Short Communication

Profiling of polar metabolites in biological extracts using diamond hydride-based aqueous normal phase chromatography

Highly polar metabolites, such as sugars and most amino acids are not retained by conventional RP LC columns. Without sufficient retention low concentration compounds are not detected due ion suppression and structural isomers are not resolved. In contrast, hydrophilic interaction chromatography (HILIC) and aqueous normal phase chromatography (ANP) retain compounds based on their hydrophilicity and therefore provides a means of separating highly polar compounds. Here, an ANP method based on the diamond hydride stationary phase is presented for profiling biological small molecules by LC. A rapid separation system based upon a fast gradient that delivers reproducible chromatography is presented. Approximately 1000 compounds were reproducibly detected in human urine samples and clear differences between these samples were identified. This chromatography was also applied to xylem fluid from soyabean (Glycine max) plants to which 400 compounds were detected. This method greatly increases the metabolite coverage over RP-only metabolite profiling in biological samples. We show that both forms of chromatography are necessary for untargeted comprehensive metabolite profiling and that the diamond hydride stationary phase provides a good option for polar metabolite analysis.

Keywords: Aqueous normal phase / ESI / Hydrophilic interaction chromatography / LC-MS / Metabolomics

Received: March 25, 2009; revised: April 8, 2009; accepted: April 9, 2009
DOI 10.1002/jssc.200900171

1 Introduction

The metabolome of either an organism or a biological fluid is a complex mixture of chemically diverse compounds covering a vast concentration range. Currently it is not possible to use a single chromatographic technique to study the full complement of the metabolome. The most commonly applied technique for profiling polar metabolites is GC-MS [1]. GC-MS is a very rapid and robust technique for high throughput analysis, however, it has a number of limitations. Derivatisation is required for compounds with polar functional groups such as alcohols, carboxylic acids and amines in order to increase their volatility and decrease their thermal instability. The addition of alkyl silyl groups from derivatisation inhibits hydrogen bonding. This reduction in polarity increases volatility which allows the separation of metabolites by GC [2]. For standard quadrupole and TOF GC-MS instruments the increase in molecular mass limits the size of the molecules capable of being detected. LC-MS provides a less biased method for the detection of metabolites as derivatisation is not required but suffers from being unable to separate the vast range of metabolites on a single column type [2]. There are two approaches for metabolite profiling by ESI/atmospheric pressure chemical ionisation (APCI)-MS. (i) Direct infusion or flow injection of a sample. This does not provide chromatographic resolution but is ideally suited for fast screening of large data sets [3–5]. Information can be lost when samples are directly infused into an ESI source, isomers cannot be distinguished and ion-suppression causes signal loss for many compounds. For example, the three metabolic breakdown products of caffeine: theobromine, paraxanthine and theophylline, are structural isomers and can only be distinguished using chromatography (Fig. 1). (ii) Chromatographic separation of metab-
olites prior to MS analysis. This is preferable for a number of reasons when analysing complex mixtures. The use of chromatography allows the assignment of the retention time, molecular mass and intensity to designate a ‘mass feature’. The use of the retention time as an extra parameter for analysis increases the number of distinguishable features and results in a more detailed comparison of metabolite profiles.

