The use of SIFT-MS in profiling the faecal volatile metabolome in horses with colic: a pilot study

Claire Turner\textsuperscript{1}, Claire Batty\textsuperscript{1}, Ebony Escalona\textsuperscript{2}, John O. Hunter\textsuperscript{3}, Christopher Proudman\textsuperscript{2}

\textsuperscript{1}Department of Life, Health and Chemical Sciences, Open University, Milton Keynes, MK7 6AA, UK; e-mail: c.turner@open.ac.uk

\textsuperscript{2}Department of Gastroenterology/School of Veterinary Science, University of Liverpool, Neston, CH64 7TE, UK.

\textsuperscript{3}Gastroenterology Research Unit, Box 262, Addenbrooke’s Hospital, Cambridge CB2 0QQ, UK.

ABSTRACT

SIFT-MS is used for the first time in profiling the volatile organic profile in faecal headspace in two groups of horses admitted to an equine hospital, one group with acute intestinal disease (colic) affecting the large colon, plus a control group of similarly managed horses admitted for non-gastrointestinal/metabolic reasons (e.g. acute orthopaedic injury). Compounds in faecal headspace which show statistically significant concentration differences between the groups are acetone and methanol. In addition, some ions at various \(m/z\) values show significantly different ion counts between the groups. Further information may be gleaned by using multivariate statistics in evaluating the differences between the two horse groups. Principal components analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) were evaluated for reducing the dimensionality of the SIFT-MS data, and OPLS-DA was found to be best at discriminating between the groups, particularly with SIFT-MS data acquired using the \(\text{H}_3\text{O}^+\) precursor ion. Analysis of these data also show the significance of ammonia as a discriminating ion. These results show that SIFT-MS may potentially be used on the headspace of horse faecal samples for detecting altered microbial fermentation associated with acute intestinal disease of the colon.

Keywords:

SIFT-MS, metabolic profiling, horse volatile metabolome

1. INTRODUCTION

Analysis of volatile organic compounds (VOCs) has now been established as a potential non-invasive method of diagnosing disease or monitoring human health [1]. Breath analysis has been around since the 1970s, when breath was condensed in a cold trap and then analysed by GC-MS [2]. Analysing breath isn’t always convenient, however, because samples are hard to store and it isn’t always possible to get the individual to the analytical instrument. Hence other samples are also analysed for the VOCs present in their headspace. These include urine, blood, faeces, skin, sputum and in fact any tissue type or body fluid. Some of these samples will provide systemic data (e.g. breath, blood and urine), while others are better at providing more localised data relevant to particular organs or parts of the body (e.g. sputum, faeces).
GC-MS is a powerful technique as it traditionally uses electron ionisation–quadrupole mass spectrometry after compound separation (although time–of-flight mass spectrometry is increasingly now being used), and there is a large library against which spectra may be searched. This will enable the identification of over 200,000 compounds. If coupled with thermal desorption or solid phase micro extraction (SPME), samples of breath or headspace may be concentrated prior to injection and separation, thus making the technique highly sensitive so that at the current time, many hundreds of compounds may be detected in breath. However, the complexity of the samples and the combined spectra that result mean that analysis is often difficult, particularly when there are many samples. Although GC-MS is good for identifying compounds, the technique is complex, time consuming and not completely quantitative. In addition, the potential for the analysis of so many different compounds may counter-intuitively make analysis more difficult as many of those present at very low concentrations may just be exogenous environmental contaminants and unrelated to a disease or even normal physiology. For this reason, other techniques have been developed, and one such technique is selected ion tube mass spectrometry, SIFT-MS. The details of this technique have been extensively covered elsewhere [3, 4] so are not included here. The advantages of SIFT-MS are that it is rapid, quantitative and relatively sensitive and the data are readily converted for multivariate analysis of spectral profiles if required.

SIFT-MS has been applied to the analysis of various compounds associated with a number of diseases including cancer [5, 6, 7, 8], diabetes [9]; tuberculosis [10]; kidney disease [11, 12]; using breath and the headspace of clinical samples [1] plus the determination of breath profiles of healthy volunteers of adults [13-18] and children [19, 20].

