JOURNAL OF BIOSCIENCE AND BIOENGINEERING
Vol. 100, No. 4, 347–354. 2005
DOI: 10.1263/jbb.100.347

REVIEW

Plant Metabolomics: Potential for Practical Operation

Eiichiro Fukusaki[1]* and Akio Kobayashi[1]

Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan[1]

Received 7 March 2005/Accepted 11 June 2005

In the postgenomic era, metabolomics is expected to be the newest useful omics science for functional genomics. However, in plant science, the present metabolomics technology cannot be considered a universal tool to perfectly elucidate perturbations imposed on sample plants although this is desired by plant physiologists. Despite it being an immature technology, metabolomics has already been used as a powerful tool for precise phenotyping, particularly for industrial application. Metabolomics is the best technology for the analysis of large mutant or transgenic libraries of model experimental plants, such as Arabidopsis, rice, etc. Here, we review the applications and technical problems of metabolomics. We also suggest the potential of metabolomics for plant postgenomic science.

[**Key words:** metabolomics, metabolome, functional genomics, genomics, transcriptomics, proteomics, mass spectrometry, informatics, chemometrics]

Metabolomics represents the exhaustive profiling of metabolites contained in organism. Proteomics and transcriptomics are both considered to be a flow of media concerning genetic information. In contrast, metabolomics should be thought as being concerned with phenotype (Fig. 1).

Recently, it has been proven that slight changes in the metabolome can be explained by perturbations imposed on plants. Perturbations include environmental change, physical stress, abiotic stress, nutritional stress, mutation, and transgenic events. Metabolomics is expected to be more useful if used in conjunction with other omics sciences such as transcriptomics or proteomics (1, 2). Current metabolomics technology cannnot be considered a universal tool to perfectly elucidate perturbations imposed on sample plants although this is desired by plant physiologists. Despite it being an of such immature technology, metabolomics has already been used as a powerful tool for precise phenotyping, particularly for industrial applications. Metabolomics is the best technology for the analysis of large mutant or transgenic libraries of model experimental plants, such as Arabidopsis, rice, etc. In fact, venture business companies for plant biotechnology are using metabolomics technology to drive the large-scale exhaustive screening of a T-DNA tagging transgenic line of Arabidopsis to determine the functions of genes with functions that had not been previously elucidated. Their final goal is to determine the relationship between useful features and their corresponding genes. Such relationships would be useful for generating commercially available transgenic plants. Such companies are rushing to submit patents that claim the identification of useful genes although some black box areas still remain. Meanwhile, it should be realized that metabolomics does not require genome information. For the most useful commercial plants, such as wheat, barley, maize, soy bean, potato, their genomes have not yet been sequenced. Metabolomics can be applied to such commercial plants without genome information. That is one of the most important advantages of metabolomics compared with transcriptomics and proteomics. However, metabolomics is a complicated interdisciplinary research field that requires bioscience, analytical chemistry, organic chemistry, chemometrics, and informatics knowledge (Fig. 2). Metaolomics analysis requires the following steps: plant cultivation, sampling, extraction, derivatization,

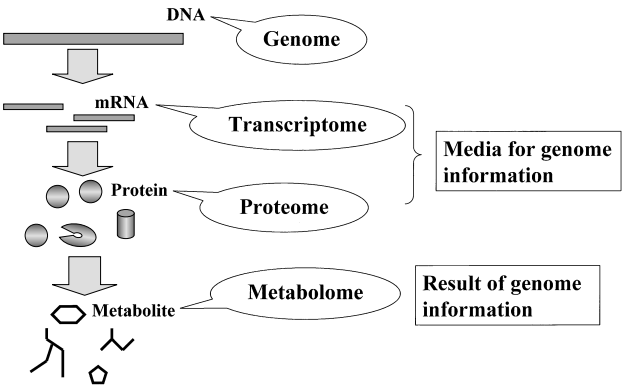

FIG. 1. Metabolomics in functional genomomics.

* Corresponding author. e-mail: fukusaki@bio.eng.osaka-u.ac.jp
phone/fax: +81-(0)6-6879-7424

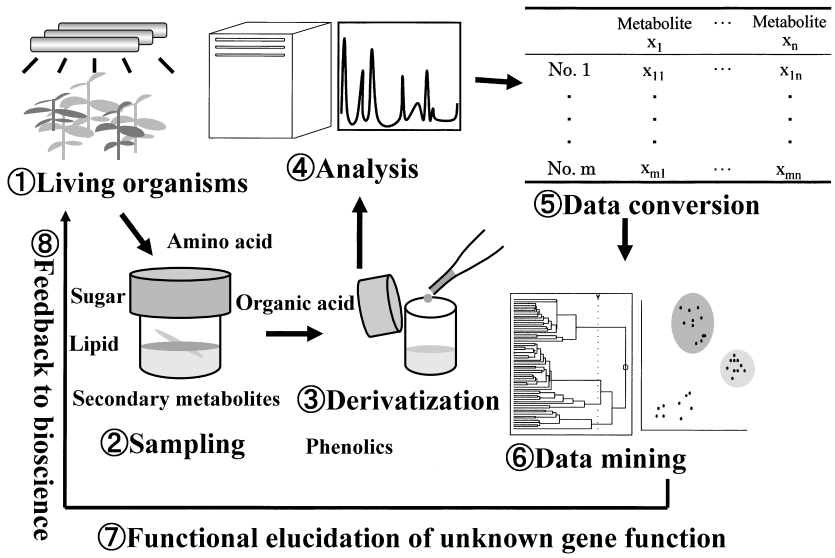

FIG. 2. General scheme of metabolomics.

