What is metabolomics all about?

Ute Roessner1,2 and Jairus Bowne1
1Australian Centre for Plant Functional Genomics, The University of Melbourne, Victoria, Australia, and 2Metabolomics Australia, School of Botany, The University of Melbourne, Victoria, Australia

BioTechniques 46:363-365 (April 2009 Special Issue) doi 10.2144/000113133
Keywords: metabolomics; systems biology; phenotyping

What is metabolomics? 

The rapidly emerging field of metabolomics combines strategies to identify and quantify cellular metabolites using sophisticated analytical technologies with the application of statistical and multi-variant methods for information extraction and data interpretation. In the last two decades, huge progress was made in the sequencing of a number of different organisms. Simultaneously, large investments were made to develop analytical approaches to analyze the different cell products, such as those from gene expression (transcripts), proteins, and metabolites. All of these so-called ‘omics approaches, including genomics, transcriptomics, proteomics, and metabolomics, are considered important tools to be applied and utilized to understand the biology of an organism and its response to environmental stimuli or genetic perturbation.

Metabolites are considered to “act as spoken language, broadcasting signals from the genetic architecture and the environment” (1), and therefore, metabolomics is considered to provide a direct “functional readout of the physiological state” of an organism (2). A range of analytical technologies has been employed to analyze metabolites in different organisms, tissues, or fluids (for review see Reference 3). Mass spectrometry coupled to different chromatographic separation techniques, such as liquid or gas chromatography or NMR, are the major tools to analyze a large number of metabolites simultaneously. Although the technology is highly sophisticated and sensitive, there are still a few bottlenecks in metabolomics. Due to the huge diversity of chemical structures and the large differences in abundance, there is no single technology available to analyze the entire metabolome. Therefore, a number of complementary approaches have to be established for extraction, detection, quantification, and identification of as many metabolites as possible (3,4).

Another challenge in metabolomics is to extract the information and interpret it in a biological context from the vast amount of data produced by high-throughput analyzers. The application of sophisticated statistical and multi-variant data analysis tools, including cluster analysis, pathway mapping, comparative overlays, and heatmaps, has not only been an exciting and steep learning process for biochemists, but has also demonstrated that current thinking needs to change to deal with large data sets and distinguish between noise and real sample-related information.

In addition, and still without consensus in the metabolomics community, is the question, “How do we deal with data that don’t make biological sense based on literature and common knowledge?” We are only beginning to even assume where metabolomics, together with the other ‘omics technologies, is going to lead us: Will we find more answers to our questions or will it bring more questions requiring more answers?

Potential and applications of metabolomics

There are four conceptual approaches in metabolomics: target analysis, metabolite profiling, metabolomics, and metabolic fingerprinting (5). Target analysis has been applied for many decades and includes the determination and quantification of a small set of known metabolites (targets) using one particular analytical technique of best performance for the compounds of interest. Metabolite profiling, on the other hand, aims at the analysis of a larger set of compounds, both identified and unknown with respect to their chemical nature. This approach has been applied for many different biological systems using GC-MS, including plants (6), microbes (7), urine (8), and plasma samples (9). Metabolomics employs complementary analytical methodologies, for example, LC-MS/MS, GC-MS, and/or NMR, in order to determine and quantify as many metabolites as possible, either identified or unknown compounds. The fourth conceptual approach is metabolic fingerprinting (or footprinting for external and/or secreted metabolites). Here a metabolic “signature” or mass profile of the sample of interest is generated and then compared in a large sample population to screen for differences between the samples. When signals that can significantly discriminate between samples are detected, the metabolites are identified and the biological relevance of that compound can be elucidated, greatly reducing the analysis time.

Since metabolites are so closely linked to the phenotype of an organism, metabolomics can be used for a large range of applications, including phenotyping of genetically modified plants and substantial equivalence testing, determination of gene function, and monitoring responses to biotic and abiotic stress. Metabolomics can therefore be seen as bridging the gap between genotype and phenotype (5), providing a more comprehensive view of how cells function, as well as identifying novel or striking changes in specific metabolites. Analysis and data mining of metabolomic data sets and their metadata can also lead to new hypotheses and new targets for biotechnology.

Metabolomics and evolution

To date, most research in evolution is based on the construction of phylogenetic trees of species using sequences of genomes, genes, mRNA, and/or proteins. However, the correlation of gene and protein expression is low and that between gene expression and metabolites even lower. However, metabolites, especially secondary metabolites, are extremely important for most organisms to defend themselves from stressful environments or predators. Although primary metabolites involved in central metabolism can be used to determine nutritional and growth status, secondary metabolite profiles may better reflect the differentiation of species and their complex response to environmental factors and other organisms. The suite of secondary metabolites in an organism can be astonishingly complex, and while certain
compounds may be found in different organisms, a vast number of compounds are very species-specific. Secondary metabolites are therefore considered as potential markers for taxonomy and phylogenetics (10).

