J. Pestic. Sci., **31**(3), 300–304 (2006)

Commentary

(Special topic)

Technical problems and practical operations in plant metabolomics

Takeshi BAMBA and Eiichiro FUKUSAKI[†,*]

Department of Applied Environmental Biology, Graduate School of Pharmaceutical Sciences, Osaka University, 1–6 Yamadaoka, Suita, Osaka 565–0871, Japan
[†] *Department of Biotechnology, Graduate School of Engineering, Osaka University, 2–1 Yamadaoka, Suita, Osaka 565–0871, Japan*

(Received June 5, 2006)

A metabolomics experiment consists of several complicated technical elements with each step ('cultivation of organisms,' 'sampling,' 'sample preparation,' 'analysis,' 'data conversion,' and 'informatics') potentially giving rise to experimental error. In order to perform metabolomics studies, it is necessary to understand the method limitations in detail and recognize possible problems at each step. Here, we review a number of technical problems associated with plant metabolomics and describe some practical knowledge for experimental design. © Pesticide Science Society of Japan

Keywords: metabolomics, metabolome, metabolite profiling, metabolic fingerprinting, mass spectrometry, chemometrics.

Introduction

Metabolomics can be defined as a comprehensive analysis in which all the metabolites of an organism are identified and quantified.[1)] This field has emerged as a follow-up to functional genomics methodology that can contribute to our understanding of complex molecular interactions in biological systems.[2)] As such, metabolomics represents the logical progression from large-scale analysis of RNA and proteins towards an understanding of systems biology.[3)]

The final goal of metabolomics is an exhaustive profiling of all metabolites present in a target organism.[1,4,5)] Although it is currently analytically impossible to profile thousands of metabolites, there are a number of approaches that are capable of acquiring significant amounts of data in a single experimental analysis. At the present, the conventional classification of these approaches originally proposed by Dr. Fiehn has been accepted world wide.[1)] "Target analysis" is similar to classical biochemical analysis of metabolites. "Metabolite profiling" implies restricted profiling focusing on specific pathway and metabolite groups.[6–8)] The third category, "metabolomics," is exhaustive profiling.[9–11)] The final grouping, "metabolic fingerprinting," refers to chemometrics tactics, in which the precise peak identification is not essential.[12–14)]

From a technological viewpoint, metabolomics is a complicated interdisciplinary science that consists of a diverse set of approaches, including bioscience, analytical chemistry, natural product chemistry, chemometrics, and informatics (Fig. 1). However, a *de facto* standard protocol for metabolomics has not yet been established because a typical metabolomics experiment involves multiple steps: 'cultivation of organism,' 'sampling,' 'extraction,' 'pretreatment,' 'derivatization,' 'analysis,' 'data conversion,' and 'data mining.' Each step is a potential source of experimental error. The complicated nature of the data acquisition and subsequent analysis is the main reason why metabolomics is not yet widely used. Here we present some of the technical problems associated with plant metabolomics research.

Cultivation of Organism, Sampling and Extraction

It is important to pay careful attention to each sample preparation step in order to reduce experimental error.[15)] The optimum preparation protocol to ensure sample reproducibility should be developed depending on a case-by-case basis by considering the following: First, it is essential to control not only the growth stage but also the exact time of sampling, the area and sample amount. Second, 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 have a very rigid tissue matrix. We use a ball mill with zirconium balls. Third, a solid matrix for pre-column work should be carefully chosen

* To whom correspondence should be addressed.
E-mail: fukusaki@bio.eng.osaka-u.ac.jp
© Pesticide Science Society of Japan

