

# NMR-based metabolomics at work in phytochemistry

R. Verpoorte · Y. H. Choi · H. K. Kim



Received: 25 July 2006 / Accepted: 16 September 2006 / Published online: 27 February 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** Metabolomics is defined as both the qualitative and quantitative analysis of all metabolites in an organism unraveling correlation with other OMICs data. Many of the technologies used in metabolomics have method-specific advantages and drawbacks in terms of diversity of metabolites detected, sensitivity, or resolution. In this paper, the potential of NMR spectrometry applied to metabolomics is reviewed using examples of *Nicotiana tabacum* and *Catharanthus roseus*

**Keywords** Metabolomics · Metabolic analysis · NMR · *Nicotiana tabacum* · *Catharanthus roseus*

## Introduction

In the past years, several genomes have been completely sequenced and the genomes of more organisms are being sequenced. The number of fully sequenced organisms will grow exponentially. This will open novel possibilities of studying the functions of the genes in these organisms. Functional genomics requires a thorough definition of the phenotype, to be able to determine what changes result from up- or down-regulation of genes. A phenotype can be defined by morphological characteristics, by physiological measurements, as well as by biochemical (transcriptomics or proteomics) and chemical (metabolomics) analysis. In recent years, more and more attention is being paid to the chemical characterization of the phenotype. Chemical charac-

terization can be done on the level of macromolecules (e.g., proteomics and characterization of polysaccharides and lignins) or low molecular weight compounds (the metabolome). The metabolome consists of two types of compounds, the primary metabolites and the secondary metabolites. The primary metabolites are compounds involved in the basic functions of the living cell, such as respiration and biosynthesis of the amino acids and other compounds needed for a living cell. Basically all organisms share the same type of primary metabolites, though per class of organisms there may be large differences in the biosynthetic pathways present, e.g., mammals lack amino acid biosynthesis, as they rely on external sources for their needs. The secondary metabolites are species specific, they play a role in the interaction of a cell with its environment, which can be other cells in the organism or external organisms, e.g., in case of plants to attract pollinators, or to defense itself against pests and diseases. So primary metabolites in plants are important for the growth and agricultural yields, whereas the secondary metabolites are, e.g., concerned with flavor and color of our food and with the resistance of plants against pests and diseases.

Obviously in studying functions of genes metabolomics will be an important tool. Metabolomics is generally defined as both the qualitative and quantitative analysis of all metabolites in an organism. In analyzing a metabolome both the primary and secondary metabolites will be detected. Plants share the primary metabolites, therefore methods that can easily identify and quantify the primary metabolites with high reproducibility are required. One or more methods are required that in all labs and at any time will give the same results, which thus enables a public

R. Verpoorte (✉) · Y. H. Choi · H. K. Kim  
Department of Pharmacognosy, Section Metabolomics,  
Institute of Biology, Leiden University, P.O. Box 9502,  
2300RA, Leiden, The Netherlands  
e-mail: verpoort@chem.leidenuniv.nl

database which can be used in the same way as the databases for protein and gene sequences. Also for the secondary metabolites such a database is needed, though in that case the occurrence of most compounds is restricted to one or a few species only. The reproducibility is also required for studying the effect of any external condition on the plant metabolome, as one needs to know the biological variation. That requires large numbers of analyses to determine the metabolomic changes as function of daily, seasonal, and developmental variation. The results should function as the control samples for all future analyses.

Thus, reproducibility is the most important criteria for developing a metabolomics technology platform. The other criteria concern the ease of quantitation and identification, the number of metabolites that can be measured, and the time needed for an analysis, including the sample preparation. Obviously a high-throughput methodology is preferred, for which robotization is important. Based on the underlying principle, the methods which one may consider for a metabolomic platform can be grouped as follows:

- chromatographic methods (liquid chromatography, gas chromatography (GC), capillary electrophoresis, thin layer chromatography);
- mass spectrometry (MS);
- nuclear magnetic resonance spectrometry.

A combination of chromatography and one or both of the other two methods also is an option. In Table 1, the various weak and strong points of these methods are summarized. Concerning the mentioned criteria, chromatographic methods have a major problem in reproducibility. Although the intra- and inter-day variability may be reasonable, there is always a problem of getting the same separation in another lab due to small changes in external conditions, chemicals, or solvents. Moreover, due to aging, column performance changes through time. Last but not least, commercial companies always come with improved columns, which usually also means an altered selectivity of the system. A simple search in the literature for the separation of, e.g., tropane alkaloids shows that every year a number of new “improved” separation

systems is described for these alkaloids. This is a poor basis for a long-term reproducible database of results of such an analysis. MS also has a problem of reproducibility, mainly because in MS analysis the compounds have to be ionized before measuring. There exist many different types of mass spectrometers with different ways to ionize the molecules, making comparison difficult. Besides the variation of the hardware and the conditions chosen, also a considerable matrix effect exists for the efficiency of the ionization, which means that a compound in different extracts may give different quantitative results.

NMR spectrometry is a physical measurement of the resonances of magnetic nuclei, such as  $^1\text{H}$ ,  $^{13}\text{C}$ , or  $^{15}\text{N}$  in a strong magnetic field. Each compound has a highly specific spectrum. The only variables are the solvent used and the magnetic field strength. The solvent is easy to standardize. The field strength is not a problem with  $^{13}\text{C}$  NMR and even in case of  $^1\text{H}$  NMR it can be overcome (see below). The high reproducibility makes that NMR among the three methods mentioned is the most suited for a public domain metabolomics database that may serve for many years to come.

Concerning the criteria of quantitation, the chromatographic methods have shown through the years to be very suitable for quantitative analysis. However, they always require calibration curves for the compounds that are quantified, as each compound will give a different detector response. Only in GC in combination with a flame ionization detector (FID) one may compare peak intensities with each other, and thus make an estimation of the absolute quantities of all compounds detected. UV-detection in HPLC, and MS in HPLC and GC will show totally different sensitivities for each compound, some even not being detected at all. For complex samples with a number of compounds, it is not feasible to make calibration curves for all compounds. That means only a relative comparison for each compound in different analyses is possible, but no absolute quantitation can be done, i.e., one cannot compare the amounts present of different compounds. This is a major disadvantage, e.g., one cannot make any conclusion about the conversion of one compound into another in case

**Table 1** Description of chromatographic methods used in metabolomics

Methods	Target metabolites	Sample preparation	Reproducibility	Resolution	Detector
TLC	General	Simple	Low	Low	UV, MS, color reagents
GC	Non-polar (low molecular)	Elaborate derivatization	Medium	High	FID, TCD, NPD, MS
HPLC	Polar (chromophore)	Elaborate	Medium	Medium	UV, RI, MS, ELSD, fluorescence, NMR
CE	Ionic	–	Medium	High	UV, MS

of a decrease of one compound and an increase in another related compound, i.e., conclusions about stoichiometry cannot be drawn from the data. For qualitative analysis the hyphenated chromatographic methods (HPLC with diode array detector, -MS, and/or NMR, GC-MS) are the most powerful as they offer both retention behavior and physical characteristics as a tool for identification. MS allows the determination of the molecular weight, and in case of high resolution also of the elemental composition, but this is not always sufficient to determine the structure. Tandem MS might be of help to identify in such case the compound through its fragmentation pattern, but in case of novel compounds further spectral data are required. The performance for NMR can be improved by using two-dimensional (2D) NMR methods (see below), which even may enable structure elucidation of novel compounds in a mixture. In fact  $^{13}\text{C}$  NMR spectra were already used in the early 1980s for fingerprinting essential oils (Formáček and Kubeczka 1982). We described the use of  $^1\text{H}$  NMR spectra for the fingerprinting of plant cell cultures (Schripsema and Verpoorte 1991). It has been used ever since for the analysis of sugars in cell cultures and media (e.g., Kraemer et al. 1999, 2002; Schripsema and Verpoorte 1991). Also other compounds have been analyzed in plant extracts by means of NMR, e.g., pyrrolizidine alkaloids (Pieters et al. 1989) and hop bitter agents (Hoek et al. 2001). With the increasing field strength of the NMR-magnets and the improved resolution the application of NMR for metabolomics became feasible.

