Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption

Amanda J. Lloyd¹, Manfred Beckmann¹, Gaëlle Favé², John C. Mathers² and John Draper¹*  
¹Institute of Biological Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, UK  
²Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

(Received 14 October 2010 – Revised 10 January 2011 – Accepted 8 February 2011 – First published online 9 May 2011)

Abstract

The lack of robust measures of dietary exposure hinders a quantitative understanding of causal relationships between diet and health. Non-targeted metabolite fingerprinting was used to explore the relationships between citrus exposure in free-living human subjects, estimated by a FFQ, and the chemical content of urine. Volunteers (study 1, n 12; study 2, n 11) were classified into high-, medium- and low-frequency citrus consumption groups. Overnight and spot fasting urine samples were obtained after exposure to a standardised citrus-free evening meal. The urine samples were analysed by flow injection electrospray-ionisation MS followed by supervised multivariate data classification analysis to discover discriminatory features associated with the level of citrus exposure. Good separation of high and low citrus consumption classes was achieved. Deeper exploration of high-ranked explanatory mass signals revealed several correlated signals derived from proline betaine. Targeted analysis of the relative levels of proline betaine in both fasting and overnight urine samples demonstrated good correlation with FFQ exposure data. Acute exposure of volunteers to orange juice resulted in the appearance of proline betaine and several biotransformed products in postprandial urine samples. Biomarker validation showed sensitivities of 80·8–92·2 % and specificities of 74·2–94·1 % (false discovery rate-adjusted P values < 0·05) for elevated proline betaine in participants who reported high citrus consumption. Proline betaine biotransformation products displayed weaker quantitative relationships with habitual citrus exposure. Targeted screening for the presence of biotransformation products of hesperidin and narirutin, known to be abundant in oranges, revealed that they were relatively poor indicators of citrus exposure.

Key words: Dietary exposure; Citrus fruits; Metabolomics; Urine; Proline betaine; Biomarkers

Research investigating links between intake of specific foods and health requires accurate assessment of dietary exposure(1,2). Conventional methods of measuring dietary exposure such as FFQ(2–5) depend upon estimates of food intake and are subject to well-recognised errors, derived largely from participant bias, which can confound interpretation of subsequent data(5,14). To address this problem, recent studies have described the targeted analysis, in blood and urine samples, of specific nutrients and metabolites derived from key foods that may have value as direct biomarkers of dietary exposure. In addition, quantification of biomarker concentrations in accessible biofluids can be used to help validate intake data obtained from FFQ and other conventional assessment methods(6–10), which is an important aspect of study design(5,11,12). To date, putative biochemical markers are available for only a relatively small number of specific foods and food components, and validation of food intake using conventional dietary assessment instruments in large cohorts of free-living participants remains a significant challenge(1,5). For example, a large number of studies have proposed that the antioxidant properties of dietary polyphenols from fruits and vegetables may protect consumers against several diseases(7,13). However, as a result of substantial metabolism after ingestion(14,15), it can be technically challenging to use the levels of specific secondary metabolites as an accurate estimate of dietary intake of purportedly beneficial foods. To address this issue, it has recently been proposed that the comprehensive chemical analysis of accessible human biofluids, using metabolomics methodology, may provide more suitable dietary intake biomarker leads(16–20). Methods utilising NMR(17,19) and particularly MS(16,18,19,21) are now implemented relatively routinely, and certain metabolite fingerprinting

Abbreviations: AUC, area under the receiver–operating characteristic curve; CRF, Clinical Research Facility; FIE–MS, flow injection electrospray-ionisation MS; FT–ICR–MS, Fourier-transform ion cyclotron resonance ultra mass spectroscopy; MEDE, Metabolomics to characterise Dietary Exposure; PC, principal component; RF, random forest.

* Corresponding author: J. Draper, fax +44 1970 621981, email jhd@aber.ac.uk
techniques are becoming high throughput, with the potential for automation\(^{(23)}\). Recently, we have developed volunteer handling, data normalisation and non-targeted metabolite fingerprinting methodology to aid measurement of metabolome changes in urine and other biofluids in response to acute dietary interventions\(^{(16,22)}\). Based on these studies that used flow injection electrospray-ionisation MS (FIE-MS)\(^{(21)}\) in conjunction with supervised multivariate data classification analysis\(^{(23)}\) and electrospray ionization-MS signal annotation tools\(^{(24)}\), we describe an approach to validate the use of FFQ dietary component descriptors without the need for prior knowledge of putative biochemical markers indicative of exposure to specific dietary components. As an example, we demonstrate that FFQ estimates of citrus exposure in small groups of free-living human subjects were correlated with distinct quantitative differences in the urine metabolome that are related to metabolites found in citrus fruits\(^{(7,14,25,26)}\).

The utility of these potential biomarkers, discovered by non-targeted fingerprinting, is compared with dietary exposure estimates made by the targeted analysis of urine samples for metabolites derived from the abundant polyphenols found in oranges\(^{(27)}\) and proposed previously as putative for metabolites derived from the targeted analysis of urine samples targeted fingerprinting, is compared with dietary exposure

### Experimental methods

#### Ethical approval and subject recruitment

The present study was approved by the Newcastle and North Tyneside 2 Research Ethics Committee (reference no. 07/H0907/136) and registered with the Newcastle upon Tyne Hospitals NHS Foundation Trust (registration no. 4392). The present study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the Newcastle and North Tyneside 2 Research Ethics Committee. Written informed consent was obtained from all subjects after a detailed explanation of the study protocol at an induction visit to the Clinical Research Facility (CRF) (Royal Victoria Infirmary, Newcastle upon Tyne, UK). The project constitutes part of the Metabolomics to characterise Dietary Exposure (MEDEx) research programme\(^{(146)}\), which aimed to develop a standardised protocol for nutritional metabolomics investigations\(^{(22)}\).

