Annu. Rev. Plant Biol. 2003. 54:669–89
doi: 10.1146/annurev.arplant.54.031902.135014

METABOLOMICS IN SYSTEMS BIOLOGY

Wolfram Weckwerth
Max-Planck-Institut für Molekulare Pflanzenphysiologie, 14424 Potsdam, Germany; email: weckwerth@mpimp-golm.mpg.de

Key Words integrative biochemical profiling, quantitative proteomics, dynamic networks, system modeling, stochastic noise, fluctuation

■ **Abstract** The primary aim of "omic" technologies is the nontargeted identification of all gene products (transcripts, proteins, and metabolites) present in a specific biological sample. By their nature, these technologies reveal unexpected properties of biological systems. A second and more challenging aspect of omic technologies is the refined analysis of quantitative dynamics in biological systems.

For metabolomics, gas and liquid chromatography coupled to mass spectrometry are well suited for coping with high sample numbers in reliable measurement times with respect to both technical accuracy and the identification and quantitation of small-molecular-weight metabolites. This potential is a prerequisite for the analysis of dynamic systems. Thus, metabolomics is a key technology for systems biology.

The aim of this review is to (*a*) provide an in-depth overview about metabolomic technology, (*b*) explore how metabolomic networks can be connected to the underlying reaction pathway structure, and (*c*) discuss the need to investigate integrative biochemical networks.

CONTENTS

INTRODUCTION

In the context of functional genomics, the nontargeted profiling of metabolites in biological samples is now regarded as a viable counterpart to protein and transcript profiling technologies (6, 31, 98, 99). The integration of methods based on gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS) for the comprehensive identification and, particularly, the accurate quantification of metabolites has attained a technical robustness that is comparable or even better than conventional mRNA or protein profiling technologies (24, 25, 72, 73, 78, 96, 106).

Most promising is the driving force of this technology to move from qualitative to comparative quantitative approaches. However, many critical parameters, such as the discrepancy between the low number of detected metabolites versus the real number of possible metabolites in plants, the extraction process, the bias against compound classes, and, most importantly, the overlap of many compartmentalized metabolic processes in tissue samples, are still unresolved and complicate the interpretation of metabolite profiles.

Nevertheless, the accurate identification and the relative quantification of a high number of metabolites in a multitude of samples make it possible to study dynamic metabolomic networks and are thus leading to observations unattainable using classical methods. Analysis of these network topologies and their control with respect to specified environmental or genetic perturbations will permit the investigation of dynamic interactions in metabolic networks and the discovery of new correlations with biochemically characterized pathways as well as pathways hitherto unknown.

Owing to our incomplete knowledge of quantitative mRNA-protein-metabolite interactions, integrative profiling approaches combining metabolomics, proteomics, and transcriptomics will greatly enhance our ability to determine relationships among components of plant systems. Furthermore, integrating quantitative data regarding growth stage and environmental conditions into analyses of the biological system is essential (12, 16). This comprehensive approach will play a major role in understanding regulation and biochemical interactions in plant metabolism at a systems level.

PLANT SYSTEMS

Plants are sessile systems unable to escape environmental pressures. As a result, they have evolved a dazzling array of flexibility in their responses to environmental conditions such as light/dark, drought, temperature, nutritional supply, microbial invasion, etc. Thus, the plant system comprises a genotype by environment response, producing a specific geno-phenotype relationship that is heavily dependent on the growth stage (12, 16). Accordingly, a gene's function should ideally be defined in the context of the systems state and environment.

Owing to this complexity, following systems responses at the molecular level using transcript and protein profiling (for instance, when a single plant gene is manipulated) is difficult. However, these primary omic technologies have been applied at the whole and subsystem levels to identify individual genes or proteins that show differential expression in response to systematic perturbation (66, 111). In most studies to date, changes in metabolite levels in response to altered gene and protein expression have not generally been monitored. This is at least in part due to the underlying paradigm that genes drive the system, whereas gene products (proteins and metabolites) are merely along for the ride. This model is inadequate because the regulation and control of metabolic fluxes clearly occur on all levels, as shown in a case study for the regulation of glycolysis (94).

Since the completion of the genome sequence, the gene annotation of *Arabidopsis thaliana* (http://luggagefast.stanford.edu/group/arabprotein/; http://mips.gsf.de/proj/thal/; http://www.tigr.org/tdb/e2k1/ath1/) provides a reasonable framework upon which to construct a theoretical metabolic pathway network. The situation is complicated by the fact that the function of more than 40% of these genes remains unknown (84). Consequently, metabolic reconstructions only provide pieced together bits of biochemical networks with many missing links. On the other hand, these reconstructions point the way to hypothetical pathways and connections (80, 81). What is missing in these model networks is a preferred pathway structure (61), which is constantly changing in response to developmental needs, environmental conditions, intra- and intercellular transients, etc. (Figure 1).

Plant tissues consist of heterogeneous cell populations and multiple cell compartments. Consequently, multiple metabolite concentration gradients and duplicate pathways are present in tissue samples. Moreover, the term steady state may be a misnomer because plants exhibit constantly changing, transient behavior in response to diurnal, circadian, and seasonal cycles (39, 42, 90, 93). These plant-specific properties of highly compartmentalized metabolic networks and the resulting complexity of metabolite connectivity present great challenges for metabolomics.

METABOLOMICS: GENOTYPE/PHENOTYPE CLASSIFICATION AND PATTERN RECOGNITION

The aim of metabolomic analysis in plant biochemistry/biotechnology is to provide comprehensive insight into the metabolic state of the plant by detecting the metabolome—the full suite of metabolites expressed in a plant (65, 100). There is a long tradition of and extensive knowledge about metabolite analysis. The application of GC/MS and LC/MS profiling to plant phenotyping (24, 25, 72, 73, 78, 96, 106) for the simultaneous detection of individual compound classes with a complex diversity of chemical properties/behavior was a major step in the development of metabolomic technology.

An excellent introduction to the field and a definition of the terms target metabolite profiling and metabolic fingerprinting are given in Reference 23. Metabolite

