lyophilization of the extract, resuspension in a suitable solvent, and introduction to the chromatography system; and cell lysis: solid-phase extraction to enrich for a particular class of compound and then processing in a manner similar to the other two methods. Exhaustive studies have been performed on optimal sample preparation methods for different sample types [155–157].

10.3.3.2 Limitations of Liquid Chromatography-Mass Spectrometry

Any analytical technique has limitations. In metabolomics, complete coverage of the metabolome is as yet not achievable. In addition to this global limitation, each platform for metabolomic analysis possesses a number of drawbacks. For critical evaluation of data, it is important to consider these drawbacks in data interpretation.

Dynamic range is a major issue in metabolomics in that the best mass spectrometers currently have a dynamic range of approximately four orders of magnitude. Given that the most abundant metabolites in a sample have millimolar concentrations (e.g., glucose has 3.5–9 mM), the minimal abundances of detected metabolites will run to the tens of micromoles while the physiological abundances of some compounds, for example, oestradiol, are in picomolar concentrations. MS sensitivity also plays a large part in dynamic range, as we will see later. It may therefore be necessary to further extract low-abundance compounds from the greater metabolome to allow detection of some classes of metabolites. Steroid hormones, for example, are generally at low levels, which makes them challenging to detect in a complex mixture.

Many compounds are labile, rapidly processed by enzymes, subject to oxidation by the atmosphere, or react chemically with other compounds in the metabolite mixture. Labile compounds such as ATP break down over relatively short periods, whereas others, like amino acids, may remain stable for months in solution. Some compounds are extremely rapidly turned over by enzymes [158]. Glycolysis is heavily optimized to rapidly process glucose to the point where at even one second post-lysis, glucose concentrations will have changed dramatically. It is therefore essential to inhibit enzymatic activity as quickly as possible without compromising the experiment, either by snap-freezing aliquots of sample or by the direct squirt method, where samples are simply pipetted, along with medium, into a cold extraction solvent. The latter obviously has the drawback that the signals from the medium will swamp the same metabolites in the cells.

While it is possible in some ways to perform extractions under a nonoxidizing atmosphere, LC-MS instrumentation is not commonly kept in an inert atmosphere and consequently the samples will still oxidize while they are waiting to be analyzed and even during separation. Glutathione and reduced glutathione are particularly prone to oxidation. Derivatization of these compounds to prevent oxidation is possible (e.g., reference [159]), but cannot be applied on a global compound basis.

Heterogeneity of metabolites was mentioned previously, but it is a significant issue for metabolomic analysis. Simply, metabolites are chemically highly diverse, so not only is there no single separation system capable of effectively retaining every metabolite; there is no mass spectrometry ionization mode capable of detecting every metabolite. Electrospray ionization [160] is soft and therefore produces molecular ions of many species, but compounds such as benzene or naphthalene will only ionize via LC-MS using atmospheric-pressure chemical ionization or photoionization [161], which is very limited for higher-molecular-weight metabolites.

Isomers are arguably the most significant problem in metabolomics. Isomers are compounds with identical empirical chemical formulas but different structures: examples are leucine and isoleucine and the hexose sugars, including glucose, fructose, and mannose. While biological samples are often known to contain only a handful of potential isomers, public chemical databases normally contain many more. These are unlikely to be present in biological samples, but they cannot normally be ruled out. There are several effective ways to distinguish between chemical isomers, but each has its limitations and there is no single effective method for distinguishing between all isomers. The most commonly applied mechanism is chromatography.

Specific chromatographic methods exist to distinguish between hexose and pentose sugars [162], sugar phosphates [163], D- and L-amino acids [164], and a plethora of other compounds. These methods are, unfortunately, specific to that group of compounds, and general methods for separating metabolites (most commonly based on reverse-phase and HILIC chromatography) are unable to separate any but the most chemically different compounds. Currently there is a growing interest in ion mobility spectrometry in the metabolomics community [165]. Ion mobility spectrometry (IMS) (see review in reference [166]) is a gas-phase method for separation of ions by charge and cross-sectional area. It is very rapid (usable in chromatographic time as an additional dimension of separation) and may be used, given a suitably high-resolution drift cell, to distinguish isomers.

The other common method for distinguishing isomers is fragmentation. Several mass spectrometry methods are available for isolating a molecular ion and breaking it into component fragments. The most common is collisionally induced dissociation (CID) or collisionally activated dissociation (CAD), which is a technique that applies thermal decomposition to a molecule. CID is usually deployed in the form of a quadrupolar collision cell or in an ion trap instrument. The former is generally a harder fragmentation than the latter, producing more small fragments. CID in an ion trap has the unique property that it can be performed in rounds, such that a precursor ion can be selected and then fragmented to produce a spectrum of fragment ions, followed by selection of a fragment that can then be fragmented, and so on. Needless to say, MS_3 (i.e., a survey scan, followed by two rounds of fragmentation) or higher provides clearer structural determination than a single round of fragmentation [120,167], but it takes longer and requires more samples. In some cases the use of MS_3 or greater is still unable to distinguish some isomers.

Sensitivity and detection limits in mass spectrometry are molecule-dependent. Certain groups, such as amines or carboxylic acid groups, are easy to ionize in a suitable buffer. Other compounds, such as uncharged molecules, aromatic species, and aldehydes, are challenging to ionize under normal conditions. Overall ionizability is determined by the totality of interactions between functional groups in a molecule. For a given amount of a compound, the signal intensity in the mass spectrometer is dependent on its

chemical structure [168]. This means that the intensity of a peak for one compound is not comparable to the intensity of a peak for another compound. Relative quantitation may be performed by comparing the intensity of a compound in one sample with the same compound in another sample, assuming that matrix effects (such as ion suppression) do not compromise the signal of either peak. For absolute quantitation, it is necessary to provide a compound with identical physicochemical properties but different mass (i.e., an isotopically labeled compound) as an internal mass standard [169].

The chromatograms produced by a typical LC-MS run are highly complex. Metabolite samples, as previously described, contain hundreds to thousands of individual metabolites. This complexity is enormously enhanced by the presence of adducts and fragments [170]. Electrospray ionization, the most common method for providing a charge to a compound in biomolecular mass spectrometry, generally provides a charge by adding a proton (for positive ions) or removing one (for negative ions). Thus the proton is the simplest and most prevalent adduct in mass spectrometry. The presence of salt in the sample, however, can provide other cations to function as charge donors. Na^+, NH_4^+, and K^+ are common adducts in blood samples analyzed on systems using high-pH ammonium buffers. Electrospray ionization is relatively nondestructive in comparison to competing ionization technologies, such as electron impact ionization or atmospheric-pressure chemical ionization (APCI).

Despite this, it is very common for labile compounds to decompose during ionization. The most common losses include H_2O, PO_4^{2-} (especially from triphosphate compounds such as ATP), CO_2, and NH_3. Compounds can also dimerize in the ion source, leading to a compound double the size of the original. Additionally, it is possible for compounds to obtain multiple charges by picking up two or more protons or other cations. Spectra are further complicated by multiple adducts combining two or more of the previously described events.

An alternative methodology for metabolomics is GC-MS, which is very complementary to LC-MS and NMR platforms.

