Metabolomics for bioactivity assessment of natural products

Yuliana, N.D.; Khatib, A.; Choi, Y.H.; Verpoorte, R.

PHYTOTHERAPY RESEARCH
Phytother. Res. **25**: 157–169 (2011)
Published online 23 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/ptr.3258

REVIEW

Metabolomics for Bioactivity Assessment of Natural Products

Nancy Dewi Yuliana,[1,2] Alfi Khatib,[2,3] Young Hae Choi[1]* and Robert Verpoorte[1]

[1]Division of Pharmacognosy, Section of Metabolomics, Institute of Biology, Leiden University, Leiden, The Netherlands
[2]Department of Food Science and Technology, Bogor Agricultural University, Bogor, Indonesia
[3]Center of Excellence for Food Safety Research, Faculty of Food Science and Technology, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

Natural products historically have been a rich source of lead molecules in drug discovery, based on their capability to create unique and diverse chemical structures. However, it is also true that the vast number of metabolites typically present in natural products and their huge dynamic range results in the loss of many possibly bioactive natural compounds, becoming an inextricable obstacle for drug development. Recently, new strategies which favour a holistic approach as opposed to the traditional reductionist methods used previously, have been introduced with the purpose of overcoming the bottlenecks in natural product research. This approach is based on the application of new technologies, including metabolomics, for example. Metabolomics allows a systematic study of a complex mixture such as a phytochemical preparation, which can be linked to observations obtained through biological testing systems without the need for isolating active principles. This may put drug discovery from natural products back in the limelight again. In this review paper, the description of some examples of successful metabolomics applications in several important fields related to drug discovery from natural sources aims at raising the potential of metabolomics in reducing the gap between natural products (NP) and modern drug discovery demand. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: natural product; drug discovery; bioactivity; metabolomics.

INTRODUCTION

In recent years the development of novel drugs has become increasingly difficult, expensive and time consuming. It is estimated that more than 800 million USD and an average of 14.2 years are needed for a novel drug application (NDA) approval. Both factors are rapidly increasing with time while the output of newly launched drugs has dropped (Butler, 2004; Dickson and Gagnon, 2004). The introduction of high throughput screening (HTS), a process which allows the testing of a large number of samples on certain molecular targets, was expected to shorten the discovery route. To fully use its high potential, HTS requires a large number of compounds to screen, in the order of thousands of samples per day (Mishra *et al*., 2008). In order to meet this demand, high throughput synthesis and combinatorial chemistry were developed. Concomitantly, further development in the field of structural biology and genomics, e.g. cloning and expression technology resulting in the production of various purified human receptors and enzymes which could be studied at a molecular level, has produced a sharp increase in the number of possible molecular targets, demanding more drug-based libraries of a wide chemical diversity to screen (Koehn and Carter, 2005). The combination of these new strategies seemed very powerful and the delivery of a large number of novel leads was expected (Ortholand and Ganesan, 2004). However, these expectations were shown to be greatly overrated as the number of new active compounds that made it to the market in the past two decades reached its lowest level ever in 2002 with an insignificant improvement in 2003 (Newman and Cragg, 2007; Newman *et al*., 2003; Ortholand and Ganesan, 2004). Sorafenib, a multikinase inhibitor indicated for advanced renal cancer, is the single synthetic drug derived from combinatorial chemistry approved by Food and Drug Administration (FDA) for clinical use (Newman and Cragg, 2007). The fact that this increased HTS screening capacity and the number of novel molecular targets has failed to produce an increase of new lead compounds poses the question as to whether this is due to limitations in the innovative aspects of the discovery process itself, e.g. generation of compounds of limited chemical diversity, or because the potential of natural products (NP) has been de-emphasized in the past few years (Butler, 2004).

The possibility of combining NP extract libraries with HTS also seems interesting, though not always an easy task. Several reasons for the incompatibility of NP and HTS have been described (Butler, 2004; Ortholand and Ganesan, 2004). The complexity of a NP extract which is a complex mixture of mostly uncharacterized compounds that may mask the bioactivity of target compounds or cause false-positive results is the most common hurdle. The large dynamic range (e.g. the

* Correspondence to: Young Hae Choi, Institute of Biology, Division of Pharmacognosy, Leiden University, Einsteinweg 55, P.O. Box 9502, Leiden, The Netherlands, 2300 RA.
Email: y.choi@chem.leidenuniv.nl

Received 02 February 2010
Revised 30 May 2010
Accepted 31 May 2010

difference in the levels of major compounds compared with that of minor compounds, differences in polarity, in boiling or evaporating points, etc.) which may affect a compound's solubility makes the integration of NP and HTS more complicated. The differences between the areas where the plants are grown, the methods of production and processing of plants are also important since these might affect the chemical profile of the NP (Jahangir *et al*., 2008; Wang *et al*., 2005). The intellectual property issue in natural product outsourcing is another problem. Thus, the drug discovery process using plant sources was almost abandoned (Dickson and Gagnon, 2004; McChesney *et al*., 2007). Companies such as Glaxo Wellcome, Smith Kline & French, and Pfizer gradually eliminated their NP screening programmes while others, such as Merck and Novartis, continued their exploration but focused on microbial sources (Ortholand and Ganesan, 2004).

The failure to increase the number of new drugs delivered to the market has placed the drug discovery programmes of pharmaceutical companies under scrutiny. Looking back to the past two decades, of the 877 novel medicines developed between 1981 and 2002, 6% were NP, 27% were NP derivatives and 16% were synthetic products developed on the basis of a NP model (Newman *et al*., 2003), indicating that NP is still an important source of novel leads for therapeutic drugs. This situation has led to a renewed interest in NP in drug discovery programmes. Therefore, it is a challenging task to make NP available for the drug discovery arena.

Most NP drug discovery work follows 'the silver bullet' approach as is usual in present Western medicine, in which the search is focused on single compounds with a particular biological activity (single target–single compound paradigm) tested at the molecular level (e.g. receptor binding inhibition or enzyme screening) (Verpoorte *et al*., 2005). This approach which has been responsible for the successful delivery of many potent drugs in the past, such as the statins to lower cholesterol level and huperzine A for Alzheimer's disease (Li *et al*., 2008b), is not actually the best approach for NP studies since the presence of synergism and pro-drugs in an extract can be misleading (Verpoorte *et al*., 2005). It is also one of the reasons for the long time line and high costs that the pharmaceutical companies have to deal with when working with NP. Shifting the silver bullet paradigm to a more holistic approach, which is actually the philosophy of some traditional medicine such as Traditional Chinese Medicine (TCM) and Ayurveda, is now evolving into a new trend (Ulrich-Merzenich *et al*., 2007; Verpoorte *et al*., 2005; Wang *et al*., 2005). In a holistic approach, a whole herbal preparation, which consists of numerous multi-components, is viewed as one active entity with effects on multiple targets in living organisms (Li *et al*., 2008b; Ulrich-Merzenich *et al*., 2007). For such an approach, the response of the whole system needs to be observed with all possible tools such as clinical trials or experimental animal models, where both classic physiological observations (e.g. blood pressure, analgesic activity, sedative) and modern molecular observation (gene expression, proteome, metabolome) can be used to obtain a better insight into different changes in the organism. This is known as a systems biology approach (Verpoorte *et al*., 2005).

Among these 'omics' methods, metabolomics, a relatively new member, is thought to be the most informative as regards understanding systems biology, since it reflects the genotype (Sumner *et al*., 2003). Therefore it is a valuable tool to use in the provision of the information required in this new approach which is particularly interesting for the study of medicinal plants. It thus offers new opportunities for NP in the drug discovery arena.

This review will discuss the chemical diversity of compounds present in NP compared with synthetic compounds, the obstacles which might limit NP integration with the new emerging drug discovery technologies, the shift in tendency from a reductionist to a holistic approach in NP research and how metabolomics can help to narrow the gap between NP and the demand of modern drug discovery, and help to turn NP and medicinal plants into a unique drug source.

NATURAL PRODUCTS AS A DRUG SOURCE: THEIR ADVANTAGES AND LIMITATIONS

Factors that distinguish NP from synthetic chemicals include their incomparable molecular diversity and biological functionality (Nisbet and Moore, 1997). Starting with only a few building blocks of NP provided by the biosynthetic pathways, the action of various 'decorating' enzymes provide a large diversity of end products with different functionalities (Nisbet and Moore, 1997; Pieters and Vlietinck, 2005; Verpoorte, 1998). Compared with products resulting from combinatorial chemistry libraries which usually lack chirality and structure rigidity, two important requirements for pharmaceutical drug-like properties (Nisbet and Moore, 1997), NP have privileged structures which are able to interact with proteins and other biological targets for specific purposes (Koehn and Carter, 2005).

