Metabolomic analysis of Catharanthus roseus using NMR and principal component analysis

Kim, H.K.; Choi, Y.H.; Verpoorte, R.; Saito, K.; Dixon, R.; Willmitzer, L.

III.6 Metabolomic Analysis of *Catharanthus roseus* Using NMR and Principal Component Analysis

H.K. Kim, Y.H. Choi, and R. Verpoorte[1]

1 Introduction

The ultimate goal of plant metabolomics is to map all metabolites in a plant both qualitatively and quantitatively. Detection of all plant metabolites seems impossible due to the large number, the chemical complexity of the metabolites and their different characteristics such as solubility and polarity. A proper analytical method should be selected in order to be able to detect as many compounds as possible in a plant. A number of analytical methods have been proposed and applied to profile the plant metabolome (Sumner et al. 2003). Basically, two types of methods can be distinguished – methods based on a chromatographic separation, e. g. HPLC, GC and TLC, and methods based on a physical characteristic of the metabolites, e. g. MS (molecular weight) and NMR (resonance of magnetic nuclei, e. g. ^{1}H or ^{13}C in a strong magnetic field). To obtain maximum selectivity, both methods can also be combined. Chromatographic methods are based on the relative behavior of the individual metabolites in a system with a mobile phase (gas or liquid) and a stationary phase. This allows a selective separation; however, reproducibility is very much dependant on the quality of both phases. For the mobile phase this is reasonable; however, for the stationary phase this is different. Many different stationary phases exist for GC and HPLC, and "improved" stationary phase are regularly introduced. Long term reproducibility is thus difficult. With respect to the requirements of speed and reproducibility, nuclear magnetic resonance (NMR) and mass spectrometry (MS) based approaches score very well if compared to chromatography. The MS based metabolomic analysis shows high separation efficiency and sensitivity, and easy coupling with chromatographic methods. These characteristics of MS analysis in the plant metabolomics allow the detection of a larger number of metabolites if compared to NMR (e. g. 3000 metabolites) (Aharoni et al. 2002; Fiehn 2002). However, there are some inevitable limitations in the MS based methods. These are mainly in terms of quantitation. Each compound will show different sensitivity which may also be different by the matrix in which it is analyzed. For absolute quantitation calibration curves are needed for each single compound. For relative occurrence of a certain compound in different materials this is of no importance. The range of metabolites covered by gas chromatography (GC)-MS is restricted more or

[1] Division of Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, P.O. Box 9502, 2300RA, Leiden, The Netherlands, e-mail: verpoort@chem.leidenuniv.nl

Biotechnology in Agriculture and Forestry, Vol. 57
Plant Metabolomics (ed. by K. Saito, R.A. Dixon, and L. Willmitzer)

less to small and volatile metabolites having a molecular weight less than 400 (e. g. mono or disaccharides, amino acids, or organic acids). Consequently, this excludes the detection of unstable and non-volatile plant secondary metabolites such as glycosides. MS can be coupled with high performance liquid chromatography (HPLC) using soft ionization methods such as electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI) (Huhman and Sumner 2002; Tolstilkov and Fiehn 2002). It is suited for all kinds of metabolites and can also detect polar or high molecular weight metabolites; however, lack of fragmentation makes it difficult to identify the metabolites – an MS/MS system is required to obtain structural information.

Despite its low sensitivity, the recent advances in NMR methods offer some advantages compared to chromatography and MS methods. The range of compounds is not limited by their volatility, presence of chromophores or polarity. The broad range of metabolites detected by NMR makes it the optimum choice for macroscopic metabolomics, a total representative view of all metabolites present. Moreover, NMR has the great advantage that the spectra are highly reproducible as it concerns a physical characteristic of a compound. In other words, the NMR data obtained at different places or time can be compared with each other. It is also possible to elucidate the structure of unknown metabolites, particularly secondary metabolites. Another advantage is that the signals of NMR spectra are based on molar concentration and can directly be compared while the intensity of metabolites in MS is highly affected by the ionization level.

Despite the advantages of NMR in plant metabolomics, the spectral complexity, lower sensitivity and costly combination with chromatography have made researchers hesitant to apply it as a tool of metabolomics. Most applications of NMR to plant metabolomics are in food quality control, e. g. identification of origin of wine (Brescia et al. 2002), coffee (Charlton et al. 2002), juice (Vogels et al. 1996) and beer (Duarte et al. 2002).

Here we will discuss several factors which should be considered for plant metabolomics using NMR. For plant materials in biological experiments, aspects such as harvesting, extraction and choice of NMR solvent are important factors to be considered in metabolomic studies. We will also show some applications of various NMR methods using several plants, and in particular *Catharanthus roseus*, as a model.

2 Experimental Consideration for Metabolomics Using NMR

2.1 Harvesting Plant Material

When plants are harvested, a plant might recognize itself as being damaged and immediate wound reactions will occur. In the short term (starting immediately after wounding) this self-defense mechanism of plants results in

oxidation or hydrolysis of metabolites. These reactions will continue and in the longer term (e. g. 12–72 h) even phytoalexin biosynthesis will produce novel compounds if the material is stored at room temperature for a certain time. To avoid these reactions, all biochemical reactions in the plant material should be stopped immediately at harvesting, for example, by freezing in liquid nitrogen followed by storage at −80 °C. Stopping biochemical reactions can also be done by heating, or by adding organic solvents or strong acid. Heating will stop enzyme activities involved in the defense reaction, but might cause decomposition of metabolites. Microwave treatment might be helpful as all material is heated to 100 °C in very short time, where conventional heating will result in a temperature gradient, in which defense reactions still may happen. To extract secologanin from *Symphoricarpos albus*, a few minutes of microwave treatment in water solution inhibited enzymatic degradation by β-glucosidase and resulted in a higher yield of secologanin (Kim et al. 2004).

