European Journal of Medicinal Plants
5(2): 191-203, 2015, Article no.EJMP.2015.018
ISSN: 2231-0894

SCIENCEDOMAIN *international*
www.sciencedomain.org

Sensitivity of NMR-based Metabolomics in Drug Discovery from Medicinal Plants

Frank Van der Kooy[1*], Beatrice Venkataya[1], Jarryd L. Pearson[1], Allan Torres[2], Chun G. Li[1] and Dennis Chang[1]

[1]National Institute of Complementary Medicine, University of Western Sydney, Australia.
[2]Nanoscale Organisation and Dynamics Group, University of Western Sydney, Australia.

Authors' contributions

This work was carried out in collaboration between all authors. Authors FVK, BV, JLP and AT designed the study, performed the analysis, wrote the protocol and wrote the first draft of the manuscript. Authors CGL and DC contributed to the preparation of the manuscript and performed final editing. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/EJMP/2015/13747
Editor(s):
(1) Marcello Iriti, Professor of Plant Biology and Pathology Department of Agricultural and Environmental Sciences Milan State University, Italy.
Reviewers:
(1) Anonymous, Beijing University of Chinese Medicine, China.
(2) Anonymous, Universidad Autónoma de Nuevo León, Mexico.
(3) Anonymous, University of the Western Cape, South Africa.
(4) Anonymous, Harbin Medical University, China.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=685&id=13&aid=6622

Original Research Article

Received 1st September 2014
Accepted 29th September 2014
Published 23rd October 2014

ABSTRACT

One of the main limitations of NMR based metabolomic analysis is its low sensitivity. The main objective of this study was to determine the sensitivity of the standard NMR based metabolomics protocol (as published in Nature Protocols) for the analysis of plant samples. To test this limitation, we prepared two sample sets from a well-known medicinal plant. Sample set one was prepared from one plant specimen, subdivided into 24 chemically equivalent samples, spiked with different concentrations of rutin and analysed on 300, 400 and 500 MHz NMR spectrometers. Sample set two was prepared from four different plant specimens of the same species reflecting the natural variation between medicinal plants. To two of these samples different concentrations of artemisinin were added, acting as our hypothetical active samples, whereas the other two samples acted as our inactive samples. Our results indicated that there was no difference in the sensitivity between the three NMR spectrometers and that the standard protocol can differentiate between samples at a spiking level of 0.2 mg/mL of rutin (328 µm). The second sample set gave differentiation at

**Corresponding author: E-mail: f.vanderkooy@uws.edu.au;*

0.05 mg/mL (177 µm) but a significant movement in the chemical shifts of artemisinin was observed. Our study demonstrated that the sensitivity of the current NMR based metabolomics protocol is not due to instrumental limitations but rather due to methodological limitations. In the same way that binning of spectra negates the better resolution of higher field magnets the same appears to be true by employing PCA analysis to spectra which effectively negates the higher sensitivity of higher field magnets. Our study also highlights that compounds can display significant movement in chemical shifts depending on its chemical environment, which can complicate identification by database comparisons. We would like to invite the NMR metabolomics community to repeat this analysis in order to confirm this finding so that the current limitations of the NMR based metabolomics protocol can be defined. This needs to be done in order to develop improved NMR metabolomics protocols.

Keywords: Chemical shift; drug discovery; medicinal plants; metabolomics; NMR.

1. INTRODUCTION

A relatively simplistic definition of metabolomics is the identification and quantification of all low molecular weight metabolites in a given sample [1]. Although this is not yet technically feasible, the term metabolomics has found acceptance in mainstream literature. For a more in-depth discussion of the terminology and the technical aspects of metabolomics [2,3,4,5] should be consulted. Throughout this manuscript we are concerned with the aspect of metabolomics, which involves a comparative analysis between two sample sets in order to identify the metabolites causing a differentiation between them.

The application of metabolomics is extremely broad and encompasses the investigation of cells, tissues and organs as well as biological fluids (urine, plasma). The main analytical tools used are liquid chromatography or gas chromatography coupled with mass spectrometry (LC/GC-MS) and nuclear magnetic resonance (NMR) spectroscopy. NMR has many advantages over MS techniques, but it also suffers from some disadvantages, including its relatively low sensitivity compared to MS, as well as peak overlap that can complicate compound identification [1,5]. Another disadvantage is that there is not yet consensus on a standard protocol for NMR-based metabolomic analysis and storing information on identified metabolites is still by and large lacking. Attempts to standardise analytical protocols [6] are ongoing and various public accessible databases for metabolite identification do however exist such as the Human Metabolome Database (http://www.hmdb.ca), Biological Magnetic Resonance Databank. (http://www.bmrb.wisc.edu/) and the Birmingham Metabolite library [7]. It is, however, important to note that most of these databases contain information regarding primary (human) metabolites and very few resources exist for plant secondary metabolite identification.

The main approach of metabolomics is to compare two data sets (e.g., control vs treatment) and by means of data reduction techniques and multivariate data analysis, differentiate or discriminate between these two datasets. This process highlights any differences between the datasets and aims to identify the compounds responsible for this differentiation. For example, biological fluids from healthy volunteers and from volunteers that have been diagnosed with a disease are analysed by GC/LC-MS [8] and/or NMR [9] techniques. The raw data will undergo data reduction techniques and multivariate statistical analysis, in order to differentiate between these two groups. The compounds responsible for the differentiation can then be identified as possible disease biomarkers. Another example is pest resistant plant cultivars vs non-resistant cultivars where metabolomics was used to study the cause of the observed resistance in order to breed improved resistant food crops [10]. In our study we focus on one application of metabolomics, namely NMR-based drug discovery from medicinal plants. The results might however be applicable to the wider NMR based metabolomics analysis of any biological material.

The classical method to identify active compounds in complex medicinal plant extracts is called bioguided fractionation. This technique makes use of multiple rounds of chromatographic fractionation followed by bioactivity testing, until the active compound(s) is/are identified. This method, although proven, is time consuming and

expensive. To improve on the classical method, NMR-based metabolomics approaches have recently been employed. The usual approach is to analyse multiple plant samples from one bioactive plant species and based on the natural variation in concentration of the active compounds and subsequent differences in the biological activity of these samples, the bioactive compound will be identified without having to resort to bioguided fractionation [11]. A different approach is to analyse different solvent extracts from a single medicinal plant sample in order to identify the bioactive compounds without the need for bioguided fractionation [12].

