METABOLOMICS: LC-MS ANALYSIS DEVELOPMENT

V.V. Tolstikov[1], N. Tanaka[2] and O. Fiehn[1]

[1] - Max Planck Institute of Molecular Plant Physiology, Dept. Lothar Willmitzer, 14424 Potsdam, Germany
[2] - Kyoto Institute of Technology, Department of Polymer Science and Engineering, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan

INTRODUCTION

Metabolomics, the comprehensive analysis of metabolites present in a biological sample, has emerged as the third major path of functional genomics besides transcriptomics (mRNA profiling) and proteomics (Fiehn, 2002; Sumner *et al.*, 2003; Jan van der Greef *et al.*, 2003). Combined with the information obtained on transcriptome and proteome, this would lead to a nearly complete molecular picture of the state of a particular biological system at a given time. Monitoring the biological system state as a function of time by above mentioned methods would allow the study of the temporal dynamics of the system homeostasis. The importance of metabolites in control, communication and as building blocks and energy transporters within the biological system provides metabolomics a unique opportunity for phenotypic and dynamic profiling. The analysis of metabolome is dramatically more complex than gene expression analysis. There is an enormous chemical diversity of an organism's metabolome as well as a large concentration range to be covered. This presents a considerable challenge for the development of an appropriate analytical technique. Over 100,000 metabolites are described in the plant kingdom, and it can be assumed, that the number of metabolites of an individual plant species may easily reach some 5,000 individual low-molecular weight compounds. In plant biology and plant functional genomics, most academic and many industrial research groups use the small weed dicotyledonous plant *Arabidopsis thaliana* or the important monocotyledonous crop rice (*Oryza sativa*) as models to tackle fundamental and applied research objectives. Both species are frequently used for plant functional genomics since complete genome sequences and functional gene annotations are publicly available. The most common and mature technique is GC-MS analysis. Over 300 metabolites were covered in a proof-of-concept study on functional genomics in Arabidopsis, using GC-MS technology. Although it has been shown that the number of detected peaks in typical GC-MS plant chromatogram can be multiplied by deconvolution algorithm, the *de novo* identification of GC-MS peaks remains cumbersome. Therefore, needs for development of the complementary technique allowing plant sample analysis without chemical modification and providing enhanced qualitative characterization of the components are clear. During the last decade, LC-MS techniques were developed employing soft ionization methods like electro spray (ESI) or photo ionization (APPI) and, simultaneously, mass spectrometers became both more sophisticated and more robust for daily use. More recently, achievements in separation sciences propose much better solutions for the separation of the complex mixtures than it was attainable before. Specifically, the advance of preparation of monolithic columns enabled the separation of the very complex mixture separations due to the high number of theoretical plates of these columns (Tanaka, 2001). The objective of this study was to develop LC-MS methods of analysis suitable for the plant metabolomics studies, and to apply this for Arabidopsis and rice plants.

MATERIALS AND METHODS

Samples

Application examples for LC-MS analysis were cold methanol extracts of the grounded leaves of *Arabidopsis thaliana* Col-0 and rice (*Oryza sativa*) grown in the green house. At this stage, no specific biological question was followed.

LC-MS^n

In our present studies, we have used the LC-MS system consisted of a Finnigan LCQ DECA mass spectrometer (ThermoFinnigan, San Jose, CA, USA), a Rheos 2000 pump (Flux Instruments AB, Karlskoga, Sweden), and an HTS PAL auto sampler (CTC Analytics, Zwingen, Switzerland). The system was operated under the Xcalibur software (version 1.3, ThermoFinnigan, USA). Data acquisition has been done in positive and negative modes. Chromatography was performed using 6.5 mM ammonium acetate (pH 5.5, adjusted by acetic acid) (A) and highest grade acetonitrile available (Biotech grade solvent, 99,93+%, (SAF), Seelze, Germany) (B), as the mobile phases. Reverse-phased (RP) split chromatography was used for capillary monolithic column at the flow rates of 50-120 µl/min at ambient temperature. LC-MS analysis was performed on monolithic capillary silica based C18 column (600 x 0.2mm ID). With the pre-column split ratio 1:10 mentioned above flow rate provided 5-12 µl/min actual flow rate through the capillary column. Separation was accomplished by stepwise gradients from 5% B to 100% B at 75 min followed by isocratic elution. Hydrophilic interaction chromatography (termed 'HILIC', see Tolstikov, 2002) was performed on PolyHYDROXYETHYL A (PolyLC, USA) columns (4,6x100 mm, 100 A, 3 µm particles size and 0,2x 100mm, 100 A, 3 µm particles size).

