Metabolomics: back to basics

Verpoorte, R.; Choi, Y.H.; Mustafa, N.K.; Kim, H.K.

Phytochem Rev (2008) 7:525–537
DOI 10.1007/s11101-008-9091-7

Metabolomics: back to basics

R. Verpoorte · Y. H. Choi · N. R. Mustafa · H. K. Kim

Received: 8 January 2008 / Accepted: 22 February 2008 / Published online: 11 March 2008

Abstract Metabolomics has developed into a major tool in functional genomics and plant systems biology. The various methods used for metabolomic analysis will be discussed from the analytical methods back to the preanalytical phase and the biological experiment. Particularly aspects of the preanalytical phase of the analysis is dealt with, including the risks of artefact formation with the various commonly used solvents. Metabolomics is like a snap shot, and conclusions from dynamic systems must be drawn with great care as demonstrated with a biosynthetic study of salicylate in *Catharanthus roseus* cell cultures.

Keywords Metabolomics · NMR spectrometry · Sample preparation · Extraction · Artefacts · *Catharanthus roseus*

Introduction

With the ever increasing speed of sequencing of genes, the emphasis of research is moving towards the function of genes. The focus was first on proteomics, as there is a logical link between the sequence of a gene and the amino acid sequence of a protein. However, such a sequence does not give much information about the function of the protein. For most proteins the function is best reflected by the phenotype. The characterization of the phenotype thus has become an important objective, and here metabolomics comes into the picture.

In the past years metabolomics has rapidly developed as an important tool in functional genomics. Numerous reviews have been published about the technology and its applications. Recently, we published on applications of NMR-based metabolomics to studying plant secondary metabolism (Verpoorte et al. 2007). Here we will discuss the complete pathway of metabolome analysis from the analytical methods back to the basics of the preanalytical phase and the biological experiment.

In the past years several papers have been published which made some quite serious comments on research in life sciences. Ioannidis (2005) shows by simulations that in most studies the experimental design is such that it is more likely that the conclusions from results are false than true. The statement was made "*Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias*". Also, Lay et al. (2006) had some quite critical remarks: "*It remains to be seen whether it will be possible to learn from prior "problems with the*

R. Verpoorte (✉) · Y. H. Choi · N. R. Mustafa · H. K. Kim
Division of Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, Leiden, The Netherlands
e-mail: verpoort@chem.leidenuniv.nl

"omics"" and "do it right" so as to leave a legacy of reliable and useable data for future researchers" and *"or whether production of difficult-to-interprete data with significant uncertainty and false-positives will continue*".

Serious allegations and our first reaction would be "this cannot be true", but just think about the history of the elucidation of the terpenoid biosynthesis. There are hundreds of papers reporting that mevalonate is incorporated into all types of terpenoids, many of which are now known to come from the MEP pathway, such as the monoterpenes, diterpenes and carotenoids (e.g. see for biosynthesis terpenoid indole alkaloids Cordell 1974). Many other examples of models and theories that proved to be wrong can be found. This does not mean that the experimental data are wrong, it is usually a matter of interpretation of the results, or the experimental design. With this in mind we have to go back to our own research and see if we also go into this trap, or that we can avoid it.

When one makes a hypothesis or a model, there is always the possibility that it might be wrong. In research the most important is to avoid making models and hypotheses into dogmas. It is obvious that the crux is the data of the measurements. If we perform an experiment the observations we make should be stored in such a way that they can be used again in future experiments, by ourselves or by others. The data in itself are unbiased and should be presented as such, independent of any interpretation. The strength of such an approach is probably best illustrated by the success of molecular biology in the past two decades. Once a gene sequence is determined it stands for ever as a fact that can be used over and over again. The same applies for amino acid sequences. By proper storage of these results, data mining becomes possible in any place at any time. Similarly in the field of chemistry, chemical structures are a form of data storage that has a long term value. Unfortunately we cannot say this about a chromatogram, a table, or graph with results of an experiment. However, raw physical data have a long term value, e.g. NMR-spectra will always be the same for a given compound. As long as we can save the results of our experiments in such "eternal" values, they are available for future use, for data mining, or for new interpretations, models, etc. The key for long-term success of all our research efforts is consequently to eternalize the observations we have made. This should result in databases for all kind of observations, similar as we have databases for sequences of genes and proteins databases, and for chemical structures.

In this review we will first briefly discuss the different analytical methods, and discuss data storage. Then we will focus on the preanalytical phase of the metabolome analysis as the extraction and sample preparation steps are the basis for highly reproducible plant metabolome analyses. Finally an example is given of a biosynthetic study, showing the limitations of metabolomics in dynamic systems.

Analytical methods

The ultimate goal of metabolomics is the qualitative and quantitative analysis of all metabolites in an organism. A quite ambitious goal, which none of the methods applied at present can meet. One has to deal with a large number of compounds with totally different chemical properties, which might be unstable and present in a wide range of quantities. Concerning the number, our estimation is that an organism contains about the same number of compounds as it has genes, e.g. for plants that would be ca. 30,000 compounds. Of course not every gene is involved in metabolism, but a number of steps do not require any enzyme, as they are chemical reactions. Moreover, a number of enzymes have multiple functions.

For the analysis of the metabolome, five major approaches are currently used, HPLC or TLC-UV, GC-MS, LC-MS, MS^n and NMR-spectrometry. Each has its strong and weak points (Table 1). Among the possible candidate technologies available for analyzing a metabolome, mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are currently considered to be the primary, most universal approaches.

Mass spectrometry is often connected to chromatographic separation like with gas chromatography (GC) and high performance liquid chromatography (HPLC). Gas chromatography coupled with MS (mainly electron impact mass spectrometry) shows high sensitivity and resolution, and a reproducible fragmentation pattern of molecules. Commercial databases are helpful for identification of metabolites (Kopka 2006; Looser et al. 2005; Santos and

Table 1 Comparison of analytical methods used in current metabolomics

	HPLC or TLC-UV	GC-MS	LC-MS	MS^n	NMR
Sample preparation	++	−	−	+	+++
Reproducibility	−	+	−	+	+++
Absolute quantitation	−	−	−	−	+++
Relative quantitation	+	++	+	++	+++
Identity	+	++	++	++	++
Compound number	ca. 30	ca. 1000	ca. 200	ca. 1000	ca. 200
Sensitivity	+	++	++	+++	−

Scale from − to +++ for major disadvantages to major advantages

Galceran 2003). GC coupled to time-of flight (TOF) mass spectrometry is now applied in metabolomics to obtain higher accuracy of *m/z* compared to a conventional quadrupole mass detector and thus improving metabolite identification (Fiehn 2003; Taylor et al. 2002; Wagner et al. 2003). To further improve resolution two dimensional GC-MS has been employed (Dallüge et al. 2003). However, the inherent limitation of GC-MS is the small range of polarity and molecular weight of metabolites that can be analyzed by GC. To overcome this problem derivatization methods to make compounds more volatile are applied prior to analysis. However, incomplete reactions, resulting in more than one peak for compounds with several reactive groups such as sugars and unstable and non-volatile compounds remain a problem (Ryan and Robards 2006). Particular complex structures such as of plant secondary metabolites, e.g. often present as glycosides, cannot be analyzed by gas chromatography.