separation and quantification, data matrix conversion, data mining, and bioscience feedback. Each step can give rise to experimental errors. At present a standard method has not been established and metabolomics is only applicable for on-demand restricted research subjects. This is the main reason why metabolomics is not understood or familiar. Here, we present the technical problems associated with metabolomics. We also suggest the potential of metabolomics for plant post-genomic science and try to explain metabolomics from the viewpoint of the plant scientist.

CLASSIFICATION OF METABOLOMICS

The final goal of metabolomics is an exhaustive profiling of all metabolites contained in target organisms. However, it is almost impossible to perform perfect profiling of at least ten thousand metabolites due to lack of appropriate technology (3). This means that current metabolomics studies are thought of as feasibility studies to indicate the potential of the actual metabolomics. At the present time, the conventional classification of metabolomics that was proposed by Fiehn has been accepted world wide (4) (Table 1).

TECHNICAL ELEMENTS OF METABOLOMICS

Metabolomics consists of several complicated technical elements described in Fig. 2 with each step possibly giving rise to an experimental error. To establish a robust system, close collaboration among researchers in the fields concerned with analytical chemistry, organic chemistry, chemometrics, informatics, and bioscience is required. Specific technical problems in each step are described in the following sections.

Plant cultivation Independent variation among samples is one of the most important problems for metabolomics. It is generally more difficult to maintain uniformity in plant cultivation than when cultivating microorganisms or breeding animals. Even if an artificial plant growth chamber is used to maintain the temperature, light, and humidity, perfect control is almost impossible due to slight variations dependent on the position in the chamber. Soil is the most common matrix for plant cultivation. Precise water control in the case of using soil is generally difficult due to lotvariance and operational error. Water-deficiency stress can often occur and such water stress would affect the metabolome. To minimize condition variance for plant cultivation, large-scale plantation using a large-volume growth chamber is preferred (5, 6). However, small-volume growth chambers are routinely used due to running costs. Great care and know-how are required to minimize variance. The simplest solution is periodical rotation of pot position in a chamber. A soil-less culture system is also useful (7). We employ a new soil-less cultivation system using a ceramictube through which water is directly supplied to plant roots for high resolution metabolomics. The ceramic culture system enables us to maintain exact control of water supply and nutrition. The reproducibility of metabolomics data has been enhanced in the case of using the ceramic culture system.

Sampling Sampling is one of the most important steps to which careful attention should be paid to reduce experimental error. Careless sampling causes significant experimental variance, which might sometimes surpass biological

Table 1. Classification of metabolomics

Classification	Definition
Target compound analysis	Quantification of specific metabolites
Metabolite profiling	Quantitative or qualitative determination of a group of related compounds or of specific metabolic pathways
Metabolomics	Qualitative and quantitative analysis of all metabolites
Metabolic fingerprinting	Sample classification by rapid, global analysis

variance. To maintain variance at a minimum, not only the growth stage but also the exact time of sampling should be controlled. In addition, the area and the amount sampled should also be considered. As expected, great care should be paid to post-harvest treatment. The analytical method and instruments should be decided dependent on the characteristics of target metabolites, including the number of metabolites being examined, and their respective quantities. The optimum preparation protocol should be developed depending on a case-by-case basis. In the case of target analysis, the preparation of appropriate internal standards is extremely important and a suitable purification scheme is also important. Metabolic profiling should cover a wide range of metabolites. Each metabolite should be considered according to its characteristics in the following categories: hydrophylic, hydrophobic, small molecule, large molecule, charged, uncharged and combinations of these. In the case of metabolic profiling, it is rather difficult to choose an appropriate internal standard for normalization because many metabolites must be targeted as analytes at the same time. The efficiency of extraction and fractionation should be normalized without internal standards through repetition of preliminary experiments. Among the many steps, the extraction procedure is the most important one. Homogenous crushing of plant materials is required to maintain the extraction efficiency. A ball mill is a more suitable apparatus for this purpose than a mixer because plant materials contain a very rigid tissue matrix. We use a ball mill with zirconium balls. Target analysis is mainly carried out in the field of plant secondary metabolism. Metabolic profiling now mainly focuses on hydrophilic and small molecules and is primarily used in the field of central primary metabolism. Gas chromatography mass spectrometry (GC-MS) and capillary electrophoresis mass spectrometry (CE-MS) are both important technologies for the analysis of hydrophilic small molecules. A research group in the Max-Planck Institute has developed a useful system for the metabolic profiling of hydrophilic small metabolites derived from the Arabidopsis leaf (8, 9). A precise protocol is available via the internet (Fiehn, O.: Metabolomic analysis: protocol for plant leaf metabolite profiling. http://www.mpimp-golm.mpg.de/fiehn/blatt-protokoll-e.html). The effect of sample preparation methods on metabolite analysis was well studied using an *Escherichia coli* system (10).