RP columns provide fast, reproducible chromatography and due to the availability of smaller silica particle sizes resolution has increased. Columns packed with sub-2 μm diameter stationary phases can produce peak widths of less than 6 s. There are commercially available RP columns such as the T3 (Waters, Milford, MA, USA), SB-Aq (Agilent, Santa Clara, CA, USA) and Synergi Hydro-RP (Phenomenex, Torrance, CA, USA) which have improved retention of polar compounds and offer different selectivity to a standard C8 or C18 stationary phase. These columns have a polar functional group such as an amide or carbamate incorporated into the alkyl ligand. The increased wettability of these phases allows their use with highly aqueous mobile phases without solvent being driven out of the pores due to surface tension effects which would result in alkyl chain collapse [6, 7]. However, since the retention mechanism still generally relies on hydrophobic interactions these columns do not retain highly polar compounds, such as sugars, carboxylic acids and most amino acids. The retention of highly polar compounds is one of the most challenging aspects in LC. Although ion pairing agents can be added to mobile phases to achieve retention of polar molecules they are problematic for use in ESI-MS. For biological fluids, such as urine or plant xylem exudate, molecules are generally polar as they are already dissolved in an aqueous solvent. Two techniques for LCMS analysis of hydrophilic compounds are hydrophilic interaction LC (HILIC) and aqueous normal phase (ANP). HILIC is described as the partitioning of the solute into the hydrated surface of the stationary phase [8]. A clear model describing ANP chromatography has not been developed, however, it is understood that the solute directly interacts with the stationary phase. This direct interaction allows both hydrophilic and hydrophobic retention[9]. Generally, more polar molecules are retained longer on the stationary phase. Retention is also affected by the pK$_a$ of the solutes and steric hindrance by the polar functional group [10]. An example of the complementarity between RP and ANP is with the analysis of two common metabolites in urine, creatine and creatinine. No retention of these compounds is observed using standard RP chromatography (Fig. 2). However, ANP retains these two compounds beyond the void volume allowing chromatographic resolution (Fig. 2).

Simple un-bonded silica and chemically-bonded amino, amide, carbamate, cyano and polyol ligands have been the most commonly used HILIC stationary phases [11]. Many of these stationary phases have been used for metabolite profiling [12–15]. Typically, HILIC columns are slow to re-equilibrate after gradient elution and do not offer the resolution that an RP column provides. This is incompatible with the fast gradient and re-equilibration times that are required for the high-throughput demands of metabolomics. Recent developments in stationary phases for HILIC and ANP, such as zwitterionic and diamond hydride silica-based stationary phases, have vastly improved the chromatography of polar compounds. These columns provide good separation and reproducible retention times which is extremely important when comparing complex chromatograms across large data sets [16–18].

This paper presents metabolic profiling data using a diamond hydride column for the separation of polar metabolites in biological fluids by ANP-ESI-quadrupole TOF (QTOF)-MS. The methodology is based on work first published by Pesek et al. [17] and both the benefits and the limitations of this method will be discussed.

2 Experimental

2.1 Materials

LCMS grade ammonium acetate and formic acid were purchased from Sigma–Aldrich (Sydney, Australia). HPLC/pesticide grade ACN and methanol from Burdock and Jackson (Ajax, Sydney, Australia) and 18.2 MΩ deionised water was used throughout experiments. The ANP column used was the Cogent diamond Hydride
2.1 mm × 100 mm, 4 μm particle size (MicroSolv Technology, Brisbane, Australia). The silica-based stationary phase contains ~2% carbon as the organic moiety on a silica hydride surface [19]. For RP chromatography a Zorbax Eclipse XDB-C18, 2.1 mm × 100 mm, 1.8 μm (Agilent) column was used. An in-house metabolite mix containing a mixture of sugars, amino acids and organic acids was analysed (Supporting Information). All compounds were purchased from Sigma–Aldrich.

2.2 Sample preparation

Two mid-stream urine samples were collected from an anonymous volunteer on consecutive mornings. The second sample was collected 3 h after the ingestion of both a coffee and a multivitamin tablet (Blackmores, Sydney, Australia). Protein precipitation of the urine sample was carried out by the addition of ACN to give a final ratio of 1:3 v/v urine/ACN. The samples were vortexed then centrifuged (10 min, 13,000, 4°C) and transferred to a vial for analysis. The high organic content is compatible with ANP chromatography so samples could be directly injected onto the LC column. Xylem fluid from soyabean (Glycine max) was collected as per [20]. Xylem samples were diluted 1:3 v/v xylem/buffer B (90% ACN, 10 mM ammonium acetate) and then centrifuged prior to injection. For RP analysis an aliquot of the urine supernatant was dried under vacuum and re-dissolved in 50% methanol.