RESULTS AND DISCUSSION

Reliable and reproducible GC-MS technique for metabolomics studies is suffering from certain limitations. Specifically, the upper mass limit of metabolites can hardly be increased over the range of trisaccharides due to volatility constraints, which is well below the mass of membrane lipids, or lots of secondary metabolites. Furthermore, metabolites usually require derivatization, which on the one hand may already adversely affect certain compound classes and on the other hand hampers de novo identification of unknown compounds tremendously.

Complementary techniques are capillary electrophoresis or liquid chromatography coupled to mass spectrometry (LC-MS). A number of problems arise from current LC-MS protocols for metabolic profiling. Some of these problems are related to the LC separation itself, others are connected to the MS detectors used. In general, mass spectrometers are able to detect a large number of the compounds simultaneously, even if these are co-eluting. Selectivity is gained by identification and quantification of the compound's specific molecular masses and fragmentation pathways. However, mass spectrometers can only detect ions, with the result, that any quantification will adamantly demand that there is perfect ionization of all compounds throughout the chromatogram. Unfortunately, this is not the case. It is well known among mass spectrometrists, that one compound can hamper the ionization efficacy of another, even if these are chemically related and co-elute at the same time. Therefore, ultimate separation is needed to ensure robust quantification, especially for very complex mixtures consisting of hundreds of components to be analysed and quantified. Figure 1 shows analyses of –15°C cold methanol extracts of 100 mg fresh weight of grounded leaves of *Arabidopsis thaliana* by monolithic C18 reversed phase LC-MS. Compared to other plant species such as rice (*Oryza sativa*), lipophilic metabolic profiles varied largely in abundance and identity of detectable metabolites which is in accordance with background knowledge of comparative genomics between different species. For example, glucosinolates only occur in Brassicaceae like Arabidopsis, and are absent from all other species. In general, many important lipophilic compound classes can be observed such as glucosinolates, flavonoids,

phenolics, anthocyanines, major and minor components of membrane lipids, porphyrins, chlorophylls and their allomers, and miscellaneous and unidentified compounds.