Derivatization and pretreatment Derivatization of target metabolites may be required depending on the analytical equipment used. Only volatile compounds are applicable for GC-MS analysis. Most hydrophilic metabolites should be derivatized by sililation or other methods. High-performance liquid chromatography (HPLC) also requires derivatization in the case of UV or fluorescence detection. Specificity and efficiency are both important factors and they should be validated in their terms of reproducibilities. Derivatizing conditions including the category of reagent and reaction conditions should be well examined. In addition, the stability of yielded derivatives should also be evaluated. Various methods of derivatization have been well reviewed (11). Utilization of a stable isotope is very important for comparative quantification because mass spectrometry is used most often for metabolomics research. Mass spectrometry is generally useful both for quantification and qualification. However, in the case of contamination of the ionization room, the ionization efficiency of the target molecule is markedly reduced (12, 13). Such deficiency is called ionization suppression. The main reason for ionization suppression is the coelution of contamination with the target due to a defect in the separation step. Ionization suppression most likely occurs in all mass spectrometries including GC-MS, LC-MS, CE-MS. Optimum time separation in chromatography prior to mass spectrometry is apparently the best solution. However, the optimum time separation of tens of metabolites would be almost impossible in practice. Consequently, stable isotope dilution-based comparative quantification is thought to be the most convenient practical solution. The principle of the method is that isotopomers of target metabolites are used as internal standards to normalize analysis variation, particularly for ionization. Isotopes can be introduced by post-harvest labeling or by *in vivo* isotope enrichment. Metabolites are extracted from one specific sample named 'the test sample'. In a similar manner, metabolites are extracted from a control sample in which all metabolites are labeled with the isotope. The test sample is then mixed with the control sample. The mixture is subjected to LC-MS, CE-MS or GC-MS. On the chromatogram or electropherogram, target metabolites and their corresponding isotopomers are coeluted coincidentally. The comparative ratio of each target metabolite is estimated by the peak ratio corresponding to each target and its isotopomer. This principle is used in the proteomics research technology, Isotope coded affinity tags (ICAT) (14). Post-sampling stable isotope labeling would also be applicable in metabolomics research although D-labeling may have some difficulties (15, 16). *In vivo* stable isotope enrichment is a promising method for stable isotope dilution. The metabolic profiling of sulfur metabolites using ^{34}S has been reported (17). ^{13}C and ^{15}N stable isotope labeling techniques are available for use in some cases (16, 18). In the future, a combination of time-course sampling and stable isotope dilution comparative quantification will be one of the de facto standard methods in dynamic metabolomics.

Separation and quantification Separation is one of the most important unit operations in metabolomics. Separation methods usually include chromatography or electrophoresis coupled to mass spectrometry. UV detection and electrochemical detection are also used for quantification. Each specific separation method is well reviewed in detail in the literature (19). Technical problems in separation processes are focused on in this paper. Metabolome data should be classified in to two independent categories that include resolution and quantification. Separation strongly affects both resolution and quantification. However, it might be costly and inconvenient to determine the best specification for both resolution and quantification. In an on-demand practical system, either resolution or quantification should be chosen as the first priority. It is important to image the matterthat is to be elucidated by metabolomics.

Resolution can be thought of as a practical index by which each separation system is evaluated in terms of the possible number of metabolites that can be separated. Resolution directly depends on the peak capacity of each system.

Among conventional separation systems, capillary electrophoresis (CE) is superior in terms of resolution. The second best is gas chromatography (GC). Liquid chromatography is worst in terms of resolution although its adaptational capacity is the widest. Complete separation of all target metabolites is desired to maintain a high degree of quantification. However, several metabolites would usually coelute or comigrate in each separation system. Liquid chromatography should be used frequently due to its wide adaptational capacity although currently capillary electrophoresis and gas chromatography tend to be used as separation tools in metabolomics due to their high resolution. Liquid chromatography would cover almost all metabolites. However, several important metabolites would be coeluted in liquid chromatography because samples derived from living organisms contain many complicated metabolites. Coelution might cause ionization suppression, which is the main reason for quantification being compromised. In addition, the solvent gradient systems that are conventionally used in liquid chromatography might affect the reproducibility of metabolite retention times, although retention time is one of the most important factors when chromatography results are subjected to a data mining procedure. The above-described difficulties are the main reason why researchers tend to hesitate to use liquid chromatography as a conventional method for metabolomics.

