METABOLOME ANALYSES:
Strategies for Systems Biology

METABOLOME ANALYSES: Strategies for Systems Biology

Edited by

Seetharaman Vaidyanathan
School of Chemistry, The University of Manchester, UK

George G. Harrigan
Pfizer, Chesterfield, MO, USA

Royston Goodacre
School of Chemistry, The University of Manchester, UK

Springer

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN-10: 0-387-25239-8 e-ISBN-10: 0-387-25240-1 Printed on acid-free paper.
ISBN-13: 978-0387-25239-1 e-ISBN-13: 978-0387-25240-7

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11054030

springeronline.com

Dedication

To my parents (SV),

To Beth, Sean and Evan
(GGH),

To Elizabeth, Tamara and
Rhozzum Connor (aka. Pickles)
(RG)

Contents

Contributing Authors

Ian N. Acworth
ESA Inc. 22 Alpha Road, Chelmsford, MA 01824, USA

Eivind Almaas
Center for Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Albert-László Barabási
Center for Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

M. Flint Beal
Weill Medical College of Cornell University, 525 East 68 St., NY 10021, USA

Christopher H. Becker
SurroMed, Inc.,1430 O'Brien Drive, Menlo Park, CA 94025, USA

Richard D. Beger
Division of Systems Toxicology, 2, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079-9502, USA

Alvin Berger
Icoria Inc. (formerly Paradigm Genetics, Inc), 108 Alexander Dr., Research Triangle Park, NC, 27709, USA

Raoul J. Bino
Plant Research International, Business Unit Bioscience, P.O. Box 16, 6700 AA, Wageningen, The Netherlands

Mikhail Bogdanov
Weill Medical College of Cornell University, 525 East 68 St., NY 10021, USA

László G. Boros
SIDMAP, LLC, 10021 Cheviot Drive, Los Angeles, CA 90064, USA

Axel A. Brakhage
Institute of Microbiology, University of Hannover, Schneiderberg 50, D-30167, Hannover, Germany

Juan I. Castrillo
The University of Manchester, School of Biological Sciences, The Michael Smith Building, Oxford Road. Manchester M13 9PT, UK

Helga David
Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark

Yvonne P. Dragan
Division of Systems Toxicology, 2, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079-9502, USA

Edward M. Driggers
Microbia, Inc., 320 Bent St., Cambridge, MA 02141, USA*

**Current address:* Ensemble Discovery Corp., 99, Erie St., Cambridge, MA 02139, USA

Thomas Ferenci
School of Molecular and Microbial Biosciences, University of Sydney G08, N.S.W. 2006, Australia

Alisdair R. Fernie
Max-Planck-Institute für Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

Paul H. Gamache
ESA Inc. 22 Alpha Road, Chelmsford, MA 01824, USA

Royston Goodacre
School of Chemistry, The University of Manchester, Faraday Towers, Sackville Street, P.O. Box 88, Manchester M60 1QD, UK

Robert D. Hall
Plant Research International, Business Unit Bioscience, P.O. Box 16, 6700 AA, Wageningen, The Netherlands

George G. Harrigan
Pfizer, Chesterfield, MO 63017, USA

Li Jia
Graduate School of Material Science, University of Hyogo, Kamigori, Hyogo, 678-1297, Japan

Rima Kaddurah-Daouk
Metabolon Inc. 800 Capitola Dr., Suite 1, Durham NC 27713, USA*

**Current address*: Duke University Medical Center, Department of Psychiatry, Box 3950, Durham NC 27710, USA

Aaron B. Kantor
SurroMed, Inc.,1430 O'Brien Drive, Menlo Park, CA 94025, USA

Bruce S. Kristal
Departments of Biochemistry and Neuroscience, Weill Medical College of Cornell University, 1300 York Ave, NY 10021, USA
and
Dementia Research Service, Burke Medical Research Institute, 785 Mamaroneck Ave, White Plains, NY 10605, USA

Arno J. Krotzky
metanomics GmbH and Co. KGaA, metanomics Health GmbH, Tegeler Weg 33 10589 Berlin, Germany

Wai-Nang Paul Lee
SIDMAP, LLC, 10021 Cheviot Drive, Los Angeles, CA 90064, USA

Eva M. Lenz
Dept. of Drug Metabolism and Pharmacokinetics, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Ralf Looser
metanomics GmbH and Co. KGaA, metanomics Health GmbH, Tegeler Weg 33 10589 Berlin, Germany

Ram Maharjan
School of Molecular and Microbial Biosciences, University of Sydney G08, N.S.W. 2006, Australia

Wayne R. Matson
ESA, Inc., 22 Alpha Road, Chelmsford, MA 01824, USA

Wayne R. Matson
ESA, Inc., 22 Alpha Road, Chelmsford, MA 01824, USA

David F. Meyer
ESA Inc. 22 Alpha Road, Chelmsford, MA 01824, USA

Jens Nielsen
Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark

Stephen G. Oliver
The University of Manchester, School of Biological Sciences, The Michael Smith Building, Oxford Road, Manchester M13 9PT, UK.

