ON THE WAY TO UNDERSTAND BIOLOGICAL COMPLEXITY IN PLANTS: S-NUTRITION AS A CASE STUDY FOR SYSTEMS BIOLOGY

HOLGER HESSE* and RAINER HOEFGEN

Max-Planck-Institute fuer Molekulare Pflanzenphysiologie, Am Muehlenberg 1, 14476 Golm, Germany

Abstract: The establishment of technologies for high-throughput DNA sequencing (genomics), gene expression (transcriptomics), metabolite and ion analysis (metabolomics/ionomics) and protein analysis (proteomics) carries with it the challenge of processing and interpreting the accumulating data sets. Publicly accessible databases and newly development and adapted bioinformatic tools are employed to mine this data in order to filter relevant correlations and create models describing physiological states. These data allow the reconstruction of networks of interactions of the various cellular components as enzyme activities and complexes, gene expression, metabolite pools or pathway flux modes. Especially when merging information from transcriptomics, metabolomics and proteomics into consistent models, it will be possible to describe and predict the behaviour of biological systems, for example with respect to endogenous or environmental changes. However, to capture the interactions of network elements requires measurements under a variety of conditions to generate or refine existing models. The ultimate goal of systems biology is to understand the molecular principles governing plant responses and consistently explain plant physiology.

Key words: Sulfur metabolism, Systems biology, Metabolome, Ionome, Transcriptome, Proteome

_____

* Corresponding author, tel: +49 331 5678247, fax: +49 331 56789 8247, e-mail: hesse@mpimp-golm.mpg.de

INTRODUCTION

Plant growth is affected by biotic and abiotic environmental factors, which induce biochemical and physiological responses to engender adaptive processes or ensure plant survival. The employment of high-throughput analytical technologies enables a systems-level understanding of the different functional components of plant cells and of entire plants by predicting their properties from the numerical data that arises from the interaction analyses of many system elements. Systems biology, as a holistic approach, will generate knowledge of cellular dynamics and function, create a detailed model of cell regulation, visualise the connections between individual pathways or metabolic networks, and provide system-level knowledge of the network of metabolic pathways,
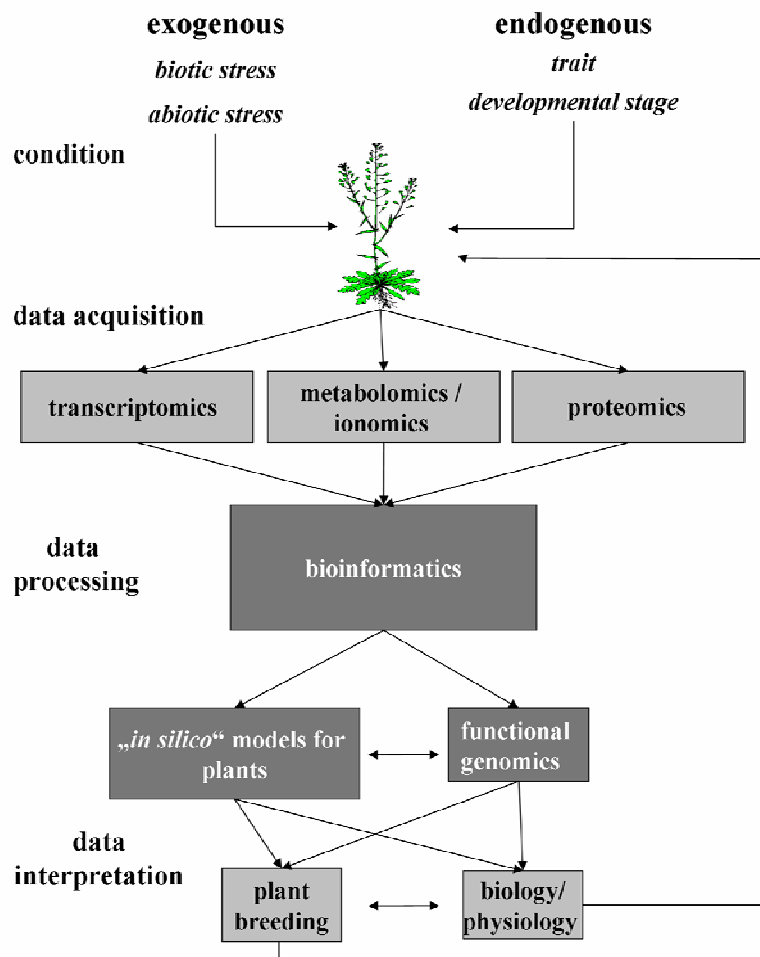

Fig. 1. A sketch of the impact of postgenomic technologies for the analysis of biological systems and plant breeding.

transport functions and signal transduction cascades that are essential for plant development and physiological function [for review 1, 2, for review 3]. Thus far, much of biology has been descriptive and empirical rather than focused on creating quantitative simulation models. To reach systems-level knowledge, mathematical and computational methods for modelling and simulating complex biological systems have to be adopted from other sciences or newly created. The ideal goal would be detailed, accurate and quantitative predictions of the behaviour of biological systems, including predictions of the effects of systems modifications, i.e. simulations. By conceiving the network architecture and thus the interrelation and regulation of its components, it can be envisioned that it will be possible to comprehend the whole system. This will allow valid predictions and ingenious manipulations (Fig. 1).

At the analytical level, systems biology relies on the comprehensive profiling of large numbers of gene expression products. These approaches are commonly referred to as transcriptomics [4, 5], proteomics [6-8], and metabolomics [5, 9-12]. The use of these technologies to obtain comprehensive data sets increased rapidly in recent years, especially with respect to the mechanisms underlying plant growth and plant responses to perturbations. The new high-throughput tools of genomics have provided the potential to systematically analyse perturbed biological systems and monitor their responses.