The next factor to consider is the drying of the plant material. Fresh plants contain approximately 70–80% of water. As this variable water content will mix with extraction solvent, it causes inaccuracy in the ratio of the extraction solvents which results in a lower reproducibility of the metabolomic profile. Moreover, in an aqueous environment the various enzymes involved in defense will be active, whereas in a dry material these metabolites are not likely to occur. For example, the extraction efficiency of sinigrin, a well known glucosinolate of *Brassica nigra* leaves, is drastically reduced in fresh material compared to that of dried ones (unpublished results). It might be due to the fact that sinigrin is degraded by myrosinase in fresh material. For this reason, freeze dried plant material would be preferable due to higher extraction reproducibility and less degradation of metabolites.

2.2 Extraction

A number of solvents can be considered for extraction from non-polar to highly polar. In fact, there are various metabolites in plants with diverse polarity such as alkaloids, amino acids, carbohydrates, fatty acids, lipids, steroids and terpenoids. It is impossible to extract all these metabolites using one single solvent. Thus, the choice of an optimum solvent is one of the most important factors in plant metabolomics. Figure 1 shows ^{1}H NMR spectra of *Catharanthus roseus* leaves extracted with different polar solvents: methanol, 0.1% trifluoroacetic acid (TFA) and methanol-KH_2PO_4 buffer (pH 6.0). The phenolic region in the ^{1}H NMR spectra shows different profiles of metabolites. In particular, the signals from alkaloids (catharanthine and vindoline) are changeable depending on the extraction solvents. In another experiment, various solvents were tested for *Arabidopsis thaliana* and *Brassica rapa* (unpublished data). More than 90% of metabolites extracted by chloroform and *n*-hexane were fatty acids or lipids. Adenosine, cytosine, phenylpropanoids, flavonoids and terpenoids were abundant metabolites extracted by methanol, acetone or

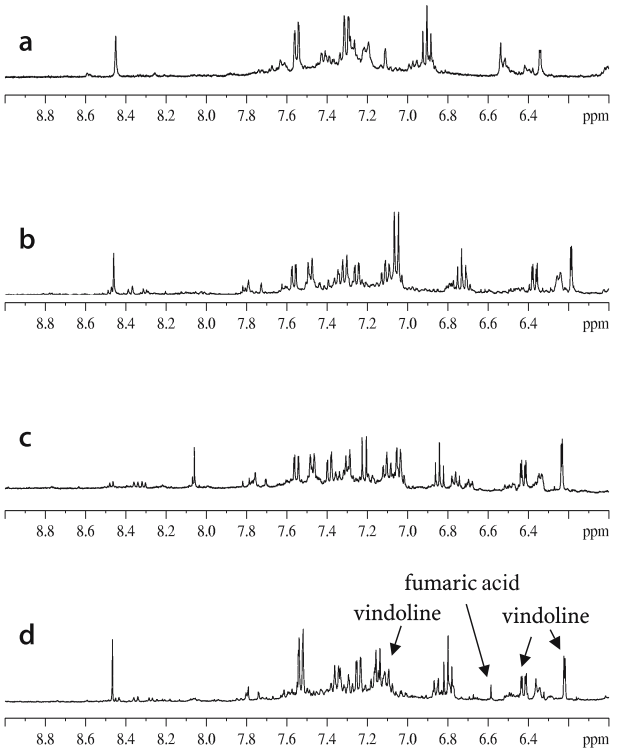

Fig. 1. ^{1}H NMR spectra (δ 9.0–6.0) of *Catharanthus roseus* leave extracted by different solvents: **a** 0.1% TFA; **b** MeOH; **c** MeOH+TFA; **d** MeOH:KH_2PO_4 buffer (1:1) pH 6.0. Leaves were extracted directly with corresponding NMR solvents (unpublished data)

acetonitrile. Amino acids and carbohydrates are well extracted by water. The extraction with the mixture of methanol-water (1:1) showed a broader range of extracted metabolites if compared to other solvents. A mixture of chloroform-methanol-water (2:1:1) was employed for several plants including *Nicotiana tabacum* (Choi et al. 2004a), *Cannabis sativa* (Choi et al. 2004b), *Catharanthus roseus* (Choi et al. 2004c), and *Ilex* species (Choi et al. 2005). In this mixture, the chloroform fraction contained a high level of fatty components, steroids, terpenoids and alkaloids (caffeine, theobromine, theophylline) while carbohydrates, phenylpropanoid glycosides and saponins were major metabolites in the water fraction. This two-phase extraction method provided a wide range of metabolites compared to a single solvent extraction. However, even with this two-phase extraction method, the extraction efficiency of medium polar metabolites such as indole alkaloids (catharanthine and ajmalicine in *Catharanthus roseus* leaves and roots) and aglycones of flavonoids was relatively low. For extraction of alkaloids it is necessary to adjust the pH of extraction solvent.

To handle hundreds of samples at one time, direct extraction with NMR solvents will be very helpful to reduce elaborate extraction procedures. It also minimizes the degradation or loss of metabolites, which occurs during elaborate extraction procedures.

For all these reasons, direct extraction using MeOD:KH_2PO_4 buffer (pH 6.0) is now routinely used for our work.

2.3 Solvent for NMR

Although NMR provides reproducible signals based on physical properties of molecules compared to other analytical methods, the signals in NMR (chemical shifts) are quite dependent on NMR solvents. Several factors should be considered to choose the solvents of NMR in plant metabolomics. The pH of the solution and concentration may influence the reproducibility of NMR spectra. Since the pH is known to affect shifts in ionizable compounds such as alkaloids (Schripsema et al. 1987), a controlled pH is thus required for metabolomic analysis. The pH control can be done using a buffer or simply adding acid. Commonly used buffers are acetate (pH range 3.7–5.6) and phosphate (pH range 5.0–7.4), in the concentration of 10−50 mmol/L. As an example, Fig. 2 shows the effect of pH on the chemical shifts of malic acid in a plant extract (*Senecio aquaticus*). The chemical shift of malic acid is highly affected not only by the pH of the NMR solvent but also by the sort of buffer. It indicates clearly the importance of pH in order to obtain reproducible chemical shifts.

For alkaloids the addition of acid can control the pH of extracts. Trifluoroacetic acid (TFA) has been used as a pH modifier to adjust pH of the NMR solvent. Several indole alkaloids including icajine, brucine, strychnine and vomicine from *Strychnos* species were analyzed by ^{1}H NMR using 1% TFA in methanol-d_4 (Frédérich et al. 2004).