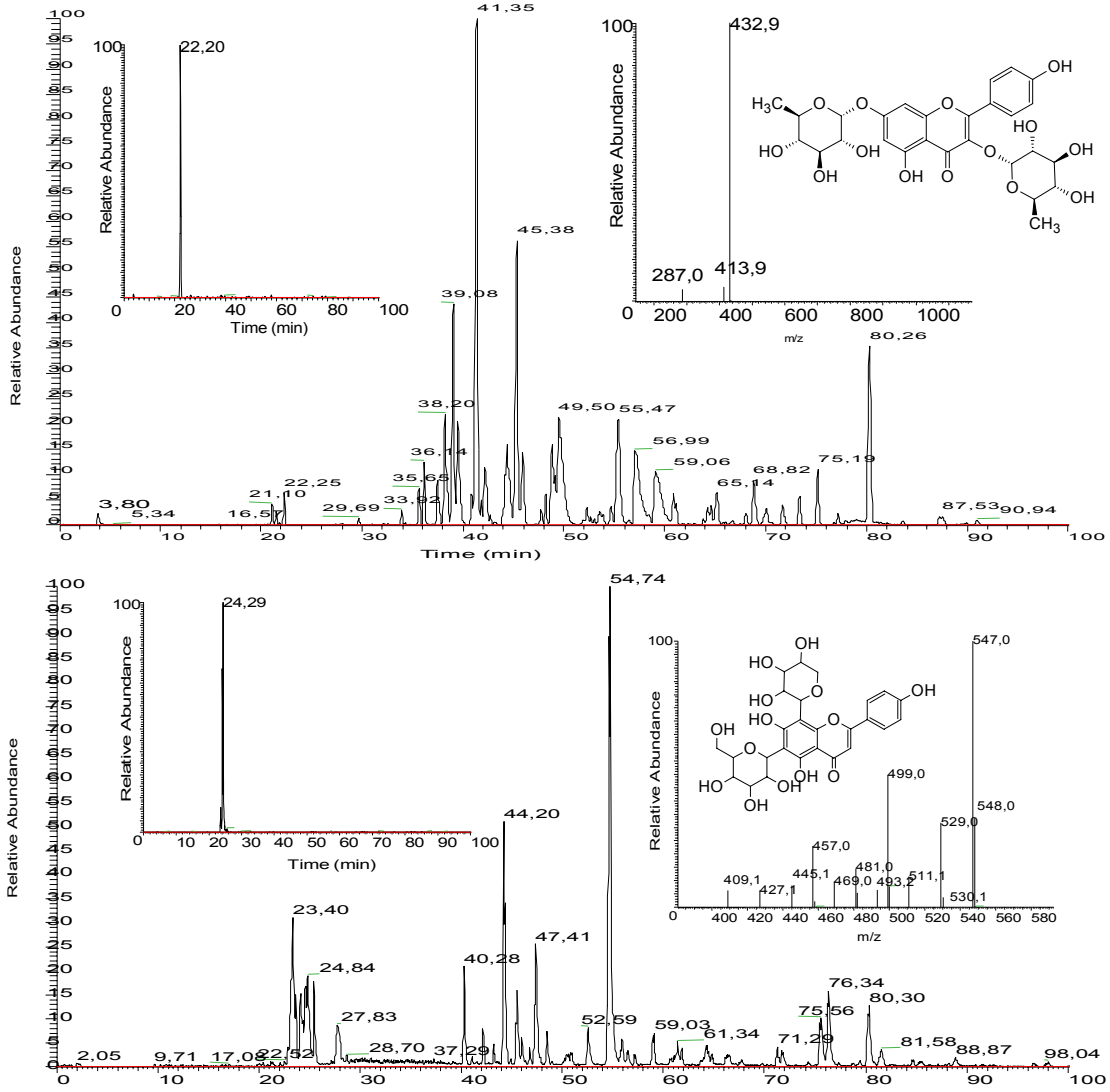

Figure 1. RP/LC-MS chromatograms. Column dimensions are 0.2mm ID x 600mm length. Upper panel: Arabidopsis leaf extract. Lower panel: Rice leaf extract. Examples for specific extracted ion chromatograms are located in the upper left corners. Compound identifications by MS^2 spectra and corresponding structures are situated in the upper right corners.

In this respect, a good coverage of the plant lipophilic and secondary metabolome is achieved. Under the current electro spray conditions, carotenoids do not produce ions and were only detectable using LC post-column flow splitter and UV-VIS detection. The identification process has been described earlier (Tolstikov, 2002) and is mainly based on the fragmentation study and off-line accurate mass measurements, off-line NMR studies, followed by a reconstruction of the molecular structure, in the case where there are no standards available to compare with. A further example of metabolic profiling with the use of currently developed LC-MS technique is shown on Figure 2, demonstrating the results of a cold acclimation experiment performed on *Arabidopsis thaliana* seedlings. Anticipated metabolic changes such as increase in raffinose family oligosaccharides were directly observable from visual