The challenge of systems-based approaches lies now in extracting information from multivariate experiments and in building models that incorporate all of the data. With the development of computational-based statistical methods, it is now possible to extract the maximum amount of information from experiments involving genome-scale data. In systems biology, bioinformatic tools are not only required to analyze the genomic data but, most importantly, to determine the experimental parameters needed for model building. Testing the derived models *in vivo* with mutants completes the circle. Thus, by combining new tools in genomic biology and bioinformatics, systems biology paves the way to a comprehension of complex biological systems.

In this review a specific example will be employed to illustrate the use of genomic tools to yield a systems biology description. The response of *Arabidopsis thaliana* to sulfur starvation was used as an example model for a systems-level analysis of plant nutrient physiology.

TRANSCRIPTOMICS

Genomics tools such as DNA microarrays (DNA chips) have enabled the simultaneous measurement of gene expression changes in response to an experimental treatment (i.e. system perturbation) or developmental changes (endogenous programmes). For *Arabidopsis thaliana*, there have been reports on more than a dozen studies that used chip technology to describe the transcriptome of exogenously perturbed systems or ontogenetic programmes. Examples include analyses of the circadian rhythm [13], hormone action [14-

16], the stress response [17-20], the cell cycle [21], developmental programs [21-25], responses to pathogens [26, 27] and toxic compounds like cesium [28, 29], plants with altered metabolisms [30], and plants under different nutrient regimes such as alterations in the levels of nitrogen [31-34], phosphate [34], iron [35-37], potassium [34, 38], and sulfur [37, 39-41]. Recent efforts have been made to collect data to establish databases of transcript profiles for Arabidopsis and also for other plant species, and these are stored and publicly available at, for example, the Nottingham Arabidopsis Stock Centre: http://nasc.nott.ac.uk/home.html; EBI: http://www.ebi.ac.uk/home.html and SMD: http://genome-www5.stanford.edu. The speed of data accumulation has shifted the emphasis of research from conducting array experiments to a need for interpretation of the data. The analysis of the expressional behaviour of genes that encode enzymes of known metabolic pathways has distinct advantages that facilitate the interpretation of microarray data (Fig. 2). The function of many of the enzymes and genes of the primary metabolism are known for plants or can be deduced from knowledge of the primary metabolism of other organisms. Array data can thus be associated with specific processes. As several of the genes encoding metabolic enzymes can be assumed to be co-regulated [13],
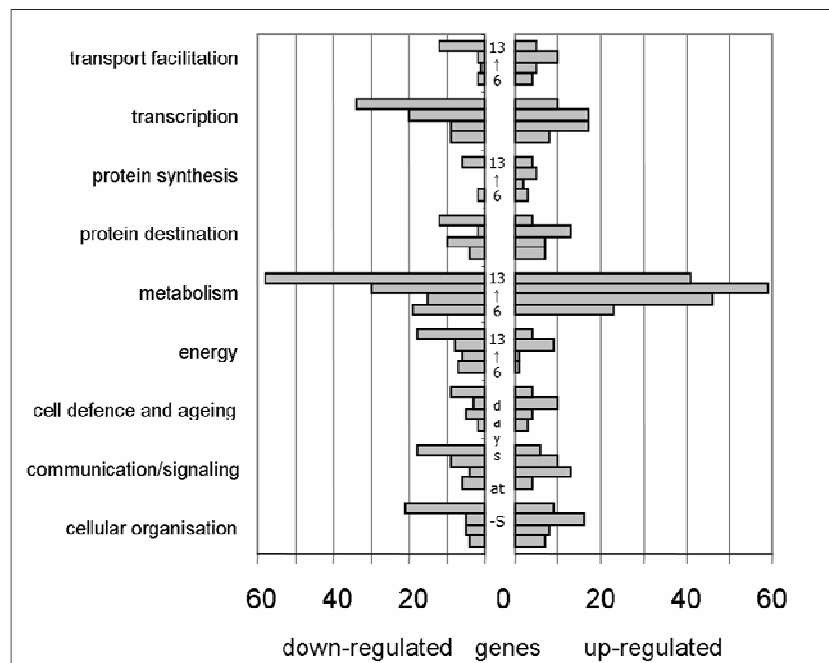

Fig. 2. Assignment to functional categories of the transcriptome response of sulfur-starved *A. thaliana* plants. Within one functional category, the number of induced and repressed genes was determined by array hybridisations. The main assignment to the functional categories is according to the MIPS Arabidopsis data base. The graph indicates the relative response development to sulfur starvation over time in the

respective functional classes. The categories display distinct response developments indicating specific regulatory programmes.

cluster and correlation analyses of expression profiles can thus serve to identify cliques of similarly regulated genes. These patterns can then be used to tentatively assign genes with unknown functions to specific metabolic responses. Further, detailed information on the analysis of macroarrays can be found in Maathius and Amtmann [42].

One example is the study of gene expression during adaptation to changing sulfur availability. Mineral deficiencies (e.g. depleted S) result in significant reductions in crop productivity. Adaptive responses to S deficiency serve to compensate for the lack of this particular nutrient and to specifically increase its uptake capacity [41]. As a signal, sulfate 're-programmes' its own metabolism and that of other pathways to initiate processes that serve to increase sulfate uptake from the soil. DNA microarrays with a cDNA EST collection of 15,442 clones comprising about 7,200 individual genes spotted on nylon filters have been used to identify novel Arabidopsis genes that respond to S-deficient conditions. Consistent with previous reports, genes encoding proteins that are directly involved in sulfate transport, reduction and assimilation are induced. Moreover, novel genes could be identified and sorted into functional categories (MIPS *Arabidopsis thaliana* data base; MATDB). In summary, over several experiments approx. 25% of the genes of the Arabidopsis genome show significant alterations, 60% up and 40% down; about 35% of those genes are of hitherto unknown function. The power of this tool became evident when the transcriptome data was overlaid on known metabolic pathways. Relationships between the sulfur-serine metabolism and tryptophan-glucosinolate-auxin metabolism complex appeared providing a potential link and explanation of one of the major phenotypical symptoms: root growth during S-starvation being triggered through auxin effects [39, 41, 43].