Soft ionization MS techniques (liquid introduction and mostly connected to HPLC) such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), or matrix assisted laser desorption (MALDI) are particularly applied in the analysis of polar metabolites (Gobey et al. 2005; Looser et al. 2005; Sumner 2003). The wide range of molecular weights and polarities, and the exact molecular weight are definitely strong points for MS in metabolomics. Moreover, the tandem MS technique offers partial structure information of the molecules. However, very different sensitivity of molecules and reproducibility affected by ionization conditions and matrix, are still unsolved problems in MS metabolomics. These problems particularly concern the quantitation of the metabolites. In addition, the lack of LC-MS databases means that the mass signals have to be identified by manual searches (Moco et al. 2006).

Nuclear Magnetic Resonance spectroscopy (NMR) used to be regarded as complementary to other analytical methods (Reily and Lindon 2005). However, this has changed, because as a tool for metabolomics, NMR has some unique advantages over chromatography and MS-based methods in metabolomics. NMR provides a very fast and detailed analysis of the biomolecular composition of a crude extract. NMR requires only relatively little sample preparation, and improvements in equipment has increased sensitivity manifold (e.g. cryoprobes) (Choi et al. 2006b; Pauli et al. 2005; Verpoorte et al. 2007). NMR is a universal detector in which the signals give direct information about the chemicals structure of the molecules.

An NMR-spectrum is a physical characteristic of a compound, and is thus highly reproducible. That is a major advantage as it means that NMR-metabolomics data stand for ever, as long as the same extraction procedures and the same NMR-solvents are used. Datamining from the raw data from previous experiments is thus always possible.

Quantitative analysis

In terms of sensitivity MS is probably the best method, but like the chromatographic methods (GC-MS and LC-MS) an absolute quantitation of the individual compounds is only possible with the aid of calibration curves of all individual compounds. Obviously this is not realistic, as it would mean calibration curves for hundreds of signals, many of which are even not yet identified. Consequently these methods only measure relative values for each of the

detected compounds, e.g. the percentage of increase or decrease. ^{1}H NMR-spectrometry has here a great advantage, as the signal intensity is only dependent of the molar concentration in the solution.

Consequently, as long as signals are well resolved in the NMR spectra, by using a proper internal standard the absolute concentration of metabolites can be easily calculated, and no calibration curves are needed for all individual compounds (Pauli et al. 2005). In spite of the lower sensitivity, the ease of quantitation gives NMR-based metabolomics a major advantage over the other analytical methods.

Identification of compounds

As mentioned above in most plant metabolomic analyses less than half of the signals are assigned. Highly accurate molecular weight and elemental composition as provided by Fourier transform ion cyclotron resonance mass spectrometry are quite useful for the identification of well known metabolites (Breitling et al. 2006), though for each molecular weight and elemental formulae quite a few structures are possible.

Comparison with reference compounds might be successful for many well known primary metabolites, but for plant specific secondary metabolites reference compounds are usually not available and the molecular weight and elemental composition alone are not sufficient for their identification. In practice many common primary metabolites can be identified in MS, particularly when coupled with a chromatographic separation.

The quality of a metabolomic analysis is not dependent on the number of detected signals but on the number of identified metabolites. One single identified metabolite is more meaningful than scores of unidentified signals. A way to solve the problem of the large number of unknowns is to only look at differences between samples and identify the compounds connected with this. In fact the past years many studies describing metabolomic approaches only identify a few marker compounds that are different between the different experimental conditions of the organism. Typically this is in the order of 20–40 compounds that are identified with certainty. Identification requires more spectral data, besides UV and MS the most important are NMR-spectra, as NMR-spectrometry is the only method that for each single compound gives a unique and characteristic spectra. The information of ^{1}H, ^{13}C and various 2D-NMR spectra allows the elucidation of the complete structure of a compound by pure logic reasoning. All other spectroscopic methods only give information of partial structural elements, but not on every single atom in a molecule as in NMR, and consequently can only be used for partial structural elucidation, or for identification of known compounds. Finally it must be noted that the stereochemistry may be different for compounds with the same basic structure.

Structure elucidation can be done in the NMR-spectra of crude extracts, which is an important advantage in the use of NMR-based metabolomics (e.g. Choi et al. 2006a; Liang et al. 2006). At least it is obvious that phytochemists in the coming years will have a major task in the isolation and structure elucidation of all the unknown signals in the various metabolomic analyses.

Long-term data storage

As mentioned above molecular biology has made great progress in the past decades, mostly because the sequences of genes can easily be stored and used for future reference. However, in metabolomics we do not have databases with a long term value yet. One of the main problems is the reproducibility of the analytical methods, and the mentioned unknown signals in the chromatograms and spectra. In the field of chromatography every day new and better columns are introduced to the market. Consequently, new methods are continuously reported for the major separation problems. This might be improvements for targeted analyses, but makes such methods not suited for long term storage of data, i.e. not suited for metabolomics. For example, when we wrote our books on “Chromatography of alkaloids” 25 years ago, we reviewed the analysis of tropane alkaloids: 69 references concerned the HPLC and 32 the GC analysis (Baerheim Svendsen and Verpoorte 1983; Verpoorte and Baerheim Svendsen 1984). Since then every year numerous new papers have been published on this separation.

For overcoming the problems with long-term storage of chromatographic data many alignment

algorithms have been proposed to correct small changes in retention behavior of compounds due to environmental and systemic factors, but in case of novel type of columns this does not help. For example, going from HPLC to UPLC means that most of metabolomic data of the previous analyses is lost. Only tables with the quantitative data of all compounds that were identified with 100% certainty remain. In spite of the tabulated data, all information on unknowns is difficult to save for future use and for any new method (or new type of chromatographic column) the whole procedure of validation, peak identification and calibration has to be repeated.