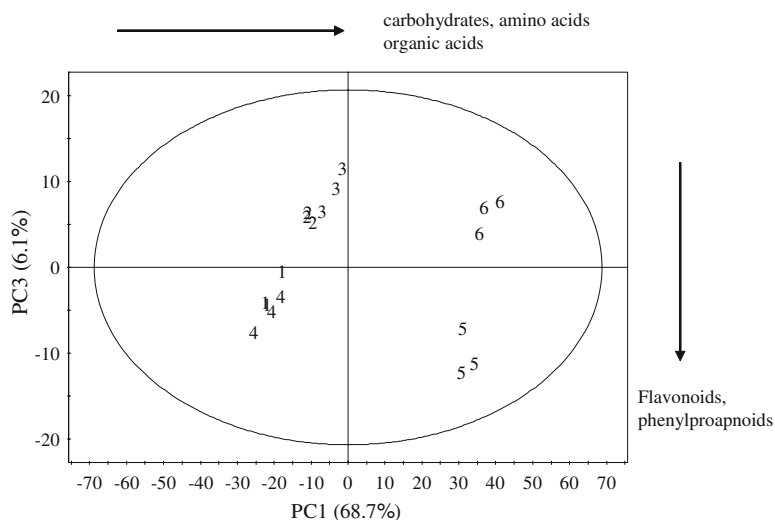
Here we will first discuss the general approach to metabolomics and the sample preparation methods

and show some examples of the use of NMR-metabolomics in phytochemical studies. The weak points and possible solutions will be discussed.

### Sample preparation

In metabolomics studies there are three phases. The first one is a qualitative phase in which as many as possible of the signals observed in the NMR are assigned to compounds. After this phase it is first necessary to determine the biological variation for the system studied, that means measuring a large number of samples, e.g., young and old leaves, different times of the day, and different stages of development of the plant. The last step is to compare the model system under different experimental conditions or compare, e.g., diseased or transgenic plants with the normal wild type plant. An important aspect is of course the extraction method. Plants contain a wide variety of compounds with totally different polarities. No single method is able to detect all metabolites in a single operation. To extract as many as possible we developed some years ago a two-phase extraction method in which chloroform-methanol-water (2:1:1) was used. This procedure extracts both polar and polar compounds in the same operation. This method gives good results, but is quite elaborate. Therefore, a series of single solvents were tested for the extraction of the metabolites from *Arabidopsis* (unpublished data, Fig. 1). By multivariate analysis, the spectra were compared in order to see which solvent covered the widest range of compounds. Methanol-water (1:1) came out as the best one in terms of the diversity (Choi et al. 2006; Hendrawati et al. 2006;

**Fig. 1** Score plot of PCA (PC1 vs PC3) based on  $^1\text{H}$  NMR spectra of the *A. thaliana* extracts. 1 Acetone, 2 acetone-MeOH (1:1), 3 MeOH, 4 AcCN, 5 MeOH-water (1:1), 6 water. The ellipse represents the Hotelling T2 with 95% confidence in score plots



Liang et al. 2006a, b; Widarto et al. 2006). By using these solvents in deuterated form, extracts can be measured directly after extraction, without any need of evaporation and reconstitution of the extract. To avoid shifts due to differences of pH (Schripsema et al. 1986) a buffered solvent ( $\text{KH}_2\text{PO}_4$ ) is used (pH 6.0).

### Transgenic tobacco plants overexpression microbial genes for salicylate biosynthesis

We have overexpressed the microbial genes encoding the isochorismate pathway leading to salicylate (SA). The genes, isochorismate synthase from *Escherichia*

*coli*, and isochorismate lyase from *Pseudomonas fluorescens*, result in the constitutive production of SA in tobacco plants. These plants have an increased resistance against viral and fungal infection (Verberne et al. 2000). This raised the question if any secondary metabolites pathways were altered, either due to the SA that is involved as signal compound in the systemic acquired resistance (SAR), or due to channeling of chorismate away from the other important chorismate pathways branches, leading among others to phenylalanine (and from there to all phenylpropanoids) and tryptophan. In a series of papers, various targeted approaches have been described, which led to the conclusion that terpenoids (by GC-analysis, Nugroho et al. 2002a), and alkaloids

**Table 2** Characteristic  $^1\text{H}$  chemical shifts of common metabolites in plants [ $\text{CD}_3\text{OD-KH}_2\text{PO}_4$  in  $\text{D}_2\text{O}$  (pH 6.0)]

