BROAD-RANGE METABOLITE ANALYSIS: INTEGRATION INTO GENOMIC PROGRAMS

ALISDAIR R. FERNIE[1*] AND LEE J. SWEETLOVE[2]

[1]Max-Planck-Institut für Molekulare Pflanzenphysiologie,
Am Mühlenberg 1, 14476 Golm, Germany

[2]Department of Plant Science, University of Oxford,
South Parks Rd, Oxford, OX1 4RB, U.K.

E-Mail: [*] fernie@mpimp-golm.mpg.de

Received: 13th April 2004 / Published: 1st October 2004

ABSTRACT

In recent years the focus of experimental biology has shifted from reductionist towards more holistic approaches. This shift has been driven by the development of genetic tools that have allowed the creation of an unprecedented base of genetic diversity and by the development of technologies allowing the rapid determination of the genetic, transcript, protein and metabolite complements of biological systems. Here we will describe experiences with broad-range metabolite analysis of potato and tomato development over the last few years: we will furthermore describe what information can be garnered from these experiments as well as describing recent attempts to analyse systems at the level of more than one molecular entity. Finally, the need for interdisciplinary collaboration and a perspective for this research field will be discussed.

METHODOLOGY

The landmark *Escherichia coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and human genome sequencings facilitated the emergence of systems biology - a science that is currently progressing on both experimental and theoretical fronts. Experimentalists carry out comprehensive and high-throughput analyses of the various molecule entities of the cell, namely, mRNAs, proteins and metabolites (for a review see [1]), whereas theoreticians

Published in „Experimental Standard Conditions of Enzyme Characterizations“, M.G. Hicks & C. Kettner (Eds.), Proceedings of the 1st Int'l Beilstein Symposium on ESCEC, Oct. 5th - 8th 2003, Rüdesheim, Germany

concentrate on the analysis of the regulatory interactions of these molecules and predicting the effects of changing the state of the system (e.g.[2, 3]). Although complementary, these approaches have normally been taken in isolation and for this reason we will discuss them separately. Transcript analysis by hybridization is a relatively mature technology and has been the focus of high quality studies in many biological systems [4-7]. These studies have allowed the determination of differential gene expression under a range of environmental or developmental conditions as well as the identification of the high level of co-ordinated, correlated changes across genes and ultimately to the establishment of gene regulatory networks. Furthermore, of the various levels of analysis it is the only one which can be said to be truly comprehensive: with the current proportional coverage of profiling technologies decreasing following the order mRNA > protein > metabolite. For this reason transcript profiling is currently the experimental approach of choice for systems biology [8, 9]. That said, the emergent technologies of proteomics [10-12] and metabolomics [13, 14] have made rapid progress in recent years and methods such as ICAT (proteins) and mass spectrometry coupled chromatography (metabolites) now allow rapid profiling of these molecules at high analytical precision. Technical aspects of the most commonly used platforms have been reviewed elsewhere (e.g.[15, 16]) so we will not detail them here but rather discuss their importance within integrated approaches.

SYSTEMS APPROACHES

Although a wealth of data can be obtained when using any of the genomic technologies described above, it is clear that transcription, translation, post-translational modifications and the turnover of mRNA, proteins and metabolites do not occur in isolation but are heavily interconnected with one another (for an example of the level of complexity involved see [17]). For this reason it makes sense to move toward integrated approaches wherein transcripts, proteins and metabolites are measured from the same sample. Such approaches have been carried out recently in the microbial and medicinal fields. In one recent study the transcriptome, proteome and protein interactions of 20 systematic perturbations of galactose utilization in the yeast *Saccharomyces cerevisiae* were monitored providing evidence that approximately 15 of 289 detected proteins are regulated post-transcriptionally, and identify explicit physical interactions governing the cellular response to each perturbation [18].

