
PART I

CONCEPTS AND METHODOLOGY

1

METABOLOMICS IN FUNCTIONAL GENOMICS AND SYSTEMS BIOLOGY

BY JENS NIELSEN

This chapter gives a brief introduction to the field of metabolomics and puts this in perspective of the current development in molecular biology, where genomics have resulted in a move from a reductionistic analysis of biological systems (or even sub-systems) to a systems (or global) view on the function of biological systems. Thus, the chapter serves as an introduction to the textbook.

1.1 FROM GENOMIC SEQUENCING TO FUNCTIONAL GENOMICS

In 1992 the first nucleotide sequence of a complete chromosome was obtained, namely the DNA sequence of chromosome III of the yeast *Saccharomyces cerevisiae*, and around the same time efforts to sequence the human genome were initiated. In 1995 the first complete genome was sequenced, namely that of the pathogenic bacterium *Haemophilus influenzae*, and in 1996 the complete genomic sequence of the yeast *S. cerevisiae* was released. Since then there has followed genomic sequences of many different organisms (Figure 1.1), and currently the number of sequences entered into GenBank is doubled every 10 months. Genomic sequences provide the blueprint for cellular function, and the complete set of genes within a genome basically defines a functional space for the organism. However, in order to further define this functional space it is necessary (1) to know the function of all the proteins and (2) to know the relationship between which genes are expressed (or which proteins are present) at different environmental conditions. Since the first complete genome was released,

Timeline of key genomics events

↑	2006	More than 300 complete genomes available
	2000	First Large compendium of protein-protein interactions (<i>S. cerevisiae</i>)
	2000	Human genome completed
	1998	First complete multicellular organism (<i>C. elegans</i>)
	1997	First genome wide transcription analysis (<i>S. cerevisiae</i>)
	1996	First complete eukaryotic genome (<i>S. cerevisiae</i>)
	1995	First complete genome (<i>H. Influenzae</i>)
	1992	First complete chromosome (<i>S. cerevisiae</i>)



Figure 1.1 A timeline of key developments in the genomics and postgenomics era. The availability of complete genomic sequence raises the question of the function of the individual genes as illustrated in the figure.

the costs of sequencing has steadily decreased and new technologies offer the possibility to dramatically decrease the costs further, opening up for complete sequencing as a tool in diagnostics. With this development, focus has shifted from genomic sequencing toward understanding the function of the individual genes (Figure 1), referred to as functional genomics. The availability of complete genomic sequences and requirement for identification of function for a large number of genes basically resulted in a paradigm shift in biology, as traditionally function was known (or studied) and research was focused on identification of the gene(s).

Bioinformatics has played a central role in functional genomics, but still experimental techniques are essential, and following the availability of complete genomic sequences a number of high-throughput experimental techniques have been developed that enables analysis of a large number of components within a living cell. These include DNA arrays for analysis of all (or a very high fraction) mRNAs, 2D-gel electrophoresis and advanced mass spectrometry for analysis of a large number of proteins, and yeast-two hybrid and other technologies for mapping of protein–protein interactions. These techniques are often referred to as omics techniques (derived from genomics), and terms such as transcriptomics, proteomics, and interactomics are used to describe these different analytical approaches. Even though all high-throughput techniques enable analysis of a large number of components (or interactions), it is, however, currently only transcriptomics that enables measurement of all the relevant components (in this case the mRNAs). Metabolomics is one of the more recently introduced “omics” technologies and as the word indicates it focus on analysis of all the metabolites within the cell under study. Similar to the use of

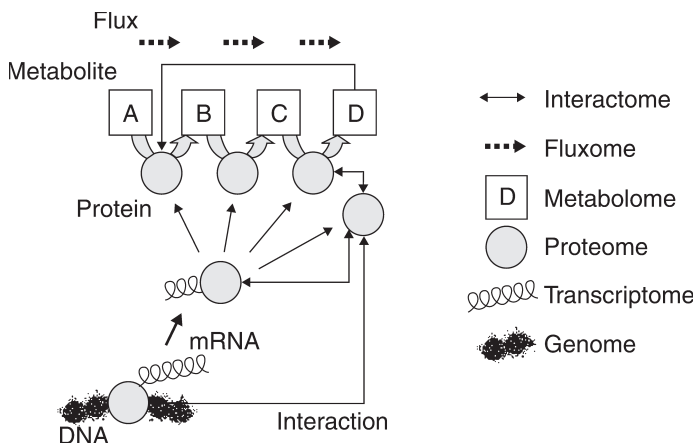


Figure 1.2 An overview of some key “omes” within a cell. The overview captures the central dogma of biology where genes are transcribed into mRNA, which is further translated into proteins. Proteins serve many different functions within the cell, but some act as enzymes that catalyze the interconversion of metabolites. The interconversion rates of metabolites are given as a set of fluxes through the different biochemical pathways operating in the cell. The different components of the cell may interact with each other resulting in the appearance of complex control loops imposed on many key functions in the cell.