10.3.4 Gas Chromatography–Mass Spectrometry

The popularity of GC-MS is due to its high chromatographic resolution, reproducibility, high throughput, and low equipment cost. Figure 10.7 shows the separation of metabolites found in orange juice, with the most abundant peak in the chromatogram matching with the library spectrum for sucrose. The most popular method for ionization of the analytes eluting from the GC column is electron impact (EI). This is a high-energy process, resulting in the majority of molecular species fragmenting into product ions. The measured EI fragment spectrum detected by the mass spectrometer is, however, highly conserved over a wide variety of analyzers and manufacturers. The apparent lack of an identifying molecular ion is made up for by a fingerprint fragment pattern that can be matched to spectral libraries.

The most comprehensive and widely used spectral libraries are those provided by the NIST database, which contains spectra for thousands of pesticides and environmental pollutants, drugs of abuse, explosives, and clinical diagnostic molecules, among many, many others [171]. The main spectral library in the current release contains spectra from over 200,000 compounds, each having been verified by expert mass spectrometrists. This makes it possible to use GC-MS to make confident molecular identifications even in the absence of a pure standard. The Golm

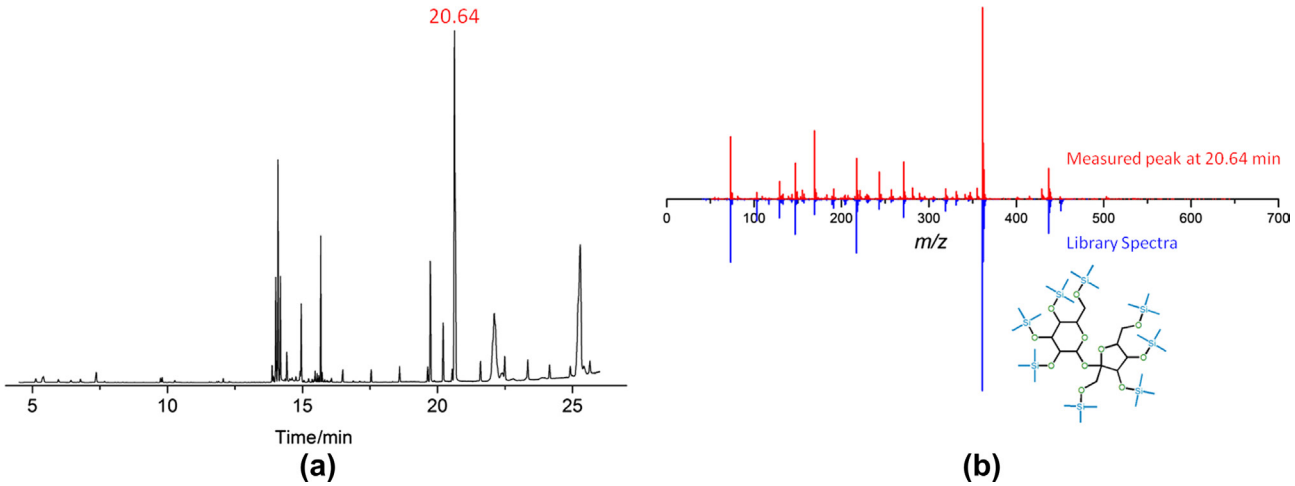

FIGURE 10.7 Gas chromatography–mass spectrometry. (a) GC-MS TIC for methanol extraction of orange juice derivitized with MSTFA and separated using a TraceGold TG-5MS (30 m × 0.25 mm ID, 0.25 µm film thickness; Thermo Scientific) in a Trace Ultra GC (Thermo Scientific) coupled to a ITQ 900 MS (Thermo Scientific). (b) EI mass spectrum at 20.64 min matching to the NIST library spectrum of octakis(trimethylsilyl) ether of sucrose.

Metabolome Database (GMD) is a GC-MS spectral library containing biologically active metabolites created by the scientific community [172]. Once a peak has been identified, a targeted method using single-ion monitoring (SIM) or multiple-reaction monitoring (MRM) can be used to increase sensitivity to achieve lower limits of detection (LoD) resulting in lower and more reproducible limits of quantification (LoQ). The addition of an internal standard, usually a deuterated labeled compound of interest, is used as a reference for quantitation [57,173].

10.3.4.1 Limitations of Gas Chromatography–Mass Spectrometry

To be compatible with GC-MS analysis, samples must be dissolved in a volatile solvent. The component molecules need to be volatile and thermally stable to be separated and detected. Many biomolecules contain polar functional groups such as alcohol (–OH), amide (–CONHR), amino (–NH2), carboxyl (–COOH), thiol (–SH), and enolizable ketones. Chemical derivitization results in the substitution of the acidic hydrogen, leading to a decrease in intermolecular interactions [174,175]. The derivitized compounds have an increased volatility, allowing samples that would decompose at high temperatures to be suitable for GC separation at moderate temperatures, and elute within the temperature limits of the column. The reduction in polar interactions with the column also results in better peak symmetry and resolution. Three classes of derivitization agents are silylation, acylation, and alkylation. The selection derivitization agent will depend on the compounds in the sample to be analyzed. This selection can be well known from the literature for a targeted compound, but the correct reagent may not be obvious for an unknown mixture when performing a global qualitative metabolomic analysis.

Silylation is the most commonly used derivitization reagent because these chemicals will react with both –OH (alcohols, carboxylic acids) and –NH (aminos and amines) [176], making them useful for mixtures of compounds with a biological origin. This also requires that the samples be dry in order to be free of water that would also react with the derivitization agent. Metabolite extraction and subsequent purification and derivitization can be very time-consuming. For urinary estrogens, this can involve days of sample preparation using multiple solid-phase extraction cartridges, enzymatic hydrolysis, and size exclusion chromatography as well as chemical derivitization [177]. In such cases a simpler LC-MS methodology may well be desirable.

The fragmentation pattern resulting from EI ionization can make identification of components simple if the spectrum matches a compound in the spectral library. However, this can be far from trivial if there are no high-quality matches or when compounds co-elute, resulting in a mix of fragments from both compounds. Co-elution of compounds can be overcome if there is a spectrum of a pure standard to identify which unique fragment ions are produced from the molecular ion. These unique ions can then be selectively filtered using SIM or MRM to detect only ions that confirm the identity of the eluting compound. Without a pure standard as a reference, identification of unknown compounds can be difficult, especially when there is no certainty that the molecular ion has been detected.

An experienced mass spectrometrist can interpret common fragments and use predictive tools to aid in identifying the molecule; however, this is very time-consuming and not feasible for global metabolite profiling, which can involve hundreds or thousands of unknown compounds. Chemical Ionization (CI) is a slightly less energetic process that leads to fewer, larger fragments and may allow the molecular ion to be observed, but this is not guaranteed. Some compounds, especially those with halogens or that have been derivitized with fluorine reagents, can be selectively detected using negative (−ve) CI, simplifying a complex mixture for simpler interpretation [178–180].

10.4 ANALYSIS OF METABOLOMIC DATA

10.4.1 NMR-Based Metabolomics Data

In 1D NMR, the Fourier-transformed data are represented as a spectra with chemical shift (δ) along the x-axis and intensity along the y-axis. The chemical shift is generally depicted as parts per million (ppm) rather than hertz (Hz), as this is independent of magnetic field strength.