METABOLOMICS/IONOMICS

Metabolomics aims to qualitatively and quantitatively determine the levels of low molecular weight compounds and to provide the metabolome, a complete metabolic profile of the cell. It is thus a manifestation of the endpoint of metabolic and physiological processes. Taken alone, measurements of the metabolome in different physiological states are likely to be more indicative for the purposes of systems biology studies than transcriptome profilings [44, 45]. Metabolomic approaches seek to profile metabolites in a nontargeted way, i.e. to reliably separate and detect as many metabolites as possible in a single analysis. Eventually, combined with information on transcript and protein abundance, this would ideally lead to a nearly complete molecular picture of the state of a particular biological system at a given time.

Understanding a significant part of plant biology requires methods allowing the sensitive detection, quantification and identification of metabolites. A huge

discrepancy can be observed between the number of metabolic genes and the number of known reactions catalyzed by these enzymes in Arabidopsis and the number of metabolites which are actually detected. A large number of metabolites have yet to be identified. The current status of metabolomics has been summarised in several reviews [e.g. 44-47]. The enormous potential of comprehensive biochemical phenotyping for the functional analysis of biological systems is recognised and numerous projects have been initiated. However, major technological limitations still need to be overcome. For instance, the chemical diversity of the metabolome necessitates the use of different analytical techniques to cover the different polarities and molecular sizes found among the metabolites occurring in a cell. A robust and reproducible analysis that provides qualitative and quantitative data and allows high sample throughput is desired. Thus far no complete metabolome has been assembled for any organism, and systematic cataloguing of model organisms is highly important. Systematic identification of the metabolites occurring in a species is particularly relevant for plants, given the wealth of natural products they produce. Thus, similarly to the 'Genome Initiative', a 'Metabolome Initiative' would be advantageous. Sequence data indicates that Arabidopsis expresses a large number of enzymes for which the substrates and products are unknown [48]. Initial efforts have been made by the plant metabolomics community to agree on conventions for data formats and the description of metabolomics experiments [49, 50]. Furthermore, a platform for a mass spectral and retention time index has been established and will be expanded on (MSRI; www.csbdb.mpimp-golm.mpg.de/gmd.html) [51]. However, for general and broad application, especially of GC technologies, the "matrix dependency" of the sample analysis is a real challenge. Compounds in different tissues of a given species or when compared between different species show slight shifts in their peak positions, thus providing problems for automated analysis. Additionally, there are problems with calibration and quantification for certain metabolites.

Profiling schemes for Arabidopsis and other plants have been developed in recent years [9, 52-54]. The experimental tools in use are element analysis via ICP-AES, ion analysis via HPLC or CE, specific HPLC analyses, as for example for amino acids and thiols, and highly random, high throughput approaches mainly based on mass spectrometry combined with various prior separation tools such as GC-MS, GC-TOF, LC-MS [44, 53-55] or others as NMR techniques [56-59]. Furthermore, the coupling of electrospray ionization (ESI) MS with CE [60] and hydrophilic interaction chromatography [61] has been successfully applied to metabolomics problems. Thus, it seems feasible to qualitatively and quantitatively catalogue several hundred to several thousand metabolites in a given tissue. Metabolomics has almost reached a resolution where it can serve to provide a system-oriented characterisation of gene function and cellular and physiological responses, though it still needs further improvement [46].

A target subsection of metabolomics is ionomics. The ionome is the sum of ions and elements which are involved in a broad range of important biological

phenomena, including electrophysiology, signalling, enzymology, osmoregulation, and transport. Efforts to understand the ionome and how it interacts with cellular systems such as the genome, proteome, metabolome, and the environment are essential to fully understand how plants integrate the organic and inorganic metabolism. The ionome includes also metals, metalloids, and nonmetals present in an organism. However, the boundaries between ionomics, metabolomics and even proteomics are blurred. Compounds such as phosphorus, sulphur, or nitrogen would fall within both the ionome and metabolome, and metals such as zinc, copper, manganese, and iron in metalloproteins would fall within the proteome, or metalloproteome. Several analytical techniques are available using flame atomic absorption spectroscopy, inductively coupled plasma spectroscopy (ICP-MS), and inductively coupled plasma atomic emission spectroscopy (ICP-AES) to determine the ionome. More detailed information about the methods and applications are described in 62, 64 and 65.

High-throughput technologies such as gas chromatography-mass spectrometry (GC-MS; [53]) made it possible to study the metabolome of S-deficient plants. From the total number of detectable metabolites, 315 peak-forming derivatised metabolites were detected, among which 110 derivatives of 82 non-redundant chemical compounds were identified [66]. While the transcript response developed gradually during S-starvation, metabolites quickly reached a new, stable homeostatic status, indicating that small alterations at the transcript level early in starvation trigger shifts in the metabolite composition. Thus, time-dependent alterations of transcript and metabolite levels are a result of continuous gene action for maintaining viability [41]. As an example, elements of the sulfur metabolism of the aspartate family and of linked pathways such as glutathione are overlaid to the known biochemical pathways of proteinogenic amino acids (Fig. 3). The levels of metabolites directly dependent on the supply of reduced sulfur, such as cysteine and glutathione, and, unexpectedly, lysine decrease. The levels of others increase: OAS and serine as precursors of cysteine and tryptophan. At the same time, metabolites of the aspartate family remain unaffected for the most part: aspartate, threonine, isoleucine and even methionine. Despite the sulfur starvation, the level of methionine seems to be kept constant as an indispensable part of the SAM methylation cycle. We showed previously that at least two of the enzymes involved in the SAM recycling to methionine are induced under sulfur starvation, adenosylhomocysteinase and SAM synthetase [41, 66].