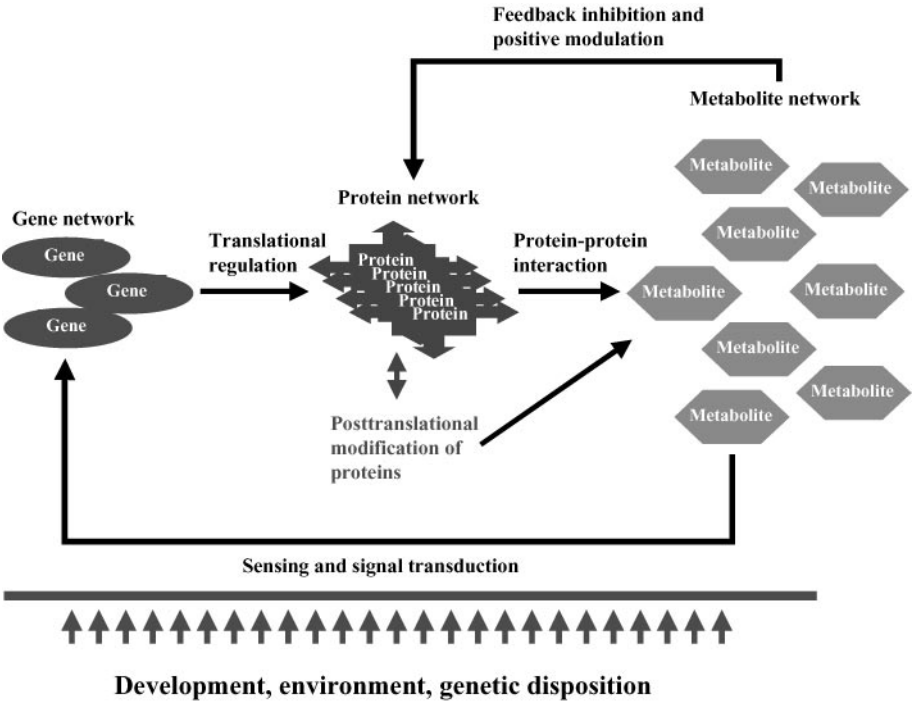

Figure 1 Amplification of a metabolic network and feedback regulation in response to developmental and environmental conditions.

target analysis utilizes specialized protocols for difficult analytes such as phytohormones, whereas metabolite profiling aims at quantifying a number of predefined targets (for example, all the metabolites of a specific pathway or a set of metabolites shared among different pathways). In contrast, the goal of metabolomic approaches is the unbiased identification and quantitation of all the metabolites present in a specific biological sample from a plant grown under defined conditions. This is different from metabolic fingerprinting, which focuses on collecting and analyzing data from crude metabolite mixtures to rapidly classify samples instead of separating individual metabolites by physical parameters. Thus, metabolomics is best suited for the investigation of metabolic networks via the quantitation of individual metabolites without a bias concerning the choice of targets to be analyzed.

The number of metabolites present in the plant kingdom is estimated to exceed 200,000—an enormous number indicating a great deal of compound structural diversity. Owing to this diversity, metabolomic approaches must apply adequate tissue sampling, homogenization, extraction, storage, and sample preparation methods in order to maintain an unbiased process. Currently, no comprehensive comparisons of extraction techniques that show high reproducibility, robustness,

and recovery for all classes of compounds have been published. For example, multiple components from homogenized tissues often are extracted using alcohols or water/alcohol mixtures (48, 91), but no systematic and rigorous validation (55) for extremes in plant tissues (such as leafs, roots, or needles) has been published. The same is true for other extraction techniques such as pressurized liquid extraction (8), supercritical fluid extraction (10, 14, 45), sonication (77), subcritical water extraction (26), microwave techniques (62), and pervaporation (85). Additionally, it is unclear what factors most affect robustness, which is defined by minimal analytical errors if protocols are carried out under slightly altered conditions. Such alterations may include subtle differences in extraction times, temperatures, solvent compositions and qualities, staff skills, tissue/solvent ratios, and others, with the potential to cause severe problems in reproducing results.

Numerous techniques exist for metabolite detection. It is questionable if data acquisition of a single physical parameter can fulfil the minimal requirements of metabolomic approaches, i.e., comprehensiveness, selectivity, and sensitivity. MS seems to be the primary candidate to fulfill these criteria, as much work has demonstrated its suitability for metabolite detection in complex matrices. However, it is well known that GC/MS, for example, is not suitable for organic diphosphates, cofactors, or metabolites larger than tri- to tetrasaccharides. Capillary electrophoresis (CE) coupled to MS provides a feasible method for the separation of ionic compounds and, hence, represents a complementary technology to reversed phase separation focusing on lipophilic compounds (83).

In a recent elegant analysis of oligosaccharides and sugar nucleotides in phloem exudates, LC/MS coupling was achieved using hydrophilic interaction chromatography, resulting in better peak shapes compared with normal phase LC (96). A comparative study using negative electro spray ionization (ESI) and LC/MS/MS revealed the diversity of saponins in different *Medicago* species (41). These authors concentrated on segregating the metabolome into several subclasses followed by parallel analysis utilizing the selectivity of MS. Alternatively, LC/LC coupling of different chromatographic columns prior to metabolite detection might work for metabolomic approaches; however, no method has been developed that is as successful as coupling ion exchange to reverse phase LC in peptide mixture analysis (105, 110).

It is obvious that GC/MS, LC/MS, LC/LC/MS, and CE/MS approaches have intrinsic biases against certain classes of compounds. For example, simple terpenes, carotenoids, or aliphatics are semi-inert to ESI, the standard technique used in conjunction with LC. Such hydrocarbons, however, are often volatile and can therefore easily be detected by a combination of GC and MS, for example, by using classical electron impact ionization. In this regard, a combination of GC/MS and LC/MS methods is adequate for analyzing a wide range of metabolites. However, especially for LC/MS, one must not forget the effects of ion suppression due to matrix effects (15, 88). Ion suppression can only be partly circumvented by reducing the size of liquid droplets (5), and it might invalidate any metabolomic approach that fails to properly preseparate metabolites prior to MS detection. Loss

of information by ion suppression and matrix effects can be circumvented by using methods complementary to MS detection in parallel. This can be accomplished by splitting LC flows (e.g., to electrochemical detection), nuclear magnetic resonance, or infrared, ultraviolet, or fluorescence spectroscopy. Ultraviolet and fluorescence detection are well-known, nondestructive tools for use in metabolite target analysis or for profiling selected classes of compounds such as amines (11), isoprenoids (74), or unsaturated fatty acids (28). Coulometric electrochemical array detectors, which are powerful and sensitive detectors of carotenoids (21), polyphenols (13), and flavonoids, could also be applied to circumvent the limitations of MS. Notably, these detectors also enable distinction of metabolite isomers (20) by spectral information, which is often hard to do using MS.

Alternatively, one might consider nuclear magnetic resonance spectroscopy (NMR). Although these approaches lack sensitivity for multiparallel analysis of hundreds of metabolites for generations of large metabolic networks, NMR has a high potential for unraveling metabolic fluxes in branched, short pathways if carried out with isotope labeling and metabolic flux balancing calculations (75, 92, 108).

Using matrix algebra and the assumption that the system reaches a steady state simplifies the formulation and the solution of complex networks. The resulting balance equations yield a predicted set of isotopomer balances for a given flux distribution. Fluxes can be iteratively modified until the flux distribution is in closest accordance with experimental observations (75).

In addition, NMR (69) spectroscopy has high discriminatory power on the level of metabolic fingerprints, for example, for rapidly assessing the mode of action of plant protectants (2). A disadvantage is that most metabolites result in several to many signals in NMR, causing a large problem in resolving individual metabolites if no chromatographic separation is performed before NMR detection and identification. Therefore, only limited information can be derived from NMR fingerprints with respect to individual metabolites. For instance, silent yeast mutants were readily discriminated from wild-type genotypes using NMR fingerprinting (69), but the data have to be complemented by classical enzyme assays to derive any biochemically meaningful conclusion. Besides NMR, direct-infusion MS is ideally suited for high-throughput classifications of sample origins (101). In such applications, partial ion suppression in electro spray ionization may eventually be helpful because it can amplify slight matrix differences, in addition to metabolites that are more abundant or unique in one of the sample populations under study.

