MEETING REPORT

Plant Metabolomics: The Missing Link in Functional Genomics Strategies

After the establishment of technologies for high-throughput DNA sequencing (genomics), gene expression analysis (transcriptomics), and protein analysis (proteomics), the remaining functional genomics challenge is that of metabolomics. Metabolomics is the term coined for essentially comprehensive, nonbiased, high-throughput analyses of complex metabolite mixtures typical of plant extracts.

This potentially holistic approach to metabolome analysis is driven primarily by recent advances in mass spectrometry (MS) technology and by the goals of functional genomics research. Achieving the broadest overview of metabolic composition is very complex and entails establishing a multifaceted, fully integrated strategy for optimal sample extraction, metabolite separation/detection/identification, automated data gathering/handling/analysis, and, ultimately, quantification. Both analytical and computational developments are essential to achieve this goal.

The First International Congress on Plant Metabolomics was held in Wageningen, The Netherlands, in April 2002, with the primary goal of bringing together those players who are already active in this field and those who soon plan to be. In so doing, opportunities are created for collaboration, overlap can be avoided, and joint strategies can be determined to meet the metabolomics challenge. Indexed abstracts from the oral and poster presentations at the meeting are now accessible at www.metabolomics.nl, and this site will continue to be used as an aid to information exchange and enhanced collaboration.

Although microbes may prove to be the richest overall source of metabolites, plants are the source of the most complex individual mixtures. Mariet van der Werf (TNO-Food, Zeist, The Netherlands) reported that it has been predicted that bacterial genomes already sequenced can support the biosynthesis of just a few hundred metabolites (e.g., 580 for Bacillus subtilis and 800 for Escherichia coli), but for individual plants, this value is likely to be in the tens of thousands.

This metabolic richness comes not just from the number of genes present (20,000 to 50,000) but also from multiple substrate specificities for many enzymes (Aharoni et al., 2000), subcellular compartmentation, and the occurrence of nonenzymic reactions. Approximately 50,000 different compounds have been elucidated in plants (De Luca and St. Pierre, 2000), and it is predicted that the final figure for the plant kingdom will approach or even exceed 200,000 (Pichersky and Gang, 2000; Fiehn, 2001, 2002). Thus, metabolomics represents a considerable challenge for plant scientists.

Metabolomics research will prove an invaluable tool for generating information of use in many fields. For functional genomics strategies, potentially fast-track methods exploiting metabolomics analyses of tagged lines or known mutants are likely to prove invaluable (Motoko Awazuhara, Chiba University, Japan [Arabidopsis]; Andy Pereira, Plant Research International, Wageningen, The Netherlands; and Jon Lightner, Exelixis Plant Sciences, Portland, OR [Arabidopsis and tomato]). Metabolomics information not only will assist in the establishment of a deeper understanding of the complex interactive nature of plant metabolic networks and their responses to environmental and genetic change but also will provide unique insights into the fundamental nature of plant phenotypes in relation to development, physiology, tissue identity, resistance, biodiversity, etc.

Some key conference presentations are described below. This report is organized according to categories representing the fields of greatest importance to the successful establishment of plant metabolomics as a technology complementary to those for gene expression and protein profiling already in existence.

TECHNOLOGY DEVELOPMENT

Most technology for metabolomics is based on MS, which was reinforced by several presentations at the meeting. Gas chromatography–MS (GC-MS) and HPLC–photodiode array–MS remain the methods of choice for quantitative and qualitative metabolite profiling. The ultimate goal of metabolomics—the ability to reliably detect and quantify every metabolite in a plant extract—is unlikely to be attained by any single analytical method available at present. Some metabolite selection occurs in all methodologies, from initial solvent extraction through chromatography to MS ionization. Nevertheless, new advances have been made.

Oliver Fiehn (Max Planck Institute for Molecular Plant Physiology, Potsdam, Germany) reported the use of rapid scanning time-of-flight (TOF) MS coupled with GC separation and integrated with peak deconvolution software. This technique increased the number of metabolites detectable by GC-MS in crude plant extracts to 500 to 1000. However, the dynamic range of TOF detectors is still restrictive when faced with mixtures containing compounds with concentration differences of several orders of magnitude.