Recently, SIFT-MS has been used to investigate the change in faecal VOCs in horse manure before and after taking an enzyme rich food supplement [21]. Adding enzymes (amylase) to feed enables the amount of carbohydrate that may be given to a horse to be increased, without affecting their health. This is significant because the horse has a digestive strategy heavily reliant upon the hindgut fermentation of dietary fibre [22]. Intestinal bacteria in the caecum and colon hydrolyse dietary fibre, releasing soluble carbohydrates. These are fermented to short chain fatty acids (acetic, propionic and butyric acid), providing horses with most of their energy requirements [23]. However, horses which are required to undertake excessive physical activity, for example racehorses, require more energy than may be obtained from hindgut fermentation alone. Dietary supplementation with readily hydrolysable carbohydrate, commonly in the form of grain, is often used to alleviate this, however, it is widely recognised that this is associated with an increased risk of intestinal disease [24-28]. Acute intestinal disease, also known as colic, is the single most important cause of mortality in horses and a significant cause of morbidity and economic loss in managed horse populations [29, 30]. It also appears to be a major cause of reduced productivity in working equids globally [31]. Hence, methods enabling the relationship between equine health, disease and diet to be investigated are likely to have a positive impact on horse welfare and associated socio-economic
factors (e.g. productivity of equine businesses).

There is a growing understanding of the complex interactions between host microbiota and metabolism [32, 33]) and diet plays a key role in this [34]The faecal metabolome reports specifically on the metabolic interplay between host, diet and intestinal microbiota and offers the potential to identify biomarkers which can act as proxy for specific bacterial populations [35, 36]. Faecal VOCs and VOC profiles have been found to be associated with a number of gastro-intestinal diseases in humans [37]. In studies in humans, urine is a convenient sample which is easy to collect and store and it has a relatively clean spectral profile (fewer ions) compared with faecal samples. However, urine samples are difficult to obtain non-invasively from horses, whereas faeces are abundant, frequently produced and therefore easily collected. Many population-based metabolomic studies report great diversity in metabotype due to factors including age, diet and genetic diversity [38-41].

In this study, we present SIFT-MS data of compounds present above the headspace of horse faeces and investigate the use of SIFT-MS in analysing changes in the equine gut volatile metabolome (i.e. the range of volatile compounds) associated with acute intestinal disease.

2. METHODS

2.1 Study population and sampling

The population of horses in our study were those admitted to the University of Liverpool’s Philip Leverhulme Equine Hospital. One group of six horses were admitted with colic and the second group of six horses, which acted as a control group, had other non-physiological problems, such as limb fractures. All horses were Thoroughbreds or Thoroughbred crosses aged 5 – 19 years old. All horses were fed mixed diets consisting of preserved forage (hay or haylage) at 60-80% dry matter intake plus concentrate feed in the form of proprietary horse cubes or cereal mix. All horses were stabled but most horses also had access to grass for a limited period each day. Horses were selected for sampling on the basis of acute onset of the problem being treated, allowing samples to be collected prior to the administration of pharmaceuticals that might influence the volatile metabolome, antibiotics in particular. Colic cases selected for sampling were those with a diagnosis of simple colonic obstruction and distension (SCOD colic[27]).

Faecal samples were obtained during the course of manual evacuation of the rectum for clinical examination of colic cases or, in the case of non-colic controls, as spontaneously voided faeces. Sampling and metadata collection were carried out under University of Liverpool ethics approval RETH000363, with the informed consent of owners. Fresh faeces from all horses were collected into polyethylene sample containers and immediately frozen in liquid nitrogen prior to transfer into a freezer at -80°C, where they were kept until analysis.

2.2 Headspace preparation
Frozen faecal samples were defrosted at room temperature. Five grams of each sample was removed from the bulk sample and placed inside a 40 cm long Nalophan sampling bag, made up of 65 mm diameter Nalophan NA tubing 25 μm thick (Kalle UK). One end of the bag was fitted with a Swagelok connector and the other end of the bag was sealed and filled with hydrocarbon free air to generate the VOC headspace. The bags were incubated at 40°C for 1 hour to allow the volatile organic compounds (VOCs) to equilibrate between headspace and solid sample. After 1 hour, the headspace was analysed by attaching the Swagelok fitting to the capillary inlet of the SIFT-MS instrument.

2.3 SIFT-MS

SIFT-MS has been described in detail previously [3] so details are not given here. Data were collected using the Mk2 instrument with a flow rate corresponding to a pressure of 0.008 Torr. After attaching the sample bag, full spectra of the count rates at each m/z value in the range m/z 10 to m/z 140 were recorded for all the samples using each of the three precursor ions, $\text{H}_3\text{O}^+$, $\text{NO}^+$ and $\text{O}_2^+$ using the full scan mode and a total sample time of 64 seconds for each precursor ion. The identities and concentrations of various components were determined using the on-line database containing reaction rate coefficients, developed from numerous detailed selected ion flow tube (SIFT) studies of various classes of compounds (alcohols, aldehydes, ketones, hydrocarbons, etc.) with the three precursor ions [3, 42]. In addition, data containing count values at each m/z value were collated and analysed (see below).