For the metabolomic approach, GC/MS profiling of plant extracts has reached a high degree of accuracy in respect to extraction, modification, time of measurement and identification, and quantification of individual metabolites (24, 25, 30, 72, 73, 106). Recent advances with respect to fast acquisitions as well as accurate mass determinations have been achieved by applying time-of-flight (TOF) technology (103). Additionally, powerful deconvolution algorithms (86, 87, 97) have been developed to find peaks without prior knowledge of their abundance, mass spectral characteristics, or retention time. Such deconvolution algorithms still do not work

in LC/MS but are urgently needed for the unbiased analysis of arbitrary complex mixtures.

Owing to a relative short acquisition time per sample (the actual time is 30 min) and automated peak identification and quantification based on characteristic compound identifier, the accurate measurement of a high number of replicates is possible using a GC/TOF system. Conceptually, this technique allows high sample throughput (103, 106). High numbers of samples are a prerequisite for generating statistically significant data. Most importantly, this instrumentation allows the accurate quantification of all co-eluting compounds irrespective of whether they are identified, unknown, or classified on the basis of characteristic fragment masses in the spectrum (86, 103). This is similar to the classical proteomics approach in which proteins are first profiled at the highest attainable resolution [up to 10,000 spots per two-dimensional gel (P. Jungblut, personal communication)], and then qualitative and quantitative differences are compared to reveal key response points to genetic or environmental perturbations (49, 56, 68, 79).

Consequently, GC/MS-based identification and quantification of individual compounds, as well as data from metabolic fingerprinting, can be used for pattern recognition and classification analysis. Recent work (24) demonstrated the discriminating power of this approach. The largest metabolic variations were found between natural *Arabidopsis* accessions (ecotypes), not between mutants and their corresponding background lines. The detection of a huge number of unknown metabolites was achieved in parallel with the accurate identification and quantification of known metabolites. Clear matrix effects were observed for accurate quantifications of standards. However, owing to the unambiguous ability to discriminate among different mutant plants and ecotypes, the information content of each individual metabolite is characteristic of the underlying metabolic regulation. Accordingly, these methods, in contrast to fingerprinting methods, have the potential to link biological questions with involved individual compounds (24, 25, 51, 72, 73, 106) and to indicate compound interrelatedness to surrounding components (see Integrative Profiling of Biological Samples, below).

TECHNICAL AND BIOLOGICAL VARIABILITY

The components of dynamic biochemical networks exhibit high biological variability owing to the inherent noise, fluctuations, transients, and oscillations that are part of plant homeostasis (7, 42, 52, 95, 104). Using metabolomic technology, one is capable of analyzing biological variability in samples in a nonbiased way.

The overall relative technical standard deviation for metabolite quantification is found to be ∼10% with a novel standardized GC/TOF analysis (W. Weckwerth and O. Fiehn, unpublished data). In that respect, arabidopsis samples show a much larger biological fluctuation in their metabolite levels—up to several fold variances. This phenomenon is usually neglected in biological systems using average metabolite levels from an adequate number of samples assuming that the same genotype

ideally should produce the same steady state or reach the same equilibrium under highly controlled conditions. However, stochastic fluctuations of metabolites as well as proteins and mRNA might account for deviations from steady state. These biological variations have to be considered within independent samples of the same genotype (112). One of the most important applications in omic technology is to compare two types of samples, a control sample and a treated or genetically modified sample, to identify individual components showing differential behavior and to therefore account for the responses of the system to the applied perturbation. This comparative analysis relies on the statistically significant detection of differences between sample groups (112). A high biological variation of individual compounds within a set of samples from the same background will hamper this approach unless a high number of replicates are used. Can we make use of this high variability in the same genotype under controlled environmental conditions for network analysis?

SYSTEM MODELING AND STOCHASTIC NOISE

Metabolites are linked via anabolic and catabolic reaction networks. By using metabolic reconstruction based on genome annotation or biochemical knowledge, researchers can investigate theoretical networks and their structural features (1, 19, 46, 47, 80). However, these predications fall short because regulatory events are precluded (although protein modifications such as phosphorylation reactions can be included into the theoretical network). Thus, we are still searching for the metabolic network that comprises the real cell state, the instantaneous active pathway, in a complex theoretical network under certain conditions (32, 61).

In any case, theoretical reaction and/or regulatory networks provide the primary background for any system modeling approach. Vance et al. (102) demonstrated how fluctuations in metabolites propagate through a theoretical reaction network, enabling the investigation of causal connectivity and overall network structure. This work is based on an approach by Arkin et al. (4) in which fluctuating input concentrations of metabolites into a small in vitro reaction network were used to examine time-dependent correlations among levels of metabolites (4). The determined correlations enabled metabolic distance maps of the entire system to be constructed, providing evidence for the structure of the underlying reaction network.

In a series of papers these inherent properties of stochastic fluctuation, noise, and/or oscillation as a special case in metabolic networks are analyzed in-depth in reference to all levels of metabolism (7, 34, 42, 52, 95, 105).

Particularly exemplary for plant systems biology is the investigation of circadian regulation in crassulacean acid metabolism (CAM) switching from net malic acid accumulation to net malic acid mobilization. Most surprising was the finding that CAM is also a free-running process under constant environmental conditions, raising the question of this biological clock's mechanisms (58). In conjunction

with this question Rascher et al. (70) showed that the circadian rhythm of the metabolic cycle of malate accumulation and mobilization is expressed as dynamic patterns of independently initiated variations in photosynthetic efficiency within a single leaf. This patchiness of photosynthetic activity correlated with malate accumulation and mobilization, and the resulting CO_2 concentration gradients in the tissue led to the identification of spatiotemporal variations of metabolism responsible for the switch of metabolic states. A further step was the investigation of stochastic noise and its influence on the regular oscillating modes of CAM (7). In a system in which endogenous rhythmicity is produced by a beat oscillator acting on a feedback-coupled metabolic pool, noise produces unexpected dynamic behavior, alternating between regular and irregular time structures. This effect can lead to phase uncoupling of a set of coupled oscillators, the single cells, producing spatiotemporal patterns.

Another example of the impact of stochastic processes in metabolism is presented by Arkin et al. (3) in a detailed stochastic model for the initial decision between two developmental pathways (lysis and lysogeny) by bacteriophage lambda. In this investigation, researchers assumed the chemical kinetics of the operator fluctuations were fast. This assumption allowed the operator states to be treated deterministically using a quasi-steady-state approximation.

The role of noise has also been considered when engineering gene networks. Here, fluctuations were added post-hoc to deterministic rate equations; therefore, the noise strength was an adjustable parameter (33, 35).

In the following sections, I explore ways in which we can make use of biological variability in samples of the same genetic background to construct metabolic networks. These networks are presented in light of their connectivity to the instantaneous pathway structure.

DIFFERENTIAL METABOLOMICS BY SNAPSHOT ANALYSIS

Transgenic or mutant plants provide an excellent means by which to look at changes in metabolic network connectivity through the specific perturbation of a gene of interest using metabolomics.