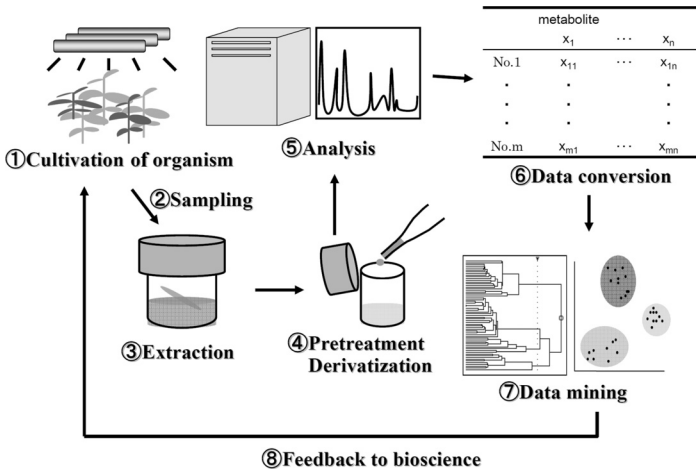

Fig. 1. General scheme of metabolomics and necessary science fields.

according to the manufacture's instruction manuals and application data. Fourth, the concentration range of target metabolites should be considered. The dynamic range of metabolite concentration can extend over 10^9 (from pmol/L to mmol/L). A simultaneous analysis for low and high abundant metabolites is practically impossible due to a lack of dynamic range of mass spectrometry. Therefore, multi-step dilution of samples is often required to prevent loss of metabolome information. Actually we conduct two-step dilutions of samples for gas chromatography/mass spectrometry (GC/MS) analysis and integrate the data obtained.

In the case of target quantification, the selection of appropriate internal standards as well as a suitable purification scheme is extremely important. Unfortunately, it can be rather difficult to choose an appropriate internal standard for normalization in metabolic profiling due to variability in the metabolites structure and physical properties. It is therefore often necessary to employ a range of internal standards that serve to cover the chemical space of the compounds being identified/quantified. It is also import to verify the efficiency of extraction and fractionation processes, often requiring the use of surrogate standards.

Derivatization

Derivatization of target metabolites may be required depending on the analytical platform used. For example, only thermally stable volatile compounds can be analyzed by GC-MS.[16] Polar metabolites generally require derivatization by silylation or other methods prior to analysis.[17] High-performance liquid chromatography (HPLC) also often requires derivatization if the analyte does not contain a chromo- or fluorophore.[18] A good handbook collecting various derivatization method is available.[19] As a matter of course, derivatizing conditions, including the category of reagent and reaction conditions, and the stability of the derivatized product greatly affect the quality of analyses. The procedures to optimize the derivatization protocol for metabolic analysis of *Arabidopsis thaliana* by GC/MS have been investigated using design-of-experient procedures.[20]

Instrumental Analysis

Choice of analytical instrumentation is important in initiating a metabolomics study, which requires the ability to analyze a range of target metabolites, quantitative linearity, resolution and throughput. It is almost impossible to implement all the specifications and the appropriate system configuration should be selected based upon analytical requirement. Some examples with their advantages and limitations are listed in Table 1. The system 1 using Fourier transform infrared spectroscopy (FT-IR)[13,21,22] or nuclear magnetic resonance (NMR)[14,23,24] is a high throughput system for metabolic fingerprinting that does not contain a separation process. This approach can cause difficulty in metabolite identification. However, abundant metabolites that have specific functional group can be quantified. The authors reported a metabolic mapping analysis of plant microscopic preparation.[25] Microscopic FT-IR imaging is also useful for metabolic fingerprinting at the cellular level. Advances in mass spectrometry in sensitivity and accuracy have lead to the development of direct injection based studies using electrospray ionization mass spectrometry (ESI MS) and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS).[9,10,12] Direct injection analysis offers a high throughput system, which is very useful for screening large numbers of samples. However, a disadvantage is that the data quality can be suspect due to ion ization suppression. Ionization suppression occurs when an analyte's ionization efficiency is suppressed owing to contaminants that coexist in the ionization source.[26,27]

The system 2 is a combination of chromatography or electrophoresis and mass spectrometry in which a detail variation can be observed at individual metabolites levels.[10,11] Among them, liquid chromatograhy/mass spectrometry (LC/MS) is thought to be the most universal method due to its broad metabolite coverage.[28–30] However, ionization suppression can be a serious problem in certain LC/MS applications. One approach to overcome this limitation is through chromato-