“omics” the term “ome” is often used to describe all the components in a given group of compounds (or interactions). Figure 1.2 gives an overview of the different “omes” in the context of cellular function; and Table 1 gives our definition of some of the most frequently analyzed “omes.”

TABLE 1.1 Definitions of Frequently Analyzed “Omes”.

Genome	The complete nucleotide sequence in the genetic material of a living cell and further the complete list of all open reading frames (ORFs) that encode proteins.
Transcriptome	The complete set of all mRNA present in the cell.
Proteome	The complete set of all proteins present in the cell. The pool includes different forms of the same protein, e.g. a protein can be present in different states (phosphorylated/non-phosphorylated), and the proteome may therefore include many more components than the transcriptome and the number of ORFs.
Metabolome	The complete set of all metabolites formed by the cell in association with its metabolism.
Fluxome	The complete set of all fluxes through the different biochemical reactions that are involved in the interconversion of metabolites.
Interactome	The complete set of interactions between different components within the cell. These interactions include protein-protein interactions, protein-DNA interactions, protein-metabolite interactions as well as other possible interactions.

1.2 SYSTEMS BIOLOGY AND METABOLIC MODELS

A fundamental problem in interpreting results from the analysis of the different “omes” is that the individual components in all the “omes” are complex functions of a large number of different cellular components (see Figure 1.2). This has called for integrated analysis, where several “omes” are measured in parallel, and mathematical models are used for the analysis of the data. This approach is referred to as systems biology, and in recent years there has been a major shift toward integrated analysis, and in particular building detailed mathematical models describing different parts that forms the basis for the complete biological system that makes up a living cell.

As an illustration of the interaction of the different components in a living cell, the transcription of a given gene is a function of the level of transcription factors, and also the activities of upstream kinases and receptors. Similarly, the level of any given protein is determined, not just by the level of its corresponding mRNA, but also by the activity of the translational apparatus, protein kinases, phosphatases, and proteases. Whereas the levels of metabolites are determined directly by the activities of many different enzymes (parts of the proteome), the individual components of the metabolome are generally far more complex functions of other components in the cell than is the case for mRNAs or proteins. Thus, the level of any metabolite in the cell is determined by the activity of all the enzymes that are involved in the synthesis and conversion of that metabolite. Detailed metabolic models (see Table 1.2 and text below) have shown that less than 30% of the metabolites are involved in only two reactions, whereas about 12% of the metabolites participate in more than 10 reactions and about 4% of the metabolites even participate in more than 20 reactions. Furthermore, most reactions in a living cell involve more than a single substrate and a single product (more than 67% in the yeast *S. cerevisiae*) and this ensures a high degree of connectivity in the metabolic network (see Figure 1.3). Thus, the metabolic network operating in a living cell is a complex myriad of reactions that are tightly connected. Due to this coupling of many different reactions within the metabolic network, even small perturbations in the proteome (e.g., an alteration in the level of a few enzymes) may result in a significant change in the levels

TABLE 1.2 Some Data from a Few Detailed Metabolic Models (From Borodina and Nielsen, 2005).

Organism	Reactions	Metabolites	Metabolic ORFs	Total ORFs
<i>H. pylori</i>	444	340	268	1638
<i>H. influenzae</i>	477	343	362	1880
<i>E. coli</i>	720	436	695	4485
<i>S. coelicolor</i>	700	501	769	8042
<i>S. cerevisiae</i>	1175	584	708	5773
<i>M. musculus</i>	1220	872	—	—