In the present study, study 1 participants were sampled during phase 2 of the MEDE project and study 2 participants were sampled during phase 3 of the MEDE project\(^{(160)}\). The volunteers were recruited through word of mouth and by advertisement in Newcastle University, UK. They were assessed for suitability via a screening questionnaire, which included the following exclusion criteria: aged under 18 years; for women being premenopausal; having a BMI < 18.5 kg/m\(^2\) or > 30 kg/m\(^2\); being a smoker, non-milk drinker and/or non-fish eater; having a history of substance abuse or alcoholism (alcohol consumption higher than 30 units/week); being allergic to any test food; suffering from any significant health problem and/or planning to change dietary or physical activity habits. Demographic data for each study participant are presented in Table S1 of the supplementary material (available online at http://www.journals.cambridge.org/bjn).

### Levels of habitual exposure to dietary citrus foods based on FFQ information

Habitual diet was characterised using the validated FFQ employed by the European Prospective Investigation into Cancer and Nutrition study\(^{(280)}\), which was modified slightly to include foods consumed frequently in the North East of England. The detailed study design and protocols will be published elsewhere\(^{(22)}\), and a detailed standard operating procedure is available on the NuGO website (http://www.nugo.org/sops/40878/41026). Volunteers were classified into three levels (low, medium and high) of habitual exposure to dietary citrus foods based on the analysis of FFQ information (Table 1) by combining exposure ratings for three specific food groups\(^{(280)}\) (see Table S2 of the supplementary material, available online at http://www.journals.cambridge.org/bjn) to provide estimates of total ‘citrus’ intake (Table 1). Individuals in the ‘low citrus’ exposure category consumed citrus food products < 2 weeks, those with ‘medium’ exposure levels consumed citrus foods almost every day, and two to three citrus portions/d were consumed by those with ‘high’ intakes.

### Sample collection and acute exposure study design

Study 1 volunteers (twelve individuals) attended two identical test days in the CRF, which were held several months apart. Volunteers were asked to collect all urine samples produced after consumption of a standardised evening meal, up to and including the morning void before attending the CRF, identified as the ‘PRE’ sample. On each study day, volunteers came to the CRF after a 12 h minimum fast, and ‘fasting’ urine samples were collected. Volunteers received a standardised test breakfast, and further urine samples were collected after 2, 4, 6 and 8 h. The test breakfast consisted of 200 ml orange juice, 190 ml tea with 14 ml skimmed milk and 12 g sugar, a 35 g butter croissant and 25 g cornflakes with 125 g semi-skimmed milk (1.7% fat) (see Favé \textit{et al.}\(^{(22)}\) for full details of all food items). A standardised light lunch, provided 4 h after the breakfast, consisted of two poached free-range eggs (approximately 2 \( \times \) 50 g), two slices of sliced white bread (2 \( \times \) 36 g) and 500 ml still mineral water (see Favé \textit{et al.}\(^{(22)}\) for full details of all food items). Study 2 volunteers (eleven individuals) attended six identical test days in the CRF, which were held at least 1 week apart, over the duration of a year. Only ‘PRE’ and ‘fasting’ samples were collected. Urine samples were frozen immediately at \(-20^\circ\text{C}\) and moved to \(-80^\circ\text{C}\) within 24 h\(^{(22)}\).
Metabolite fingerprinting and data analysis – feature selection

FIE-MS was carried out as described previously\textsuperscript{21,22}. Aliquots of thawed urine samples (50 μl) were diluted in 450 μl of pre-chilled methanol–water (3·5:1), vortexed, shaken for 15 min at 4°C and then centrifuged for 5 min at 14 000 g. For each urine sample, data were acquired in alternating positive and negative ionisation modes over four scan ranges (mass:charge ratio \((m/z)\) 15–110, 100–220, 210–510 and 500–1200), with an acquisition time of 5 min, on a linear trap quadrupole linear ion trap (Thermo Electron Corporation, San Jose, CA, USA). The resulting mass spectrum was the mean of twenty scans about the apex of the infusion profile. Raw data dimensionality was reduced by electronically extracting signals with ±0·1 Da mass accuracy, which resulted on average in one characteristic curve (AUC) and Welch’s \(t\) test, were used in feature selection, to produce a full feature rank list based on their statistical score values \textsuperscript{23}. RF feature selection was obtained by calculating importance scores, being the mean decrease in accuracy over all classes when a feature is omitted from data. The AUC used the area under curve of the sensitivity (true-positive rate) against the specificity (false-positive rate), and Welch’s \(t\) test ranked the features by their absolute values of the false discovery rate-corrected \(P\) values. Randomised re-sampling strategies using bootstrapping were applied in the process of classification and feature selection to counteract the effect of any unknown structured variance in the data. In the present data analysis, 100 bootstraps were used for classification and feature selection with RF using 1000 trees. Pearson’s correlation coefficients between selected variables were calculated using the \(R\)-function cor (correlation function). Variables with correlation coefficients \(\geq 0·7\) were considered to belong to a cluster indicative of different ionisation or potential biotransformation/breakdown products of a metabolite.