Zoltán N. Oltvai
Department of Pathology, Northwestern University, Chicago, IL 60611, USA

Ugo Paolucci
Dementia Research Service, Burke Medical Research Institute, 785 Mamaroneck Ave., White Plains, NY 10605, USA

C.H. Ric de Vos
Plant Research International, Business Unit Bioscience, P.O. Box 16, 6700 AA, Wageningen, The Netherlands

Johann M. Rohwer
Triple-J group for Molecular Cell Physiology, Department of Biochemistry, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

Uwe Sauer
Institute of Biotechnology, Swiss Federal Institute of Technology (ETH) Zürich, 8093 Zürich, Switzerland

Laura K. Schnackenberg
Division of Systems Toxicology, 2, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079-9502, USA

Yevgeniya Shurubor
Dementia Research Service, Burke Medical Research Institute, 785 Mamaroneck Ave., White Plains, NY 10605, USA

Jacky L. Snoep
Triple-J group for Molecular Cell Physiology, Department of Biochemistry, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
and
Molecular Cell Physiology, Vrije Universiteit, Amsterdam, The Netherlands

Ralf Steuer
University, Potsdam, Nonlinear Dynamics Group, Am Neuen Palais 10, 14469 Potsdam, Germany

Shigeru Terabe
Graduate School of Material Science, University of Hyogo, Kamigori, Hyogo, 678-1297, Japan

Richard N. Trethewey
metanomics GmbH and Co. KGaA, metanomics Health GmbH, Tegeler Weg 33 10589 Berlin, Germany

Ewa Urbanczyk-Wochniak
Max-Planck-Institute für Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

Seetharaman Vaidyanthan
School of Chemistry, The University of Manchester, PO Box 88, Manchester M60 1QD, UK

Harrie A. Verhoeven
Plant Research International, Business Unit Bioscience, P.O. Box 16, 6700 AA, Wageningen, The Netherlands

Wolfram Weckwerth
Max-Planck-Institute of Molecular Plant Physiology, 14424 Potsdam, Germany

Rebecca Williams
Dept. of Drug Metabolism and Pharmacokinetics, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Lothar Willmitzer
Max-Planck-Institute für Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

Ian D. Wilson
Dept. of Drug Metabolism and Pharmacokinetics, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Nicola Zamboni
Institute of Biotechnology, Swiss Federal Institute of Technology (ETH) Zürich, 8093 Zürich, Switzerland

Haihong Zhou
SurroMed, Inc.,1430 O'Brien Drive, Menlo Park, CA 94025, USA

Foreword

The value of obtaining information on entire classes of analytes is now widely recognized among biological researchers. This unbiased ('omic) approach allows for observation of whole systems, and it is being employed in myriad applications spanning the entire spectrum of biology. There is, of course, no substitute for the hypothesis-driven experiment in validating new concepts. With an 'omics approach, however, it is possible to develop hypotheses for testing from an astonishingly complete understanding of a system and to monitor the results of hypothesis-driven experiments in a far more comprehensive fashion. Unbiased research was developed and most enthusiastically embraced by the genomics community. Looking back on the 'omic revolution from the future we might expect to observe that genomics defined a new course for biological research and made many fundamental advances in biological knowledge. It would not be surprising, however, to find that most of the practical tools developed through 'omics research were developed by applying the principles of genomics to profiling metabolites. Metabolites are particularly valuable for practical applications because they represent the integrated consequence of endogenous metabolism and the response to environmental stimuli. Thus, metabolic profiling provides a method for gaining insight into how biological entities function and into how they adapt or fail in the context of their surroundings. Profiling metabolites is not a new concept- metabolites have been used as useful indices of phenotype for many decades- but the improved analytical and informatic technologies exponentially increase the power of the approach. Research fields that have and will continue to benefit greatly from metabolomic profiling include functional genomics, nutrition, metabolic disease research, clinical care, drug discovery and development, agricultural biotechnology

and toxicology to name a just few. A major advantage for metabolic profiling over other 'omic strategies in advancing our understanding of these fields is that metabolites are inherently linked to phenotype and, importantly, 100 years of biochemical knowledge has been assembled around biochemical pathways. This latter point should allow a much faster translation of profile data to knowledge than is possible with genomics.

Advances in metabolic profiling have been driven in large part by improved analytical and informatics capabilities. The previous volume of this book outlined several of the primary technologies for profiling metabolites including mass spectrometry and NMR. While mass spectrometry and NMR will continue to serve as the core technologies for broad-based metabolic profiling schemes, the goals of metabolic profiling (generating quality data on a wide variety of metabolites simultaneously) do not favor any analytical platform over another. Older chromatographic platforms are equally likely to find use in this field, depending on the biological applications. This edition contains further examples of techniques and applications for spectrometry and NMR, but also contains several examples of new analytical technologies. While the advances in metabolic profiling capabilities are undeniable, the next phase of development for the field should encourage a broad range of researchers to adopt this obviously powerful research strategy. Only proof-of-principle biological results can accomplish this, and it is these examples the current practitioners of metabolic profiling should pursue.