Probably the best and most exciting applications of metabolomics tools to distinguish different fungal species have been summarized by Smedsgaard and Nielsen (11). Direct infusion electrospray mass spectrometry (DiMS) was used for rapid chemical classification of filamentous fungi. Crude fungal extracts of a number of different subspecies were directly injected into a mass spectrometer and resulting mass profiles compared using chemometric analysis tools (4). More than 80% of the analyzed species could be classified based on their mass profile compared with a conventional phenotypic identification.

In our laboratory, we are using metabolomics to determine novel mechanisms for adaptation and tolerance of plants to abiotic stresses, such as drought, salinity, cold, frost, and mineral deficiencies or toxicities (www.acpfg.com.au). Our major plants of interest are cereals, such as barley and wheat, but we also look at model plants or plants that are known to exhibit a greater level of tolerance to a certain stress condition. The comparison of different species’ responses to different stresses demonstrated that there are a number of responses that are stress- and/or plant-specific and a few that are common between stresses and/or plants. Therefore, we decided to compare the levels of metabolites in the leaves of four different species: the moss Physcomitrella patens, the model plant Arabidopsis thaliana, and the crop plants Hordeum vulgare L. and Triticum aestivum L. We compared the metabolite levels in unstressed plants in order to investigate if there is a correlation between tolerance levels and metabolite profiles. We used GC-MS to profile ~140 known metabolites (12) and normalized the data for comparison between the species. Multivariate analysis of the resulting data set using principle component or hierarchical cluster analysis demonstrated that the metabolite profiles of the four species are very distinct, with barley and wheat leaf profiles being the most similar (Figure 1A). The first principle component separated wheat and barley from the other two species, accounting for 58% of the variability of the whole data set. Figure 1B shows a heat map representation of the same data set comparing the levels of metabolites of the different species. Most metabolites are at a much lower level in moss and Arabidopsis compared with barley and wheat (raw data not shown). There are a few exceptions, including urea, glycerol, tyramine, allantoin, tocopherol, xylitol, fucose, and inositol, which are found at much higher levels in the moss than in all other species. This raises the question of whether those metabolites may be responsible for the high tolerance of moss to abiotic stresses (13).

This example demonstrates the potential of metabolomics to be used for the identification and classification of organisms. The examples mentioned above may only be the start. We believe it is worthwhile to pursue a more systematic study to compare metabolite profiles between a larger number of organisms using complementary analytical approaches to cover as many metabolites as possible, and to investigate if the metabolite profiles are related to phylogenetic and evolutionary relationships between organisms. This type of study may result in novel insights in the evolution of pathways, survival mechanisms, and life in general.

**Metabolomics in a systems biology context**

As we have described in this paper, metabolomics aims ideally at the analysis of all small molecules in a cell. This is only a portion of the cellular products within a cell. For a systems biology approach, metabolomics only provides the measurement of a portion of all elements in a biological system. Yet, systems biology comprises not
only the ability to measure all elements of a system, such as DNA, mRNA, proteins, metabolites, and structural elements such as cell walls and membranes, but also to determine the relationship of those elements to one another as part of the system’s response to environmental or genetic perturbation. After integrating all of the different levels of information, the intention is to model the behavior of the system using computational methods that may allow the description of the behavior of the system under any kind of perturbation. A systems biology approach requires biologists, physicists, computer scientists, engineers, chemists, and mathematicians to learn a common language that allows them to communicate with each other. Another important requirement for a successful systems biology approach is creating an environment that provides access to all of the high-throughput platforms needed to obtain and measure the properties and elements of the system of interest. Also, an effective systems biology approach must offer the opportunity and scale for fast development and employment of new global technologies and powerful computational tools that allow gathering, classifying, analyzing, integrating, and ultimately, modeling of biological information.

Systems approaches to human diseases, such as cancer, cardiovascular disease, and obesity, will give the opportunity to greatly facilitate the success of selecting a novel target for treatments and drug development. In the future, systems biology may enable us to develop new approaches in medicine that will be predictive, preventative, and personalized. The aim would be to achieve the ability to determine a probabilistic health history for each individual, and within that framework, systems biology will be a strategy for the discovery and development of new therapeutic as well as preventative drugs.

In summary, studying the response of various organisms to different stresses and environments at the genetic, transcript, protein, and metabolite levels using different methods and comparing these results with those of other organisms will strengthen their integration into a systems biology framework. As the framework develops, the greater synergy between organisms will provide a much clearer picture of the function of cells, organs, and organisms, bringing us closer to understanding their roles in nature.

Acknowledgments

The authors thank the Australian Centre for Plant Functional Genomics for funding. We would like to give special thanks to Tim Erwin, who developed the heatmap presented in Figure 1B.

The authors declare no competing interests.

References


Received 9 February 2009; accepted 10 February 2009.

Address correspondence to Ute Roessner, Metabolomics, Australia School of Botany, The University of Melbourne, 3010 Victoria, Australia. Email: u.roessner@unimelb.edu.au