The role of metabolomics in systems biology

Jens Nielsen and Michael C. Jewett

Abstract

The metabolome comprises the complete set of metabolites, the non-genetically encoded substrates, intermediates, and products of metabolic pathways, associated to a cell. By representing integrative information across multiple functional levels and by linking DNA encoded processes with the environment, the metabolome offers a window to map core attributes responsible for different phenotypes. Given increasing demand to quantitatively identify the metabolome and understand how trafficking of metabolites through the metabolic network impact cellular behavior, metabolomics has emerged as an important complementary technology to the cell-wide measurements of mRNA, proteins, fluxes, and interactions (e.g. protein-DNA). Metabolomics is already a powerful tool in drug discovery and development and in metabolic engineering. While maintaining these strengths, the field promises to play a heightened role in systems biology research, which is transforming the practice of medicine and our ability to engineer living organisms.

1 Metabolomics

Metabolome analysis, originally proposed by Oliver et al. in 1998, seeks to identify and quantify the entire collection of intracellular and extracellular metabolites. Conceptually, there are two basic analytical methodologies used in metabolomics (Fig. 1) (Villas-Bôas et al. 2005a). Mainly exploited for classification, *metabolite profiling* strategies investigate qualitative scanning of a large number of metabolites (i.e. more than 100). Here, the pattern of metabolites (or even spectra from chromatography or mass spectrometry) is used to find discriminatory elements via high-throughput detection followed by data deconvolution methods (Kell 2004; Goodacre et al. 2004). Metabolite profiling comprises of metabolic fingerprinting, which covers the endometabolome (intracellular metabolites), and metabolic footprinting, which covers the exometabolome (metabolites in the growth media or extracellular fluid) (Fig.1). The other general method used in metabolomics is *target analysis*. Here, absolute, or at least semi-quantification and unambiguous detection of metabolites are achieved. While historically target analysis has been reserved for interrogating relatively small numbers of metabolites (e.g. less than 20), new developments enable quantitative analysis of more expanded metabolome coverage (e.g. Villas-Bôas et al. 2005b; Soga et al. 2003; Roessner et al. 2000).

Topics in Current Genetics, Vol. 18
J. Nielsen, M. C. Jewett (Eds.): Metabolomics
DOI 10.1007/4735_2007_0228 / Published online: 13 June 2007

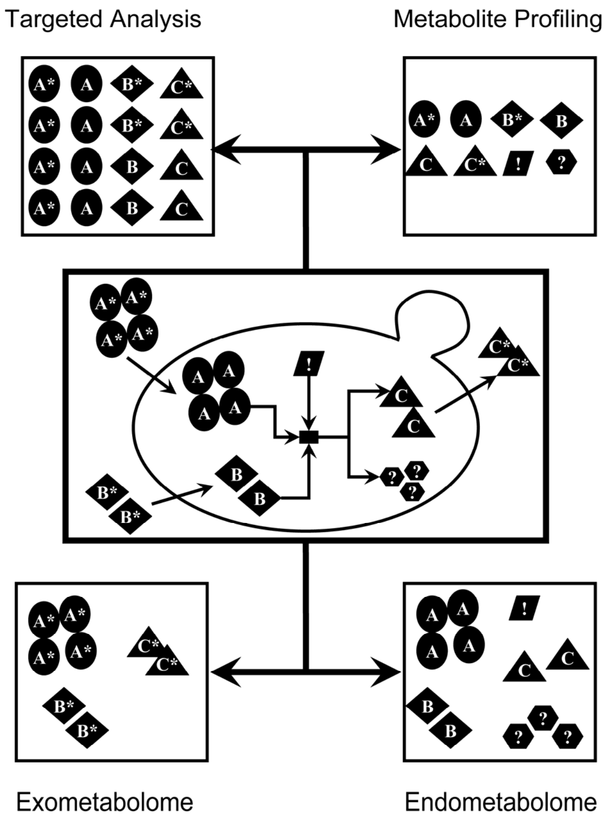

Fig. 1. Strategies for metabolome analysis. The metabolome is comprised of two parts, the endometabolome (intracellular metabolites) and the exometabolome (extracellular metabolites). Metabolome analysis seeks to identify cellular metabolites through targeted analysis (identification and quantification of pre-defined metabolites) or metabolite profiling (scanning of all metabolites identified by a specific analytical technique). Extracellular metabolites: A*, B*, and C*. Intracellular metabolites: A, B, C, !, ?. Note: ! and ? are unidentified metabolites.