Table 1. Comparison of metabolomics systems

System	Equipment	Advantage	Defect
1. Metabolic fingerprinting	FT-IR, NMR	Convenient, High-throughput, Reproducible	Lack of metabolic information, Low resolution
2. Metabolite profiling	GC-MS, CE-MS, LC-MS	Available for metabolic information, High resolution	Low quantitative
3. Stable isotope dilution based profiling	GC-MS, CE-MS, LC-MS	Highly quantitative	Costly, Restricted in targets

graphic separation prior to mass spectrometric analysis. Although, it is not possible to achieve complete separation of complex mixtures, chromatographic separation can significantly reduce ion suppression.

Recently, a high-efficiency HPLC system consisting of monolithic silica columns composed of network type silica skeletons has been developed.[31,32] These systems possess a number of advantages including: (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 approximately 100,000～200,000 theoretical plates. The usefulness of the system has been proved by metabolic profiling of Arabidopsis metabolites.[29] We employed a monolith column HPLC system to accomplish the perfect separation of naturally derived polyprenol geometric isomers.[33,34] In the near future, a two dimensional micro HPLC system using monolithic silica capillary column is expected to be available.[35,36]

Another approach for solving the ion suppression problem is the utilization of a stable isotope dilution-based comparative quantification, which is 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 or CE/MS analysis. On the chromatogram or electropherogram, target metabolites and their corresponding isotopomers are co-eluted. 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 proteomics research, isotope coded affinity tags (ICAT).[37] We conducted methylation based post-sampling stable isotope labeling for metabolomics. In that study, we employed ^{13}C stable isotopes because deuterium labeling may have some problems.[38] *In vivo* stable isotope enrichment is also a promising method for stable isotope dilution. We applied *in vivo* ^{15}N stable isotope enrichment to achieve accurate quantification of rather labile nitrogen containing metabolites.[39]

Data Conversion

In metabolomics, raw data from instrumental analysis, such as chromatogram, electropherogram, spectra, *etc.*, should be converted to appropriate data matrices.[14,16,40,41] Only essential components should be selected from the raw data and subjected to further mining steps. A brief explanation is addressed using the example of GC-MS analysis. According to the conventional analysis, a peak list, in which each peak is integrated and identified to be a corresponding metabolite, should be prepared prior to multivariate analysis.[16] However, peak list preparation can be very laborious. Firstly, peak integration might not be accurate due to peak broadening and co-elution. Secondly, perfect peak identification is generally impossible, potentially leading to serious mistakes in the data analysis. We developed a data mining protocol in which chromatogram patterns are directly analyzed without peak list formation (Fig. 2).[42] In our method, all chromatogram data points are used for matrix formation. The method includes all small peaks in the chromatogram and is not affected by peak co-elution. In addition, throughput of the method is very high because peak list preparation is not essential.

On-demand programs are required for data conversion and normalization corresponding to the needs of the individual equipment. The *de facto* standard methods for mass spectra have already been established. However, a proper protocol of chromatogram pattern recognition has not been developed due to some difficulties of noise reduction, base line correction and so on. A trial for data mining of GC/MS mass chromatogram has been reported.[43–45] Several data conversion software packages have been developed independently by a different groups and can be browsed *via* the Internet.[46–48]