Targeted metabolite analysis – validation of features

Selected variables revealed by FIE-MS data mining were investigated further using targeted Nano-Flow (TriVersa NanoMate; Advion BioSciences Limited, Norfolk, UK) linear trap quadrupole-Fourier-transform ion cyclotron resonance mass spectrometry ultra (FT-ICR-MS; where ultra referred to

<table>
<thead>
<tr>
<th>Volunteer ID</th>
<th>Grapefruit</th>
<th>Oranges</th>
<th>Fruit juice</th>
<th>Total 'citrus'†</th>
<th>'Citrus' exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>213</td>
<td>0·000</td>
<td>0·066</td>
<td>0·066</td>
<td>0·132</td>
<td>Low</td>
</tr>
<tr>
<td>223</td>
<td>0·000</td>
<td>0·066</td>
<td>0·066</td>
<td>0·132</td>
<td>Low</td>
</tr>
<tr>
<td>218</td>
<td>0·000</td>
<td>0·140</td>
<td>0·000</td>
<td>0·140</td>
<td>Low</td>
</tr>
<tr>
<td>220</td>
<td>0·000</td>
<td>0·140</td>
<td>0·140</td>
<td>0·280</td>
<td>Low</td>
</tr>
<tr>
<td>217</td>
<td>0·430</td>
<td>0·140</td>
<td>0·140</td>
<td>0·710</td>
<td>Medium</td>
</tr>
<tr>
<td>221</td>
<td>0·000</td>
<td>0·066</td>
<td>0·780</td>
<td>0·846</td>
<td>Medium</td>
</tr>
<tr>
<td>215</td>
<td>0·000</td>
<td>0·066</td>
<td>1·000</td>
<td>1·066</td>
<td>Medium</td>
</tr>
<tr>
<td>214</td>
<td>0·066</td>
<td>0·430</td>
<td>1·000</td>
<td>1·496</td>
<td>High</td>
</tr>
<tr>
<td>216</td>
<td>0·066</td>
<td>0·430</td>
<td>1·000</td>
<td>1·496</td>
<td>High</td>
</tr>
<tr>
<td>204</td>
<td>0·066</td>
<td>1·000</td>
<td>1·000</td>
<td>2·066</td>
<td>High</td>
</tr>
<tr>
<td>219</td>
<td>0·000</td>
<td>0·066</td>
<td>2·500</td>
<td>2·566</td>
<td>High</td>
</tr>
<tr>
<td>222</td>
<td>0·140</td>
<td>2·500</td>
<td>0·066</td>
<td>2·706</td>
<td>High</td>
</tr>
</tbody>
</table>

ID, individuals.

* The scoring system is described in Table S1 of the supplementary material (available online at http://www.journals.cambridge.org/bjn).

† Sum of columns 2, 3 and 4 per volunteer.

Table 1. Frequency of exposure to dietary citrus of twelve free-living volunteers*
Table 2. Discrimination of habitual dietary citrus exposure by positive ionisation mode flow injection electrospray-ionisation (FIE) MS fingerprint analysis of overnight (PRE) and fasting urine samples in free-living volunteers

<table>
<thead>
<tr>
<th>Mass range (m/z)</th>
<th>PRE urine</th>
<th>Fasting urine</th>
</tr>
</thead>
<tbody>
<tr>
<td>15–110</td>
<td>1.26</td>
<td>1.27</td>
</tr>
<tr>
<td>100–220</td>
<td>1.73</td>
<td>2.24</td>
</tr>
<tr>
<td>210–510</td>
<td>1.20</td>
<td>1.10</td>
</tr>
<tr>
<td>500–1200</td>
<td>1.13</td>
<td>0.88</td>
</tr>
</tbody>
</table>

m/z, Mass:charge ratio; $T_w$, discriminant function 1 eigenvalue; AUC, area under the receiver-operating characteristic curve; Margin, random forest (RF) classification margin.

* Principal component-linear discriminant analysis (PC-LDA) and RF classification of data acquired by FIE-MS (positive ion mode) analysis of pre-test day overnight urine samples 'PRE' and 'fasting' urine samples, after a 12 h (minimum) fast, from twelve individuals. For PC-LDA, ‘dietary citrus exposure’ from Table 1 was the class structure applied consisting of ‘high’, ‘medium’ and ‘low’ citrus consumers (twelve volunteers, two repeat samples provided). Pairwise RF comparisons were made between five ‘high’ and four ‘low’ citrus consumers.

Discrimination of habitual dietary citrus exposure by positive ionisation mode flow injection electrospray-ionisation MS fingerprints (m/z 100–220) of the (a) pre-test day overnight ‘PRE’ urine and (b) ‘fasting’ urine samples, after a 12 h (minimum) fast, using volunteer ‘citrus consumption’ – high, medium and low – as the class structure (see Table 1). Eigenvalues ($T_w$) are given in brackets. DF1, discriminant function 1; DF2, discriminant function 1.

Results

Classification of habitual consumption of citrus foods using metabolite fingerprint analysis of urine

PC-linear discriminant analysis was used to determine how well ‘PRE’ and ‘fasting’ urine samples from study 1 volunteers were constructed using urine samples from four volunteers chosen at random and reconstituted in methanol–water (80:20, v/v). For each spray on the TriVersa Nanomate, a sample volume of 13.0 µl was used, and 2 µl of air were aspirated after the sample. Gas pressure was maintained between 0.2 and 0.6 psi, with the voltage at 1.4–1.7 kV (generally higher for negative ionisation mode) to achieve currents at 0.2 and 0.6 psi, with the voltage at 1.4–1.7 kV (generally rated after the sample. Gas pressure was maintained between (brackets. DF1, discriminant function 1; DF2, discriminant function 1.

