Precision mapping of the metabolome

Rainer Breitling¹, Andrew R. Pitt² and Michael P. Barrett³

¹ Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9751 NN Haren, The Netherlands
² Sir Henry Wellcome Functional Genomics Facility, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK
³ Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

The global study of the structure and dynamics of metabolic networks has been hindered by a lack of techniques that identify metabolites and their biochemical relationship in complex mixtures. The recent application of Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) to metabolomic analysis suggests a way to tackle the problem. A lower-cost alternative to high-field FTICR-MS, the Orbitrap mass analyzer, promises accelerated activity in this area. Here, we show how the ultra-high mass accuracy and resolution provided by this new generation of mass spectrometers can help to identify metabolites and connect them into metabolic networks. Data from perturbation studies and isotope-tracking experiments can complement this information to create metabolic maps de novo and chart unexplored areas of metabolism.

The challenge of metabolomics

The study of metabolism is a founding discipline of biochemistry. Researchers elucidated the structure of central metabolic pathways in classical experiments and began to determine cell-wide protein interaction networks long before they sequenced whole genomes. In the post-genome era, however, metabolomics (see Glossary) has generally lagged behind its cousins: genomics, transcriptomics and proteomics. In this article we will discuss how recent developments in high-resolution mass spectrometry (Fourier transform ion cyclotron resonance and Orbitrap mass analyzers) could change this situation.

The immense chemical diversity of small biomolecules makes comprehensive metabolome screens, comparable to gene expression microarrays, difficult [1–3]. The lack of unifying principles (such as the genetic code) that would assist molecule identification, comparison and causal connection is another important challenge. Hence, although whole-genome information has been used to infer the global metabolic network structure within entire organisms, such efforts have been limited to well-studied areas of metabolism. Moving beyond the traditional biochemical map requires new experimental approaches. Developing analytical methods that can detect and quantify large numbers of metabolites (see Glossary) in parallel was a necessary first step, but a majority of all small molecules detected in a cellular sample remain unidentified. Furthermore, the biochemical connectivity between such novel metabolites has been impossible to assess. Even well known and intensely studied biosynthetic pathways have proven difficult to survey comprehensively. For example, the last biosynthetic enzyme of mammalian cholesterol biosynthesis was identified only in 2002 [4]. Major parts of secondary metabolism, which determine important aspects of the biological function and phenotype of an organism, remain entirely unexplored.

Recent advances in ultra-high mass accuracy (see Glossary) mass spectrometers (Box 1) could change this situation. For some time, mass spectrometry has been the technology of choice for resolving and quantifying complex metabolite samples [5]. However, the identification of the metabolites underlying the measured masses and, even more so, their connection in metabolic networks was time-consuming and impossible on a large scale. We argue that two major advantages provided by the new ultra-high mass accuracy equipment can enable ab initio determination of metabolic networks: (i) the ability to identify molecular formulae (see Glossary) based on exact masses; and (ii) the inference of biosynthetic relationships between masses directly from the mass spectrum. Mass spectrometers with the necessary performance parameters (mass accuracy around 1 ppm and resolution above 100 000) are now available online 24 October 2006.

Glossary

**Mass resolution:** The resolution for a single peak of mass \( m \) is commonly expressed as \( m/\Delta m \), where \( \Delta m \) is the width of the peak at half its height (FWHM) resolution. Masses differing by the mass resolution yield distinguishable peaks. For metabolites, the resolution can be around 100 000 in ultra-high resolution mass spectrometers, which means that they can, in principle, resolve millions of peaks (although the dynamic range of the instruments will prevent such extreme performance).

**Mass accuracy:** Difference between measured and actual mass relative to the actual mass. Using FTICR-MS or an Orbitrap mass analyzer this can be as low as 1 part per million (ppm), which means that a mass of 100 Da is measured accurately up to four decimal places.

**Metabolite:** Small biological molecule, usually with a mass below 1500 Da. Typical examples are sugars, nucleotides, alkaloids and lipids.

**Metabolomics:** The comprehensive identification and quantification of all metabolites and their relationships in a biological sample.

**Molecular formula:** The absolute number of atoms of each element that make up a molecule. Mass spectrometry can not discriminate between molecules that have the same molecular formula but different structure (isomers) unless additional information (e.g., from chromatography or mass-spectrometric fragmentation patterns) is used.

**Monoisotopic mass:** The mass of a molecule when all constituent atoms are of the most common isotope. For small molecules this mass will give the largest peak in a mass spectrum.
within the realm of many researchers and will change the way we think about metabolomics.