This is where NMR-spectroscopy has a major advantage; it is the only method where the physical characteristics of compounds always will be the same. An NMR-spectrum of an extract is just the sum of all the highly reproducible individual spectra of all compounds present. It makes the ^{1}H NMR-based metabolomics the method of choice for a long-lasting database. The success of the NMR-analysis of urine (metabonomics) shows that when one standard procedure is used rapid progress can be made in storing all analytical data for long term use (Lindon et al. 2005). Datamining from the raw data from previous experiments is thus always possible (see below). In case of plants crude extracts, we were able to extract information from the ^{1}H NMR fingerprints we made of a series of our cell cultures in the 1980's (Schripsema and Verpoorte 1991; Schripsema et al. 1991). We have now a database with 1D and 2D spectra including more than 500 plant metabolites and some 20,000 spectra of various plant extracts to which we applied metabolomics. This database is very useful for data mining and comparison with newly recorded spectra. For transcriptomics one has been able to define a unified method: the MIAME protocols for reporting transcriptomic data (Brazma et al. 2001). A similar approach is needed for metabolomics data, though it will be more difficult to define a unified method, because the variation in structures of plant metabolites is very large and consequently different analytical methods will always be required. That includes both extraction and the subsequent instrumental analysis. The difficulties in making metabolomic databases is probably best illustrated by the fact that after ca. 40 years of GC(-MS) and MS and 30 years of HPLC still no universal databases are available that cover all natural products. Considering the number of variables it seems that general standard protocols for NMR-measurements would be easiest to realize.

Preanalytical procedures

Every analysis starts with the harvesting of the material. In this step already a number of problems are present. First of all the time of harvesting (diurnal changes) and the stage of development of a plant are important parameters for the metabolome. Possible variation needs to be determined by measuring series of samples from different times of the day and different stages of development. Once the biological variation is known, the standard conditions for experiments can be set and one may measure the variation due to experimental conditions.

Harvesting means stressing and usually also wounding the plant. This will cause changes in the plant metabolism. Some are very fast (within minutes, e.g. hydrolysis of glucosides like glucosinolates or cyanogenic glucosides), others are dependent on induction of pathways on the level of the genes. To have reproducible samples, the metabolism must be immediately stopped at harvesting, i.e. by freezing in liquid nitrogen. However, freezing destroys the cells and after thawing all kind of biochemical conversions may occur. This should be blocked during thawing, i.e. by extraction of the frozen material by a denaturating solvent, or by a brief treatment with microwaves, killing all enzyme activity. Freeze-drying is thus a gentile way to avoid degradation due to enzymes, but in the subsequent extraction denaturing solvents are required. Treatment with microwaves do destroy the enzymes. For example in the analysis of flower colors we found that in fresh flowers that were subjected to a short microwave treatment only had one major anthocyanidin glycoside, whereas in a classical extraction, without any denaturation step, hydrolysis occurred and several glycosides were found (T.J.C. Luijendijk and R. Verpoorte, unpublished results).

The next step is the isolation of the metabolites from their matrix. Solvent extraction is the most common, alternatives are (steam) distillation for volatile compounds and supercritical fluid extraction.

For solvent extraction many options exist. Each solvent has a different profile of compounds that are dissolved, depending on polarity or specific interactions with the solvent. No universal extraction solvents exist. To chose an extraction method one should first consider what the compounds of interest are. In plants one may distinguish three types of compounds:

- Non-polar compounds which are part of cell membranes, the cuticula, or occur in specialized cells, such as glandular hairs. Most of these compounds are either terpenoids or fatty acids and their derivatives.
- Medium polar compounds, which includes most of the secondary metabolites. These compounds are connected with the interaction of the plant with its environment, e.g. as defense against pests and diseases. Such compounds are made to have a biological activity, they are thus expected to have drug-like properties, e.g. to be able to pass cellular membranes.
- Polar compounds. Much of the primary metabolism concerns water soluble compounds, e.g. sugars and products connected with the biosynthesis of amino acids and the production of energy.

None of the metabolomics method is able to deal with this broad variety of compounds, usually only parts of these are measured by the extraction method employed. Commonly methods are applied that extract medium polar or polar compounds. For the non-polar compounds lipidomics has been developed as a sub-metabolomics approach (Han and Gross 2003). The analysis of essential oils as it has been performed since the very beginning of gas chromatography more than 40 years ago, can in fact be considered as sub-metabolomics *avant-la-lettre* (Baerheim Svendsen and Karlsen 1966).

Extraction solvents

For extraction of plants a number of solvents are available. In selecting a solvent the following parameters should be considered:

- Selectivity of the solvent. The solvent parameters as described by Snyder (1974) for chromatographic separations are useful in selecting solvents based on their selectivity and polarity.
- Polarity determines to a great extend the type of compounds that are best extracted. We compared a series of polar solvents for the extraction of *Arabidopsis* using ^{1}H NMR followed by principal component analysis. Each solvent shows its own range of metabolites that it dissolves (Verpoorte et al. 2007). A small difference in polarity already brings about large differences in the pattern of extracted metabolites. This is illustrated by the results in measuring the metabolic changes in jasmonate treated *Brassica nigra* leaves. When 50% of methanol was used for the extraction hardly any difference was found if compared to controls. However, in the case of 80% methanol extraction a clear difference is observed in the region of the NMR spectra covering aromatic compounds (Fig. 1).

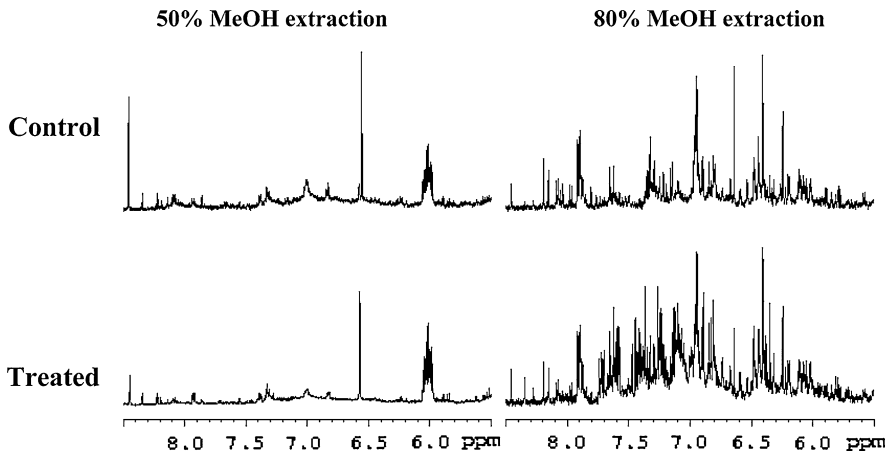


Fig. 1 Comparison of ^{1}H-NMR spectra of control *Brassica nigra* and treated with jasmonic acid obtained from 50 to 80% methanol