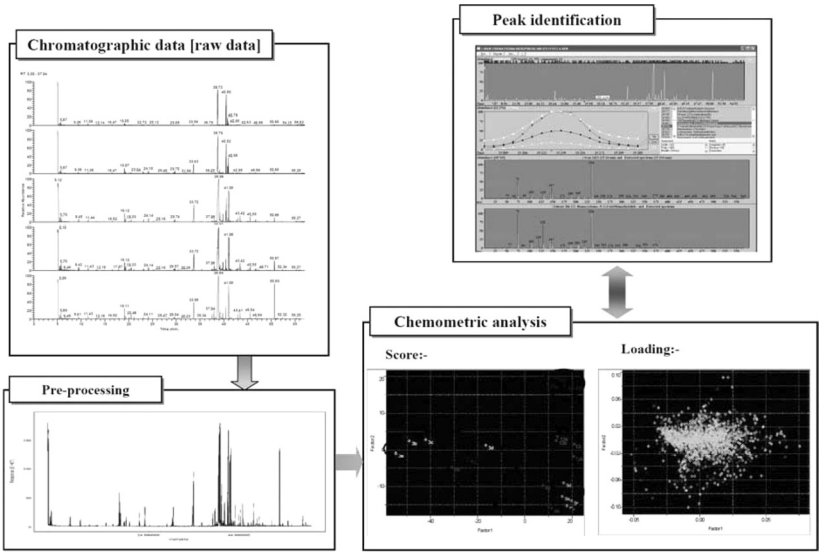

Fig. 2. General scheme of data mining.

Data Mining

The validated data are analyzed by a multivariate analytical algorithm depending on data structure and mining intention. Multivariate analysis methodology used includes multiple regression, discriminant analysis, principal component analysis (PCA), hierarchical cluster analysis, self organizing mapping and many others.[9,10,13,43,45,49] Among these methods, PCA tends to be used the most often in 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. As described above, metabolomics strongly requires a correct application of chemometrics that is a interdisciplinary science crossing over both analysis and informatics.[42–45] Careful validation of the data prior to data mining is essential.

Conclusion

Truly comprehensive profiling of metabolites is currently not possible. In practice, a combination of different analytical platforms including multiple processes is required (Fig. 3). For example, high throughput analysis should be performed at first for understanding significant variation among the whole samples. Next focused metabolome analysis or target profiling should be conducted, in which high accuracy and moderate reproducibility is required. On-demand fractionations, pre-treatments, and other analyses should be considered. By incorporating experimental results from multiple analytical platforms, one can attempt to reconstruct the full metabolome. In either case, MECE (maturely exclusive and collectively exhaustive) concept is important for design of analytical system in metabolomics.

Acknowledgments

A part of this work was performed in the project “Development of

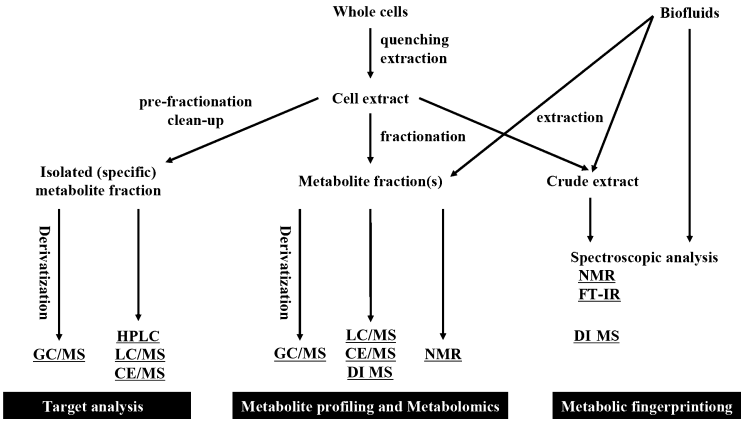

Fig. 3. Practical strategies for metabolomics. DI: direct-infusion.

Fundamental Technologies for Controlling the Production of Industrial Materials by Plants" supported by NEDO (New Energy and Industrial Technology Development Organization, Japan). This work was also supported in part by MEXT (Ministry of Education, Culture, Sports, Science and Technology) and JST_CREST (Japan Science and Technology Agency). The authors are very grateful to Mr. T. Yamaji (GL Sciences Inc.) for helpful discussion and experimental support.