Discrimination of habitual dietary citrus exposure level by metabolite fingerprinting of urine. Principal component-linear discriminant analysis of study 1 positive ion mode flow injection electrospray-ionisation MS fingerprints (m/z 100–220) of the (a) pre-test day overnight ‘PRE’ urine and (b) ‘fasting’ urine samples, after a 12 h minimum fast, using volunteer ‘citrus consumption’ – high, medium and low — as the class structure (see Table 1). Eigenvalues ($T_w$) are given in brackets. DF1, discriminant function 1; DF2, discriminant function 1.
in each of the habitual citrus exposure levels (high, medium and low) were discriminated using both positive and negative ionisation modes and four overlapping mass ranges (m/z 15–110, 100–220, 210–510 and 500–1200). In positive ionisation mode (Table 2), the mass range m/z 100–220 had the most classification power using both ‘PRE’ and ‘fasting’ urine samples. The same mass range was the most informative using negative ionisation mode data (see Table S3 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). Good discrimination of each citrus exposure class is evident in PC-linear discriminant analysis score plots comparing FIE-MS fingerprints of ‘PRE’ (Fig. 1(a)) and ‘fasting’ urine samples (Fig. 1(b)). The eigenvalue (T²w) for separation between high and low citrus consumers in the first discriminant function dimension (DF1) is > 2 for fasting urine FIE-MS fingerprint models (m/z 100–220), which indicates robust classification of the habitual consumption of citrus foods (Fig. 1(b)) in study 1 volunteers.

![Fig. 2.](https://doi.org/10.1017/S0007114511001164)

**Fig. 2.** Identification of signals explanatory of habitual dietary citrus exposure level following analysis of urine samples by positive ion mode flow injection electrospray-ionisation MS. Random forest (RF) importance scores of the top-ranked positive-mode discriminatory signals in a pairwise comparison between ‘high’ and ‘low’ citrus fruit consumption for study 1 (–Δ–) and study 2 (–×–) data using (a) pre-test day overnight ‘PRE’ urine and (b) ‘fasting’ urine samples, after a 12 h minimum fast; (c) top fifteen signals in both PRE and fasting urine samples in study 1 and study 2 discriminating ‘high’ and ‘low’ citrus consumers based on FFQ data. Black shading and white type indicate that the m/z signal is ranked in the top fifteen in three or all of the datasets; medium shading indicates that the m/z signal is ranked in the top fifteen in both urine samples for that particular study; (d) a clade of a hierarchical cluster analysis of the fifty top-ranked signals discriminating ‘high’ and ‘low’ citrus consumers following analysis of either PRE or fasting study 1 urine sample. Data shown are based on the correlation coefficient using the Pearson correlation method. *m/z 160 is ranked 17th in PRE data and m/z 198 is ranked 29th. Information relating to putative individuals is presented in Fig. 3 and Table 3. ID, individual.
Identification of urine metabolite fingerprint signals potentially explanatory of habitual citrus food consumption level

A total of five volunteers in study 1 were considered to be high-level habitual consumers of citrus foods, while four individuals were classified as reporting low-level exposure to citrus foods (Table 1). Analysis of FFQ data from study 2 volunteers (see Table S1 of the supplementary material, available online at http://www.journals.cambridge.org/bjn) revealed that a similar number of volunteers could be considered either high or low habitual consumers of citrus foods (six individuals were categorised as high consumers and five were low consumers). For each volunteer in study 1, two independent fasting and two independent overnight void (PRE) urine samples were available. Fasting and overnight void urine samples were collected from study 2 volunteers on six independent occasions spread throughout a 14-month period. In both studies, volunteers had consumed a freely chosen diet for several weeks before collecting urine samples, and thus for the purpose of the present analyses, each was considered an independent class (i.e. high or low citrus consumer) replicate. Metabolite fingerprints from ‘PRE’ and ‘fasting’ urine samples (representative of high and low habitual citrus exposure classes (study 1 samples, eighteen: ten high citrus and eight low citrus; study 2 samples, sixty-six: thirty-six high citrus and thirty low citrus) were subjected to pairwise comparison, using machine-learning techniques in which a combination of three methods (RF, AUC and Welch’s t test) were employed to rank features for discrimination power. To maximise predictability, re-sampling using the bootstrap method was applied. As a ‘rule of thumb’ we have shown, using a range of other FIE-MS datasets, that the threshold for significance in a pairwise analysis lies within an importance score range of 0.0015–0.00325. The curve inflection occurring at approximately 0.002 shows that the top fifteen to twenty of the m/z signals conferred the majority of discriminatory power in both PRE and fasting urine samples (Fig. 2(a) and (b)) in both studies. Fig. 2(c) shows that seven common signals (m/z 104, 166, 182, 183 and 201; shaded in black) were explanatory of high v. low citrus exposure levels in both PRE and fasting urine samples and for both studies (highly ranked in three or all four of the datasets) (expanded lists are shown in Table S4 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). Of these signals, m/z 104, 169 and 201 did not correlate with other highly ranked signals (data not shown). However, m/z 166, 167, 182 and 183 were strongly correlated with each other (Fig. 2(d)), together with m/z 144 and 145 (ranked in the top fifteen of the study 1 data and slightly lower ranked in the study 2 data), suggesting that these signals may be isotopes and/or salt adducts of ionised metabolites (Fig. 2(d)). In PRE urine sample, two further signals (m/z 160 and 198) formed part of the same correlation grouping.