We investigated potato plants suppressed in the expression of a specific sucrose synthase isoform proposed to be active in phloem metabolism (57). For a high number of sample snapshots of the control plant and the antisense plant, the set of identified and quantified metabolites was systematically searched through for Pearson's correlations. These correlations provide the basis for constructing connectivity networks of metabolites based on the Pearson's correlation coefficient (shown in Figure 2). This coefficient was used to quantify the distance of the biological connectivity of all the measured metabolites and enabled the construction of metabolite distance maps that can be visualized. Using this approach both the wild-type and the antisense plants showed metabolite connectivity networks, which can then be investigated with respect to the altered behavior of their individual constituents (106). Principle component analysis alone was not able

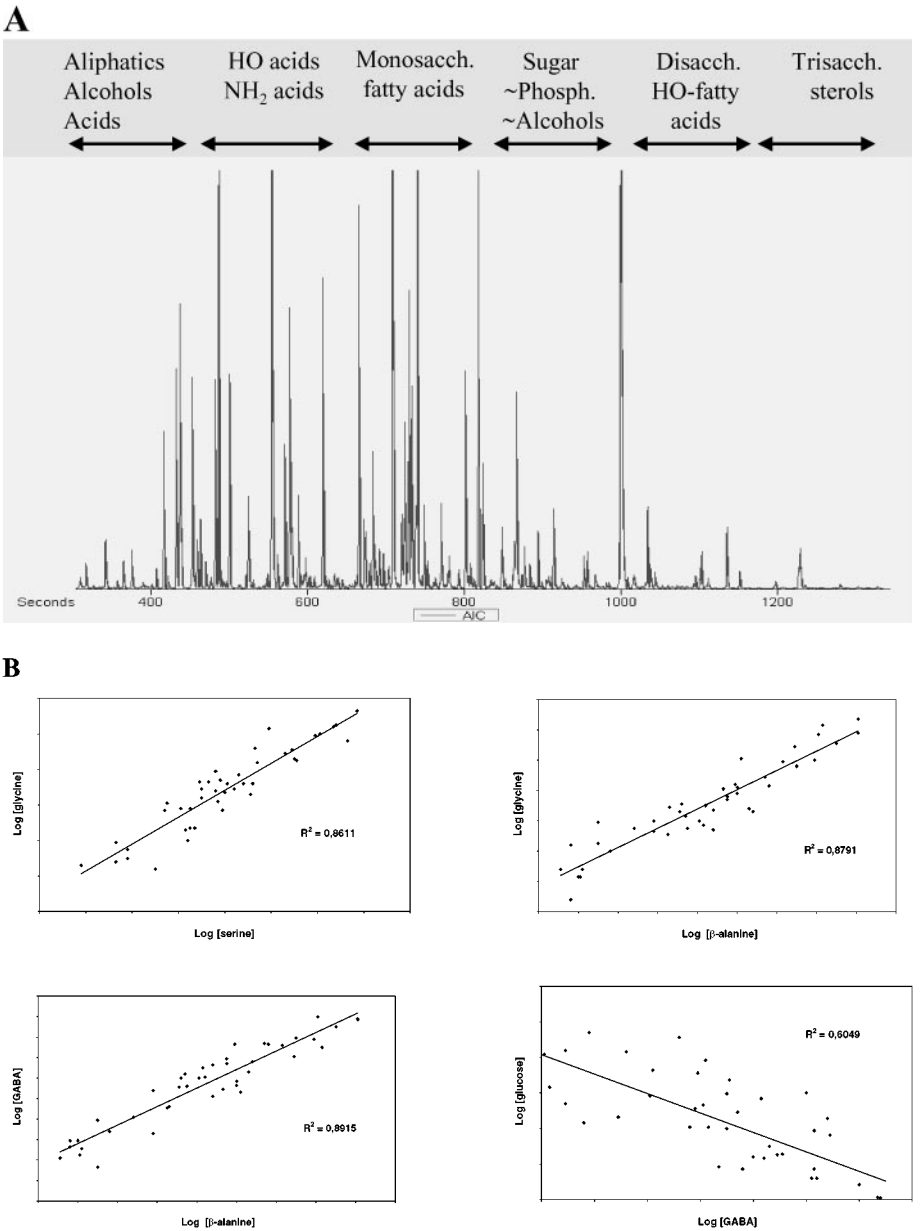

Figure 2 (*A*) Typical GC/TOF chromatogram resolving several compound classes from a crude plant extract. (*B*) Examples of metabolite scatter plots using metabolite data from 44 potato-tuber samples of the same genotype grown and sampled under controlled conditions and at the same developmental stage. (*C*) Connectivity network comprising the subset of metabolites from example *B*. Glucose and mannitol are negatively correlated to the highly connected amino-acid cluster.

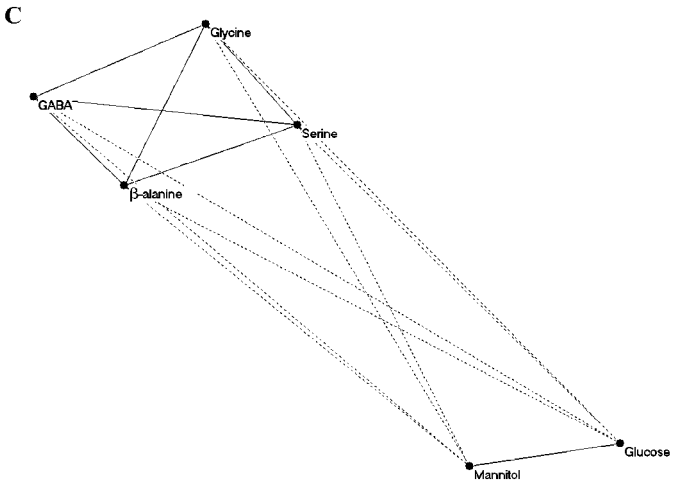

Figure 2 (*Continued*)

to differentiate between the wild-type and antisense plants. However, using the distance maps and connectivity of metabolites, we found alterations in the antisense plants in comparison to the wild-type plants. As expected, sucrose as well as fructose and glucose and their corresponding phosphates showed an alteration concerning the number of correlations with a specified threshold and, more importantly, concerning the distance to other metabolites comparing the wild-type and antisense plants (W. Weckwerth and O. Fiehn, unpublished data). This alteration is consequently an effect of a suppressed sucrose synthase II expression.

These metabolic networks are difficult to understand but contain the inherent information of causal connectivity in the underlying reaction network (4, 76, 103). Although we are not able to provide a complete interpretation of the results yet, owing to fundamental lack of knowledge concerning biochemical network dynamics, as discussed above, it is feasible to define key points in metabolic networks as a response to specified gene alteration.

CAUSAL CONNECTIVITY AND SNAPSHOT CORRELATION NETWORKS

To gain information from differential metabolic network analysis we must connect these experimental data to their underlying regulatory network structures. In the following section, I provide a simple theoretical model for the analysis of metabolite snapshot correlation networks.