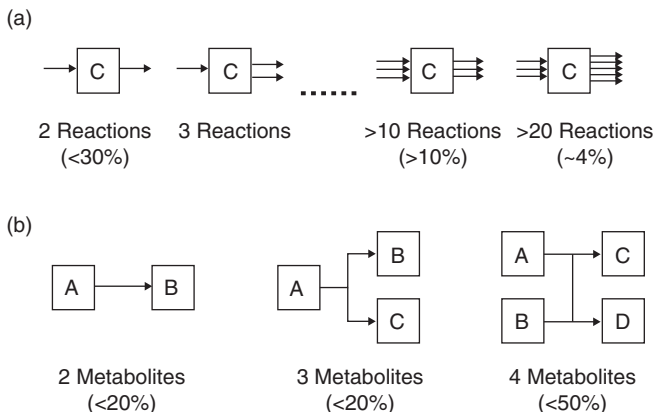


Figure 1.3 Illustration of the tight coupling of the different reactions in the metabolic network operating in a living cell. (a) Distribution of the number of reactions spanning the different metabolites. (b) Distribution of the number of metabolites being involved in the different reactions in the metabolic network.

of many metabolites. The biological reason for this may well be that this ensures a stable operation of the metabolic network with respect to the occurrence of mutations, i.e., upon a decrease in the activity of a particular enzyme, the response may be an increase in the level of the substrates of that enzyme, thereby ensuring that the change in the flux may only be slightly altered. Thus, evolution may have favored the establishment of metabolic networks that are tightly coupled and hence are robust to different kinds of perturbations.

As mentioned above the objective of systems biology is to represent cellular function through mathematical models, and many different types of mathematical models have been developed for the description of a wide range of cellular processes. Due to the conserved nature of the central metabolism in different biological systems, the function of metabolism has been extensively studied, and also the genes encoding enzymes involved in the central metabolism are very well annotated for most organisms. This has formed the basis for reconstruction of complete metabolic networks of several different organisms (see Table 1.2). This reconstruction process relies on genomic information and biochemical information of the studied organism (Palsson, 2006). These reconstructed metabolic networks serve as scaffolds for metabolic models that can be used to predict cellular function and study the role of individual reactions, and also for analysis of “omics” data (Borodina and Nielsen, 2005; Palsson, 2006). In the context of metabolomics these models are particularly useful as they provide links between the different metabolites in the metabolic network. They can also be used to calculate the fluxes through different parts of the metabolism, and through combination with metabolome analysis; it is hereby possible to correlate metabolite levels and fluxes, which enables identification of key control points in the metabolism.

1.3 METABOLOMICS

Being the intermediates of biochemical reactions, metabolites play a very important role in connecting the many different pathways that operate within a living cell. As mentioned above the level of metabolites represents integrative information of the cellular function, and, hence, defines the phenotype of a cell or tissue in response to genetic or environmental changes. Analysis of cellular function at the molecular level requires recruitment of several different analytical techniques. Whereas comprehensive methods for analysis at the transcriptional level (transcriptome) and at the translational level (proteome) are currently in a rapid state of development, and high-throughput analytical methods are already in use, methods for analysis of the metabolomics approaches are, however, so far less common, and currently there is no single method that enables analysis of the metabolome. Although metabolite profiling has long been applied for medical and diagnostic purposes as well as for phenotypic characterization, it is not until recently that increasing efforts have been undertaken to develop methods to screen of a high number of intracellular metabolites in the context of functional genomics (Fiehn, 2001).

Metabolome analysis covers the identification and quantification of all intracellular and extracellular metabolites with molecular mass lower than 1000 Da¹, using different analytical techniques. In common with the transcriptome and the proteome, the metabolome is context-dependent, and the levels of each metabolite depend on the physiological, developmental, and pathological state of a cell, tissue, or organism. However, an important difference is that, unlike mRNA and proteins, it is difficult or impossible to establish a direct link between genes and metabolites. The convoluted nature of cell metabolism, where the same metabolite can participate in many different pathways, complicates the interpretation of metabolite data.

The genome, transcriptome, and proteome elucidations are based on target chemical analyses of biopolymers composed of four different nucleotides (genome and transcriptome) or 22 amino acids (proteome). Those compounds are highly similar chemically, and facilitate high-throughput analytical approaches. Within the metabolome, there is, however, a large variance in chemical structures and properties. Thus, the metabolome consists of extremely diverse chemical compounds from ionic inorganic species to hydrophilic carbohydrates, volatile alcohols and ketones, amino and nonamino organic acids, hydrophobic lipids, and complex natural products. That complexity makes it virtually impossible to simultaneously determine the complete metabolome (Chapter 2). To further add to the complexity of metabolome analysis is the very rapid turnover of metabolites, i.e., many metabolites are present in low concentrations and there are very high fluxes through the metabolite pools. It