We will now focus on the examples of high-efficiency HPLC separations made possible by monolithic silica columns composed of network type silica skeletons (20, 21). Micro-HPLC systems with a monolithic silica capillary column possess the following advantages: (i) small consumption of stationary and mobile phases, (ii) high detection sensitivity for a certain amount of samples, (iii) high-speed separation with a low pressure drop, and (iv) the possible use of a long column of 1–2 m that can provide around 100,000–200,000 theoretical plates. The disadvantages are (i) the smaller sample capacities of monolithic silica columns compared with particle-packed columns, (ii) the necessity of skill and knowledge to operate a capillary HPLC system to obtain high separation efficiency, and (iii) insufficient supply of good columns and instruments for capillary HPLC. We employed a molithic column HPLC system to accomplish the perfect separation of naturally derived polyprenol regioisomers (22). It has been almost impossible to separate polyprenol regiomers by means of conventional separation modes. In addition, a monolithic capillary. HPLC system was proven to be useful for plant metabolic profiling (23). A sophisticated two-dimensional monolithic capillary HPLC system has been developed to provide more than 10-fold higher resolution than conventional HPLC systems (24, 25). Refinement of the system is required for user-friendly operation. Supercritical fluid chromatography (SFC) is one of the promising technologies for separation of hydrophobic metabolites. In particular, it would be useful for the separation of hydrophobic polymers that are very difficult to analyze by conventional HPLC systems. The separation power of SFC has been proven through the analysis of polyprenols (26, 27). A mutidimensional detection system is a powerful tool for the analysis of complicated elution patterns, which include mass spectrometry, photo diode arrays, GC by GC and so on. A sophisticated deconvolution system for GC-MS has been developed in which coeluted metabolites can be separated and identified by mass spectrometry (28). Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), which is the newest mass spectrometry technology, is a powerful tool for identifying biomarkers. Infusion FT-ICR-MS analysis without preseparation by chromatography has been achieved for exhaustive metabolic profiling (29).

Quantitative performance is also one of the most important factors for metabolomics in practice. This depends on the dynamic range of linearity in each analytical system. Surveying the best option in terms of a detection system is required. The influence of the existence of several types of contamination should also be considered. Mass spectrometry has a serious drawback, known as ionization suppression, which might diminish quantitative reproducibility as described above. Perfect time separation by high-resolution chromatography would be a unique solution when using mass spectrometry. An infusion operation using FT-ICR-MS can be regarded as one convenient solution in which resolution has priority over quantitation. Normalization of ionization suppression would be essential for infusion analysis. The reproducibility of UV detection and electrochemical detection might be less affected by coelution despite their low sensitivities compared with mass spectrometry. Consequently, classical detection systems are still regarded as important in the case when quantitation has priority over resolution. Spectrometry is also useful in some cases. The conventional spectrometries that are available include Fourier transform nuclear magnetic resonance (FT-NMR), Fourier transform infrared spectroscopy (FT-IR), and Near field infrared spectroscopy (NIR). Especially, FT-NMR was recently used in plant metabolomics (30–32). In FT-NMR analysis, complicated separation steps such as chromatography or electrophoresis are not essential. In addition, the dynamic range of FT-NMR detection is rather wide. FT-NMR tends to be used in metabolic fingerprinting because of its superior specificity. NMR analyses require a relatively large amount of sample, and the analysis results might be affected by contamination, which is a serious problem in practice. Recently, a combination of several chromatographies and FT-NMR has been developed for precise metabolite profiling or target analysis (33, 34). Diffusion-ordered NMR spectroscopy (DOSY), which is one of the newest methodologies, is a possible tool for the profiling of complicated mixtures of samples without separation (35). The method involves two dimensional NMR application using pulse magnetic field gradient technologies. In detail, a self-diffusion constant (D) is characteristic of DOSY for discrimination depending on the molecular weight of each metabolite measured. Thus, DOSY can afford almost similar information to that obtained by a combination of size-exclusion chromatography and FT-NMR. FT-IR is a well-known conventional spectrometry for structure elucidation of organic compounds. FT-IR can also be used in metabolomics, especially for metabolic fingerprinting, in addition to FT-NMR. FT-IR is useful for the quantification of compounds that have specific functional groups although it is not optimal for complicated metabolite profiling due to the lack of a separation procedure.

FT-IR has been proven to be useful for the profiling of complicated mixtures derived from industrial raw materials or food. FT-IR would be useful in the field of large-scale exhaustive profiling as an easy and high-throughput method. Actually, FT-IR has been used for the metabolic fingerprinting of tomato and Arabidopsis (36, 37). Recently, a combination of microscopy and FT-IR has been proven to be a powerful tool for histochemical metabolic profiling (38).

Data conversion Multivariate analyses tend to be used in metabolomics research because rather complicated linear and nonlinear relationships must be elucidated. Analogue raw chromatographic data must be converted to digital matrix data. Raw data obtained from chromatography or electrophoresis can be converted into a matrix data table that can be subjected to multivariate analysis via peak identification and integration. In this case, the target metabolite should be used as an independent variable and peak area should be used as a dependent variable. In the case of using spectroscopy data , some specific know-how is required to prepare an appropriate matrix table that can be subjected to data mining, because spectroscopic instruments (FT-IR, FT-NMR) generally provide analytical data in a specific data format. All matrix tabl data sets must be adjusted into exactly the same form, which is essential for multivariate analysis. In the case that data points are different among the data to be analyzed, data point adjustment must be conducted. Spectral data might often include some disturbance according to each specificity and environmental perturbation. Such disturbance should be corrected for by appropriate data preprocessing. Data obtained by GC-MS or LC-MS should also be corrected by appropriate preprocessing. In fact, appropriate preprocessing is a prerequisite for data mining. Preprocessing involves (i) noise reduction, (ii) baseline correction, (iii) resolution enhancement, and (iv) normalization. Several tactics, such as smoozing, spectral difference, differentiation, baseline correction, peak separation, mean center,and scaling are often performed in common preprocessing. Recently, a sophisticated algorithm concerning the preprocessing of GC-MS chromatogram data was reported (39).