To address the question of how the unique structural properties of NP are able to complement the chemical space of synthetic compounds, a comparison of the major structural differences between combinatorial compounds, drugs and NP molecules has been made. It is obvious that current commercial drugs and NP share more similarities in several topological pharmacophore patterns. The important differences between NP and synthetic compounds include the number of chiral centres, the prevalence of aromatic rings, the presence of complex ring systems and the degree of saturation of the molecule as well as the number and ratios of different heteroatoms (Feher and Schmidt, 2003). Natural products have a higher number of chiral centres, a higher number of oxygen atoms instead of nitrogen, sulfur and halogen containing groups present than in synthetic compounds. Natural products have a lower ratio of aromatic ring atoms to total heavy atoms, and a higher number of solvated hydrogen bond donors and acceptor functional groups. These characteristics are favourable attributes for a drug candidate profile (Feher and Schmidt, 2003; Henkel *et al*., 1999; Newman and Cragg, 2007; Newman *et al*., 2003). With these unique properties, in general, NP actually have more drug-like properties than synthetic compounds (Feher and Schmidt, 2003; Koehn, 2008).

However, a NP-based drug discovery project also poses some challenges, mostly connected with their presence within a complex matrix with all kinds of

active compounds (Bindseil *et al*., 2001; Koehn, 2008; Koehn and Carter, 2005; McChesney *et al*., 2007; Ortholand and Ganesan, 2004; Pieters and Vlietinck, 2005; Schmid *et al*., 1999):

1. The first challenge is dereplication, which is the rapid identification of known and unknown compounds from partially purified mixtures. In a number of test systems, common plant products which have been found to be active, e.g. unsaturated fatty acids on adenosine and GABA receptors or tannins on enzyme-based bioassays, have to be identified at the very beginning of the study of an active extract.
2. Interaction between metabolites present in extracts. Extracts consist of mixtures of many compounds requiring an elaborate purification to isolate new pure compounds. Another problem of complex extracts is the presence of antagonism or synergism which might occur between components in a mixture, posing a major intellectual challenge which can be painstakingly time consuming.
3. Structure determination by spectroscopic methods.
4. Large scale production of the active component is sometimes an obstacle as the identified NP lead might be a minor compound of the extract, or the supply of the material for extraction might be limited.
5. Sourcing of authenticated plants, a very important issue for quality control. For example, some plants have congener substitutes which have an active principal content which may be different to the original one.

Last but not least, biodiversity and intellectual property issues can be another complicating factor related to this context.

PREVIOUS EFFORTS TO INCREASE THE NUMBER OF COMPOUNDS FOR DRUG CANDIDATES

Because of the demand to speed up the drug discovery timeline and to reduce costs, the *in vitro* screening system has become the preferred method. The use of animal tests or isolated organs was abandoned for more specific testing targets at a molecular level (reductionist mode) (Verpoorte *et al*., 2006). Following this trend, there are two crucial prerequisites to start the drug discovery process: a large high quality library of candidates to be screened, and a reliable and sensitive assay to screen the activity. Combinatorial chemistry was developed to address the first need (Williard *et al*., 1996). Based on the total chemical space, this technology is theoretically able to provide 10^{30}–10^{100}compounds (Macarron, 2006; Myers, 1997).

The second need was met by high throughput screening (HTS) as the response to the availability of hundreds to thousands of compounds resulting from combinatorial chemistry. HTS is a fast and cost-effective technology designed to select drug candidates (Bleicher *et al*., 2003; Gómez-Hens and Aguilar-Caballos, 2007). It is defined as an automatic approach in drug discovery which is characterized by miniaturization to allow the use of minute quantities of samples and reagent. Continuous innovations keep taking place to maximize the quantity of screened compounds and to minimize the cost. The miniaturization of a screening assay aimed at increasing HTS capacity is achieved by increasing the screening plate density and minimizing screening volume (Landro *et al*., 2000). The system which enables the screening of 120000–240000 compounds per day has been created by the use of high density 1536-well plates. Higher capacity plates of 3456 wells are also available but are used rarely due to the limitation determined by the low liquid volume delivery (Bajorath, 2002; Smith, 2002; Williams, 2004). Compounds found to be active in HTS are called 'hits', but the distance from a hit to a lead is very great. This requires more extensive screening for other activities, including toxicological evaluations. In this process similar compounds will be produced to optimize the activity profile.

However, compounds obtained from the screening are not always (physico) chemically and pharmacologically suitable for further development as a lead. While the *in vitro* profile is maybe very promising, ADMET (absorption, distribution, metabolism, elimination, toxicity) properties are often poor (Bleicher *et al*., 2003). In most cases the compound will require modifications to become a lead and drug candidate. Thus, the crucial bottleneck of lead development is not only hit identification, but the potential of the candidate to become a new drug (Bleicher *et al*., 2003).

Therefore it is necessary not to focus only on quantity but also to consider the quality of both library and assays. One suggestion made on this matter was to introduce simple ADMET parameters at the early stage of screening and the more sophisticated ones at a later stage (Walters and Namchuk, 2003). The appropriate application of virtual screening as a complement to a screening campaign to enrich the hit rate and reduce the number of false-positives has been suggested as well (Bajorath, 2002; Bleicher *et al*., 2003; Entzeroth, 2003; Smith, 2002). Most virtual screening is guided by the structure of the target protein and 3-dimensional descriptors of small molecular structures that fit into the protein active site. Further approaches such as the identification of active compounds by clustering and partitioning techniques, compound filtering to eliminate undesired compounds (e.g. toxicant and groups with poor aqueous solubility and passive absorption) can also be applied (Bajorath, 2002). Understanding the mechanism of action of a biological target is another approach. Several techniques have been reported, such as immobilization and stabilization of the target protein upon its ligand binding, as well as the use of nuclear magnetic resonance (NMR) spectroscopy coupled to other analytical methods to learn about the drug–protein interaction (Entzeroth, 2003). For the targets without sufficient bio-structural information, such as in the area of G-protein coupled receptors (GPCRs), a ligand motif-based library design approach can be applied (Bleicher *et al*., 2003).

The aforementioned efforts, however, did not successfully fulfil the high expectation of increasing productivity. The tremendous amount of money spent was not counterbalanced by the number of new approved drugs (Smith, 2002). However, all these methods will result in all companies having the same basic information on targets and probably ending up with the same compound, so the chance for real innovation with this

approach will be limited. This has encouraged reconsidering NP as a drug source.

EFFORTS TO COMBINE NATURAL PRODUCTS WITH NEW TECHNOLOGIES

Many studies reported possible approaches which could potentially help to overcome the above mentioned problems in the NP drug discovery process. Solutions to address the incompatibility of NP extracts with HTS technology were proposed, including the generation of a high quality NP compound library, the improvement of fractionation methods, and of dereplication and identification steps.

Constructing a high quality natural product library

Several strategies to construct better NP libraries and to improve methods to screen these libraries have been reviewed. The advantages and limitations of a crude extract library, pre-fractionated library, or pure compound library have been described (Koehn, 2008). Crude extract libraries are simple to make, the obstacles appearing in the identification and dereplication steps. A pre-fractionated library can be effective for overcoming these hurdles since the less complex fractions require fewer purification steps and the interference of unwanted compounds is reduced, while more investment is required both for fraction preparation and for the HTS assay, since several samples are obtained from a single extract. The pure compound library is preferred since it is able to move forward directly after finding a hit.

The strategy employed to build a fraction library as a compound source for HTS from Chinese herbal formulations reported to treat atherosclerosis has been described. Medium pressure liquid chromatography with a gradient solvent system followed by reverse phase semi-preparative HPLC was used as a filtering method to remove unfavourable compounds lacking desirable ADME properties and to prepare semi-crude extracts with different polarity. The fractions were characterized and identified by LC-DAD-ESI-MS. A component detection algorithm (CODA) was used to reduce noise and background of the chromatograms which then were presented as a histogram describing compound molecular distribution. Small molecules (less than 500 Da) were characterized by retention time, molecular weight, UV absorbance and crude drug origin. In this way, 36 major compounds of each fraction were identified and after cell based multiple bioassay screening, validation and dereplication steps, nine hits were found (Liu *et al*., 2008).

Improving fractionation methods

Another study focused on sample fractionation using an automatic multiple solid phase extraction (SPE) method (Schmid *et al*., 1999). Several different solid phases were used, thus reducing the complexity of an extract from a microbial fermentation broth, leading to a more reliable and reproducible screening result.

Improving dereplication and identification steps

A fast dereplication and identification of novel active compounds can be achieved by direct coupling of the bioassay to liquid chromatography (Bajorath, 2002; van Elswijk and Irth, 2002). An example is the post-column acetylcholine esterase (AChE) inhibitor assay. After centrifugal partition chromatography (CPC) prefractionation and bioactivity testing, the most active *Narcissus* 'Sir Winston Churchill' fraction was injected into an online biochemical detection system coupling HPLC-UV-MS for AChE inhibitor activity. This system enabled a fast dereplication, allowing the identification of known compounds. It proved very useful for the detection of novel active compounds in the presence of known active compounds. For example, besides the well known galantamine, another active compound, ungiminorine, was identified (Ingkaninan *et al*., 2000). A similar technique was applied to screen for natural phosphodiesterase and estrogen receptor ligands (Schenk *et al*., 2003; Schobel *et al*., 2001).