Can we assume that each plant leaf tissue sample provides the average metabolite level of millions of metabolically coupled cells? As shown by studies of CAM

metabolism (58, 70; see System Modeling and Stochastic Noise), we can dissect leaf-cell photosynthesis to reveal an assembly of coupled individual oscillators indicating a high level of synchronization between the cells and tissue as a whole (70). In studies with yeast Richard et al. (71) investigated the oscillation of glycolytic metabolites. Based on experimental data and kinetic modeling this system was shown to explain intracellular as well as intercellular network connectivity resulting in the synchronization of billions of yeast cells (9, 17, 18, 109). Here, the postulate of communicators and "mere slaves" was stated defining key substances such as ATP/ADP and NAD/NADH to mediate network fluctuation/oscillation.

Simplifying that plant tissues can be understood as coupled cellular units with defined pathway structures, we investigated exemplarily the actions of a small theoretical network of yeast glycolysis (89) by introducing a fluctuating glucose input. Using numerical solutions of the rate equations and simulating snapshot sampling as described above, we found metabolite scatter plots resulting from propagating metabolite-level fluctuation through a complex network (see Figure 3). According to Vance et al. (102), the fluctuation is attenuated a few steps into the reaction network. Most surprisingly, we observed strong correlations ($r_{xy} = 0.9$) between intermediates that are linearly connected but far from each other (i.e., separated by a high number of reactions) (see Figure 3*B*). This might be understood as the connectivity of the intermediates of a network such as G6P and F6P and the efflux of compounds such as acetaldehyde (ACA) and the connectivity of these compounds via cofactors such as ATP/ADP and NADH/NAD (Figure 3*A*) (also see System Modeling and Stochastic Noise, above).

The mean of the steady-state level in this model is unchanged. This is a prerequisite for the simple assumption that a snapshot correlation network includes information regarding the underlying regulated biochemical network.

Based on the proposed model of intrinsic noise or fluctuation in a biochemical network, a single gene perturbation inevitably leads to a changed correlation matrix of the metabolic snapshot network (for instance, Pearson's correlation). This change enables the comparison of a control and a treated plant on the basis of their metabolite snapshot correlation matrices and, consequently, the resulting network topology (1, 19, 47, 54, 106).

However, interpretation of these metabolic networks is most precarious owing to the paucity of knowledge about the underlying gene and protein expression network. For this reason, we propose an integrated approach to investigate a biological system as a whole.

INTEGRATIVE PROFILING OF BIOLOGICAL SAMPLES

Besides a qualitative description of proteins expressed in an organism, the need for quantitation of protein abundance in response to specific perturbations is a prerequisite in systems biology. Protein quantification is especially important because correlations between RNA levels and protein levels are remarkably low in all studies published to date (29, 44, 82).

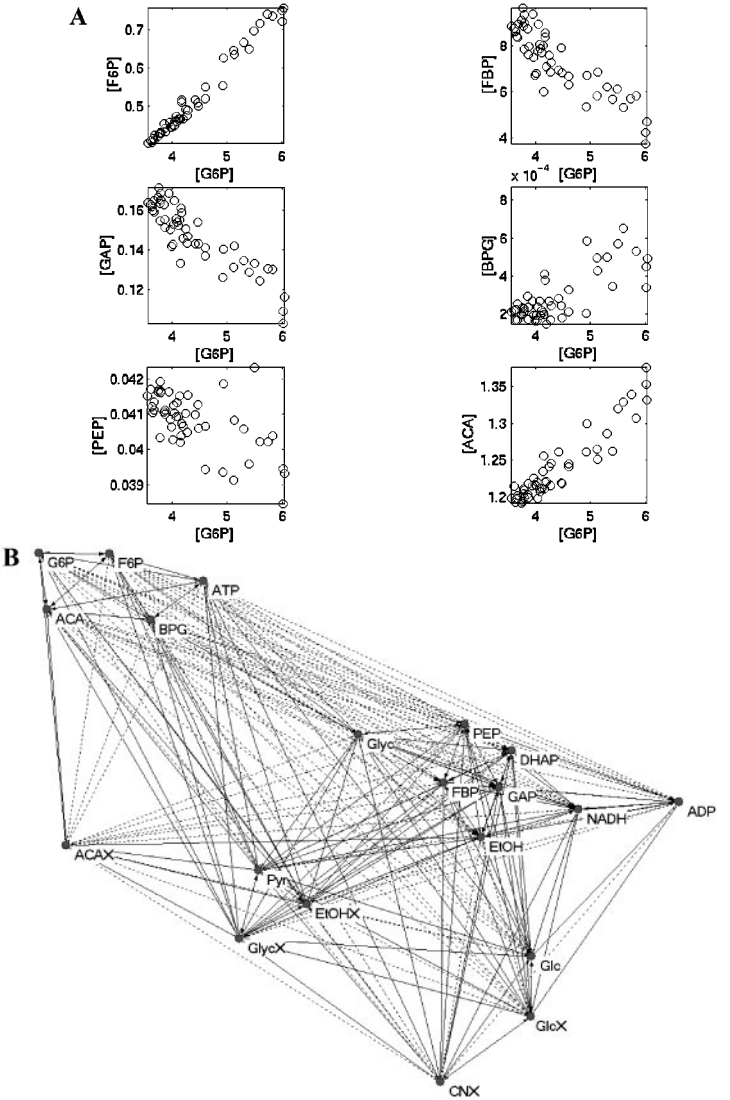

Figure 3 (*A*) Metabolite scatter plots as a result of in silico snapshot sampling in a theoretical pathway network corresponding to the model by Hynne et al. (43). Fluctuation of glucose input is propagated through the network of yeast glycolysis. The rate equations are adapted from Reference 43, and samples are collected independently for analysis of metabolite-metabolite correlations. It is interesting to note that a high correlation of G6P, F6P, and ACA is observed, whereas almost no correlation is found for phosphoenolpyruvate (PEP), which is a precursor of ACA (for details see text). (*B*) The complete Pearson matrix of metabolite correlations is used to construct a snapshot metabolite correlation network. Metabolites are placed according to their correlation coefficient: The higher the coefficient is, the closer the metabolites are, e.g., F6P, G6P, and ACA (for details see text).

In that respect, posttranslational regulation is considered a major regulatory event in metabolism, but techniques to identify and quantify posttranslational modification of proteins on a systems level are still at a preliminary state of development (22, 59, 60, 63, 64, 107).

The idea of correlating transcript RNA and protein levels can be extended to the metabolite level: Which increase or decrease of metabolite levels is connected to which protein expression level and/or posttranslational modification via linear or nonlinear correlation? Answers to questions such as these point the way to hypothetical or known biochemical relationships able to explain flux alterations or increasing metabolite pool size.

Owing to the arguments above and our technical inability to either resolve metabolic compartments accurately or unambiguously dissect the underlying network of reactions, we propose an integrative extraction for metabolites, proteins, and mRNA from one biological sample to reveal correlations inside of complex fluctuating biochemical networks. Figure 4 shows the principle of this idea.

We used ESI-LC/LC/MS/MS to analyze the protein fraction (53, 105). We identified a set of ~300 proteins online via a data-dependent run on an ion trap MS and a subsequent database search (W. Weckwerth, unpublished data).