Metabolites	Chemical shifts (ppm) and coupling constants (Hz)
Alanine	$\delta$ 1.48 (H-3, d, $J = 7.2$ Hz)
Ascorbic acid	$\delta$ 4.52 (H-4, d, $J = 2.0$ Hz)
Caffeic acid	$\delta$ 7.52 (H-7, d, $J = 15.9$ Hz), $\delta$ 7.12 (H-2', d, $J = 2.1$ Hz), $\delta$ 7.03 (H-6', d, $J = 8.4, 2.0$ Hz), $\delta$ 6.88 (H-5', d, $J = 8.4$ Hz), $\delta$ 6.29 (H-8', d, $J = 15.9$ Hz)
Chlorogenic acid	$\delta$ 7.60 (H-7', d, $J = 15.9$ Hz), $\delta$ 7.15 (H-2', d, $J = 2.1$ Hz), $\delta$ 7.05 (H-6', d, $J = 8.4, 2.0$ Hz), $\delta$ 6.87 (H-5', d, $J = 8.4$ Hz), $\delta$ 6.36 (H-8', d, $J = 15.9$ Hz), $\delta$ 5.33 (H-3, td, $J = 10.0, 4.8$ Hz), $\delta$ 4.21 (H-5, br q, $J = 3.1$ Hz), $\delta$ 3.80 (H-4, dd, $J = 9.9, 3.2$ Hz), $\delta$ 2.17 (H-2 $\beta$ , H-6 $\beta$ , m), $\delta$ 2.04 (H-2 $\alpha$ , H-6 $\alpha$ , m)
Citramalic acid <sup>a</sup>	$\delta$ 2.80 (H- $\beta$ , d, $J = 15.7$ Hz), $\delta$ 2.61 (H- $\beta'$ , d, $J = 15.7$ Hz)
Citric acid <sup>a</sup>	$\delta$ 2.74 (H- $\beta$ , d, $J = 17.6$ Hz), $\delta$ 2.56 (H- $\beta'$ , d, $J = 17.6$ Hz)
Fatty acids	$\delta$ 0.88 (H- $\omega$ , t, $J = 7.5$ Hz)
Ferulic acid	$\delta$ 7.56 (H-7', d, $J = 15.9$ Hz), $\delta$ 7.19 (H-2', d, $J = 2.1$ Hz), $\delta$ 7.10 (H-6', d, $J = 8.4, 2.0$ Hz), $\delta$ 6.88 (H-5', d, $J = 8.4$ Hz), $\delta$ 6.33 (H-8', d, $J = 15.9$ Hz)
Fumaric acid	$\delta$ 6.56 (s)
$\alpha$ -Glucose	$\delta$ 5.20 (H-1, d, $J = 3.8$ Hz)
$\beta$ -Glucose	$\delta$ 4.59 (H-1, d, $J = 7.9$ Hz)
GABA	$\delta$ 2.28 (H-2, d, $J = 7.5$ Hz), $\delta$ 1.88 (H-3, q, $J = 7.5$ Hz), $\delta$ 1.88 (H-4, t, $J = 7.5$ Hz)
Glutamic acid	$\delta$ 2.07 (H-3, m), $\delta$ 2.36 (H-4, m)
Glutamine	$\delta$ 2.15 (H-3, m), $\delta$ 2.47 (H-4m)
<i>p</i> -Hydroxy benzoic acid	$\delta$ 7.91 (H-3, H-5, d, $J = 8.8$ Hz), $\delta$ 6.92 (H-2, H-6, d, $J = 8.8$ Hz)
Inositol	$\delta$ 4.00 (H-2, t, $J = 2.8$ ), $\delta$ 3.61 (H-4 and 6, t, $J = 9.9$ ), $\delta$ 3.44 (H-1 and 3, dd, $J = 9.9, 2.9$ Hz), $\delta$ 3.24 (H-5, t, $J = 9.3$ Hz)
Isocitric acid <sup>a</sup>	$\delta$ 4.00 (H- $\alpha$ , d, $J = 4.2$ Hz), $\delta$ 3.04 (H- $\beta$ , td, $J = 11.7, 4.2$ Hz), $\delta$ 2.56 (H- $\gamma$ , dd, $J = 15.4, 7.3$ Hz), $\delta$ 2.56 (H- $\gamma'$ , dd, $J = 15.4, 7.7$ Hz)
Kaempferol	$\delta$ 8.04 (H-2' and H-6', d, $J = 8.6$ Hz), $\delta$ 6.74 (H-3' and H-5', d, $J = 8.6$ Hz), $\delta$ 6.52 (H-8, d, $J = 2.0$ Hz), $\delta$ 6.28 (H-6, d, $J = 2.0$ Hz)
$\alpha$ -Linolenic acids	$\delta$ 0.96 (H- $\omega$ , t, $J = 7.5$ Hz), $\delta$ 1.30 (CH <sub>2</sub> brs)
Malic acid <sup>a</sup>	$\delta$ 4.32 (H- $\alpha$ , dd, $J = 6.6, 4.7$ Hz), $\delta$ 2.80 (H- $\beta$ , dd, $J = 16.6, 4.7$ Hz), $\delta$ 2.61 (H- $\beta'$ , dd, $J = 16.6, 6.6$ Hz)
Proline	$\delta$ 4.08 (H-2, dd, $J = 8.6, 6.4$ Hz), $\delta$ 2.34 (H-3, m)
Quercetin	$\delta$ 6.99 (H-5', d, $J = 8.6$ Hz), $\delta$ 6.52 (H-8, d, $J = 1.8$ Hz), $\delta$ 6.28 (H-6, d, $J = 1.8$ Hz)
Rhamnose in flavonoid	$\delta$ 1.10 (3- <i>O</i> -rhamnosylglucoside, H-6, d, $J = 6.1$ Hz)
Sinapic acid	$\delta$ 7.48 (H-7, 16 Hz), $\delta$ 6.93 (H-2, H-6, s), $\delta$ 6.37 (H-8, 16 Hz)
Steroids	$\delta$ 0.78 (H-18, s), $\delta$ 0.85 (H-19, s), $\delta$ 0.89 (H-21, d, $J = 7.0$ Hz)
Succinic acid <sup>a</sup>	$\delta$ 2.54 (s)
Sucrose	$\delta$ 5.40 (H-1, d, $J = 3.8$ Hz), $\delta$ 4.17 (H-1, d, $J = 8.5$ Hz)
Threonine	$\delta$ 1.32 (H-5, d, $J = 6.6$ Hz)
Tryptophan	$\delta$ 7.20 (H-5', t, $J = 7.4$ Hz), $\delta$ 7.29 (H-6', t, $J = 7.5$ Hz), $\delta$ 7.32 (H-2', s), $\delta$ 7.54 (H-7', d, $J = 8.1$ Hz), $\delta$ 7.73 (H-4', d, $J = 7.9$ Hz)
Tyrosine	$\delta$ 7.19 (H-2, H-6, t, $J = 8.5$ Hz), $\delta$ 6.86 (H-3, H-5, t, $J = 7.5$ Hz)
Valine	$\delta$ 1.00 (H-3a, d, $J = 6.8$ Hz), 1.05 (H-3b, d, $J = 6.8$ Hz)

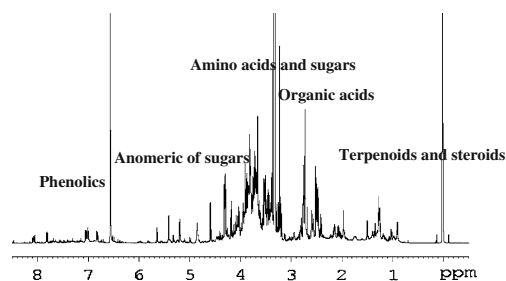
<sup>a</sup> Changeable by pH and concentration

(by HPLC-analysis, Nugroho et al. 2002b) were not effected, whereas flavonoids (by HPLC-analysis, Nugroho et al. 2002c) showed lower levels, similar conclusions were drawn from a non-targeted approach using centrifugal partitioning chromatography to prefractionate the extract followed by LC-MS to analyze the fractions (Halim et al. 2003). This was a major effort as each of the different groups of compounds required the validation of a suited method.

The next step was to use this model system for an NMR-based metabolomics approach, to learn how this approach compared with the elaborate targeted approach.

The first step was the qualitative analysis. A number of compounds could easily be identified. To facilitate the identification we started to build a database with NMR spectra of common plant metabolites. This database contains now more than 300 compounds, and more than 600 spectra (both 1D and 2D spectra). Some of the very common diagnostics signals are summarized in Table 2. They represent, e.g., phenolics, the anomeric protons of sugars, amino acids, and organic acids. In each area of the NMR, one may look for specific groups of compounds (Fig. 2).

Using this method extensive studies have been made of the metabolomes of the tobacco plants (Choi et al. 2004a, 2006). We used the two-phase or deuterated solvent extraction method, and measured the NMR spectra. For analyzing the large data set of NMR spectra, multivariate analysis (using SIMCA-P, Umetrics, Umeå, Sweden) was used. From the results we were able to clearly distinguish these in wild type and transgenic plants. To introduce the NMR spectra into the biostatistics program, they have to be processed into digital quantifiable data. This is done by dividing the spectra into buckets (or bins) of 8 or 16 Hz, peak height in each bucket is integrated and introduced into a text file used for further statistical analysis. This processing can be done by the NMR-software (e.g., AMIX, Bruker Biospin GmbH, Rheinstetten, Germany). For robustness of the bucketing high reproducibility of chemical shift is required. By using a buffered NMR-solvent, the reproducible chemical shift



**Fig. 2** Typical  $^1\text{H}$  spectra of *Arabidopsis* [400 MHz,  $\text{CD}_3\text{OD}-\text{KH}_2\text{PO}_4$  in  $\text{D}_2\text{O}$  (pH 6.0)] (adapted from Hendrawati et al. 2006)

can be obtained. However, there are still problem with organic acids such as citric, citramalic, isocitric, malic, and succinic acid. The chemical shift of these signals may vary considerably depending on pH of the solution. In order to solve the problem the region in which they occur can be bucketed by larger size (e.g., 40 Hz). However, one should always be aware of the risk that a compound may show changes in its NMR spectra, e.g., due to complexity of the mixture, or due to concentration effects (Rhee et al. 2004). Table 3 shows the example of the variation of chemical shifts of citric acid and malic acid in different concentration.

