

Metabolomics: analytic methods driving exponential understanding

Metabolomics is distinct from genomics and proteomics in several significant respects. Where genes and proteins are unique to specific organisms, metabolites are ubiquitous throughout the plant and animal kingdoms. Compared with proteomics and genomics, which quantify the building blocks and machinery of organisms, metabolomics provides information about the interaction between the more established 'omics and what has been termed as the 'wild card' of systems biology – the organism's environment¹.

While the relationship between genes and proteins is well understood, metabolites' connections to an organism's upstream building blocks and machinery are less direct and more variable. Because of the elements of time and environment, defining a metabolome requires that conditions be specified as well. This interrelationship provides metabolomics' singular advantages, and its greatest challenges.

Challenges

The goals of metabolomic studies are quantitative assessment of biochemical differences reflected in the metabolome, differential analysis between and among sample groups, and identifying compounds responsible for observed changes. Counter-balancing these objectives are formidable hurdles:

- Complexity and diversity of biological samples.
- Chemical diversity of small molecule metabolites.
- Wide concentration dynamic range, as high as 10¹⁴.
- Multiple sources of variability from sample, analysis methods, workflow, reagents, etc.
- Lack of analytical standards, particularly for unknown metabolites.

- Incomplete information – the majority of compounds detected by LC/MS are unknowns.
- Time and expense of unknown structure elucidation.
- The need for robust, reliable data handling and bioinformatics.
- Throughput issues for preparing and analysing large numbers of samples and standards.
- Variability in the significance of the appearance or concentration of metabolites, including crossover among biological pathways.

Many of these issues are common to other types of high-throughput biochemical analysis. The difference is that in most assays investigators know what they are looking for, and can confidently ignore the chaff. From a naïve genomics perspective, one is never certain of the identity and significance of unknowns, particularly given their widely variable concentrations.

Sources of variability for metabolomics differ somewhat from conventional biochemical analysis. Many are related to the presence of significant low-abundance metabolites and the need to multiplex the characterisation of hundreds of analytes:

- Instrumental: mass and retention time stability,

By Dr Ian D. Jardine



robustness and stability of detector response, sufficient resolution to resolve isobaric interference.

- Chemical/data processing: background from column/solvents, multiple signals per compound, setting detection thresholds.
- Biological: different response rates to a stimulus between individual subjects, organism's stress and feeding status, other health factors.
- Study design: proper controls and randomised sampling/analysis to minimise systematic errors, sampling, sample preparation and storage.
- Statistical analysis: limited sampling, over-fitting data.

Several analytic platforms have emerged for conducting metabolomics studies. All have advantages and disadvantages, but they share several fundamental capabilities:

- Reproducibility and comprehensiveness to incorporate known and unknown metabolites.
- Separation of analytes from complex matrices.
- Analyte identification post-separation.
- Reliance on unbiased methodology to avoid missing critical metabolites.
- Biological understanding².

Modern metabolomics relies strongly on unequivocal compound identification. Lacking this information forces reliance on pattern-matching, which does not provide the biological understanding, the connectedness between proteomics and genomics, that is the very rationale for metabolomics.

Significance of LC-MS

Metabolomics can be reduced to the analytical biochemistry of small molecule metabolites. While metabolomics has benefited from advances in analytical instrumentation and information technology, many of the same general methods used two decades ago – when the field was known as 'metabonomics' – remain valuable today although not always for the same reasons.

Early metabolomics work employed nuclear magnetic resonance (NMR) spectrometry, principally to analyse amino acids in the blood of newborns. NMR worked well enough for quantifying 20 or 30 compounds in blood or urine, and providing structural and concentration data. But NMR sensitivity is far too low for low-abundance metabolites, and lacks a separation component. Today, NMR is reserved for studying known, high-abundance targets.

Progress in modern metabolomics has been fuelled mostly by the application of liquid chromatography and high-resolution mass spectrometry. These techniques have enabled the routine, simultaneous characterisation of several hundred metabolites – an order of magnitude improvement over the best NMR method. Metabolon, Inc, which specialises in metabolomics studies, routinely quantifies up to 800 metabolites in urine or plasma, and is pushing well beyond 1,000 analytes.

Ion chromatography is also useful for separating highly polar compounds, but is not perfectly compatible with MS. Buffered mobile phases that optimise IC separations interfere with metabolite

ionisation within the electrospray ionisation source. When introduced, techniques providing low ionisation suppression with superior separation, will breathe new life into IC-MS for metabolomics but will not displace the consensus metabolomic analytical platform, LC-MS.