Figure 2. EICs of creatine and creatinine in a urine sample. Sample was analysed by RP (top two chromatograms) and ANP (bottom two chromatograms). No retention of these two compounds is observed by RP, however, good retention and separation is achieved using ANP.
2.3 LC-MS system details and method

2.3.1 LC

The LC system used was an Agilent 1200 series (Santa Clara, CA, USA) comprising a vacuum degasser, binary pump, with a thermostated autosampler and column compartment. The organic mobile phase solvent (B) was composed of 90% ACN v/v with 0.1% ammonium acetate w/v and 0.1% acetic acid v/v. The aqueous mobile phase (A) was composed of 100% deionised water with 0.1% ammonium acetate w/v and 0.1% acetic acid v/v (pH 3.4). A moderate pH buffer was used as this method was developed for both positive and negative ion MS. The column flow-rate was 0.4 mL/min and column temperature was 50°C. The optimal gradient started at 100%B then linearly decreased to 40%B over 10 min, followed by a 1 min hold at 40%B. The column was then re-equilibrated at 100%B for 6 min. Gradient times from 5 to 60 min were trialled to determine the best compromise between the number of features detected and run-time.

Some analyses were carried out using RP chromatography for comparison with the ANP data. A 10 min linear gradient from 95:5 v/v water/ACN to 5:95 v/v water/ACN at 0.4 mL/min was used and the column temperature was held at 50°C. Both mobile phases contained 0.1% formic acid v/v.

2.3.2 MS

An Agilent 6520 QTOF MS system (Santa Clara, CA, USA) was used with a dual sprayer ESI source. ESI conditions were: nebuliser pressure 45 psi, gas flow-rate 10 L/min, gas temperature 300°C, capillary voltage 4000 V, fragmentor 150 V and skimmer 65 V. Instrument was operated in the extended dynamic range mode and data collected in m/z range 70–1700 amu. Both positive and negative ion modes were employed.

2.4 Data analysis

Mass features were extracted using Agilent’s MassHunter Qual software. The feature extraction (chromatographic peak deconvolution algorithm) searched for peaks above a 1000 count threshold (Fig. 3). These data were then imported into GenespringMS (Agilent) where retention times were aligned and peak intensities normalised. The default normalisation protocol within GenespringMS uses a percentile shift, which assumes that in the absence of instrument variation, the total concentration of metabolites in each sample is equal. This assumption is rarely correct for most biological extracts. For example, when collecting xylem fluid it is difficult to collect samples with the same volume and overall metabolite concentration. The same is observed with urine analysis; gender, age, time of day, lifestyle, diet and health or disease state are all variables which can affect urine composition [21]. These factors result in large sampling variations. For urine, creatinine concentration is sometimes used as a normalisation factor however if the disease state effects the excretion of creatine this method is no longer valid [22]. One approach for normalisation is based on the area under the total ion chromatogram (TIC). This is a more appropriate method for these samples. Integration of the TIC allows for the use of a correction factor for each sample.

3 Results and Discussion

3.1 Optimisation of LC conditions

Si–OH based stationary phases used for HILIC generally result in broad peaks with poor retention time reproducibility and are slow to re-equilibrate. The diamond hydride column separates compounds with much better efficiency giving sharp peaks, good retention time stability and fast column re-equilibration times.

Of the gradients trialled from 5 to 60 min, the 10 min gradient offered the best compromise between the numbers of compounds detected and sample throughput. There was no significant advantage in using shallower gradients (Table 1). The flow-rate was also optimised. Increasing the flow-rate from 0.4 to 0.5 mL/min did not improve peak shape, however, an ~30% loss in sensitivity was observed. The diamond hydride ANP column also provided good retention time stability. Five overlaid extracted ion chromatograms (EICs) for arginine in a urine sample show little shift (±0.05 min) in retention time.
time (Fig. 4). RSDs for retention times of five injections from both urine samples ranged between 0.1–0.9% and 4–6% for area RSDs (n = 5).