The NMR metabolomics approach consists mainly of two steps. Step 1 is to perform ^{1}H NMR analysis on all samples followed by data reduction and multivariate data analysis (PCA or OPLS) in order to identify the signals causing the differentiation between the sample sets. Step 2 focusses on a reduced number of samples where 2D NMR analyses are performed in order to identify the compounds responsible for the differentiating signals identified in Step 1 [1,5]. Needless to say is that step1 is therefore the most important aspect as all subsequent (time consuming) work such as compound identification, depends on this relatively simple 1D analysis performed during the first step. One of the most common questions being asked when NMR based metabolomics is discussed is the following: "How sensitive is NMR?" or "At what concentration will I be able to see differentiation between my sample sets?" To the best of our knowledge, this question has not yet been answered adequately. To define or to explain this question more precisely: the question does not refer to the instrumental sensitivity of NMR, which can be defined as the concentration of compound that will give a limit of detection of at least five times signal to noise (the instrumental sensitivity of NMR will increase with the use of higher field instruments). The question refers to the concentration of compound that will cause a positive differentiation between two sample sets on a PCA generated scoring scatter plot using the standard metabolomics protocol [5]. This concentration can only be similar to or higher than the instrumental limit of detection of the NMR being used in the study. In order to find a quantitative answer to this question we tested a sample set prepared and analysed according to the standard NMR based metabolomics protocol for plant samples [5]. We prepared a chemically identical sample set from a single *Artemisia afra* specimen and added different concentrations of rutin. All samples were manually analysed on a 300, 400 and 500 MHz NMR spectrometers and all data handling and PCA analysis were done according to the standard protocol [5].

Due to the first sample set being chemically equivalent except for the added rutin at different concentrations, we also prepared a second sample set from four different *Artemisia afra* specimens. This will be a closer reflection of the expected natural chemical variation between different medicinal plant samples of the same species. *Artemisia afra* is known not to contain the well-known bioactive compound artemisinin [13] although the chemistry is closely related to *Artemisia annua* from where artemisinin was first discovered [14]. *Artemisia afra* therefore acted as a "placebo" but is devoid of the active compound artemisinin. For this test we assumed that artemisinin has not yet been discovered and that we are using NMR based metabolomics in order to identify this bioactive compound. To two of these samples we added different concentrations of artemisinin (active samples) and to the remaining two samples solvent only (control samples – inactive) and the same procedure was followed as described above. Our main objectives therefore were a) to determine at which concentration will separation on a PCA scoring scatter plot occur using the different spectrometers and b) at what concentration will the signals from the added reference compounds be present in the top 10 signals of the loadings scatter plot.

2. MATERIALS AND METHODS

2.1 Plant Material

The aerial parts of *Artemisia afra* Jacq. *ex* Willd (Asteraceae) were collected in the botanical garden of the University of Pretoria (South Africa) on 7/11/2008 (sample 1), 15/10/2008 (sample 2), 15/12/2008 (sample 3) and 22/11/2011 (sample 4). Plant material was stored at room temperature in the dark until used.

2.2 Chemicals

Deuterated methanol (MeOD) (cat. no.DLM-24-50) and sodium deuteroxide (cat. no. DLM-45-50) were obtained from Cambridge Isotope Laboratories and deuterated water (D_2O) (cat. no. 5150) from Spectra Stable Isotopes. 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP) (cat. no. 26991-3), KH_2PO_4 (cat. no.

30407) from Riedel-de Haen and rutin (cat. no. R5143-50G) and artemisinin (cat. no. 361593-100MG) were obtained from Sigma Aldrich.

2.3 Sample Preparation

2.3.1 Sample Set 1

This sample set was prepared from *Artemisia afra* (sample 4) as described in Kim et al. [5] with the exception that only one large extract was prepared. In short, the protocol describes the extraction of 50-100 mg of plant material with a 1:1 mixture of MeOD: $D_2O$$KH_2PO_4$ buffer pH 6 containing 0.01% TSP (1.5 mL). We used the same plant material to solvent ratio (100 mg: 1.5 mL solvent) and in order to prepare 24 samples, 2.4 g of plant material was extracted with 36 mL of MeOD: D_2O KH_2PO_4 buffer pH 6 containing 0.01% TSP. The mixture was sonicated and centrifuged as described in the protocol. After filtration, 800 µL of extract was transferred into 5 mm NMR tubes to yield 24 chemically equivalent samples. To three of these samples 100 µL of MeOD was added to act as control samples whereas 100 µL of MeOD containing different concentrations of rutin was added in triplicate to the rest of the samples to yield a concentration range of 0.05, 0.10, 0.20, 0.40, 0.60, 0.80 and 1.00 mg/mL as the final concentrations of rutin. All samples were vortexed well in order to ensure homogeneity and were stored at 4°C until analysis.

2.3.1 Sample Set 2

Four different *A. afra* samples (samples 1-4) were weighed (100 mg each) into 2 mL Eppendorf tubes in duplicate. 1.5 mL of a 4:1 ratio of MeOD: D_2O KH_2PO_4 buffer pH 6 containing 0.01% TSP was added to each sample. (Due to the low aqueous solubility of artemisinin this test was performed with a 4:1 ratio of NMR solvents). The samples were vortexed, sonicated and centrifuged according to the protocol [5], with 800 µL of resulting supernatant transferred to 5 mm NMR tubes for analysis. To two of these samples (sample 1 and 2) 100 µL of MeOD was added, while in samples 3 and 4, 100 µL of different concentrations of artemisinin dissolved in MeOD was added (0.05, 0.10, 0.20 and 0.40 mg/mL). Sample 1 and 2 therefore acted as our "inactive" samples whereas samples 3 and 4 served as our "active" samples. This yielded two control samples which are chemically different reflecting the natural variation one can expect in medicinal plants; and two spiked samples which in turn are chemically different to the control samples and in addition contain different concentrations of artemisinin. All samples were stored at 4°C until analysis.

2.4 NMR Analysis

For the analysis of the first sample set we used a Varian 300 MHz, Bruker 400 MHz and Bruker 500 MHz spectrometers. All samples were manually analysed at 298 K. A standard proton pulse sequence was used with the following parameters. Flip angle 60°, relaxation delay 1.5 sec., complex points 16 000, dummy scans = 4, number of scans = 64, zero filling 64K, line broadening 0.3 Hz. The 300 MHz Varian spectrometer was manually tuned before analysis and MeOD was used as the lock signal. The first sample was manually shimmed and the shims of all subsequent samples slightly adjusted to ensure the highest lock level. The Varian Mercury 300 MHz spectrometer was equipped with a 5 mm 4-nuclei (1H, 13C, 31P and 19F) inverse probe. The samples analysed on the Bruker 400 and 500 MHz spectrometers were automatically tuned and matched. MeOD was used as lock signal and the samples were manually shimmed. The 400 MHz spectrometer was equipped with a 5 mm-BBFO probe while the 500 MHz was equipped with a 5 mm BBI probe. Due to limited NMR accessibility the second sample set was only analysed on the 300 and 400 MHz NMR spectrometers using the same parameters as described above.