PROTEOMICS

The third signpost of functional genomics approaches is the systematic analysis of protein composition and activity to determine the proteome. Together with transcriptomics and proteomics, this should provide the tools to describe the functionality of the whole system, the organism. In particular, proteome analysis

is vital, as in the majority of cases, any observed phenotype is a direct result of the action of proteins rather than of the genome sequence. Proteomic methods reveal the proteins translated from the mRNA molecules that are the direct result of gene expression. Currently, a general weakness in proteomics compared with
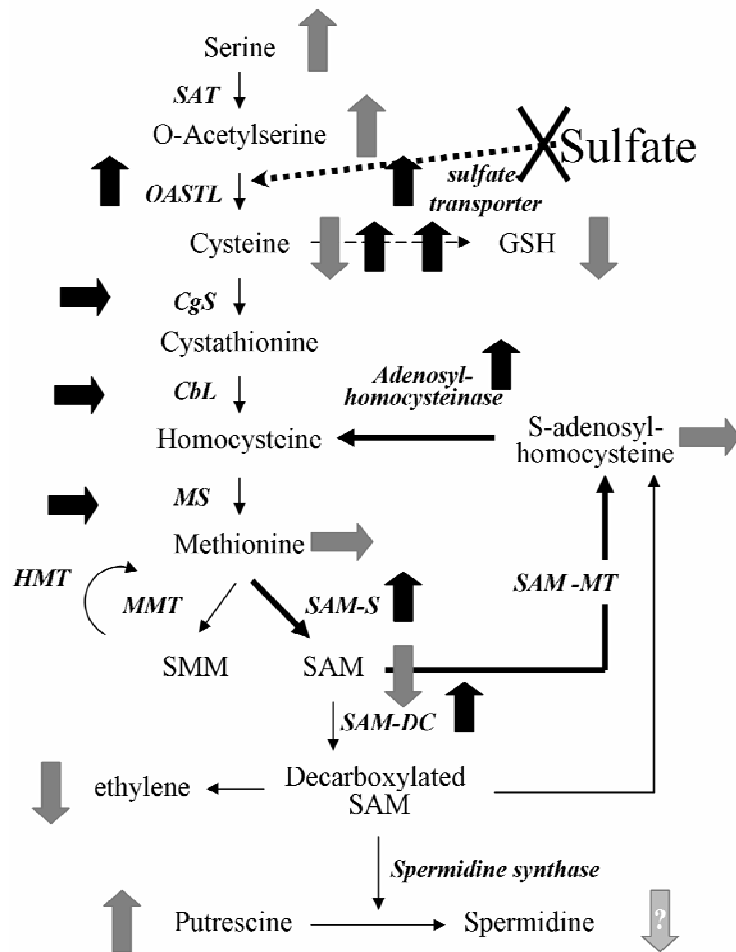

Fig. 3. Metabolic changes overlaid to selected sulfur-induced biosynthetic pathways. The biochemical pathway of cysteine and methionine biosynthesis and its recycling was adapted to the relative changes of transcripts (black arrows) and metabolites (grey arrows) upon sulfur depletion with arrows indicating an increase (up), decrease (down) or no change (horizontal). *OASTL*: O-acetylserine (thiol)lyase; *SAT*: serine acetyltransferase; *CgS*, cystathionine γ-synthase; *CbL*, cystathionine β-lyase; *MS*, methionine synthase; *SAM*, S-adenosyl methionine; *SAM-S*, S-adenosylmethionine synthetase; *SMM*, S-methyl methionine; *HMT*, homocysteine S-methyltransferases; *MMT*, methionine *S*-methyltransferase; *SAM-DC*, SAM-decarboxylase; *SAM-MT*, SAM methyltransferase.

transcriptomic techniques is the low throughput, often limited by the cumbersome nature of current two-dimensional gel electrophoresis (2-D) techniques and low recoveries from 2-D gels and non-representative results (due to the exclusion of several protein types) [67, 68]. Furthermore, despite recent improvements, this technology remains poorly suitable to separate highly hydrophobic, basic or low-abundant proteins [for review 69]. Thus, subcellular membrane proteome, and especially their integral protein moieties, remain poorly accessible. A variety of techniques are used for protein identification, the most common being matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). However, the hybrid electrospray ionization (ESI) method of quadrupole TOF MS, with its increased mass accuracy, is gaining popularity. Recently, in an attempt to increase throughput and to resolve proteins not easily analysed by a 2DE approach, so-called multi-dimensional liquid chromatography (cation exchange followed by reverse-phase column separation) coupled to ESI-MS/MS was successfully introduced [70, 71]. A key challenge of proteomics in comparison to the static sequence-driven problems of genomics is that the protein complement is extremely dynamic in amount, composition, interaction and activity [72]. However, as the speed of experimental proteomic data generation increases, it appears possible to meet this challenge, but this needs to be supported with an appropriate advance in the bioinformatic tools to handle these data.

By contrast to DNA and RNA, there are currently no techniques available to amplify proteins of low abundance (the equivalent of PCR for DNA) or to identify them with very high-dynamic resolution. Moreover, all of the cells that compose a plant have the same genome, but each cell type has a different composition of proteins. In addition to these cell-to-cell variations in proteome expression, we must also consider proteomic variation over time. As a plant cell matures or responds to an environmental perturbation, its expressed mRNAs and proteome respond through adaptive alterations.

Further, the functional consequences of the proteome are dependent on the determination of whether a protein is present, and in what amount and at what activity level. The activity of many proteins is regulated by interactions with other proteins that form complexes or catalyze structural modifications. Comprehensive proteomics, therefore, requires not only a list of the protein functions but also a detailed understanding of protein-protein interactions within a cell and protein modifications that regulate function, such as phosphorylation and glycosylation. Methods for the global measurements of various protein modifications, including mass spectrometry and the use of modification-specific antibodies, are only now being developed. By contrast to genomics, in which standardized tools and standards for data storage and sharing are available, the current methodology for handling large proteomics data sets is less uniform and organized.

To gain a better understanding of the events that occur during sulfate starvation, we used a directed proteomics approach to identify proteins that are modulated

in amount. This attempt was made using two-dimensional gel electrophoresis focusing on proteins that were absent or reduced or increased in level. Unexpectedly, only a few protein changes were observed in S-deficient plants relative to the controls despite the observed massive changes in transcriptome and metabolome and general reduction of total protein content (data not shown). These proteins have yet to be identified by mass spectroscopy, but these results already give a glimpse of the complexity of responses to be expected. Other processes, such as post-translational modifications or protein-protein interactions, have more influence on setting a new cellular homeostasis during adaptation than expected from transcriptome data.

