METABOLOME ANALYSIS

An Introduction

SILAS G. VILLAS-BÔAS
AgResearch Limited
Grasslands Research Centre
New Zealand

UTE ROESSNER
Australian Centre for Plant Functional Genomics
School of Botany, University of Melbourne, Australia

MICHAEL A. E. HANSEN
JORN SMEDSGAARD
JENS NIELSEN
Center for Microbial Biotechnology, BioCentrum-DTU
Technical University of Denmark

WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication

METABOLOME ANALYSIS

THE WILEY BICENTENNIAL–KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

William J. Pesce

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

Peter Booth Wiley

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

METABOLOME ANALYSIS

An Introduction

SILAS G. VILLAS-BÔAS
AgResearch Limited
Grasslands Research Centre
New Zealand

UTE ROESSNER
Australian Centre for Plant Functional Genomics
School of Botany, University of Melbourne, Australia

MICHAEL A. E. HANSEN
JORN SMEDSGAARD
JENS NIELSEN
Center for Microbial Biotechnology, BioCentrum-DTU
Technical University of Denmark

WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Metabolome analysis : an introduction / Silas G. Villas-Bôas … [et al.].
p. ; cm.
Includes bibliographical references.
ISBN-13: 978-0-471-74344-6
1. Metabolites. 2. Genomics. I. Villas-Bôas, Silas G. (Silas Granato)
[DNLM: 1. Metabolism. 2. Cell Physiology. 3. Genomics–methods.
4. Systems Biology–methods. QU 120 M587973 2007]
QP171.M48 2007
572.8'6–dc22 2006022114

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To
our colleagues,
families and
friends

CONTENTS

PREFACE

It has been less than a decade the word "metabolome" was first used referring to all low molecular mass compounds synthesized and modified by a living cell or organism. As a consequence, metabolomics emerged as a new field in the biological science, achieving tremendous development and popularity in the last couple of years. Many would say that metabolomics is a new word for an old science, because it revives the classical biochemical concepts and studies what became "unfashionable" during the genomics era, and it makes extensive use of analytical techniques idealized much earlier than the massive genome sequencing programmes. But, the applicability of metabolomics combined with genomic information or other system wide approaches make this field unique in modern science, both because of its multidisciplinary requirement, where biologist, chemists, engineers, physicists, mathematicians, and statisticians have to join forces to solve common problems; or by its ambition in connecting the different levels of biological information at the molecular level.

As a postgenomics tool, metabolomics is a young field in science but in an exponential growth phase. There is already a peer reviewed journal in its second year of publication, totally dedicated to publish works in the metabolomics field (*Metabolomics, Springer*), an international Metabolomics Society that was formed in 2004 (www.metabolomicssociety.org), and six annual international conferences focused entirely on metabolomics developments and studies (the International Conference on Plant Metabolomics and the Scientific Meeting of the Metabolomics Society).

Despite of all the advances in the metabolomics area, there has been a lack of a concise and basic literature focused on metabolome analysis, particularly an introductory text that can be used as a general guide for a novice interested to start exploring this new field or as a textbook for graduate and undergraduate students

attending specialized courses. We, professionals with different scientific backgrounds, therefore joint efforts to write this textbook, aiming to guide the reader to the main steps involved in metabolite analysis, and covering different biological materials (e.g., from plant and animal tissues to microbial and cell cultures, body fluids, and extracellular media), as well as presenting and discussing the principles of the most used methodologies for sample preparation, separation techniques, and detection methods.

The reader will find the book divided into two parts: Part I presents and discusses the concepts and methodology behind metabolite analysis. We first introduced the metabolomics field and its new terminologies (Chapter 1), followed by a general introduction to the diverse biochemical world of small molecules, where the basic concepts of cell metabolism are presented and the differences between primary and secondary metabolites as well as the dynamics of biochemical reactions and metabolite turnover are discussed (Chapter 2). Then, progressively, the reader is taken through the several steps of metabolome analysis, starting with reviewing the diversity of techniques used for sampling and sample preparation (Chapter 3), followed by a global overview of modern analytical methods used in the separation, detection, and identification of metabolites (Chapter 4) and ending with Chapter 5 that is fully dedicated to the most challenging aspect of metabolomics—the data analysis.

Part II of the book is aimed to illustrate the applicability of metabolomics and to discuss specific particularities and requirements of metabolomics in certain groups of organisms. Thereby, we review successful cases of metabolome analysis, illustrating yeast metabolomics (Chapter 6); reviewing specialized sampling devices for microbial metabolomics (Chapter 7); discussing the plant systems and reviewing the major achievements in plant metabolomics (Chapter 8); illustrating the applicability of metabolomics in the classification of filamentous fungi (Chapter 9); and finishing the book with a complete review of metabolomics applied to human and other mammals (Chapter 10).

Our goal as authors was to write a concise and practical focused book as an introduction to metabolome analysis. A book focused on an integrated analytical approach combining the whole analytical chain from sampling over extraction and separation to state-of-the art mass spectrometry and data processing. Although we included a few review chapters in the second part of the book, it is important to emphasize that this book was not intended to be a review book but a textbook that introduces the principles rather than the latest results. The readers will find in the next pages bits of biochemistry, bits of molecular biology, bits of analytical chemistry, bits of mathematics and statistics, and even bits of chemical engineering. That was the challenges that we faced when decided to write this book: to organize the work-flow in metabolome analysis covering all different biological systems and all interdisciplinary aspect. We believe in metabolomics as a field *per se* rather than an additional tool in science. We borrow tools from different sciences to build this new field: METABOLOMICS. Now we invite you to try it.