For instance, using this method, we analyzed two different *Arabidopsis* ecotypes, C24 and Col2, with respect to their metabolite and protein content and

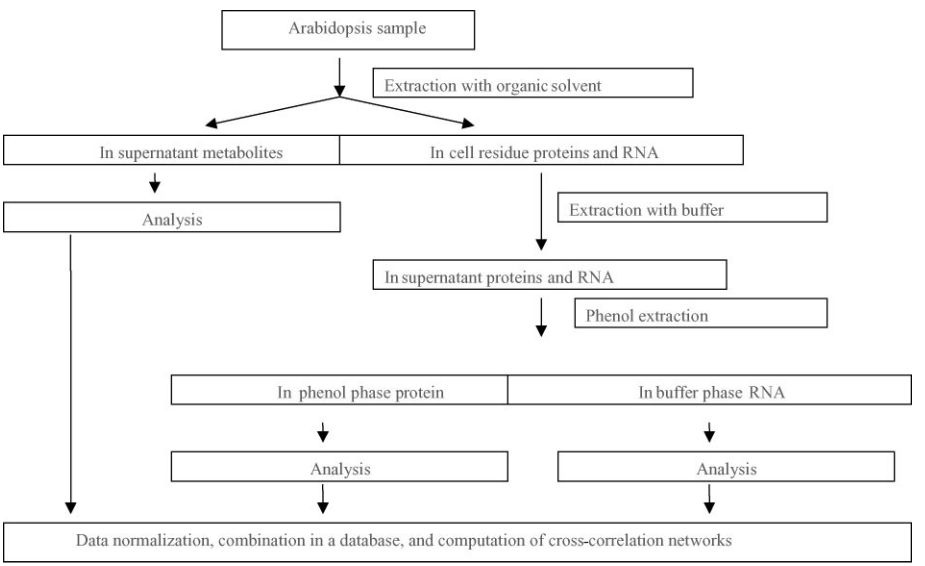

Figure 4 Extraction procedure enabling the cross-correlation analysis of metabolites, proteins, and mRNA from a single sample. Owing to high variances among component levels in samples, dynamic networks can be constructed from a high number of samples (replicates) and analyzed with respect to control sample networks (for details see text).

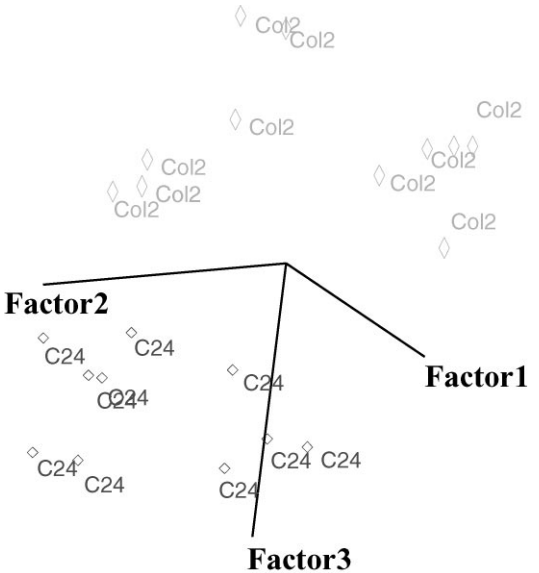

Figure 5 Cross-correlation of metabolites and proteins within leaf tissue. In principle component analysis, two *Arabidopsis* ecotypes, col2 and C24 (each with a set of 10 snapshot samples), can be clearly discriminated.

defined the variance within one set of samples. Owing to a well-defined technical error (∼10%), we assigned the values indicating the variation between protein levels (∼39 %) and metabolite levels (∼ 42%) in approximation to biological variances. Furthermore, these variances were specific for the corresponding genotype and, using principle component analysis, led to clear discrimination between the two ecotypes (Figure 5) (W. Weckwerth, unpublished data).

This principle is not restricted to metabolites, proteins, or mRNA. The systematic description of a biological system also requires the quantitative information of growth stage, environmental parameters, and/or diurnal and circadian rhythmicity. Only when these combined data of a plant system are known with respect to linear or nonlinear co-regulations can we assign functions to genes at a systems level.

CONCLUSION

In this review, two important aspects of comprehensive nontargeted profiling technologies are described: (*a*) the identification of previoulsy unknown components and the investigation of their relation to biochemical processes and (*b*) the quantitation of components. Using this process, we can expect to discover real dynamics of a biological system in response to specific perturbations. Metabolomics is best suited for this systems biology approach owing to its comprehensive information

content concerning dynamic metabolic networks. However, the discrepancy of detected compounds compared with the number of proposed compound structures in plants is disillusioning. High-resolution MS (40) and high-resolution chromatography must be combined to increase the number of detectable metabolites in an unbiased way. A further drawback of metabolomic technology yet to be overcome is the vast number of unknown compound structures. Combinations of NMR and MS can be used for structure elucidation to slowly remedy this shortcoming.

Despite great progress at these levels, it remains to be seen if models can be extended to anything more than approximate predictions of a narrow range of metabolism (50). Rather pessimistic is the finding that small changes in protein expression or flux alteration may affect all correlations in a metabolic network. Because we do not yet know the exact rate equations, in vivo constants, and connectivity, we are not able to calculate these system changes (27). Nevertheless, collecting quantitative data at the metabolite, protein, and mRNA level and correlating this with quantitative descriptions of developmental stage and environment will give a phenomenological description of the whole system. Time-dependent snapshot sampling will reveal directed correlations of metabolic processes.

Using empirical observations, we try to establish models able to describe and predict the behavior of systems. Whether we model the reality or only a shadow of it (67), we form a better understanding of the intricate biochemical processes and their scattering in living systems.

ACKNOWLEDGMENTS

I gratefully acknowledge Megan McKenzie for revising the manuscript. I thank Oliver Fiehn for the fruitful discussions we have shared during the past few years.

The *Annual Review of Plant Biology* is online at http://plant.annualreviews.org

LITERATURE CITED

1. Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature* 406:378–82
2. Aranibar N, Singh BK, Stockton GW, Ott KH. 2001. Automated mode-of-action detection by metabolic profiling. *Biochem. Biophys. Res. Commun.* 286:150–55
3. Arkin A, Ross J, McAdams HH. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics* 149:1633–48
4. Arkin A, Shen PD, Ross J. 1997. A test case of correlation metric construction of a reaction pathway from measurements. *Science* 277:1275–79
5. Bahr U, Pfenninger A, Karas M, Stahl B. 1997. High sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry. *Anal. Chem.* 69:4530–35
6. Beale M, Dupree P, Lilley K, Beynon J, Trick M, et al. 2002. GARNet, the Genomic Arabidopsis Resource Network. *Trends Plant Sci.* 7:145–47
7. Beck F, Blasius B, Luttge U, Neff R, Rascher U. 2001. Stochastic noise interferes coherently with a model biological