A similar approach combining large-scale perturbation analyses, in combination with computational methodologies, genomic data, cis-regulatory analysis, and molecular embryology was used to define a regulatory network underpinning sea urchin development and revealing how given cells generate their ordained fates in the embryo [8]. These experiments are by nature multidisciplinary and multi-laboratory in that they combine molecular and cellular biologies with protein biochemistry and bioinformatics. While this does not pose a problem in itself, it is vital to note that in order to gain a meaningful insight into interactions between the various molecular entities it is imperative that experiments are performed on samples that are spatially, temporally and micro-environmentally identical. In the following section we intend to describe similar approaches we have taken in plants, albeit on a smaller scale, and to propose a framework by which large-scale systematic approaches can be achieved in plants.

Many recent plant studies have provided data sets comprising transcripts, enzyme activities and metabolites including those focused on various aspects of carbon metabolism interactions including the responses to nitrate [19, 20] and to diurnal changes [21], as well as studies of individual branches of secondary metabolism such as the recent study of the early stages of triterpene saponin biosynthesis [22]. Even though these studies are restricted in their coverage, the potential of this approach is still visible. Whilst in some cases changes in transcripts are accompanied by changes in enzyme activities and shifts in metabolism [20, 22], in other cases they are not (for example the impact of sugars and light and carbon on nitrate reductase: [19]), indicating a major role for post-transcriptional modification in the regulation of the pathways involved. Expansion of such approaches using the tools at hand has vast potential in aiding the understanding of the complex change underlying diverse patterns, such as development and circadian rhythms. We have recently studied potato tuber development and plants exhibiting altered sucrose metabolism using both metabolic and transcript profiling techniques [23]. A comparison of the discriminatory power of metabolite and RNA profiling to distinguish between different potato tuber systems suggests that metabolite profiling has a higher resolution than RNA profiling. Furthermore, when performing comparisons across the molecular entities we established a correlation between 571 of the 26, 616 possible metabolite: transcript pairs. Most of these observations were novel and notably included several strong correlations to nutritionally important compounds, such as vitamins and essential amino acids [23]. We therefore believe that this combinatorial approach is of high potential value in the identification of candidate genes for modifying the metabolite content of biological systems.

INTERPRETATION OF RESULTS HARVESTED FROM SYSTEMS BIOLOGICAL APPROACHES

Although it is probably too early to comment on the effectiveness of reverse genetic versus environmental system perturbation it is important to note that there is a fundamental difference in the data obtained via these methods. Although genetic changes give insight into mechanisms of system robustness and into gene/protein functionality, unless you have a range of gene inhibitions, they do not allow an interrogation of control (since removal of a gene-product merely tells you how well the system copes without it). In contrast, environmental perturbations will result in effects at many genetic loci and it will therefore be very hard to attribute function to any particular gene/protein. However, the ability to perturb the system to a variety of extents may allow the identification of regulatory control points. The interpretation of data sets resulting from the latter example is however clearly far more complicated. Ideally, as many different (types of) system perturbation should be applied (see [18]) and in many instance the use of transgenic lines expressing a range of activities of an enzyme would be preferable to single knock-out mutants. A further advantage of using both environmental and genetic perturbations, is that the identification of common patterns of changes following the different experimental approaches, allows greater surety that the recorded response is a direct result of the desired perturbation, rather than a pleiotropic artefact of the method used to elicit the perturbation.

Inverting this argument, the use of systems perturbations can be utilized to infer common mechanisms by which plant cells respond to different treatments, for example in studying the metabolic complements of variously modified potato tuber systems, Roessner et al. [24] revealed that those expressing a yeast invertase at an apoplastic location could be faithfully phenocopied by feeding glucose (and to a lesser extent fructose) to potato tuber discs. Whilst this example is somewhat trivial (since the resultant conclusions are what would be expected) it highlights the possibilities open for large-scale systematic approaches.

WHAT ELSE DO WE NEED TO ASSAY?

As mentioned above proteomic and metabolomic coverage is far from complete. However, with few exceptions the majority of pathways of plant primary metabolism can be studied in detail with the tools presently available at the protein [25, 26] and intermediary metabolite level [14, 24] in addition to the RNA level.