2.5 Data Reduction and PCA Analysis

The Free induction decays (FID's) of all samples were transferred to MestReNova software (Version 8.1.2-11880) for post analysis data reduction. All FID's were Fourier transformed, manually phased in zero and first phase, baseline corrected and referenced to TSP at 0.0 ppm. The following data analysis steps were taken:

- All spectra were overlayed in order to visually inspect if there were any outliers e.g. caused by inadequate shimming or phasing. Any outliers were removed from further analysis.
- The residual solvent peaks were removed from the spectra (3.28-3.36 ppm for MeOD and 4.6-5.0 ppm for D_2O).
- Spectra were binned into 0.04 p.p.m. sized bins from 0.0 – 9.0 ppm resulting in ~ 200 variables per spectrum.

- The spectra were normalised to the total sum of intensities.
- The binned spectra were converted into an ASCII file format and imported into Simca Umetrics (Version 13.0.2.0, Jan 11 2013) software for PCA analysis.
- Paretto scaling was used during PCA analysis.

3. RESULTS

In our current study we did not attempt to improve the differentiation between the samples by plotting different PCs (for example PC2 vs PC3 or PC1 vs PC3) and we also did not use different scaling techniques or different multivariate data analysis techniques e.g., PLS or OPLS. The reason for this was that we wanted to test the existing standard protocol as is, in order to define the limitations and highlight the areas where the protocol can or should be improved. The difference in instrumental sensitivity between the three spectrometers was also calculated by determining the signal to noise ratio of TSP. The increase of instrumental sensitivity of the 400 and 500 MHz spectrometers as compared to the 300 MHz spectrometer were 3.0 and 5.4 respectively, which was somewhat lower than expected.

3.1 Analysis of Sample Set 1

Sample set 1 consisted of samples which were chemically equivalent except for the differing concentrations of the added reference compound. Since the analysis of PCA generated scoring plots is subjective, we considered positive separation when all spiked samples were on one side of a straight vertical line (separated across PC1) or a horizontal line (separated across PC2), with the control samples on the opposite side of this line. We also considered the samples being separated if clear clustering occurred with all spiked samples grouping very close together. Fig. 1a-f gives the PCA scoring plots of samples analysed on the 300 MHz spectrometer. As expected the separation between the spiked samples and the control samples improved with an increase in concentration of the reference compound. Separation was not detected at a spiking level of 0.1 mg/mL or lower, with initial separation observed at 0.2 mg/mL and clear clustering at 0.8 and 1.0 mg/mL. According to [5], it is assumed that there is no need to analyse duplicate or triplicate samples due to the high reproducibility of NMR analysis. However, based on Fig. 1a-f it is clear that good clustering only occurs adequately at the highest concentrations tested (0.8 and 1.0 mg/mL) and that instrumental variation leads to significant separation between replicate samples at lower concentrations. Although the overall trend is visible, the control samples and all other spiked samples are separated across PC1 and/or PC2 indicating that instrumental variation (e.g., shimming, phasing, baseline correction etc.) plays a significant role in obtaining good results at the lower concentrations tested. Therefore, it may be advisable to analyse at least duplicate samples instead of single samples if at all possible. Similarly, the 400 MHz scoring plots showed separation at 0.2 mg/mL whereas clustering was observed at 0.6 mg/mL and at higher concentrations. For the 500 MHz dataset, separation occurred at 0.1 mg/mL, whereas clear clustering was only observed for the control samples for all tested concentrations; clustering of the spiked samples was not detected for the 500 MHz dataset.

Loading plot interpretation is analogous to scoring plot interpretation, although it is considered to be even more subjective. Due to NMR based metabolomics being hailed as an unbiased technique, we had to define what an unbiased positive result would be. It was decided that from the ~200 data points presented in the loading plots, only the top ten data points contributing to the separation should be analysed. Initially, scoring plots with positive separation were identified, and from each of these, as shown in Table 1, the top ten signals listed. If no separation occurred in the scoring plot (i.e. no separation of samples on the horizontal or vertical straight lines, or clustering) we did not analyse the loadings plot. Fig. 2 illustrates the loadings plot of the 0.8 mg/mL spiked samples. From Fig. 1e it can be seen that clustering of these samples occurs in the upper right quadrant and therefore it is from this quadrant that the top ten signals were obtained and listed in Table 1. Additionally, bins in which pure rutin gave a contribution were also listed and matched to the top ten loadings plot signals. A positive result was therefore considered to be matching signals between the top ten loadings plot signals and the signals arising from rutin.

Investigation of the separate loading plots of each concentration plotted against the control samples confirmed that at a concentration of 0.2 mg/mL the correct signals are detected and

that rutin is partly responsible for the observed separation. Table 1 gives the top ten loadings plot signals as compared with the signals of pure rutin at the different concentrations tested. Based on these results, it can be concluded that for the 300, 400 and 500 MHz, differentiation occurred at a concentration of 0.2 mg/mL and higher, and that this differentiation is partly caused by rutin. As expected, at higher concentrations more signals matched than at the lower concentrations. The 500 MHz results did however give separation at 0.1 mg/mL on the scoring plot but none of the signals in the loadings plot matched the signals of rutin. The instrumental sensitivity of the 500 MHz which is 5.4 times higher than the 300 MHz, therefore does not appear to have any improved benefit in differentiating between samples at lower spike levels. In theory the 500 MHz spectrometer should have differentiated between samples at a concentration of 0.03 mg/mL due to it being 5.4 times more sensitive than the 300 MHz spectrometer.