Even under controlled pH, ^{1}H NMR signals of some metabolites are largely affected by their concentration in the solution. Recently, we encountered this problem with ungeremine (Rhee et al. 2004). Lower concentration of ungeremine in the solution resulted in the downfield shift of each proton by 0.1−0.01 ppm. Other examples are organic acids such as citric acid and malic acid. These acids can be found in the plant as major products of the TCA cycle. They show characteristic signals in the range of δ 2.5–δ 3.0. As shown in Fig. 3, the chemical shifts of the organic acids are also largely changed by their concentration.

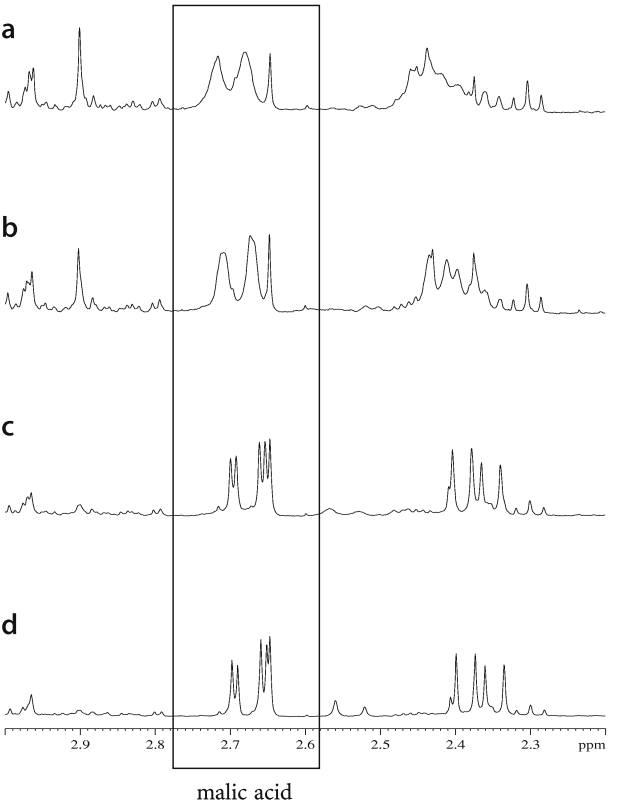

Fig. 2. ^{1}H NMR spectra of aqueous extract of *Senecio* measured in different solvents: **a** pH 6.0 in acetate buffer; **b** pH 6.0 in phosphate buffer; **c** pH 7.0 in phosphate buffer; **d** pH 8.0 in phosphate buffer. Note that the resolution of malic acid is affected by different solvent and pH (unpublished data)

3 Application of NMR for Plant Metabolome

3.1 ^{1}H NMR

Because of its relatively high sensitivity and the universal occurrence of protons in organic metabolites, ^{1}H NMR is a good starting tool for a metabolomic study. ^{1}H NMR spectroscopy has been shown to provide a wealth of information about the main metabolites in plants. From 10 to 50 mg of dried plant material, a ^{1}H NMR spectrum can be generated within 10 min. The spectrum covers approximately 50–100 metabolites, of which 10–20 compounds are easily identified. Basically the identification of the metabolites is possible by means of chemical shifts and coupling constants.

By visual inspection of the ^{1}H NMR spectrum, one has a first view of the whole metabolome of the plant material. Figures 4 and 5 show the ^{1}H NMR

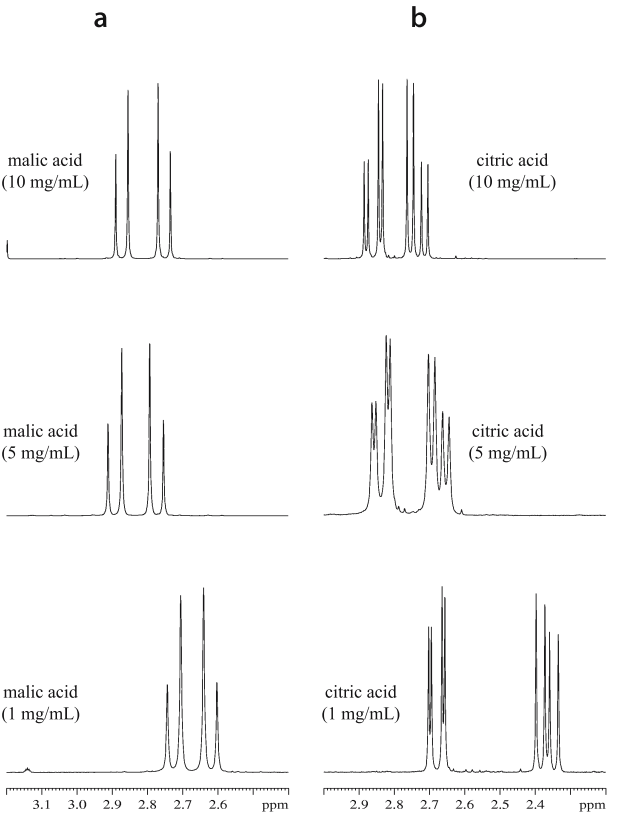

Fig. 3. ^{1}H NMR spectra of: **a** malic acid; **b** citric acid in different concentrations – *top*: 10 mg/mL, *middle*: 5 mg/mL, *bottom*: 1 mg/mL in MeOD-KH_2PO_4 buffer (pH 6.0). Note that the signals of these organic acids were upfield shift in the lower concentration. Protons of citric acid appear at δ 2.7–3.0 (d, J = 17.6 Hz) and δ 2.6–2.8 (d, J = 17.6 Hz), malic acid at δ 2.8–2.6 (dd, J = 16.6 Hz, 4.7 Hz) and δ 2.7–2.3 (dd, J = 16.6 Hz, 6.6 Hz) (unpublished data)