These studies have allowed description of the mitochondrial proteome and have uncovered new molecular mechanisms involved in mitochondrial defence, as well as cataloging degradation of sensitive protein components including TCA cycle enzymes in response to oxidative stress. Such studies have also highlighted the degree of systemic change following a relatively simple perturbation such as altered hexose supply. Furthermore, metabolic profiling of potato tubers expressing more efficient pathways of sucrose degradation revealed that the tuber contained all the necessary biosynthetic machinery for the de novo biosynthesis of amino acids - knowledge that is a prerequisite for any rational attempt to modify free amino acid content in this tissue. In the majority of instances approaches such as these are adequate to (or even preferable to sifting through data from esoteric pathways) answer the biological question raised. That said, it is clear that for certain approaches, such as unravelling circadian or developmental patterns, truly holistic approaches are needed. For such questions further developments are required in the areas of proteomics and metabolomics.

Whilst new proteomic methodologies (e.g. MudPIT in conjunction with ICAT) dispense with 2D gels and allow high-throughput analysis of thousands of proteins [27, 28] and protein CHIP technology may ultimately allow a simultaneous analysis of the entire proteome [29, 30], increasing coverage of the metabolome presents a greater challenge. Indeed metabolite analysis within systems approaches faces something of a dichotomy, since on the one hand it is important to increase the scope of the metabolites measured (without compromising the accuracy of the measurements) and on the other it is necessary to gain greater understanding of the subcellular levels of metabolites.

Although LC-MS based methodologies have recently been developed allowing the measurement of several important classes of secondary metabolites including alkaloids, flavonoids, glucosinolates, isoprenes, oxylipins, phenylpropanoids, pigments and saponins, and high-throughput spectrophotometric assays have been developed for GC-MS unfriendly primary metabolites such as phosphorylated intermediates and acetyl CoA (reviewed in [16]), it is clear that expansion of the coverage of metabolite profiling methods remains a daunting task. The problem of obtaining information on subcellular information is particularly acute for metabolites which lack the targeting signatures of proteins and turnover too rapidly to allow measurement following aqueous fractionation procedures regularly used in protein analysis (see [25]).

Despite these problems, methods have been developed to obtain the subcellular information on metabolite levels in intact plants - which is ultimately essential to allow accurate modelling of metabolism. The first of these methods, non-aqueous fractionation of lyophilized material involves the separation of small cell portions on an organic density gradient and the use of simultaneous equations to estimate metabolite concentrations with respect to marker enzymes of various organelles in the cell [31]. A second method involves the production of chimeric proteins that differentially fluoresce upon the binding of a certain metabolite. Such proteins can now be created by the fusion of periplasmic binding proteins to green fluorescent proteins and subsequent monitoring of fluorescence energy resonance transfer (FRET), allowing imaging of changes in the concentration of selected metabolites in real time [32]. Although both methods have the potential to provide subcellular spatial information, they both have severe drawbacks - current methods of non-aqueous fractionation only allow the discrimination of three compartments the vacuole, plastid and cytosol, whereas multiple independent chimeric proteins are required for each metabolite measured using the FRET approach - suggesting further research effort will be needed to refine such procedures. However, it is clearly preferable to use the estimated plastid metabolite concentrations than the average cellular concentrations when modelling plastidial metabolism, and the coupling of the non-aqueous fractionation method to GC-MS-based metabolite profiling methods, gave insight into the subcellular distribution of a wider range of metabolites than had been determined to date [33].

In addition to understanding steady-state metabolite levels, it is imperative that high-throughput methods of determining cellular flux between these metabolites are developed for plants. Although frameworks for such experiments have existed for many years in medicinal and microbial sciences (see [34], flux studies in plants are normally carried out using low resolution, highly time consuming protocols based on following the redistribution of radiolabelled substrates. Whilst these are useful in gaining information on the bulk flow through the major pathways of primary metabolism (for an example see [35]) they offer little information about other pathways and focus largely on pathway endpoints. More comprehensive methods that are commonly used in microbial sciences, utilize a combination of stable carbon isotope labelling and NMR or MS-based detection systems, in order to determine positional information of the fed label throughout metabolism.

This method offers the advantage that the labelling pattern of metabolic intermediates can be studied and the position of labelling within the carbon skeleton of end products can allow retrospective evaluation of the metabolic route by which they were formed.