It has been reported that more features are detected using RP chromatography in comparison to HILIC [14–16]. In this study, the ANP diamond hydride column detected similar numbers of features as the RP column. Using RP chromatography, a mean of 849 features was determined with an RSD of 1.67% (n = 5). The ANP method gave a mean of 964 features with an RSD of 1.34% (n = 5). The deconvoluted chromatograms from the 10 min gradient were imported into GenespringMS to determine the number of features conserved over the five replicate injections. It was found that 582 features were consistently detected. This number however is dependant on the ability of the software (GenespringMS) to align chromatograms and correctly assign a unique mass feature to all detected compounds therefore it is not surprising that the number or conserved features across all replicates is less than the average reported in Table 1.

However, the diamond hydride column does not provide good chromatography for all metabolites using the conditions described here. For example, citric acid and other tri-carboxylic acids such as the plant phytosiderophore nicotianamine, gave broad irreproducible peak shapes. Further method optimisation may improve the peak shape of these compounds. Good peak shape for citric acid using this column has been reported using an alternative gradient and different mobile phase pH [17].

Studies using an in-house metabolite mix (Supporting Information) containing a range of sugars, amino acids and organic acids showed that amino acids elute between 3 and 8.5 min, sugars between 1 and 5 and some small organic acids, such as succinate, malate and fumarate throughout the whole gradient. A detailed characterisation of the chromatography of these compounds using various gradients has been previously published [17].

A paper comparing other HILIC stationary phases has been published [13]. It was found that an aminopropyl column at pH 9.45 provided the most effective separation of polar cellular metabolites. However, the run-times are much longer (>30 min) than the method described here and the high mobile phase pH used reduces the overall column lifetime. There are also methods published using the Seaquant ZIC-HILIC column (Merck, Melbourne, Australia) for the separation of water soluble metabolites [16, 18, 23, 24]. These methods use similar run times and mobile phases and have provided comparable chromatography to the diamond hydride column.

### Table 1. Number of features detected with different gradient length

<table>
<thead>
<tr>
<th>Replicate injection number</th>
<th>Number of features deconvoluted in each gradient number</th>
<th>5 min</th>
<th>10 min</th>
<th>15 min</th>
<th>20 min</th>
<th>30 min</th>
<th>60 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>901 969 1043 1066 1048 1047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>873 984 1059 1029 1033 1034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>913 952 1025 1020 1064 1071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>894 954 1026 1040 1028 1037</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>889 963 1035 1030 1055 959</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>894 964 1038 1037 1046 1030</td>
<td>1.66</td>
<td>1.34</td>
<td>1.35</td>
<td>1.71</td>
<td>1.44</td>
<td>4.08</td>
</tr>
<tr>
<td>%RSD</td>
<td>1.66 1.34 1.35 1.71 1.44 4.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Urine analysis

Approximately 1400 unique mass features from the five replicate injections of the two urine samples were imported into GenespringMS for analysis. This software contains a number of filtering tools as well as multivariate and univariate statistics for determining differences between LC-MS profiles. After filtering the urine profile data on relative frequency (feature is removed if not present in all technical replicates) and a ten-fold change filter, 144 unique mass features were identified in the sample taken after ingestion of both a multivitamin and coffee. The data were presented as a dendrogram by GenespringMS through the use of hierarchical cluster analysis with Pearson correlation (Fig. 5). Riboflavin was one of the 144 mass features identified as having a greater than ten-fold difference between the two sample sets (Fig. 5). The fold-change is calculated using the peak intensities determined by the molecular feature extractor. This difference can clearly be seen when examining