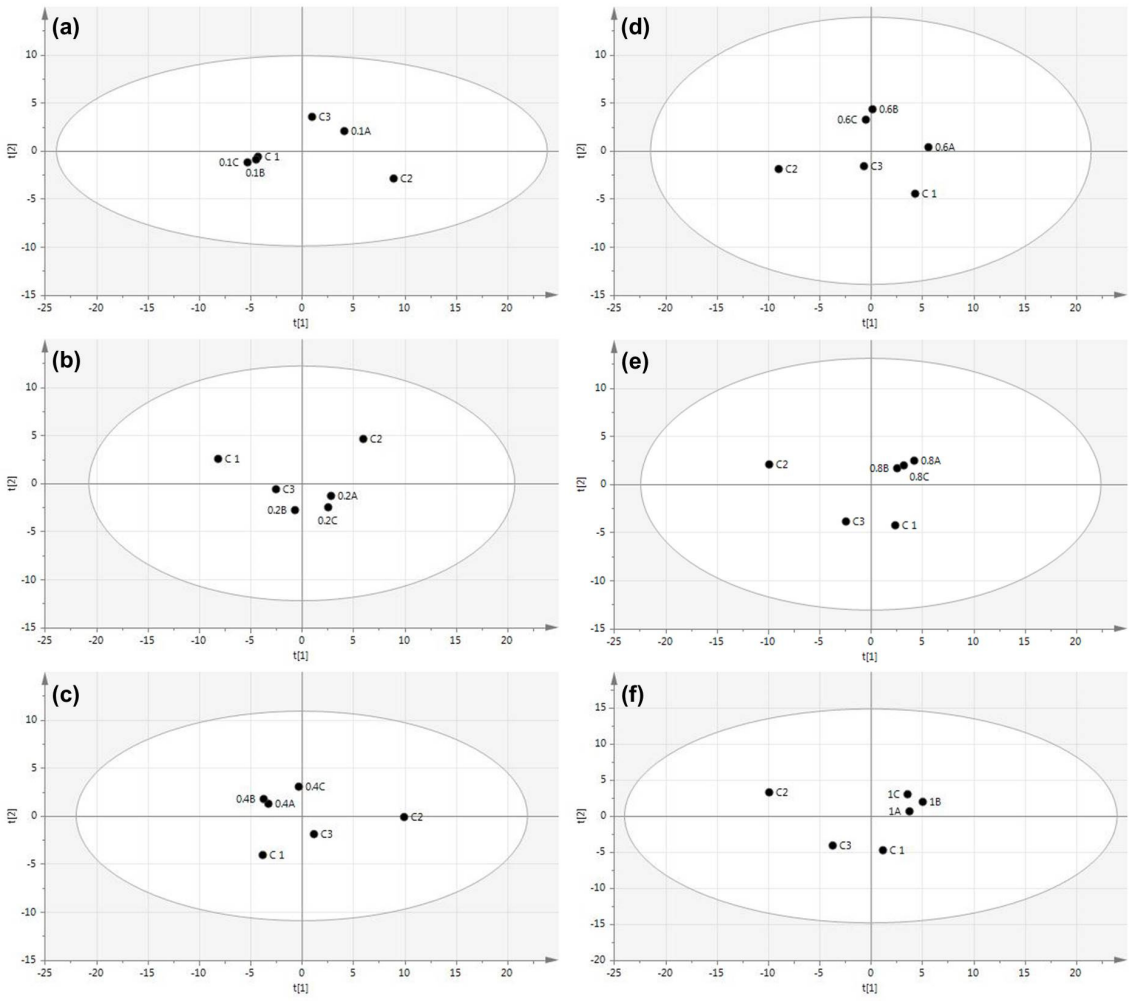

Fig. 1. Scoring plot illustrating the differentiation between control samples and rutin spiked samples. Rutin samples were spiked at (a) 0.1 mg/mL, (b) 0.2 mg/mL, (c) 0.4 mg/mL, (d) 0.6 mg/mL, (e) 0.8 mg/mL, (f) 1.0 mg/mL concentrations. From the scoring plots, clear clustering can be seen at 0.8 and 1.0 mg/mL

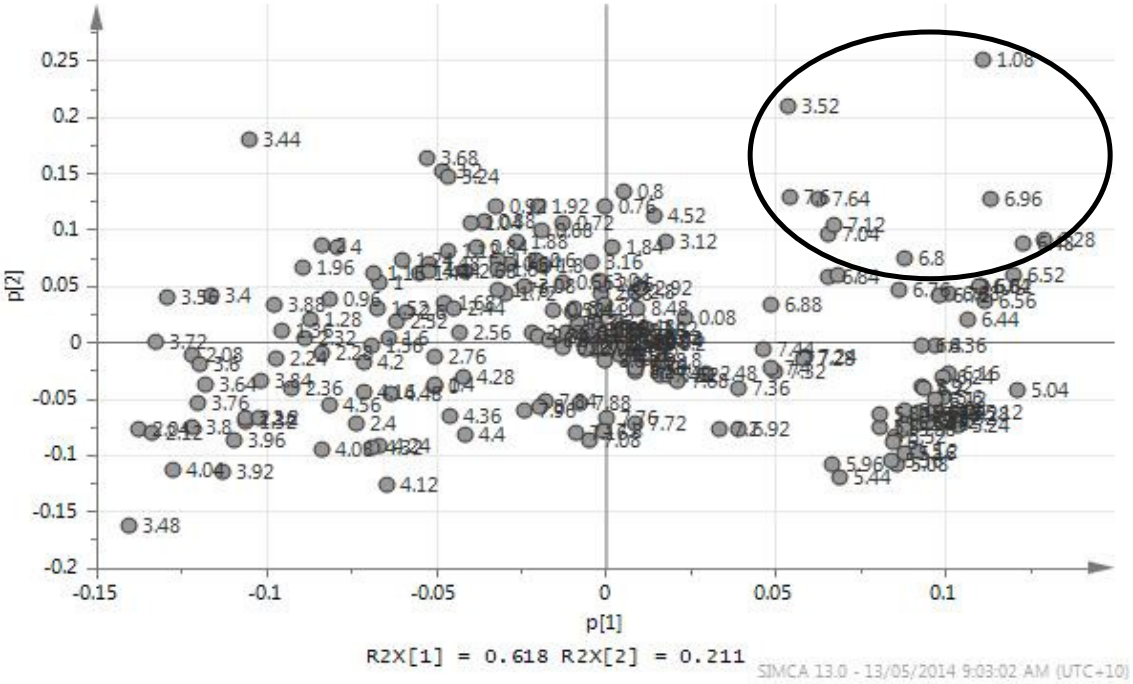

Fig. 2. Loadings plot obtained from plotting the 0.8 mg/mL rutin spiked samples as compared to the control group. The spiked samples grouped together in the upper right hand quadrant and therefore the loadings plot signals in this quadrant was listed as contributing the most to the observed separation