spectra of a healthy plant and phytoplasma-infected *C. roseus* plant. In the chloroform fraction (Fig. 4), most of the signals come from the aliphatic chains of fatty acids and methyl and methylene groups of triterpenoids or steroids. The expanded aromatic region shows the characteristic signal of vindoline: H-9 at δ 6.89 (d, J = 8.2 Hz), H-10 at δ 6.29 (dd, J = 8.5 Hz, 2.3 Hz), H-12 at δ 6.07 (d, J = 2.2 Hz), H-14 at δ 5.85 (ddd, J = 10.2 Hz, 4.9 Hz, 1.7 Hz). Together with these signals, other signals of vindoline such as OCH_3 of C-11 at δ 3.79 (s), OCH_3 of C-22 at δ 3.78 (s), H-18 at δ 0.49 (t, J = 7.4 Hz) could also be observed. The intensities of H-9 signal at 6.89 indicate that in infected plants vindoline content is two times higher than in healthy plants. Figure 5 shows the ^{1}H NMR spectra of the aqueous fraction of the *C. roseus.* Most of the signals in the crowded region at δ 3.0–5.0 come from the carbohydrates present in high amounts in the plant. Besides the signals of carbohydrates and amino acids,

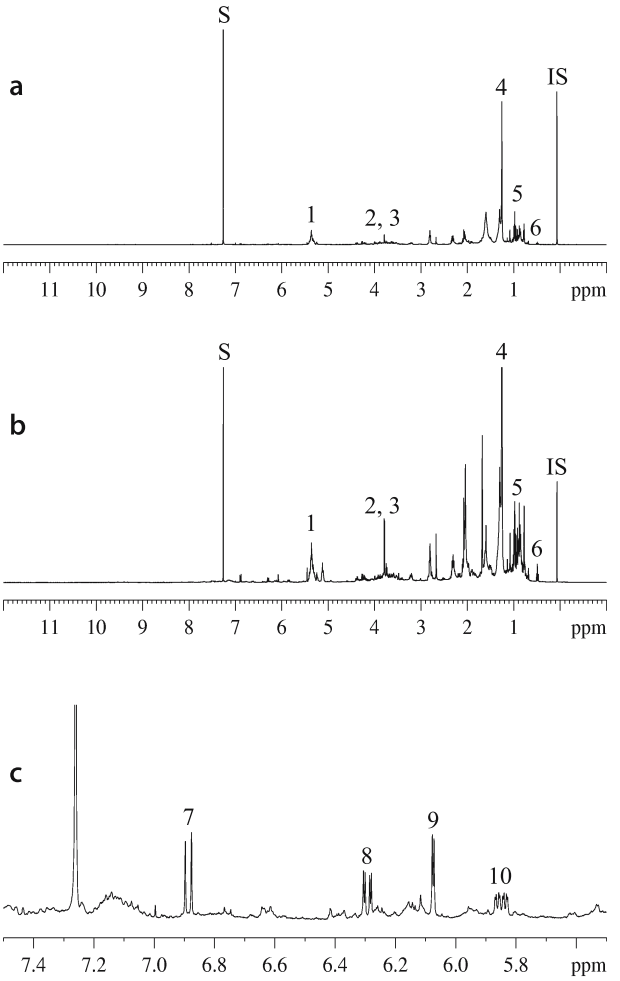

Fig. 4. ^{1}H-NMR spectra of $CHCl_3$ extract of: **a** healthy *C. roseus* leaves; **b** phytoplasma (BIL) infected *C. roseus* leaves; **c** the expansion of δ 5.5–7.5. 1; olefinic signals of fatty components or terpenoids, 2; OCH_3 of C-11 of vindoline, 3; OCH_3 of C-22 of vindoline, 4; long chain CH_2 of fatty material, 5; steroidal or triterpenoidal CH_3, 6; H-18 of vindoline, 7; H-9 of vindoline, 8; H-10 of vindoline, 9; H-12 of vindoline, 10; H-14 of vindoline, S; residual chloroform signal, IS; internal standard (HMDSO). (With kind permission of American Society of Plant Biologists, reproduced from Choi et al. 2004c)

characteristic signals from secologanin and loganic acid, important precursors of the indole alkaloids, can be found, e. g. for the H-3 of secologanin is at δ 7.57 and δ 7.49 and of loganic acid at δ 7.57 and δ 7.49. Other signals from secondary metabolites such as phenolic acids and chlorogenic acid also could be detected in the aqueous extracts. The compounds found in the *C. roseus* plant by ^{1}H NMR are listed in Table 1. In the infected leaves, the contents of

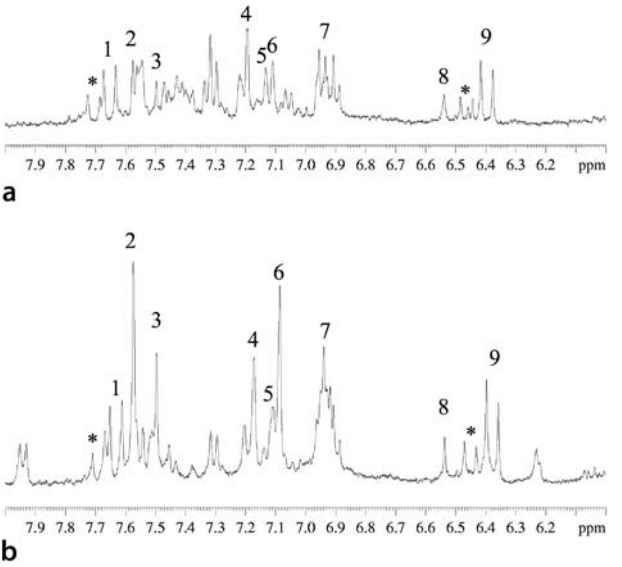

Fig. 5a,b. ^{1}H-NMR spectra of methanol-water fraction of phytoplasma (UDINESE) infected *C. roseus* leaves in the rage of δ 6.0–8.0. 1; H-7′ of chlorogenic acid, 2 and 3; H-3 of secologanin, 4; H-2′ of chlorogenic acid, 5; H-6′ of chlorogenic acid, 6; H-3 of loganic acid, 7; H-5′ of chlorogenic acid and aromatic signals of polyphenols, 8; fumaric acid, 9; H-8′ of chlorogenic acid, *; possible signals of chlorogenic acid derivatives. (With kind permission of American Society of Plant Biologists, reproduced from Choi et al. 2004c)

secologanin, loganic acid, chlorogenic acid and sugar are higher than in the healthy plants.