While metabolic profiling has many advantages over genomics and proteomics in terms of utility, it is not without its own set of pitfalls and tradeoffs. Metabolites possess such an astonishingly broad spectrum of physical and chemical properties that no single analytical platform has, or is likely to, accurately quantify and identify all metabolites simultaneously from a biological sample. This fact forces some degree of compromise on the part of researchers, who can choose to trade quantitation for analytical breadth or *vice versa.* In general, research striving to be as inclusive as possible, and therefore sacrificing some degree of accuracy or the identification of compounds, is termed unbiased metabolomics. Research striving to be as accurate as possible on a known subset of the metabolome is termed focused metabolomics. There are also difficulties in the interpretation of data once they are generated. High-content datasets are notoriously prone to produce false discoveries as a result of the number of predictors relative to the degrees of freedom, and metabolic profiling is not exempt from this problem. As metabolic profiling matures, innovative solutions to these problems need to be developed.

Since the publication of the previous volume of this book, the National Institutes of Health announced the NIH Roadmap which outlines the key

themes and initiatives the NIH feels will advance public health in the coming years (Zerhouni, 2003). Among the initiatives singled out in the Roadmap for attention and, critically, public funding is metabolomic research and analytical technology development. The fact that the NIH has chosen to publicly back the concept of metabolic profiling and to commit to funding the development of new technologies is an indication that the field is entering a new phase of development and growth. The growing interest in metabolic profiling in the academic community is another sign that the field is beginning to mature. A keyword search on PubMed using the common terms for metabolic profiling demonstrates the rapid acceleration of publication in the field. While the number of papers meeting these search criteria (just shy of 1,000 as of this writing) lags far behind similar results for genomics, transcriptomics and proteomics, there are many signs that metabolome analyses will catch up in the coming years. Several prominent peer-reviewed publications are actively recruiting manuscripts involving metabolomic research and the new journal *Metabolomics* will begin publishing manuscripts in early 2005. These developments point to a recognition of metabolic profiling/metabolome analyses as an emerging, and important, new field.

It is undeniable that, at the time of this printing, capital investment in biochemical profiling and the publications produced by the approach lag far behind those for genomics, transcriptomics or proteomics. There are many encouraging indications that this disparity will not persist for long. The adoption of biochemical profiling as a central discovery platform should accelerate dramatically as more researchers enter the field, as access to grant money and investments continues to increase, and as proof-of-principle biological results develop and become widely recognized.

Zerhouni E. The NIH roadmap. *Science* 302: 63 (2003).

Steven M. Watkins

President and CSO
Lipomics Technologies, Inc,
West Sacramento, CA 95691

Acknowledgments

SV thanks the University of Manchester and the UK BBSRC for the opportunity and financial assistance. Contributions to the cover design by Sukanya is gratefully acknowledged, as is the help provided by present and past members of the research group, including Irena Spasic, Consuelo Lopez-Diez and Steve O'Hagan, at various times during the compilation of this volume. GGH acknowledges Margann Wideman of Pfizer for her continued support. RG would like to thank the University of Manchester and the UK BBSRC for allowing the academic freedom and financial assistance to investigate metabolic profiling. Heartfelt thanks are also expressed to all present and past members of the research group for their hard work and enthusiasm. Needless to say the editors are greatly indebted to all the authors for their invaluable contributions, without whom this volume would not have been possible.

Chapter 1

INTRODUCTION

Metabolome analyses for systems biology

Seetharaman Vaidyanathan[1], George G. Harrigan[2] and Royston Goodacre[1]
[1]School of Chemistry, The University of Manchester, Faraday Towers, Sackville Street, P.O. Box 88, Manchester M60 1QD, UK. [2]Pfizer, Chesterfield, MO 63017, USA

We are currently in a phase of scientific enquiry that is increasingly driven by the need to analyse biological systems much more holistically. Much of the excitement with respect to this need is due to the realization among practitioners of the traditional reductionist approach, including biochemists and molecular biologists, that there is more to biological systems than can be adequately accounted for by reductionist enquiries alone. Although not entirely novel, a 'systems' perspective in biology affords challenges and prospects which are only now being fully addressed in detail.

Tracking changes in the metabolic complement of the system (the low molecular weight component – the metabolome) that relate to its behaviour is progressively gaining momentum (Oliver *et al.*, 1998; Tweeddale *et al.*, 1998; Fell, 2001; Fiehn, 2001; ter Kuile and Westerhoff, 2001; Harrigan and Goodacre, 2003; Goodacre *et al.*, 2004; Kell, 2004). This particular aspect forms the subject matter of this edited volume. Following in the footsteps of its predecessor (Harrigan and Goodacre, 2003), this volume is compiled to give an overview of the scientific activity that is in progress in this particular field of enquiry. It is by no means comprehensive, but is aimed at capturing the excitement of the current practitioners of the field and relates to their experiences. In keeping with this objective, the authors' views are preserved and presented with minimal edits. Consequently, while the appearance of similar views strengthens its foundation, the appearance of conflicting views only reflects the growing nature of the field and emphasizes the need for active discussions that are inevitable in any emerging field.