3.2 Analysis of Sample Set 2

Four different samples of *Artemisia afra* were prepared according to the protocol. All four samples were chemically different and in addition two samples contained a known bioactive compound, artemisinin, at different concentrations. Clear separation between all four unspiked samples was achieved reflecting the natural chemical variation between the samples. Separation also occurred after spiking two of these samples with different concentrations of artemisinin (Fig. 3). Although the intra-sample variation became smaller (variation between S3A and S3B), after the addition of artemisinin the inter-sample variation still played a bigger role in the separation (variation between S3 and S4). Nevertheless, the loading plots revealed that this separation is partly caused by signals arising from artemisinin (Table 2a). As expected the number of matching signals increased with concentration with the methyl groups of artemisinin playing the biggest role. The H-12 (Fig. 4) singlet at 6.07 ppm does not appear to play any role in the separation although it is clearly visible in the NMR spectrum at all concentrations tested. For the 400 MHz analysis the same trend was observed although there were more matching signals at the lower concentrations tested. The singlet at 6.07 ppm again did not play any role in the observed separation. Upon closer inspection of the raw NMR data we noticed that there were substantial differences between the chemical shifts of pure artemisinin as compared to the chemical shifts of artemisinin added to the samples. It is well known that some phenolic acids can display movement of chemical shifts based on their concentration, sample pH or temperature [5,15] but in our experiments we kept the pH and temperature constant while varying the concentration of artemisinin (A very slight concentration dependant shift was observed for rutin). Fig. 5 illustrates that the movement of chemical shift does not appear to be due to the varying concentrations of artemisinin but rather due to the chemical environment in which artemisinin exists. The movement of chemical shift therefore appears to be concentration independent. Furthermore the movement in chemical shift is constant for each sample, for example, the chemical shift of the H-12 proton of artemisinin added to sample 3 was 6.038 ppm (constant at all concentrations).

Table 1. Top ten loading plot signals of each concentration vs the signals from pure rutin. Rutin gives resonance signals in the following bins: 1.08, 3.4-3.8, 6.28, 6.48, 6.96, 7.6, 7.64. Matching signals are given in bold font

Spiking level in mg/mL	clustering	Top ten loading plot signals (ppm)	Matching signals
300 MHz loading plot data			
0.05	no	n/a	n/a
0.10	no	n/a	n/a
0.20	yes	**1.08**, 1.92, 2.00, 3.20, **3.40**, **3.44**, **3.52**, **3.60**, **3.68**, 4.00	6
0.40	yes	**1.08**, **3.52**,**6.28**, 6.32, **6.96**, 7.04, 7.12, **7.60**, **7.64**, 7.68	6
0.60	yes	**1.08**, 2.48, **3.52**, 4.52, 5.40, **6.28**, **6.96**, **7.60**, **7.64**, 8.48	6
0.80	yes	**1.08**, **3.52**, **6.28**, **6.48**, 6.80, **6.96**, 7.04, 7.12, **7.60**, **7.64**	7
1.00	yes	**6.28**, 6.32, **6.48**, 6.52, 6.56, 6.60, 6.64, **6.96**, **7.60**, **7.64**	5
400 MHz loading plot data			
0.05	no	n/a	n/a
0.10	no	n/a	n/a
0.20	yes	**1.08**, 1.40, 2.52, 3.12, **6.28**, 6.88, **6.96**, 7.04, **7.60**, **7.64**	5
0.40	yes	**1.08**, 1.20, 3.12, **6.28**, **6.48**, **6.96**, 7.16, **7.60**, **7.64**, 7.68	6
0.60	yes	**1.08**, 1.20, 3.12, **3.52**, **6.28**, **6.48**, 6.52, **6.96**, **7.60**, **7.64**	7
0.80	yes	**1.08**, 1.20, 1.40, 3.12, **6.28**, **6.48**, **6.96**, **7.60**, **7.64**, 7.68	6
1.00	yes	**1.08**, 3.12, 4.08, 4.20, 4.52, **6.28**, **6.48**, **6.96**, **7.60**, **7.64**	6
500 MHz loading plot data			
0.05	no	n/a	n/a
0.10	yes	4.04, 4.08. 4.12, 4.20, 4.24, 4.28, 4.32, 4.36, 4.40, 4.48	0
0.20	yes	**1.08**, 4.04, 4.08, 4.20, 4.28, 4.44, 4.52, 6.92, **6.96**, **7.64**	3
0.40	yes	1.36, 1.40, 1.56, **3.48**, 4.04, 4.08, 4.20, 5.12, 5.16, **6.28**	2
0.60	yes	**1.08**, **3.48**, 4.52, 5.04, 5.08, **6.28**, **6.96**, 7.00, **7.60**, **7.64**	6
0.80	yes	**1.08**, **3.52**, 4.52, 5.04, 5.08, 5.12, **6.28**, **6.96**, **7.60**, **7.64**	6
1.00	yes	**1.08**, **3.40**, **3.52**, 4.52, 5.04, 5.08, **6.28**, **6.96**, **7.60**, **7.64**	7

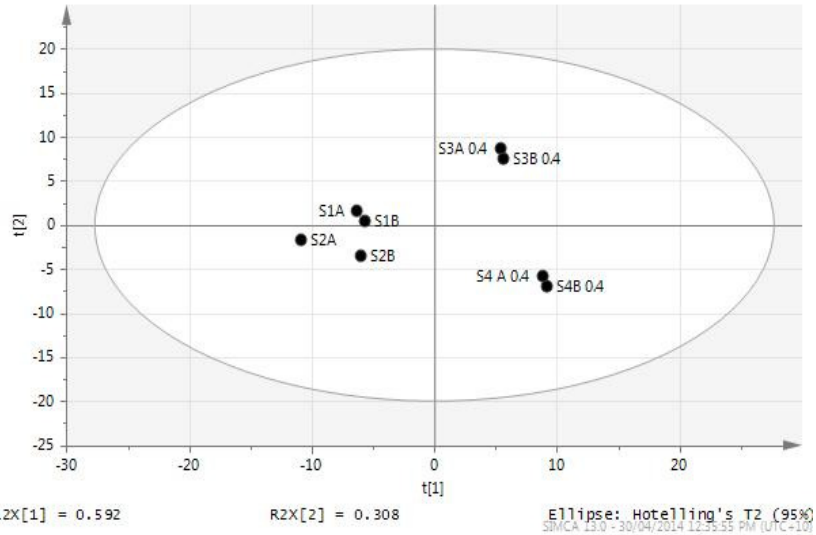

Fig. 3. Scoring plot of four *Artemisia afra* samples (S1-S4) with S1 and S2 receiving solvent only and S3 and S4 spiked with 0.4 mg/mL of artemisinin. Separation based on the natural chemical differences between these samples is clearly visible. The natural chemical variation between S3 and S4 plays a big role in their separation even after the addition of 0.4 mg/mL of artemisinin