The signals responsible for the separation found in the score plot of principle component analysis (PCA) were identified from the loading plot (Fig. 3). Assigning these signals to various known tobacco constituents resulted in the conclusion that the transgenic plants had lower levels of flavonoids, chlorogenic acid, and sugars, and higher levels of malic acid and alanine (Choi et al. 2004a). The alkaloid and terpenoid levels were not changed. Similar conclusions as for the targeted approach (see above), but as extra information the increased levels of sugars and alanine were observed in this approach. This nicely showed the great potential of studying metabolic changes in a single NMR-analysis.

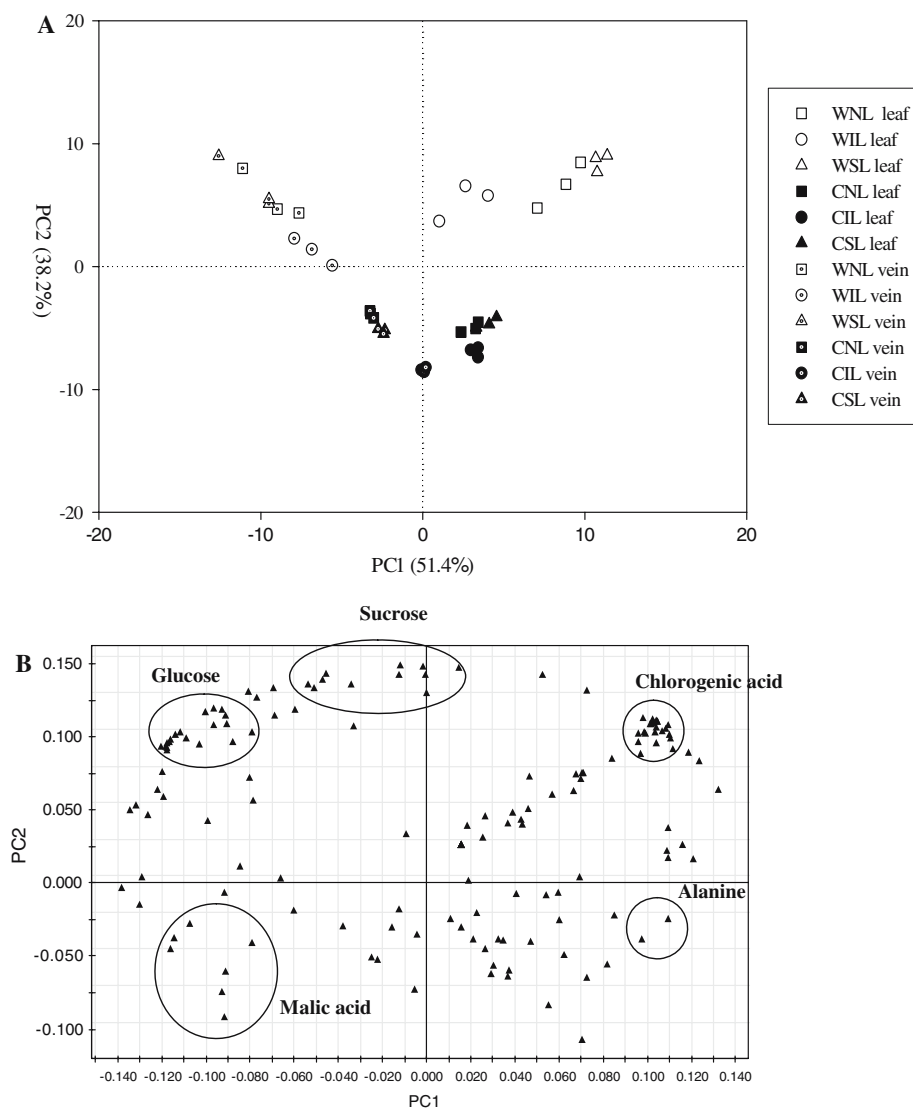
### Phytoplasma infected *Catharanthus roseus*

The same approach as for tobacco was applied to *Catharanthus roseus* plants infected with different

**Table 3**  $^1\text{H}$  chemical shifts of H- $\beta$  of citric and malic acids in different concentrations [ $\text{CD}_3\text{OD}-\text{KH}_2\text{PO}_4$  in  $\text{D}_2\text{O}$  (pH 6.0)]

Concentration (mg/mL)	Citric acid		Malic acid	
	H- $\beta$	H- $\beta'$	H- $\beta$	H- $\beta'$
1	$\delta$ 2.73 (d, $J$ = 15.0 Hz)	$\delta$ 2.62 (d, $J$ = 15.0 Hz)	$\delta$ 2.68 (dd, $J$ = 14.0, 3.3 Hz)	$\delta$ 2.37 (dd, $J$ = 14.0, 10.0 Hz)
5	$\delta$ 2.89 (d, $J$ = 15.0 Hz)	$\delta$ 2.77 (d, $J$ = 15.0 Hz)	$\delta$ 2.84 (dd, $J$ = 15.0, 5.0 Hz)	$\delta$ 2.67 (dd, $J$ = 15.0, 7.0 Hz)
10	$\delta$ 2.87 (d, $J$ = 13.3 Hz)	$\delta$ 2.74 (d, $J$ = 13.3 Hz)	$\delta$ 2.85 (dd, $J$ = 15.0, 5.0 Hz)	$\delta$ 2.72 (dd, $J$ = 15.0, 7.0 Hz)

**Fig. 3** Score (a) and loading plot (b) of methanol–water fraction for wild type and CSA-line #16 leaves and veins following PCA (PC1 vs PC2). *WNL* wild type non-inoculated leaf, *WIL* wild type inoculated leaf, *WSL* wild type systemic leaf, *CNL* CSA non-inoculated leaf, *CIL* CSA inoculated leaf, *CSL* CSA systemic leaf, *WNV* wild type non-inoculated vein, *WIV* wild type inoculated vein, *WSV* wild type systemic vein, *CNV* CSA non-inoculated vein, *CIV* CSA inoculated vein, *CSV* CSA systemic vein (adapted from Choi et al. 2004a)