Sometimes, the information on the molecular weight and UV spectra provided by the combination of LC-DAD-MS is not sufficient for the identification of active compounds. The use of electrospray ionization Fourier Transform Ion Cyclotron Resonance MS (ESI-FTICR-MS) as a rapid and informative approach for direct affinity screening of bovine carbonic anhydrase II (bCAII) ligands from extracts was reported (Poulsen *et al*., 2006). The bCAII-ligand affinity is based on the non-covalent binding between the enzyme and the ligand. The electrospray ionization technique is known as a gentle ionization process that can transfer intact non-covalent complexes from the solution phase into the gas phase while the high resolution of the FTICR-MS allows the very accurate deduction of mass-to-charge ratio (m/z) of the protein–ligand complex from which even the molecular weight of this ligand can be calculated. This method was applied by separating small molecules from macromolecules by size exclusion chromatography (SEC), after the protein was added to the extract. This protein–ligand complex was then subjected to online SEC-ESI-FTICR-MS or directly infused into ESI-FTICR-MS, allowing the rapid determination of molecular weight of the active compound with an error below 0.08 Da. Bioactive compounds which were present in the mixture could be detected down to a level of 0.02% dry weight. HPLC coupled to an on-line postcolumn antioxidant detection system in combination with absorbance spectrometry and quadrupole time-of-flight high-resolution MS (QTOF-MS) allowed the identification of compounds responsible for the antioxidant capacity of raspberry (*Rubus idaeus*) and their level during raspberry ripening stages and in different cultivars and different genetic background (Beekwilder *et al*., 2005).

In a recent report, NMR based screening of bioactive NP and structure elucidation of the active compound in a single step was described (Politi *et al*., 2005). This study aimed at detecting ligands from plants that bind to a specific target protein, lectin. The interaction between a ligand present in the plant extract was detected by comparing the saturation transfer difference (STD) and NOESY spectra of the target protein with and without extract. The most important structural information was provided by two-dimensional NMR data such as

TOCSY, HSQC, HMBC and selective 1D-COSY. Several LC-NMR hyphenation techniques and their application to advance plant extract profiling have been described (Exarchou *et al*., 2005). The advantage and limitation of continuous flow LC-NMR, stop flow LC-NMR, LC-NMR/MS, LC-SPE-NMR, and application of cryogenic and microflow technology have also been reviewed. The performance of these LC-NMR coupling technologies depends on many factors such as the physical connection of LC-NMR, flow-through probe design and factors affecting NMR sensitivity (solvent compatibility, solvent suppression, the adequate sample volume). Practical examples of the application of various hyphenated LC-NMR techniques for rapid screening, dereplication and structure determination of active compounds from natural extract have been published (Wolfender *et al*., 2006).

Virtual screening

The application of virtual screening of known NP has been proposed (Rollinger *et al*., 2008). Natural products are still an incredible rich source of new drugs and because of the better understanding of protein–ligand interaction properties, NP virtual screening can constitute an alternative way to accelerate the drug discovery process. Many advantages have been described including higher capacity, a reduction in the number of experiments for testing and isolating compounds, the ability to calculate the possible interaction of compounds and targets, and the possibility of applying additional filters such as ADME and drug-like properties to improve the hit quality. Strategies to combine virtual screening and classical methods in NP exploration were proposed: (1) Direct screening of a 3D database of NP using available target information and a pharmacophore model of specific activity, followed by the isolation of hits once the results are confirmed by experimental data. (2) Extracts were roughly screened experimentally to select the active ones, or the specific activity was explored through an ethnopharmacological approach, the result of which was then virtually screened using a pharmacophore model of the target, followed by isolation.

METABOLOMICS

Metabolomics is one of the new 'omics' fields, joining genomics, transcriptomics and proteomics (Rochfort, 2005). The general aim of metabolomics is the qualitative and quantitative analysis of all metabolites (the metabolome), present in an organism, at a specific time and under specific influence (Colquhoun, 2007; Hall, 2006; Mendes *et al*., 2005; Ulrich-Merzenich *et al*., 2007; Verpoorte *et al*., 2007). A metabolome is the entire set of small molecules present in a cell, tissue, or organism under certain conditions and at a certain point in time (Ulrich-Merzenich *et al*., 2007).

Metabolome components can be viewed as gene expression end products, thus its quantitative and qualitative measurement may provide a broad insight of the biochemical status of an organism and can be used to monitor gene function (Sumner *et al*., 2003). The use of transcriptomics and proteomics profiles to predict gene function has been criticized because of limitations such as the fact that the changes in the transcriptome or proteome do not always correlate to biochemical phenotypes, or that the translated protein may or may not be enzymatically active (Nobeli and Thornton, 2006; Rochfort, 2005; Sumner *et al*., 2003). The lack of a database source may result in limited information since the identification of mRNA and proteins is based on sequence similarity or database matching. Therefore, among the 'omics' technologies, metabolomics can be viewed as the most functional approach (Sumner *et al*., 2003). The combined information obtained from genomics, proteomics and metabolomics should be able to provide an integrated understanding of a cell or organism (Sumner *et al*., 2003; Ulrich-Merzenich *et al*., 2007).

Detection method

Compared with proteins and RNA, secondary metabolites have a much more diverse atomic arrangement which results in broad variations in chemical and physical properties, such as molecular weight, polarity, solubility and volatility (Dunn and Ellis, 2005), and can be present in a large range of concentrations (Moco *et al*., 2006). This generates technological difficulties in the simultaneous profiling of all metabolites (Sumner *et al*., 2003). There are thus considerable challenges in metabolomics experiments such as sampling, sample preparation, instrumental analysis, and then in the interpretation of the tremendous amount and diversity of data arising from the metabolomics analysis (Bhalla *et al*., 2005; Rochfort, 2005), the construction of consolidated metabolite libraries and the development of metabolite-specific data management systems (Moco *et al*., 2006). Therefore, metabolomics based work still depends on further improvement of sampling, analysing and computational technologies (Bhalla *et al*., 2005; Rochfort, 2005).

The identification and quantification techniques applied to metabolomic based work have been comprehensively reviewed (Bhalla *et al*., 2005; Fiehn, 2002; Kell, 2004; Sumner *et al*., 2003; van der Greef and Smilde, 2005; Verpoorte *et al*., 2008). Many analytical methods with better reliability, robustness, selectivity and resolution have been developed in which instead of targeting a few specific compounds, the aim is to identify and quantitate multiple targets in order to obtain an overview of compound classes (Fiehn, 2002).

The progressive development of ionization and mass detection methods, and the ability to detect a wide range of molecular weights and polarities have become a strong point for the use of MS and tandem MS techniques in metabolomics (Verpoorte *et al*., 2008). GC coupled to MS (GC-MS) has been used to identify and elucidate the structures of 326 different compounds from *Arabidopsis thaliana* leaf (Fiehn *et al*., 2000) and to identify and classify 322 tomato volatiles from 94 genotypes according to their biochemical precursor similarity (Tikunov *et al*., 2005).

Liquid chromatography (LC) is an alternative tool particularly for non-volatile plant metabolites, though it requires more technical support than GC (Dunn and Ellis, 2005; Nobeli and Thornton, 2006). An open access database dedicated to LC-MS based metabolomics of

tomato fruit (*Solanum lycopersicum*) has been presented. A C18 base RP LC-photodiode array detection (PDA)-electrospray ionization (ESI)-quadrupole time-of-flight (QTOF)-MS method was developed for the detection and putative identification of predominantly semipolar secondary metabolites. Some previously unreported novel compounds from tomato fruit were identified by this means (Moco *et al*., 2006).

The use of laser desorption/ionization mass spectrometry on porous silica (DIOS) and cluster analysis to discriminate the metabolic footprints of mutant yeast and wild type yeast has been reported (Vaidyanathan *et al*., 2005). Compared with matrix-assisted laser desorption/ionization (MALDI) mass spectrometry, DIOS-MS is much simpler, allowing the analysis of small molecules in the absence of matrix ions with little or no fragmentation and better tolerability to moderate amounts of contaminants in the biological sample. Both methods are also especially effective for analysis of proteins and their derivatives (Thomas *et al*., 2001).

Although MS sensitivity is probably the best among the analytical methods, there are some limitations for its application in metabolomics. For example, the different sensitivity of the detection of molecules and its reproducibility are important issues. Because of different detector response, calibration curves are needed for each individual compound for absolute quantification, this being totally unrealistic considering the hundreds of signals that appear in a metabolomic profile. In this case, NMR spectrometry is the method of choice, in the absence of which only a relative quantification of the individual compounds in different experiments can be made (Verpoorte *et al*., 2008).

The application of NMR for metabolic profiling of plant tissue extracts has been reviewed by several authors (Colquhoun, 2007; Verpoorte *et al*., 2007; Ward *et al*., 2007). Reproducibility is the most important criteria for developing a metabolomics technology platform, and NMR is the most suited choice, even though its sensitivity is not as high as MS (1 μM–1 mM in NMR tube) (Colquhoun, 2007; Verpoorte *et al*., 2007). The simple and fast sample preparation, short measurement time, the availability of less automatization and advanced data analysis methods, plus the possibility to elucidate structures of known or unknown compounds in a complex mixture using advanced two-dimensional (2D) NMR methods, are further advantages (Colquhoun, 2007; Verpoorte *et al*., 2007).