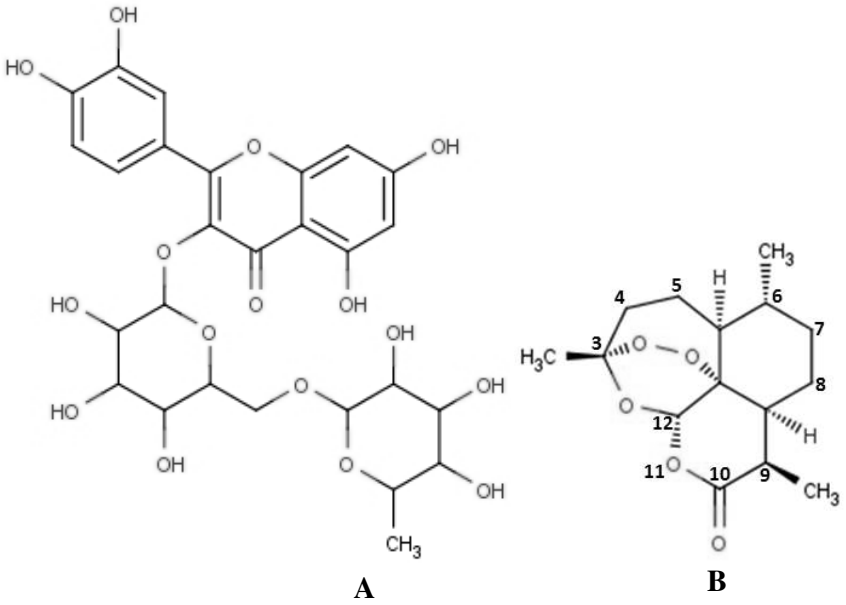

Fig. 4. Chemical structures of the added reference compounds a) rutin and b) artemisinin

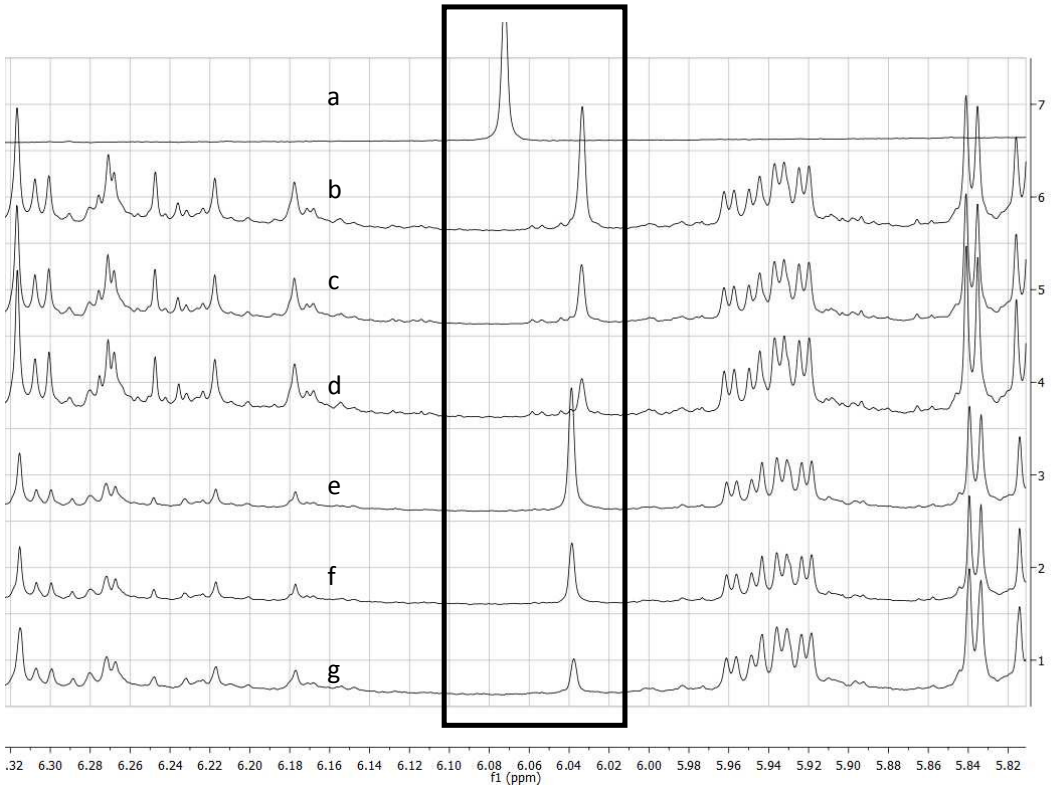

Fig. 5. NMR spectra of samples S3 and S4, highlighting the H-12 proton chemical shift. The movement of the chemical shift is evident as compared to the perfect alignment of the signals arising from unknown compounds between 5.92-5.96 ppm and the rest of the spectrum. Spectra represented are of (a) pure artemisinin; spiked S4 samples with (b) 0.4 mg/mL, (c) 0.2 mg/mL and (d) 0.1 mg/mL artemisinin and spiked S3 samples with (e) 0.4 mg/mL, (f) 0.2 mg/mL and (g) 0.1 mg/mL artemisinin

Table 2a. Top ten loading plot signals of spiked vs control samples. The signals from pure artemisinin falls into the following bins: 0.96-1.24, 1.40-1.48, 1.80-1.96, 2.00-2.20, 2.36-2.48, 6.04-6.12. Matching signals are given in bold font.

Spiking level (mg/mL)	clustering	Top ten loading plot signals (ppm)	Matching signals
300 MHz loading plot data			
0.05	yes	**1.12**, 1.36, 4.20, 4.12, 5.00, 5.04, 6.76, 6.80, 6.96, 7.08	1
0.10	yes	**1.12**, 1.36, **2.12**, **2.16**, 6.76, 6.80, 6.84, 6.96, 7.00, 7.08	3
0.20	yes	**1.12**, 1.36, **2.16**, 3.44, 6.00, 6.80, 6.84, 7.00, 7.08, 7.12	2
0.40	yes	**0.96**, **1.08**, **1.12**, **1.16**, 1.36, **1.40**, **1.44**, 1.76, **1.80**, 6.00	7
400 MHz loading plot data			
0.05	yes	**1.12, 1.16**, 2.16, 3.44, 6.80, 6.84, 6.96, 7.00, 7.12, 7.60	2
0.10	yes	**1.12**, **1.16**, 1.36, **1.44**, **1.80**, **1.84**, **2.16**, 3.44, 6.80, 7.08	6
0.20	yes	**0.96**, **1.08**, **1.12**, **1.16**, 1.36, **1.40**, **1.44**, 1.76, **1.80**, 3.44	7
0.40	yes	**0.96**, **1.08**, **1.12**, **1.16**, 1.36, **1.40**, **1.44**, **1.84**, **2.16**, 3.44	8

whereas the same concentrations of artemisinin added to sample 4 was 6.030 ppm, as compared to the chemical shift of pure artemisinin at 6.070 ppm (Fig. 5). Taking the movement of chemical shifts into account Table 2b gives the number of matching signals between the spiked samples at different concentrations and artemisinin.