clock and produces specific dynamic behaviour. *Proc. R. Soc. London Ser. B* 268: 1307–13
8. Benthin B, Danz H, Hamburger M. 1999. Pressurized liquid extraction of medicinal plants. *J. Chromatogr. A* 837:211–19
9. Bier M, Bakker BM, Westerhoff HV. 2000. How yeast cells synchronize their glycolytic oscillations: a perturbation analytic treatment. *Biophys. J.* 78:1087–93
10. Blanch GP, Caja MM, del Castillo MLR, Santa-Maria G, Herraiz M. 1999. Fractionation of plant extracts by supercritical fluid extraction and direct introduction in capillary gas chromatography using a programmable temperature vaporizer. *J. Chromatogr. Sci.* 37:407–10
11. Bouchereau A, Guenot P, Lather F. 2000. Analysis of amines in plant materials. *J. Chromatogr. B* 747:49–67
12. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, et al. 2001. Growth stage-based phenotypic analysis of arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13:1499–510
13. Brenes M, Garcia A, Garcia P, Garrido A. 2000. Rapid and complete extraction of phenols from olive oil and determination by means of a coulometric electrode array system. *J. Agric. Food Chem.* 48:5178–83
14. Castioni P, Christen P, Veuthey JL. 1995. Supercritical-fluid extraction of compounds from plant-origin. *Analysis* 23: 95–106
15. Choi BK, Hercules DM, Gusev AI. 2001. Effect of liquid chromatography separation of complex matrices on liquid chromatography-tandem mass spectrometry signal suppression. *J. Chromatogr. A* 907:337–42
16. Cooper M, Chapman S, Podlich D, Hammer G. 2002. The GP problem: quantifying gene-to-phenotype relationships. *In Silico Biol.* 2:151–64
17. Dano S, Hynne F, De Monte S, d'Ovidio F, Sorensen PG, Westerhoff H. 2001. Synchronization of glycolytic oscillations in a yeast cell population. *Faraday Discuss.* 120:261–76
18. Dano S, Sorensen PG, Hynne F. 1999. Sustained oscillations in living cells. *Nature* 402:320–22
19. Fell DA, Wagner A. 2000. The small world of metabolism. *Nat. Biotechnol.* 18: 1121–22
20. Ferruzzi MG, Nguyen ML, Sander LC, Rock CL, Schwartz SJ. 2001. Analysis of lycopene geometrical isomers in biological microsamples by liquid chromatography with coulometric array detection. *J. Chromatogr. B* 760:289–99
21. Ferruzzi MG, Sander LC, Rock CL, Schwartz SJ. 1998. Carotenoid determination in biological microsamples using liquid chromatography with a coulometric electrochemical array detector. *Anal. Biochem.* 256:74–81
22. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, et al. 2002. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* 20:301–5
23. Fiehn O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* 48:155–71
24. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. 2000. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18:1157–61
25. Fiehn O, Kopka J, Trethewey RN, Willmitzer L. 2000. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* 72:3573–80
26. Gamiz-Gracia L, de Castro MDL. 2000. Continuous subcritical water extraction of medicinal plant essential oil: comparison with conventional techniques. *Talanta* 51:1179–85
27. Giersch C. 2000. Determining elasticities in situ. In *Technological and Medical Implications of Metabolic Control Analysis*, ed. AJ Cornish-Bowden, ML Cardenas, pp. 2–25. Amsterdam: Kluwer

28. Gobel C, Feussner I, Schmidt A, Scheel D, Sanchez-Serrano J, et al. 2001. Oxylipin profiling reveals the preferential stimulation of the 9-lipoxygenase pathway in elicitor-treated potato cells. *J. Biol. Chem.* 276:6267–73
29. Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19:1720–30
30. Halket JM, Przyborowska A, Stein SE, Mallard WG, Down S, Chalmers RA. 1999. Deconvolution gas chromatography mass spectrometry of urinary organic acids—potential for pattern recognition and automated identification of metabolic disorders. *Rapid Commun. Mass Spectrom.* 13:279–84
31. Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R. 2002. Plant metabolomics: the missing link in functional genomics strategies. *Plant Cell* 14:1437–40
32. Hanisch D, Zien A, Zimmer R, Lengauer T. 2002. Co-clustering of biological networks and gene expression data. *Bioinformatics.* In press
33. Hasty J, Isaacs F, Dolnik M, McMillen D, Collins JJ. 2001. Designer gene networks: towards fundamental cellular control. *Chaos* 11:207–20
34. Hasty J, McMillen D, Isaacs F, Collins JJ. 2001. Computational studies of gene regulatory networks: in numero molecular biology. *Nat. Rev. Genet.* 2:268–79
35. Hasty J, Pradines J, Dolnik M, Collins JJ. 2000. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. USA* 97:2075–80
36. Deleted in proof
37. Deleted in proof
38. Deleted in proof
39. Huber SC, Huber JL. 1996. Role and regulation of sucrose-phosphate synthase in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:431–44
40. Hughey C, Rodgers R, Marshall A. 2002. Resolution of 11000 compositionally distinct components in a single electrospray ionization fourier transform ion cyclotron resonance mass spectrum of crude oil. *Anal. Chem.* 74: 4145–49
41. Huhman DV, Sumner LW. 2002. Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry* 59:347–60
42. Hutt MT, Luttge U. 2002. Nonlinear dynamics as a tool for modelling in plant physiology. *Plant Biol.* 4:281–97
43. Hynne R, Dano S, Sorensen PG. 2001. Full-scale model of glycolysis in Saccharomyces cerevisiae. *Biophys. Chem.* 94:121–63
44. Jansen RC, Nap JP, Mlynarova L. 2002. Errors in genomics and proteomics. *Nat. Biotechnol.* 20:19
45. Jarvis AP, Morgan D. 1997. Isolation of plant products by supercritical-fluid extraction. *Phytochem. Anal.* 8:217–22
46. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411:41–42
47. Jeong H, Tombor B, Albert R, Oltval ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. *Nature* 407:651–54
48. Johansen HN, Glitso V, Knudsen KEB. 1996. Influence of extraction solvent and temperature on the quantitative determination of oligosaccharides from plant materials by high-performance liquid chromatography. *J. Agric. Food Chem.* 44: 1470–74
49. Jungblut P, Thiede B. 1997. Protein identification from 2-DE gels by MALDI mass spectrometry. *Mass Spectrom. Rev.* 16:145–62
50. Kell D, Mendes P. 2000. Snapshots of systems: metabolic control analysis and biotechnology in the postgenomic era. In *Technical and Medical Implications of Metabolic Control Analysis*, ed. AJ Cornish-Bowden, ML Cardenas, pp. 2–25. Amsterdam: Kluwer
51. Kell DB, Darby RM, Draper J. 2001.