BIOINFORMATICS OR *in silico* ANALYSIS

The coherent modelling of living organisms is the ultimate goal of systems biology. Modelling should greatly rationalise our attempts to understand plants. For example, genes with similar responses at the level of expression over a range of conditions are often clustered together to form functional groups. It can be assumed that these associated genes are under the control of common transcription factors. Recent computer simulations of partial or whole genetic networks have demonstrated the complexity of network behaviours and emergent properties that were not apparent from the examination of a few isolated interactions alone [73]. The modelling of intact higher plants will be especially challenging because of the differential responsiveness of various cell types to a given perturbation. The collection of the comprehensive data needed for modelling might initially be most successful using single-cell microorganisms or higher plant cells grown in defined liquid cultures. The modelling of *Escherichia coli* and yeast is already under way [74, 75], and this might act as a blueprint for the modelling of other cell types and organisms. To model the plant response accurately, a multitude of software programs of the sorts widely used by engineers (e.g. parameter optimization, flux balance analysis, systems analysis, and computer model simulations) need to be adapted for the study of plants. The integration of modelling and experimental work will yield many new insights, with greater complexity, and hopefully have a greater impact on global problems such as the necessary improvements in plant breeding for better yield and pest tolerance.

A step in this direction was to overlay the transcriptome and metabolome results on the known biochemical pathways in a biased, knowledge-driven approach. This allows research to be focused on certain pathways and interactions of pathways (Fig. 3; [41, 76]), a task that can be achieved, for example, using web tools such as AraCyc (http://www.arabidopsis.org/tools/aracyc/) to visualise the results on biochemical pathways.

Bioinformatics tools allow biologists to move beyond cataloguing and simple linear interpretations to increase our understanding of how network components interact [9, 46, 77-81]. Statistical tools are available or being established to

exploit, extract and mine raw data to perform correlation analyses and deduce matrices and networks (Fig. 4). Furthermore, as both datasets rely on ratios between the experimental and control state, it is possible to fuse metabolome and transcriptome databases. Combined analyses have been performed, although for the most part on just a few metabolites to transcriptome data or on pairwise correlations [82, 83]. MetaGeneAlyse (http://metagenealyse.mpimp-golm.mpg.de) [84] and MapMan [85] are analytical tools for analysing the perturbation of a system in transcript and metabolic data with distinct statistical methods.

The reconstruction of a response network is based on similarities in the patterns of element behaviour as a measure of their coherence [86]. The network features and elements will be deduced from this [80]. It is assumed that such a network no longer mirrors biochemical pathways *per se* (though it might in part) but rather describes families of co-behaving (coherent) elements (vertices, nodes) and their correlation via connecting lines (edges) [79]. Typically, biological networks are expected to show inhomogeneous connectivity patterns distinct from a random network [80] with elements of highest connectivity (hubs), while other elements remain lowly connected (Fig. 4). These hubs will be points of high interest for further investigations and often do not appear among the usually selected genes or metabolites with highest ratios for alteration. This will allow functional relationships to be deduced from the network. Furthermore, this approach can be easily applied to other stresses, e.g. nutrient and environmental, challenging the ability of a plant to adapt, or also to investigations of plant developmental programmes.
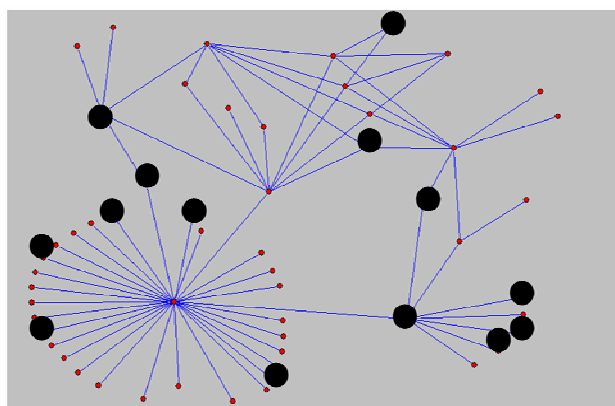

Fig. 4. Correlation connections for a selected part of the transcriptome for sulfur-starved *Arabidopsis* plants. A profile of the strongest connections contains many transcriptional factors (labelled with black circles).

Case studies have been done for sulfur metabolism in Arabidopsis. Transcriptome and metabolome data were used in tandem with bioinformatics tools to describe in a holistic way the biochemical, molecular and physiological

response of a plant to nutrient starvation [87-89]. In the first attempt, it was possible to show that the genes and metabolites involved in glucosinolate metabolism were co-ordinately modulated [87]. Thus, by understanding such gene-to-metabolite networks, it was possible to identify, for example, the gene function of 3 genes encoding sulfotransferases with thus far unknown function; they are now known to be involved in glucosinolate biosynthesis [89].

CONCLUSION

For a deeper, even quantitative understanding of biology, the integration of mRNA, proteomic and metabolomic data within systems biology studies via the use of continuous models is required [90]. This would allow for a substantial improvement in the transcriptional and translational data interpretation and for progress in identifying regulatory networks. The goal is to achieve a better understanding of cellular mechanisms. Transcriptome information cannot be used on its own to predict biological dynamics [91]. High-throughput analytical procedures provide an ever-expanding amount of data and derived information [9, 86]. The amount, the variability of the data, and the incomparability of experimental conditions provides a challenge for the analytical procedures and the data analysis using bioinformatics [92]. Through this, it is to be expected that the body of accumulating information will give rise to a better understanding of biological systems as a whole and will allow us to interpret and to forecast the responses and manifestations of biological systems [80, 81, 93, 94]. The goal, eventually, will be to describe the wiring scheme of metabolic and physiological processes in plants [95, 96] or even cross-species [97]. Through this progress, the responses of plants to genetic manipulations and environmental perturbations will become increasingly predictable. This will make systems biology attractive as a tool for the creation of hypotheses.