1. THE PANOMICS ROUTE TO SYSTEMS BIOLOGY

The central dogma of molecular biology over the last few decades has advocated that the flow of information from the genes to function (or phenotype) is linear and is translated through transcripts, then proteins and finally metabolites. Most scientists have tended to analyse these in isolation with little emphasis on cross-talk between these different levels of molecular organisation. By contrast, the central dogma of systems theory dictates that there is more to a system than the sum of its parts. Indeed, the interaction of a system's parts can result in an emergent state that is not adequately accounted for by investigating the parts independently of each other (Weiner, 1948; Bertalanffy, 1969). Systems biology thus attempts to account for biological system behaviour that cannot be adequately explained by investigations at the molecular level alone (Ideker *et al.*, 2001; Kitano, 2001). Two routes to the evolution of this thinking within biological scientific enquiry can be identified (Levesque and Benfey, 2004; Westerhoff and Palsson, 2004) – i) the panomics route that relies on the generation of high-throughput data on the components of the system (the parts list) and ii) *in silico* routes that attempt to provide information on the interactions that the parts of the system might be involved in to effect a function.

The panomics route to systems biology has its roots in molecular biology. Molecular biology investigations over the past few decades have resulted in the identification of the molecular make-up of cells and the construction of a likely route to the storage, replication, processing and execution of information within cells. A linear hierarchy, in which information is stored in DNA, processed by RNA and proteins, and executed by proteins and metabolites, has become the basis for our understanding of cellular function. Consequently, it has become essential to catalogue these molecular entities in order to understand system behaviour. The genomic era ushered in large-scale DNA sequencing of living organisms, with the aim of explaining biological complexity and versatility in terms of genetic make-up. However, it is now known that whilst a few thousand genes can code for a eukaryotic cell (6000 for yeast (Goffeau *et al.*, 1996)), only two to three times as many is required to construct an entire multicellular organism (Bird *et al.*, 1999) and as little as five times more is required to construct a human being (McPherson *et al.*, 2001; Venter *et al.*, 2001). In addition, discoveries such as short-term information storage in proteins (Bray, 1995), the significant role of post-transcriptional and post-translational modifications in cell function, and the existence of metabolite-mediated regulation of cell function (Winkler *et al.*, 2004), now serve to question the rigor of classically defined hierarchical organisation and illustrate the limitations of genomic

enquiries. Clearly, it has become essential to catalogue other players in the cell factory to define gene function in the post-genomic era. This has now given birth to trancriptomes, proteomes and metabolomes, each relating to the make up of the cell associated with the respective components, RNA, proteins and metabolites.

Whilst transcriptomic and proteomic investigations are facilitating gene-function and annotation efforts, metabolomic investigations are lagging behind. An overview of the gains to be had by directing investigations at the metabolome level is provided in the following three chapters which address microbial (Chapter 2), plant (Chapter 3) and animal (Chapter 4) systems. These chapters also set the scene by providing an indication of the scope and context of metabolome analyses as applicable to different biological systems

Castrillo and Oliver (Chapter 2) elegantly provide the justification and need for directing enquiries at the metabolome level, taking a microbial system, the 'well characterized' yeast, as their model system. The complexity and metabolic diversity of plants, especially with respect to secondary metabolites, offers unique challenges to the characterization of their metabolomes. Hall and colleagues introduce us to some of these aspects in Chapter 3, and discuss metabolome analyses as applied to plant systems. In the following chapter Kaddurah-Daouk and colleagues give an insight into the application of metabolome analyses to the identification of (surrogate) biomarkers and therapeutic targets in animal systems, elaborating on issues pertaining to the study of disorders of the central nervous system.

1.1 Strategies for capturing metabolome-wide changes

Various strategies and challenges pertaining to the tracking of metabolome-wide changes in different biological systems under different application contexts are discussed in the next seven chapters (Chapters 5-11). Most strategies for capturing comprehensive metabolomic data employ a separation technique followed by sensitive detection, typically using mass spectrometry (MS). Separation techniques include two-dimensional thin layer chromatography (2D-TLC), capillary electrophoresis (CE), gas-chromatography (GC) and liquid chromatography (LC). Whilst the objective in such strategies is to capture comprehensive metabolome-wide changes, often the nature of the techniques and sample preparation protocols bias the type of metabolites detected, restricting the analyses to sub-metabolomes. Ferenci and Maharjan discuss the development and application of 2D-TLC in the context of profiling microbial metabolomes (Chapter 5). This is an economically viable solution, useful for comparing metabolomes. CE strategies are discussed by Jia and Terabe (Chapter 6), with respect to, but by no means restricted to, microbial metabolomes. In Chapter 7, Trethewey

and colleagues give an overview of current practices in GC-MS and LC-MS approaches to profiling metabolomes, as applicable to plant, microbial and health care investigations. The development and application of electrochemical techniques in combination with LC separations is discussed in Chapter 8 by Ackworth and collegues. Zhou and colleagues elaborate on the application of LC-MS strategies in Chapter 9 with emphasis on biomarker discovery using MS, within a clinical and drug discovery and developmental context.