These high-ranked nominal mass bins within this clade (Fig. 2(d)) were investigated in detail by ultra-high mass resolution FT-ICR-MS. Table 3 summarises the accurate mass FT-ICR-MS analysis of the correlated explanatory mass bins in both PRE and fasting urine samples. Querying the identity of the accurate mass signals in MZedDB24) suggested that these correlated explanatory signals were ionisation adducts and isotopes of proline betaine (stachydrine) and of hydroxyproline betaine (Table 3). A comparison of spectra derived from FIE-MS/MS fragmentation of m/z 144 (Fig. 3(a)), with an authentic sample of synthetic proline betaine [M+H]1+ (Fig. 3(b)) confirmed this annotation. In addition, the correlated explanatory signals, proposed to be salt adduct and isotopes of proline betaine (Table 3), were also confirmed by FIE-MS/MS fragmentation with standards (data not shown). FIE-MS/MS spectra of the nominal mass bin containing predicted hydroxyproline betaine [M+H]1+ (m/z 160) from a single fasting individual (Fig. 3(c)) substantially matched that of the FIE-MS/MS spectra of an authentic sample of 4-hydroxyproline betaine [M+H]1+ (m/z 160). However, the presence of fragment ions at m/z 60, 102 and 116 in the spectra derived from FIE-MS/MS analysis of this particular individual’s urine suggested the presence of more than one chemical in this nominal mass bin. The FIE-MS/MS spectra of m/z 160 from a ‘pool’ derived from urine collected from four random fasting volunteers (Fig. 3(c)) showed an enhancement of fragment ions m/z 60 and 116. In addition, the correlated signal proposed to be a K+ adduct of hydroxyproline betaine (Table 3) was also confirmed by FIE-MS/MS fragmentation (data not shown). The structures of proline betaine and 4- and 3-hydroxyproline betaine are shown in Fig. 4(a).

Table 3. Identity of highly ranked and correlated signals potentially explanatory of habitual dietary citrus exposure using Fourier-transform ion cyclotron resonance mass spectroscopy (FT-ICR-MS)

<table>
<thead>
<tr>
<th>Nominal mass</th>
<th>Accurate mass in FT-ICR-MS</th>
<th>Putative identification</th>
<th>Ionisation product</th>
</tr>
</thead>
<tbody>
<tr>
<td>144</td>
<td>144-10 187</td>
<td>Proline betaine</td>
<td>[M+H]1+</td>
</tr>
<tr>
<td>145</td>
<td>145-10 518</td>
<td>Proline betaine</td>
<td>[M+H]1+</td>
</tr>
<tr>
<td>166</td>
<td>166-08 377</td>
<td>Proline betaine</td>
<td>[M+Na]1+</td>
</tr>
<tr>
<td>167</td>
<td>167-08 713</td>
<td>Proline betaine</td>
<td>[M+Na]1+</td>
</tr>
<tr>
<td>182</td>
<td>182-05 772</td>
<td>Pro line betaine</td>
<td>[M+K]1+</td>
</tr>
<tr>
<td>183</td>
<td>183-06 109</td>
<td>Proline betaine</td>
<td>[M+K]1+</td>
</tr>
<tr>
<td>160</td>
<td>160-09 674</td>
<td>4-Hydroxyproline betaine</td>
<td>[M+H]2+</td>
</tr>
<tr>
<td>198</td>
<td>198-05 267</td>
<td>4-Hydroxyproline betaine</td>
<td>[M+K]2+</td>
</tr>
</tbody>
</table>
Acute exposure to a test breakfast containing orange juice demonstrates proline betaine, hesperidin and narirutin biotransformation and excretion

The possibility of biotransformation and excretion of proline betaine was examined in spot urine samples collected 2 and 8 h after consumption of a standard breakfast including 200 ml orange juice in study 1 volunteers. Signals discriminat- ing fasting urine samples from either a 2 or 8 h postprandial urine sample in both positive- and negative-ion mode FIE-MS data are shown in Table 4. Comparison of signals at both 2 and 8 h after consumption of orange juice allowed an assessment of the potential contribution of metabolite signals derived from the colonic fermentation of ingested food residues (present in 8 h but not in 2 h samples). Explanatory signals common to both urine sampling times in positive-ion data (italicised in Table 4) corresponded with those derived from proline betaine and hydroxyproline betaine. The two explanatory mass bins, m/z 223 and m/z 319, in negative-ion

Fig. 3. Confirmation of signals explanatory of habitual dietary citrus exposure level following: analysis of urine samples by flow injection electrospray-ionisation MS (FIE-MS/MS). FIE-MS/MS spectra of the nominal mass bin contain the following: (a) FIE-MS/MS pooled m/z 144, putative proline betaine [M+H]⁺ (m/z 144) from a pool of four randomised fasting volunteers; (b) proline betaine [M+H]⁺ standard, the FIE-MS/MS of an authentic sample of synthetic proline betaine [M+H]⁺; (c) FIE-MS/MS individual m/z 160, putative hydroxyproline betaine [M+H]⁺ (m/z 160) from a single individual; (d) 4-hydroxyproline betaine [M + H]⁺, the FIE-MS/MS of an authentic sample of synthetic 4-hydroxy proline betaine [M+H]⁺; (e) FIE-MS/MS pooled m/z 160, putative hydroxyproline betaine [M+H]⁺ (m/z 160) from a pool of four randomised fasting volunteers.
Proline betaine and habitual citrus exposure