DATA MINING BY MULTIVARIATE ANALYSIS

In metabolomics, multivariate analysis with an appropriate algorithm should be performed depending on data structure and mining intention. Multivariate analysis methodology used includes multiple regression, discriminant analysis, principal component analysis, hierarchical cluster analysis, factor analysis, canonical analysis. Among these methods, exploratory analysis tends to be used most often in plant metabolomics. The mission of the analysis is mainly for the characterization of data structure and preliminary mining of significant tendencies included in the data. Exploratory data analysis should be performed before conducting further analysis, such as multiple regression or classification. Biologists tend to hesitate to use multivariate analysis due to some difficulties concerning basic linear algebra and statistics, although multivariate analysis is one of the most important tools in metabolomics. Principal component analysis (PCA), hierarchical cluster analysis (HCA) and self-organizing mapping (SOM) are the most important multivariate analysis methods that are used often. Their features and operation methods are described in the following sections.

Principal component analysis (PCA) The definition of principal component analysis is the analysis of data that has been transformed from the original axes to the principal axes. PCA is a useful technique to reduce the dimensionality of large data sets, such as those from metabolomics analysis. PCA is also useful to identify significant signals in noisy data. The mathematical technique used in PCA is called eigen analysis. The eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products can be solved from data matrix obtained from metabolite analysis. In many cases, the data matrix for PCA should be prepared from data obtained by GC-MS, LC-MS or CE-MS. Therefore, the target metabolites should be used as an independent variable and the amount of the corresponding metabolite should be used as dependent variable. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix. PCA can identify and indicate useful information from the metabolome using a few principal components. In fact, the application of PCA to a metabolome data set provides two quantities: the score and the loading. The loading allows the evaluation of the contribution that each metabolite makes to the total information of the metabolome. The loading is useful to understand differences among samples in each metabolite level. The PCA score is defined as the coordinate of data vectors in the base of the principal component analysis. The score plot, limited to the most significant principal components, gives a visual image of the differences of samples from an all-around view point. The first principal axis is the direction in which the data are primarily distributed in n-dimensional space. In the field of biology or biochemistry, the principal component is defined as a constituent that is most abundant or that is most important. Differences between informaticians and biologists sometimes cause misunderstandings and each group should appreciate the viewpoint of the other.

Hierarchical cluster analysis (HCA) HCA is also used frequently in metabolomics. HCA is a method of cluster analysis based on the multivariate distance between every pair of data points. In HCA, the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. HCA is subdivided into agglomerative methods, which are proceeded by a series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings. Agglomerative techniques are commonly used in metabolomics. HCA may be represented by a two-dimensional diagram known as a dendrogram, which illustrates the fusions or divisions made at each successive stage of analysis. The n agglomerative

hierarchical clustering procedure produces a series of partitions of the data, P_n, P_{n-1}, ... P_1. The first P_n consists of n single object clusters, the last P_1, consists of a single group containing all n cases. At each particular stage, the two clusters which are closest together are joined. Different ways of defining distance between clusters are available.

The linkage methods are divided into single linkage clustering (nearest neighbor technique), complete linkage clustering (farthest neighbor technique), average linkage clustering, average group linkage. HCA is familiar to biologists because it is often applied for phylogenetic studies based on the sequence homologies of several orthologic genes such as 16S rDNA. This method is accurate but is very computer-intensive when a data set has a huge number of data points. For large numbers of data points, the k-mean clustering (kMC) or batch learning self-organizing mapping (BL-SOM) is preferred. HCA is sometimes used after the data have first been transformed into their principal component.

Self-organizing mapping (SOM) SOM is one of the noncluster exploring data analysis methods employing neural network technology. In addition to PCA and HCA, SOM is used in omics sciences including genomics and transcriptomics. It should be possible to apply the SOM algorithm to metabolomics (40). The original SOM algorithm (41, 42) requires rather a long time for calculation, and it might afford different clustering results in its topology depending on the order of data input. Recently, the original SOM has been improved to a batch-learning SOM (BL-SOM), which would not be affected by input order. BL-SOM can be conducted in the laboratory using a Personal Computer because the algorithm does not require high CPU (central processing unit) power. BL-SOM is widely used for genomics and transcriptomics (43, 44). HCA that presents a dendrogram based on the distance of each sample point would be useful for visceral comparison. However, HCA might afford false images when each data set is in close formation with other sets. This unwelcome phenomena might be caused when the linking method used is mismatched for the data set structure. PCA might not work well in case the of nonlinear or noncontiguous data structures. For example, the fold change in the amounts of metabolites obtained by time-course sampling should be analyzed by SOM or KMC. However, SOM can provide no information on why clusters are separated. Careful consideration is required for the practical usage of SOM.