3.2 J-resolved NMR

NMR-based metabolomic studies typically employ one-dimensional NMR methods to minimize sample acquisition times and therefore maximize throughput. However, the spectral complexity and overlapping signals of one dimensional ^{1}H NMR limits the number of metabolites that can be identified and quantified. Moreover, few databases for ^{1}H NMR spectra of plant metabolites are available if compared to MS spectra. There are some databases related to ^{13}C NMR spectra (e. g. NMRshiftDB). However, ^{13}C NMR spectrometry has limitations in the field of plant metabolomics in the aspect of acquisition time and quantitation. It takes more than 14 h to obtain informative ^{13}C NMR spectra from the same concentration of samples from which ^{1}H NMR spectra are obtained within 10 min. In addition, broad band decoupling adopted to increase the sensitivity of ^{13}C NMR signals cause non reproducible signal increase (up to 200%) by the nuclear Overhauser effect. Therefore other two-dimensional NMR methods should be considered for application to plant

Table 1. ^{1}H Chemical shifts of metabolites of *Catharanthus roseus* leaves detected from NMR

Number	Chemical shifts (ppm) and coupling constants (Hz)	Metabolites
1	1.00 (d, J = 7.0)	H-10 of loganic acid
2	1.33 (d, J = 6.7)	H-4 of threonine
3	1.48 (d, J = 7.4)	H-3 of alanine
4	2.49 (s)	Succinic acid
5	3.56 (s)	Glycine
6	4.22 (d, J = 8.8)	Anomeric proton of fructose (sucrose)
7	4.64 (d, J = 9.5)	Anomeric proton of β-glucose
8	5.24 (d, J = 3.7)	Anomeric proton of α-glucose
9	5.42 (d, J = 3.8)	Anomeric proton of glucose (sucrose)
10	6.39 (d, J = 15.9)	H-8′ of phenylpropanoid
11	6.54 (s)	Fumaric acid
12	6.93 (d, J = 8.5)	H-5′ of chlorogenic acid
13	7.09 (d, J = 1.1)	H-3 of loganic acid
14	7.11 (d, J = 8.5)	H-6′ of chlorogenic acid
15	7.18 (s)	H-2′ of chlorogenic acid
16	7.57 (s)	H-3 of secologanin
17	7.64 (d, J = 15.9)	H-7′ of phenylpropanoid
18	9.65 (s)	Aldehyde proton of secologanin

extracts. Among the 2D NMR methods, the J-resolved technique is an interesting option. It greatly improves the resolution of the ^{1}H NMR spectra within comparably shorter time (25 min) than other 2D NMR techniques and it is easy to build up a database since a projection of the 2D spectrum on the chemical shift axis results in a spectrum in which most of protons are observed as singlet (Viant 2003).

One of the advantages of J-resolved spectra is that it provides spin multiplicities which are sometimes difficult to determine in the 1D ^{1}H NMR due to overlapping of signals. J-resolved spectra separate the chemical shift and spin–spin coupling data onto different axes, F1 for spin–spin coupling and F2 for chemical shifts. Complex aromatic signals in the ^{1}H NMR spectrum of *C. roseus* leaves appeared as less congested signals in the 2D J-resolved spectra (Fig. 6). It was quite difficult to identify indole alkaloids such as catharanthine and vindoline in the plants due to overlapping with other signals in the ^{1}H NMR (e. g. phenolics). However, the resolution of these signals in the J-resolved spectra is dramatically increased. The signals of vindoline H-9 at δ 7.0 (d, J = 2.2 Hz) are clearly separated from other signals. Also the signals from catharanthine H-9 at δ 7.6 (d, J = 7.8 Hz), H-12 at δ 7.2 (d, J = 8.0 Hz), H-10 at δ 7.2 (t, J = 8.0 Hz), H-11 at δ 7.1 (t, J = 8.0 Hz) and H-9 at δ 6.3 (d, J = 2.6 Hz) can be clearly identified.

The enhanced resolution obtained from J-resolved NMR spectra can be applied to monitor minor metabolic change in the plants which might be difficult to detect by 1D ^{1}H NMR spectra (Choi et al., unpublished data).