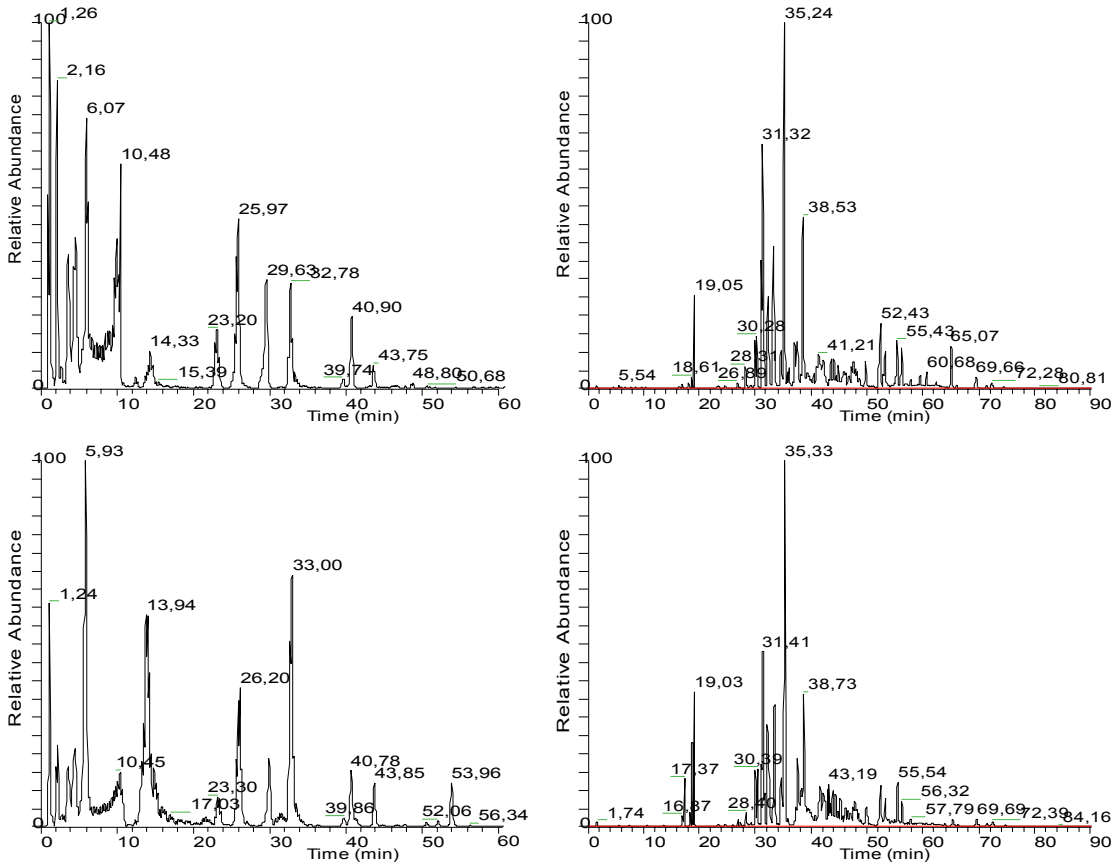

Figure 2. RP and HILIC -MS chromatograms. Upper panel: control plant. Lower panel: cold acclimated plant. Left column: HILIC capillary column. Right column: RP capillary monolithic column.

inspection. A more detailed investigation revealed that more than 20 peaks/components were statistically significantly different between cold acclimated plants samples compare to controls. The relative quantification of the detected components has been done with the assistance of LCQuan software (ThermoFinnigan, USA, included in Xcalibur 1.3 package) by calculation of peak areas for each pre-defined target compound. External standards have been used for calibrations. Peak height/area of a number of components changed reflecting changes in the metabolite concentrations in the sample. Eventually, such changes can be interpreted in terms of alteration of biochemical processes. With respect to metabolome coverage, the total number of detectable peaks is particularly interesting. A careful manual and scan-wise inspection of peaks using the instrument's Xcalibur software revealed about 200 distinct components found on chromatogram for the Arabidopsis extract recorded for RP/LC-MS analysis with the use of monolithic column, positive/negative ionization switching and MS/MS triggering. By applying mass spectral deconvolution with Mass Frontier 3.0 software, 160 peaks were found in positive electro spray mode. Unfortunately, MS deconvolution on continuous positive/negative ionization switching is not possible with any software available to date. This would be highly advantageous since many components are much more efficiently ionized in negative mode than in positive that strictly depends on the chemical structure of the ionized compounds.