Data processing

Data resulting from metabolomics-based work are typically high dimensional data which require multivariate data analysis (MVDA) methods for interpretation. Most metabolomics data analysis methods are based on the classification of samples into different groups (e.g. treatment, genotype), both by supervised (e.g. discriminant function analysis, artificial neural networks) or unsupervised type data analysis method, e.g. principal component analysis (PCA), hierarchical cluster analysis (HCA) (Nobeli and Thornton, 2006). It is also possible to use MVDA to make a regression modeling between two blocks of data, usually denoted as X and Y. In metabolomics based NP studies, X may represent signals from different metabolites present in plant extracts which are sampled at regular time intervals, and Y the responses (e.g. quality of product, bioactivity, yield). The model then can be used to predict Y from X which is obtained from new observations. The most common MVDA method for this modeling is the partial least square (PLS) method (Wold *et al*., 2001a). Some extended versions of PLS to improve the predictability power were developed recently, such as orthogonal signal correction (OSC), orthogonal-PLS (OPLS) and orthogonal-orthogonal-PLS (O2PLS) (Gabrielsson *et al*., 2006; Wold *et al*., 2001b).

VARIOUS APPLICATIONS OF METABOLOMICS IN BIOACTIVITY SCREENING OF NATURAL PRODUCT

Metabolomics to improve identification and dereplication steps

The reductionist approach as described above is still widely applied in many NP drug discovery projects, i.e. in bioassay guided isolation work. In this approach, the rapid dereplication of known actives and identification of novel actives is essential. Addressing this issue, some papers have described the way in which a metabolomics approach consisting of the use of various analytical methods followed by appropriate MVDA can be used to shorten the reductionist bioassay driven isolation route, especially in the identification and dereplication step.

In one study, the hierarchical cluster analysis of the chromatographic profiles of 500 lipidic extracts of bacteria, resulted in the identification of two new oligoprenylsesquiterpenes and two new *N*-(2-arylethyl) amides, though the bioactivity profile was not reported here. HCA was chosen instead of PCA since the number of principal components representing more than 50% of the variance was much too high for PCA. PCA is more suitable for a large data set of similar compounds, which was not the case in this project. The use of this clustering method was able to advance the dereplication step and to avoid redundant analyses for selection of strains containing new NP (Böröczky *et al*., 2006).

The most time-consuming step in metabolomics based work is the identification of metabolites in a prefractionated mixture. The availability of a compound spectral library is crucial for rapid identification. A free web-based tool containing data related to primary and secondary metabolites from various species based on NMR or MS data, the Madison Metabolomics Consortium Database ((MMCD) was developed (Cui *et al*., 2008). The sensitivity of MS and the feasibility of NMR spectroscopy, including the application of two dimensional (2D) NMR followed by MVDA application, make these techniques particularly well suited for the characterization of crude unfractionated NP. The analysis based on double quantum filtered correlation spectroscopy (DQF-COSY) was introduced to simplify the analysis of 2D NMR spectra and has demonstrated its utility to detect previously unreported terpenoid indole alkaloids from *Tolypocladium cylindrosporum* grown in different media. Graphical analysis of the obtained spectra through multiplicative stacking and graphical manipulation of the bitmap spectra allowed signals only

present in a specific media to be distinguished. Further structure elucidation was then performed by heteronuclear single quantum correlation (HSQC) and heteronuclear multiple bond correlation (HMBC) spectra analysis. This method is considerably more efficient than spectra subtraction since it results in a clear colour alteration and partially obsolescence of a common signal present in two spectra even if those are present in a distinct different concentration (Schroeder *et al*., 2007). The large size of data limits the application of this method. Three different clustering methods, i.e. tree clustering analysis (TCA), K-means clustering and multidimensional scaling, were used to cluster together partially purified plant and marine organism extracts which contain the same bioactive compound or groups of compounds. These clustering methods were chosen to overcome limitations of PCA to handle the data with a large number of clusters or subsets within the data (Pierens *et al*., 2005). Only a few of the secondary metabolites from *Myxococcus xanthus* have been investigated, in particular some antibiotics such as myxovirescin, saframycin and althiomycin. However, only one metabolite unique to *M. xanthus* was known, that is DK xanthanes. Intraspecific diversity in the metabolite profiles of 98 *Myxococcus xanthus* strains originated from 78 locations worldwide was rapidly explored by means of fast chromatographic separation (UPLC)-coupled ESI-TOF-MS and PCA. As a result, 37 candidates for novel compounds were identified but their bioactivities were not reported (Krug *et al*., 2008).

Metabolomics for quality control of phytomedicines

At present, the most popular application of metabolomics in NP research is for the quality control of phytomedicines. Different metabolite profiles may occur due to variations between species or variety, adulteration, environmental changes during growth and harvesting, post harvesting treatment, extraction and method of preparation. These factors may significantly alter the efficacy of phytomedicines prepared with these herbs (Wang *et al*., 2005). The application of ^{1}H NMR combined with multivariate data analysis to detect variations between different herbal preparations has been reported. The most common methods that have been applied for these ends are unsupervised principal component analysis (PCA) and supervised partial least square analysis/partial least square analysis with discriminant analysis (PLS/PLS-DA). These methods were applied to differentiate various chamomile (*Matricaria recutita* L.) preparations made with plants grown in different locations, the percentage of the desired plant part and extraction method (Wang *et al*., 2004). Discrimination between different *Cannabis sativa* cultivars based on the content of Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) was achieved (Choi *et al*., 2004), as well as discrimination between 11 *Ilex* species including three varieties (Choi *et al*., 2005). Similarly, discrimination between *Ephedra* species (Kim *et al*., 2005) and commercial feverfew (*Tanacetum parthenium*) preparations from different suppliers (Bailey *et al*., 2002) was obtained. The discrimination of commercial preparations of St John's Wort (*Hypericum perforatum*) from different batches of the same supplier, including variation of flavonoid content which have been previously related to the antidepressant activity was also reported (Rasmussen *et al*., 2006). An example of application of GC and GC-MS combined with multivariate data analysis to correlate artemisinin concentration with the growth stage of *Artemisia annua* L. and to distinguish between transgenic and wild type has been reported (Ma *et al*., 2008).

Metabolomics can also be used to detect adulterations of herbal preparations with similar species but with low levels of the desired active compound(s). ^{1}H NMR and PCA were used to verify that the antimalaria capsules produced by one specific company did not contain *Artemisia annua* as labeled, but *Artemisia afra* which has never been reported to contain the active compound artemisinin or any of its derivatives (van der Kooy *et al*., 2008). Instead of ^{1}H NMR, the use of HPLC fingerprints followed by application of PCA and PLS-DA was applied to examine the authenticity of the traditional Chinese medicines *Pericarpium Citri Reticulatae* and *Pericarpium Citri Reticulatae Viride* which contain *Citrus reticulata* '*Chachi*', *Citrus reticulata 'Dahongpao'* and *Citrus erythrosa Tanaka* as well that of commercial preparations containing a mixture of many tangerine peels which had been made with new tangerine mutations of uncertain quality (Yi *et al*., 2007).

Metabolomics to link chemical profile and bioactivity pattern of phytomedicines

As previously described, a reductionist approach focusing on a single active compound is necessary in some cases, but in the case of St John's Wort extract, where no single compound or even group of compounds have been found to be responsible for its activity, the possibility of synergism and pro-drugs is evident, and a holistic *in vivo* approach would be more valuable to guarantee its therapeutic efficacy (Ulrich-Merzenich *et al*., 2007; Verpoorte *et al*., 2005). Animal experiments and clinical trials are two classical ways to conduct *in vivo* experiments, but the time required and the costs involved make this option practically non-viable. Furthermore, if these tests are performed with extracts of uncertain composition, the results will invariably be unreliable or inconsistent. Recently new approaches are evolving which might help to get a better view of the mode of action of herbal or natural products and lead to the possibility of obtaining proof and guarantee of their pharmacological activity over different batches. These approaches are based on considering the NP holistically, i.e. on the bulk of chemical constituents rather than focusing on certain individual components or groups of compounds. On the other hand, the changes caused by a sample on the transcriptome, proteome and/or metabolome patterns can be compared with those observed after treatment with known drugs, rather than looking only at certain specific changes. Genomics, proteomics and metabolomics platforms will be crucial since they allow the identification of detectable protein/genes or compounds which might relate to the biological activity of complex extract (Ulrich-Merzenich *et al*., 2007). This may shift the paradigm in the development and application of complex plant/phytocompound mixtures in modern medicine (Shyur and Yang, 2008). Moreover, by measuring the effect of a complex extract administered to a living organism, either *in vitro* or *in vivo*, it is

possible to identify the compound or combination of compounds correlated with activity. Besides, new or unsuspected mode(s) of action and targets might also be revealed in this approach (Verpoorte *et al*., 2005).

Several papers discussed a metabonomics approach to study various aspects of natural products extracts such as bioavailability, toxicity and proof of efficacy. Metabonomics is defined as 'the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification' (Nicholson *et al*., 2002). This term is in general equivalent to the NMR spectroscopic analysis of biofluids to identify changes in the concentrations and patterns of endogenous metabolites induced by exogenic compounds (Nobeli and Thornton, 2006) but mostly used in medicinal or toxicological studies although sometimes the term metabolomics is also used.