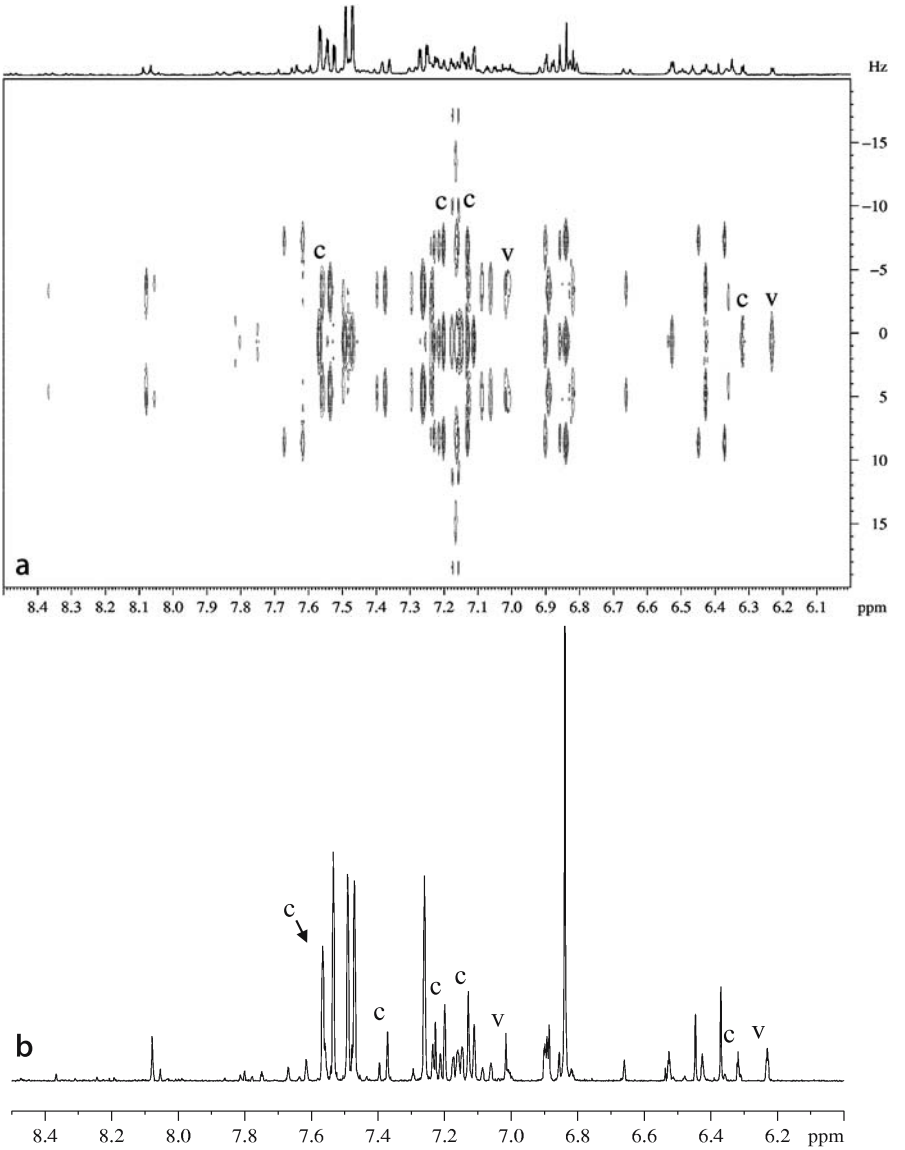

Fig. 6. a J-resolved NMR spectra of MeOD-KH_2PO_4 buffer (pH 6.0) extraction of *C. roseus* leaves. **b** Projection of J-resolved spectra. Vindoline signals (v) of H-9 and H-10, catharanthine signals (c) of H-9, 10, 11, 12 were indicated (unpublished data)

3.3 2D NMR for Structural Confirmation

Although ^{1}H NMR provides a wealth of structural information, extensive overlapping in ^{1}H NMR spectrum often makes it difficult to identify metabolites in

plants. Therefore, 2D NMR methods are essential to identify the metabolites. There are a number of 2D NMR methods which can be applied. Homonuclear correlated spectroscopy (COSY) and total correlated spectroscopy (TOCSY) are helpful to obtain information of connectivity and correlation between protons (Braunschweiler and Ernst 1983; Bax and Davis 1985). The most crowded region in the ^{1}H NMR spectra of plant extracts is in the range of δ 3.0 to δ 5.0

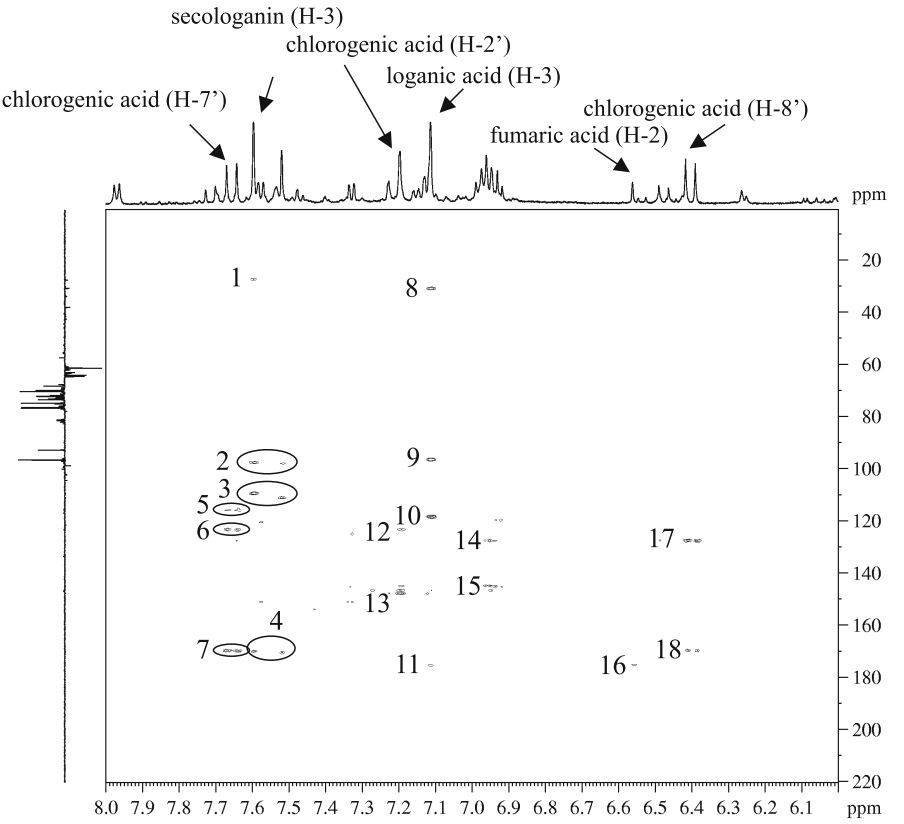

Fig. 7. HMBC spectrum of water fraction of phytoplasma (UDINESE) infected *C. roseus* leaves in the rage of δ 5.5–7.5. *1*; correlation of H-3 and C-5 of secologanin, *2*; correlation of H-3 and C-1 of secologanin, *3*; correlation of H-3 and C-4 of secologanin, *4*; correlation of H-3 and carbonyl group 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. (With kind permission of American Society of Plant Biologists, reproduced from Choi et al. 2004c)

where various amino acids and carbohydrates have their signals. The complex signals in this region can be assigned by COSY and TOCSY.