Although used to a limited extent in studies of substrate cycles in primary metabolism [36], perhaps the best example of this technology to date is its use in understanding of the metabolism of storage lipids and proteins in developing *Brassica napus* [37]. In this paper the authors demonstrated the different bioenergetic contributions that amino acids make to the cytosolic and plastidial acetyl CoA pools, highlighting the potential of flux analysis in the understanding of both pathway importance and location. A further example of the importance of both flux measurements and understanding of subcellular location is provided by our recent finding that enzymes of glycolysis are functionally associated with the mitochondria in Arabidopsis [38]. We established this using a combination of proteomic analyses of a highly purified mitochondrial fraction which we then confirmed by enzyme activity assays. The sensitivity of these activities to protease treatments indicated that the glycolytic enzymes are present on the outside of the mitochondrion. Furthermore, when supplied with appropriate cofactors, isolated, intact mitochondria were capable of the metabolism of C-13-glucose to C-13-labelled intermediates of the trichloroacetic acid cycle, suggesting that the complete glycolytic sequence is present and active in this subcellular fraction. Whilst highly novel in plants such associations have been previously reported in *Tetrahymena pyriformis* [39] and more recently in human heart [40]. It is clear that the understanding of such localized pathways is of critical importance for the design of reliable metabolic models.

In addition to the development of efficient flux phenotyping platforms is the continued development of the analysis of higher order modules. Whilst protein-protein interactions which inform assignation of gene function and represent another method of pathway definition have been the focus of much recent research effort [41-43], others such as protein-DNA [44] and protein-lipid [4] interactions have only just begun to be studied at this level. If the analysis of these higher order modules can be carried out in parallel with analysis at the various independent levels under a range of environmental and developmental conditions, then a far greater understanding of factors governing metabolic regulation would be achieved. The hope being that once regulatory properties of the system are known at this level we can build accurate models of metabolism that could be interrogated *in silico* to facilitate the design of rational engineering strategies to generate desired properties in plants.

INTERPRETATION OF DATA SETS

The analysis of molecular entities in a high-throughput manner and on a global scale places a particular premium on the extraction of meaning from extremely large data sets. Currently, the largest data sets tend to be those that contain mRNA transcript abundances and many of the techniques of data analysis that we will describe have been applied principally to transcriptomic data. Nevertheless, it is worth pointing out that the problems and solutions are the same, whether one is dealing with mRNA-transcripts, proteins or metabolites. Microarrays are now available that can be plausibly described as genomic (the Affymetrix ATH1 genechip reports on some 24,000 *Arabidopsis thaliana* genes) and their use generates data sets that consist of tens of thousands of data-points. With such a high data density, it is difficult to display the entire data set in a way that allows meaningful interpretation. Instead, methods must be used to filter the data to retrieve only that data which satisfies criteria relevant to the experimental query. These criteria can be manifold, but generally the most informative are those based on either the degree of change of the data value between two conditions or on similarity of change of data value.

Many microarray-based experiments are a simple pairwise comparision between two conditions, generally a control and a genetic variant or a different treatment (environmental condition, pathogen attack etc.) and the aim is to identify genes that are differentially expressed between the two conditions. These genes can be identified by filtering the data according to a simple heuristic rule such as an expression cut off. Typically, this only considers genes to be differentially expressed if there is a change in relative mRNA transcript abundance above a defined threshold; typically 1.5-2.0 fold (see for example [6, 24]). While this method allows a rapid filtering of the data, it will introduce a high number of false negatives since many biologically-relevant changes may occur at a level that is below this arbitrary threshold. The need to identify such changes has led to the application of more sophisticated statistical methods to the analysis of array data [46]. These include: the nonparametric t-test, the Wilcoxon rank sum test and the ideal discriminator method [47]. Each of these methods performs differently, with some being more conservative than others. Such statistical approaches greatly reduce the number of false negatives and impart some much needed rigour to the analysis of differential gene expression on microarrays.