A summary on the implementation of metabolomics, particularly NMR based metabolomics coupled to multivariate data analysis tools to detect links between the chemical profile of a herbal preparation and its bioactivity is provided in Fig. 1. A quick review of the literature reflects the enormous amount of studies that are carried out nowadays using this technology, some representative examples of which are summarized below.

Identification of active compounds and quantitative prediction of bioactivity. Chemometric and ^{1}H NMR analyses were used to discriminate *Artemisia annua* L. samples obtained from different sources. PCA assignment showed that artemisinin is the main discriminant factor for clustering. PLS analysis was applied to predict the quantitative antiplasmodial activity (IC_{50}) and toxicity ($ToxED_{50}$) value of different *A. annua* extracts. The predictive values were in close agreement with the actual experimental values (Bailey *et al*., 2004). The same approach was used to demonstrate the correlation between the spectral pattern and the opioid receptor binding properties of St. John's Wort (*Hypericum perforatum*) extracts prepared with different solvents (Roos *et al*., 2004). GC-MS-based metabolic profiling coupled with the antiinflammatory activity data of ginger (*Zingiber officinale*) from different sources and of some other *Zingiber* species showed that two other closely related *Zingiber* species had similar antiinflammatory activity to *Z. officinale*, but there was a lack of correlation between the content of gingerol, a compound specific to *Z. officinale,* and the observed total antiinflammatory activity. Therefore, it was concluded that other compounds, unidentified to date, which contributed to this activity had to be detected and quantified in order to guarantee the bioactivity of a particular *Zingiber* sample (Jiang *et al*., 2006). The application of ^{1}H NMR and PCA analysis allowed the differentiation of six samples of *Galphimia glauca*, a Mexican herbal used to treat central nervous disorders, collected from six different areas. Galphimines and 1,3,4,5-tetra-*O*-galloylquinic acid contents were found to be the main discriminating factors for PCA separation. Further analysis with PLS-DA confirmed that the level of both compounds was correlated with sedative and anxiolytic effects in mice (Fig. 2) (Cardoso-Taketa *et al*., 2008). This technique was also applied successfully to distinguish thrips-resistant and -susceptible F2 hybrids of *Senecio jacobaea* and *Senecio aquaticus*. Pyrrolizidine alkaloids, kaempferol and jacaranone were found to be higher in resistant plants. This is of interest since the last two compounds have positive effects on human health as anticancer agents (Leiss *et al*., 2009).

Proof of efficacy and mode of action identification of herbal medicines. The *in vitro* cellular effects of *Cannabis* from different cultivars and preparations were analysed using a proteomic and transcriptomic platform (Wang *et al*., 2005). In another case, the *in vivo* effect of *Ginkgo* on rats was studied by applying a ^{1}H-NMR and supervised MVDA method to measure metabolites in their urine (Wang *et al*., 2005). The effect of three traditional Chinese medicines on insulin resistant APOE*3-Leiden transgenic mice which are highly responsive in their plasma lipids to dietary treatments was also studied. LC-MS measurement combined with MVDA was applied to three TCM preparations to discover which was the most effective TCM but had the lowest toxicity (Wang *et al*., 2005).

The use of metabolomic profiling to study the mode of action of active natural products recently emerged as well, predominantly in antimicrobial activity analysis. As an example, HPLC/DAD/ESI-MS application followed by PCA confirmed that dihydrocucurbitacin F-25-*O*-acetate was the main component having an antimicrobial activity in *Hemsleya pengxianensis.* The possible antibacterial mode of action on *Staphylococcus aureus* CCTCC AB9105 was predicted by comparing the metabolite profile of *S. aureus* treated by the plant extract and commercial antibacterials with a known mode of action (Biao-Yi *et al*., 2008).

The use of metabonomics to prove the efficacy of herbal medicines was reported in some papers. The evaluation of the antiageing effects of the total flavone content of Herba Epimedii (TFE), a famous Chinese herbal medicine used to treat some age-related diseases was conducted by ^{1}H NMR analysis and PCA assignment. Ten metabolites were identified as ageing markers in the rat urine profile: creatinine, lactate, alanine, acetate, acetone, succinate, allantoin, methylamine, dimethylamine and trimethylamine-*N*-oxide. It showed that after TFE intervention, the urine profile of 24-month-old rats was shifted towards that of 18-month-old rats (Wu *et al*., 2008). In another case, the study of *Epimedium brevicornum* Maxim, was undertaken. This herbal medicine, one of the most popular in China is used for toning the kidney and strengthening bones, but the kind of constituents responsible for its pharmacological effect, and the mechanism through which it affects the entire body metabolism is unknown. The plasma metabolites of rats from three different treatments (pre-hydrocortisone intervention, post-hydrocortisone intervention and *Epimedium brevicornum* Maxim. treated) were analysed using UPLC-MS followed by PCA. The distinct increase of the level of three metabolites was found after hydrocortisone intervention, one of them identified as ethylindole-3-crylic acetate. The result showed that after *Epimedium brevicornum* Maxim treatment, the plasma metabolite profile was restored closer to that of the pre-hydrocortisone intervention. Of the four compounds of this herbal preparation detected in plasma (epimedin C, icariin, icariside II, and 2″-O-rhamnosoyl icaride), epimedin C and icariin were also found in urine, and it was concluded that these might be the active ones (Famei Li *et al*., 2007).

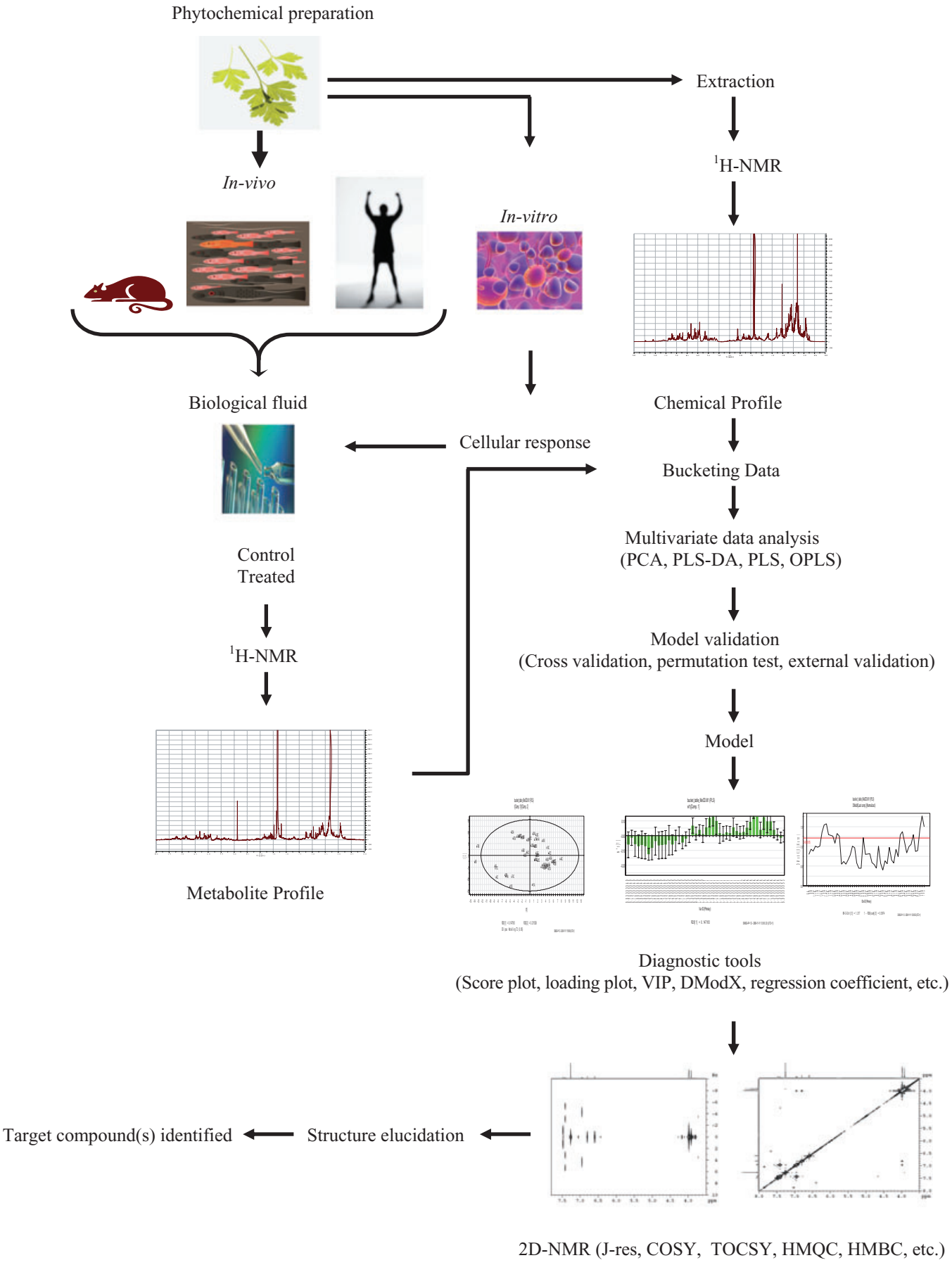

Figure 1. Schematic illustration of metabolomics and multivariate data application in bioactivity screening of phytochemical preparation. Multivariate data analysis permits the correlation of the metabolite profile obtained from the bioactivity test with the chemical profile of the phytochemical preparation, thus the efficacy involving either single or multi-synergistically active compounds can be viewed. This figure is available in colour online at http://wileyonlinelibrary.com/journal/ptr