For the time being comprehensive metabolomics will continue to employ several orthogonal methods. Metabolon, for example, analyses its samples with positive ion LC/MS, negative ion LC/MS, and GC/MS (underivatized). While overlap occurs with this approach, the three methods together facilitate the analysis of close to 1,000 compounds.

LC/MS/MS is a very rapid technique, as short as 15 minutes per analysis using the latest UHPLC separations and rapid MS. Generic LC/MS has made NMR all but obsolete, and demoted ion chromatography-MS to second-tier status.

Metabolomics capable of 800-metabolite analysis and higher demands very high-performing liquid chromatography that is robust across methods and laboratories and reproducible across samples. Due to the complexity of metabolomics samples, the chromatographic method must provide exactly the same retention times from run to run. By contrast, retention times are not as critical when analysing a peptide digest by LC/MS. If retention time or peak shapes differ between runs the MS will sort things out. That is not possible when studying hundreds of small molecules, only 60% of which are known.

Despite the desirability of high-performance LC/MS/MS, it is unlikely that one approach will exclusively serve the analytical needs of metabolomics.

While MS detection is highly desirable due to its sensitivity and ability to deliver unequivocal molecular masses, other supportive or confirmatory detection modes are useful for niche applications. In a pilot study, electrochemical detection, which is specific for redox metabolites, was coupled with MS to profile plasma redox metabolites in uremia³. Similarly, charged aerosol detection provides a 'universal' analysis mode with quantitation.

Recent successes/analytical approaches

A genome-wide association study (GWAS)⁴ conducted by Metabolon, Inc and academic partners in Europe defines the state of the art for metabolomics and established a strong association between metabolite concentrations and genetics.

Researchers have used GWAS to identify genetic risk factors for disease, but studies had previously been limited by size and lack of biological under-

standing of the results. The Metabolon paper provides deep metabolomic understanding for cardiovascular and kidney disorders, type 2 diabetes, cancer, gout, venous thromboembolism and Crohn's disease.

The investigation was based on serum collected from two large epidemiologic studies, one German and one British. Researchers identified 37 genetic *loci* associated with blood metabolite concentrations, of which 25 showed unusually high effects. Twenty-three *loci* described new genetic associations with metabolic traits while 14 extended knowledge of known associations. In all, researchers analysed more than 250 metabolites, covering 60-plus biochemical pathways, in samples from 2,820 individuals at 24 minutes per sample. The study, according to Metabolon, "establishes biochemistry, perhaps the most easily measured genetic trait, as an intermediate to provide a biological link that contributes to the understanding of the genetic effects and more effectively impacts discovery and development of individualised biomarkers and therapies⁵".

Professor Dean Jones, in the Department of Medicine at Emory University, explained the idea of the 'exposome' at the 2012 Metabolomics Society annual meeting in Washington, DC. He defines exposome as "all exposures from conception onwards, including those from lifestyle, diet and the environment" – the sum total of an organism's experiences. Exposome factors include diet, exercise, prescription drugs, etc.

Jones employs a data processing algorithm, adaptive processing LC/MS (apLCMS), which when applied to dual chromatography with high-resolution mass detection measures 44,000 ions in 20 minutes from human plasma. This approach enabled Jones in one year to double the number of ions detectable by LS/MS from approximately 20,000 to 45,000. Most are unknowns and include analyte fragments, but those peaks confirmed by MS correspond to more than half of KEGG human metabolites, drugs, environmental chemicals and products of the microbiome. Jones writes that "high-resolution metabolomics, with resolution comparable to that for the human genome, can quantify life-course environmental exposures from the prenatal period onwards as conceptualised in the exposome".

Similarly, in a study of the impact of the exposome to *drosophila* longevity, Jones detected more than 20,000 ions and demonstrated significant differences in ion prevalence corresponding to age and longevity between the sexes and genetic lineages.

In July, 2012, the National Heart, Lung and Blood Institute awarded Emory University a \$5 million grant to employ mass spectrometry to identify and validate new cardiovascular disease risk biomarkers based on metabolites. Professor Jones will use the grant to study metabolites in blood correlating with blood vessel function and coronary artery disease.

Nature/metabolon paper

Scientists who were confounded by proteomics in the early 1990s, then hopeful about genomics, have been disappointed to learn that unravelling the relevance of the genome has taken more time and effort than anyone imagined. A study led by investigators at Metabolon, Inc⁶ demonstrated, for the first time, the potential for using metabolomics to reach back into the genome and make sense out of it.