A variety of analytical platforms have been utilized for metabolite detection (Villas-Bôas et al. 2005a). While most quantitative strategies couple a separation technique (e.g. capillary electrophoresis (CE), liquid chromatography (LC), and gas chromatography (GC)) with mass spectrometry (MS) or NMR based detection, it is not uncommon to only make use of direct infusion MS for metabolite profiling. From a practical standpoint, our inability to quantitatively extract and detect highly diverse families of metabolites in their original state over a large dynamic range with a single or even limited-set of analytical techniques makes analyzing the complete set of all metabolites associated to a cell impossible. Thus, metabolomics is more appropriately used to describe an "area of science rather than an analytical approach" (Villas-Bôas et al. 2005c).

2 Applications of metabolomics

The application of metabolomics has been widely pursued. Historically, metabolite profiling has been used for medical and diagnostic purposes (Horning and Horning 1971; Gates and Sweeley 1978) as well as strain classification and characterization (Frisvad and Filtenborg 1983). The latter has been particularly true for fungi and plants, which have extremely diverse metabolic landscapes. As an example, detection and quantification of mycotoxins from fungi has been a focal point for characterization studies (for overview of mycotoxins see Bennett and Klich 2003). The increase in public awareness over the safety of food and feed during the last several years has led to the establishment of many new laws and guidelines with respect to mycotoxins (FAO 2004). For the latest developments in analysis and detection methods, we refer the reader to a review detailing this topic (Krska et al. 2005). A key development is that unconventional biosensor methods, which typically rely on metabolite profiling; such as, electronic nose or tongue technology, have a strong potential to mature into key techniques for the detection of mycotoxins and toxigenic fungi (Logrieco et al. 2005).

Metabolome analysis is also an important tool in functional genomics, revealing the roles of genes from comprehensive analysis of the metabolome (Fiehn 2001; Trethewey 2001). For example, metabolite profiling and target analysis have been effectively used to classify molecular signatures responsible for the phenotype of silent and unknown mutations (Raamsdonk et al. 2001; Allen et al. 2003; Weckwerth et al. 2004). In one illustration, Weckwerth et al. (2004) demonstrated the application of target analysis, using GC-TOF-MS for quantification of more than 1,000 metabolites, to characterize the features responsible for a silent plant phenotype. Exploiting statistical tools, metabolic correlations were determined between identified metabolites (e.g. trehalose-erythritol) and used to reveal network maps which suggested hypotheses for the impact of an exact phenotype on carbohydrate and amino acid metabolism.

Hierarchical metabolomics, first reported in plants, is also well suited to guide targeted analysis of metabolism (Catchpole et al. 2005). Catchpole et al. (2005) used metabolome coverage of conventional and genetically modified (GM) potato crops to reveal that, apart from anticipated engineered differences, metabolic compositions were comparable among several types of cultivars. First, they applied metabolic fingerprinting of potato tuber extracts to classify several potato genotypes. Second, target analysis of defined and specific classes of metabolites using LC-MS and GC-TOF-MS was exploited to identify specific fructans responsible for the global classifications. Finally, data analysis tools were applied to remove the influence of anticipated differences in the GM crops and show that the GM and conventional crops were within the variation observed from investigating several unmodified metabolic phenotypes. Hierarchical analysis provides a rapid and relatively inexpensive screen for many functional genomics and screening applications.

3 The role of metabolomics in systems biology

In addition to finding utility in drug discovery, strain classification and functional genomics, metabolomics is emerging as a powerful tool in systems biology (Jewett et al. 2006; Wang et al. 2006). Systems biology is the quantitative study of an organism, viewed as a complex web of interacting and interchanging molecular participants (DNA, mRNA, proteins, and metabolites) and their environment. The overarching vision is that studying defined biological systems as a whole, through the combination of mathematical modeling and experimental biology, will provide insights into cellular behavior that are not apparent from investigating the components alone (Nielsen and Jewett, submitted). As a result of pioneering advances in genome sequencing, today's systems biology has dramatically enhanced our ability to study the relationships among active molecular players of the cell for describing and predicting cellular behavior. It promises to transform the practice of medicine and our ability to engineer living organisms by facilitating drug discovery, treating disease, and improving bioprocesses (Hood and Perlmutter 2004; Stephanopoulos et al. 2004; Weston and Hood 2004).