4. DISCUSSION AND CONCLUSION

Due to the advantages of NMR spectroscopy over other techniques it is worthwhile to investigate the current limitations in order to define the causes for these limitations so that solutions can be presented which will improve this technique. Our current study aimed at doing exactly this by trying to give a numerical value to one of the main limitations of NMR based metabolomics namely its low sensitivity. In addition we also tested the effectiveness of the current protocol for drug discovery from medicinal plants.

We found that the sensitivity of the current protocol for the chosen plant material and reference compound rutin, was 0.2 mg/mL as measured on the 300, 400 and 500 MHz spectrometers, while artemisinin spiked to four chemically different samples gave separation at 0.05-0.10 mg/mL (due to movement of chemical shifts an exact concentration cannot be given). Although these values give us an indication of how sensitive the protocol is,

Table 2b. With the observed movement of chemical shifts for artemisinin in the samples, the methyl and H-12 proton signals have moved into the following bins: 0.96-1.00, 1.12-1.20, 1.36-1.40, 6.00-6.40. Due to peak overlap the movement of chemical shifts for the single proton signals could not be detected.

Spiking level (mg/mL)	Clustering	Top ten loading plot signals (ppm)	Matching signals
300 MHz loading plot data			
0.05	yes	**1.12**, **1.36**, 4.20, 4.12, 5.00, 5.04, 6.76, 6.80, 6.96, 7.08	2
0.10	yes	**1.12**, **1.36**, 2.12, 2.16, 6.76, 6.80, 6.84, 6.96, 7.00, 7.08	2
0.20	yes	**1.12**, **1.36**, 2.16, 3.44, **6.00**, 6.80, 6.84, 7.00, 7.08, 7.12	3
0.40	yes	**0.96**, **1.08**, **1.12**, **1.16**, **1.36**, **1.40**, 1.44, 1.76, 1.80,**6.00**	7
400 MHz loading plot data			
0.05	yes	**1.12, 1.16,** 2.16, 3.44, 6.80, 6.84, 6.96, 7.00, 7.12, 7.60	2
0.10	yes	**1.12**, **1.16**, **1.36**, 1.44, 1.80, 1.84, 2.16, 3.44, 6.80, 7.08	3
0.20	yes	**0.96**, 1.08, **1.12**, **1.16**, **1.36**, **1.40**, 1.44, 1.76, 1.80, 3.44	5
0.40	yes	**0.96**, 1.08, **1.12**, **1.16**, **1.36**, **1.40**, 1.44, 1.84, 2.16, 3.44	5

we have to state clearly that these values will be dependent on the sample matrix and the nature of the reference compound. For example, artemisinin contains three methyl groups, which will give a signal response three times stronger than single protons due to the molar equivalency of NMR spectroscopy.

Nevertheless we can conclude that the current NMR based metabolomics analyses of plants will be able to differentiate between two sample sets at a concentration of low µg concentrations (± 50- 200 µg/mL) or low µM levels (177-328). It was also clear that at the lower concentrations tested, instrumental variation played a more significant role in differentiation between replicates of the same samples. It is therefore advisable to analyse at least duplicate samples when employing this protocol.

Our main findings in this study are that we do not only lose the improved resolution (during binning of the spectra) but also the higher sensitivity of higher field NMR instruments using the current protocol in the first step of metabolomic analysis. We must remember that the enhanced sensitivity and resolution of higher field NMR instruments will still be advantageous during the second step of metabolomics namely during compound identification. A more direct method to identify differences between two chemical profiles is to overlay and visually inspect the spectra. For example, the H-12 proton of artemisinin is clearly visible by visual inspection (Fig. 5) and yet the methodology did not detect this signal at the lower concentrations tested; the 400 MHz analysis did not detect this signal at all. Overlaying spectra works well if the number of samples is five or less or for the trained eye 10 samples or less. Visually analysing more than 10 samples becomes very difficult and error-prone. This is the main reason why data reduction techniques and multivariate data analyses are currently being used. It is acknowledged that there will be some loss of data (e.g. loss of resolution during binning), but at the moment it is the best way of automatically performing this task on large data sets. Another disadvantage of using the current protocol is that we also lose the improved sensitivity of higher field magnets. It appears that it is rather the integral ratio of signals (integral value of spiked compound versus the integral value of all other compounds in the sample) that leads to positive differentiation between sample sets and not the absolute integral value of these signals. This is a fixed ratio and the instrumental sensitivity does not play a role in this differentiation. This fixed ratio will be completely dependent on the material used and the compound added. This limitation to the current protocol needs to be addressed. Future work will therefore focus on finding ways to make full use of the higher instrumental sensitivity of higher field magnets during the first step of metabolomics analysis. Crutchfield et al. [16] has recently published a LC-MS technique in which data is presented in overlayed panes without making use of multivariate data analysis. It should be possible to adapt this technique for NMR generated data.

Our application analyses performed on four different samples revealed that the current protocol can potentially work well to identify bioactives in medicinal plants. It detected the bioactive compound artemisinin at concentrations at which it occurs in nature [17]. We should, however, clearly state that this will only work if the bioactive compound is present at relatively high concentrations (>0.1%). If we look at other natural products and the concentration at which they occur in nature then we have to conclude that at the moment this technique needs to be improved before it can used more widely. For example, Palmyra et al. [18] found that the content of salicin (precursor of salicylic acid) varies between 0.08 and 12.6% in willow bark. The anticancer drug Paclitaxel (taxol) occurs at a concentration of <0.01-0.05% in dried *Taxus brevifolia* needles [19], vinblastine from *Catharanthus roseus* at a concentration of 0.00024-0.00059% [20] and galanthamine from *Narcissus pseudonarcissus* at a concentration of around 0.2% [21]. Based on these examples we can conclude that the current NMR based metabolomics protocol might have worked well for the discovery of artemisinin, salicin and maybe even galanthamine but the concentration of taxol and vinblastine are far beyond the detection limit of this technique.