Table 4. Top twenty positive- and negative-ion features (m/z 100–550) discriminating between fasting and either a 2 or 8 h postprandial urine sample after exposure to a standard breakfast containing orange juice

<table>
<thead>
<tr>
<th>Rank*</th>
<th>Positive-ion signals</th>
<th>Negative-ion signals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fasting v. 2 h</td>
<td>Fasting v. 8 h</td>
</tr>
<tr>
<td>1</td>
<td>166</td>
<td>160</td>
</tr>
<tr>
<td>2</td>
<td>160</td>
<td>301</td>
</tr>
<tr>
<td>3</td>
<td>167</td>
<td>166</td>
</tr>
<tr>
<td>4</td>
<td>391</td>
<td>167</td>
</tr>
<tr>
<td>5</td>
<td>184</td>
<td>171</td>
</tr>
<tr>
<td>6</td>
<td>407</td>
<td>134</td>
</tr>
<tr>
<td>7</td>
<td>145</td>
<td>198</td>
</tr>
<tr>
<td>8</td>
<td>144</td>
<td>170</td>
</tr>
<tr>
<td>9</td>
<td>312</td>
<td>144</td>
</tr>
<tr>
<td>10</td>
<td>381</td>
<td>145</td>
</tr>
<tr>
<td>11</td>
<td>198</td>
<td>149</td>
</tr>
<tr>
<td>12</td>
<td>182</td>
<td>279</td>
</tr>
<tr>
<td>13</td>
<td>183</td>
<td>102</td>
</tr>
<tr>
<td>14</td>
<td>279</td>
<td>273</td>
</tr>
<tr>
<td>15</td>
<td>270</td>
<td>208</td>
</tr>
<tr>
<td>16</td>
<td>365</td>
<td>165</td>
</tr>
<tr>
<td>17</td>
<td>115</td>
<td>109</td>
</tr>
<tr>
<td>18</td>
<td>163</td>
<td>254</td>
</tr>
<tr>
<td>19</td>
<td>282</td>
<td>263</td>
</tr>
<tr>
<td>20</td>
<td>257</td>
<td>478</td>
</tr>
</tbody>
</table>

* Rank in a random forest classification of fasting v. postprandial urine samples.

Downloaded from https://www.cambridge.org/core. IP address: 54.70.40.11, on 10 Sep 2018 at 23:42:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . doi:10.1017/S0007114511001164

Proline betaine and its biotransformation products as potential biomarkers of habitual, in addition to acute, citrus exposure

Within 2 h of acute exposure to orange juice (from the standard breakfast), proline betaine and hydroxylated derivatives were present in urine and persisted in detectable concentrations for at least 8 h. In addition, in both ‘fasting’ and ‘PRE’ urine samples, positive-ion signals derived from both...
of these chemicals were strongly explanatory of citrus intakes estimated by FFQ (Table 3). The targeted analysis of proline betaine m/z signals in urine samples from individual volunteers representing low, medium and high habitual citrus exposure classes demonstrated a potential quantitative relationship between exposure level and signal intensity (Fig. 7). The false discovery rate-adjusted P values indicate a significant difference between the high and low consumers for all three proline betaine m/z signals (m/z 144, 166 and 182) in both ‘PRE’ and ‘fasting’ urine samples (P < 0.05; see Table S6 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). However, it was not possible to distinguish between medium and either high- or low-level consumers (false discovery rate-adjusted P values > 0.05; see Table S6 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). The elevation of excretion of these three proline betaine m/z signals in the high citrus consumers compared with low citrus consumers in PRE urine samples showed a sensitivity of 84.7–92.2% and a specificity of 74.2–94.1%, depending on the adduct (see Table S7 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). In the fasting urine samples, the elevation of excretion of these three proline betaine m/z signals showed a sensitivity of 80.8–89.2% and a specificity of 79.6–89.0% (see Table S7 of the supplementary material, available online at http://www.journals.cambridge.org/bjn). In addition, negative-ion signals associated with sulphonated or glucuronidated proline betaine biotransformation products were also present at low levels in fasting and PRE urine samples (data not shown).

Discussion

In the present study, we used a non-targeted metabolomics approach to discover and structurally identify urinary biochemical markers of citrus exposure, in a small group of volunteers. Subsequently, we confirmed this observation in a second group of volunteers of similar size, but who provided a larger number of replicate urine samples collected at intervals over a 14-month period. We observed that proline betaine, an abundant component of citrus fruits, was strongly explanatory of both acute and habitual exposure to citrus-containing foods. Previous reports have described the rapid excretion of proline betaine in urine following acute exposure to either the pure chemical or orange juice. A recent investigation of proline betaine excretion...
using NMR analysis of postprandial urine samples suggested that it was cleared from the body rapidly and could not be detected easily 14 h after consuming orange juice (31). Subsequent validation of urinary proline betaine as a potential biomarker of citrus consumption was undertaken using samples and data from the International Study of Macro- and Micro-Nutrients and Blood Pressure study, in which participants were dichotomised into citrus consumers and citrus non-consumers on the basis of two consecutive multipass 24 h dietary recalls repeated after 3 weeks and analysis of two 24 h urine sample collections made concurrently (31). The previous report concluded that proline betaine was an effective biomarker of citrus exposure, where 24 h dietary recall data indicated that citrus products had been consumed within the previous 24 h. The rapid clearance kinetics of proline betaine reported by Heinzmann et al. (31) might seem to limit the utility of this metabolite as a biomarker of citrus consumption. However, as well as confirming that proline betaine can be detected by MS in 2 and 8 h postprandial urine samples after acute exposure to orange juice, we demonstrated that this metabolite is present at elevated levels in overnight void (‘0’) and ‘fasting’ urine samples in individuals reporting habitual high intake of citrus foods. Furthermore, the present study shows that the quantitative relationships between habitual citrus intake, estimated by FFQ and the levels of proline betaine in fasting urine samples, are not dependent on the knowledge of citrus fruit consumption on the day of urine collection nor is it compromised by unreported factors associated with the timing of citrus intake before urine sampling.