2.4 Univariate data analysis

SIFT-MS data were analysed firstly by examining individual ions present in each sample. The combination of spectra from the three precursor ions were used to identify compounds and their headspace concentrations where possible. Some were relatively easy to identify as their presence is well established; the identity of others was less certain. After looking at individual ions, the sample data were separated into the four groups described above and the mean values for each of the ions were calculated within each of the four groups. T-tests on log transformed data (concentrations, which are shown to be log normally distributed in such systems, [16, 18] or un-transformed data which are normally distributed (ion intensities) were carried out to determine which ions showed statistically significant differences between the two groups.

2.5 Multivariate analysis

Data for each precursor ion from SIFT-MS scans were normalised, and converted into spreadsheet format and imported into SIMCA-P+ (version 12, Umetrics, Umeå, Sweden) for multivariate analysis. Pareto scaling of data was applied. Principal component analysis (PCA) models were generated for all horses in the study for each carrier ion ($\text{H}_3\text{O}^+$, $\text{O}_2^+$ and $\text{NO}^+$). The first two components were fitted and scores plots generated. Significant outliers were identified using Hotelling’s $T^2$ range, observations outside the 95% critical limit were excluded from further analysis. Model fit was described by $R^2$, the proportion of variance in the data
explained by the model. Predictive ability of the model was explained by $Q^2$, the proportion of variation predicted by the model in a seven group cross-validation.

Associations between variables and class/group (colic cases vs. non-colic controls) were explored with an OPLS (orthogonal partial least squares) model in order to maximise separation of observations based on class and to measure non-correlated (orthogonal) variation. As for PCA models, Hotellings $T^2$ range was used on scores plots to evaluate outliers, the quality of model fit by $R^2$, and model predictive ability by $Q^2$. An S-plot of loadings was used to evaluate the magnitude (x-axis) and reliability (y-axis) of each variable (m/z value) in predicting class assignment. Jack-knifed 95% confidence intervals were added to a loadings bar chart to evaluate the significance of association between each variable and class assignment.

3. RESULTS & DISCUSSION

3.1 Compound and ion data

An investigation into the $\text{H}_3\text{O}^+$, $\text{NO}^+$ and $\text{O}_2^+$ spectra obtained from the samples enables the identification of the compounds given in Table 1. Many more ions are present, however it is difficult to be certain of all their identities due to likely overlaps of ions. The spectrum from the headspace of a horse faecal sample using the $\text{H}_3\text{O}^+$ precursor ions is shown in Fig. (1).

Table 1. Compounds clearly identifiable in the headspace of horse faecal samples from the horses with and without colic. The mean values, and concentration range in parts-per-billion-by volume, ppbv, are given.

<table>
<thead>
<tr>
<th>Compound (precursor ion)</th>
<th>Leahurst colic Mean, (range), ppbv</th>
<th>Leahurst controls Mean, (range), ppbv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaldehyde ($\text{H}_3\text{O}^+$)</td>
<td>3500, (2350 – 4700)</td>
<td>2600, (520 -7300)</td>
</tr>
<tr>
<td>Acetic acid ($\text{NO}^+$)</td>
<td>300, (0 - 670)</td>
<td>155 (0 – 350)</td>
</tr>
<tr>
<td>Acetone ($\text{NO}^+$)</td>
<td>930, (300, 1250)</td>
<td>390, (170 – 600)</td>
</tr>
<tr>
<td>Acetonitrile ($\text{H}_3\text{O}^+$)</td>
<td>55, (0 – 215)</td>
<td>34, (0 – 120)</td>
</tr>
<tr>
<td>Ammonia ($\text{H}_3\text{O}^+$)</td>
<td>2060, (175 – 4250)</td>
<td>1500 (215 – 3950)</td>
</tr>
<tr>
<td>Ethanol ($\text{H}_3\text{O}^+$)</td>
<td>2010, (540 – 8500)</td>
<td>2850, (830 – 8900)</td>
</tr>
<tr>
<td>Formaldehyde ($\text{H}_3\text{O}^+$)</td>
<td>0, (0 – 0)</td>
<td>30, (0, 140)</td>
</tr>
<tr>
<td>Terpenes (all isomers) ($\text{H}_3\text{O}^+$)</td>
<td>2710, (180 – 3500)</td>
<td>1900, (420 – 5800)</td>
</tr>
<tr>
<td>Methanol ($\text{H}_3\text{O}^+$)</td>
<td>1940, (430 – 4900)</td>
<td>780, (340 – 1730)</td>
</tr>
<tr>
<td>Methanethiol ($\text{H}_3\text{O}^+$)</td>
<td>23, (0 -62)</td>
<td>45, (9 -224)</td>
</tr>
<tr>
<td>Compound</td>
<td>M/z (ppb)</td>
<td>M/z (ppb)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Methyl phenol (H$_3$O$^+$)</td>
<td>27, (0 – 88)</td>
<td>23, (0 – 75)</td>
</tr>
<tr>
<td>(1 and 2)-Propanol (H$_3$O$^+$)</td>
<td>1875, (420 – 5750)</td>
<td>2750, (225 – 12000)</td>
</tr>
<tr>
<td>Toluene (H$_3$O$^+$)</td>
<td>550, (110 – 1900)</td>
<td>430, (140 -1450)</td>
</tr>
<tr>
<td>2-butanone (NO$^+$)</td>
<td>53, (0 – 180)</td>
<td>21, (0 - 95)</td>
</tr>
<tr>
<td>Carbon disulphide (O$_2$$^+$)</td>
<td>1025, (60 – 1740)</td>
<td>675, (100 – 1300)</td>
</tr>
<tr>
<td>Benzene (O$_2$$^+$)</td>
<td>194, (0 – 519)</td>
<td>116, (0 – 437)</td>
</tr>
</tbody>
</table>