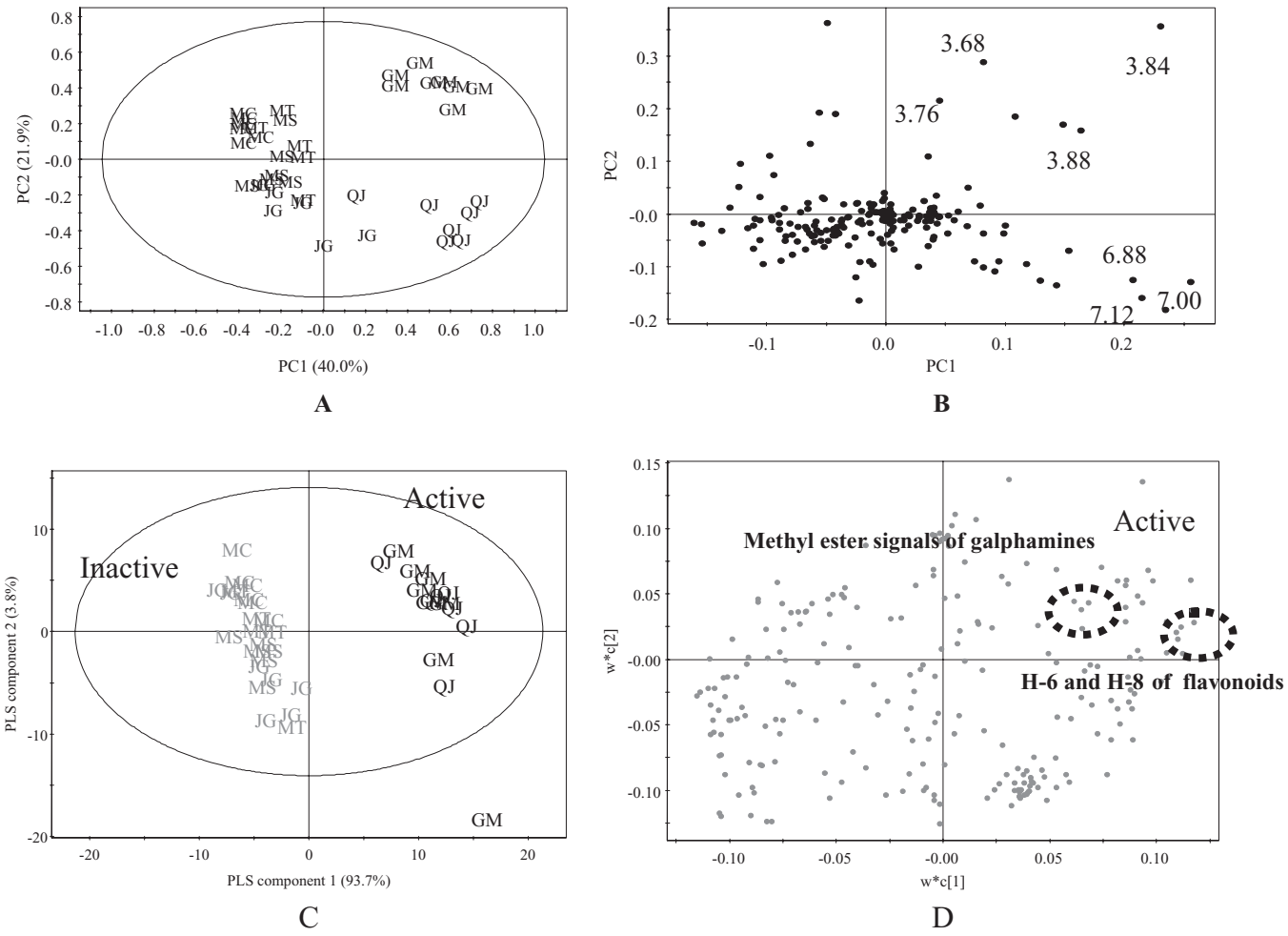

Figure 2. (A) Score plot of PCA of *Galphimia glauca* extract from six different locations. GM and QJ samples were clearly separated from others. (B) Signals of the galphimine C-27 methyl ester proton at δ 3.68–3.88 ppm as appeared on the loading plot were the main factors for the separation. (C) Partial least square discriminant analysis (PLS-DA) allows further separation base on the anxiolytic activity. (D) Galphimines and quercetin analogues were clearly shown as responsible compounds for the separation (adapted from Cardoso-Taketa *et al.*, 2008, with permission from Thieme Medical Publisher).

The efficacy and the mechanism of action of the traditional Chinese medicine preparation known as Xindi soft capsules (consisting of sea buckthorn flavonoids and sea buckthorn berry oil) used to treat blood stasis, was studied by UPLC combined with quadrupole time-of-flight tandem mass spectrometry (UPLC-QTOF-MS) followed by PCA and PLSDA analysis. Blood stasis includes haematological disorders such as haemorrhage, nasal congestion, thrombosis, and local ischaemia (microclots).

The comparison of the urinary metabolite pattern obtained for five groups of rats (healthy control group, acute blood stasis model group, low dose group of Xindi soft capsule, middle dose group of Xindi soft capsule, high dose group of Xindi soft capsule) showed that the treated groups were located between the acute blood stasis model group and the healthy control group. Confirmation by haemorheological analysis proved that within the treated group, especially in the case of the highest dose group, the urine metabolite pattern tends toward that of the control group. Some potential biomarkers were identified as well, such as cholic acid, phenylalanine and kynurenic acid (Zhao *et al.*, 2008).

Bioavailability and fate of natural compounds assessment. The metabonomics approach was applied successfully to study the bioavailability and fate of natural compounds. A traditional Chinese medicine formulation, Dangguibuxue decoction (ODD) consisting of huangqi (*Radix astragali*)-danggui (*Radix angelicae sinensis*), is widely used as a tonic. ODD is a very complex mixture containing hundreds of components. More than 70 compounds have currently been isolated and identified but only a few constituents were reported to be responsible for its activity. Metabolic fingerprinting by LC/DAD-MS was suitable for analysis of the bioavailability of the metabolites of the ODD components *in vivo* by comparing the LC/DAD-MS fingerprint of the ODD formula with the metabolite fingerprint of plasma of rabbits treated with ODD. Identification of detected compounds was done by MS and UV spectra comparison with references from the literature. Some flavonoids, phtalides and triterpene saponins, which are known as ODD active compounds were detected, showing that they are well absorbed into the body (Wang *et al.*, 2007).

Safety and toxicity of herbal medicines assessment. There are also many reports on the application of the metabonomics approach to solve the two other major issues related to the quality control of herbal medicines – safety and toxicity. The mechanism of the toxicity of aristolochic acid was studied by analysing the urine of rats submitted to intraperitoneal injection of aris-

tolochic acid and several toxins with known modes of action. The PCA, confirmed by biochemical parameter analysis, allowed researchers to conclude that aristolochic acid induced the damage of renal proximal tubules and medulla, accompanied by a slight hepatic lesion (Zhang *et al*., 2006). Similarly, the toxicity of Hei-Shun-Pian (the processed lateral root of *Aconitum carmichaelii* Debx., Ranunculaceae) was evaluated using the ^{1}H NMR spectral metabolic profile of urine and plasma samples of rats after 5 days oral treatment with the decoction of Hei-Shun-Pian at different doses. Hei-Shun-Pian is a traditional Chinese medicine with analgesic, antipyretic, antirheumatoid arthritis and anti-inflammatory effects. The C-19 diterpenoid alkaloid aconitine and a series of its derivatives are suggested to be the active principles, but some of these are also very toxic and have narrow therapeutic windows (LD_{50}/ED_{50}). Studying the PCA and PLS analysis of ^{1}H NMR data, it was found that the primary differences between the plasma of treated and the control groups was the increase in the concentration of lipids and the decrease in the concentration of phosphatidylcholine (PtCho) and *O*- and *N*-acetyl glycoproteins, while taurine, acetate, creatinine, 2-oxoglutarate, dimethylamine and hippurate levels in urine contributed to the distinction between high and medium dose groups and control group (Li *et al*., 2008a).

CONCLUSIONS AND PERSPECTIVES

In response to the renewed interest in NP, several strategies have been designed to improve natural product lead finding. Strategies aimed at improving the compatibility of natural products with high throughput screening requirements by improving the NP libraries shorten the discovery route, using advanced analytical methods and combining them with online bioassays, and selecting new approaches which can rapidly provide evidence of efficacy.

The reductionist approach is still widely used in NP drug discovery studies. However, metabolomics coupled with MVDA can be applied in those cases to speed up the limiting steps in dereplication and identification of active compounds. This must be supported by the availability of a high quality open access NP library based on NMR or MS data. Metabolomics is undoubtedly an important tool for the quality control of herbal preparations. It captures the differences in the composition due to different cultivars, growth areas, methods of preparation and adulterations, all of which might significantly affect the content of active principles in the preparation.