The more metabolites measured, the more complete the picture of the relevance of genomic variations. This is what drives the need to analyse and identify even more compounds, and of greater variety, simultaneously. At the current rate, we expect the number of metabolites falling into the category of routine analysis to reach 1,500-2,000. These will require even better analytical methods and information systems as well. Much work remains on interpreting the fragmentation of unknown compounds to allow cataloging them in compound libraries.

One such development will be the routine use of accurate mass MS. Remarkably, Metabolon's work was accomplished using conventional, nominal mass LC/MS/MS. The company has recently upgraded to an orbitrap mass instrument, which is expected to increase its metabolite capabilities significantly. Orbitrap MS has been around for nearly 90 years. In their current embodiment, the instruments are capable of a mass accuracy of 1ppm, resolution up to 200,000 and dynamic range capabilities of 5,000.

According to Dr Bruce S. Kristal of Brigham and Women's Hospital, Boston, Massachusetts, high-resolution, accurate-mass orbitrap LC-MS has quadrupled actionable information obtained from his experiments.

Lipidomics

Lipidomics – the study of the lipid fraction of the metabolome – is increasingly viewed as a distinct sub-discipline. Approximately 20% of individuals performing metabolomics primarily study lipids.

Lipids are remarkably diverse and include species with widely diverse chemical structures,

polarities and potential to ionise. As such, lipids present unique challenges related to their physical and chemical similarities but structural heterogeneity. Just one category – triglycerides – encompass numerous combinations of chain lengths, sites and degrees of unsaturation and stereochemistry. One sample may contain several lipid categories.

Baseline separation of structurally-related lipids by LC is often problematic. One potential solution is to let the mass spectrometer perform the separation through a technique known as infusion. It involves preliminary fractionation of major lipid classes by LC, followed by direct elution into the ionisation source of the MS. With the assistance of deconvolution software, high-sensitivity, highly-resolving MS will sort out closely-related lipids through their unique fragmentation patterns, most (but not all) of which are already catalogued.

Lipidomics case studies

The research group led by Dr Kristal investigates how changes in the fat components and glycemic index of diet reflect in an organism's mitochondrial lipidome, and ultimately its health. The researchers employ LC/MS profiling to identify the sets of mitochondrial lipids in liver and other tissue that serve as dietary biomarkers to monitor long-term disease risk.

Investigators developed LC-MS lipidomics-profiling methodology that is remarkably robust, high-throughput, broadly applicable, untargeted, qualitative and semi-quantitative. Its success relies on high-resolution, accurate-mass (HR/AM) and higher-energy collisional dissociation (HCD) fragmentation capabilities of an orbitrap LC-MS and powerful software for data reduction and differential expression analysis.

Diversity in lipid side-chains is structurally and functionally significant in mitochondrial physiology. Hundreds of lipid species exist in a single sample. Many of these molecules are uncharacterised, requiring accurate mass peaks for positive identification.

In Dr Kristal's approach, lipids are profiled using HR/AM full-scan measurements of the parent molecule and HCD fragmentation data generated from additional experiments or alternating scans in the same experiment. Experiments were run in both positive and negative ionisation mode for comprehensive coverage.

Because it resolves ions of the same nominal mass, HR/AM deconvolutes the heterogeneity of lipid samples. The results are class-specific and lipid-specific diagnostic fragments, with mass accuracies of 3ppm or better – sufficient to reconstruct lipid molecules. Since HCD fragments all ions in the

chromatogram, data may be reanalysed to characterise lipids that were not initially of interest.

Recent publications⁷⁻⁹ by Dr Kristal's group demonstrate the robust and reproducible performance of this unbiased method, and its ability to profile many more lipids in a much shorter time-frame than was ever thought possible.

A research group led by Maria Carvalho at the Max Planck Institute in Dresden, Germany, examined the effects of diet and development on the *Drosophila* lipidome¹⁰. This study, which incorporated an element of exposomics, used high-resolution MS to measure the molar concentrations of 250 lipids belonging to 14 molecular classes. Carvalho looked at six tissues, at 27 developmental stages, of animals raised on four different diets.

She found that dietary lipids most profoundly affect phospholipid composition throughout the animal. Furthermore the fruit fly "differentially regulates uptake, mobilisation and tissue accumulation of specific sterols", and that lipid metabolism shifts significantly between larval and pupal stages. Furthermore, dietary sterols form into lipoproteins with varying efficiencies and tissue-specific accumulation patterns. These observations suggest metabolomic variation within tissues, with perhaps some organ sources representing more significant targets for metabolomic analysis for certain conditions.