Unlike previous reports that used only positive ionisation mode liquid chromatography–MS procedures only (33) or NMR fingerprinting (32) to detect proline betaine, we have demonstrated that biotransformed proline betaine derivatives are detectable in urine samples and are explanatory of habitual citrus exposure levels. Betonicine (4-hydroxyproline betaine) is a component of citrus fruits, present at a lower concentration than proline betaine (33), and thus its appearance in postprandial urine samples following exposure to orange juice (m/z signals 160 and 198 in positive-ion mode data; Table 4) is unsurprising. Analysis of biotransformation products in rat urine samples has suggested hydroxylation of proline betaine at carbon 3 (34). The present FIE-MS/MS analysis of m/z 160 revealed three additional fragment ions (m/z 60, 116 and 102). Of the three fragment ions, two (m/z 60 and 102) matched abundant signals in the previously reported fragmentation

Table 5. Confirmation of sulphonated and glucuronated negative-ion mode derivatives of proline betaine following analysis of urine samples by flow injection electrospray-ionisation MS

<table>
<thead>
<tr>
<th>MS2 ions</th>
<th>MS2 fragment identification</th>
<th>MS2 ions</th>
<th>MS2 fragment identification</th>
<th>MS2 ions</th>
<th>MS2 fragment identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z 223</td>
<td></td>
<td>m/z 319</td>
<td></td>
<td>m/z 143*</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>[M – CH₂N⁺(CH₃)₂]</td>
<td>–</td>
<td>175</td>
<td>Glucuronide – H</td>
<td>–</td>
</tr>
<tr>
<td>143†</td>
<td>[M – SO₃H]</td>
<td>143†</td>
<td>[M – Glucuronide]</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>141</td>
<td>[M – SO₂H₂]</td>
<td>113</td>
<td>Glucuronide – H – CO₂ – H₂O</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>115</td>
<td>[M – SO₃₋₂₋₀] – CO</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>113</td>
<td>[M – SO₃₋₂₋₀ – CO] or [M – SO₃₋₂₋₀ – (CH₃)₂]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>97</td>
<td>[M – SO₃₋₂₋₀ – HO]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>81</td>
<td>[M – SO₃₋₂₋₀ – CH₃ – H₂O]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

* Proline betaine standard with abundant signals at m/z 125, 115, 113 and 99.
† MS2 analysis of the MS2 product ion m/z 143 produced the same fragmentation as the proline betaine standard.

**Proline betaine and habitual citrus exposure**

![Fig. 6. Orange juice flavonoid aglycone conjugate signals in postprandial urine samples following exposure to orange juice (m/z signals 160 and 198 in positive-ion mode data; Table 4) is unsurprising. Analysis of biotransformation products in rat urine samples has suggested hydroxylation of proline betaine at carbon 3 (34). The present FIE-MS/MS analysis of m/z 160 revealed three additional fragment ions (m/z 60, 116 and 102). Of the three fragment ions, two (m/z 60 and 102) matched abundant signals in the previously reported fragmentation.**
spectra of an authentic sample of synthetic 3-hydroxyproline betaine [M+H]+, but the origin of the fragment ion at m/z 116 is currently unknown. Therefore, our data (Fig. 3(e)) suggest that 3-hydroxyproline betaine is also present in human urine, which is probably derived by biotransformation of proline betaine. Additionally, we demonstrate, for the first time (using negative ionisation mode FIE-MS fingerprinting), that proline betaine is also conjugated in human subjects to form sulphate and monoglucuronide derivatives. This relatively complex phase II metabolism was not identified in a previous study of citrus fruit metabolism using NMR probably because of the relative insensitivity of NMR-based methodology. The present study shows that proline betaine conjugates are present in urine within 2–8 h after consuming orange juice, and that signal intensities of these derivatives are substantially lower than those reflecting the presence of non-modified proline betaine. Further work is required to describe quantitatively the kinetics of appearance and excretion of sulphate and glucuronide derivatives of proline betaine to help evaluate the possible utility of these biotransformation products as putative biomarkers of citrus exposure. In contrast to proline betaine, citrus food flavonone biomarkers would be subject to considerable diurnal variation dependent on the timing of major phases of colonic fermentation activity required for their absorption. This is in agreement with a recent review, which also

Fig. 7. Box plots of the top positive-ion mode explanatory metabolite signals between ‘high’, ‘medium’ and ‘low’ habitual citrus fruit consumers. (a) m/z 144, fasting urine; (b) m/z 144, PRE urine; (c) m/z 166, fasting urine; (d) m/z 166, PRE urine; (e) m/z 182, fasting urine; (f) m/z 182, PRE urine; n 12 volunteers; between ‘high’ and ‘low’ citrus consumers, false discovery rate (FDR)-corrected P values < 0.05; between ‘high’ or ‘low’ and ‘medium’ consumers, FDR-corrected P values > 0.05 (see Table S6 of the supplementary material, available online at http://www.journals.cambridge.org/bjn, for complete test results). The box indicates the interquartile range; the red horizontal bar indicates the median; vertical bars indicate the maximum and minimum values up to 1.5 x interquartile range; error bars represent the standard error of twelve volunteers. ‘PRE’, pre-test day overnight urine voids; ‘Fasting’, spot urine sample after a 12 h (minimum) fast. TIC, total ion count.
concludes that the biotransformation products of the flavono-
none glycosides hesperidin and narirutin are unlikely to be
suitable biomarkers of habitual exposure to citrus.