Whilst comprehensive analysis would be informative for gaining metabolome-wide knowledge of the system, there are instances when capturing dominant changes in the metabolome through the detection of changes in a few metabolites as biomarkers can provide sufficient information for identifying system wide disturbances. These are usually effected with fingerprinting approaches that involve the direct detection of the system-wide changes with minimal sample pre-treatment or analyte separation, usually with the application of MS, nuclear magnetic resonance (NMR), Fourier transform infrared (FT-IR) or Raman spectroscopies (Harrigan and Goodacre, 2003; Goodacre *et al.*, 2004). In Chapter 10, Beger and colleagues discuss analytical strategies using NMR, highlighting its application in toxicology investigations.

A characteristic feature of 'omic approaches is the parallel and simultaneous high-throughput analysis of several analytes. This places unique demands on experimental design, with the requirement for careful considerations of biological, analytical and data processing issues. Kristal and colleagues (Chapter 11) elaborate on some of these issues and share the lessons they have learnt from metabolic profiling of a model nutritive system in animals.

2. METABOLIC INTERACTIONS FROM A SYSTEMS PERSPECTIVE – THE *IN SILICO* ROUTE TO SYSTEMS BIOLOGY

A metabolomic "parts" list will benefit functional genomic investigations, and can be associated with system-level perturbations. However, knowledge of gene function or, as identified earlier, a catalogue of all the genes, transcripts, proteins and metabolites associated with a system is unlikely to suffice in explaining system behaviour. In addition to establishing which components are involved in a given cellular or biological event, systems-level understanding requires information on how the different components interact to influence system behaviour. A second route to

systems biology (Levesque and Benfey, 2004; Stelling, 2004; Westerhoff and Palsson, 2004) that deals with *in silico* analysis of cellular processes and systems-level data that aim to capture system structure and dynamics can also be identified. At the metabolome level, this route promises to provide information on metabolic interactions from a systems perspective.

In Chapter 12, David and Nielsen focus their discussion on the construction, properties and application of genome scale models developed for fungal systems, and debate their significance in gaining systems level understanding of cellular function. Snoep and Rohwer (Chapter 13) present kinetic modeling of biological systems and elaborate on the concept of metabolic control analysis.

It is now increasingly recognized that complex entities such as biological systems can be represented as networks, the large-scale behaviour of which, if predicted, would enable the understanding of systems behaviour. Complex interactions of intracellular molecules can be captured by this network concept. Oltvai and colleagues (Chapter 14) discuss metabolic networks, presenting the underlying principles, approaches, and utilization of such information regarding these networks. It has been observed with plant systems that metabolites tend to vary in concert with other metabolites. The resulting correlation in metabolite levels within a data set can be used to construct metabolic correlation networks that can be useful in understanding systems behaviour. Weckwerth and Steuer discuss this aspect in Chapter 15. Another *in silico* route to understanding system behaviour is to combine information available from different 'omic platforms to look for patterns that can be associated with systems behaviour. Fernie and colleagues take this route and describe the pair-wise analysis of transcript and metabolite profiles to study potato tuber metabolism and discuss the potential of this approach in Chapter 16.

Metabolic flux ratio analysis can provide information of metabolic network operation, as opposed to network composition. In Chapter 17, Zamboni and Sauer describe flux ratio analysis and discuss the potential of comparative fluxome profiling, illustrating this type of analysis in microbial systems.

3. THE PATH AHEAD – CONCLUDING REMARKS

The final four chapters (Chapters 18-21) deal with the application of metabolome analyses in different contexts to summarize the potential scope of the technique in different application areas. Boros and Lee, in Chapter 18, detail the utility of stable isotope-labeled approaches (SIDMAP) in capturing metabolic changes. They show how SIDMAP can provide valuable

information in investigations of the effect of endogenous and exogenous agents on intermediary metabolism in tumor cells, and debate the role of metabolic profiling in targeted drug design. In the next chapter (Chapter 19), Lenz and colleagues provide an overview of metabonomic investigations in the pharmaceutical industry and discuss the potential this approach holds in toxicological studies and the study of disease models. Lipids constitute a significant proportion of the metabolic complement of biological systems, and play key roles in its functioning. Berger, in Chapter 20, explains why and how this subset of the metabolome contributes to our understanding of system behaviour. In the final (but by no means less important) chapter of the volume (Chapter 21), Driggers and Brakhage discuss the role of metabolic profiling in the study of fungal virulence and show the value of combining metabolome level data with transcriptome level information for assessing this system.

By now, one aspect of *Systems Biology* can be well appreciated, i.e., that it is *an integrative approach.* The route to obtaining systems level information, be it through molecular investigations or through global analysis of networks and interactions, is clearly complementary, and metabolome level data will have to be analysed alongside data obtained from other 'omic platforms to make meaningful observations on system-wide behaviour. Without doubt, data integration and bioinformatics tools for countering the challenges posed by such integration of data from different platforms will have to be addressed before meaningful interpretations can be made. Not withstanding, the potential in profiling metabolomes and investigating metabolome-wide network behaviour in understanding systems behaviour is clearly evident. We hope that this volume convinces you of this exciting potential and that you enjoy reading it!

REFERENCES

Bertalanffy Lv. *General System Theory, Foundations, Development, Applications*, George Braziller, New York, 1969.