Conclusion

At this stage of its evolution metabolomics is rightly concerned with numbers of metabolites. Increasing the number of accessible metabolites is essential for establishing a comprehensive metabolome for organisms of interest, for establishing their relevance to the state of the organism, and for correlating metabolites with genomic changes. One can think of this as the 'discovery' stage for metabolomics. Once the critical molecules for a particular system are identified, it is likely that follow-on studies in medical diagnostics, pharmacology and 'omics-based disease studies will focus on a very small subset of molecules, perhaps as few as 20. At this stage investigators can switch to higher-throughput instrumentation such as triple quad MS. At that point metabolomics will have entered the mainstream.

The future of metabolomics is intimately interwoven with development of instrumentation and software capable of rapidly processing many hundreds, if not thousands, of metabolites in exploratory studies, and subsequently homing in on several dozen targets for routine analysis. Characterising unknown metabolites is another

top priority, as most peaks in a metabolomic LC analysis are unknowns.

Furthermore the diagnostic and prognostic capabilities of metabolomics, and its relevance to drug development, will depend on algorithms for relating metabolomic differences to the activation state of genes, and the biochemistry of metabolic precursors such as proteins. The trajectory of the evolution of instruments and software suggests routine elucidation of metabolomes consisting of 2-3,000 distinct molecules is less than five years away.

Reducing metabolomics to the status of 'supercharged' analytical biochemistry – basically an engineering problem – is tempting but an oversimplification. Most biochemical analysis in health-related fields involves known compounds and well-characterised impurities. Metabolomics' greatest present and, presumably future achievement, is its concern with identifying all relevant metabolites quickly and quantitatively and uncovering their relevance. To developers of analytical instrumentation, these needs translate to higher resolution, sensitivity, reproducibility and method robustness.

Over the years several analytical platforms have served the immediate needs of metabolomics research. NMR, for example, was sufficient when the number of metabolites of interest measured in the tens. As the field evolved other methods, such as ion chromatography, assumed niche status. Ultimately, the broadest functionality and speed will be achieved through LC separations and tandem MS for compound identification. LC/MS has its limitations as well, but it is the platform most likely to achieve the most comprehensive results. **DDW**

Dr Ian D. Jardine is Vice-President and Chief Technology Officer, Life Sciences Mass Spectrometry, Thermo Fisher Scientific. Previously, he was Vice President of Global R&D for Thermo Fisher Scientific. He joined the company in 1988 as director of analytical biochemistry, and then director of marketing for mass spectrometry. He then became President of the MS business. Dr Jardine earned his doctorate in organic chemistry/mass spectrometry from the University of Glasgow and then completed a fellowship at the Johns Hopkins University School of Medicine. Prior to joining Thermo, he held an Assistant Professorship at Purdue University and a Professorship at the Mayo Clinic and Mayo Medical School.

References

- 1 Phenomenom Discoveries. Delivering on the Promise of Metabolomics. Business Briefing: Pharmtech, 2004.
- 2 Phenomenom Discoveries. Delivering on the Promise of Metabolomics. Business Briefing: Pharmtech, 2004.
- 3 Acworth, I et al. Metabolomic Profiling of Uremia With an LC-EC Array-MS Parallel Platform. Application note, Thermo Fisher Scientific (and references therein).
- 4 Suhre, K et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477, 54-60, 1 September 2011.
- 5 Metabolon press release. Accessed at: <http://www.metabolon.com/news/PressReleases.aspx?year=2011> on July 31, 2001.
- 6 Suhre, K et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*. 2011 Aug 31;477(7362):54-60.
- 7 Bird, SS et al. Qualitative characterization of the rat liver mitochondrial lipidome using LC-MS profiling and high energy collisional dissociation (HCD) all ion fragmentation. *Metabolomics*. 2012.
- 8 Bird, SS et al. Serum Lipidomics Profiling Using LC-MS and High-Energy Collisional Dissociation Fragmentation: Focus on Triglyceride Detection and Characterization. *Analytical Chemistry*. 2011.
- 9 Bird, SS et al. Lipidomics Profiling by High-Resolution LC-MS and High-Resolution LC-MS and High-Energy Collisional Dissociation Fragmentation: Focus on Characterization of Mitochondrial Cardiolipins and Monolysocardiolipins. *Analytical Chemistry*. 2010.
- 10 Carvalho, M et al. Effects of diet and development on the *Drosophila* lipidome. *Molecular Systems Biology* 8:600 (2012).