- Boiling point is important in case solvents have to be evaporated. One should also take into account that metabolites might be volatile, for example a phenolic compound like salicylate is volatile, and without proper measures the recovery of such a compound can be completely erratic. Recoveries varying from 30-60% have been reported for the analysis of salicylic acid in plants, thus resulting in very unreliable data, but by buffering the solution this problem could be solved (Verberne et al. 2002) (Fig. 2).
- Toxicity and environmental considerations are important issues. For example benzene should be avoided due to its carcinogenicity in spite of its higher resolution in NMR and better chromatographic results for certain groups of metabolites.
- Possible interference with the analytical procedure, e.g. strong UV-absorption in HPLC, or suitability for subsequent derivatization step (e.g. water-free solvents like pyridine or DMF) as required for the GC-MS based metabolomics.
- Possible contaminations in solvents that may interfere with the analysis, e.g. antioxidants such as butylated hydroxytoluene (BHT, MW 220), 2,6-di-*tert*-butylphenol (MW 206) and propylgallate (MW 212) (in ethers, chloroform). Water as an impurity varies depending on the solvent, but may cause overlapping with metabolite signals in the NMR spectra, (e.g. water in methanol and chloroform give major signals at δ(4.8 and δ(1.5, respectively).

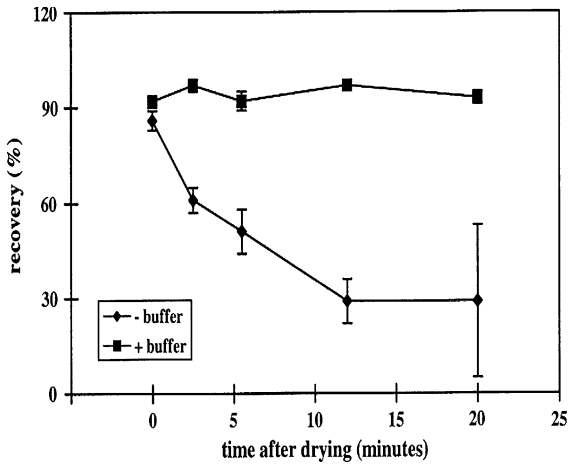


Fig. 2 Salicylic acid in *n*-hexane-EtOAc (1:1) with or without buffer pH 5.5 concentrated in Speedvac (adapted from Verberne et al. 2002)

- Possible contaminations that may cause artifact formation: peroxides (in ethers), dichlorocarbene (in chloroform), ethylchloroformate (in chloroform), phosgene (in chloroform) (Fig. 3), dichlorobromomethane (in dichloromethane; chloroform) (Fig. 4), ethanol (in chloroform 1–2% of ethanol as stabilizer; diethyl ether; ethyl acetate), formaldehyde (in chloroform; ethyl acetate), acetaldehyde (in chloroform; ethyl acetate). Examples of various reaction products with the degradation products of chloroform have been reported in the analysis of nitrogen containing compounds (alkaloids) (Fig. 3). The reaction of tertiary amines with dichloromethane has been reported to be quite fast, a compound like strychnine will be converted for 100% into the quartenary dichloromethane compound when extracted from plant material in a Soxhlet apparatus with dichloromethane (Bisset and Choudhury 1974) (Fig. 4). Peroxides may result in the formation of *N*-oxides, which can in their turn be converted in carbinolamines and result in ring opening (Fig. 4).

In case of aqueous solvents the pH is an important factor, acidic or basic conditions may lead to all kind of artifacts, for example the common metabolite chlorogenic acid (5-cinnamoyl-quinic acid) is easily converted in the 3- and 4-isomers (Fig. 5) (Hanson 1965), or may form a lactone. In Fig. 6 the effect of pH and oxidation on the decomposition of dimeric indole alkaloids is shown, from our own experience we know that the isolation of these dimeric alkaloids from the plant always results in considerable degradation, even under very mild conditions. One single alkaloid as bisnordihydrotoxiferine may easily results in a mixture of at least 81 different products! (Verpoorte and Baerheim Svendsen 1976, 1978) (Figs. 4, 6).

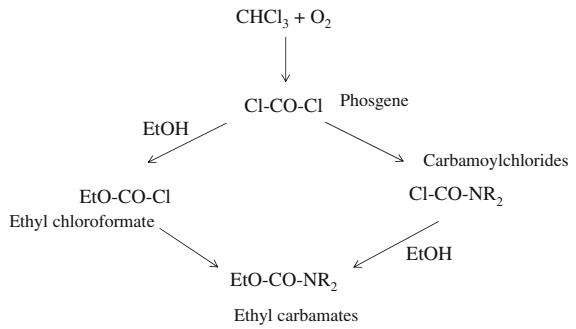
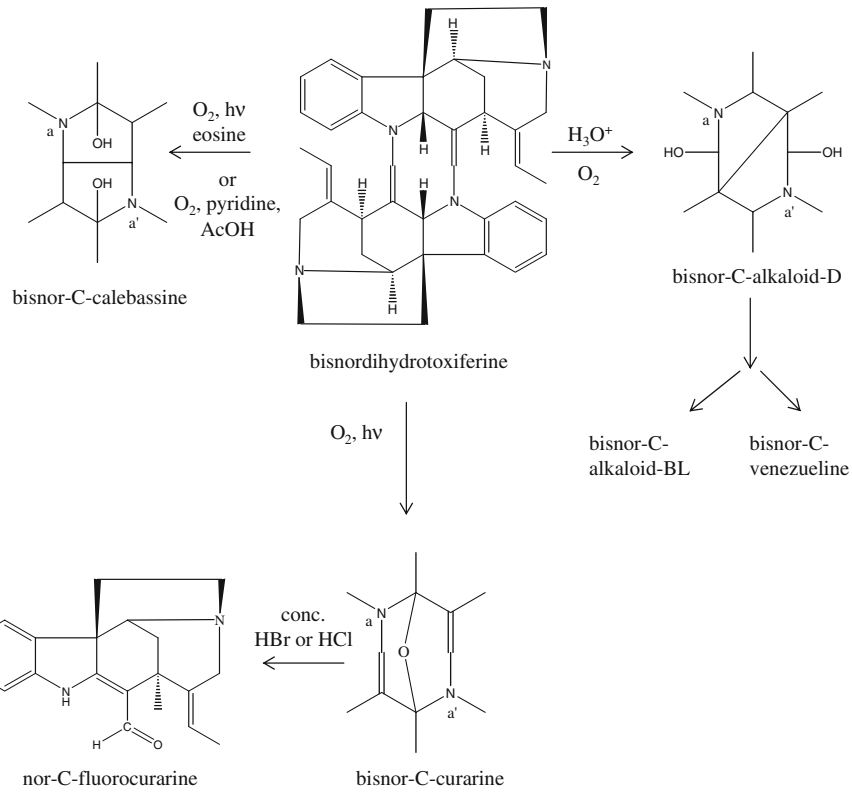