*0 refers to below detectable levels (about 10 ppbv)

There is a statistically significant difference in the concentrations of acetone and methanol between the horses with colic and the control horses at the Leahurst hospital. Headspace concentrations of both compounds were higher in the animals with colic than the control animals with non-gastrointestinal, orthopaedic injuries.

When looking at the intensities at the various m/z values, the colic and control groups show that using the H$_3$O$^+$ precursor, m/z 87 is the only m/z where differences in ion counts are significant with p<0.05. There are a number of possibilities for the identity of m/z 87, however, it is most likely to be pentanal or a structural isomer such as methyl butanal, given the presence of an ion at m/z 85 on NO$^+$ which also corresponds to the 5 carbon aldehyde. However, ion intensities at m/z 85 using the precursor NO$^+$ is not significantly different between the groups. If looking at p values <0.1, m/z values with differences between the groups are m/z 47, which is the protonated ethanol ion and m/z 77 (protonated acetone plus 2H$_2$O). Using the NO$^+$ precursor, m/z 88 (acetone associated with NO$^+$) is significant at p<0.05.

There are only six colic horses and six controls and inter-individual variability is high, as it is in humans, e.g. in breath analysis studies [13]. However, there are still some significantly different compounds between the groups. That acetone is significantly different is interesting. Considering the colic horses and non-colic controls, it is clear that the colic horses have higher concentrations of acetone in their faecal headspace than the controls. In the case of breath in humans [9] high acetone is linked with a relatively low blood glucose concentration. Its significance in horse faeces is not yet known, but it may reflect changes in microbial gut populations [43]. The alcohols, methanol and ethanol also appear to be significantly different between the colic and non-colic horses, which is evidence for the gut metabolism and fermentation being altered between the two groups. This is not surprising in the case of horses with colic, and is consistent with the hypothesis that SCOD colic is associated with dysbiosis [44, 45]. The exact change in microbial populations affect the fermentation should be confirmed by microbial analysis of faecal samples; the change in the concentrations of the alcohols can only hint at this.

### 3.2 Multivariate statistics
Although there are some differences between the concentrations of compounds and ion intensities between the horse groups, they are not sufficient to fully distinguish between them on the basis of looking at individual ions and individual samples. Hence it is likely that the use of all the data using multivariate statistics is a more suitable way of using the information contained within the SIFT-MS spectra. For this reason, multivariate techniques were employed to investigate the SIFT-MS profiles for the two groups of samples.

Modelling all the observations using PCA revealed two outliers: LC1, which was discovered to be a sample of caecal contents rather than faeces, and LN4, a horse which had been administered antibiotics for seven days prior to sampling and which showed oral stereotypic behaviour that has been demonstrated to affect urine metabolic profile on NMR spectroscopy (Escalona, unpublished results). Both observations were excluded from further analysis.

Scores plots for PCA models for the remaining 10 observations with each of the three precursor ions are illustrated in Fig. (2a, c, and e). Scores plots for all three OPLS-DA models demonstrate good ability to separate observations according to class i.e. colic vs. non-colic (Fig. (2, b, d, and f)). Model diagnostics for each of the six models are presented in Table 2. These demonstrate that PCA models for all three carrier ions show moderate model fit but poor predictive ability; OPLS-DA models show similar levels of model fit but for H$_3$O$^+$ and O$_2^+$ carrier ions, predictive ability of the model is good (high Q$^2$(cum) value).