Bird DM *et al.* The *Caenorhabditis elegans* genome: A Guide in the post genomics age. *Annu. Rev. Phytopathol.*, 37: 247-265 (1999).

Bray D. Protein molecules as computational elements in living cells. *Nature*, 376: 307-312 (1995).

Fell DA. Beyond genomics. *Trends Genet.*, 17: 680-682 (2001).

Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genom.*, 2: 155-168 (2001).

Goffeau A *et al.* Life with 6000 genes. *Science*, 274: 546-567 (1996).

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.*, 22: 245-252 (2004).

Harrigan GG, Goodacre R. *Metabolic Profiling: Its role in biomarker discovery and gene function analysis*, Kluwer academic publishers, Boston (2003).

Ideker T, Galitski T, Hood L. A new approach to decoding life: Systems Biology. *Annu. Rev. Genomics Hum. Genet.*, 2: 343-372 (2001).

Kell DB. Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.*, 7: 296-307 (2004).

Kitano H. *Foundations of Systems Biology*, MIT Press, Cambridge, MA, 2001.

Levesque MP, Benfey PN. Systems Biology. *Curr. Biol.*, 14: R179 (2004).

McPherson JD *et al.* A physical map of the human genome. *Nature*, 409: 934-941 (2001).

Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.*, 16: 373-378 (1998).

Stelling J. Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.*, 7: 513-518 (2004).

ter Kuile BH, Westerhoff HV. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.*, 500: 169-171 (2001).

Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J. Bacteriol.*, 180: 5109-5116 (1998).

Venter JC *et al.* The sequence of the human genome. *Science*, 291: 1304-1351 (2001).

Weiner N. *Cybernetics or control and communication in the Animal and the Machine*, MIT Press, Cambridge, MA (1948).

Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. *Nat. Biotechnol.*, 22: 1249-1252 (2004).

Winkler WC *et al.* Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428: 281-286 (2004).

Chapter 2

TOWARDS INTEGRATIVE FUNCTIONAL GENOMICS USING YEAST AS A REFERENCE MODEL

Metabolomic analysis in the post-genomic era

Juan I. Castrillo and Stephen G. Oliver
School of Biological Sciences. The Michael Smith Building. University of Manchester.Oxford Road. Manchester M13 9PT, UK.

1. INTRODUCTION

Metabolites have been the subject of investigation since the early stages of modern biology. Thus, classical studies on identification of enzymes and metabolic intermediates performed in yeast in the 1920s-1930s (e.g. Embden-Meyerhoff unified theory of glycolysis, citric acid cycle, AMP, ATP) constitute the foundations of modern enzymology and biochemistry (Lehninger, 1975; Alberts *et al.*, 2002). The main interest of these studies focused on the elucidation of the complete map of central metabolic pathways and intermediary metabolites of an organism. This objective, satisfactorily fulfilled for the case of a few organisms (bacteria, yeast), may constitute a major task in more complex organisms (e. g. plants, mammalian cells), with particular metabolites (e.g. secondary metabolites and regulatory compounds) still to be identified. For the case of eukaryotes, yeast central metabolic pathways and methods for determination of metabolites are used as a reference from which to approach more complex biological systems (Gancedo and Gancedo, 1973; Saez and Lagunas, 1976; Rose and Harrison, 1987-1995; Fell, 1997; Alberts *et al.*, 2002).

The current 'genomic revolution' is generating large amounts of valuable information, primarily in the form of new genome sequences and genome-wide expression data (microarray-transcriptome data), with significant

advances on proteome studies as well (Castrillo and Oliver, 2004 and references therein). However metabolomics, the comprehensive analysis of the complete pool of cellular metabolites (the 'metabolome') closely interacting with the other genomic levels, and directly reflecting the cell's phenotype, is sometimes inadvertently overlooked in post-genomic studies (Adams, 2003; Harrigan and Goodacre, 2003; Goodacre *et al.,* 2004).

In the new post-genomic era studies will progressively have to evolve from the punctual, isolated discovery of biological information to the integration of present and new data in a structured manner, towards the comprehension of the cell as a global entity in which different genomic levels (genome, transcriptome, proteome, metabolome, Oliver *et al.,* 1998; Castrillo and Oliver, 2004) exert their respective functions not independently but interacting coordinately with the others, through specific regulatory mechanisms, direct response to the environmental conditions, in an integrative, 'Systems Biology' perspective (Kitano, 2002; Kafatos and Eisner, 2004).

The purpose of this chapter is to present a comprehensive view of metabolomics as an essential, intrinsic component of integrative studies in the post-genomic era. In the first section of the chapter basic metabolic profiling techniques and applications will be presented. In the second part, relevance of metabolites and metabolic regulation will be reported, along with new mechanisms involving participation of metabolites in global expression and regulatory control. Finally in the last section attention is focused on the favourable characteristics of yeast as a reference model organism for integrative genomic approaches, including metabolomics, for application in Systems Biology studies.