Fig. 3 Possible artifacts from chloroform



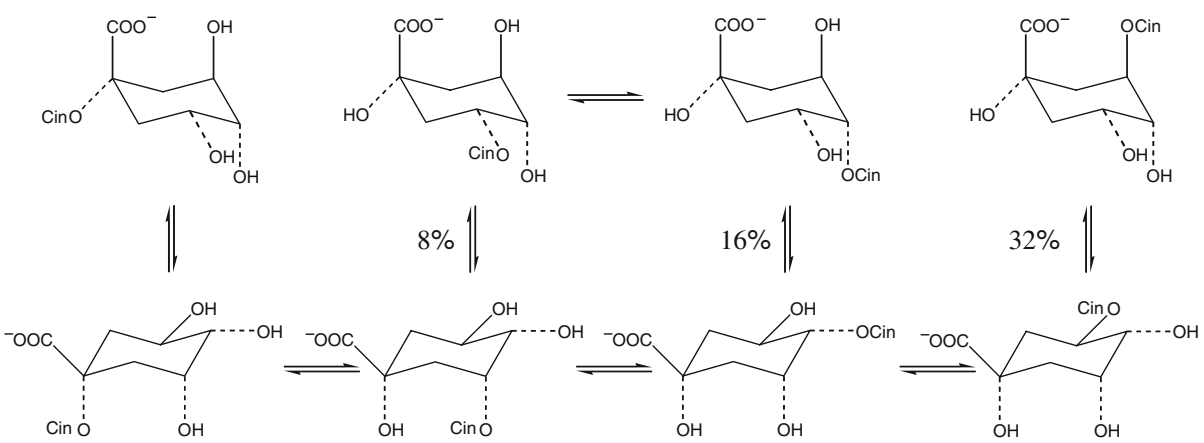


Fig. 5 Intramolecular migration cinnamoyl group in chlorogenic acid at 90°C and pH 7 for 3 min (adapted from Hanson 1965)

Another problem in NMR-based metabolomics comes from hydrogen exchange of OH, NH, or SH with deuterated solvents like water or alcohols. Also aromatic protons such as in phloroglucinol-type compounds and aldehydic protons can be replaced by those of deuterated NMR solvents. Figure 7 shows an example in which storage for some time in deutererated solvent reduces the signal intensity of H-6 and H-8 in the ^{1}H-NMR spectrum of naringenin due to the replacement with deuterium of the deuterated NMR solvent (Choi et al. unpublished data).

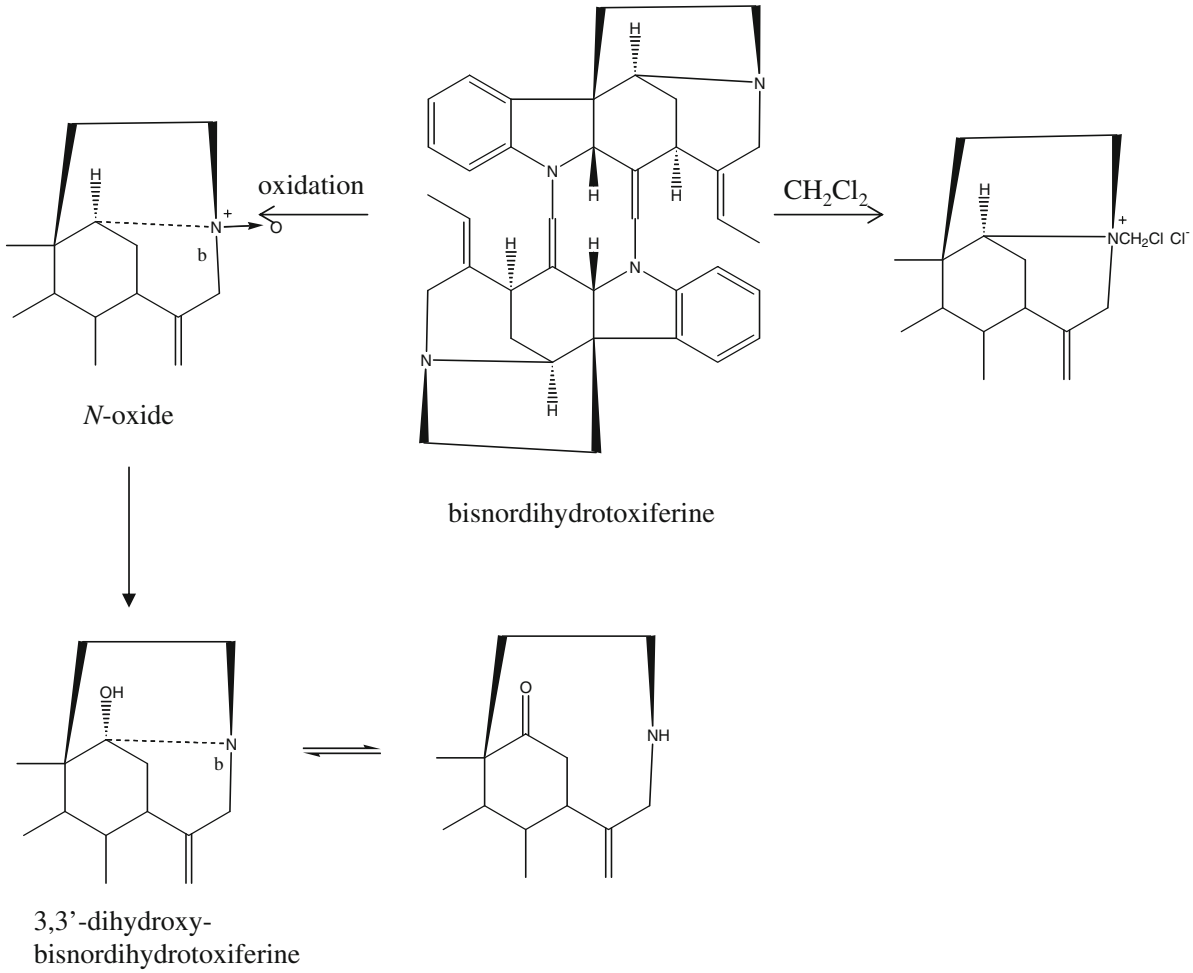


Fig. 6 Possible oxidation products and artifacts of bisnordihydrotoxiferine (adapted from Verpoorte and Baerheim Svendsen 1976, 1978)

Obviously the choice of solvents is extremely important and the choice should be for simple solvents with least possible chance of artifact formation such as water and alcohols, though alcohols may react with, for example, carbinolamines, or form esters. In case of methanol as extraction solvent it will be difficult to know whether a methoxy group is naturally occurring or comes from the solvent. Ethanol does not have this problem as ethoxy groups in nature are rare.