An internal standard (IS)—tetramethylsilane (TMS), Trimethylsilylpropionic acid (TSP), 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), or their deuterated analogs— is used as a reference for 0 ppm and for determining the signal produced per proton. These molecules are chemically inert, stable, and soluble; have multiple equivalent nuclei resulting in a single sharp peak; are easily removed from the sample after measurement; and have resonances upfield of most organic compounds [181]. Ideally the IS should not evaporate at the temperatures used for analysis and should not be affected by pH or temperature.

The number of peaks present for a pure metabolite depends on the number of different environments the metabolites' nuclei are in. Because the spectra are dependent on the chemical environment of the nuclei within the molecule, NMR is a powerful method for structural determination. The chemical environment includes the shielding effect of neighboring atoms, the saturation of the molecule, and the number of neighbors [121]. The main determinant of the chemical shift is how shielded the proton is from the magnetic field. Electrons have spin; their magnetic vector opposes the overall magnetic field, shielding the nucleus. Therefore, if the electron density of a nucleus is high, the shielding effect will be higher. Electronegative atoms (such as oxygen, nitrogen,

and chlorine) deshield the nuclei, resulting in a shift to the left (an increase in ppm, known as downfield). If there is more than one electronegative atom neighboring the nucleus, the deshielding effect is additive [121].

Another affect of NMR is spin-spin splitting, also known as J-coupling, which can be exploited in structure determination. If you have two protons, with different chemical environments, on adjacent carbons, their different spin states can slightly add to or slightly subtract from the overall magnetic field experienced by the other proton nucleus. If one proton is spin-up (aligned with the magnetic field), it will slightly add to the magnetic field strength experienced by the neighboring proton.

Conversely, if it is spin-down (opposed to the magnetic field), it will slightly cancel out the magnetic field strength experienced by the neighboring proton. The neighboring proton will have its peak split into two equal peaks, one at a slightly higher ppm (downfield—due to increased local magnetic field) and one at a slightly lower ppm (upfield) than the peak would be if it hadn't been subjected to these effects. The splitting pattern follows the n + 1 rule: if a proton has one neighboring proton with a different chemical environment, it will be split in two, as just described. If it has two neighboring protons with a different environment, it will be split into a triplet with a 1:2:1 ratio. This is because 25% of the time both of these neighbors will spin up, adding to the local magnetic field; 25% of the time both protons will spin down, subtracting from the local magnetic field; and 50% of the time one will spin up and one will spin down, canceling out the effect on the local magnetic field.

Similarly, if there are three coupled protons, a quartet is observed (1:3:3:1); for four coupled protons a quintet is observed (1:4:6:4:1). The pattern follows Pascal's triangle. The area of all the split peaks added together will equal the original singlet, so quantitation is still possible. The effect is generally transmitted only through three bonds or fewer. However, coupling can be observed over more bonds if the molecule has a very rigid structure. This can therefore be used to determine the number of neighbors, helping to determine the structure [182]. The peaks are split by J Hz, where J is the coupling constant. This is independent of magnetic field, as it is caused by the magnetic field of another nucleus, not the magnet itself [121].

Chemical shift, intensity, and coupling patterns are used together to identify the metabolites in the sample. If it is suspected that a peak belongs to a certain metabolite, the other peaks known to be present for that metabolite must be observed in the spectrum, unless that part of the spectrum is overcrowded. Metabolite identification is difficult in mixtures. Note that many NMR chemical shifts are dependent on solvent, sample concentration, and other acquisition parameters (see section 10.3.3 on limitations). Metabolite identification is difficult or impossible for the wide variety of minor and trace metabolites, especially organism-specific metabolites. However, for common and abundant metabolites, automatic identification and quantification are possible using the appropriate software.

Quantitation is based on integration of the peaks and comparison with the reference peak intensity (internal standard of known concentration). The number of nuclei that contribute to the signal is also crucial [118]. For example, the ratio of a methylene (CH2) group to a methyl group (CH3) should be 2:3 for a single metabolite. Depending on whether the methyl shift or the methylene shift is used for quantitation (depending on spectral resolution), a different factor will be required.

$$\text{amount of x} = ((\text{area of x} / \text{area of IS}) \times (\text{nuclei in IS} / \text{nuclei in x})) \times \text{conc of IS}$$

The amount of x can be in grams or moles depending on the amount used for the IS. The areas are integrated using NMR software. The number of nuclei for IS depends on the structure (12 for TMS, 9 for TMS and DSS). The number of nuclei for the metabolite of interest depends on the structure and formula of the metabolite and the number of protons contributing to the signal at that specific chemical shift.

10.4.1.1 Spectral Deconvolution and Peak Picking

One method of dealing with overlapping spectra is deconvolution, which mathematically fits a Lorentzian or Gaussian peak shape to peaks suffering from partial overlap. Software that makes use of this method includes MestReNova (Mestrelab Research, Santiago de Compostela, Spain) and ACD/Labs Spectrus processor (ACD/Labs, Toronto, Cananda) [118]. Chenomx (Alberta, Canada) and Amix (Bruker Biospin, Bremen, Germany) both offer compound identification and quantification with reference to spectral libraries [70].

10.4.1.2 Binning or Bucketing

Binning, or bucketing, is a commonly used method for dealing with the crowded spectra obtained from 1D NMR of complex samples [70,118]. In its simplest form, this method involves splitting the spectrum into bins of a specific size (e.g., 0.02 ppm sections) and integrating the spectrum in the bin, regardless of whether it is from a well-defined peak or an area of overlap or is just baseline. This method cannot be used for absolute quantitation, but the integrals of the bins can be compared between samples using chemometric methods and useful information about where in the spectra differences between cases and controls (or similar) are located. Further analysis can then be performed to identify and quantify metabolites in these regions.

Care must be taken in the interpretation of binned data. As discussed previously, variations in ionic strength and pH between samples can cause the resonances of some peaks to be shifted. Slight alterations in shifts are acceptable as long as the new chemical shift stays in the same bucket; if it moves between buckets, a different bin will be increased and the original will be decreased, which can lead to

misinterpretation when comparing different samples [118]. This is becoming less of a problem as alignment algorithms improve [119].

A more robust method of binning is adaptive-intelligent binning. In this case the width of the bin is not kept constant, but is adapted to take account of pH- or ionic-strength-induced shifts. It uses local minima to identify the best location to split the bins [118].

10.4.1.3 Data Normalization

Urine samples can vary significantly in the concentration of solutes depending on fluid status or the individual. Many studies use the creatinine concentration in the NMR spectra to normalize the concentrations of other metabolites [70].

The area of the full spectrum can be summed and used to normalize the solutes. However, the abundance of metabolites across different samples is not constant. For example, a diabetic patient may have significant glucosuria, which makes the other metabolites falsely low using this method [118].

By calculating the most likely dilution factor in comparison to a reference spectrum, probabilistic quotient normalization (PQN) can also be used to scale spectra to the same virtual overall concentration in order to minimize changes in NMR spectra [118,183].

10.4.2 MS-Based Metabolomics Data

When using only two-dimensional datasets, such as when using direct infusion mass spectrometry, data analysis is usually performed by binning the resulting spectrum and comparing sets of spectra using multivariate statistics [184]. These sorts of studies are generally applied in a metabolic fingerprinting process, as the lack of separation of isomers and ion suppression make this method challenging for metabolic profiling and pool analysis.