Two “new” technological approaches, not involving chromatography of metabolites, were highlighted during the congress. These were NMR analysis of crude extracts and direct examination of crude extracts by MS, either in the form of quadrupole (Q)
MEETING REPORT

TOF-MS or ultra-high-resolution Fourier transform ion cyclotron MS (FT-MS). Direct infusion of extracts into MS instruments using “soft” electrospray or atmospheric pressure chemical ionization sources is an attractive way to generate “fingerprints” of the molecular ions of the metabolites present. Such use of FT-MS was demonstrated in presentations by Asaph Aharoni (Plant Research International, The Netherlands; Aharoni et al., 2002) and Dayan Goodenowe (Phenome- nome Discoveries, Saskatoon, Canada). This powerful, and relatively expensive, mass analyzer is capable of generating mass data of sufficient accuracy for definitive empirical formula determination. Several hundred ions were observed for each ionization method in both positive and negative modes.

Lively debate regarding the validity of this technology ensued, with particular reference to the lack of differentiation between isomers, the presence of fragment and adduct ions, and problems of quantification caused by ion suppression. Such problems are not unique to FT-MS, and the combination of FT-MS data with other chromatographic MS data is potentially powerful.

Several presenters from the United Kingdom (Marianne Defernez, Institute of Food Research, Norwich; Mike Beale, Institute of Arable Crops Research-Long Ashton; Nigel Bailey, Imperial College, London; and Adrian Charlton, Central Institute of Arable Crops Research-Long Ashton, Wageningen, The Netherlands) described the use of proton NMR of crude plant extracts, followed by multivariate analysis, to cluster data sets to highlight differences. This type of analysis gives a comprehensive summation fingerprint of all (hydrogen-containing) metabolites extracted and can provide direct structural information regarding individual metabolites in the mixture, particularly when two-dimensional techniques are applied. Therefore, it is suitable for high-throughput, rapid, first-pass screening. Subtraction of data sets generates virtual NMR spectra and hence important structural data on compound(s) contributing to differences between samples.

DATA HANDLING

A number of contributors described “industrial-scale” throughput or multiple-partner collaborations in which the need was evident for whole-process data capture integrated with the Laboratory Information Management System (LIMS) and followed through to well-structured archives and databases. Different groups reported generating volumes of data on the order of 10 gigabytes per day, and a number of trends suggest that the volume of data produced will continue to increase. Richard Trethewey (Metanomics, Berlin, Germany) discussed faster analytical machines and accelerated plant life cycles, precision cultivation of large numbers of replicates, enhanced biological resolution (organelle separation), more extraction and detection procedures, and the identification of more analytes.

Mariet van der Werf emphasized the need for an increased ability to analyze more replicates with better-defined biological questions. A number of large commercially based data-handling systems were described. Eve Wurtele (Iowa State University, Ames) described the development of the publicly available GeneExpression- Toolkit (http://www.math.iastate.edu/danwell/ GET/GET.html). This software package aims to integrate data from literature, microarray, proteomics, and metabolomics analyses using fuzzy cognitive maps to extract metabolic and regulatory networks.

Data Preprocessing

Data from most analytical instruments require significant preprocessing before they can be analyzed statistically. Standardization of techniques and quality control are necessary, and many contributors emphasized the need for post-sample-collection techniques such as noise reduction, deconvolution, profile alignment, reference to internal standards, and peak labeling using spectral libraries. Oliver Fieth reported better rates of peak identification from GC-TOF with a commercial program than with the often-used Atomic and Molecular Data Information System. A number of contributors noted profile alignment issues, and Arjen Lommen (State Institute for Quality Control of Agricultural Products, Wageningen, The Netherlands) reported significant progress with automatic comparison of NMR and full-scan MS data after data alignment.

Databases

Presentations (particularly one by Pedro Mendes, Virginia Bioinformatics Institute, Blacksburg, VA) and contributions during the workshop on database issues identified a range of types of database and the desire that these should be well integrated. This would ensure consistency of terminology and standardization of interchange formats and allow querying across databases. Sets of related profiles from contrasting analytical procedures must be linked with the experimental design, because these biological and environmental data are necessary for meaningful interpretation. There was general agreement at the workshop that such databases will be based on standard commercial relational database management systems and that object-oriented systems are unlikely to be appropriate.