Novel extraction methods are supercritical fluid extraction (Kim et al. 2001), with e.g. carbon dioxide, or the use of ionic liquids (reviewed by Marsh et al. 2004; Seddon 1997). However, there is little experience yet with the use of such solvents in high-through put analytical methods as required for metabolomics.

To have the broadest spectrum of compounds, crude extracts are used as such in the metabolomic analysis. In case of a targeted approach one or more clean-up steps can be applied, such as liquid-liquid partitioning or solid-phase extraction. We also successfully applied pre-fractionation using centrifugal partition chromatography (Halim and Verpoorte, unpublished results) but these methods are outside the scope of this review.

Consequently, in the field of metabolomics there is no perfect method to extract all metabolites without any artifact formation and chemical degradation. In order to minimize the problems of extraction methods our group has been developed two kinds of extraction methods for NMR-based metabolomics (Choi et al.

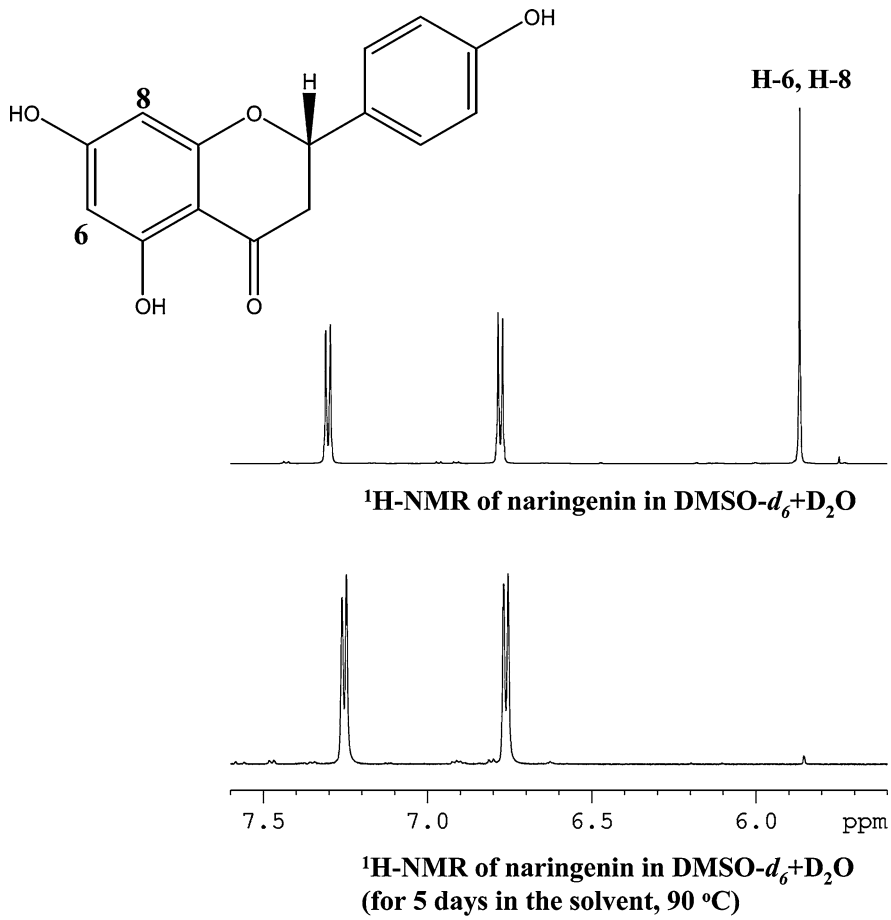


Fig. 7 Chemical structure of naringenin and ^{1}H-NMR spectra of naringenin in DMSO-d_6 and D_2O

2006b). The simple protocols for the methods are as follows.

(1) Two-phase extraction method for NMR analysis

- Add 2 ml of chloroform, 1 ml of methanol, and 1 ml of water to 50–100 mg of plant material.
- Ultrasonicate for 5 min. and centrifuge at 1308*g* for 20 min at 4°C.
- Take upper phase of methanol-water fraction and lower phase of chloroform-methanol fraction.
- Evaporate both fraction using rotary evaporator and add 1 ml of chloroform-*d* to chloroform–methanol fraction and 1 ml of methanol-d_4 and H_2O-d_2 (KH_2PO4 buffer, pH 6.0) to methanol-water fraction for NMR measurement.

(2) Single solvent extraction method for NMR metabolomics

- Add 0.75 ml of methanol-d_4 and 0.75 ml of H_2O-d_2 (KH_2PO_4 buffer, pH 6.0) to 50–100 mg of plant materials.
- Vortex for 1 min.
- Ultrasonication for 10 min.
- Centrifuge at 10968*g* for 10 min at r. t.
- Transfer 800 µl supernatant to NMR tube.

Metabolomics in dynamic system

As mentioned above each method has its advantages and disadvantages (see Table 1), we will not further discuss this. Rather there is more general point that needs to be considered in connection with the way experiments are made. Transcriptomics, proteomics, and metabolomics are measurements made at a selected time point, they are like a snap shot. However, in a dynamic system this may cause problems in the interpretation of the results. For example, is a high level of a certain compound due to

storage of an end product of *de-novo* biosynthesis, or of a large flux through a pathway? For real evaluation of a biological process a dynamic film rather than a static picture is needed. This is where fluxomics comes into the picture: measuring the flux through pathways should be the final goal. In measuring the metabolic changes one should also take into account the time scale of different biological processes (Table 2).