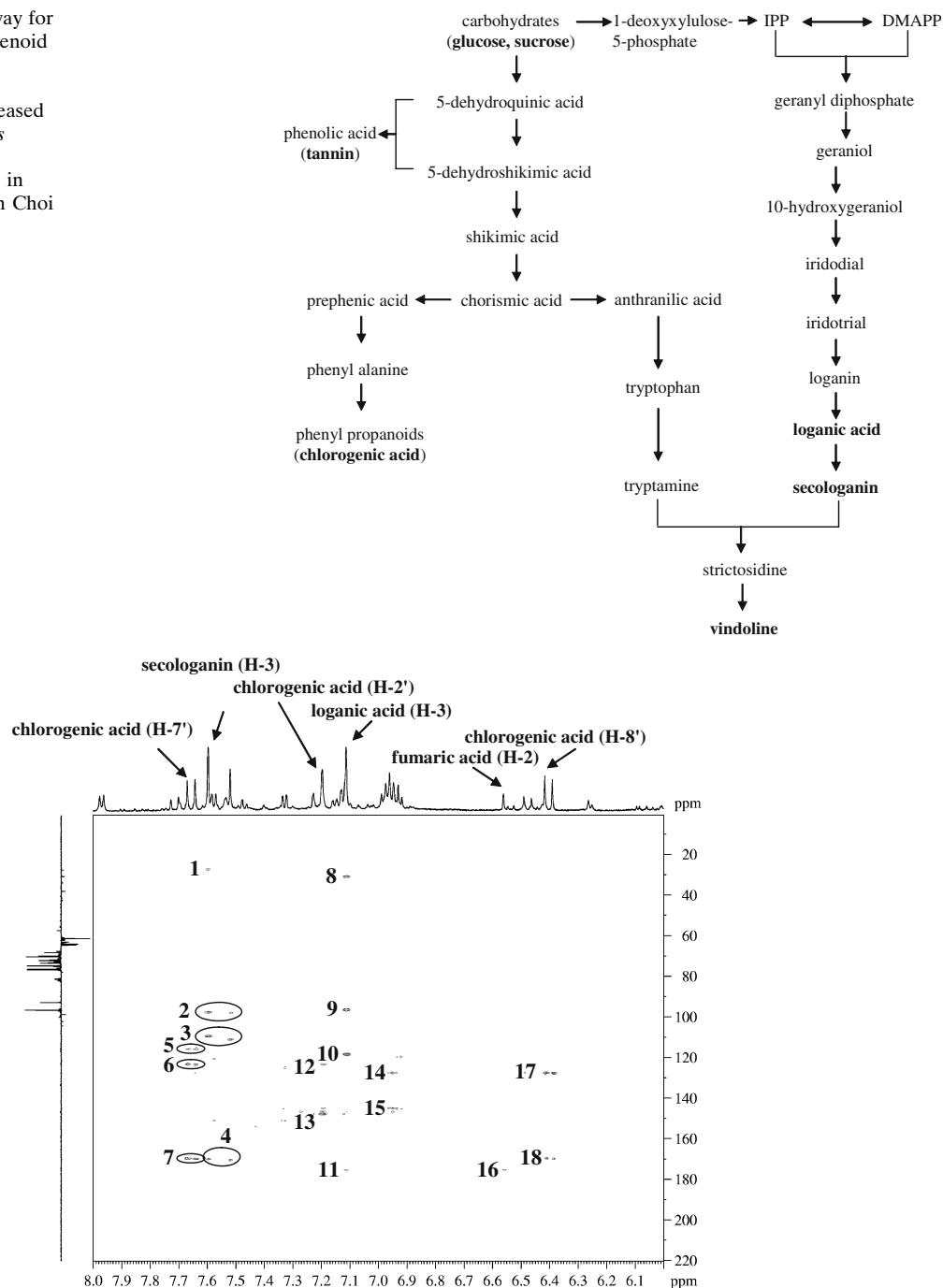
types of phytoplasmas (Choi et al. 2004b). Again a clear separation could be obtained for the plants, analysis of the markers for the differences led to the conclusions as shown in Fig. 4. Also, in this case, the macroscopic view on the metabolome via the NMR-analysis clearly showed major changes in the infected plants. In this case, a considerable effort was made to identify a number of compounds using 2D NMR spectrometry, this among others allowed the identification of the iridoids (see Fig. 5).

### Brassicaceae NMR-based metabolomics

From the results with the tobacco and *C. roseus* it became clear that biostatistical software (e.g., SIMCA-P)

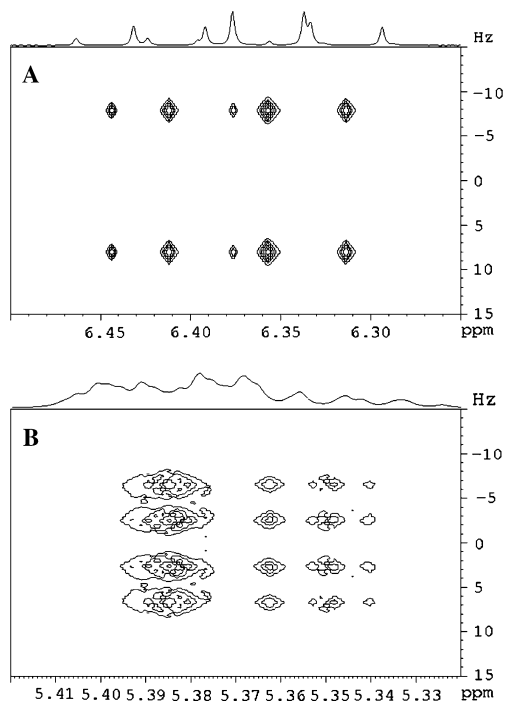
especially developed for metabolomics-like applications might be an advantage. Also, we learned that in general the two-phase extraction system was quite elaborate and thus not the preferred one for a high-throughput systems. Moreover, the chloroform phase mostly only showed strong signals due to fatty acids, which can be quantified as group, but individual acids cannot be observed in the 1D NMR spectra. On the other hand, the methanol–water phase gave an excellent overview of the plant's metabolome, and particularly on a broad range of secondary metabolites. A direct extraction with deuterated methanol–water solvent thus seems to be most suited (unpublished data). For the identification of compounds further 2D NMR methods were explored to identify compounds in the crude extracts. Also, the problem of comparing spectra

**Fig. 4** Schematic pathway for the biosynthesis of terpenoid indole alkaloid, phenylpropanoid, and phenolic acid. The increased metabolites in *C. roseus* leaves infected by phytoplasma are shown in **bold font** (adapted from Choi et al. 2004b)



**Fig. 5** HMBC spectra of water fraction of phytoplasma (UDI-NESE) infected *C. roseus* leaves. In HMBC spectra (C), **1** correlation of H-3 and C-5 of secologanin, **2** correlation of H-3 and C-1 of secologanin, **3** correlation of H-3 and C-4 of secologanin, **4** correlation of H-3 and carbonyl group of secologanin, **5** correlation of H-7' and C-2' of chlorogenic acid, **6** correlation of H-7' and C-6' of chlorogenic acid, **7** correlation of H-7' and carbonyl group of chlorogenic acid, **8** correlation of H-3 and C-5 of loganic acid, **9** correlation of H-3 and C-1 of

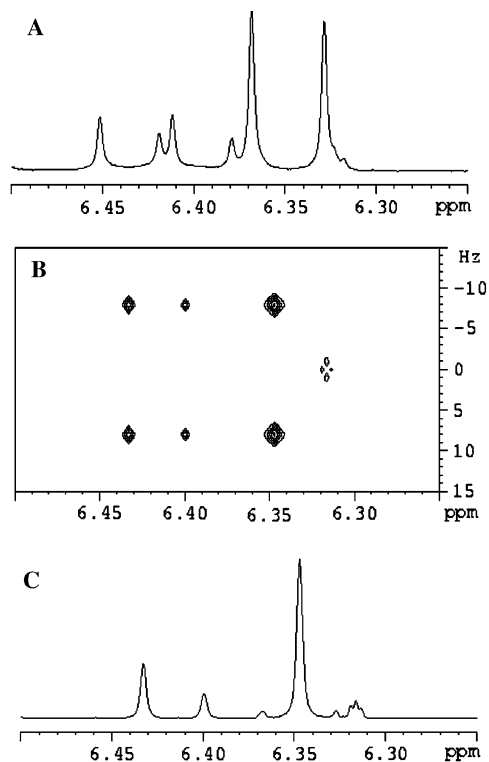
loganic acid, **10** correlation of H-3 and C-4 of loganic acid, **11** correlation of correlation of H-3 and carbonyl group of loganic acid, **12** correlation of H-2' and C-1' of chlorogenic acid, **13** correlation of H-2' and C-3', **14** correlation of H-2 and C-1 of gallic acid derivatives, **15** correlation of H-2 and C-3 of gallic acid derivatives, **16** correlation of H-2 and carbonyl group of fumaric acid, **17** correlation of H-8' and C-1' of chlorogenic acid, **18** correlation of H-8' and carbonyl group of chlorogenic acid (adapted from Choi et al. 2004b)



**Fig. 6** Two-dimensional J-resolved NMR spectra of phenylpropanoids of *B. rapa* leaves in the range of  $\delta$  6.25–6.50 (a);  $\delta$  5.32–5.42 (b). The spectra were measured in methanol- $d_4$  (adapted from Liang et al. 2006a)

obtained at different field strength needed to be solved. These matters were addressed in a series of experiments with tobacco, *Brassica rapa* and *Arabidopsis thaliana*.