**Molecular identification**

Molecules of the same nominal (integer) mass but different molecular formulae will have slightly different exact masses. For example, ethanol (C₂H₅OH) and methyl hydrazine (CH₃N₂) both have a nominal mass of 44, but their exact masses (44.026220 and 44.037449) differ by 0.011 Da (225 ppm). If only the most common biological elements are permitted (carbon, hydrogen, oxygen, nitrogen, sulfur, and phosphate), only a finite number of combinations of atoms will give a match to an observed mass with a given degree of accuracy. In the case of metabolites, which are small molecules by definition, it has been shown that a mass accuracy of 1 ppm will often enable an (almost) unique assignment to a molecular formula [6,7] (Figure 1). Such high accuracy is not routinely accessible by high-resolution time-of-flight mass spectrometers, even using an internal mass lock system, but can be achieved by the new generation of mass analyzers (Box 1) [8–11]. Because the working principle of these analyzers ensures that their mass accuracy increases with decreasing molecular weight [12,13], they are particularly suitable for metabolomics. The identified formulae can then be used to search for corresponding matches in databases of organic chemicals, such as PubChem (http://pubchem.ncbi.nlm.nih.gov/) or the Dictionary of Natural Products (http://www.chemnetbase.com/scripts/dnpweb.exe). However, even at an accuracy of 1 ppm there are still limitations to this approach. For example, even relatively small compounds such as C₅H₇O₂P₂ (1,2-ethanediylbis-phosphine oxide) and C₃H₂N₄S (1,3,4-thiadiazol-2-ylamino)formonitrile) have monoisotopic masses (300 and 1000 Da) [14–18]. Current FTICR-MS and Orbitrap analyzers do not achieve accurate isotope ratio measurements for many metabolites (particularly those between 300 and 1000 Da) [14–18].

**Box 1. Working principles of high-resolution mass spectrometers**

Currently, two types of mass analyzer are able to achieve, routinely, high-resolution and mass accuracy in the 1 ppm range for biomolecular samples. In both instruments, the ionized metabolite mixture is trapped in an orbital trajectory. The frequency of their orbit depends on the mass-over-charge ratio of the ions and can be measured precisely, which is the basis of the exceptional accuracy. In Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) [8,13], the trapping is achieved in a strong magnetic field, which exerts a force on the charged particles that is perpendicular to their direction of motion and thus confines them to a circular path (Figure 1a). An additional electric field contains the analytes in the direction parallel to the magnetic field. Initially, the ions have low kinetic energy and thus orbit close to the centre of the trap. They are excited by a radio pulse, which brings them to a wider orbit and, at the same time, focuses them in a coherent package. Because the ion packet now passes closely to the two detection electrodes, it induces a sinusoidal image current that can be measured. Because ions of multiple masses, and thus of different orbital frequency, are present in the trap, the detected signal is a superposition of multiple sinusoids, which are deconvoluted by fast Fourier transform and converted from a frequency spectrum to a mass spectrum. During the observation period, the excited ions travel a distance of many kilometers on their orbit, which gives an intuitive understanding of the physical basis for the potentially much higher mass resolution (see Glossary) compared with time-of-flight mass spectrometers [13]. Because the resolving power (and many other figures of merit of FTICR-MS) are proportional to the strength of the magnetic field [13], superconducting magnets (3–12 Tesla) are required, which makes the maintenance costs of FTICR-MS spectrometers relatively high. The Orbitrap is a recent alternative that traps ions without a magnetic field [20,32], which is easier to maintain and yields almost the same performance [9,12]. Here, ions are trapped in a radial electric field between a central and an outer cylindrical electrode (Figure 1b). As an ion packet enters the trap, the electric field is switched on and attracts the ions to the central electrode. If the ions are sufficiently fast, they will not plunge onto the electrode but remain on a circular orbit around it, in the same way as planets are captured in their orbit around central stars. In contrast to the FTICR-MS spectrometer, the frequency on this circular orbit depends, in a complex way, on the kinetic energy of the ions. To overcome this problem, the electrodes are produced in a special spindle-shaped form, which modifies the electric field in such a way that ions are attracted towards the equator of the spindle. Ions injected off-centre will thus begin harmonic (pendulum) oscillations around the equator, superimposed on their circular motion. The frequency of this oscillation is proportional to the mass-over-charge ratio and is detected and processed by fast Fourier transform, as in FTICR-MS.

![Figure 1](https://www.sciencedirect.com)
measurements (intensity accuracies of only 5–10% have been reported [19]) but improvements in time-of-flight mass spectrometers are showing great promise in this area. However, the high resolving power of top-end FTICR machines enables the discrimination of minor isotope peaks, for example, a compound containing one $^{34}$S atom can be resolved from a molecule containing two $^{13}$C atoms [20]. This would enable the correct identification of C$_2$H$_8$O$_2$P$_2$ and C$_3$H$_2$N$_4$S, as described above (Figure 2).