Our work on the biosynthesis of salicylate (SA) may illustrate this. *Catharanthus roseus* cell cultures produce 2,3-dihydroxybenzoic acid (DHBA) after elicitation with biotic elicitors in quite high levels (Moreno et al. 1994). Also, salicylate (SA) is produced in small amounts, as it is a signal compound for systemic acquired resistance. Because of the importance of SA as signal compound in plant defense, much research has been done on its biosynthesis. However the pathway(s) involved is still under debate. Originally SA was thought to be derived from phenylalanine but the complete pathway has not yet been proven. In microorganisms SA is made along another pathway, in which the formation of isochorismate is the first step (Verberne et al. 1999). We have shown that this microbial pathway can be expressed in plants (Verberne et al. 2000), we also have shown in a retrobiosynthetic study that DHBA in *C. roseus* is derived from this isochorismate pathway (Budi Muljono et al. 2002a, b), and the gene encoding the first step, isochorismate synthase (ICS) was cloned from *C. roseus* (Van Tegelen et al. 1999a, b). Recent elicitation experiments with new cell lines of *C. roseus* fed with 1-^{13}C-labeled glucose showed that the ^{13}C-labeling pattern of SA is in line with the isochorismate pathway. In the same experiment the DHBA produced was analyzed, confirming the earlier finding that also this compound is derived from the isochorismate pathway. However, the labeling percentages in SA and DHBA which were isolated 24 h after elicitation were different (Mustafa 2007). At $t = 0$, the time point when the cells were elicited with a *Pythium aphanidermatum* preparation, both SA and DHBA are below the detection limit. As the ICS activity is induced by the elicitation, one might draw the conclusion that both compounds are coming from the same pathway. But the different labeling means that it is a more complex process. Following the SA and DHBA production through time after elicitation, it is clear that SA is already detectable at an earlier stage, whereas most of the DHBA is formed later (Budi Muljono et al. 2002a, b). With other words, SA might be formed by already-present enzymatic machinery, whereas DHBA is the result from a pathway that is induced on the level of the genes. In our opinion this is a clear example showing the need to study fluxes in studies of plant defense systems. Recent studies by Saito's group (Yamazaki et al. 2004) using novel software for translating percentages of labels into the involvement of various primary metabolite pathways for the production of secondary metabolites is an important step in this direction.

Table 2 Time scale various biological processes

Level	Example	Processing time (s)
Molecular	Ion channel gating	10^{-6}
Cellular	Mitosis	10^{3}
Plant-elicitor	pH	10^{2}
	Expression or regulatory gene	10^{3}
	Expression of structural gene	10^{4}
	Biosynthesis proteins	5×10^{4}
	Biosynthesis phytoalexins	10^{5}
Plant development	Aging	10^{7}

Conclusions and perspectives

The analysis of a plant metabolome is a major challenge. No single method is capable of analyzing all compounds in a plant but it is clear that with the unbiased approach of measuring as many different metabolites as possible (NMR or MS), a lot of novel information can be obtained. The information might be more completed by the use of a more targeted analytical approach (LC-MS, GC-MS). The classical targeted phytochemical methods will thus play a major role in further studies of the results of metabolomic studies. The various multivariate data analysis methods which are available are powerful tools to extract the relevant information out of the huge data sets available. The strength of metabolomics should be that metabolomic data collected many years ago can be used again and again to compare these with the latest results. This requires rigidly standardized protocols for the analysis and for the storage of the original analytical data for future use. The integration with transcriptomics data will be a major challenge for the bioinformatics. Gene induction through external

signals and subsequent production of secondary metabolites, each with different time scales, form a quite complex system of interacting networks. It involves 4 dimensions, 3 of space and one of time. The space concerns for example the cellular and subcellular compartmentation; time concerns for example diurnal and developmental changes in the metabolism. In fact one may ask the question "what is the metabolome of a plant?". Every single cell has its individual metabolome, and for example in a leaf many different cell types occur. These micrometabolomes should be the real target, which requires single cell analysis, or methods such as laser dissection to collect cells of the same type from a tissue (Day et al. 2005). But eventually it might turn out that one needs to analyze nanometabolomes, the metabolomes occurring in the various cellular compartments (e.g. nuclei, plastids, mitochondria, vesicles, vacuoles) to unravel the biochemistry of the cell.

To better understand biochemical processes fluxomics might be more suited than metabolomics, which only shows the situation at one single time point. For static traits of course metabolomics are suitable, e.g. quality control of botanicals. Also constitutively expressed traits of plants can be identified by means of metabolomics, e.g. resistance by comparison the metabolome of resistant and non-resistant cultivars.

In the coming years there will be a major role for phytochemistry to further improve the analysis of the metabolome and to identify all unknown compounds. Collaboration with bioinformatics will be crucial to better plan experiments and to extract all possible information from the huge amount of data that is produced by the analysis of plant metabolomes. The importance of data storage, in our view, requires a change in the way of publishing experimental work. With the present day possibilities each experimental study should be published with the full raw experimental data as a supplement which can be used by others for data mining, and comparison with other experiments, thus avoiding any bias in the interpretation by producing graphs or tables.

References

Baerheim Svendsen A, Karlsen J (1966) Direkte Gaschromatografie der Bestandteile der Aetherische Oelen in Pflanzenmaterial. Planta Medica 14:376

Baerheim Svendsen A, Verpoorte R (1983) Chromatography of alkaloids, part I, Thin-layer Chromatography, vol 23A. Chromatography Library, Elsevier, Amsterdam, 534 pp

Bisset NB, Choudhury AK (1974) Alkaloids and iridoids from *Strychnos nux-vomica* fruits. Phytochemistry 13:265–269

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. Nat Genet 29:365–371

Breitling R, Ritchie S, Goodenowe D, Stewart ML, Barrett MP (2006) *Ab initio* prediction of metabolic networks using Fourier transform mass spectrometry data. Metabolomics 2:155–164

Budi Muljono RA, Scheffer JJC, Verpoorte R (2002a) Isochorismate is an intermediate in 2,3-dihydroxybenzoic acid biosynthesis in *Catharanthus roseus* cell cultures. Plant Physiol Biochem 40:231–234

Budi Muljono RA, Talou JR, Scheffer JJC, Verpoorte R (2002b) Biosynthesis of 2,3-dihydroxybenzoic acid in transgenic *Catharanthus roseus* cell cultures overexpressing isochorismate synthase. Heterocycles 56: 341–352

Choi YH, Kim HK, Linthorst HJM, Hollander JG, Lefeber AWM, Erkelens C, Nuzillard J-M, Verpoorte R (2006a) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. J Nat Prod 69: 742–748

Choi YH, Kim HK, Verpoorte R (2006b) Metabolomics. In: Kayser O, Quax W (eds) Medicinal plant biotechnology from basic research to industrial applications. Wiley-VCH, Weinheim, pp 9–28

Cordell GA (1974) The biosynthesis of indole alkaloids. Lloydia 37:219–98

Dallüge J, Beens J, Brinkman UAT (2003) Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. J Chromatogr A 1000:69–108