One of the major problems of NMR-metabolomics is the overlap of many signals, particularly in some crowded regions. The use of 2D NMR spectrometry can improve this considerably. In 2D J-resolved spectra, the second dimension gives for each signal the coupling constant (Fig. 6a, Liang et al. 2006a). In this way, the congestion is to a great extent solved. Moreover, the coupling constant is an excellent diagnostic tool for the classification of signals as belonging to a certain class of compounds (e.g., cinnamic acid derivatives, Fig. 6a) (Liang et al. 2006a). Moreover, by projecting the signals of 2D J-resolved spectra on the F2 axis (the chemical shift), each signal is reduced to a singlet, i.e., each proton is now observed as a singlet (Fig. 7). The spectra are comparable in appearance to  $^{13}\text{C}$  NMR spectra. This simplifies the spectra enormously, as well as the identification of known compounds as the overlap of signals is much less. The most important aspect of these projected 2D J-resolved spectra is that all secondary effects are eliminated, consequently the spectra are independent of the field

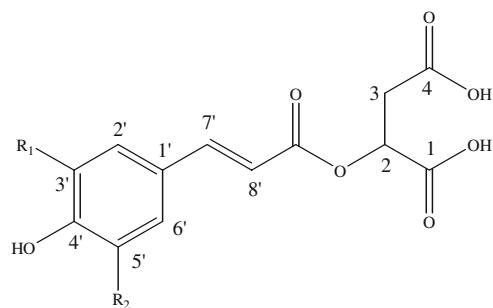


**Fig. 7**  $^1\text{H}$  NMR (a), 2D J-resolved (b), and projected 1D J-resolved (c) spectrum of healthy *Nicotiana tabacum* leaves in the range of  $\delta$  6.25–6.50 (adapted from Choi et al. 2006)

strength (Choi et al. 2006). This makes them to excellent candidates for storage in a large database for long-term use in metabolomics. The disadvantage of these spectra is that because of the projection the intensity of the signals of the different protons cannot be compared anymore. For absolute quantitation, the intensities of the different peaks in the second dimension need to be added.

For the identification of compounds 2D NMR spectrometry methods are of great value. Signals can be identified as well as their coupling with other protons or carbons. Figure 6 shows part of the spectrum of *B. rapa* in which from the 2D J-resolved spectrum it was clear that the congested signals at  $\delta$  6.3–6.5 were due to cinnamic acid derivatives as they all have protons with a large coupling constant (16 Hz, Liang et al. 2006a). A single separation step of the extract was made to have an enriched phenolic compound fraction, and thus reduce the NMR time needed to run the various 2D NMR spectra. Further 2D NMR spectra were used to elucidate their structures. From the COSY and HMBC spectra, it can be learned that malic acid moiety attaches to carboxylic group of





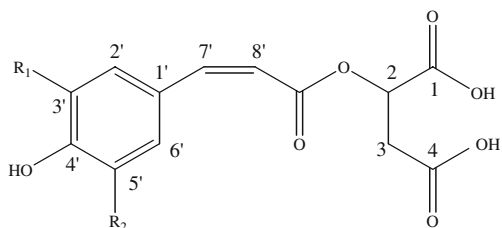
$R_1 = \text{OCH}_3, R_2 = \text{OH}$ , *trans*-5-hydroxyferuloyl malate

$R_1 = \text{OH}, R_2 = \text{H}$ , *trans*-caffeoyl malate

$R_1 = \text{H}, R_2 = \text{H}$ , *trans*-coumaroyl malate

$R_1 = \text{OCH}_3, R_2 = \text{H}$ , *trans*-5-hydroxyferuloyl malate

$R_1 = \text{OCH}_3, R_2 = \text{OCH}_3$ , *trans*-5-sinapoyl malate



$R_1 = \text{OCH}_3, R_2 = \text{OH}$ , *cis*-5-hydroxyferuloyl malate

$R_1 = \text{OH}, R_2 = \text{H}$ , *cis*-caffeoyl malate

$R_1 = \text{H}, R_2 = \text{H}$ , *cis*-coumaroyl malate

$R_1 = \text{OCH}_3, R_2 = \text{H}$ , *cis*-5-hydroxyferuloyl malate

$R_1 = \text{OCH}_3, R_2 = \text{OCH}_3$ , *cis*-5-sinapoyl malate

**Fig. 8** Chemical structures of phenylpropanoids identified in *B. rapa* leaves

phenylpropanoids. The presence of the attached H-2 of malic acid was confirmed by 2D J-resolved spectra (Fig. 6b). From the various 2D NMR experiments, a whole series of cinnamic acid derivatives was identified as shown in Fig. 8, both *cis*- and *trans*-isomers were found (Liang et al. 2006a). Whether the *cis*-derivatives are artifacts or already formed in the plant remains to be determined.

Using the NMR-based metabolomics approach we studied the effects of a methyl jasmonate treatment of *B. rapa* plants (Liang et al. 2006b). The score plot of the PCA (Fig. 9) shows a clear separation between the treated and non-treated plants. From the loading plot, it can be concluded that the phenylpropanoids discussed above increased in the treated plants, also an

indole glucosinolate such as neoglucobrassicin and indole acetic acid increased in the treated plants.

### Virus infected tobacco

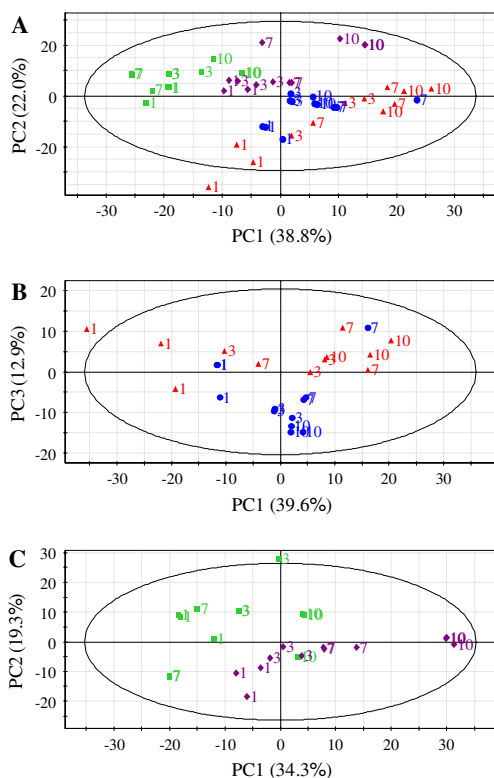
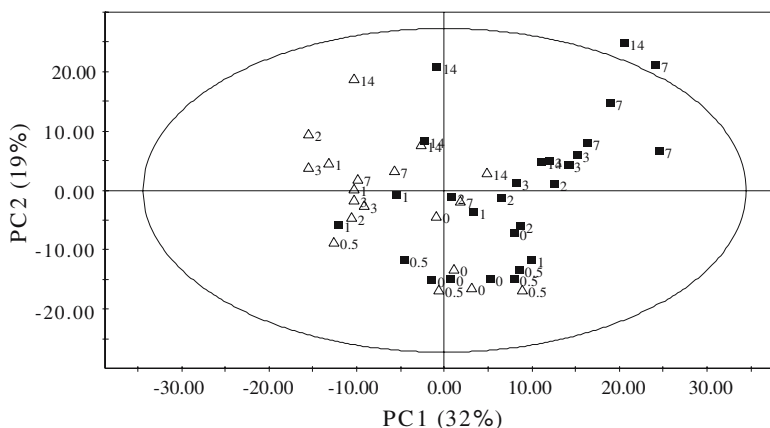
The various tools we developed since our first tobacco experiments were applied on a study of tobacco mosaic virus (TMV) infected tobacco plants. This system has been studied extensively on the level of genes and proteins to clarify the SAR in plants. As already discussed above SA does play an important role in SAR. Interestingly very little work has been done on the role of secondary metabolism in SAR, as still the function of several of the pathogen-related proteins is unknown, they might play a role in secondary metabolism. We thus applied our NMR-metabolomics platform technology on this system (Choi et al. 2006). In Fig. 10, the PCA score plot is shown. As expected the metabolome of the leaf changes through time, similar changes are observed in the infected leaves (mainly in PC1), but on top of that in PC2 or PC3 clear differences between leaves from the infected and control plants can be observed.