While differential expression of genes (or indeed changes in protein and metabolite abundances) provides the primary level of information as to the changes that accompany a biological event, there is much more to be gained from such data sets than just establishing what goes up and what goes down. Often, we are not so much interested in which genes are altered in expression as we are in the *relationships* between those changes. In particular, statistical techniques can be used to group data together on the basis of common patterns of changes. Such groupings are particularly useful in the context of deciphering gene function, since it is likely that elements that share the same function, or participate in the same process, are coordinately regulated [48]. Co-responses of elements of unknown- and known-function allow novel functional associations to be proposed [49]. This type of analysis has been used to propose novel functions of genes [6], their protein products [49] and also to discriminate coordinately regulated groups of metabolites [14] and is one of the principal ways in which system network structure can be established. Two methods have been used to group data points: hierarchical clustering analysis (HCA) [50] and principal components analysis (PCA) [51].

The former considers data objects as points in n-dimensional space or as *n*-dimensional vectors (where *n* is the number of samples for comparison) and measures the distance (or similarity) between these objects in *n*-dimensional space. This distance matrix can then be clustered using standard clustering algorithms and the results presented as a dendrogram [52]. Recently, more sophisticated clustering algorithms have been developed that automatically calculate the optimal distribution of data objects over clusters and overcome problems related to robustness and the establishment of optimal linear ordering of the cluster [53]. The complementary method of PCA also establishes n-dimensional vectors and focuses on the vector that gives the greatest separation between samples, the so-called principal component. The results are usually displayed as a two-dimensional plot with the first principal component on the x-axis and the second principal component on the y-axis. A refinement of this approach is to use a supervised projection method such as discriminant function analysis (DFA) [54] which exploits user-defined information (such as which samples are replicates of one another) to determine within- and between-group variation. This information is then used in combination with principal components to define discriminant functions that separate the groups.

COMPUTATIONAL BIOLOGY

The ultimate aim of systems biology is the construction of a complete mathematical description of the system. The idea is to exploit the increasing experimental knowledge of the behaviour of the system in terms of different molecular entities, to devise a set of rules that accurately model the behaviour of the system. There are clear advantages to such a mathematical approach. First, such a model represents a way of storing and describing our current understanding of a system. Second, mathematical models allow *in silico* interrogations of the behaviour of the system. Such *in silico* experiments have the advantage of both speed and power over conventional "wet" experimentation. Speed, in that any number of parameters can be varied and the effect of these variations on the system behaviour or output can be calculated virtually instantaneously. The power of the approach is derived from the fact that the effects of such changes are calculated in the context of the entire system, the behaviour of which is too complex to be understood intuitively. The realization of this aim of constructing a single model that describes an entire system is still some way off. However, models are being devised that describe discrete parts of the system and can be thought of as modules. It is likely that these modules can be joined together to form a higher order model that represents a more complete coverage of the system.

Most of the mathematical models that have been devised, concentrate on describing the control of flux through a metabolic pathway in terms of the amounts of the enzymes present and tend to ignore other levels of the regulatory hierarchy (e.g. regulatory-gene expression). The reason for this concentration on enzymes is that there are well established mathematical frameworks, such as metabolic control analysis (MCA), that describe the control of metabolic pathways in terms of enzymes and metabolites. Such frameworks are an essential starting point for mathematical models. Indeed, most of the models of plant metabolism to date have used MCA as the basic conceptual tool (e.g. [55, 56]). These models generally use kinetic parameters of enzymes to derive the control structure. An alternative approach is that of metabolic flux analysis (MFA) which can be used to interpret *in vivo* metabolic flux data and derive the metabolic network and its control structure [57]. Both approaches suffer from the fact that the experimental inputs required are generally not known in their entirety for a given system. In the case of models based on kinetic parameters of enzymes, a 'mix and match' approach is often taken where kinetic constants of enzymes from different organisms are combined.

Although core pathways such as glycolysis are highly conserved, there are nevertheless key differences in their regulation between different organisms and such an approach is unlikely to adequately describe the subtleties of different control structures in different organisms. In the case of metabolic flux analyses, the situation is slightly better as there are a number of theoretical approaches such as metabolic-flux balancing [58] and optimization approaches [59] that can be used to derive unknown fluxes. Recently, a new approach to the modelling of metabolic pathways has emerged that is based upon the principle of stoichiometric analysis [60].