Day RC, Grossniklaus U, Macknight RC (2005) Be more specific! Laser-assisted microdissection of plant cells. Trends Plant Sci 10:397–406

Fiehn O (2003) Metabolic networks of *Cucurbita maxima* phloem. Phytochemistry 62:75–886

Gobey J, Cole M, Janiszewski J, Covey T, Chau T, Kovarik P, Corr J (2005) Characterization and performance of MALDI on a triple quadruple mass spectrometer for analysis and quantification of small molecules. Anal Chem 77:5463–5654

Hanson KR (1965) Chlorogenic acid biosynthesis. Chemical synthesis and properties of the mono-*O*-cinnamoylquinic acids. Biochemistry 4:2719–2731

Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2:696–701

Kim J, Choi YH, Yoo K-P (2001) Supercritical fluid extraction of alkaloids. In: Pelletier SW (ed) Alkaloids: chemical and biological perspectives. Pergamon Press, Amsterdam, pp 415–432

Kopka J (2006) Gas chromatography mass spectrometry. In: Saito K, Dixon RA, Willmitzer L (eds) Biotechnology in

agriculture and forestry 57 plant metabolomics. Springer, Leipzig, pp 3–20

Lay JO Jr, Borgmann S, Liyanage R, Wilkins CL (2006) Problems with "Omics". Trends Anal Chem 25: 1046–1056

Liang Y-S, Lefeber AWM, Erkelens C, Kim HK, Choi YH, Verpoorte R (2006) Identification of phenylpropanoids in *Brassica rapa* leaves treated with methyl jasmonate using two-dimensional nuclear magnetic resonance spectroscopy. J Chromatogr A 1112:148–155

Lindon JC, Keun HC, Ebbels TMD, Pearce JMT, Holmes E, Nicholson JK (2005) The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. Pharmacogenomics 6:691–699

Looser R, Krotzky A, Trethewey RN (2005) Metabolite profiling with GC-MS and LC-MS. In: Vaidyanathan S, Harrigan G, Goodacre R (eds) Metabolome analyses strategies for systems biology. Springer, New York, pp 103–118

Marsh KN, Boxall JA, Lichtenthaler R (2004) Room temperature ionic liquids and their mixtures – a review. Fluid Phase Equilib 219:93–98

Maudens KE, Wille SMR, Lambert WE (2007) Traces of phosgene in chloroform: consequences for extraction of anthracyclines. J Chromatogr B 848:384–390

Moco S, Bino RJ, Vorst O, Verhoeven HA, De Groot J, Van Beek TA, Vervoort J, De Vos CHR (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol 141:1205–1218

Moreno PRH, Van der Heijden R, Verpoorte R (1994) Elicitor-mediated induction of isochorismate synthase and accumulation of 2,3-dihydroxybenzoic acid in *Catharanthus roseus* cell suspension and shoot cultures. Plant Cell Rep 14:188–191

Mustafa NR (2007) Retrobiosynthetic study of salicylic acid in *Catharanthus roseus* cell suspension cultures. PhD thesis, Leiden University, Leiden

Pauli GF, Jaki BU, Lankin DC (2005) Quantitative ^{1}H NMR: development and potential of a method for natural products analysis. J Nat Prod 68:133–149

Reily MD, Lindon JC (2005) NMR spectroscopy: principles and instrumentation. In: Robertson DG, Lindon J, Nicholson JK, Holmes E (eds) Metabonomics in toxicity assessment. CRC Press, Boca Raton, pp 75–104

Ryan D, Robards K (2006) Metabolomics: the greatest omics of them all? Anal Chem 78:7954–7958

Santos FJ, Galceran MT (2003) Modern developments in gas chromatography-mass spectrometry-based environmental analysis. J Chromatogr A 1000:125–151

Schripsema J, Verpoorte R (1991) Investigation of extracts of plant cell cultures by ^{1}H NMR. Phytochem Anal 2: 155–162

Schripsema J, Erkelens C, Verpoorte R (1991) Intra- and extracellular carbohydrates in plant cell cultures investigated by ^{1}H-NMR. Plant Cell Rep 9:527–530

Seddon KR (1997) Ionic liquids for clean technology. J Chem Tech Biotechnol 68:351–356

Snyder LR (1974) Role of the solvent in liquid-solid chromatography – a review. Anal Chem 46:1384–1393

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

Taylor J, King RD, Altmann T, Fiehn O (2002) Application to metabolomics to plant genotype discrimination using statistics and machine learning. Plant J 18:241–248

Van Tegelen LJP, Bongaerts RJM, Croes AF, Verpoorte R, Wullems GJ (1999a) Isochorismate synthase isoforms from elicited cell suspension cultures of *Rubia tinctorum*. Phytochemistry 51:263–269

Van Tegelen LJP, Moreno PRH, Croes AF, Verpoorte R, Wullems GJ (1999b) Purification and cDNA cloning of Isochorismate synthase from elicited cell suspension cultures of *Catharanthus roseus*. Plant Physiol 119:705–712

Verberne MC, Budi Muljono RA, Verpoorte R (1999) Salicylic acid biosynthesis. In: Hooykaas PJJ, Hall MA, Libbenga KR (eds) Biochemistry and molecular biology of plant hormones. New comprehensive biochemistry, vol 33. Elsevier, Amsterdam, pp 295–312

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

Verberne MC, Brouwer N, Delbianco F, Linthorst HJM, Bol JF, Verpoorte R (2002) Method for the extraction of the volatile compound salicylic acid from tobacco leaf material. Phytochem Anal 13:45–50

Verpoorte R, Baerheim Svendsen A (1976) The alkaloids of *Strychnos dolichothyrsa*. J Nat Prod 39:357–362

Verpoorte R, Baerheim Svendsen A (1978) Alkaloids from *Strychnos dolichothyrsa*. J Pharm Sci 67:171–174

Verpoorte R, Baerheim Svendsen A (1984) Chromatography of alkaloids, part II. Gas and liquid chromatography, vol 23B. Chromatography Library, Elsevier, Amsterdam, 534 pp

Verpoorte R, Choi YH, Kim HK (2007) NMR-based metabolomics at work in phytochemistry. Phytochem Rev 6: 3–14

Wagner C, Sefkow M, Kopka J (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. Phytochemisty 62:887–900

Xianlin Han, Gross RW (2003) Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. J Lipid Res 44:1071–1079

Yamazaki Y, Kitajima M, Arita M, Takayama H, Sudo H, Yamazaki M, Aimi N, Saito K (2004) Biosynthesis of Camptothecin in silico and in vivo tracer study from [1-^{13}C] Glucose. Plant Physiol 134:161–170