Another finding in this study was that the reference compound artemisinin displayed some movement in chemical shifts. This difference in chemical shift was big enough to cause the methodology to detect the right signal (only at high concentrations) but at the wrong place. For example the loadings plot highlighted 6.0-6.4 ppm as causing differentiation between the samples for the 300 MHz dataset at a spiking level of 0.4 mg/mL (Table 2a). This signal is indeed caused by artemisinin but it has moved into an adjacent bin. The same signal of the pure reference compound falls into the 6.04-6.08 bin

(chemical shift 6.07 ppm). The other prominent signals of artemisinin, specifically the methyl signals, displayed a similar movement. It appears that these signal shifts is not caused by a difference in concentration or temperature. We used buffered solvents and kept the temperature constant. Instrumental variation can also be excluded as the exact same shift occurred in both the 300 and 400 MHz results. This has a number of implications. It appears that other compounds present in the samples (which are different or at different concentrations between sample 3 and sample 4) have an influence on the chemical shift of artemisinin. If this is the case it will complicate matters when database searches are conducted in order to identify compounds (databases contains spectra generated from pure reference compounds). It can also lead to positive separation between two sample sets caused by the same compound. This specific signal of artemisinin is also being used to quantify artemisinin in *Artemisia annua*. Movement of this signal will therefore potentially complicate this otherwise rapid quantitation method [22]. To explain this observation will be difficult. One explanation is that other compounds in the samples interact chemically with artemisinin causing the observed shifts. LC-MS analysis revealed this not to be the case giving the exact same retention time and fragmentation pattern for artemisinin in all samples tested. Although we cannot offer an explanation for this observation we should ask the question if this phenomenon is limited to only artemisinin or if it is more widespread. If it is more widespread it might seriously affect the use of NMR in metabolomics analysis.Fig. 5 does give us some indication as none of the non-artemisinin signals showed any significant movement in chemical shifts (comparing signals between sample 3 and sample 4). This phenomenon might therefore be limited to sesquiterpene lactones although this has to be investigated in future studies.

CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

ACKNOWLEDGEMENTS

We would like to thank the School of Science and Health, University of Western Sydney for sponsoring this work through a seed grant and also the centralised NMR facility of the University of Western Sydney for the use of the NMR equipment. Although the data is not used in this manuscript we also like to thank Dr Wilford Lie at the University of Wollongong for the use of their 500 MHz spectrometer and Mr Shaun Sullivan who analysed some of the first samples.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Van der Kooy F, Maltese F, Choi YH, Kim HK, Verpoorte R. Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. Planta Med. 2009;75:763-775.
2. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22:245–252.
3. Villas-Boas S.G, Rasmussen S, Lane G.A. Metabolomics or metabolite profiles? Trends Biotechnol. 2005;23:385–386.
4. Daviss B. Growing pains for metabolomics. The Scientist, Canada: LabX Media Group. 2005;19:25–28.
5. Kim HK, Choi YH, Verpoorte R. NMR-based metabolomic analysis of plants. Nature Prot. 2010;5:536-549.
6. Fiehn O, Robertson D, Griffin J, Van der Werf M, Nikolau B, Morrison N, et al. The metabolomics standards initiative (MSI). Metabolomics. 2007;3:175–178.
7. Ludwig C, Easton JM, Lodi A, Tiziani S, Manzoor S, Southam AD, et al. Birmingham metabolite library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). Metabolomics. 2012;8:8-18.
8. Zhou J, Yu S, Wang Y, Gu X, Wu Q, Xue Y, et al. Serum metabolite profiling of B-cell non-Hodgkin's lymphoma using UPLC-QTOFMS and GC-TOFMS. Metabolomics; 2013. DOI 10.1007/s11306-013-0596-8
9. Ahmed SSJ. Systems biology in unruptured intracranial aneurysm: a metabolomics study in serum for the

detection of biomarkers. Metabolomics. 2014;10:52–62.
10. Sana TR, Fischer S, Wohlgemuth G, Katrekar A, Jung K-H, Ronald PC, et al. Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen *Xanthomonas oryzae* pv. Oryzae. Metabolomics. 2010;6:451-465.
11. Cardoso-Taketa AT, Pereda-Miranda R, Choi YH, Verpoorte R, Villarreal ML. Metabolic profiling of the Mexican anxiolytic and sedative plant *Galphimia glauca* using nuclear magnetic resonance spectroscopy and multivariate data analysis. Planta Med. 2008;74:1295-1301.
12. Yuliana ND, Khatib A, Verpoorte R, Choi YH. Comprehensive extraction method integrated with NMR metabolomics: a new bioactivity screening method for plants, adenosine A1 receptor binding compounds in *Orthosiphon stamineus* Benth. Anal Chem. 2011;83:6902–6906.
13. Van der Kooy F, Verpoorte R, Meyer JJM. Metabolomic quality control of claimed anti-malarial *Artemisia afra* herbal remedy and *A. afra* and *A. annua* plant extracts. S Afr J Bot. 2008;74:186-189.
14. Liu NQ, Cao M, Frédérich M, Choi YH, Verpoorte R, Van der Kooy F. Metabolomic investigation of the ethnopharmacological use of *Artemisia afra* with NMR spectroscopy and multivariate data analysis. J Ethnopharm. 2010;128:230-235.
15. Mitra A, Seaton PJ, Aliassarpour R, Williamson T. Unprecedented concentration dependent chemical shift variation in H-NMR studies: A caveat in the investigations of molecular recognition and structure elucidation. Tetrahedron. 1998;54:15489-15498.
16. Crutchfield CA, Olson MT, Gourgari E, Nesterova M, Stratakis CA, Yergey AL. Comprehensive analysis of LC/MS data using pseudocolor plots. J Am Soc Mass Spectrom. 2013;24:230-237.
17. De Ridder S, Van der Kooy F, Verpoorte R. *Artemisia annua* as a self-reliant treatment for malaria in developing countries. J Ethnopharm. 2008;120:302-314.
18. Kenstaviciene P, Nenortiene P, Kiliuviene G, Zevzikovas A, Lukosius A. Kazlauskiene D. Application of high-performance liquid chromatography for research of salicin in bark of different varieties of *Salix*. Medicina. 2009;45:644-651.
19. Van Rozendaal ELM, Lelyveld GP, Van Beek TA. Screening of the needles of different yew species and cultivars for paclitaxel and related toxoids. Phytochem. 2000;53:383-389.
20. Chu IH, Bodnar JA, Bowman RN, White EL. Determination of vincristine and vinblastine in *Catharanthus roseus* plants by high performance liquid chromatography/electrospray ionization mass spectrometry. J. Liq. Chromatogr. Relat. Technol. 1997;20:1159-1174.
21. Lubbe A, Gude H, Verpoorte R, Choi YH. Seasonal accumulation of major alkaloids in organs of pharmaceutical crop *Narcissus* Carlton. Phytochem. 2013;88:43-53.
22. Liu NQ, Choi YH, Verpoorte R, Van der Kooy F. Comparative quantitative analysis of artemisinin by chromatography and qNMR. Phytochem Anal. 2010;21:451–6.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=685&id=13&aid=6622