**Molecular connectivity**

Isotope-distribution analysis exploits relationships between mass peaks to interpret spectra, and accurate mass information enables still more extensive use of such relationships. Metabolites in a biological sample are related: they share biochemical ancestry and are linked within a closely connected metabolic network. Within a metabolome, molecules will generally be related to precursors and derivatives. Ultra-high mass accuracy also implies ultra-high mass difference accuracy. Accurate mass differences, in turn, contain information about chemical relationships between molecules. A limited repertoire of chemical transformations account for the majority of biochemical reactions within cells. For example, hydrogenation–dehydrogenation (mass difference $\pm$ H$_2$ = 2.01565 Da), oxidation (mass difference O = 15.9949 Da) and phosphorylation (mass difference PO$_3$H = 79.9663 Da) are typical common reactions. A more complex case is provided by condensation reactions, with a mass difference H$_2$O = 18.01055 compared to the sum of the two precursor masses. Detecting a difference corresponding to a common transformation between two metabolites implies that they are likely to be chemically related and might well be connected directly by an enzymatic reaction. A systematic screen for commonly occurring mass differences (metabolic-difference analysis) can create a comprehensive metabolite map of a sample (Figure 3). Such an approach is not new in mass spectrometry. Difference analysis is widely used in the interpretation of tandem mass spectra, where ‘neutral losses’ are identified and used to infer, for example, the sequence of a fragmented peptide. Metabolic-difference analysis is particularly powerful for analyzing homologous series of compounds and has recently been used to interpret the pattern of membrane phospholipids in the protozoan Trypanosoma brucei, a parasite that causes sleeping sickness [7] (Box 2). A similar approach was followed to analyze a series of homologous hydrocarbons in complex petrochemical samples [21]. In both cases, the ultra-high accuracy provided by FTICR-MS was necessary for the successful identification of intermolecular relationships.
Combination with orthogonal information

The study of metabolite fragmentation by tandem mass spectrometry can confirm and complement inferences based on metabolic-difference analysis. For example, a molecule of mass 785.59348 Da that is related to other masses by the addition and loss of C2H4 units, as well as by various dehydrogenation reactions, could be predicted with good confidence to be dioleoylphosphatidylcholine. If the fragmentation pattern in a tandem mass spectrum shows a signal for the choline headgroup as well as loss of an oleic acid side chain, this identification would be confirmed. Both FTICR-MS and Orbitrap mass analyzers are available in hybrid instruments with linear ion traps, which can make such a tandem strategy straightforward. FTICR-MS has the additional benefit of being able to select the ion for tandem spectrometry with high resolution in the ion cyclotron resonance cell and fragment it inside the cell, which has obvious advantages in a complex metabolomic sample.

Sample preparation, pre-processing and ionization all lead to a bias towards specific classes of chemicals, so that the final mass spectrum of a given sample will not represent the complete metabolome. Although this makes ‘one-shot’ metabolomic resolution unlikely, it does actually offer advantages. Biochemically connected compounds tend to be chemometrically similar [22]; therefore, they will tend to occur together in the spectrum derived from a particular sample preparation method. This information can be used to assist in molecular identification and to place molecules in a functional framework.

Metabolic difference analysis can also be combined with perturbation experiments, which monitor the correlation patterns of metabolites after external or internal changes to the system [23,24]. Combining this with iso-tope tracer studies, which follow the fate of metabolites that are labelled with stable isotopes [25,26], is also an obvious option to yield networks of chemically related metabolites. The prospect of implementing pulse-chase experiments to follow the fate of metabolic precursors in time is appealing. The additional knowledge about potential biochemical connections that is provided by ultra-high mass accuracy will constrain the possible interpretations of the spectra, enabling the analysis to be extended to areas where little or no previous pathway information exists. Shared correlation behaviour or isotope dynamics will also serve to distinguish true and false positive connections in the biochemical network, inferred from the mass spectra.

Ultimately, a systematic use of the multitude of parameters involved in sample preparation and analysis can assist in metabolite identification. Computational tools for the automatic integration of the various levels of information will provide great assistance in metabolome exploration.