¹This cut-off molecular weight is obviously not very strict as many secondary metabolites have molecular weights above 1000 Da, and still they are considered to be metabolites. However, it is necessary to have some kind of discrimination between metabolites and macromolecules that are the major constituents of the cell, i.e., proteins, DNA, RNA, lipids, etc.

is therefore important to quench the metabolism rapidly and this calls for dedicated methods for quenching and extraction of metabolites from living cells. Therefore, the metabolomics encompass sample preparation (Chapter 3), sample analysis (Chapter 4), and data analysis (Chapter 5). Basically each metabolome study requires an evaluation of the sample preparation and the extraction procedure and how they couple to a combination of different analytical techniques in order to achieve as much information as possible, and we will illustrate this in a number of examples at the end of the textbook (Chapters 6–9).

As there are no single analytical method for analysis of the metabolome, different terms are often used in the field of metabolomics (see Table 1.3). There is a general consensus that the term metabolome describes the total sum of metabolites a given biological system can either use or form by its metabolism. The metabolome is often divided into the exometabolome and the endometabolome, where the former represents metabolites outside the cell and the latter represents intracellular metabolites. Whereas this distinction between exo- and endometabolome is quite useful for microbial systems where it is easy to separate the cells from the extracellular medium, it is less useful for multicellular systems where it may be difficult to isolate the cells from complete tissue. However, still it is conceptually important to differentiate between these two as the exometabolome often plays a very different

TABLE 1.3 Some Definitions Used in Metabolome Analysis (Adapted from Nielsen and Oliver, 2005).

Metabolome	The complete set of all metabolites used by or formed by the cell in association with its metabolism. The metabolome comprises both the endometabolome (the complete set of intracellular metabolites) and the exometabolome (the set of metabolites excreted into the growth medium or extracellular fluid).
Metabolomics	Approaches to analyze the metabolome or a fraction of the metabolome. Metabolomics involves sampling, sample preparation, chemical analysis, and data analysis.
Metabolic fingerprinting	Spectra from NMR or MS analysis that provides a fingerprint of metabolites produced by a cell. The fingerprint typically does not provide information about specific metabolites.
Metabolic footprinting	Analysis of the exometabolome. This may be either through analysis of specific metabolites or through spectra that do not provide information about specific metabolites (in analogy with metabolite fingerprinting).
Metabolite profiling	Analysis of a group of specific metabolites, e.g. a class of metabolites such as carbohydrates or amino acids. The analysis does not need to be quantitative, but often it is at least semiquantitative.
Metabolite target analysis	Quantitative analysis of metabolites participating in a specific part of the metabolism.

physiological role than the endometabolome. Two terms that are often used to describe analysis of a part of the metabolome are metabolite profiling and metabolic fingerprinting. These two terms are often used as synonyms with no clear distinction, but here we will use the definitions given in Table 1.3, which is adapted from Fiehn (2001) (see also Nielsen and Oliver, 2005). According to these definitions, metabolite profiling is the analysis of a given set of metabolites, e.g., a set of amino and organic acids, whereas metabolic fingerprinting is an unspecific analysis of a sample, e.g., a range of mass peaks obtained by mass spectrometry. The former provides direct physiological information, and the data can be integrated into metabolic models, whereas the latter provides a fingerprint that only can be used for grouping of different samples, e.g., using cluster analysis. As one may use nonspecific analysis of both the exo- and the endometabolome, the term metabolic footprinting has been introduced to describe analysis of the exometabolome in microbial cultures (Allen et al., 2003). The term footprinting indicates that the microbial cells leaves a footprint in the extracellular medium when they take up nutrients and secrete metabolites in connection with their growth process. Even though metabolic fingerprinting (or footprinting) does not provide information about the levels of specific metabolites, these analysis techniques may still be used for classification of mutants (or growth conditions) and permit the assignment of functions to orphan genes through the concept of guilt-by-association. It is, however, difficult to integrate this kind of data with other types of data, e.g., transcriptome data, and even though the concept of guilt-by-association is useful for classification of and hence can be used in functional genomics, it is less useful in systems biology where quantitative data are required. There are basically two solutions to this fundamental problem: (1) one may identify the peaks (or metabolites) that are playing a key role in distinguishing the different mutants (e.g., by using MS–MS) or (2) one may restrict the analysis to a group of metabolites which can be measured quantitatively (e.g., by CE–MS, LC–MS, or GC–MS), i.e., using metabolite profiling. Whereas the first solution provides some insight into the qualitative response of metabolism to the genetic change, it is associated with the risk of not identifying the quantitative effects of a given mutation. The other solution may produce a quantitative phenotype for a given mutation, but miss metabolites that are the key to the analysis. Some new developments in CE–MS (Soga et al., 2003) and GC–MS (Roessner et al., 2000; Weckwerth et al., 2004; Villas-Boas et al., 2005) do, however, enable true quantitative analysis of a relatively large number of metabolites.

Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are the most frequently employed methods of detection in the analysis of the metabolome (Chapter 4). NMR, in particular, is very useful for structure characterization of unknown compounds and has been applied for the analysis of metabolites in biological fluids and cells extracts. However, in certain circumstances, the ^1H NMR spectrum is insufficient on its own to provide information that will fully characterize a metabolite, but it may still provide a valuable metabolic fingerprint. This is obvious the case where analytes contain functional groups that are deficient in protons or where the protons can readily chemically exchange with the solvent, the signals thus being broadened beyond detection. Alternatively, other nuclei

can also be used, such as ^{13}C NMR. However, ^{13}C NMR spectroscopy presents relatively low sensitivity, i.e., in the range of μmol to mmol . In addition, ^{13}C NMR analysis may take several hours for a single sample, as a consequence of its low sensitivity, and the equipment costs are much higher compared to MS-based techniques.

The most important advantages of MS is its high sensitivity, and high-throughput in combination with the possibility to confirm the identity of the components present in the complex biological samples as well as the detection and, in most of the cases, the identification of unknown and unexpected compounds. Furthermore, the combination of separation techniques (e.g., chromatography) with MS tremendously expands the capability of the chemical analysis of highly complex biological samples. The basic information of mass spectra is characterized by its simplicity. The spectrum displays masses of the ionized molecule and its fragments, and those masses are simply the sums of the masses of the component atoms. In some cases, a mass spectrum contains a wealth of specific analytical and structural information, much more information than the expert in the field can currently utilize; unfortunately that abundance of information can discourage the novice who turns to compendia of mass spectrometric information for help. Nevertheless, it is comparatively simple to handle the mass spectra and there are several available software applications that make the interpretation of mass spectrometric data relatively easy.

1.4 FUTURE PERSPECTIVES

From the recent past it became obvious that metabolomics is a scientific field which develops with an enormous speed which makes it already difficult to follow the increasing numbers of scientific publications presenting the development of novel instrumentation, methodologies, or exciting applications in biology. With this development metabolomics has attracted increasingly interests, not only by biologists but also by the public and politicians as its value has been convincing from many successful applications. In near future, many institutions and laboratories worldwide will have established the physical and intellectual capacities to apply metabolomics in their research programs. Metabolomics will become more and more advanced, which will concurrently lead to certain confidence in the way it is applied and in the validity of the data obtained. In plant research, potential applications for metabolomics are enormous as described in Chapter 8, and for this reason the Plant Metabolomics Society has been founded some years ago (www.plantmetabolomics.nl) and four international conferences so far were held by the society, which has given the opportunity to share exciting new developments in the field. This society has been followed by the recently founded Metabolomics Society (www.metabolomicssociety.org).

As discussed above, the strength of metabolome analysis is that metabolite levels present a high degree of integrative information. This is, however, also a drawback as it is inherently difficult to interpret the results. In those cases where the levels of

many different metabolites have been measured, it is often difficult to bring the data into a physiological context that matches our current understanding of metabolism (measurement of many metabolites is, however, valuable for discovery of new pathways). Some studies have succeeded in mapping measurements of several metabolites onto metabolic charts and have hereby demonstrated how metabolite profiling can be combined with transcriptome analysis for mapping responses when the cells are exposed to different environmental conditions (Hirai et al., 2004; Villas-Bôas et al., 2005). However, as mentioned above, metabolism is far more connected than is shown by maps downloaded from KEGG (www.genome.jp/kegg) or other databases. Therefore, if a large number of metabolites are measured, it is necessary to adopt a more structured approach to data analysis. This is provided through the integration of experimental data with mathematical models, and as metabolism has been particularly well described for many microorganisms (Kell, 2004), it makes sense to start such model-driven data analyses using such single-celled systems. Recently, it has been demonstrated how a detailed metabolic model for *E. coli* could form the basis for integrating transcriptome data with computational data (Covert et al., 2004). Furthermore, by converting a genome-scale metabolic model to a metabolic graph, it has shown possible to use genome-scale metabolic models for identification of parts of the metabolic network that are transcriptionally coregulated (Patil and Nielsen, 2005), and this concept can easily be extended to the integration of transcriptome, proteome, and metabolome data.