Statistical Analysis

The use of some univariate statistics was apparent, with a number of presentations applying analysis of variance to particular metabolites in different samples. All pairwise correlations of the phenolics in birch samples under study were investigated by Jyrki Loponen (University of Turku, Finland), and mining of coresponses in yeast was described by Steve Oliver (University of Manchester, UK). Multivariate statistics commonly means principal component analysis, which appears to be the default technique for many in the field.

Hierarchical cluster analysis was used by a number of contributors, and some
used partial least-squares and linear discriminant analysis. Mining with evolutionary computing techniques was demonstrated by Roy Goodacre (University of Wales, Aberystwyth, UK), and genetic algorithms and fuzzy cognitive maps were discussed by Eve Wurtele. Mike Beale described the addition of data sets from well-characterized mutants to a comparative analysis database with the aim of making confident predictions of areas of metabolic abnormality.

The consensus seemed to be that there are challenges across the range of data-handling tasks. Presentation of the conclusions of the database workshop and continuing discussions on the topic will be facilitated via the metabolomics World Wide Web site (www.metabolomics.nl).

METABOLIC PROFILING AND FINGERPRINTING

How can metabolomic approaches be used? This question can be answered only if terms are defined carefully so as not to confuse analytical strategies aimed at answering different biological questions. For example, discrimination between plant genotypes or bacterial pools can be achieved using metabolite fingerprinting without the need for compound identification. Roy Goodacre demonstrated this using Fourier transform infrared spectroscopy in 384-well plates with a throughput capacity of 7000 bacterial or plant samples per day followed by data mining using genetic programming (Goodacre et al., 2000).

Metabolite profiling aims at a quantitative assessment of a predefined number of target metabolites. Often, such profiles are restricted to certain pathways or compound classes. For example, Tony Larson (University of York, UK) presented liquid chromatography (LC)-based profiling methods for low-level (femtomole) detection of CoA-acyl esters. Lipid profiling also was the focus of an analysis of glossy mutants of maize with altered composition of epicuticular waxes. Basil Nikolau (Iowa State University, Ames) described the complexity of wax layer biosynthesis. The wax extraction method was optimized to rapid (30 to 60 s) hexane extractions, resulting in clear GC scans for fatty alcohols, aldehydes, esters, alkanes, ketones, and acids. Most surprisingly, differences in wax layer composition were observed in a cross between maize inbred lines (B73 and A188) in the F1 generation. Maternal inheritance for wax composition was seemingly different from paternal heredity, and both F1 lines were quite different from the parental background.

Metabolomic analysis attempts to avoid bias against certain compound classes and to allow for the analysis of every metabolite individually. Although this aim has not been reached, clear progress in this direction was reported. For example, a combination of different metabolite-profiling tools was used by Ric de Vos (Plant Research International, Wageningen, The Netherlands) for the analysis of metabolic effects in high-flavonoid genetically modified tomatoes.

Using LC/photodiode array detection as well as LC-QTOF-MS and direct infusion–QTOF–MS, it was confirmed that flavonoid contents were up to 70-fold higher compared with those of common cultivars, and glycosidic structures were presented for all aglycone intermediates that accumulated. These analyses were performed in tandem with GC-MS profiling of volatile compounds using headspace solid-phase microextraction. Surprisingly, in addition to an increase in flavonoids, the volatile compound methylsalicylate was found to be increased in transgenic plants overexpressing the Lc/C1 transcription factor.

Oliver Fiehn presented metabolomic studies using GC-TOF and LC-Ion Trap MS analysis to assess network effects in primary metabolism in potato plants by metabolite–metabolite coreresponse analysis. An effect of the antisense inhibition of Suc synthase, which was associated with an apparently silent phenotype, was characterized by a comparison with the corresponding Desiree (wild-type) cultivar with respect to network sizes and relationships within these networks. Apparent changes in amino acid pathways were presented, and hypotheses were derived for the regulation of several classes of carbohydrates.

MEtabolomics in the realm of integrated functional genomics

In the “postgenomics” era, there is a strong movement toward the functional characterization of sequenced genomes and comprehensive investigations of biological systems in response to external stimuli. To achieve these goals, some groups are adopting integrated approaches that include analyses at multiple levels, including the genome, transcriptome, proteome, and now also the metabolome.