The basis of this approach is to define elementary flux modes - non-decomposable sub-networks that account for every possible flux within the network. Elementary flux modes are non-decomposable in the sense that each mode contains a minimal set of enzymes such that if only enzymes belonging to this set are operating, then complete inhibition of one of these enzymes would lead to a complete cessation of pathway flux [2]. This approach allows one to mathematically define and describe all metabolic routes that are both stoichiometrically and thermodynamically feasible and is an extremely useful tool for the definition of network structure [61]. When applied on a sufficiently large scale the approach allows cellular behaviour to be reconstructed from network topology and thus represents a genuine systems analysis [62]. Although stoichiometric analysis concentrates on enzymes and metabolic pathways, the related approach of gene circuit analysis [63] deals with regulatory modules that operate at the genetic level. An integration of these two approaches would potentially describe the majority of regulatory features that are known to occur in a metabolic network and bring us ever closer to a truly holistic description of a biological system. However, to achieve such an integration will require much concerted effort between experimentalists and theoreticians.

CONCLUSION

The availability and continued improvement of high-throughput analytical techniques has brought about a distinct shift in the way biologists are approaching the solution of metabolic control networks. Instead of the reductionist enzyme-by-enzyme approach, we are instead attempting to take a more system-wide approach, in which a comprehensive analysis of a broad range of molecular responses across the system are made and are used as the basis of a holistic understanding of the system driven by computational methods.

While this new approach has the potential to provide a quantum leap in our ability to understand the control of metabolic networks, fulfilment of that potential will ultimately depend on a number of key developments. First, several analytical challenges need to be met to ensure that analysis of all types of molecular entity is as comprehensive as possible. These challenges include a broadening of the coverage afforded by protein and metabolite profiling technologies with the latter representing a considerable obstacle. In addition, the analysis of these molecules needs to be refined to include single cells and different sub-cellular compartments. Second, the field of computational biology needs to continue its progress in developing increasingly sophisticated tools and approaches to the extraction of biological meaning from genomic data sets. It is clear that the true success of the systems biology approach will be determined by the extent to which theoretical and computational biologists can work together with experimental biologists. The latter need to adopt new experimental strategies that are specifically tailored to the systems approaches, while the former need to ensure that computational tools are made available to the wider community such that the systems approach becomes a standard part of the experimental arsenal and its use extends beyond a handful of test cases.

REFERENCES

[1] Oliver, S.G. (2002) Functional genomics: lessons from yeast. *Phil. Trans. R. Soc. Lond. B* **357**: 17-23.

[2] Schuster, S., Dandekar, T., Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17**: 53-60.

[3] Brazhnik, P., de la Fuente, A., Mendes, P. (2002) Gene networks: how to put the function in genomics. *Trends Biotechnol.* **20**: 467-472.

[4] Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson, J., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D., Brown, P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**: 83-87.

[5] Furlong, E.E.M., Andersen, E.C., Null, B., White, K.P., Scott, M.P. (2001) Patterns of gene expression during Drosophila mesoderm development. *Science* **293**: 1629-1633.

[6] Ramonell, K.M., Somerville, S. (2002) The genomics parade of defense responses: to infinity and beyond. *Curr. Opin. Plant Biol.* **5**: 291-294.

[7] Ruuska, S.A., Girke, T., Benning, C., Ohlrogge, J.B. (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* **14**: 1191-1206.

[8] Davidson, E.H. et al. (2002) A genomic regulatory network for development. *Science* **295**: 1669-1678.

[9] Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929-934.

[10] Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M. (1996) Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* **93**: 14440-14445.

[11] Mann, M., Hendrickson, R., Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**: 437-473.

[12] Lilley, K.S., Razzaq, A., Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr. Opin. Chem. Biol.* **6**: 46-50.

[13] Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**: 1157-1161.

[14] Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., Fernie, A.R. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**: 11-29.

[15] Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromova, P., Yu, J., Palsdottir, H., Magnusson, N., Ornoft, T.F. (2000) Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**: 2-16.