REFERENCES

1. Schuster, S., Fell, D.A. and Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. **Nat. Biotechnol.** <u>18</u> (2000) 267-268.
2. Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R. and Hood, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. **Science** <u>292</u> (2001) 929-934.
3. Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A. and Palsson, B.O. Metabolic pathways in the post-genome era. **Trends Biochem. Sci.** <u>28</u> (2003) 250-258.

4. Holtorf, H., Guitton, M.-C. and Reski, R. Plant functional genomics. **Naturwissenschaften** <u>89</u> (2002) 235-249.
5. Oliver, S.G., Winson, M.K., Kell, D.B. and Baganz, F. Systematic functional analysis of the yeast genome. **Trends Biotechnol.** <u>16</u> (1998) 373-378.
6. Blackstock, W.P. and Weir, M.P. Proteomics: quantitative and physical mapping of cellular proteins. **Trends Biotechnol.** <u>17</u> (1999) 121-127.
7. Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossingnol, M., Santoni, V., de Vienne, D. and Zivy, M. Proteomics for genetic and physiological studies in plants. **Electrophoresis** <u>20</u> (1999) 2013-2026.
8. van Wijk, K.J. Update on plant proteomics. Challenges and prospects of plant proteomics. **Plant Physiol.** <u>126</u> (2001) 501-508.
9. Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. Metabolite profiling for plant functional genomics. **Nat. Biotechnol.** <u>18 </u>(2000) 1157-1161.
10. Trethewey, R.N., Krotzky, A.J. and Willmitzer, L. Metabolic profiling: a rosetta stone for genomics? **Curr. Opin. Plant Biol.** <u>2</u> (1999) 83-85.
11. Trethewey, R.N. Gene discovery via metabolic profiling. **Curr. Opin. Biotechnol.** <u>12</u> (2001) 135-138.
12. Trethewey, R.N. Metabolite profiling as an aid to metabolic engineering in plants. **Curr. Opin. Plant Biol.** <u>7</u> (2004) 196-201.
13. Harmer, S.L., Hogenesch, L.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A. and Kay, S.A. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. **Science** <u>290</u> (2000) 2110-2113.
14. Goda, H., Shimada, Y., Asami, T., Fujioka, S. and Yoshida, S. Microarray analysis of brassinosteroid-regulated genes in *Arabidopsis*. **Plant Physiol.** <u>130</u> (2002) 1319-1334.
15. Müssig, C., Fischer, S. and Altmann, T. Brassinosteroid-regulated gene expression. **Plant Physiol.** <u>129</u> (2002) 1241-1251.
16. Rashotte, A.M., Carson, S.D., To, J.P. and Kieber, J.J. Expression profiling of cytokinin action in *Arabidopsis*. **Plant Physiol.** <u>132</u> (2003) 1998-2011.
17. Kreps, J.A., Wu, Y., Chang, H.S., Zhu, T., Wang, X. and Harper, J. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. **Plant Physiol.** <u>130</u> (2002) 2129-2141.
18. Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. **Plant J.** <u>31</u> (2002) 279-292.
19. Hammond J.P., Bennett, M.J., Bowen, H.C., Broadley, M.R., Eastwood, D.C., May, S.T., Rahn, C., Swarup, R., Woolaway, K.E. and White, P.J. Changes in gene expression in Arabidopsis shoots during phosphate

starvation and the potential for developing smart plants. **Plant Physiol.** <u>132</u> (2003) 578-596.

20. Oono, Y., Seki, M., Nanjo, T., Narusaka, M., Fujita, M., Satoh, R., Satou, M., Sakurai, T., Ishida, J., Akiyama, K., Iida, K., Maruyama, K., Satoh, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray. **Plant J.** <u>34</u> (2003) 868-887.

21. Menges, M., Hennig, L., Gruissem, W. and Murray, J.A.H. Cell cycle-regulated gene expression in *Arabidopsis*. **J. Biol. Chem.** <u>277</u> (2002) 41987-42002.

22. Tepperman, J.M., Zhu, T., Chang, H.S., Wang, X. and Quail, P.H. Multiple transcription-factor genes are early targets of phytochrome A signaling. **Proc. Natl. Acad. Sci. USA** <u>98</u> (2001) 9437-9442.

23. Che, P., Gingerich, D.J., Lall, S. and Howell, S.H. Global and hormone-induced gene expression changes during shoot development in *Arabidopsis*. **Plant Cell** <u>14</u> (2002) 2771-2785.

24. Honys, D. and Twell, D. Comparative analysis of the *Arabidopsis* pollen transcriptome. **Plant Physiol.** <u>132</u> (2003) 640-652.

25. Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. and Grossniklaus, U. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. **Genes Dev.** <u>17</u> (2003) 1540-1553.

26. Puthoff, D.P., Nettleton, D., Rodermel, S.R. and Baum, T.J. *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. **Plant J.** <u>33</u> (2003) 911-921.

27. Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G. and Katagiri, F. Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. **Plant Cell** <u>15</u> (2003) 317-330.

28. Hampton, C.R., Bowen, H.C., Broadley, M.R., Hammond, J.P., Mead, A., Payne, K.A., Pritchard, J. and White, P.J. Cesium Toxicity in Arabidopsis. **Plant Physiol.** <u>136</u> (2004) 3824-3837.

29. Sahr, T., Voigt, G., Paretzke, H.G., Schramel, P. and Ernst, D. Caesium-affected gene expression in *Arabidopsis thaliana*. **New Phytologist** <u>165</u> (2005) 747-754.

30. Laule, O., Fürholz, A., Chang, H.S., Zhu, T., Wang, X., Heifetz, P.B., Gruissem, W. and Lange, M. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. **Proc. Natl. Acad. Sci. USA** <u>100</u> (2003) 6866-6871.

31. Wang, R., Guegler, K., LaBrie, S.T. and Crawford, N.M. Genomic analysis of a nutrient response in Arabidopsis reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. **Plant Cell** <u>12</u> (2000) 1491-1509.