By using the various 2D NMR methods, a number of signals that are clearly involved in the SAR response and direct defense against the infection could be identified (see above). In Fig. 11, the various changes are summarized. Apparently some of the effects occur in the infected leaves and SAR leaves, whereas others only in the SAR or the infected leaves. The meaning of the various changes need of course further study. But obviously this systems biology type of approach using NMR-based metabolomics, gives many novel leads for further study.

### NMR-based metabolomics platform

These various examples nicely show the great potential of NMR as a tool for having a macroscopic view on the plant metabolome. It is thus of great interest in functional genomics studies as well as in systems biology studies on all kind of biological processes. But such a platform has also many other fields of application, such as quality control of botanicals and food (Choi et al. 2005; Kim et al. 2005; Yang et al. 2006); phenotyping of plants for breeders rights (Fernie et al. 2004; Fiehn et al. 2000; Fukusaki and Kobayashi 2005; Roessner et al. 2002); studies on the compounds related to pharmacological activity in medicinal plants (Joshi et al. 2002); dereplication of plant extracts in bioprospecting

**Fig. 9** Score plot of PCA of J-resolved NMR data of control ( $\Delta$ ) and methyl jasmonate treated ( $\blacksquare$ ) *B. rapa* leaves. The number after symbol shows the time (days) after EtOH treatment as control (adapted from Liang et al. 2006b)

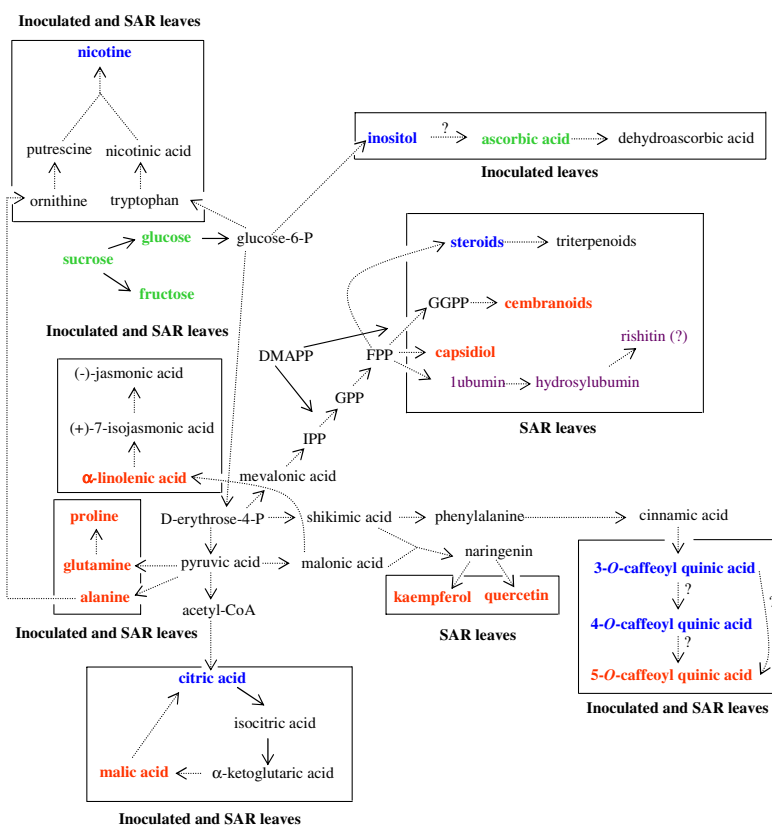


**Fig. 10** Score plot of PCA of healthy and infected *Nicotiana tabacum* leaves by TMV.  $\bullet$  lower leaves of healthy plants,  $\blacksquare$  upper leaves of healthy plants,  $\blacktriangle$  local-infected leaves of TMV-infected plants,  $\blacklozenge$  SAR leaves of TMV-infected plants. **a** PCA for lower and upper leaves in healthy plants and inoculated and SAR leaves, **b** PCA for lower leaves in healthy plants and inoculated leaves in the infected plants, **c** PCA for upper leaves in healthy plants and SAR leaves in the infected plants. The ellipse represents the Hotelling T2 with 95% confidence in score plots. Numberings on the plot are the dates after infection (adapted from Choi et al. 2006)

screening programs (Nord et al. 2001); equivalence of genetically modified plants with non-GM plants (Fernie et al. 2004; Manetti et al. 2004; Noteborn et al. 2000; Roessner et al. 2001, 2002).

The great success of the NMR-based metabolomic studies on urine for, e.g., diagnosis of diseases (Bollard et al. 2005; Kochhar et al. 2006; Lindon et al. 2003) and studying toxicity of compounds in animal models (Gavaghan et al. 2000, 2002; Potts et al. 2001) can serve as a good example how plant metabolomics may further develop. Analyzing the much more complex metabolome of plants if compared with the various body fluids such as urine and serum, will be a major challenge for the coming years. Though the number of compounds that are detected in an extract is limited, still the NMR spectra give a good picture of what really is present. Minor compounds might not be seen, but the major trends are clear. By combining the NMR-metabolomic data with a more targeted approach (LC-MS, GC-MS) for those compounds that seem to be of interest one would have the broadest possible information on the metabolome. It can be compared with a microscope zooming in on smaller units. The NMR gives us the general overall view, the macroscopic view. Of course the question could be raised how many compounds can be seen in NMR. The answer is many, but the number of compounds actually present in an extract is much smaller. In fact not all the common compounds we have in our database are detected in all plant extracts. By using both 1D and 2D NMR spectrometry, one may say that about 50–100 compounds can be detected, and the absence (or below threshold) of another 50–100 can also be concluded. For the further development of the NMR platform first of all a database with all common compounds is required. But

**Fig. 11** Proposed metabolomic alterations in the *Nicotiana tabacum* leaves infected by TMV. ● decreased, ● increased, ● transient increased, ● previous results from other *Nicotiana* species (e.g., *N. undulata*, or *N. rustica*, *N. glutinosa*), ● based on general plant biosynthesis (adapted from Choi et al. 2006)



that also requires agreement on the solvents used for the NMR-metabolomics. In our experience, buffered methanol–water (1:1) is an excellent solvent for studying a broad range of primary and secondary metabolites. This might be further completed by a non-polar solvent, though the compounds found in such extracts are usually lipids and terpenoids. Lipids can be identified as total groups, but not as individual compounds by means of 1D NMR. For terpenoids, the possibilities for identification of individual compounds are better.

Our present database of more than 300 common compounds of plants showed to be very useful for identification of compounds in the NMR spectra. A further computerized identification using the projected 2D J-resolved spectra would be an important improvement.