For example, Steve Oliver emphasized the need for comprehensive profiling methods of the transcriptome, proteome, and metabolome for the identification of gene function and the further delineation of known metabolic genes. He also discussed a strategy designated the functional analysis by coreresponses in yeast (FANCY) and its use in the functional characterization of Saccharomyces cerevisiae deletion mutants. Coreresponses of unknown gene deletions measured by NMR can be correlated to similar responses of known gene deletions for inference of function. Although this work focused on yeast, the concepts and approach presumably would be identical in plants.

Lloyd Sumner (Samuel Roberts Noble Foundation, Ardmore, OK) described a divisional program involving integrated functional genomics of the model legume Medicago truncatula. Initial data were presented for the profiling and identification of >300 proteins from specific tissues of M. truncatula. Multiple metabolic profiling technologies were emphasized for greater visualization of the metabolome. Sumner provided specific examples, including LC-UV-MS of flavonoids and saponins, GC-MS of polar and lipophilic extracts, and capillary electrophoresis-MS of anions and
MEETING REPORT

amino acids. Metabolite profiling also was illustrated in the study of temporal development.

Denise Jacobs (Leiden University, The Netherlands) presented data that correlated proteome changes with alkaloid accumulation in periwinkle (Catharanthus roseus) cell cultures, which showed that as many as one-third of the 2000 proteins visualized by two-dimensional gel electrophoresis were correlated with alkaloid accumulation. Several protein identifications via matrix-assisted laser-desorption ionization (MALDI)-TOF were presented. However, Jacobs emphasized the limitations in protein identification using MALDI-TOF peptide mass mapping and discussed future approaches.

Several other groups are involved with or are developing integrated functional genomics programs. Integrated academic programs include those at Iowa State University (Eve Wurtele and Basil Nikolau), the University of Massachusetts (Yuen Yee Tam and Jennifer Normany), the Max Planck Institute (Oliver Fiehn and Wolfram Weckwerth), and Plant Research International (Asaph Aharoni, Harrie Verhoeven, Ric de Vos, Raoul Bino, and Robert Hall). Integrated commercial programs were discussed in detail by Scott Harrison (Paradigm, Research Triangle Park, NC) and Ji-Sook Song (Unigen/Eugentech, Seoul, Korea). Furthermore, a start is being made on linking metabolomics data to biological activity relevant to the food industry (Dries de Bont, Numico, Wageningen, The Netherlands; Claire Daykin, Unilever, Vlaardingen, The Netherlands).

Finally, a prevailing thought was voiced suggesting that integrated functional genomics will be truly successful only if corresponding informatics tools are generated and implemented (see Data Handling).

THE FUTURE

It is clear that plant metabolomics is in its infancy and that there is still a great deal to do. However, an excellent foundation upon which to build and develop metabolomics into a key biological tool already exists. The technical challenges are significant, but the tremendous enthusiasm to meet these challenges is clearly evident. Nevertheless, simple limitations such as the nonavailability of reference compounds and the need for appropriate, dedicated bioinformatics tools represent major challenges, and these can be approached effectively with sufficient speed only through a coordinated and collaborative effort.

This meeting made important progress in this effort, and continued interaction will be facilitated through the metabolomics World Wide Web site (www.metabolomics.nl). A working group has been initiated to assist in future coordinated efforts (see the World Wide Web site for details). Follow-up conferences in Germany in April 2003 (organized by Oliver Fiehn) and in the United States in 2004 (organized by Basil Nikolau) will maintain the necessary momentum. The primary biological driving force will always be the value of the information produced and its application in many fields of genetics and biology. In the future, metabolomics will play a key role in complementing data sets obtained from the existing “-omics” technologies.

REFERENCES


Robert Hall
Plant Research International
BU Cell Cybernetics
P.O. Box 16, 6700 AA
Wageningen, The Netherlands
r.d.hall@plant.wag-ur.nl

Mike Beale
IACR-Long Ashton
Bristol BS41 9AF, United Kingdom

Oliver Fiehn
MPI Molecular Plant Physiology
Postfach 14424
Potsdam, Germany

Nigel Hardy
Department of Computer Science
University of Wales, Penglais
Aberystwyth SY23 3DB, United Kingdom

Lloyd Sumner
Samuel Roberts Nobel Foundation
Plant Biology, P.O. Box 2180
Ardmore, OK 73402

Raoul Bino
Plant Research International
BU Cell Cybernetics
P.O. Box 16, 6700 AA
Wageningen, The Netherlands