References

1) O. Fiehn: *Plant. Mol. Biol.* **48**, 155–171 (2002).
2) R. Hall, M. Beale, O. Fiehn, N. Hardy, L. Sumner and R. Bino: *Plant Cell* **14**, 1437–1440 (2002).
3) W. Weckwerth: *Annu. Rev. Plant. Biol.* **54**, 669–689 (2003).
4) L. W. Sumner, P. Mendes and R. A. Dixon: *Phytochemistry* **62**, 817–836 (2003).
5) R. J. Bino, R. D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B. J. Nikolau, P. Mendes, U. Roessner-Tunali, M. H. Beale, R. N. Trethewey, B. M. Lange, E. S. Wurtele and L. W. Sumner: *Trends Plant Sci.* **9**, 418–425 (2004).
6) S. Landini, M. Y. Graham and T. L. Graham: *Phytochemistry* **62**, 865–874 (2003).
7) J. Burns, P. D. Fraser and P. M. Bramley: *Phytochemistry* **62**, 939–947 (2003).
8) M. Stobiecki, I. Matysiak-Kata, R. Franski, J. Skala and J. Szopa: *Phytochemistry* **62**, 959–969 (2003).
9) M. Y. Hirai, M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara and K. Saito: *Proc. Natl. Acad. Sci. USA* **101**, 10205–10210 (2004).
10) G. S. Catchpole, M. Beckmann, D. P. Enot, M. Mondhe, B. Zywicki, J. Taylor, N. Hardy, A. Smith, R. D. King, D. B. Kell, O. Fiehn and J. Draper: *Proc. Natl. Acad. Sci. USA* **102**, 14458–14462 (2005).
11) T. Soga, R. Baran, M. Suematsu, Y. Ueno, S. Ikeda, T. Sakurakawa, Y. Kakazu, T. Ishikawa, M. Robert, T. Nishioka and M. Tomita: *J. Biol. Chem.* **281**, 16768–16776 (2006)
12) R. Goodacre, E. V. York, J. K. Heald and I. M. Scott: *Phytochemistry* **62**, 859–863 (2003).
13) H. E. Johnson, D. Broadhurst, R. Goodacre and A. R. Smith: *Phytochemistry* **62**, 919–928 (2003).
14) J. L. Ward, C. Harris, J. Lewis and M. H. Beale: *Phytochemistry* **62**, 949–957 (2003).
15) S. G.Villas-Boas, S. Mas, M. Akesson, J. Smedsgaard and J. Nielsen: *Mass Spectrom. Rev.* **24**, 613–646 (2005).
16) C. Wagner, M. Sefkow and J. Kopka: *Phytochemistry* **62**, 887–900 (2003).
17) O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R. N. Trethewey and L. Willmitzer: *Nat. Biotechnol.* **18**, 1157–1161 (2000).
18) J. Lopez-Cervantes, D. I. Sanchez-Machado and J. A. Rosas-Rodriguez: *J. Chromatogr. A* **1105**, 106–110 (2006).
19) K. Blau: "Handbook of Derivatives for Chromatography," ed. by K. Blau and J. M. Hanlket, 2nd ed., John Wiley & Sons Inc., Great Britain, 1993.
20) J. Gullberg, P. Jonsson, A. Nordstrom, M. Sjostrom and T. Moritz: *Anal. Biochem.* **331**, 283–295 (2004).
21) E. Gidman, R. Goodacre, B. Emmett, A. R. Smith and D. Gwynn-Jones: *Phytochemistry* **63**, 705–710 (2003).
22) H. E. Johnson, D. Broadhurst, D. B. Kell, M. K. Theodorou, R. J. Merry and G. W. Griffith: *Appl. Environ. Microbiol.* **70**, 1583–1592 (2004).
23) N. J. Bailey, M. Oven, E. Holmes, J. K. Nicholson and M. H. Zenk: *Phytochemistry* **62**, 851–858 (2003).
24) K. H. Ott, N. Aranibar, B. Singh and G. W. Stockton: *Phytochemistry* **62**, 971–985 (2003).