LC and GC MS produces much more complex datasets with three dimensions to consider. Each eluted metabolite will have a mass, intensity, and retention time. MRM data are easier to interpret as they are restricted to specific peaks preselected for analysis. For quantitative analysis, peaks simply have to be integrated and compared between samples or to an internal standard.

In untargeted, accurate mass LC-MS, extracting real metabolite peaks from background electrical and chemical noise is highly challenging; it requires complex chemometric and bioinformatic pipelines to sift, filter, and annotate the raw data [185]. Known metabolites can be relatively easily detected with extracted ion chromatograms (thin slices across chromatographic time, showing a single peak or a small number of peaks for a specific mass-to-charge ratio) (see Figure 10.8), but detecting all the metabolite signals

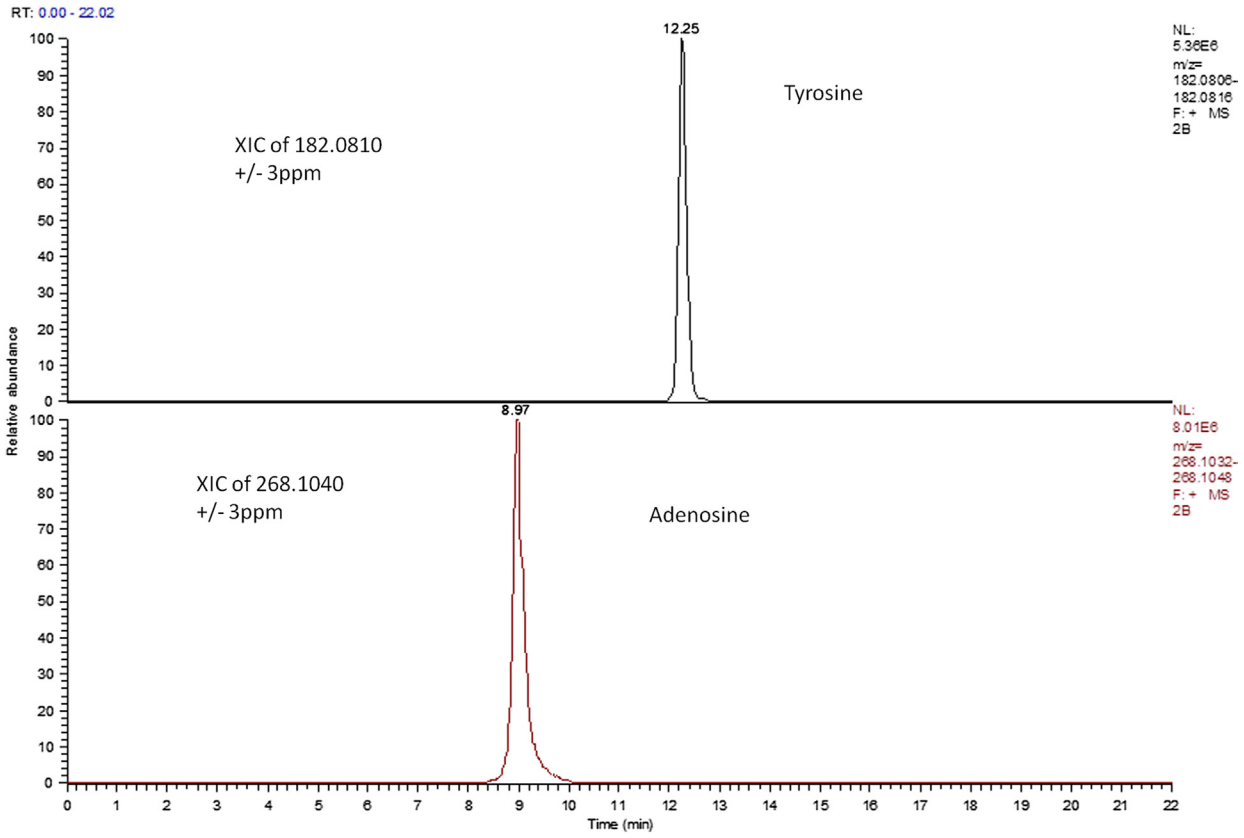

FIGURE 10.8 Extracted ion chromatogram. Accurate mass of tyrosine (top) and adenosine (bottom). In these instances, a 3 ppm mass tolerance is sufficient to exclude all other detected compounds from the chromatogram.

requires an unbiased method for peak picking. Several are available, including those incorporated into the XCMS [186], MzMine [187], MAVEN [116], ProGenesis CoMet, and Sieve software packages.

Once peaks are extracted, they must be processed; noisy peaks must be removed; and the entire dataset must be correlated, quantified, and identified. A good metabolomics pipeline must support all of these steps, and many pipelines are available with several alternatives for each one. It is therefore extremely important to apply the correct method for the appropriate data.

Gap filling should be performed on the data. Peak-picking algorithms are not infallible, and often peaks with nonoptimal shapes are excluded in one or more samples from a group. In cases where a peak is detected for some members of a replicate set, it is very valuable to re-search the original raw data for peaks that, rather than not being in the sample, have merely been missed by the peak detection algorithm [185]. At a later point in the processing, missing values should be annotated as n/a rather than 0, as this skews the probabilities for detected values and unrealistically represents the baseline noise level in the dataset. Missing values may also be mathematically imputed given sufficient other samples in each replicate set [188].

Normalization of metabolite data is also extremely challenging, as there are no canonical "housekeeping" metabolites available to perform normalization on. Various techniques have been applied, such as quantile normalization [189], normalization to internal standards [190], normalization to quality control samples [81], and normalization to creatinine concentration for urine samples [191]. In addition to normalization, scaling and log transformation may be applied to metabolomics datasets to reduce the dynamic range of the data and minimize skew in intensity distribution [192].

Due to the high dimensionality of the data and generally the low number of samples compared to the number of variables, it is essential that correction for multiple testing is performed on the data [193], especially if univariate statistics such as t-tests are applied to the data. The most conservative modification to the probabilities is Bonferroni correction [193]; Benjamini-Hochberg [194] correction is a less stringent method for reducing errors due to multiple testing.

10.4.3 Contextualization of Metabolomics Data

Once raw data are translated to a list of quantified metabolites, the most time-consuming and difficult part of the entire metabolomics practice starts: interpretation. Fitting the thousands of detected compounds and hundreds of significantly changing metabolites in a single experiment into a meaningful biological context is extremely challenging.

For fingerprinting analysis, patterns may be generated using multivariate statistical methods such as principal component analysis (PCA) [70], partial least squares discriminant analysis (PLS-DA) [70], or support vector machines (SVMs) [195], which will allow unsupervised (PCA) or supervised (PLS-DA and SVM) clustering of metabolomic data.

When attempting to move to the level of elucidating pathways and networks, however, it is worthwhile to begin by focusing on the most significantly changing metabolites, whether or not they are identified; after all, the most exciting discoveries are likely to be those metabolites that have not as yet been described. In general, it is helpful to examine the metabolic pathways that the compounds fit into. If most metabolites that change in one direction or another belong to the same or related pathways, this is a good indication that this pathway is important to the question being addressed.