Furthermore, there are several kinds of C-H correlation spectra. Heteronuclear multiple quantum coherence (HMQC) and heteronuclear single quantum coherence (HSQC) spectra give information of direct C-H correlations (J_1). These two-dimensional NMR spectra are very useful for identifying anomeric carbons of carbohydrates (δ 90–δ 110), C-6 and C-8 of flavonoids (δ 95–δ 110) and methyl protons of terpenoids (δ 10–δ 25). For long range correlations (J_2 and J_3) in molecules, heteronuclear multiple bond correlation (HMBC) is applied to confirm structures of plant metabolites. Figure 7 shows an example of the HMBC spectrum to identify the metabolites in the aqueous fraction of *C. roseus* leaves. In case of amino acids, H-2 or H-3 is correlated with the carbonyl group of the amino acid. The proton of alanine at δ 1.48 (H-3d, J, = 4.8 Hz) correlates with the carbon at δ 178.6, glutamic acid at δ 2.14 (m) and δ 2.38 (m) correlates with δ 179.2, and glycine at δ 3.56 (s) correlates with δ 174.7. For iridoids such as loganic acid and secologanin, several correlations between protons and carbons (see figure) confirm the identity of these compounds.

4 Principal Component Analysis

The goal of metabolomic studies is either to characterize an organism or to determine the effect of certain conditions on the organism. It thus requires one to determine first the biological variability of the system followed by determining any significant change. This requires the comparison of a large number of spectra. Thus unbiased or non-targeted analysis is required for these huge data sets. For this purpose multivariate analysis and in particular principal component analysis (PCA) are suited. PCA is an unsupervised clustering method requiring no knowledge of the data set. It reduces the dimensionality of multivariate data while preserving most of the variance within it (Goodacre et al. 2000). All samples are plotted on the coordinates consisting of raw variables (chemical shifts in the case of NMR applications) and a line is constructed based on the best approximation of the data in the least squares sense. Each sample is projected onto this line. The co-ordinate value along the line is a PC1 score. Other PCs can be calculated by the line orthogonal to former PCs (Eriksson et al. 2001). Generally, the separation takes place in the first two components (PC1 and PC2). For the PCA, care must first be taken to choose an appropriate scaling method. The unit variance scaling method uses a reciprocal of standard deviation. It results in normalizing the effect of big and small signals. However, noise in the spectra might have a bigger effect on the result than expected. No scaling is used in PCA combined with NMR spectra because it could preserve the original effect of each variable but the effect of minor metabolites (in particular plant secondary metabolites) is probably neglected. The Pareto scaling method is preferred for application to the analysis of NMR spectra. It gives each

variable a variance numerically equal to its initial standard deviation instead of unit variance. Therefore, the Pareto scaling is an intermediate between no scaling and unit variance scaling. The principal components can be displayed graphically as a scores plot. This plot is useful for observing any grouping in the data set. Coefficients by which the original variables must be multiplied to obtain the PC are called loadings. The numerical value of a loading of a given variable on a PC shows how much the variable has in common with that component. Thus, for NMR data, loading plots can be used to detect the metabolites responsible for the separation in the data.

Figure 8 shows an example of PCA of healthy and phytoplasma-infected *C. roseus*. PCA score plots of healthy and infected plants by ten different phytoplasmas show that healthy *C. roseus* leaves are clearly separated from the phytoplasma infected leaves in both chloroform fraction (Fig. 8a) and water fraction (Fig. 8b). Loading plots explain that, in the chloroform fraction, the

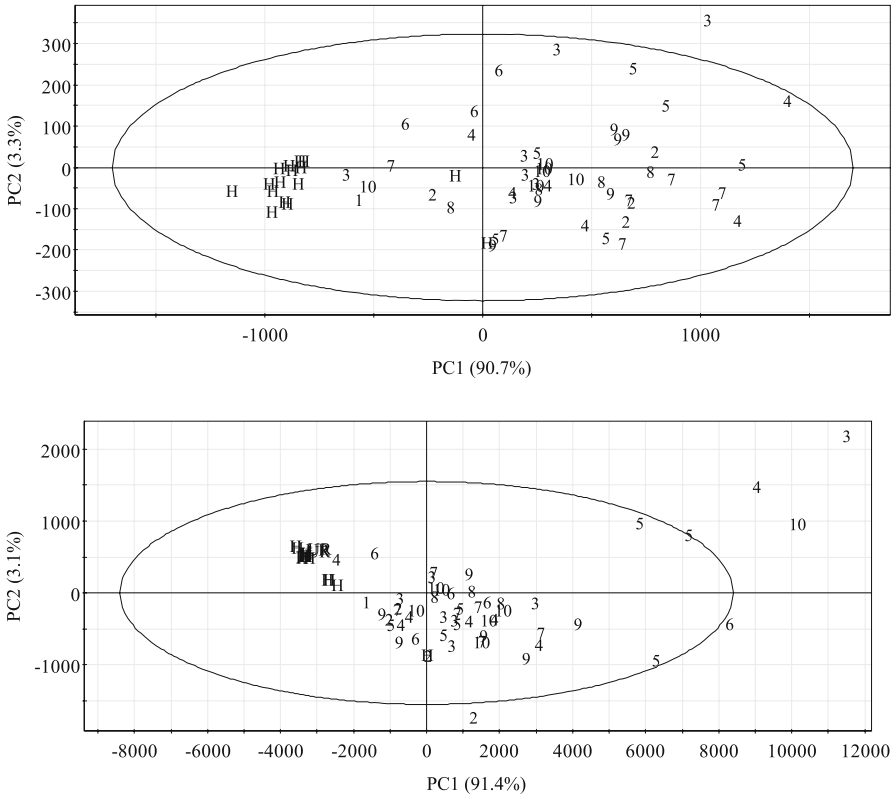

Fig. 8. Score plot of principal component analysis of: **a** $CHCl_3$ extracts; **b** water extracts of healthy and phytoplasma infected *C. roseus* leaves. *1–10*: Infected plants by 10 different phytoplasmas, H; healthy plant. The *ellipse* represents the Hotelling T2 with 95% confidence in score plots. (With kind permission of American Society of Plant Biologists, reproduced from Choi et al. 2004c)

responsible components for separation are: fatty component – the signals at 1.2–1.4 (CH_2), 5.0–5.5 (olefinic CH_2), and indole alkaloids such as vindoline – the signals at 3.79 (OCH_3 of C-11), 3.78 (OCH_3 of C-22), 0.49 (H-18). It indicates that *C. roseus* leaves infected by phytoplasma contain less fatty components and higher vindoline compared to healthy leaves. When the intensity of each signal was compared, it is clear that infected leaves have two to four times increased level of vindoline relative to healthy plants. For the water extract, the score plot shows that healthy leaves are well separated from infected plants by both PC1 and PC2 (Fig. 8b). The healthy leaves have lower PC1 and higher PC2 relative to infected ones. The loading plot of PC1 and PC2 explained that most of infected *C. roseus* leaves have higher amounts of sucrose, chlorogenic acid, loganic acid, secologanin, and polyphenols compared to healthy plants.