An S-plot of loadings for the H$_3$O$^+$ OPLS model (Fig. (3a)) identifies some product ions to be strongly associated with class; the significance of the association for each is explored with 95% confidence intervals in Fig. (3b). The following ions potentially demonstrate significant correlation with class at high magnitude: m/z 18, m/z 54, m/z 69, and m/z 77 and. Most likely identification of these ions is: m/z 18 and m/z 54 both correspond to ammonia and ammonia with hydrates due to the sample being moist, (NH$_4^+$ and NH$_4^+$.2H$_2$O), m/z 69: methanol (again, with 2 hydrates) and m/z 77: acetone (with 1 hydrate). Identification of acetone as a discriminator was confirmed using an O$_2^+$ carrier ion (detected at m/z 58) and an NO$^+$ carrier ion (detected at m/z 88).

Table 2. Multivariable model diagnostics for faecal headspace SIFT-MS data from horses with colic, and non-colic controls.

<table>
<thead>
<tr>
<th>Precursor ion</th>
<th>PCA</th>
<th>OPLS-DA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$Q^2$</td>
</tr>
<tr>
<td>H$_3$O$^+$</td>
<td>0.55</td>
<td>0.09</td>
</tr>
<tr>
<td>O$_2^+$</td>
<td>0.56</td>
<td>-0.15</td>
</tr>
<tr>
<td>NO$^+$</td>
<td>0.63</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
4. CONCLUSIONS

Data have been presented showing the identity and concentration range of a number of volatile compounds present in the headspace of horse faeces. Our analysis has identified volatile organic compounds which differentiate between metabolically healthy horses and those with simple colonic obstruction and distension colic, a serious and potentially fatal disease in horses. Multivariate modelling using OPLS-DA gave very clear differentiation between the SIFT-MS metabolic profiles of colic horses and the control population.

Hence both the compound data and the multivariate data provide evidence of a change in fermentation in the hindgut of horses with colic, and SIFT-MS is very well suited for this analysis by looking at individual compounds – their identity and concentration, but also by the use of multivariate statistics. The use of SIFT-MS using both analysis of spectra and associated univariate and multivariate analysis has indicated that there are significant differences in the concentrations of acetone, ammonia, methanol and methanol between the two groups; potential biomarkers for equine colic.

Because SIFT-MS data are relatively simple, being two dimensional consisting of m/z values and ion intensities, they are very well suited to analysis by multivariate statistics in which data from the whole spectrum may be used in obtaining profiles. This contrasts with GC-MS, which may be more sensitive and be better at compound identification, but data analysis is cumbersome and time consuming. Although these are significant differences in the concentrations of some individual compounds, this is unlikely to be a diagnostic test. The use of multivariate statistics is more likely to be suitable for assessing and monitoring colic.

ACKNOWLEDGEMENTS

Claire Turner carried out the sample and data analysis and was principal author of the manuscript. Claire Batty analysed the samples and assisted with data analysis. Ebona Escalona took the samples. John Hunter provided interpretation of data through understanding gut fermentation. Christopher Proudman took samples, provided interpretation of results through an understanding of horse physiology and analysed data. All authors contributed to the design of the study.

Ebony Escalona is a Horserace Betting Levy Board Research Training Scholar.

REFERENCES


6. Španěl, P., Smith, D., Quantification of trace levels of the potential cancer biomarkers formaldehyde, acetaldehyde and propanol in breath by SIFT-MS, *J. Breath Research*, **2008**, 2, 046003.


List of figures

Figure 1. SIFT-MS spectrum using the precursor \( \text{H}_3\text{O}^+ \) of headspace of faecal sample of a horse with colic.

Figure 2. Scores plots for PCA models (a, c, d) and OPLS-DA models (b, d, f) for SIFT-MS analysis of faecal headspace using the carrier ions \( \text{H}_3\text{O}^+ \) (a and b), \( \text{O}_2^+ \) (c and d), and \( \text{NO}^+ \) (e and f). Samples obtained from horses with colic (black triangles, “LC” label) and non-colic controls (red triangles, “LN” labels).

Figure 3. (a) S-plot illustrating correlation (y-axis) between each detected ions and class (colic vs. non-colic) and magnitude of association (x-axis) for OPLS-DA model with \( \text{H}_3\text{O}^+ \) carrier ion. (b) Loadings plot for the same model with 95% confidence intervals allowing interpretation of significance of observed associations; 95% confidence intervals not including 0 indicate significant positive or negative association between ion and class (colic vs. non-colic).
Fig. (1).
Fig. (2).
Fig. (3)