To explain the chemodiversity offered by Nature in a more intelligent and innovative way than random studies, HTS can be employed. For this systems biology based approach, metabolomics is viewed as a crucial tool. Systems biology may allow the connection between the chemical profile of herbal preparations and the response given by a living system, either *in vitro* or *in vivo*. Such a holistic approach to obtain proof of efficacy has emerged as the new paradigm to detect novel NP in drug discovery. Significant and obvious advantages are that synergistic interaction can be detected, as well as prodrugs. At the same time metabolomics will provide information on the mode of action and include toxic side effects. The capability quickly to analyse a broad range of metabolites makes metabolomics important in the growing field of personalized medicine. It allows the metabolic profile of individual patients to be linked to their response to specific drug therapy as a base to develop a concept on markers of the efficacy, toxicity and clearance of drugs, permitting a better prescription for an individual patient based on their genetic makeup and phenotype status.

It is evident that metabolomics has been shown to be a potent facilitator in the recovery of NP as excellent sources for novel leads, and even further, as a means to discover novel targets, as the strength of the 'omics' based approaches lies greatly in the fact that results can be stored for long term data mining.

Conflict of Interest

The authors have declared that there is no conflict of interest.

REFERENCES

Bailey NJC, Sampson J, Hylands PJ *et al*. 2002. Multi-component metabolic classification of commercial feverfew preparations via High-Field ^{1}H-NMR spectroscopy and chemometrics. *Planta Med* **68**: 734–738.

Bailey NJC, Wang Y, Sampson J *et al*. 2004. Prediction of anti-plasmodial activity of *Artemisia annua* extracts: application of ^{1}H NMR spectroscopy and chemometrics. *J Pharm Biomed Anal* **35**: 117–126.

Bajorath J. 2002. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* **1**: 882–894.

Beekwilder J, Jonker H, Meesters P *et al*. 2005. Antioxidants in raspberry; On-line analysis links antioxidant activity to a diversity of individual metabolites. *Agric Food Chem* **53**: 3313–3320.

Bhalla R, Narasimhan K, Swarup S. 2005. Metabolomics and its role in understanding cellular responses in plants. *Plant Cell Rep* **24**: 562–571.

Biao-Yi Z,Yu Y, Zeng-Liang Y. 2008. Investigation of antimicrobial model of *Hemsleya pengxianensis* W.J. Chang and its main active component by metabolomics technique. *J Ethnopharmacol* **116**: 89–95.

Bindseil KU, Jakupovic J, Wolf D *et al*. 2001. Pure compound libraries; a new perspective for natural product based drug discovery. *Drug Discov Today* **6**: 840–847.

Bleicher KH, Bohm H-J, Muller K *et al*. 2003. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**: 369–378.

Böröczky K, Laatsch H, Döbler IW *et al*. 2006. Cluster analysis as selection and dereplication tool for the identification of new natural compounds from large sample sets. *Chem Biodiversity* **3**: 622–634.

Butler MS. 2004. The role of natural product chemistry in drug discovery. *J Nat Prod* **67**: 2141–2153.

Cardoso-Taketa AT, Pereda-Miranda R, Choi YH *et al*. 2008. Metabolic profiling of the Mexican anxiolytic and sedative plant *Galphimia glauca* using nuclear magnetic resonance spectroscopy and multivariate data analysis. *Planta Med* **74**: 1295–1301.

Choi YH, Kim HK, Hazekamp A *et al*. 2004. Metabolomic differentiation of *Cannabis sativa* cultivars using ^{1}H NMR spectroscopy and principal component analysis. *J Nat Prod* **67**: 953–957.

Choi YH, Sertic S, Kim HK *et al*. 2005. Classification of *Ilex* species based on metabolomic fingerprinting using nuclear magnetic resonance and multivariate data analysis. *J Agric Food Chem* **53**: 1237–1245.
Colquhoun IJ. 2007. Use of NMR for metabolic profiling in plant systems. *J Pestic Sci* **32**: 200–212.
Cui Q, Lewis IA, Hegeman AD *et al*. 2008. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotech* **26**: 162–164.
Dickson M, Gagnon JP. 2004. Key factors in the rising cost of new drug discovery and development. *Nat Rev Drug Discov* **3**: 417–429.
Dunn WB, Ellis DI. 2005. Metabolomics: Current analytical platforms and methodologies. *Trends Anal Chem* **24**: 285–294.
Entzeroth M. 2003. Emerging trends in high-throughput screening. *Curr Opin Pharmacol* **3**: 522–529.
Exarchou V, Krucker M, van Beek TA *et al*. 2005. LC-NMR coupling technology: recent advancements and applications in natural products analysis. *Magn Reson Chem* **43**: 681–687.
Famei Li, Xiumei Lu, Huiping Liu *et al*. 2007. A pharmacometabonomic study on the therapeutic basis and metabolic effects of *Epimedium brevicornum* Maxim on hydrocortisone-induced rat using UPLC-MS. *Biomed Chromatogr* **21**: 397–405.
Feher M, Schmidt JM. 2003. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci* **43**: 218–227.
Fiehn O. 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* **48**: 155–171.
Fiehn O, Kopka J, Trethewey RN *et al*. 2000. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal Chem* **72**: 3573–3580.
Gabrielsson J, Jonsson H, Airiau C *et al*. 2006. OPLS methodology for analysis of pre-processing effects on spectroscopic data. *Chemom Intell Lab Syst* **84**: 153–158.
Gómez-Hens A, Aguilar-Caballos MP. 2007. Modern analytical approaches to high-throughput drug discovery. *Trends Anal Chem* **26**: 171–182.
Hall RD. 2006. Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* **169**: 453–468.
Henkel T, Brunne RM, Müller H *et al*. 1999. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew Chem Int Ed* **38**: 643–647.
Ingkaninan K, Hazekamp A, de Best CM *et al*. 2000. The application of HPLC with on-line coupled UV/MS: Biochemical detection for isolation of an acetylcholinesterase inhibitor from *Narcissus* Sir Winston Churchill. *J Nat Prod* **63**: 803–806.
Jahangir M, Kim HK, Choi YH *et al*. 2008. Metabolomic response of *Brassica rapa* submitted to pre-harvest bacterial contamination. *Food Chem* **107**: 362–368.
Jiang H, Xie Z, Koo HJ *et al*. 2006. Metabolic profiling and phylogenetic analysis of medicinal *Zingiber* species: Tools for authentication of ginger (*Zingiber officinale* Rosc.). *Phytochemistry* **67**: 1673–1685.
Kell DB. 2004. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **7**: 296–307.
Kim HK, Choi YH, Erkelens C *et al*. 2005. Metabolic fingerprinting of *Ephedra* species using ^{1}H-NMR spectroscopy and principal component analysis. *Chem Pharm Bull* **53**: 105–109.
Koehn FE. 2008. High impact technologies for natural products screening. In *Natural Compounds as Drugs Volume I*. Birkhäuser, Basel, 175–210.
Koehn FE, Carter GT. 2005. The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **4**: 206–220.
Krug D, Zurek G, Revermann O *et al*. 2008. Discovering the hidden secondary metabolome of *Myxococcus xanthus*: a study of intraspecific diversity. *Appl Environ Microbiol* **74**: 3058–3068.
Landro JA, Taylor ICA, Stirtan WG *et al*. 2000. HTS in the new millennium: The role of pharmacology and flexibility. *J Pharmacol Toxicol Methods* **44**: 273–289.
Leiss K, Choi Y, Abdel-Farid I *et al*. 2009. NMR metabolomics of thrips (*Frankliniella occidentalis*) resistance in *Senecio* hybrids. *J Chem Ecol* **35**: 219–229.
Li L, Sun B, Zhang Q *et al*. 2008a. Metabonomic study on the toxicity of Hei-Shun-Pian, the processed lateral root of *Aconitum carmichaelii* Debx. (Ranunculaceae). *J Ethnopharmacol* **116**: 561–568.
Li P, Qi L-W, Liu EH *et al*. 2008b. Analysis of Chinese herbal medicines with holistic approaches and integrated evaluation models. *Trends Anal Chem* **27**: 66–77.
Liu L, Li Y-F, Cheng Y-Y. 2008. A method for the production and characterization of fractionated libraries from Chinese herbal formulas. *J Chromatogr B* **862**: 196–204.
Ma C, Wang H, Lu X *et al*. 2008. Metabolic fingerprinting investigation of *Artemisia annua* L. in different stages of development by gas chromatography and gas chromatography-mass spectrometry. *J Chromatogr A* **1186**: 412–419.
Macarron R. 2006. Critical review of the role of HTS in drug discovery. *Drug Discov Today* **11**: 277–279.
McChesney JD,Venkataraman SK, Henri JT. 2007. Plant natural products: Back to the future or into extinction? *Phytochemistry* **68**: 2015–2022.
Mendes P, Camacho D, de la Fuente A. 2005. Modelling and simulation for metabolomics data analysis. *Biochem Soc Trans* **33**: 1427–1429.
Mishra KP, Ganju L, Sairam M *et al*. 2008. A review of high throughput technology for the screening of natural products. *Biomed Pharmacother* **62**: 94–98.
Moco S, Bino RJ, Vorst O *et al*. 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* **141**: 1205–1218.
Myers PL. 1997. Will combinatorial chemistry deliver real medicines? *Curr Opin Biotechnol* **8**: 701–707.
Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **70**: 461–477.
Newman DJ, Cragg GM, Snader KM. 2003. Natural products as sources of new drugs over the period 1981–2002. *J Nat Prod* **66**: 1022–1037.
Nicholson JK, Connelly J, Lindon JC *et al*. 2002. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* **1**: 153–161.
Nisbet LJ, Moore M. 1997. Will natural products remain an important source of drug research for the future? *Curr Opin Biotechnol* **8**: 708–712.
Nobeli I, Thornton JM. 2006. A bioinformatician's view of the metabolome. *Bioessays* **28**: 534–545.
Ortholand J-Y, Ganesan A. 2004. Natural products and combinatorial chemistry: back to the future. *Curr Opin Chem Biol* **8**: 271–280.
Pierens GK, Palframan ME, Tranter CJ *et al*. 2005. A robust clustering approach for NMR spectra of natural product extracts. *Magn Reson Chem* **43**: 359–365.
Pieters L, Vlietinck AJ. 2005. Bioguided isolation of pharmacologically active plant components, still a valuable strategy for the finding of new lead compounds? *J Ethnopharmacol* **100**: 57–60.
Politi M, Chávez MI, Cañada FJ *et al*. 2005. Screening by NMR: A new approach for the study of bioactive natural products? The example of *Pleurotus ostreatus* hot water extract. *Eur J Org Chem* **2005**: 1392–1396.
Poulsen S, Davis R, Keys T. 2006. Screening a natural product-based combinatorial library using FTICR mass spectrometry. *Bioorg Med Chem* **14**: 510–515.
Rasmussen B, Cloarec O, Tang H *et al*. 2006. Multivariate analysis of integrated and full-resolution ^{1}H-NMR spectral data from complex pharmaceutical preparations: St. John's Wort. *Planta Med* **72**: 556–563.
Rochfort S. 2005. Metabolomics reviewed: A new omics platform technology for systems biology and implications for natural products research. *J Nat Prod* **68**: 1813–1820.
Rollinger JM, Stuppner H, Langer T. 2008. Virtual screening for the discovery of bioactive natural products. In *Natural Compounds as Drugs Volume I*, Petersen F, Amstutz R (eds). Birkhäuser Verlag: Basel, 211–249.
Roos G, Röseler C, Büter KB *et al*. 2004. Classification and correlation of St. John's Wort extracts by nuclear magnetic resonance spectroscopy, multivariate data analysis and pharmacological activity. *Planta Med* **70**: 771–777.
Schenk T, Breel GJ, Koevoets P *et al*. 2003. Screening of natural products extracts for the presence of phosphodiesterase inhibitors using liquid chromatography coupled online to parallel biochemical detection and chemical characterization. *J Biomol Screen* **8**: 421–429.
Schmid I, Sattler I, Grabley S *et al*. 1999. Natural products in high throughput screening: automated high-quality sample preparation. *J Biomol Screen* **4**: 15–25.