## References

- Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E (2005) NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed* 18:143–162
- Choi HK, Choi YH, Verberne M, Lefeber AWM, Erkelens C, Verpoorte R (2004a) Metabolic fingerprinting of wild type and transgenic tobacco plants by  $^1\text{H}$  NMR and multivariate analysis technique. *Phytochemistry* 65:857–864
- Choi YH, Tapias EC, Kim HK, Lefeber AWM, Erkelens C, Verhoeven JJJ, Brzin J, Zel J, Verpoorte R (2004b) Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using  $^1\text{H}$  NMR spectroscopy and multivariate data analysis. *Plant Physiol* 135:2398–2410
- Choi YH, Sertic S, Kim HK, Wilson EG, Michopoulos F, Lefeber AWM, Erkelens C, Kricun SDP, Verpoorte R (2005) Classification of *Ilex* species based on metabolomic fingerprinting using NMR and multivariate data analysis. *J Agric Food Chem* 53:1237–1245
- Choi YH, Kim HK, Linthorst HJM, Hollander JG, Lefeber AWM, Erkelens C, Nuzillard J-M, Verpoorte R (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. *J Nat Prod* 69:742–748
- Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* 5:763–769
- Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolic profiling for plant functional genomics. *Nat Biotechnol* 18:1157–1161
- Formáček V, Kubezka K-H (1982) Essential oils analysis by capillary gas chromatography and carbon-13 NMR spectroscopy. Wiley, Chichester
- Fukusaki E, Kobayashi A (2005) Plant metabolomics: potential for practical operation. *J Biosci Bioeng* 100:347–354

- Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk: ApfCD mouse. *FEBS Lett* 484:169–174
- Gavaghan CL, Wilson ID, Nicholson JK (2002) Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correlation and PLS-DA. *FEBS Lett* 530:191–196
- Halim VA, Verberne MC, Verpoorte R (2003) Differential metabolic profiling by HPLC-PDA-MS of wild type and transgenic tobacco plants constitutively producing salicylic acid. *J Liq Chromatogr Relat Technol* 26:369–383
- Hendrawati O, Yao Q, Kim HK, Linthorst HJM, Erkelens C, Lefeber AWM, Choi YH, Verpoorte R (2006) Metabolic differentiation of *Arabidopsis* treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci* 170:1118–1124
- Hoek AC, Hermans-Lokkerbol ACJ, Verpoorte R (2001) An improved NMR method for the quantification of  $\alpha$ -acids in hops and hop products. *Phytochem Anal* 12:53–57
- Joshi L, Van Eck JM, Mayo K, Di Silvestro R, Blake NME, Ganapathi T, Haridas V, Gutterman JU, Arntzen CJ (2002) Metabolomics of plant saponins: bioprospecting triterpene glycoside diversity with respect to mammalian cell targets. *OMICS* 6:235–246
- Kim HK, Choi YH, Erkelens C, Lefeber AWM, Verpoorte R (2005) Metabolic fingerprinting of *Ephedra* species using  $^1\text{H}$ -NMR spectroscopy and principal component analysis. *Chem Pharm Bull* 53:105–109
- Kochhar S, Jacobs DM, Ramadan Z, Berruex F, Fuerholz A, Fay LB (2006) Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabolomics. *Anal Biochem* 352:274–281
- Kraemer KH, Schenkel EP, Verpoorte R (1999) Glucosylation of ethanol in *Ilex paraguariensis* cell suspension cultures. *Plant Cell Rep* 18:509–513
- Kraemer KH, Stival AL, Ferreira AG, Schenkel EP, Verpoorte R (2002) *Ilex paraguariensis* cell suspension culture characterization and response against ethanol. *Plant Cell Tissue Organ Cult* 68:257–263
- Liang Y-S, Lefeber AWM, Erkelens C, Kim HK, Choi YH, Verpoorte R (2006a) Identification of phenylpropanoids in *Brassica rapa* leaves treated with methyl jasmonate using two-dimensional nuclear magnetic resonance spectroscopy. *J Chromatogr A* 1112:148–155
- Liang Y-S, Choi YH, Kim HK, Linthorst HJM, Verpoorte R (2006b) Metabolomic analysis of methyl jasmonate treated *Brassica rapa* leaves by two dimensional NMR spectroscopy and multivariate analysis. *Phytochemistry* (in press)
- Lindon JC, Nicholson JK, Holmes E, Antti H, Bollard ME, Keun H, Beckonert O, Ebbels TM, Reilly MD, Robertson D, Stevens GJ, Luke P, Breaud AP, Cantor GH, Bible RH, Niederhauser U, Senn H, Schlotterbeck G, Sidelmann UG, Laursen SM, Tymiak A, Car BD, Lehman-McKeeman L, Colet JM, Loukaci A, Thomas C (2003) Contemporary issues in toxicology the role of metabolomics in toxicology and its evaluation by the COMET project. *Toxicol Appl Pharmacol* 187:137–146
- Manetti C, Bianchetti C, Bizzarri M, Casciani L, Castro C, D'Ascenzo G, Delfini M, Di Cocco ME, Lagana A, Miccheli A, Motto M, Conti F (2004) NMR-based metabonomic study of transgenic maize. *Phytochemistry* 65:3187–3198
- Nord LI, Kenne L, Jacobsson SP (2001) Multivariate analysis of  $^1\text{H}$  NMR spectra for saponins from *Quillaja saponaria* Molina. *Anal Chim Acta* 446:199–209
- Noteborn HPJM, Lommen A, Van der Jagt RC, Weseman JM (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J Biotechnol* 77:103–114
- Nugroho LH, Peltenburg-Looman AMG, Verberne MC, Verpoorte R (2002a) Is accumulation of sesquiterpenoid phytoalexins induced in tobacco plants constitutively producing salicylic acid? *Plant Sci* 162:989–993
- Nugroho LH, Peltenburg-Looman AMG, De Vos H, Verberne MC, Verpoorte R (2002b) Nicotine and related alkaloids accumulation in constitutive salicylic acid producing tobacco plants. *Plant Sci* 162:575–581
- Nugroho LH, Verberne MC, Verpoorte R (2002c) Activities of enzymes involved in the phenylpropanoid pathway in constitutive salicylic acid producing transgenic tobacco plants. *Plant Physiol Biochem* 40:755–760
- Pieters LA, Hartmann T, Janssens J, Vlietinck AJ (1989) Comparison of capillary gas chromatography with  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance spectroscopy for the quantitation of pyrrolizidine alkaloids from *Senecio vernalis*. *J Chromatogr* 462:287–391
- Potts BCM, Deese AJ, Stevens GJ, Reilly MD, Robertson DG, Theiss J (2001) NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse. *J Pharm Biomed Anal* 26:463–476
- Rhee IK, Appels N, Hofte B, Karabatak B, Erkelens C, Stark LM, Flippin LA, Verpoorte R (2004) Isolation of the acetylcholinesterase inhibitor ungeremine from *Nerine bowdenii* isolated by preparative HPLC coupled on-line to a flow assay system. *Biol Pharm Bull* 27:1804–1809
- Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13:11–29
- Roessner U, Willmitzer L, Fernie AR (2002) Metabolic profiling and biochemical phenotyping of plant systems. *Plant Cell Rep* 21:189–196
- Schripsema J, Verpoorte R (1991) Investigation of extracts of plant cell cultures by  $^1\text{H}$ -NMR. *Phytochem Anal* 2:155–162
- Schripsema J, Verpoorte R, Svendsen AB (1986) Trifluoroacetic acid. A  $^1\text{H}$  NMR shift reagent for alkaloids. *Tetrahedron Lett* 27:2523–2526
- Verberne M, Verpoorte R, Bol J, Mercado-Blanco J, Linthorst HJM (2000) Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nat Biotechnol* 18:779–783
- Widarto HT, Van der Meijden E, Lefeber AWM, Erkelens C, Kim HK, Choi YH, Verpoorte R (2006) Metabolomic differentiation of *Brassica rapa* leaves attacked by herbivore using two dimensional nuclear magnetic resonance spectroscopy. *J Chem Ecol* (in press)
- Yang SY, Kim HK, Lefeber AWM, Erkelens C, Angelova N, Choi YH, Verpoorte R (2006) Application of two dimensional nuclear magnetic resonance spectroscopy to quality control of Ginseng commercial products. *Planta Med* 72:364–369