Other data mining methods Several other multivariate analysis methods in addition to PCA, HCA and SOM are also available for plant metabolomics. Soft independent modeling of class analogy (SIMCA) is a method for the classification and prediction of unknown samples by means of the principal component models that are prepared in each category of training sets (known samples). SIMCA is useful for the profiling of a large number of samples. K-nearest neighbor (KNN) and k-mean cluster analysis (kMN) are available for sample classification. Principal component regression (PCR) or partial least squares regression (PLS) are also useful in several cases. A de facto standard protocol for data mining in metabolomics has not been established. Therefore, data mining of metabolomics is now being operated in an on-demand fashion. At present, metabolomics can be considered almost more an art than a science.

PRACTICAL OPERATION OF METABOLOMICS

Metabolomics could be integrated with other omics sciences including transcriptomics or proteomics. However, de facto standard protocols for integrated analytical systems have not been established. The throughput of proteomics might be less than one hundredth of that of metabolomics. Transcriptomics is an extremely expensive protocol despite its rather low throughput. However, metabolomics has already been considered as an apparently useful technology despite a lack of integration with other omics science in the case of restricted aims. The crucial points for the practical operation of metabolomics are described below. The application of metabolomics is also discussed.

Quantitative performance or resolution power? It is obvious that both quantitative performance and resolution are desirable in general analytical chemistry. However, on occasion it must be decided which one is of primary importance. To identify biomarkers indicating extreme perturbations imposed on test organisms, resolution power is of primary importance although the analytical system might not be quantitative. In this case, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) is currently the best tool. On the other hand, in the case of the classification of unknown samples based on finger-printing patterns, the first priority is quantitative performance and repeatability. GC-MS, LC-MS, FT-IR are almost suitable in this case.

Focused (biased) or exhaustive (nonbiased)? When a certain working hypothesis has been established, a specific biosynthetic pathway can be proposed. In such cases, a few samples upon which perturbations were imposed, including physical stresses, biotic stress, mutation, transgenic stress should be evaluated in a small-scale experiment. Metabolomics should focus on the specific biosynthetic pathway or on the specific category of metabolites. This means that target analysis and small-scale metabolic profiling should be conducted in depth. Time-course sampling or a combination of several sets of perturbations that are imposed should be considered. Snap-shot analyses for unknown samples without any significant hypothesis should be conducted in the nonbiased mode, which involves identifying biomarkers.

Reproducibility and repeatability When an exactly equal result is obtained in the same experiment under the same specification with the same instrument by the same person, this is defined as repeatability. When a similar experiment is performed at a different place with different equipment by a different person and a similar result is obtained, this is termed reproducibility. Repeatability must be the minimum requirement for experiments. Thus, the specification of the system should be lowered until the repeatability is assured. In detail, the following solution can be proposed: (i) lowering the threshold of observation sensitivity; (ii) decreasing the number of index compounds; and (iii) reducing the accuracy of detection. Generally, it is difficult to maintain repeatability at a high level because many factors that might compromise repeatability are involved in metabolomics schemes. Careful consideration must be made

and technical know-how should be applied at each step.

With respect to reproducibility, only a general tendency might be reproducible upon re-examination. It is very difficult to maintain an exact reproducibility of metabolomics at the present moment because of an incompatibility between the mass spectrometry equipment used by the original researcher and the equipment used by other researchers. In particular, it is almost impossible to reproduce metabolic fingerprinting experiments because the clustering topology is affected by slight differences in mass spectrometric peak patterns. Apart from mass spectrometry differences, other processes, such as plant growth, sampling, derivatization might also induce experimental error. Careful consideration is required in the case of exchanging metabolomics experimental data between researchers. In fact, metabolomics research imposes several biases regarding the class of targets or analytical protocols due to the general technical problems being encountered at present. This means that most metabolomics research should be termed biased metabolomics. Recently, several attempts to impose a strong bias on the class of analytes have been reported. Focused metabolomics includes lipidomics (focusing on lipids) (45–47), glycomics (focusing on glycans or glycosides) (48, 49), peptidomics (focusing on peptides) (50, 51) and RNomics (focusing on siRNA or miRNA) (52, 53). New biased metabolomics will be established based on need in the future.

FUTURE PERSPECTIVES

A de facto standard protocol of metabolomics has not been established although metabolomics is obviously becoming a promising tool for functional genomics as described above. The lack of a standard protocol makes bioscientists hesitate to use metabolomics as a research tool. Researchers who are developing new metabolomics protocols must explain their own technology as clearly as possible for bioscientists. Bioscientists should identify areas in which metabolomics can contribute meaningfully. Only through good collaboration, will metabolomics receive its deserved position in functional genomics.

ACKNOWLEDGMENTS

This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO). This work was also supported in part by the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS).