There are several pathway-mapping packages available—for example, MetExplore [23], MassTrix [196], Pathos [197], and iPath [198]—which place already identified compounds in their constituent pathways. Pathos and MetExplore can also display up- and downregulation of the particular compounds via color coding in the generated maps. These can be very useful for observing general patterns of change in metabolites.

10.4.4 Annotation/Identification

While many of the metabolites detected may be assigned identities, it is important to classify them as either identifications or annotations depending on the surety of the assignment. According to the metabolite standards initiative [199], only compounds verified by accurate mass, MRM and retention time, or 2D NMR spectral match can be classified as identifications, while those identified by mass or substructure (fragmentation) are categorized as annotations. Although this does not cover all eventualities (especially those situations where retention time and accurate mass are not sufficient to guarantee a single identity matched to a standard), it does provide a reasonable guide to the confidence of an assignment. Of course, heuristics can be applied in context to allow greater confidence: if a number of compounds are detected in the same pathway, with the same direction of change, it is more likely that their identifications are correct.

10.5 CONCLUSION

Metabolomics is a collection of powerful tools for the analysis of phenotype, both by hypothesis generation and by hypothesis testing. Building on the strengths of the "omics technologies that came before, metabolomics uniquely comprises analytical technologies that can provide diagnostic

patterns via fingerprinting, absolute quantitation of targeted metabolites via pool analysis, relative quantitation of large portions of the metabolome using metabolite profiling, and tracing of the biochemical fate of individual metabolites through a metabolic system via flux analysis. Each of these technologies is supported by the two most commonly used and powerful techniques currently available for metabolomics: mass spectrometry and NMR.

Mass spectrometry-based metabolomics techniques are the most sensitive for simultaneous analysis of a large number of compounds. While limited in quantitation capabilities without appropriate labeled standards, the sheer quantity of information available in a single LC-MS or GC-MS experiment can provide detailed information on the patterns of metabolite change in an entire metabolic network.

NMR metabolomics complements mass spectrometry. It is limited in terms of sensitivity, but is uniquely capable of elucidating molecular structure. An important additional feature of NMR is that it is quantitative, capable of providing absolute levels of detected compounds when appropriate techniques are used.

Interpreting and deriving context from complex metabolomic datasets are highly challenging and represent a major area of research. Still, great strides are being made in the integration of metabolome data with genomics and proteomics. Metabolomics offers the promise that, in the future, biochemical analysis of the entire path from genotype to phenotype will be measured and explored for new insights into biology and medicine.

REFERENCES

[1] Lander ES et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860–921.

[2] Crick F. Central dogma of molecular biology. Nature 1970;227(5258):561–3.

[3] Fu J et al. System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet 2009;41(2):166–7.

[4] Burton PR et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–78.

[5] Duerr RH et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314(5804):1461–3.

[6] Harold D et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41(10):1088–93.

[7] De Gobbi M et al. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science 2006;312(5777):1215–7.

[8] Hugot JP et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411(6837):599–603.

[9] Velaga M et al. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab 2004;89(11):5862–5.

[10] Oliver SG et al. Systematic functional analysis of the yeast genome. Trends Biotechnol 1998;16(9):373–8.

[11] Gates SC, Sweeley CC. Quantitative metabolic profiling based on gas chromatography. Clin Chem 1978;24(10):1663–73.

[12] Jellum E. Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids. J Chromatogr B Biomed Sci Appl 1977;143(5):427–62.

[13] Politzer I, Dowty B, Laseter J. Use of gas chromatography and mass spectrometry to analyze underivatized volatile human or animal constituents of clinical interest. Clin Chem 1976;22(11):1775–88.

[14] Bales J et al. Metabolic profiling of body fluids by proton NMR: Self-poisoning episodes with paracetamol (acetaminophen). Magn Reson Med 1988;6(3):300–6.

[15] Gartland K et al. Pattern recognition analysis of high resolution 1H NMR spectra of urine. A nonlinear mapping approach to the classification of toxicological data. NMR Biomed 1990;3(4):166–72.

[16] Resnick L, Nicholson J, Laragh J. Calcium metabolism in essential hypertension: relationship to altered renin system activity. In: Fed proc; 1986.

[17] Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst 2005;130(5):606–25.

[18] Soga T et al. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2003;2(5):488–94.

[19] Burtis CA, Ashwood ER, Bruns DE, editors. Tietz Fundamentals of Clinical Chemistry. 6th ed. WB Saunders; 2008.

[20] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 5th ed. Garland, Science; 2008. 1392

[21] Stryer LBJM, Tymoczko JL. Biochemistry. 7th ed. W. H. Freeman; 2011. 1026

[22] Jeong H et al. The large-scale organization of metabolic networks. Nature 2000;407(6804):651–4.

[23] Cottret L et al. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. Nucleic Acids Res 2010;38(Suppl. 2):W132–7.

[24] Förster J et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 2003;13(2):244–53.

[25] Wishart DS et al. HMDB: the human metabolome database. Nucleic Acids Res 2007;35(Suppl. 1):D521–6.

[26] Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics 2001;2(3):155–68.

[27] Kim J-W, Dang CV. Cancer's molecular sweet tooth and the Warburg effect. Cancer Res 2006;66(18):8927–30.

[28] Chaneton B et al. Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature 2012.

[29] Vizán P, Mazurek S, Cascante M. Robust metabolic adaptation underlying tumor progression. Metabolomics 2008;4(1):1–12.

[30] Kenny LC et al. Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. Hypertension 2010;56(4):741–9.

[31] Lanza IR et al. Quantitative metabolomics by 1H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. PLoS One 2010;5(5):e10538.

[32] Jung JY et al. 1H-NMR-based metabolomics study of cerebral infarction. Stroke 2011;42(5):1282–8.

[33] Lockman KA et al. Oxidative stress rather than triglyceride accumulation is a determinant of mitochondrial dysfunction in *in vitro* models of hepatic cellular steatosis. Liver Int 2012;32(7):1079–92.