2. METABOLIC PROFILING. EXPERIMENTAL STRATEGIES AND APPLICATIONS

2.1 Methods of analysis of metabolites: Requirements.

The metabolic state of a cell is defined by the identity and concentrations of both intracellular and extracellular metabolites present or acting upon the cell. These will vary in a tightly regulated way in response to the environmental or developmental changes. In order to establish a reliable picture of a cell's metabolic state, covering a wide range of metabolites, comprehensive and efficient methods are required. This is intrinsically difficult due to the heterogeneity of different families of metabolites, their high reactivity (i.e. the turnover rates of intermediary metabolites range from

several seconds to milliseconds; Fell, 1997), and the different ranges of concentrations over which they exert their physiological effects (Table 1 and references therein).

Table 1. Ranges of internal and external metabolite concentrations. Physiological ranges of selected groups of yeast and fungal metabolites (Gancedo and Gancedo, 1973; Atkinson and Mavituna, 1991; Martinez-Force and Benitez, 1991; de Koning and van Dam, 1992).

Metabolites	Range	
	(aerobic)	(anaerobic)
Internal intermediary metabolites		
Glycolytic intermediates (aerobic – anaerobic)	mM	μM
Amino acids	mM	
Nucleotides (AMP, ADP, ATP)	mM	
Vitamins	[μM – mM]	
External metabolites/compounds		
Substrates/nutrients (C, N, P, S sources, mineral salts trace elements, vitamins)	[μM – mM]	
Products (e.g. ethanol, acetate, organic acids)	[μM – mM]	
Secondary metabolites (amino acids, peptides, other signalling molecules, e.g. heterocyclic compounds)	[nM – μM]	

In vivo studies can be applied in limited cases (e.g. fluorescence spectrophotometry, dual beam spectrophotometry or NMR; Fell, 1997), but in the majority of cases it will be necessary to work with extracts and, if the measurements are to truly represent the situation within the living cell, a number of requirements have to be fulfilled. These requirements have been established through the work of several researchers (e.g. Saez and Lagunas, 1976; De Koning and van Dam, 1992; Fell, 1997; Hajjaj *et al.*, 1998; Castrillo *et al.*, 2003) and they can be summarized as:

1) *Fast sampling.* Due to the low turnover rates of metabolites fast sampling (including extracellular medium and cells) coupled to methods to stop further reactions and fix the concentration of metabolites (quenching) is mandatory (Theobald *et al.*, 1993; Fell, 1997; Lange *et al.*, 2001).

2) *Quenching of metabolites.* A number of different methods are used, including rapid drop to low temperatures (-40 °C or lower), sudden pH change or mixing with organic solvents (Fell, 1997; Hajjaj *et al.*, 1998; Castrillo *et al.*,2003; Mashego *et al.*, 2003; Villas-Boas *et al.*, 2003).

3) *Efficient extraction of internal metabolites.* Due to their heterogeneity, there is no universal method that allows the extraction of all metabolites with maximum efficiency. Extraction is usually performed at neutral pH in mixtures of organic compounds (e.g. chloroform) or in boiling ethanol, in order to obtain a representative sample of the variety of chemically

compatible metabolites (e.g. soluble metabolites) present in the cell (Gonzalez *et al.,* 1997; Villas-Boas *et al.,* 2003).

4) *Concentration step.* The quenching and extraction steps result inevitably in the dilution of the metabolites, whose concentration can fall below the sensitivity limit of subsequent analytical techniques. A concentration step is, therefore, necessary. This is usually performed by evaporation of the solvent. After that, the extracts can be stored for short periods at -80 °C but, since different types of metabolites can exhibit different stabilities, immediate analysis is strongly recommended (Castrillo *et al.,* 2003).

5) *Preparation of the sample and analyte determination.* Due to the different ranges of concentrations of metabolites (Table 1) and the dilution and concentration steps inherent to the extraction method, the preparation of the sample from the concentrated extract has to be carefully designed to allow determination of the largest group of metabolites within the dynamic range and sensitivity of the analytical technique to be used. Among the most extensively used are: enzymatic and immunoassays methods (Fell, 1997; Gonzalez *et al.,* 1997), NMR (Brindle *et al.,* 1997; Griffin, 2004), and mass spectrometry methods (e. g. electrospray ionization mass spectrometry, ES-MS; Vaidyanathan *et al.,* 2001; Allen *et al.,* 2003). These can be used with high versatility, either individually (e.g. direct infusion electrospray mass spectrometry; Castrillo *et al.,* 2003) or combined with selected chromatographic techniques (e.g. GC-MS, GC-Q-ToF-MS; Villas-Boas *et al.,* 2003), coupled to tandem mass spectrometry (MS/MS) or even combined with the use of substrate labelling with stable isotopes (e.g. isotopomer ratio analysis of labelled extracts using LC-ES-MS/MS; Mashego *et al.,* 2004). More recently, a significant improvement in the level sensitivity has been obtained by the development of a new mass spectrometry technique, Fourier Transform Ion Cyclotron Mass Spectrometry (FT-ICR) which opens the possibilities to new advanced metabolome studies (Aharoni *et al.,* 2002).