In the present study, we provided a fruit-free standardised
meal(16,22) on the evening before collection of overnight
(PRE) and fasting urine samples and were able to distinguish
between low, medium and high habitual intakes of citrus
foods (estimated by FFQ) based on urinary proline betaine
measurements, despite studying samples from only a relatively
small number of volunteers. It is observed that even though
six replicate samples were available for each volunteer in
study 2, the inclusion of extra replicates did not improve
significantly the classification robustness over that achieved
in study 1 volunteers. In support of the potential of this
metabolite as a biomarker of citrus intake, we have observed
strong links between orange juice consumption in a standard
breakfast and urinary proline betaine excretion 2–8 h later.
In addition, we have provided preliminary evidence that proline
betaine may be metabolised in human subjects to a number of
derivatives including sulphates and glucuronides – an obser-
vation contrary to the assumption that proline betaine is
metabolically inert(31). A potential quantitative relationship
between high and low dietary citrus consumption and urinary
excretion of proline betaine signals (m/z 144, 166 and 182) in
positive-ion data was demonstrated (adjusted P values <0·05;
see Table S6 of the supplementary material, available online at
http://www.journals.cambridge.org/bjn). However, differences
in intensity levels of these adducts between medium
citrus consumers and either low or high consumers were not
statistically significant (P>0·05). This may be because individ-
uals who consume either a high amount (at least once a day)
or very low amounts of a particular food generally are able
more accurately estimate their consumption (using a FFQ
reporting system) than individuals who consume these foods
at ‘medium’ levels. In addition, of course, it is technically
easier to detect the larger differences between ‘high’ and
‘low’ intakes. Relatively high sensitivities and specificities
(80·8–92·2 and 74·2–94·1%, respectively, see Table S7 of
the supplementary material, available online at http://www.
journals.cambridge.org/bjn) for the three elevated proline
betaine adducts (H+, Na+ and K+) were demonstrated in partici-
pants who reported high citrus consumption (in both pooled
overnight urine and spot fasting urine samples), thus
further validating the potential biomarker status of this metab-
olite. This potential quantitative relationship between high
and low dietary citrus consumption and urinary excretion of
proline betaine could be further explored using standard
Triple Quad technology(36,37).

In conclusion, the present metabolomics-based study pro-
vides prima facie evidence that urinary excretion of proline
betaine (and possibly some of its metabolites) is a potentially
useful biomarker of habitual citrus consumption. However, we
have not attempted to test the utility of this metabolite as a
biomarker for different types of citrus fruit nor have we
examined, extensively, dose– and time–response relation-
ships between citrus food consumption and patterns of
urinary excretion, which would be necessary to establish the
sensitivity and specificity of our proposed biomarker. These
areas require further investigation.

Acknowledgements
The present study was supported by the UK Food Standards
Agency (project N05073). The authors thank the volunteers
for their commitment, the CRF, Newcastle, for nursing support;
Claire Kent, Heather E. Gifford, Julie Coaker and Linda Penn
for their practical support; and Marks & Spencer for donating
the chocolate éclairs (standardised evening meal). We thank
Kathleen Tailliat for FIE-MS analysis and for technical sup-
port, and Dr Wanchang Lin (Manchester School of Biomedici-
é) for support with the statistical analysis of data. The
authors’ contributions to the study were as follows: A. J. L.
conducted the data analysis, produced figures, researched
the literature and wrote the manuscript; M. B. developed the
urine extraction procedures, designed the metabolite finger-
printing experiments, supervised MS support staff, pre-
processed data for analysis and edited the manuscript; G. F.
undertook the volunteer recruitment, coordinated the volun-
teer CRF visits and supervised CRF support staff, refined
sampling methodology and edited the manuscript; J. C. M.
coordinated the study, supervised the study in Newcastle
University, designed the volunteer handling protocols and
edited the manuscript; J. D. coordinated the study, supervised
the study in Aberystwyth, designed the figures and wrote the
manuscript. None of the authors has a conflict of interest with
respect to the study.

References
in nutritional epidemiology: applications, needs and new
the food frequency questionnaire? Cancer Epidemiol
Biomarkers Prev 14, 2826–2828.
the food frequency questionnaire: point. Cancer Epidemiol
Biomarkers Prev 15, 1757–1758.
Public Health Nutr 5, 821–827.
dietary assessment methods in nutritional epidemiology:
weighed records v. 24-h recalls, food-frequency question-
6. Holmes MD, Powell IJ, Campos H, et al. (2007) Validation of
a food frequency questionnaire measurement of selected
nutrients using biological markers in African-American men.
and phenolic acids as biomarkers of intake for polyphenol-rich
validation of a self-administered semiquantitative food-
Biochemical validation of food frequency questionnaire-
estimated carotenoids, alpha-tocopherol, and folate intakes
among African Americans and non-Hispanic Whites in the