32. Wang, Y.-H., Garvin, D.F. and Kochian, L.V. Nitrate-induced genes in tomato roots. Array analysis reveals novel genes that may play a role in nitrogen nutrition. **Plant Physiol.** 127 (2001) 345-359.

33. Colebatch, G., Kloska, S., Trevaskis, B., Freund, S., Altmann, T. and Udvardi, M.K. Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. **Mol. Plant Microbe Interact.** 15 (2002) 411-420.

34. Wang, Y.-H., Garvin, D.F. and Kochian, L.V. Rapid induction of regulatory and transporter genes in response to phosphorus, potassium, and iron deficiencies in tomato roots. Evidence for cross talk and root/rhizosphere-mediated signals. **Plant Physiol.** 130 (2002) 1361-1370.

35. Thimm, O., Essigmann, B., Kloska, S., Altmann, T. and Buckhout, T.J. Response of Arabidopsis to iron deficiency stress as revealed by microarray analysis. **Plant Physiol.** 127 (2001) 1030-1043.

36. Negishi, T., Nakanishi, H., Yazaki, J., Kishimoto, N., Fujii, F., Shimbo, K., Yamamoto, K., Sakata, K., Sasaki, T., Kikuchi, S., Mori, S. and Nishizawa, N.K. cDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. **Plant J.** 30 (2002) 83-94.

37. Wang, R., Okamoto, M., Xing, X. and Crawford, N.M. Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, Iron, and sulfate metabolism. **Plant Physiol.** 132 (2003) 556-567.

38. Maathuis, F.J.M., Filatov, V., Herzyk, P., Krijger, G.C., Axelsen, K.B., Chen, S., Green, B.J., Li, Y., Madagan, K.L., Sánchez-Fernández, R., Forde, B.G., Palmgren, M.G., Rea, P.A., Williams, L.E., Sanders, D. and Amtmann, A. Transcriptome analysis of root transporters reveals participation of multiple gene families in the response to cation stress. **Plant J.** 35 (2003) 675-692.

39. Hirai, Y.M., Fujiwara, T., Awazuhara, M., Kimura, T., Masaaki, N. and Saito, K. Global expression profiling of sulphur-starved Arabidopsis by DNA macroarray reveals the role of O-acetyl-L-serine as a general regulator of gene expression in response to sulphur nutrition. **Plant J.** 33 (2003) 651-663.

40. Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T. and Takahashi, H. Transcriptome profiling of sulphur-responsive genes in Arabidopsis reveals global effects of sulphur nutrition on multiple metabolic pathways. **Plant Physiol.** 132 (2003) 597-605.

41. Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H. and Hoefgen, R. Transcriptome analysis of sulphur depletion in *Arabidopsis thaliana*: interlacing of biosynthetic pathways provides response specificity. **Plant J.** 33 (2003) 633-650,

42. Maathuis, F.J.M. and Amtmann, A. Transcriptional profiling of membrane transporters. in: **Plant Nutritional Genomics**, (Broadley, M. and White, PJ. Eds.), Plant Sciences Division, School of Biosciences, University of Nottingham, UK, 2005, 170-200.
43. Nikiforova, V., Gakière, B., Kempa, S., Adamik, M., Willmitzer, L., Hesse, H. and Hoefgen, R. Towards dissecting nutrient metabolism in plants: a systems biology case study on sulfur metabolism. **J. Exp. Bot.** 55 (2004) 1861-1870.
44. Fiehn, O. Metabolomics – the link between genotypes and phenotypes. **Plant Mol. Biol.** 48 (2002) 155-171.
45. Sumner, L.W., Mendes, P. and Dixon, R.A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. **Phytochemistry** 62 (2003) 817-836.
46. Stitt, M. and Fernie, A.R. From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions. **Curr. Opin. Biotechnol.** 14 (2003) 136-144.
47. Nielsen J. and Oliver S. The next wave in metabolome analysis. **Trends Biotechnol.** 23 (2005) 544-546.
48. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature** 408 (2000) 796-815.
49. Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S. and Sumner, L.W. Potential of metabolomics as a functional genomics tool. **Trends Plant Sci.** 9 (2004) 418-425.
50. Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R., Kopka, J., Lane, G.A., Lange, B.M., Liu, J.R., Mendes, P., Nikolau, B.J., Oliver, S.G., Paton, N.W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L.W., Wang, T., Walsh, S., Wurtele, E.S. and Kell, D.B.A. Proposed framework for the description of plant metabolomics experiments and their results. **Nature Biotechnol.** 22 (2004) 1601-1606.
51. Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R. and Kopka, J. GC–MS libraries for the rapid identification of metabolites in complex biological samples. **FEBS Lett.** 579 (2005) 1332-1337.
52. Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. and Willmitzer, L. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. **Plant J.** 23 (2000) 131-142.
53. Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A.R. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. **Plant Cell** 13 (2001) 11-29.

54. Wagner, C., Sefkow, M. and Kopka, J. Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. **Phytochemistry** 62 (2003) 887-900.

55. Fiehn, O. and Weckwerth, W. Deciphering metabolic networks. **Eur. J. Biochem.** 270 (2003) 579-588.

56. Ward, J.L., Harris, C., Lewis, J. and Beale, M.H. Assessment of ^{1}H-NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. **Phytochemistry** 62 (2003) 949-957.

57. Ott, K.-H., Aranibar, N., Singh, B. and Stockton, G. Metabolomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. **Phytochemistry** 62 (2003) 971-985.

58. Defernez, M. and Colquhoun, I.J. Factors affecting the robustness of metabolite fingerprinting using ^{1}H-NMR spectra. **Phytochemistry** 62 (2003) 1009-1017.

59. Le Gall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J. and Verhoeyen, M.E. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H-NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. **J. Agric. Food Chem.** 51 (2003) 2447-2456.

60. Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M. and Nishioka, T. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. **Anal. Chem.** 74 (2002) 2233-2239.