By using ^{1}H NMR in combination with PCA, it is clearly shown that the metabolites related to the biosynthesis of terpenoid indole alkaloid (loganin acid, secologanin, vindoline) and phenylpropanoids (chlorogenic acid, polyphenol) are present in higher amounts in the phytoplasma-infected leaves.

5 Concluding Remarks

Several analytical methods may be used for metabolomic profiling of plants; however, ^{1}H NMR spectra offers a wealth of information of metabolites compared to other methods. Decoupled NMR spectra (J-resolved) provide even more information since metabolites can be accurately integrated and it can exclude broad resonances from macromolecules and spin–spin coupling data. In addition, combinations of two dimensional NMR methods are quite helpful to identify the metabolites in plant extract.

To be able to compare all the data generated from NMR from different experiments and different laboratories, a large database is required. However, that requires a high degree of reproducibility, which can be achieved by using a standardized method for sample preparation and data acquisition.

So far, NMR has been successfully used for the metabolomic fingerprinting and profiling of plants and is successfully applied in quality control of among others, food and botanicals. The use of NMR metabolomics in functional genomics will be the challenge for the coming year.

References

Aharoni A, Ric de Vos CH, Verhoeven HA, Maliepaard CA, Kruppa G, Bino R, Goodenowe D (2002) Non-targeted metabolomic profiling using Fourier transform ion cyclotron mass spectrometry (FTMS). OMICS J Intergrative Biol 6:217–234

Bax A, Davis DG (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J Magn Reson 65:355–360

Braunschweiler L, Ernst RR (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J Magn Reson 53:521–528

Brescia MA, Caldarola V, de Giglio A, Benedetti D, Fanizzi FP, Sacco A (2002) Characterization of the geographical origin of Italian red wine based on traditional and nuclear magnetic resonance spectrometric determinations. Anal Chim Acta 458:117–186

Charlton AJ, Farrington WHH, Brereton P (2002) Application of ^{1}H NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee. J Agric Food Chem 50:3098–3103

Choi H-K, Choi YH, Verberne M, Lefeber AWM, Erkelens C, Verpoorte R (2004a) Metabolic fingerprinting of wild type and transgenic tobacco plants by ^{1}H NMR and multivariate analysis technique. Phytochemistry 65:857–864

Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AWM, Verpoorte R (2004b) Metabolomic differentiation of *Cannabis sativa* cultivars using ^{1}H NMR spectroscopy and principal component analysis. J Nat Prod 67:953–957

Choi YH, Tapias EC, Kim HK, Lefeber AWM, Erkelens C, Verhoeven JTJ, Brzin J, Verpoorte R (2004c) Metabolomic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using ^{1}H-NMR spectroscopy and multivariate data analysis. Plant Physiol 135:2398–2410

Choi YH, Sertic S, Kim HK, Wilson EG, Michopoulosa F, Lefeber AWM, Erkelens C, Verpoorte R (2005) Classification of *Ilex* species based on metabolomic fingerprinting using NMR and multivariate data analysis. J Agric Food Chem 53:1237–1245

Duarte I, Barros A, Belton PS, Righelato R, Spraul M, Humpfer E, Gil AM (2002) High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. J Agric Food Chem 50:2475–2481

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) Multi-and megavariate data analysis. Umetrics Academy, Umeå, Sweden

Fiehn O (2002) Metabolomics-the link between genotypes and phenotypes. Plant Mol Biol 48:155–171

Frédérich M, Choi YH, Angenot L, Harnischfeger G, Lefeber AWM, Verpoorte R (2004) Metabolomic analysis of *Strychnos nux-vomica*, *Strychnos icaja* and *Strychnos ignatii* extracts by ^{1}H nuclear magnetic resonance spectrometry and multivariate analysis techniques. Phytochemisty 65:1993–2001

Goodacre R, Shann B, Gilbert RJ, Timmins EM, McGovern AC, Kell DB, Logan NA (2000) Detection of the dipicolic acid biomarker in *Bacillus* spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. Anal Chem 72:119–127

Huhman DV, Sumner LW (2002) Metabolic profiling of saponins in *Medicago sativa* and *Medicago trumcatula* using HPLC coupled to an elcetrospreay ion-trap mass spectrometer. Phytochemisty 59:347–360

Kim HK, Choi YH, Luijendijk TJC, Vera Rocha RA, Verpoorte R (2004) Comparison of secologanin extraction methods and quantitative analysis of secologanin from *Symphoricarpos albus* by using ^{1}H-NMR. Phytochem Anal 15:257–261

Rhee IK, Appels N, Hoete B, Karabatak B, Erkelens C, Stark LM, Flippin LA, Verpoorte R (2004) Isolation of the acetylcholinesterase inhibitor ungeremine from *Nerine bowdenii* by preparative HPLC-coupled on-line to a flow assay system. Biol Pharm Bull 27:1804–1809

Schripsema J, van Beek TA, Verpoorte R, Erkelens C, Perera P, Tibell C (1987) A reinvestigation of the stereochemistry of tubotaiwine using NMR spectroscopy. J Nat Prod 50:89–101

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics:large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817–836

Tolstikov VV, Fiehn O (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectroscopy. Anal Biochem 301:298–307

Viant MR (2003) Improved methods for the acquisition and interpretation of NMR metabolomic data. Biochem Biophys Res Commun 310:943–948

Vogels JTWE, Terwel L, Tas AC, van den Berg F, Dukel F, van der Greef J (1996) Detection of adulteration in orange juices by a new screening method using proton NMR spectroscopy in combination with pattern recognition techniques. J Agric Food Chem 44:175–180