Remaining challenges

Theoretically, the resolving power of the FTICR-MS and Orbitrap is sufficiently high to resolve even the most complex metabolite mixtures using direct infusion. This
can lead to rapid experimentation and minimizes bias introduced during pre-processing. However, such a direct infusion approach has important practical limitations, and a pre-separation step can be useful for several reasons. First, it provides an additional dimension of information (retention time) that can help the identification of molecules and separation of isomers. Molecules with the same mass but distinct structures and physiological function (e.g. glucose-1-phosphate, glucose-6-phosphate and fructose-6-phosphate) can only be discriminated and quantified when such additional information is available; even then, discrimination of stereoisomers can be difficult. Second, molecules frequently compete during the ionization step, so that the presence of one compound in high abundance or with a high ionization potential can strongly repress the signal of another (ion suppression), which substantially interferes with quantification. Such effects are reduced if the ionized mixture is less complex. Finally, pre-separation can increase the dynamic range of the analysis. Only a limited number of ions can be analyzed at the same time inside an ion trap because molecular crowding leads to interactions between the ions (space–charge effects) that distort the signal and substantially reduce mass accuracy. This is particularly the case for FTICR-MS, where all ions share the same orbit and, in extreme cases, mass errors in excess of 10 ppm can result. A pre-separation step spreads the spectrum over a larger number of scans and enables optimal ion accumulation at each time point. Both FTICR-MS and Orbitrap mass analyzers can be coupled to many different ion sources – including (nano)electro spray, atmospheric pressure chemical, and matrix-assisted laser desorption ionization – and various pre-separation protocols – liquid chromatography, gas chromatography or capillary electrophoresis. This is facilitated by hybrid mass spectrometers that couple the high-resolution mass analyzer to a preceding quadrupole or ion trap [12,27–29], where ions are accumulated, selected and potentially fragmented before analysis. Such a setup also enables automatic gain control, which minimizes space–charge effects and leads to concentration-independent accuracy throughout a wide dynamic range [10]. Optimizing the pre-processing steps (including the challenging sample preparation [5]) to ensure maximally unbiased transfer of metabolites to the mass analyzer will be crucial for further advances.

Box 2. Case studies

The number of metabolomics studies using ultra-high resolution mass spectrometry is currently limited. A first proof-of-principle study examined the metabolite composition of ripening strawberries and transgenic tobacco flowers by direct injection FTICR-MS, without previous separation [33]. Studies on Arabidopsis thaliana (thale cress) metabolism followed [34–36]. In a recent paper, the same technology was applied to the metabolome of the sleeping sickness parasite Trypanosoma brucei and its host environment [7]. In the trypanosome study, metabolite mass differences were measured with high precision and used successfully to confirm metabolite identification and to predict biochemical relationships between masses. Other FTICR-MS studies have focused on particular metabolite species, for example, phospholipid profiles of mouse brain, heart and liver [37]. Studies using Orbitrap analyzers in comprehensive metabolomics experiments have still to be published but initial reports using standard mixtures of small molecules show promising performance parameters, with mass accuracy around 2 ppm [12]. Accurate Orbitrap measurements combined with tandem mass spectrometry have been used successfully for sphingolipid profiling in yeast [27] and for tracking the metabolic fate of anti-depressant drugs in human liver microsome preparations [38]. In the latter case, the combination of accurate mass analysis, fragmentation patterns and predictive metabolite identification software enabled the reconstruction of biotransformation pathways with high confidence.

Figure 3. Mass differences of metabolites help the inference of metabolic relationships. (a) High-resolution mass spectrum of the Trypanosoma brucei metabolome, focusing on the mass range containing phospholipids. Mass 809.5935 is a phosphatidylcholine with 38 side-chain carbons and 4 unsaturated bonds (a possible formula is shown on top). It is related to other metabolites in the sample by characteristic accurate mass differences. (b) Knowledge of the exact mass differences identifies the related metabolites and places them in a fully connected metabolic network, of which only a small part is shown here. (c) Average abundances for all members of the phosphatidylcholine series. Longer side-chains contain larger numbers of unsaturated bonds. Patterns like this only become obvious when the data are organized on the basis of exact mass information.
Another major challenge is posed by data processing and analysis. The higher the resolution, the larger the data files that need to be analyzed. If additional dimensions are added by chromatographic pre-separation and tandem mass spectra, the task rapidly becomes daunting. New bioinformatics techniques need to be developed to automate and optimize both the data acquisition and the interpretation processes, to make routine high-throughput applications at the highest resolution possible. The first steps in this direction have recently been published [30,31].

The ultra-high resolution provided by the new generation of mass spectrometers promises that the resulting spectra will lead to exciting new discoveries in territories of the metabolic map that have, up to now, been inaccessible.

Acknowledgements
We thank three anonymous referees for their constructive comments on the manuscript.

References