Genomic computing. Explanatory analysis of plant expression profiling data using machine learning. *Plant Physiol.* 126: 943–51
52. Kepler TB, Elston TC. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–36
53. Koller AWM, Lange BM, Andon NL, Deciu C, Haynes PA, et al. 2002. Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. USA* 99:11969–74
54. Kose F, Weckwerth W, Linke T, Fiehn O. 2001. Visualizing plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics* 17:1198–208
55. Krull IS, Swartz M. 1999. Analytical method development and validation for the academic researcher. *Anal. Lett.* 32: 1067–80
56. Lilley KS, Razzaq A, Dupree P. 2002. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr. Opin. Chem. Biol.* 6:46–50
57. Loureiro M. 1999. *Role of SNF1 kinase and sucrose synthase in the metabolism of potato*. PhD thesis. Max Planck Inst. Mol. Plant Physiol. Potsdam, Germany. 200 pp.
58. Luttge U. 2000. The tonoplast functioning as the master switch for circadian regulation of crassulacean acid metabolism. *Planta* 211:761–69
59. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, et al. 2002. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. USA* 99:7900–5
60. Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A. 2002. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* 20:261–68
61. Marcotte EM. 2001. The path not taken. *Nat. Biotechnol.* 19:626–27
62. Namiesnik J, Gorecki T. 2000. Sample preparation for chromatographic analysis of plant material. *JPC-J. Planar Chromatogr.-Mod. TLC* 13:404–13
63. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. 1999. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* 96:6591–96
64. Oda Y, Nagasu T, Chait BT. 2001. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* 19:379–82
65. Oliver SG, Winson MK, Kell DB, Baganz F. 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16:373–78
66. Peck SC, Nuhse TS, Hess D, Iglesias A, Meins F, Boller T. 2001. Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* 13:1467–75
67. Platon. 427-347 B.C.E. Plato's Cave Allegory.
68. Prime TA, Sherrier DJ, Mahon P, Packman LC, Dupree P. 2000. A proteomic analysis of organelles from *Arabidopsis thaliana*. *Electrophoresis* 21:3488–99
69. Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, et al. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19:45–50
70. Rascher U, Hutt MT, Siebke K, Osmond B, Beck F, Luttge U. 2001. Spatiotemporal variation of metabolism in a plant circadian rhythm: the biological clock as an assembly of coupled individual oscillators. *Proc. Natl. Acad. Sci. USA* 98: 11801–5
71. Richard P, Teusink B, Hemker MB, Van-Dam K, Westerhoff HV. 1996. Sustained oscillations in free-energy state and hexose phosphates in yeast. *Yeast* 12:731–40
72. Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, et al. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally

modified plant systems. *Plant Cell* 13:11–29
73. Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* 23:131–42
74. Romer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, et al. 2000. Elevation of the provitamin A content of transgenic tomato plants. *Nat. Biotechnol.* 18:666–69
75. Roscher A, Kruger NJ, Ratcliffe RG. 2000. Strategies for metabolic flux analysis in plants using isotope labelling. *J. Biotechnol.* 77:81–102
76. Samoilov M, Arkin A, Ross J. 2001. On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* 11:108–14
77. Sargenti SR, Vichnewski W. 2000. Sonication and liquid chromatography as a rapid technique for extraction and fractionation of plant material. *Phytochem. Anal.* 11:69–73
78. Sauter H, Lauer M, Fritsch H. 1991. Metabolic profiling of plants—a new diagnostic-technique. *Acs Symp. Ser.* 443:288–99
79. Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP, Kieselbach T. 2002. Proteome map of the chloroplast lumen of Arabidopsis thaliana. *J. Biol. Chem.* 277:8354–65
80. Schuster S, Fell DA, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18:326–32
81. Selkov E, Maltsev N, Olsen GJ, Overbeek R, Whitman WB. 1997. A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene* 197:GC11–26
82. Smith RD, Pasa-Tolic L, Lipton MS, Jensen PK, Anderson GA, et al. 2001. Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. *Electrophoresis* 22:1652–68
83. Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, Nishioka T. 2002. Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.* 74:2233–39
84. Somerville C, Dangl L. 2000. Genomics—plant biology in 2010. *Science* 290:2077–78
85. Starmans DAJ, Nijhuis HH. 1996. Extraction of secondary metabolites from plant material: a review. *Trends Food Sci. Technol.* 7:191–97
86. Stein SE. 1999. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* 10:770–81
87. Stein SE, Scott DR. 1994. Optimization and testing of mass-spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* 5:859–66
88. Sterner JL, Johnston MV, Nicol GR, Ridge DP. 2000. Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples. *J. Am. Soc. Mass Spectrom.* 35: 385–91
89. Steuer R, Kurths J, Fiehn O, Weckwerth W. 2003. Observing and interpreting correlations in metabolomic networks. *Bioinformatics.* In press
90. Stitt M, Muller C, Matt P, Gibon Y, Carillo P, et al. 2002. Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot.* 53:959–70
91. Streeter JG, Strimbu CE. 1998. Simultaneous extraction and derivatization of carbohydrates from green plant tissues for analysis by gas-liquid chromatography. *Anal. Biochem.* 259:253–57
92. Szyperski T. 1998. C-13-NMR, MS and metabolic flux balancing in biotechnology research. *Q. Rev. Biophys.* 31:41–106
93. Taybi T, Cushman JC, Borland AM.

2002. Environmental, hormonal and circadian regulation of crassulacean acid metabolism expression. *Funct. Plant Biol.* 29:669–78

94. ter Kuile BH, Westerhoff HV. 2001. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* 500:169–71
95. Thattai M, van Oudenaarden A. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* 98:8614–19
96. Tolstikov VV, Fiehn O. 2002. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* 301:298–307
97. Tong CS, Cheng KC. 1999. Mass spectral search method using the neural network approach. *Chemoms. Intell. Lab. Syst.* 49:135–50
98. Trethewey RN. 2001. Gene discovery via metabolic profiling. *Curr. Opin. Biotechnol.* 12:135–38
99. Trethewey RN, Krotzky AJ, Willmitzer L. 1999. Metabolic profiling: a Rosetta Stone for genomics? *Curr. Opin. Plant Biol.* 2:83–85
100. Tweeddale H, Notley-McRobb L, Ferenci T. 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J. Bacteriol.* 180:5109–16
101. Vaidyanathan S, Rowland JJ, Kell DB, Goodacre R. 2001. Discrimination of aerobic endospore-forming bacteria via electrospray-ionization mass spectrometry of whole cell suspensions. *Anal. Chem.* 73:4134–44
102. Vance W, Arkin A, Ross J. 2002. Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. USA* 99:5816–21
103. Veriotti T, Sacks R. 2001. High-speed GC and GC/time-of-flight MS of lemon and lime oil samples. *Anal. Chem.* 73:4395–402
104. Vilar JMG, Kueh HY, Barkai N, Leibler S. 2002. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA* 99:5988–92
105. Washburn MP, Wolters D, Yates JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19:242–47
106. Weckwerth W, Tolstikov V, Fiehn O. 2001. Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS analysis reveals silent metabolic phenotypes. *Proc. ASMS Conf. Mass Spectrom. Allied Topics, 49th*, pp. 1–2. Chicago: Am. Soc. Mass Spectrom.
107. Weckwerth W, Willmitzer L, Fiehn O. 2000. Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* 14:1677–81
108. Wiechert W. 2002. Modeling and simulation: tools for metabolic engineering. *J. Biotechnol.* 94:37–63
109. Wolf J, Passarge J, Somsen OJG, Snoep JL, Heinrich R, Wasterhoff HV. 2000. Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys. J.* 78:1145–53
110. Wolters DA, Washburn MP, Yates JR. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73:5683–90
111. Wu SH, Ramonell K, Gollub J, Somerville S. 2001. Plant gene expression profiling with DNA microarrays. *Plant Physiol. Biochem.* 39:917–26
112. Zien A, Fluck J, Zimmer R, Lengauer T. 2002. Microarrays: How many do you need? *Proc. RECOMB'02*, pp. 321–30. Washington