[34] OuYang D et al. Metabolomic profiling of serum from human pancreatic cancer patients using 1H NMR spectroscopy and principal component analysis. Appl Biochem Biotechnol 2011;165(1):148–54.
[35] Vulimiri SV et al. Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. Chem Res Toxicol 2009;22(3):492–503.
[36] Barrett MP, Bakker BM, Breitling R. Metabolomic systems biology of trypanosomes. Parasitology 2010;137(9):1285.
[37] Vincent IM et al. A molecular mechanism for eflornithine resistance in African trypanosomes. PLoS Pathog 2010;6(11):e1001204.
[38] Vincent IM et al. Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against *Trypanosoma brucei*. PLoS Negl Trop Dis 2012;6(5):e1618.
[39] Meyer H, Liebeke M, Lalk M. A protocol for the investigation of the intracellular *Staphylococcus aureus* metabolome. Anal Biochem 2010;401(2):250–9.
[40] Slupsky CM et al. *Streptococcus pneumoniae* and *Staphylococcus aureus* pneumonia induce distinct metabolic responses. J Proteome Res 2009;8(6):3029–36.
[41] Takahashi N, Washio J, Mayanagi G. Metabolomics of supragingival plaque and oral bacteria. J Dent Res 2010;89(12):1383–8.
[42] Calvani R et al. Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. Int J Obes 2010;34(6):1095–8.
[43] Proctor LM. The human microbiome project in 2011 and beyond. Cell Host Microbe 2011;10(4):287–91.
[44] Mateo R et al. Capillary column gas-chromatographic identification of sugars in honey as trimethylsilyl derivatives. J Chromatogr 1987;410(2):319–28.
[45] Morrison B, Scotland CJ, Fleck A. The determination of plasma glucose in a diabetic clinic. Clin Chim Acta 1972;39(2):301–6.
[46] Vesper HW, Archibold E, Myers GL. Assessment of trueness of glucose measurement instruments with different specimen matrices. Clin Chim Acta 2005;358(1–2):68–74.
[47] Wahjudi PN et al. Measurement of glucose and fructose in clinical samples using gas chromatography/mass spectrometry. Clin Biochem 2010;43(1–2):198–207.
[48] Hannestad U, Lundblad A. Accurate and precise isotope dilution mass spectrometry method for determining glucose in whole blood. Clin Chem 1997;43(5):794–800.
[49] Hsu MC, Chen DP, Liu RH. Detection of abused drugs in urine by GC-MS. J Food Drug Anal 2009;17(4):233–45.
[50] Nair H et al. Clinical validation of a highly sensitive GC-MS platform for routine urine drug screening and real-time reporting of up to 212 drugs. J Toxicol 2013;7.
[51] Segura J, Ventura R, Jurado C. Derivatization procedures for gas chromatographic-mass spectrometric determination of xenobiotics in biological samples, with special attention to drugs of abuse and doping agents. J Chromatogr B Biomed Sci Appl 1998;713(1):61–90.
[52] Castagnetta LAM et al. Tissue content of hydroxyestrogens in relation to survival of breast cancer patients. Clin Cancer Res 2002;8(10):3146–55.
[53] Houghton E et al. Application of gas-chromatography mass-spectrometry to steroid analysis in equine sports–problems with enzyme hydrolysis. Org Mass Spectrom 1992;27(10):1061–70.
[54] Kelly GE et al. Metabolites of dietary (soya) isoflavones in human urine. Clin Chim Acta 1993;223(1–2):9–22.
[55] Liere P et al. Validation of an analytical procedure to measure trace amounts of neurosteroids in brain tissue by gas chromatography-mass spectrometry. J Chromatogr B 2000;739(2):301–12.
[56] Fiehn O, Kind T. Metabolite profiling in blood plasma. In: Weckwerth W, editor. Metabolomics. Humana Press; 2007. p. 3–17.
[57] Halket JM et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 2005;56(410):219–43.
[58] Lisec J et al. Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc 2006;1(1):387–96.
[59] Roessner U et al. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. Plant J 2000;23(1):131–42.
[60] Chan ECY, Pasikanti KK, Nicholson JK. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. Nat Protoc 2011;6(10):1483–99.
[61] Serkova NJ et al. Early detection of graft failure using the blood metabolic profile of a liver recipient. Transplantation 2007;83(4):517.
[62] Constantinou M et al. Application of nuclear magnetic resonance spectroscopy combined with principal component analysis in detecting inborn errors of metabolism using blood spots: a metabonomic approach. Anal Chim Acta 2004;511(2):303–12.
[63] Beckonert O et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2007;2(11):2692–703.
[64] Walsh MC et al. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr 2006;84(3):531–9.
[65] Wishart DS et al. The human cerebrospinal fluid metabolome. J Chromatogr B 2008;871(2):164–73.
[66] Brindle JT et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med 2002;8(12):1439–45.
[67] Otvos JD, Jeyarajah EJ, Cromwell WC. Measurement issues related to lipoprotein heterogeneity. Am J Cardiol 2002;90(8):22–9.
[68] Nicholson JK et al. Metabolic phenotyping in clinical and surgical environments. Nature 2012;491(7424):384–92.
[69] Serkova NJ, Standiford TJ, Stringer KA. The emerging field of quantitative blood metabolomics for biomarker discovery in critical illnesses. Am J Respir Crit Care Med 2011;184(6):647.
[70] Lutz NW, Sweedler JV, Wevers RA. Methodologies for metabolomics: experimental strategies and techniques. Cambridge University Press; 2013.
[71] Griffin JL. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. Curr Opin Chem Biol 2003;7(5):648–54.
[72] Kamleh M et al. Metabolomic profiling of *Drosophila* using liquid chromatography Fourier transform mass spectrometry. FEBS Lett 2008;582(19):2916–22.
[73] Kamleh MA et al. Towards a platform for the metabonomic profiling of different strains of *Drosophila melanogaster* using liquid chromatography–Fourier transform mass spectrometry. FEBS J 2009;276(22):6798–809.
[74] Malmendal A et al. Metabolomic profiling of heat stress: hardening and recovery of homeostasis in *Drosophila*. Am J Physiol Regul Integr Comp Physiol 2006;291(1):R205–12.
[75] Welthagen W et al. Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC × GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. Metabolomics 2005;1(1):65–73.