Schobel U, Frenay M, Van Elswijk DA *et al*. 2001. High resolution screening of plant natural product extracts for estrogen receptor a and f3 binding activity using an online HPLC-MS biochemical detection system. *J Biomol Screen* **6**: 291–303.
Schroeder FC, Gibson DM, Churchill ACL *et al*. 2007. Differential analysis of 2D NMR spectra: new natural products from a pilot-scale fungal extract library. *Angew Chem Int Ed* **119**: 919–922.
Shyur L-F, Yang N-S. 2008. Metabolomics for phytomedicine research and drug development. *Curr Opin Chem Biol* **12**: 66–71.
Smith A. 2002. Screening for drug discovery: The leading question. *Nature* **418**: 453–459.
Sumner LW, Mendes P, Dixon RA. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62**: 817–836.
Thomas JJ, Shen Z, Crowell JE *et al*. 2001. Desorption/ionization on silicon (DIOS): A diverse mass spectrometry platform for protein characterization. *Proc Natl Acad Sci USA* **98**: 4932–4937.
Tikunov Y, Lommen A, de Vos CHR *et al*. 2005. A novel approach for nontargeted data analysis for metabolomics: Large-scale profiling of tomato fruit volatiles. *Plant Physiol* **139**: 1125–1137.
Ulrich-Merzenich G, Zeitler H, Jobst D *et al*. 2007. Application of the '-Omic-' technologies in phytomedicine. *Phytomedicine* **14**: 70–82.
Vaidyanathan S, Jones D, Broadhurst D *et al*. 2005. A laser desorption ionisation mass spectrometry approach for high throughput metabolomics. *Metabolomics* **1**: 243–250.
van der Greef J, Smilde AK. 2005. Symbiosis of chemometrics and metabolomics: past, present, and future. *J Chemometr* **19**: 376–386.
van der Kooy F,Verpoorte R, Meyer JJM. 2008. Metabolomic quality control of claimed anti-malarial *Artemisia afra* herbal remedy and *A. afra* and *A. annua* plant extracts. *S Afr J Bot* **74**: 186–189.
van Elswijk DA, Irth H. 2002. Analytical tools for the detection and characterization of biologically active compounds from nature. *Phytochem Rev* **1**: 427–439.
Verpoorte R. 1998. Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development. *Drug Discov Today* **3**: 232–238.
Verpoorte R, Choi Y, Kim H. 2007. NMR-based metabolomics at work in phytochemistry. *Phytochem Rev* **6**: 3–14.
Verpoorte R, Choi Y, Mustafa N *et al*. 2008. Metabolomics: back to basics. *Phytochem Rev* **7**: 525–537.
Verpoorte R, Choi YH, Kim HK. 2005. Ethnopharmacology and systems biology: A perfect holistic match. *J Ethnopharmacol* **100**: 53–56.
Verpoorte R, Kim H, Choi Y. 2006. Plants as source for medicines: New perspectives. In *Medicinal and Aromatic Plants: Agricultural, Commercial, Ecological, Legal, Pharmacological and Social Aspects*, Bogers R, Craker L, Lange D (eds). Wageningen UR: Wageningen, 261–273.
Walters WP, Namchuk M. 2003. Designing screens: how to make your hits a hit. *Nat Rev Drug Discov* **2**: 259–266.
Wang M, Lamers R-JAN, Korthout HAAJ *et al*. 2005. Metabolomics in the context of systems biology: bridging traditional Chinese medicine and molecular pharmacology. *Phytother Res* **19**: 173–182.
Wang P, Liang Y, Zhou N *et al*. 2007. Screening and analysis of the multiple absorbed bioactive components and metabolites of Dangguibuxue decoction by the metabolic fingerprinting technique and liquid chromatography/diode-array detection mass spectrometry. *Rapid Commun Mass Spectrom* **21**: 99–106.
Wang Y, Tang H, Nicholson JK *et al*. 2004. Metabolomic strategy for the classification and quality control of phytomedicine: a case study of chamomile flower (*Matricaria recutita* L.). *Planta Med* **70**: 250–255.
Ward JL, Baker JM, Beale MH. 2007. Recent applications of NMR spectroscopy in plant metabolomics. *FEBS J* **274**: 1126–1131.
Williams G. 2004. Advances in high throughput screening. *Drug Discov Today* **9**: 515.
Williard X, Pop I, Bourel L *et al*. 1996. Combinatorial chemistry: a rational approach to chemical diversity. *Eur J Med Chem* **31**: 87–98.
Wold S, Sjöström M, Eriksson L. 2001a. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* **58**: 109–130.
Wold S, Trygg J, Berglund A *et al*. 2001b. Some recent developments in PLS modeling. *Chemom Intell Lab Syst* **58**: 131–150.
Wolfender J-L, Queiroz EF, Hostettmann K. 2006. The importance of hyphenated techniques in the discovery of new lead compounds from nature. *Expert Opin Drug Discov* **1**: 237–260.
Wu B, Yan S, Lin Z *et al*. 2008. Metabonomic study on ageing: NMR-based investigation into rat urinary metabolites and the effect of the total flavone of *Epimedium*. *Mol Biosyst* **4**: 855–861.
Yi L-Z, Yuan D-L, Liang Y-Z *et al*. 2007. Quality control and discrimination of Pericarpium Citri Reticulatae and Pericarpium Citri Reticulatae Viride based on high-performance liquid chromatographic fingerprints and multivariate statistical analysis. *Anal Chim Acta* **588**: 207–215.
Zhang X, Wu H, Liao P *et al*. 2006. NMR-based metabonomic study on the subacute toxicity of aristolochic acid in rats. *Food Chem Toxicol* **44**: 1006–1014.
Zhao X, Zhang Y, Meng X *et al*. 2008. Effect of a traditional Chinese medicine preparation Xindi soft capsule on rat model of acute blood stasis: A urinary metabonomics study based on liquid chromatography-mass spectrometry. *J Chromatogr B* **873**: 151–158.