Realizing the promise of systems biology is hampered by the integrated and complex nature of cellular networks. First, there is not a one-to-one correlation between genes and metabolites (Nielsen and Oliver 2005). Not only can metabolites participate in many different biochemical reactions, but also multiple mRNAs can be formed from one gene, multiple proteins from one mRNA, and multiple metabolites from one enzyme. Second, interactions between proteins and small molecules, translational regulation, and other post-transcriptional mechanisms weaken the linkage between transcriptional state and metabolic phenotype. Third, the highly connected nature of cellular networks means that small perturbations rapidly traverse the cellular landscape; hence, impacting the overall functional operation of the network. Based on the yeast *Saccharomyces cerevisiae* genome-scale metabolic model containing about 800 metabolites and 1200 enzymatic reactions, for example, the average path length to get from any metabolite or enzyme to any other metabolite or enzyme is only about 5 (Patil and Nielsen 2005).

Given their central role as signals capturing information from all functional levels of the cell (Nielsen 2003) and also as nodes in dense metabolic networks (Jewett et al. 2006), determining concentrations of specifically identified metabolites is core to the systems biology agenda. We envision that the most powerful approach for using metabolomics data for systems biology is within the context of complex interactions, cellular pathways, molecular participants, and environmental stimuli that they connect. We believe that metabolic networks, which are well established, provide an effective and efficient scaffold for organizing systems biology data. Emerging computational strategies that exploit topological information from genome-scale metabolic models have already proven to be a compelling approach for inferring co-regulated cellular network structures (Patil and Nielsen 2005; Çakir et al. 2006).

Examples using metabolomics in systems biology mainly focus on quantifying metabolite levels and flows in primary metabolism. By relying on predefined connections between genetic sequences and metabolites, the information observed by acquiring a snapshot of the cellular metabolic composition is upgraded. Elucidating a metabolic image of the central carbon metabolism has provided insights for linking normal anabolic and catabolic trafficking with other branches of metabolism. The use of LC-MS, for example, has been exploited to map metabolic activity and flexibility through dynamic analysis of intracellular metabolites during the yeast cell-cycle (Wittmann et al. 2005) and the effect of culture age on metabolite pools (Mashego et al. 2005). Even though quantification of biomolecules involved in central metabolism offers many insights into key nodes of metabolism, other applications have also laid the foundation for target analysis of metabolic hubs that lay one step beyond the central metabolism. To quantify metabolites containing an amino or carboxylic acid group, Villas-Bôas et al. (2005b) applied a sensitive GC-MS method coupled to a statistical data-mining strategy for the integrated analysis of clearly identified and quantified intra- and extracellular metabolites in *S. cerevisiae* (approximately 60). By isolating statistically significant differences among metabolite levels from four biological conditions, they observed discriminatory metabolic features which hinted at the potential for future integration with comparative omic analyses. Highlighting the generality of this method, Panagiotou et al. (2005) have utilized this analytical approach to determine the influence of aerobic and anaerobic cultivation conditions on the metabolic state of *Fusarium oxysporum*.

Equally important in guiding a systems-level understanding of the overlapping layers of global regulation and network flexibility are efforts to experimentally measure the flow of material through central metabolism. Characterization of metabolic operation is achieved by using ^{13}C-labeled substrates followed by determination of characteristic metabolite patterns which can indicate directional flow (Sauer 2006). The most general approach uses proteinogenic amino acid analysis to infer labeling patterns and flux distributions. However, the application of rapid sampling and quenching has recently been applied to directly analyze intracellular metabolites from *S. cerevisiae* without being impeded by the high metabolic turnover rates (van Winden et al. 2005). This approach generates direct data without inference; however, caution must be exercised due to the rapid dynamics of exchange between metabolites and amino acids incorporated into cellular proteins (Grotkjær et al. 2004).

High-throughput efforts for comparative flux analysis offer an unprecedented view of the rigidity, flexibility, and performance of metabolic networks. In one illustration, Blank et al. considered flux data from over 30 mutants in *S. cerevisiae* to investigate potentially flexible fitness reactions during growth on glucose (Blank et al. 2005). Combination of measurements with mathematical modeling revealed that metabolic network robustness to single gene knockouts was principally a result of genetic redundancy, duplicate genes, with alternative pathways, redirection of carbon flow, having less importance. This approach was taken further in a larger scale, systematic flux analysis of 137 null mutants of *Bacillus subtilis* (selected from all major functional categories) on its preferred substrate (Fischer and Sauer 2005). As in the previous illustration, this strategy enabled identification of fundamental design principles of *in vivo* network operation. A

Springer

http://www.springer.com/978-3-540-74718-5

Metabolomics
A Powerful Tool in Systems Biology
Nielsen, J.; Jewett, M.C. (Eds.)
2007, XVI, 284 p. 53 illus., 13 illus. in color., Hardcover
ISBN: 978-3-540-74718-5