LIST OF CONTRIBUTORS

Dr. David Wishart, Deptments of Biological Sciences & Computings Sciences, 2-21 Athabasca Hall, University of Alberta, Edmonton, AB Canada, T6G 2E8

Dr. Jens Nielsen, Center for Microbial Biotechnology, Building 223, BioCentrum-DTU, Technical University of Denmark, DK-2800, Kongens Lyngby, Denmark

Dr. Jørn Smedsgaard, Center for Microbial Biotechnology, Building 221, BioCentrum-DTU, Technical University of Denmark, DK-2800, Kongens Lyngby, Denmark

Dr. Michael Adsetts Edberg Hansen, Center for Microbial Biotechnology, Building 223, BioCentrum-DTU, Technical University of Denmark, DK-2800, Kongens Lyngby, Denmark

Dr. Silas Granato Villas-Bôas, AgResearch Limited, Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand

Dr. Ute Roessner, Australian Centre for Plant Functional Genomics, School of Botany, the University of Melbourne, 3010 Victoria, Australia

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 subsystems) 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,

Metabolome Analysis: An Introduction, by Silas G. Villas-Bôas, Ute Roessner, Michael A. E. Hansen, Jorn Smedsgaard and Jens Nielsen

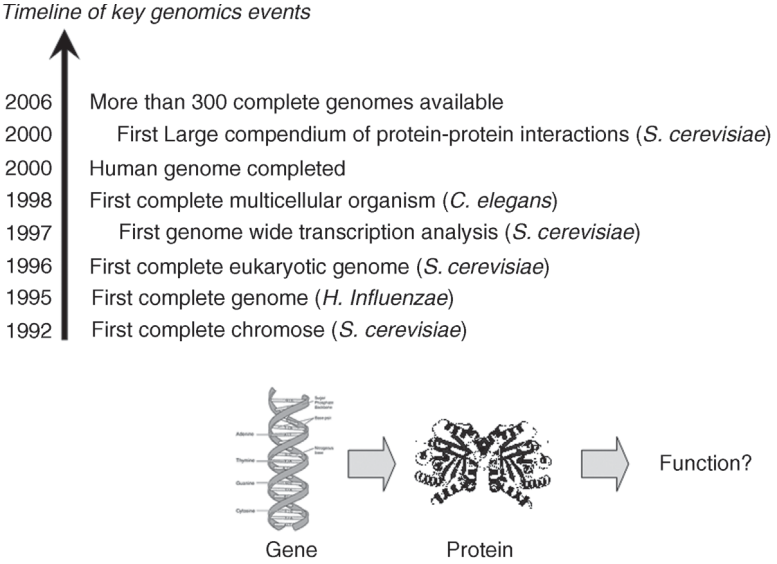

Figure 1.1 A timeline of key developments in the genomics and postgenomics era. The availability of complete genomic sequence raises the question of the function of the individual genes as illustrated in the figure.

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

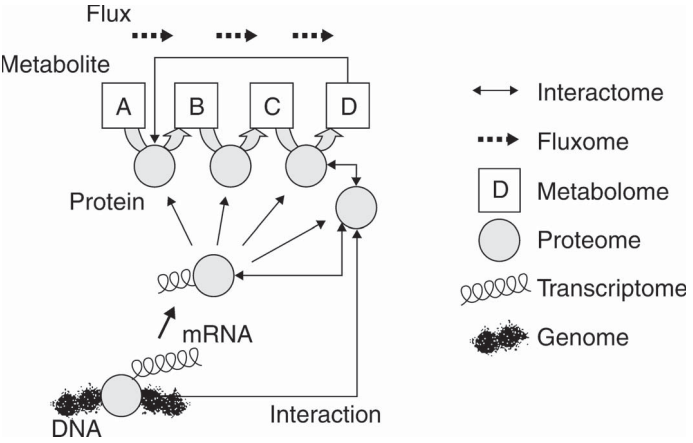

Figure 1.2 An overview of some key "omes" within a cell. The overview captures the central dogma of biology where genes are transcribed into mRNA, which is further translated into proteins. Proteins serve many different functions within the cell, but some acts as enzymes that catalyze the interconversion 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
H. pylori	444	340	268	1638
H. influenzae	477	343	362	1880
E. coli	720	436	695	4485
S. coelicolor	700	501	769	8042
S. cerevisiae	1175	584	708	5773
M. musculus	1220	872	—	—

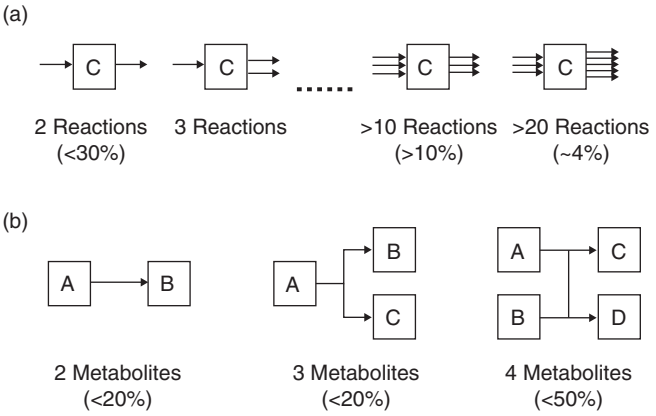

Figure 1.3 Illustration of the tight coupling of the different reactions in the metabolic network operating in a living cell. (a) Distribution of the number of reactions spanning the different metabolites. (b) Distribution of the number of metabolites being involved in the different reactions in the 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 Da[1], 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

[1]This 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 date 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 ^{1}H 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,