REFERENCES

1. **Edwards, D. and Batley, J.:** Plant bioinformatics: from genome to phenome. Trends Biotechnol., **22**, 232–237 (2004).
2. **Weckwerth, W.:** Metabolomics in systems biology. Annu. Rev. Plant Biol., **54**, 669–689 (2003).
3. **Dixon, R. A. and Strack, D.:** Phytochemistry meets genome analysis, and beyond. Phytochemistry, **62**, 815–816 (2003).
4. **Fiehn, O.:** Metabolomics — the link between genotypes and phenotypes. Plant Mol. Biol., **48**, 155–171 (2002).
5. **Trethewey, R. N.:** Metabolite profiling as an aid to metabolic engineering in plants. Curr. Opin. Plant Biol., **7**, 196–201 (2004).
6. **Trethewey, R. N. and Fukusaki, E.:** Industrial metabolic profiling. Bio Indstry, **21**, 41–46 (2004). (in Japanese)
7. **Fukusaki, E., Ikeda, T., Suzumura, D., and Akio, K.:** A facile transformation of *Arabidopsis thaliana* using ceramic supported propagation system. J. Biosci. Bioeng., **96**, 503–505 (2003).
8. **Fiehn, O., Kopka, J., Trethewey, R. N., and Willmitzer, L.:** Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal. Chem., **72**, 3573–3580 (2000).
9. **Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L.:** Metabolite profiling for plant functional genomics. Nat. Biotechnol., **18**, 1157–1161 (2000).
10. **Maharjan, R. P. and Ferenci, T.:** Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. Anal. Biochem., **313**, 145–154 (2003).
11. **Blau, K. and Halket, J. M.:** Handbook of derivatives for chromatography, 2nd ed. John Wiley & Sons, Chichester (1993).
12. **Mueller, C., Schaefer, P., Stoertzel, M., Vogt, S., and Weinmann, W.:** Ion suppression effects in liquid chromatography: electrospray-ioniztion transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass sectra libraries. J. Chromatogr. B, **773**, 47–52 (2002).
13. **King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T.:** Mechanistic investigation of ionization suppression in electrospray ionization. J. Am. Soc. Mass Spectrom., **11**, 942–950 (2000).
14. **Han, D. K., Eng, J., Zhou, H., and Aebersold, R.:** Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol., **19**, 9469–9451 (2001).
15. **Zhang, R., Sioma, C. S., Wang, S., and Regnier, F. E.:** Fractionation of isotopically labeled peptides in quantitative proteomics. Anal. Chem., **73**, 5142–5149 (2001).
16. **Fukusaki, E. i., Harada, K., Bamba, T., and Kobayashi, A.:** An isotope effect on the comparative quantification of flavonoids by means of methylation-based stable isotope dilution coupled with capillary liquid chromatograph/mass spectrometry. J. Biosci. Bioeng., **99**, 75–77 (2005).
17. **Mougous, J. D., Leavell, M. D., Senaratne, R. H., Leigh, C. D., Williams, S. J., Riley, L. W., Leary, J. A., and Bertozzi, C. R.:** Discovery of sulfated metabolites in mycobacteria with a genetic and mass spectrometric approach. Proc. Natl. Acad. Sci. USA, **99**, 17037–17042 (2002).
18. **Wu, L., Mashego, M. R., van Dam, J. C., Proell, A. M., Vinke, J. L., Ras, C., van Winden, W. A., van Gulik, W. M., and Heijnen, J. J.:** Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. Anal. Biochem., **336**, 1641–1671 (2005).
19. **Tomita, M. and Nishioka, T.:** Forefront of metabolomics research. Springer Verlag Tokyo, Tokyo (2003). (in Japanese)
20. **Tanaka, N., Kobayashi, H., Ishizuka, N., Minakuchi, H., Nakanishi, K., Hosoya, K., and Ikegami, T.:** Monolithic silica columns for high-efficiency chromatographic separations. J. Chromatogr. A, **965**, 35–49 (2002).
21. **Tanaka, N., Kobayashi, H., Nakanishi, K., Minakuchi, H., and Ishizuka, N.:** Monolithic LC columns. Anal. Chem., **73**, 420A–429A (2001).
22. **Bamba, T., Fukusaki, E., Nakazawa, Y., and Kobayashi, A.:** Rapid and high-resolution analysis of geometric polyprenol homologues by connected octadecylsilylated monolithic silica columns in high-performance liquid chromatography. J. Sep. Sci., **27**, 293–296 (2004).
23. **Tolstikov, V. V., Lommen, A., Nakanishi, K., Tanaka, N., and Fiehn, O.:** Monolithic silica-based capillary reversed-

phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. Anal. Chem., **75**, 6737–6740 (2003).
24. **Tanaka, N., Kimura, H., Tokuda, D., Hosoya, K., Ikegami, T., Ishizuka, N., Minakuchi, H., Nakanishi, K., Shintani, Y., Furuno, M., and Cabrera, K.:** Simple and comprehensive two-dimensional reversed-phase HPLC using monolithic silica columns. Anal. Chem., **76**, 1273–1281 (2004).
25. **Wienkoop, S., Glinski, M., Tanaka, N., Tolstikov, V., Fiehn, O., and Weckwerth, W.:** Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. Rapid Commun. Mass Spectrom., **18**, 643–650 (2004).
26. **Bamba, T., Fukusaki, E., Kajiyama, S., Ute, K., Kitayama, T., and Kobayashi, A.:** High-resolution analysis of polyprenols by supercritical fluid chromatography. J. Chromatogr. A, **911**, 113–117 (2001).
27. **Bamba, T., Fukusaki, E., Nakazawa, Y., Sato, H., Ute, K., Kitayama, T., and Kobayashi, A.:** Analysis of long-chain polyprenols using supercritical fluid chromatography and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. J. Chromatogr. A, **995**, 203–207 (2003).
28. **Halket, J. M., Przyborowska, A., Stein, S. E., Mallard, W. G., Down, S., and Chalmers, R. A.:** Deconvolution gas chromatography/mass spectrometry of urinary organic acids — potential for pattern recognition and automated identification of metabolic disorders. Rapid Commun. Mass Spectrom., **13**, 279–284 (1999).
29. **Aharoni, A., Ric de Vos, C. H., Verhoeven, H. A., Maliepaard, C. A., Kruppa, G., Bino, R., and Goodenowe, D. B.:** Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. Omics, **6**, 217–234 (2002).
30. **Baileya, N. J. C., Ovenb, M., Holmesa, E., Nicholsona, J. K., and Zenkc, M. H.:** Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ^{1}H NMR spectroscopy and chemometrics. Phytochemistry, **62**, 851–858 (2003).
31. **Ott, K.-H., Aranibar, N., Singh, B., and Stockton, G. W.:** Metabolomic classifies pathways affected by bioactive compouds. Artificial neural network classification of NMR spectra of plant extracts. Phytochemistry, **62**, 971–985 (2003).
32. **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**, 949–957 (2003).
33. **Wolfender, J. L., Ndjoko, K., and Hostettmann, K.:** The potential of LC-NMR in phytochemical analysis. Phytochem. Anal., **12**, 2–22 (2001).
34. **Griffin, J. L.:** Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. Curr. Opin. Chem. Biol., **7**, 648–654 (2003).
35. **Johnson, C. S., Jr.:** Diffusion ordered NMR spectroscopy: principles and applications. Prog. NMR Spectrosc., **34**, 203–255 (1999).
36. **Gidmana, E., Goodacreb, R., Emmettc, B., Smitha, A.R., and Gwynn-Jonesa, D.:** Investigating plant-plant interference by metabolic fingerprinting. Phytochemistry, **63**, 705–710 (2003).
37. **Johnson, H. E., Broadhurst, D., Goodacre, R., and Smith, A. R.:** Metabolic fingerprinting of salt-stressed tomatoes. Phytochemistry, **62**, 919–928 (2003).
38. **Bamba, T., Fukusaki, E., Nakazawa, Y., and Kobayashi, A.:** *In-situ* chemical analyses of trans-polyisoprene by histochemical staining and Fourier transform infrared microspectroscopy in a rubber-producing plant, *Eucommia ulmoides* Oliver. Planta, **215**, 934–939 (2002).
39. **Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M., and Moritz, T.:** A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal. Chem., **76**, 1738–1745 (2004).
40. **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. Natl. Acad. Sci. USA, **101**, 10205–10210 (2004).
41. **Kohonen, T.:** Self-organized formation of topologically correct feature maps. Biol. Cybern., **43**, 59–69 (1982).
42. **Kohonen, T.:** The self-organizing map. Proc. IEEE, **78**, 1464–1480 (1990).
43. **Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T., and Ikemura, T.:** Informatics for unveiling hidden genome signatures. Genome Res., **13**, 693–702 (2003).
44. **Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., and Ikemura, T.:** Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. Gene, **276**, 89–99 (2001).
45. **Houjou, T., Yamatani, K., Imagawa, M., Shimizu, T., and Taguchi, R.:** A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom., **19**, 654–666 (2005).
46. **Lu, Y., Hong, S., Tjonahen, E., and Serhan, C. N.:** Mediator-lipidomics: databases and search algorithms for PUFA-derived mediators. J. Lipid Res., **46**, 790–802 (2005).
47. **Taguchi, R.:** Systems for lipidomics: metabolomics focused on lipids. Tanpakushitsu Kakusan Koso, **49**, 1911–1916 (2004). (in Japanese).
48. **Hirabayashi, J.:** Oligosaccharide microarrays for glycomics. Trends Biotechnol., **21**, 141–143 (2003).
49. **Ratner, D. M., Adams, E. W., Disney, M. D., and Seeberger, P. H.:** Tools for glycomics: mapping interactions of carbohydrates in biological systems. Chembiochem, **5**, 1375–1383 (2004).
50. **Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., De Loof, A., and Schoofs, L.:** Peptidomics. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., **803**, 3–16 (2004).
51. **Minamino, N.:** Peptidome: the fact-database for endogenous peptides. Tanpakushitsu Kakusan Koso, **46**, 1510–1517 (2001). (in Japanese).
52. **Huttenhofer, A., Cavaille, J., and Bachellerie, J. P.:** Experimental RNomics: a global approach to identifying small nuclear RNAs and their targets in different model organisms. Methods Mol. Biol., **265**, 409–428 (2004).
53. **Marker, C., Zemann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J. P., Green, P., Bachellerie, J. P., Brosius, J., and Huttenhofer, A.:** Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. Curr. Biol., **12**, 2002–2013 (2002).