[16] Fernie, A.R. (2003) Metabolome characterization in plant system analysis. *Funct. Plant Biol.* **30**: 1-10.

[17] Zhao, J., Williams, C.C., Last, R.L. (1998) Induction of Arabidopsis tryptophan pathway enzymes and camalexin by amino acid starvation, oxidative stress and an abiotic elicitor. *Plant Cell* **10**: 359-370.

[18] Ideker, T., Galitski, T., Hood, L. (2001) A new approach to decoding life: Systems biology. *Annu. Rev. Genom. Hum. Genet.* **2**: 343-372.

[19] Muller, C., Scheible, W.R., Stitt, M., Krapp, A. (2001) Influence of malate and 2-oxoglutarate on the NIA transcript level and nitrate reductase activity in tobacco leaves. *Plant Cell Environ.* **24**: 191-203.

[20] Scheible, W.R., Krapp, A., Stitt, M. (2000) Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant Cell Environ.* **23**: 1155-1167.

[21] Masclaux-Daubresse, C., Valadier, M.H., Carrayol, E., Reisdorf-Cren, M., Hirel, B. (2002) Diurnal changes in the expression of glutamate dehydrogenase and nitrate reductase are involved in the C/N balance of tobacco source leaves. *Plant Cell Environ.* **25**: 1451-1462.

[22 Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T., Dixon, R.A. (2002) A genomics approach to the early stages if triterpene biosynthesis in *Medicago trunculata*. *Plant J.* **32**: 1033-1048.

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

[24] Roessner, U., Willmitzer, L., Fernie, A.R. (2001) High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. *Plant Physiol.* **127**: 749-764.

[25] Sweetlove, L.J., Heazlewood, J.L., Herald, V., Holtzapffel, R., Day, D.A., Leaver, C.J., Millar, A.H. (2002) The impact of oxidative stress on Arabidopsis mitochondria. *Plant J.* **32**: 891-904.

[26] Millar, A.H., Sweetlove, L.J., Giege, P., Leaver, C.J. (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol.* **127**: 1711-1727.

[27] Whitelegge, J.P. (2002) Plant proteomics: BLASTing out of a MudPIT. *Proc. Natl. Acad. Sci. USA* **18**: 11564-11566.

[28] McDonald, W.H., Ohi, R., Miyamoto, D.T., Mitchison, T.J., Yates, J.R. (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Intl. J. Mass Spec.* **219**: 245-251.

[29] Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., Snyder, M. (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genetics* **26**:283-289.

[30] Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science* **293**: 2101-2105.

[31] Gerhardt, R., Stitt, M., Heldt, H.W. (1983) Subcellular metabolite determination in spinach leaves through non-aqueous fractionation. *Physiol. Chem.* **364**: 1130-1131.

[32] Fehr, M., Frommer, W.B., Lalonde, S. (2002) Visualisation of maltose uptake in living yeast cells by fluorescent nanosensors. *Proc. Natl. Acad. Sci. USA* **99**: 9846-9851.

[33] Farre, E.M., Tiessen, A., Roessner, U., Geigenberger, P., Trethewey, R.N., Willmitzer, L. (2001) Analysis of the compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids, and sugar alcohols in potato tubers using a non-aqueous fractionation method. *Plant Physiol.* **127**: 685-700.

[34] Szyperski, T. (1998) 13C-NMR, MS and metabolic flux balancing in biotechnology research. *Quarterly Rev. Biophys.* **31**: 41-106.

[35] Lytovchenko, A., Sweetlove, L., Pauly, M., Fernie, A.R. (2002) The influence of cytosolic phosphoglucomutase on photosynthetic carbohydrate metabolism. *Planta* **215**: 1013-1021.

[36] Dieuaide-Noubhani, M., Raffard, G., Canioni, P., Pradet, A., Raymond, P. (1995) Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from 13C- or 14C- labeled glucose. *J. biol. Chem.* **270**: 13147-13159.

[37] Schwender, J., Ohlrogge, J.B. (2002) Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiol.* **130**: 347-361.