The requirements listed above allow the extraction and analysis of a number of cell metabolites in order to obtain a global picture of the metabolic state of the cell (by high-throughput analysis of global external and internal metabolic profiles). However, eukaryotic cells, like yeast, contain a number of compartments and the internal metabolites are not uniformly distributed among them. For advanced studies, including quantification of metabolites in specific cellular compartments or free and bound metabolites, specific assumptions of relative volumes of water in these different compartments, in addition to well-designed strategies for organelle isolation and analysis regimes are required (Fell, 1997, Farre *et al.,* 2001).

2.2 Metabolic profiling of internal and external metabolites: Applications.

The concentrations and variations in the levels of metabolites reflect the metabolic state of the cell, and the metabolome is considered the closest level of analysis to the cell's phenotype (Oliver, 1997; Trethewey *et al.*, 1999; Raamsdonk *et al.*, 2001). Hence, metabolic profiling is applied to evaluate variations in metabolic states, competing favourably with, or being complementary to, other 'omic techniques (Adams, 2003; Harrigan and Goodacre, 2003). Metabolic profiling of internal metabolites (metabolic fingerprinting) is currently being used in a wide variety of organisms (yeast, plants, mammalian cells) for different applications (Trethewey *et al.*, 1999; Fiehn *et al.*, 2000; Raamsdonk *et al.*, 2001; Watkins and German, 2002). Metabolic profiling of external metabolites (metabolic footprinting) is being increasingly used (Allen *et al.*, 2003; Kell and Mendes, 2000), and more discoveries are sustaining their physiological relevance, not only in microorganisms (Petroski and McCormick, 1992; Demain, 1998) but also in human cell biology (Hebert, 2004). In functional genomics studies, new methods for metabolic profiling in different organisms (Fiehn *et al.*, 2000; Watkins and German, 2002; Adams, 2003) are used for the elucidation of the function of new genes and metabolic pathways (Teusink *et al.*, 1998; Raamsdonk *et al.*, 2001; Trethewey, 2001; de la Fuente *et al.*, 2002; Weckwerth and Fiehn, 2002). For applied purposes metabolic profiling is used in the investigation of molecules for nutritional assessments (e.g. studies on the interaction of diet and health, or for the assessment of GM foods), evaluation of health and disease states (biomarkers, e g. in cancer cells) for application in diagnostics, as indicators of disease progression and for the screening of new drugs (Griffin *et al.*, 2001; Schilter and Constable, 2002; Watkins and German, 2002; Fiehn and Spranger, 2003; Griffin and Shockcor, 2004; Lee and Boros, 2003; Heaton *et al.*, 1999,; Kaddurah-Daouk and Kristal, 2001; Stockton *et al.*, 2002).

3. METABOLOMIC STUDIES IN FUNCTIONAL GENOMICS

3.1 Role of metabolism and metabolites in Functional Genomics: Regulation.

Primary metabolism can be defined as the coordinated biochemical conversion of substrates through tightly regulated metabolic pathways in

order to generate energy and building blocks for growth and the maintenance of cellular functions. It is usually divided into catabolism and anabolism with participation of common amphibolic reactions (Lehninger, 1975; Castrillo and Oliver, 2004). Based on this definition only, the role of metabolism and metabolites in Functional Genomics could be underestimated, and be considered of secondary importance to the flow of genetic information and the regulation of gene expression. In the flow of information from gene (DNA) to RNA to proteins (e.g. enzymes, which catalyse the specific metabolic reactions) metabolites could be regarded as inert molecules with negligible participation in regulation. However, a comprehensive revision on participation of metabolites in regulation and control offers a more complete perspective of the importance of metabolomics in Functional Genomics, as can be seen from the following observations:

1) *Central metabolic pathways.* Internal metabolites exert rapid short-term regulation of metabolic fluxes by modulation of enzymatic activity. The changes in fluxes along the major metabolic pathways have long been reported to be tightly regulated by the concentration of specific internal metabolites (e.g. fructose-1,6-diphosphate, ATP, ADP, citrate) through rapid activation and inhibition of key enzymes by reversible covalent modification as well as by allosteric effects (metabolic effectors; see e.g. Monod *et al.*, 1963; Fell, 1997; Muller *et al.*, 2003; Plaxton, 2004). These key metabolites (e.g. sugar-phosphates, adenylates, cAMP), which collectively regulate carbohydrate metabolism, have no direct involvement in carbon regulation of gene expression. In these cases, assimilation of carbon nutrients is regulated by specific sensing and signal transduction pathways involving other specific protagonists.

2) *External signals - metabolite sensors.* A cell has to maintain the stability of the intracellular environment (homeostasis) in response to changes in the external conditions. The nature and variations of levels of external metabolites (i.e. substrates, sometimes called catabolites; products; other external compounds) constitute the primary level of environmental information (signals) detected by the cell through its specific sensing mechanisms (usually by means of metabolite-protein interactions, ligand-receptor at the membrane level; Hancock, 1997).

3) *Signal transduction pathways - internal metabolites.* Once an external signal (presence, absence or change in metabolite concentrations) is detected, intracellular signal transduction pathways are triggered (Hancock, 1997; Sprague *et al.*, 2004). In the widely accepted model of mechanism, the metabolite binds to a specific protein which can modify other regulatory proteins post-transcriptionally, resulting in changes in the levels and/or mechanisms of action of other regulatory proteins (e.g. transcription factors)