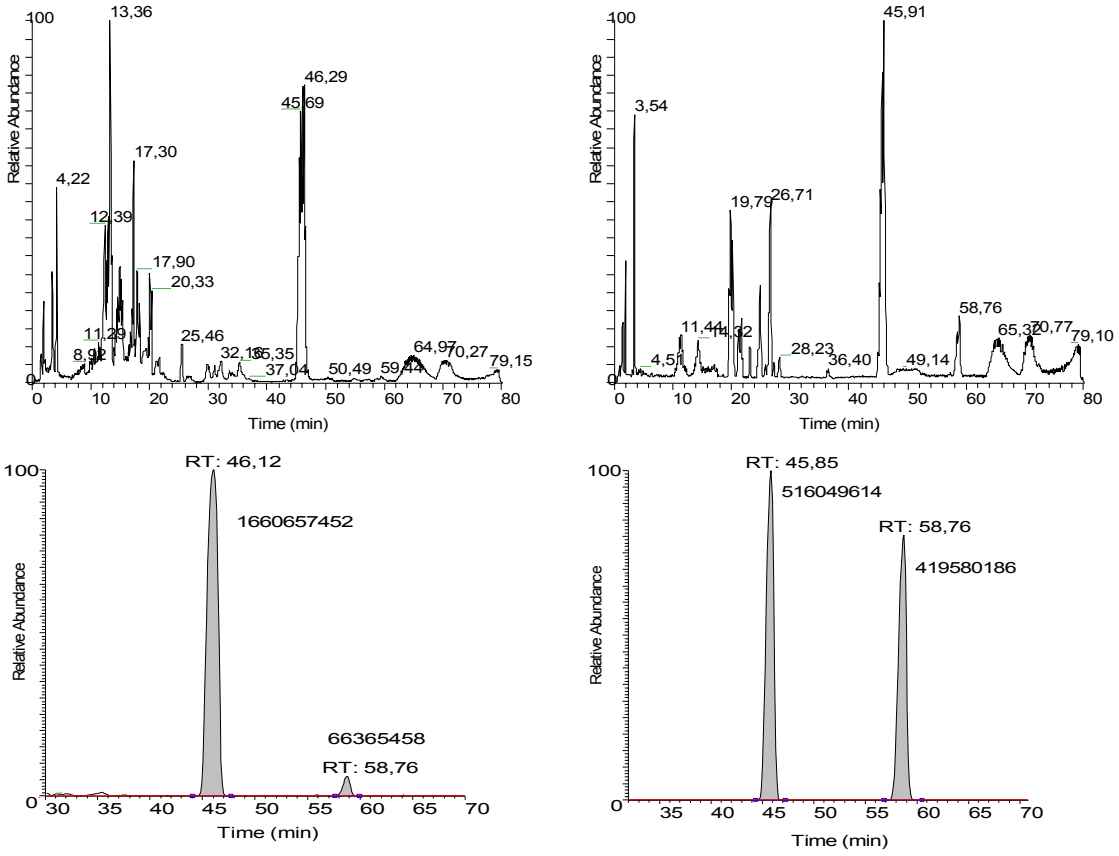

Figure 3. HILIC-MS chromatograms. Upper panel: full chromatogram. Lower panel: sucrose (Rt 46 min) and raffinose (Rt 59 min) peak area on extracted chromatograms for positive ions. Left column: rice leaf extract. Right column: rice grain extract.

Figure 3 illustrates HILIC -MS analyses of methanol extracts of rice (*Oryza sativa*) leaf and rice grain. The analysis of compounds of specific interest is done either on extracted ion chromatograms or on MS^2 transitions. On the lower panel, the determination of a reduced amount of sucrose by extracted ion monitoring is demonstrated concomitant with an elevated amount of raffinose in grain compare to leaf extract.

CONCLUSION

In conclusion, we have briefly described LC-MS methods for plant metabolomic studies based on RP and HILIC chromatography. This technique is complementary to established GC-MS metabolomic profiling and provides unique information on compound identity and mass ranges. We believe the reported data supports the successful integration of capillary silica based monolithic C18 columns into current reversed phase LC-MS protocols.

ACKNOWLEDGEMENTS

This work was financially supported by Metanomics GmbH & Co. KGaA, Berlin, Germany, and by the project fund “Development of micro HPLC for post-genomic analysis", provided by the Ministry of Economy, Trade and Industry, through Kansai Bureau of Economy, Trade and Industry, and Osaka Science and Technology Centre.

REFERENCES

Fiehn, O**.** (2002). Plant.Mol.Biol. 48**:** 155-171.

Jan van der Greef, Eugene Davidov, Elwin Verheij, Jack Vogels, Rob vander Heijden, Aram S. Adourian, Matej Oresic, Edward W. Marple and Stephen Naylor (2003) The Role of Metabolomics in Systems Biology. George G. Harrigan & Royston Goodacre (eds.) Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Boston/Dordrecht/London: Kluwer Academic Publishers 171-198.
Sumner, L.W., Mendes, P. and Dixon, R.A. (2003). Phytochemistry, 62: 817-836.
Tanaka, N., Kobayashi, H., Nakanishi, K., Minakuchi, H., Ishizuka, N. (2001) Anal. Chem. 73: 420A-429A.
Tolstikov, V.V., Fiehn, O. (2002). Anal.Biochem. 301:298-307.
Tolstikov, V., Costisella, B., Weckwerth, W., Zhang, B., Fiehn, O. (2002) Proceedings 50th ASMS Conference on Mass Spectrometry and Allied Topics, Carbohydrates and Glycoconjugates, 146.