[38] Giege, P., Heazlewood, J.L., Roessner-Tunali, U., Millar, A.H., Fernie, A.R., Leaver, C.J., Sweetlove, L.J. (2003) Enzymes of glycolysis are functionally associated with the mitochondrion in Arabidopsis cells. *Plant Cell* **15**: 2140-2151.

[39] Srere, P.A. (1987) Complexes of sequential metabolic enzymes. *Annu. Rev. Biochem.* **56**: 89-104.

[40] Taylor, S.W., Fahy, E., Zhang, B., Glenn, G.M., Warnock, D.E., Wiley, S., Murphy, A.N., Gaucher, S.P., Capaldi, R.A., Gibson, B.W., Ghosh, S.S. (2003) Characterisation of the human heart mitochondrial proteome. *Nat. Biotechnol.* **21**: 281-286.

[41] Panicot, M., Minguet, E.G., Ferrando, A., Alcazar, R., Blazquez, M.A., Carbonell, J., Altabella, T., Koncz, C., Tiburcio, A.F. (2002) A polyamine metabolon involving aminopropyl transferase complexes in Arabidopsis. *Plant Cell* **14**: 2539-2551.

[42] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**: 399-403.

[43] Gavin, A.C. et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**: 141-147.

[44] Smith, J., Khin, E.K., Zaitseva, E.M., Freebern, W., Dzekunova, I., Gardner, K. (2002) Genome wide analysis of protein/DNA interactions. *FASEB J.* **16**: A1104.

[45] Howell, N.K., Herman, H., Li-Chan, E.C.Y. (2001) Elucidation of protein-lipid interactions in a lysozyme-corn oil system by Fourier transform Raman spectroscopy. *J. Agric. Food Chem.* **49**: 1529-1533.

[46] Pan, W. (2002) A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**: 546-554.

[47] Troyanskaya, O.G., Garber, M.E., Brown, P.O., Botstein, D., Altman, R.B. (2002) Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics* **18**: 1454-1461.

[48] Walhout, A.J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K.C., Schetter, A.J., Morton, D.G., Kemphues, K.J., Reinke, V., Kim, S.K., Piano, F., Vidal, M. (2002) Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**: 1952-1958.

[49] Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**: 45-50.

[50] Brazma, A., Vilo, J. (2000) Gene expression data analysis. *FEBS Lett.* **480**: 17-24.

[51] Jolliffe, I.T. (1986) *Principal Components Analysis*. Springer-Verlag, New York.

[52] Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**: 14863-14868.

[53] Lukashin, A.V., Fuchs, R. (2001) Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters. *Bioinformatics* **17**: 405-414.

[54] Windig,W., Haverkamp, J., Kistemaker, P.G. (1983) Interpretation of sets of pyrolysis mass spectra by discriminant-analysis and graphical rotation. *Analyt. Chem.* **55**: 81-88.

[55] Pettersson, G. Ryde-Pettersson, U. (1988) A mathematical model of the Calvin photosynthesis cycle. *Eur. J. Biochem.* **175**: 661-672.

[56] Thomas, S., Mooney, P.J., Burrell, M.M., Fell, D.A. (1997) Metabolic Control Analysis of glycolysis in tuber tissue of potato (*Solanum tuberosum*): explanation for the low control coefficient of phosphofructokinase over respiratory flux. *Biochem. J.* **322**: 119-127.

[57] Wiechert, W., Mollney, M., Isermann, N., Wurzel, M., de Graaf, A.A. (1999) Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioengng.* **66**: 69-85.

[58] Bonarius, H., Schmid, G., Tramper, J. (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* **15**: 308-314.

[59] Torres, N., Voit, E., Gonzalez-Alcon, C. (1996) Optimization of nonlinear biotechnological processes with linear programming: application to citric acid production by *Aspergillus niger*. *Biotechnol Bioengng.* **49**: 247-258.

[60] Cornish-Bowden, A., Cardenas, L. (2002) Systems biology: Metabolic balance sheets. *Nature* **420**: 129-130.

[61] Schuster, S., Fell, D.A., Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**: 326-332.

[62] Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**: 190-193.

[63] de la Fuente, A., Brazhnik, P., Mendes, P. (2002) Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genet.* **18**: 395-398.