61. Tolstikov, V.V. and Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. **Anal. Biochem.** 301 (2002) 298-307.

62. Rea, P.A. Ion genomics. **Nature Biotechnol.** 21 (2003) 1149-1151.

63. Lahner, B., Gong, J., Mahmoudian, M., Smith, E.L., Abid, K.B., Rogers, E.E., Guerinot, M.L., Harper, J.F,. Ward, J.M., McIntyre, L., Schroeder, J.I. and Salt, D.E. Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. **Nature Biotechnol.** 21 (2003) 1215-1221.

64. Salt, D.E. Update on Plant Ionomics. **Plant Physiol.** 136 (2004) 2451-2456.

65. Lahner, B. and Salt, D.E. Mapping links between the genome and ionome in plants. in: **Plant Nutritional Genomics**, (Broadley, M. and White, P.J. Eds.), Plant Sciences Division, School of Biosciences, University of Nottingham, UK, 2005, 150-170.

66. Nikiforova, V.J., Kopka, J., Tolstikov, V., Fiehn, O., Hopkins, L., Hawkesford, M.J., Hesse, H. and Hoefgen, R. Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of Arabidopsis plants. **Plant Physiol.** 138 (2005) 304-318.

67. Molloy, M.P., Phadke, N.D., Maddock, J.R. and Andrews, P.C. Two-dimensional electrophoresis and peptide mass fingerprinting of bacterial outer membrane proteins. **Electrophoresis** 22 (2001) 1686-1696.

68. Lester, P. and Hubbard, S. Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics. **Proteomics** <u>2</u> (2002) 1392-1405.

69. Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and Weiss W. The current state of two-dimensional electrophoresis with immobilized pH gradients. **Electrophoresis** <u>21</u> (2000) 1037-1053.

70. Wolters, D.A., Washburn, M.P. and Yates, J.R. 3rd. An automated multi-dimensional protein identification technology for shotgun proteomics. **Anal. Chem.** <u>73</u> (2001) 5683-5690.

71. Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. 3rd. Direct analysis of protein complexes using mass spectrometry. **Nature Biotechnol.** <u>17</u> (1999) 676-682.

72. Shen, Y. and Smith, R. Proteomics based on high-efficiency capillary separations. **Electrophoresis** <u>23</u> (2002) 3106-3124.

73. Jönsson, H., Shapiro, B.E., Meyerowitz, E.M. and Mjolsness, E. (2003) Signaling in multicellular models of plant development. In: **On Growth, Form, and Computers** (Kumar, S., Bentley, P.J., Eds.) Academic Press, London, UK, 156-161.

74. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. and Gilles, E.D. Metabolic network structure determines key aspects of functionality and regulation. **Nature** <u>420</u> (2002) 190-193.

75. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Ziv, B.-J., Gerber, G.K., Hannett, N.M., Harbison, C., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.-B., Volkert, T.L., Fraenkel, E., Gifford, D.K. and Young, R.A. Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. **Science** <u>298</u> (2002) 799-804.

76. Buckhout, T.J. and Thimm, O. Insights into metabolism obtained from microarray analysis. **Curr. Opin. Plant Biol.** <u>6</u> (2003) 288-296.

77. Girke, T., Ozkan, M., Carter, D. and Raikhel, N.V. Towards a modeling infrastructure for studying plant cells. **Plant Physiol.** <u>132</u> (2003) 410-414.

78. Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. Microarray analysis of developing Arabidopsis seeds. **Plant Physiol.** <u>124</u> (2000) 1570-1581.

79. Jasny, B.R. and Ray, L.B. Life and the art of networks. **Science** <u>301</u> (2003) 1863.

80. Bray, D. Molecular networks: The top-down view. **Science** <u>301</u> (2003) 1864-1865.

81. Alon, U. Biological networks: The tinkerer as an engineer. **Science** <u>301</u> (2003) 1866-1867.

82. Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J.V., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and Madden, K.T.

Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. **Nature Biotechnol.** 21 (2003) 150-156.

83. Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L. and Fernie, A.R. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. **EMBO Rep.** 4 (2003) 989-993.

84. Daub, C.O., Kloska, S. and Selbig, J. MetaGeneAlyse: analysis of integrated transcriptional and metabolite data. **Bioinformatics** 19 (2003) 2332-2333.

85. Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Müller, L.A., Rhee, S.Y. and Stitt, M. MapMan: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. **Plant J.** 37 (2004) 914-939.

86. Kitano, H. Systems biology: A brief overview. **Science** 295 (2002) 1662-1664.

87. Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. and Saito, K. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. **Proc. Nat. Acad. Sci. USA** 101 (2004) 10205-10210.

88. Nikiforova, V.J., Daub, C.O., Hesse, H., Willmitzer, L. and Hoefgen, R. Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response. **J. Exp. Bot.** 56 (2005) 1887-1896.

89. Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M., Suzuki, H., Sakurai, N., Shibata, D., Tokuhisa, J., Reichelt, M., Gershenzon, J., Papenbrock, J. and Saito K. Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. **J. Biol. Chem.** 280 (2005) 25590-25595.

90. Gill, R.T., Katsoulakis, E., Schmitt, W., Taroncher-Oldenburg, G., Misra, J. and Stephanopoulos, G. Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis sp.* strain PCC 6803. **J. Bacteriol.** 184 (2002) 3671-3681.

91. Graves, P. and Haystead, T. Molecular biologist's guide to proteomics. **Microbiol. Mol. Biol. Rev.** 66 (2002) 39-63.

92. Katagiri, F. Attacking complex problems with the power of systems biology. **Plant Physiol**. 132 (2003) 417-419.

93. Sweetlove, L.J., Last, R.L. and Fernie, A.R. Predictive metabolic engineering: a goal for systems biology. **Plant Physiol.** 132 (2003) 420-442.

94. Minorsky, P.V. Achieving the *in silico* plant. Systems biology and the future of plant biological research. **Plant Physiol.** 132 (2003) 404-409.

95. Chong, L. and Ray, L.B. Whole-istic biology. **Science** 295 (2002) 1661.

96. Quackenbush, J. Microarrays – guilt by association. **Science** <u>302</u> (2003) 240-241.
97. Stuart, J.M., Segal, E., Koller, D. and Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. **Science** <u>302</u> (2003) 249-255.