![Figure 4. Five overlaid EIC arginine (m/z 175.1195) from repeat injections of a urine sample.](image-url)
the EICs from the raw data files (Fig. 5). The chromatographic peak is not obvious in the TIC and can only be found after peak deconvolution and comparison. Riboflavin is also well retained by RP chromatography but the more polar compounds, such as a high intensity ion with $m/z$ 144.1022 at 8.7 min, are not retained by RP and has a much lower intensity. A library search suggested this compound was proline betaine (only identified through library searching) and is found to be 99-fold higher in the urine sample taken post-caffeine and multivitamin ingestion. Generally, the low concentration compounds detected by HILIC and ANP are not detected when using RP chromatography as they co-elute in the void volume where ion-suppression reduces signal intensity. The same occurs with metabolites detected by RP. An expected difference between the two sample sets is the presence of caffeine and its metabolites. However, using ANP chromatography these mass features were not deconvoluted by MassHunter as they are not retained on the column. The elution of caffeine in the void volume.

**Figure 5.** Top panel is the GenespringMS representation of mass features after filtering. Samples are technical replicates of the two urine samples (pre/post-multivitamin and coffee) run on the ANP column in positive ionisation mode. The vertical rows represent the mass feature profile for each sample. Each rectangle represents the intensity of the mass feature; darker rectangles have higher intensities. The samples can be visually compared across horizontal rows. Bottom panel are overlayed EICs at $m/z$ 377.1482 (riboflavin) showing the large differences between samples sets in the raw data.
results in a low intensity non-Gaussian peak at RT 1.5 min (Fig. 6). Caffeine is easily detected when samples are analysed using RP chromatography (Fig. 6). This illustrates the importance of selecting the correct chromatography to target the groups of metabolites of interest and also how using both forms of chromatography results in a wider range of compounds detected. There is no single column and ionisation mode which enables the simultaneous separation and detection of all metabolites. Orthogonal chromatography using both RP and diamond hydride-based ANP with fast polarity switching may overcome some of these limitations [25, 26]. Furthermore, as there are more than 1000 compounds detected in one run, these compounds are not all resolved (e.g. Fig. 3) and this will also lead to ion-suppression/enhancement [27].

4 Conclusions

The diamond hydride column offers a fast and reproducible method for polar profiling of biological samples by LC-MS. The two urine samples taken from the same person on consecutive days provided a complex mixture of metabolites which when chromatographed on the diamond hydride-based ANP method for polar metabolite profiling was able to distinguish many features between the two samples. In order to cope with these differences an accurate normalisation protocol is required. Quality control samples are important in order to assess system suitability. To date we have found normalisation to be one of the greatest challenges when comparing chromatograms from biological samples. In order to increase confidence, larger datasets with more replication are required.

The experiment described here was a simple test of the methodology. The complexity of the dataset increases as more biological variation is introduced through different patients and more biological replicates. For example, a study involving urine samples from many patients would result in many thousands of features for analysis. This type of study requires highly reproducible chromatography as well as stable instrument calibration. The analyst needs to be confident that differences observed are due to biological variation and not the instrument and/or software. The software relies on these parameters to filter data. The method described here allows the analysis of complex mixtures of polar compounds for metabolomic studies.

The Victorian node of Metabolomics Australia is funded through Bioplatforms Australia Pty Ltd, a National Collaborative Research

![Figure 6. Top panel: EIC for caffeine (m/z 195.0890) using ANP chromatography. Bottom panel EIC (m/z 195.0890) for caffeine using RP chromatography, inset is a zoomed in area over the peak.](image-url)
Infrastructure Strategy (NCRIS), 5.1 biomolecular platforms and informatics, co-investment with the Victorian State government and The University of Melbourne. The authors would like to thank Theodore Sana and Steven Fisher (Agilent technologies Santa Clara) for discussions on urine sample preparation and Professor Joseph and Dr Maria Pesek (San Jose State University, USA) for discussions on chromatography. We would also like to thank Professor Peter Gresshoff and Dugald Reid (University of Queensland, Australia) for providing the xylem sap samples.

The author declared no conflict of interest

5 References