[76] McMurray RJ et al. Nanoscale surfaces for the long-term maintenance of mesenchymal stem cell phenotype and multipotency. Nat Mater 2011;10(8):637–44.
[77] McNamara LE et al. Skeletal stem cell physiology on functionally distinct titania nanotopographies. Biomaterials 2011;32(30):7403–10.
[78] Tsimbouri PM et al. Using nanotopography and metabolomics to identify biochemical effectors of multipotency. ACS Nano 2012;6(11):10239–49.
[79] Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. BioEssays 2004;26(1):99–105.
[80] Zelena E et al. Development of a robust and repeatable UPLC− MS method for the long-term metabolomic study of human serum. Anal Chem 2009;81(4):1357–64.
[81] Kamleh MA et al. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. Anal Chem 2012;84(6):2670–7.
[82] Horgan GW. Sample size and replication in 2D gel electrophoresis studies. J Proteome Res 2007;6(7):2884–7.
[83] Parsons HM et al. Spectral relative standard deviation: a practical benchmark in metabolomics. Analyst 2009;134(3):478–85.
[84] Kaderbhai NN et al. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT–IR and direct injection electrospray mass spectrometry. Comp Funct Genomics 2003;4(4):376–91.
[85] Keymanesh K, Darvishi MH, Sardari S. Metabolome comparison of transgenic and non-transgenic rice by statistical analysis of FTIR and NMR spectra. Rice Sci 2009;16(2):119–23.
[86] Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 2006;131(8):875–85.
[87] Harrigan GG, Goodacre R. Metabolic profiling: its role in biomarker discovery and gene function analysis. Springer; 2003.
[88] Birkemeyer C et al. Metabolome analysis: the potential of *in vivo* labeling with stable isotopes for metabolite profiling. Trends Biotechnol 2005;23(1):28–33.
[89] Yang J et al. Discrimination of Type 2 diabetic patients from healthy controls by using metabonomics method based on their serum fatty acid profiles. J Chromatogr B 2004;813(1):53–8.
[90] Bictash M et al. Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. J Clin Epidemiol 2010;63(9):970–9.
[91] Balog J et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. Sci Trans Med 2013;5(194):194ra93.
[92] Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. J Bacteriol 1998;180(19):5109–16.
[93] Taylor RL, Machacek D, Singh RJ. Validation of a high-throughput liquid chromatography–tandem mass spectrometry method for urinary cortisol and cortisone. Clin Chem 2002;48(9):1511–9.
[94] Huang X, Regnier FE. Differential metabolomics using stable isotope labeling and two-dimensional gas chromatography with time-of-flight mass spectrometry. Anal Chem 2008;80(1):107–14.
[95] Bajad SU et al. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J Chromatogr A 2006;1125(1):76–88.
[96] Troufflard S et al. *In vivo* ^{13}C NMR determines metabolic fluxes and steady state in linseed embryos. Phytochemistry 2007;68(16):2341–50.
[97] Creek DJ et al. Metabolomic analysis of trypanosomatid protozoa. Mol Biochem Parasitol 2012;181(2):73–84.
[98] Hirayama A et al. Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. Cancer Res 2009;69(11):4918–25.
[99] t'Kindt R et al. Metabolomics to unveil and understand phenotypic diversity between pathogen populations. PLoS Negl Trop Dis 2010;4(11):e904.
[100] Kettunen J et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 2012;44(3):269–76.
[101] Inouye M et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Syst Biol 2010;6(1).
[102] Soininen P et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst 2009;134(9):1781–5.
[103] Würtz P et al. Characterization of systemic metabolic phenotypes associated with subclinical atherosclerosis. Mol BioSyst 2011;7(2):385–93.
[104] Mäkinen V-P et al. 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. Mol Syst Biol 2008;4(1).
[105] Tukiainen T et al. A multi-metabolite analysis of serum by ^{1}H NMR spectroscopy: early systemic signs of Alzheimer's disease. Biochem Biophys Res Commun 2008;375(3):356–61.
[106] Ala-Korpela M, Kangas AJ, Inouye M. Genome-wide association studies and systems biology: together at last. Trends Genet 2011;27(12):493–8.
[107] Ala-Korpela M, Kangas AJ, Soininen . Quantitative high-throughput metabolomics: a new era in epidemiology and genetics. Genome Med 2012;4(4):36.
[108] Inouye M et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet 2012;8(8):e1002907.
[109] Breitling R, Achcar F, Takano E. Modeling challenges in the synthetic biology of secondary metabolism. ACS Syn Biol 2013;2:373–8.
[110] Zamboni N, Sauer U. Novel biological insights through metabolomics and ^{13}C-flux analysis. Curr Opin Microbiol 2009;12(5):553–8.
[111] Wisselink HW et al. Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *Saccharomyces cerevisiae*. Metab Eng 2010;12(6):537–51.
[112] Munger J et al. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nat Biotechnol 2008;26(10):1179–86.
[113] Portais JC et al. Metabolic flux determination in C6 glioma cells using carbon-13 distribution upon [1-13C] glucose incubation. Eur J Biochem 1993;217(1):457–68.
[114] Sauer U. Metabolic networks in motion: 13C-based flux analysis. Mol Syst Biol 2006;2(1).
[115] Chokkathukalam A et al. MzMatch–ISO: an R tool for the annotation and relative quantification of isotope-labeled mass spectrometry data. Bioinformatics 2013;29(2):281–3.
[116] Clasquin MF, Melamud E, Rabinowitz JD. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine.

Curr Protc Bioinform 2012. http://dx.doi.org/10.1002/0471250953.bi1411s37 [Chapter 14:Unit 14.11].
[117] Barton RH et al. High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. Int J Epidemiol 2008;37(Suppl. 1):i31–40.
[118] Barding Jr GA, Salditos R, Larive CK. Quantitative NMR for bioanalysis and metabolomics. Anal Bioanal Chem 2012;404(4):1165–79.
[119] Fonville JM et al. Evaluation of full-resolution J-resolved 1H NMR projections of biofluids for metabonomics information retrieval and biomarker identification. Anal Chem 2010;82(5):1811–21.
[120] van der Hooft JJ et al. Spectral trees as a robust annotation tool in LC–MS based metabolomics. Metabolomics 2012;8(4):691–703.
[121] Jacobsen N. NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology. Wiley-Blackwell; 2007.
[122] Hoult D, Ginsberg N. The quantum origins of the free induction decay signal and spin noise. J Magn Reson 2001;148(1):182–99.
[123] Teahan O et al. Impact of analytical bias in metabonomic studies of human blood serum and plasma. Anal Chem 2006;78(13):4307–18.
[124] Barton RH et al. The influence of EDTA and citrate anticoagulant addition to human plasma on information recovery from NMR-based metabolic profiling studies. Mol BioSyst 2010;6(1):215–24.
[125] Psychogios N et al. The human serum metabolome. PLoS One 2011;6(2):e16957.
[126] Mckay RT. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: an examination of the pulse sequence components and evolution. Conc Magn Reson Part A 2011;38(5):197–220.
[127] Meiboom S, Gill D. Modified spin-echo method for measuring nuclear relaxation times. Rev Sci Instrum 1958;29(8):688–91.
[128] Becker ED et al. The choice of optimal parameters for measurement of spin-lattice relaxation times. II. Comparison of saturation recovery, inversion recovery, and fast inversion recovery experiments. J Magn Reson(1969) 1980;37(3):381–94.
[129] Tal A, Frydman L. Single-scan multidimensional magnetic resonance. Prog Nucl Magn Reson Spectrosc 2010;57(3):241–92.
[130] Le Guennec A et al. Fast determination of absolute metabolite concentrations by spatially encoded 2D NMR: application to breast cancer cell extracts. Anal Chem 2012;84(24):10831–7.
[131] Mori S et al. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. J Magn Reson Ser B 1995;108(1):94–8.
[132] Thrippleton MJ, Edden RA, Keeler J. Suppression of strong coupling artefacts in *J*-spectra. J Magn Reson 2005;174(1):97–109.
[133] Ludwig C, Viant MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochem Anal 2010;21(1):22–32.
[134] Xia J et al. MetaboMiner–semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. BMC Bioinform 2008;9(1):507.
[135] Chignola F et al. The CCPN metabolomics project: a fast protocol for metabolite identification by 2D-NMR. Bioinformatics 2011;27(6):885–6.
[136] Cloarec O et al. Statistical total correlation spectroscopy (STOCSY): a new approach for individual biomarker identification from metabonomic NMR datasets. Anal Chem 2005;77:1282–9.
[137] Crockford DJ et al. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. Anal Chem 2006;78(2):363–71.
[138] Piotto M, Saudek V, Sklenář V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR 1992;2(6):661–5.
[139] Wishart DS. Quantitative metabolomics using NMR. TrAC Trends Anal Chem 2008;27(3):228–37.
[140] Raiford DS, Fisk CL, Becker ED. Calibration of methanol and ethylene glycol nuclear magnetic resonance thermometers. Anal Chem 1979;51(12):2050–1.
[141] Paul W, Steinwedel H. A new mass spectrometer without a magnetic field. Zeitschrift fuer Naturforschung (West Germany). Divided into Z. Nautrforsch., A, and Z. Naturforsch., B: Anorg Chem Org Chem. Biochem Biophys 1953;8:448–50.
[142] Goudsmit S. A time-of-flight mass spectrometer. Phys Rev 1948;74(5):622.
[143] Mamyrin B et al. The massreflect ron, a new non-magnetic time-of-flight mass spectrometer with high resolution. Zh Eksp Teor Fiz 1973;64:82–9.
[144] Hu Q et al. The Orbitrap: a new mass spectrometer. J Mass Spectrom 2005;40(4):430–43.
[145] Comisarow MB, Marshall AG. Fourier transform ion cyclotron resonance spectroscopy. Chem Phys Lett 1974;25(2):282–3.
[146] Fiehn O et al. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem 2000;72(15):3573–80.
[147] Kind T, Fiehn O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinform 2007;8(1):105.
[148] Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinform 2006;7(1):234.
[149] McReynolds JH, Anbar M. Isotopic assay of nanomole amounts of nitrogen-15 labeled amino acids by collision-induced dissociation mass spectrometry. Anal Chem 1977;49(12):1832–6.
[150] Zakett D, Flynn R, Cooks R. Chlorine isotope effects in mass spectrometry by multiple reaction monitoring. J Phys Chem 1978;82(22):2359–62.
[151] Yost R, Enke C. Selected ion fragmentation with a tandem quadrupole mass spectrometer. J Am Chem Soc 1978;100(7):2274–5.
[152] Annesley TM. Ion suppression in mass spectrometry. Clin Chem 2003;49(7):1041–4.
[153] Lu W et al. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Anal Chem 2010;82(8):3212–21.
[154] Kamleh A, Barrett MP, Wildridge D, Burchmore RJ, Scheltema RA, Watson DG. Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. Rapid Commun Mass Spectrom 2008;22(12):1912–8.
[155] Bruce SJ et al. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal Chem 2009;81(9):3285–96.
[156] Kim HK, Verpoorte R. Sample preparation for plant metabolomics. Phytochem Anal 2010;21(1):4–13.
[157] Villas-Bôas SG et al. Global metabolite analysis of yeast: evaluation of sample preparation methods. Yeast 2005;22(14):1155–69.
[158] Koning Wd, Dam Kv. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal Biochem 1992;204(1):118–23.