As has been shown in a number of cases and will be shown in this textbook, metabolome analysis has proven successful for phenotypic mapping of cells, and thereby for the clustering of different mutants. However, as pointed out recently by Nielsen and Oliver (2005), it is a requirement for a wider use of metabolome analysis, and particularly for integration of these data with mathematical models as mentioned above, that there is a shift toward *truly quantitative analysis of specific metabolites obtained under well-defined conditions*. By “true quantitative analysis” they mean not only measurement of relative levels, but also measurement of actual concentrations of the different metabolites. This calls for

- Definition of appropriate data standards
- Development of standard analytical methods
- Development of appropriate libraries of mass spectra of GC–MS and LC–MS for standard analytical methods.

Definition of data standards is important for enabling comparison of data from different experiments, and from transcriptome analysis the true value of accumulating large data-sets has been demonstrated in several cases. Thus, in analogy with the MIAME standards for transcription analysis, it is interesting to define data standards for metabolome analysis, and there are already movements in this direction (Jenkins et al., 2004), and obviously the above-mentioned Metabolomics Society will play an important role in defining standards and building libraries. This is not an easy task because, for example, many different synonyms are used for one and the same metabolite and many different methodologies are used to analyze metabolites. Therefore,

ways for the standardization of metabolomics experiments have to be defined and accepted by the community, and anthologies have to be determined and used commonly. The driving force behind these initiatives is the desire of each metabolomics user to increase the number of identified metabolites and hereby increase the amount of information extractable from measurements. In addition, a functional database for public metabolomics data will attract computer scientists and bioinformaticians to develop novel methods for analysis of these huge data-sets leading, for example, to the development of new and useful software packages for data visualization, mining, and information extraction. This again will be of great help and use for the biologists. In recent years, there have been some reports on standard analytical methods that enable quantitative analysis of a large number of metabolites and there is a trend toward defining mass spectral libraries for these methods (Villas-Bôas et al., 2005; Halket et al., 2005; Schauer et al., 2005), which will clearly support further advancement of the research field.

In conclusion, it is an extremely exciting time for metabolomics as a new, rapidly growing scientific field. Most interestingly in near future will be the development of a common language among biologists, biochemists, geneticists, molecular biologists, analytical chemists, bioinformaticians, and computer scientists for best and most satisfactory outcomes of any metabolomics approach. We hope that our textbook will assist in this development and spur further developments in metabolomics.

REFERENCES

- Allen J, Davej HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB. 2003. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnol* 21:692–696.
- Borodina I, Nielsen J. 2005. From genomes to *in silico* cells via metabolic networks. *Curr Opin Biotechnol* 16:1–6.
- Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BØ. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429:92–96.
- Fiehn O. 2001. Combining genomics, metabolome analysis and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* 2:155–168.
- Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exper Bot* 56:219–243.
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K. 2004. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101:10205–10210.
- Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Wurtele ES, Kell DB. 2004. A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 22:1601–1606.

- Kell DB. 2004. Metabolomics and systems biology: Making sense of the soup. *Curr Opin Microbiol* 7:296–307.
- Nielsen J, Oliver S. 2005. The next wave in metabolome analysis. *Trends Biotechnol* 23:544–546.
- Palsson BO. 2006. *Systems Biology*, Cambridge University Press, New York, NY, USA.
- Patil K, Nielsen J. 2004. Uncovering transcriptional regulation of metabolism using metabolic network topology. *Proc Natl Acad Sci USA* 102:2685–2689.
- Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J* 23:131–142.
- Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L, Fernie AR, Kopka J. 2005. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett* 579:1332–1337.
- Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. 2003. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2:488–494.
- Villas-Boas SG, Moxley JF, Åkesson M, Stephanopoulos G, Nielsen J. 2005. High-throughput metabolic state analysis: The missing link in integrated functional genomics. *Biochem J* 388:669–677.
- Weckwerth W, Loureiro ME, Wenzel K, Fiehn O. 2004. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci USA* 101:7809–7814.