25) T. Bamba, E. Fukusaki, Y. Nakazawa and A. Kobayashi: *Planta* **215**, 934–939 (2002).
26) R. King, R. Bonfiglio, C. Fernandez-Metzler, C. Miller-Stein and T. Olah: *J. Am. Soc. Mass Spectrom.* **11**, 942–950 (2000).
27) C. Mueller, P. Schaefer, M. Stoertzel, S. Vogt and W. Weinmann: *J. Chromatogr. B* **773**, 47–52 (2002).
28) V. V. Tolstikov and O. Fiehn: *Anal. Biochem.* **301**, 298–307 (2002).
29) V. V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka and O. Fiehn: *Anal. Chem.* **75**, 6737–6740 (2003).
30) I. D. Wilson, R. Plumb, J. Granger, H. Major, R. Williams and E. M. Lenz: *J. Chromatogr. B* **817**, 67–76 (2005).
31) N. Tanaka, H. Kobayashi, K. Nakanishi, H. Minakuchi and N. Ishizuka: *Anal. Chem.* **73**, 420A–429A (2001).
32) N. Tanaka, H. Kobayashi, N. Ishizuka, H. Minakuchi, K. Nakanishi, K. Hosoya and T. Ikegami: *J. Chromatogr. A* **965**, 35–49 (2002).
33) T. Bamba, E. Fukusaki, Y. Nakazawa and A. Kobayashi: *J. Sep. Sci.* **27**, 293–296 (2004).
34) T. Bamba, E. Fukusaki, H. Minakuchi, Y. Nakazawa and A. Kobayashi: *J. Lipid Res.* **46**, 2295–2298 (2005).
35) N. Tanaka, H. Kimura, D. Tokuda, K. Hosoya, T. Ikegami, N. Ishizuka, H. Minakuchi, K. Nakanishi, Y. Shintani, M. Furuno and K. Cabrera: *Anal. Chem.* **76**, 1273–1281 (2004).
36) T. Ikegami, T. Hara, H. Kimura, H. Kobayashi, K. Hosoya, K. Cabrera, N. Tanaka: *J. Chromatogr. A* **1106**, 112–117 (2006).
37) D. K. Han, J. Eng, H. Zhou and R. Aebersold: *Nat. Biotechnol.* **19**, 946–951 (2001).
38) E. Fukusaki, K. Harada, T. Bamba and A. Kobayashi: *J. Biosci. Bioeng.* **99**, 75–77 (2005).
39) J. K. Kim, K. Harada, T. Bamba, E. Fukusaki and A. Kobayashi: *Biosci. Biotech. Biochem.* **69**, 1331–1340 (2005).
40) A. Nordstrom, G. O'Maille, C. Qin and G. Siuzdak: *Anal. Chem.* **78**, 3289–3295 (2006).
41) T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita and T. Nishioka: *J. Proteome Res.* **2**, 488–494 (2003).
42) E. Fukusaki, K. Jumtee, T. Bamba, T. Yamaji and A. Kobayashi: *Z. Naturforsch.* **61c**, 267–272 (2006).
43) P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom and T. Moritz: *Anal. Chem.* **76**, 1738–1745 (2004).
44) P. Jonsson, A. I. Johansson, J. Gullberg, J. Trygg, J. A. B. Grung, S. Marklund, M. Sjostrom, H. Antti and T. Moritz: *Anal. Chem.* **77**, 5635–5642 (2005).
45) P. Jonsson, E. S. Johansson, A. Wuolikainen, J. Lindberg, I. Schuppe-Koistinen, M. Kusano, M. Sjostrom, J. Trygg, T. Moritz and H. Antti: *J. Proteome Res.* **5**, 1407–1414 (2006).
46) MSFACTs, http://www.noble.org/PlantBio/MS/MSFACTs/MSFACTs.html
47) metAlign, http://www.metalign.nl
48) MZmine, http://mzmine.sourceforge.net/
49) L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam and S. G. Oliver: *Nat. Biotechnol.* **19**, 45–50 (2001).