[159] Shimada K, Mitamura K. Derivatization of thiol-containing compounds. J Chromatogr B Biomed Sci Appl 1994;659(1):227–41.
[160] Dole M et al. Molecular beams of macroions. J Chem Phys 1968;49:2240.
[161] Robb DB, Covey TR, Bruins AP. Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. Anal Chem 2000;72(15):3653–9.
[162] Guignard C et al. Analysis of carbohydrates in plants by high-performance anion-exchange chromatography coupled with electrospray mass spectrometry. J Chromatogr A 2005;1085(1):137–42.
[163] Antonio C et al. Quantification of sugars and sugar phosphates in *Arabidopsis thaliana* tissues using porous graphitic carbon liquid chromatography-electrospray ionization mass spectrometry. J Chromatogr A 2007;1172(2):170–8.
[164] Desai MJ, Armstrong DW. Analysis of native amino acid and peptide enantiomers by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. J Mass Spectrom 2004;39(2):177–87.
[165] Dwivedi P et al. Metabolic profiling by ion mobility mass spectrometry (IMMS). Metabolomics 2008;4(1):63–80.
[166] Kanu AB et al. Ion mobility–mass spectrometry. J Mass Spectrom 2008;43(1):1–22.
[167] Rojas-Cherto M et al. Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. Anal Chem 2012;84(13):5524–34.
[168] Gross JH. Mass spectrometry: a textbook. 2nd ed. Springer; 2011. p. 774
[169] Honour JW. Development and validation of a quantitative assay based on tandem mass spectrometry. Ann Clin Biochem 2011;48(2):97–111.
[170] Huang N et al. Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data. J Am Soc Mass Spectrom 1999;10(11):1166–73.
[171] NIST standard reference database 1A2011: national institute of standards and technology.
[172] Kopka J et al. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 2005;21(8):1635–8.
[173] Kvitvang HFN et al. Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. Anal Chem 2011;83(7):2705–11.
[174] Knapp DR. Handbook of analytical derivatization reactions. Wiley-Intercience; 1979.
[175] Evershed RP, Blau K, Halket JM. Handbook of derivatives for chromatography. 2nd ed. Wiley; 1993.
[176] Pierce AE. Silylation of organic compounds. Pierce Chemical Company; 1968.
[177] Xu X et al. A liquid chromatography-mass spectrometry method for the quantitative analysis of urinary endogenous estrogen metabolites. Nat Protoc 2007;2(6):1350–5.
[178] Fine DD et al. Quantitation of estrogens in ground water and swine lagoon samples using solid-phase extraction, pentafluorobenzyl/trimethylsilyl derivatizations and gas chromatography–negative ion chemical ionization tandem mass spectrometry. J Chromatogr A 2003;1017(1–2):167–85.
[179] Pawlosky RJ, Sprecher HW, Salem N. High sensitivity negative ion GC-MS method for detection of desaturated and chain-elongated products of deuterated linoleic and linolenic acids. J Lipid Res 1992;33(11):1711–7.
[180] Peters FT, Kraemer T, Maurer HH. Drug testing in blood: validated negative-ion chemical ionization gas chromatographic–mass spectrometric assay for determination of amphetamine and methamphetamine enantiomers and its application to toxicology cases. Clin Chem 2002;48(9):1472–85.
[181] Wishart DS et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6(2):135–40.
[182] Hahn E, Maxwell D. Spin echo measurements of nuclear spin coupling in molecules. Phys Rev 1952;88(5):1070.
[183] Dieterle F et al. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem 2006;78(13):4281–90.
[184] Allen J et al. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol 2003;21(6):692–6.
[185] Castillo S et al. Algorithms and tools for the preprocessing of LC–MS metabolomics data. Chemometrics Intell Lab Syst 2011;108(1):23–32.
[186] Smith CA et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 2006;78(3):779–87.
[187] Pluskal T et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinform 2010;11(1):395.
[188] Hrydziuszko O, Viant MR. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. Metabolomics 2012;8(1):161–74.
[189] Sung Min P, Hae Won L, Young-Ran Y. Quantile normalization approach for liquid chromatography–mass spectrometry-based metabolomic data from healthy human volunteers. Anal Sci 2012;28:801.
[190] Nordström A et al. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. Anal Chem 2006;78(10):3289–95.
[191] Chen Y, Shen G, Zhang R, He J, Zhang Y, Xu J, et al. Combination of injection volumes calibration by creatinine and MS signals' normalization to overcome urine variability in LC-MS-based metabolomics studies. Anal Chem 2013;85(16):7659–65.
[192] van den Berg RA et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC genomics 2006;7(1):142.
[193] Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006;2(4):171–96.
[194] Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9(7):811–8.
[195] Mahadevan S et al. Analysis of metabolomic data using support vector machines. Anal Chem 2008;80(19):7562–70.
[196] Suhre K, Schmitt-Kopplin . MassTRIX: mass translator into pathways. Nucleic Acids Res 2008;36(suppl 2):W481–4.
[197] Leader DP et al. Pathos: a web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. Rapid Commun Mass Spectrom 2011;25(22):3422–6.
[198] Letunic I et al. IPath: interactive exploration of biochemical pathways and networks. Trends Biochem Sci 2008;33(3):101–3.
[199] Sumner LW et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 2007;3(3):211–21.