

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 Draper^{1*}

¹*Institute of Biological Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, UK*

²*Human 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⁾. Conventional methods of measuring dietary exposure such as FFQ^(2,3) 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^(4,5). 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⁽¹³⁾. 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^(17,20) 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⁽²¹⁾. 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 biomarkers for citrus exposure^(7,13,14,25,26). Additionally, to develop metabolomic procedures suitable for future epidemiological studies, we demonstrate the value of overnight and, particularly, fasting urine samples for the discovery of potential chemical biomarkers indicative of habitual 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 (MEDE) research programme⁽¹⁶⁾, 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⁽¹⁶⁾. 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/m² or >30 kg/m²; 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⁽²⁸⁾, 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⁽²⁸⁾ (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 week, 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 *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 × 50 g), two slices of sliced white bread (2 × 36 g) and 500 ml still mineral water (see Favé *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°C and moved to -80°C within 24 h⁽²²⁾.

Table 1. Frequency of exposure to dietary citrus of twelve free-living volunteers*

Volunteer ID	FFQ score				'Citrus' exposure
	Grapefruit	Oranges	Fruit juice	Total 'citrus'†	
213	0.000	0.066	0.066	0.132	Low
223	0.000	0.066	0.066	0.132	Low
218	0.000	0.140	0.000	0.140	Low
220	0.000	0.140	0.140	0.280	Low
217	0.430	0.140	0.140	0.710	Medium
221	0.000	0.066	0.780	0.846	Medium
215	0.000	0.066	1.000	1.066	Medium
214	0.066	0.430	1.000	1.496	High
216	0.066	0.430	1.000	1.496	High
204	0.066	1.000	1.000	2.066	High
219	0.000	0.066	2.500	2.566	High
222	0.140	2.500	0.066	2.706	High

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.

Metabolite fingerprinting and data analysis – feature selection

FIE-MS was carried out as described previously^(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 signal per mass bin (in the following referred to as a 'nominal mass'). Data were log transformed and normalised to total ion current before data analysis⁽²¹⁾.

Comprehensive data mining was carried out following the FIEmspro workflow validated previously in Aberystwyth⁽²³⁾ and accessible at the uniform resource locator <http://users.aber.ac.uk/jhd/>. Principal component (PC) analysis transformed the sample data matrix into a coordinate system, where each new projection (PC) is a linear combination of the original variables, thus reducing data dimensionality. This was followed by supervised PC-linear discriminant analysis in which plots of the first two discriminant functions allowed visualisation of the goodness of class separation. Eigenvalues (T_w), defined as the ratio of the between- and within-group standard deviations of the discriminant variables, were used to evaluate the performance of PC-linear discriminant analysis; discrimination was considered good for eigenvalues of >2.0 and poor for eigenvalues <1.0 ⁽²³⁾. Random forest (RF), which is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees, was also employed in the analysis of the multivariate data. In RF, each tree is trained on a bootstrap sample of the training data, and

predictions are made by a majority vote of the trees. The best split is based on the variables in the training set instead of all variables. To assess the classification performance, the RF classification margin, defined as the proportion of votes for the correct class minus the maximum proportion of votes for the other classes, was used. Average margin values ≥ 0.3 indicate adequate classification in metabolomic experiments⁽²³⁾.

Feature selection techniques were used to select the nominal mass signals (m/z), which were responsible for discriminating between different sample classes. A combination of three methods, RF, area under the receiver-operating 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⁽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 >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 spectroscopy ultra (FT-ICR-MS; where ultra referred to

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*

Mass range (<i>m/z</i>)	Discrimination of habitual dietary citrus exposure level					
	PRE urine			Fasting urine		
	T_w value	AUC	Margin	T_w value	AUC	Margin
15–110	1.26	0.78	0.13	1.27	0.72	0.14
100–220	1.73	0.94	0.26	2.24	0.90	0.28
210–510	1.20	0.77	0.13	1.10	0.70	0.08
500–1200	1.13	0.55	0.01	0.88	0.67	0.07

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 12h (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.

the high-sensitivity ICR-cell). For each biological class, pools 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 μ l was used, and 2 μ l 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 80–120 and –100 to –60 nA in positive and negative ionisation mode, respectively. Operating the FT-ICR-MS in narrow mode, a resolution of 100 000 was chosen, and mass range was scanned for 1 min. A minimum of three samples per class or treatment containing the specific selected mass were required for successful accurate mass verification. The system was calibrated with the linear trap quadrupole Fourier-transform calibration solution prepared according to the instrument manufacturer's instructions.

For metabolite signal identification, the accurate mass values were then queried against MZedDB, an interactive accurate mass annotation tool that can be used to provide tentative annotation of signals by means of neutral loss and/or adduct formation rules⁽²⁴⁾. Further metabolite signal identification

was obtained using FIE-MS/MS in which the scan window was set for twenty scans, with an isolation width of *m/z* 1 and normalised collision energy of 40 V. The activation coefficient *Q* of 0.250 was chosen and an activation time of 30 ms, with wideband activation turned on and source fragmentation of 20 V. Mass range settings were dependent upon the molecular weight of the target ion. Chemical standards investigated with FT-ICR-MS and FIE-MS/MS were obtained commercially at highest purity, and solvents were of HPLC-MS grade. Standards were prepared by dissolving 1 mg of the metabolite standard in 1 ml extraction solvent and reconstituted in methanol–water (80:20, v/v). To confirm the identity of salt adducts, additional aqueous Na⁺ or K⁺ ion-containing solutions (50 mM as bicarbonate in the case of Na⁺ and chloride in the case of K⁺) were added to aid adduct formation.

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

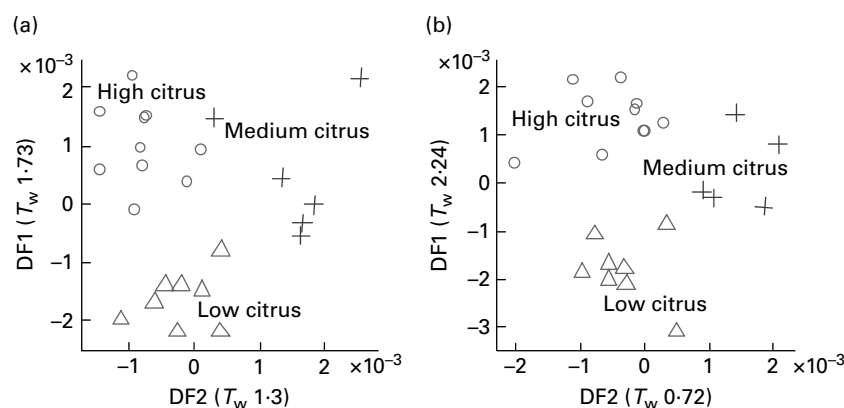


Fig. 1. Discrimination of habitual dietary citrus exposure level by metabolite fingerprinting of urine. Principal component-linear discriminant analysis of study 1 positive ion mode flow injection electrospray-ionisation MS fingerprints (*m/z* 100–220) of the (a) pre-test day overnight 'PRE' urine and (b) 'fasting' urine samples, after a 12h 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 (T_w) 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.

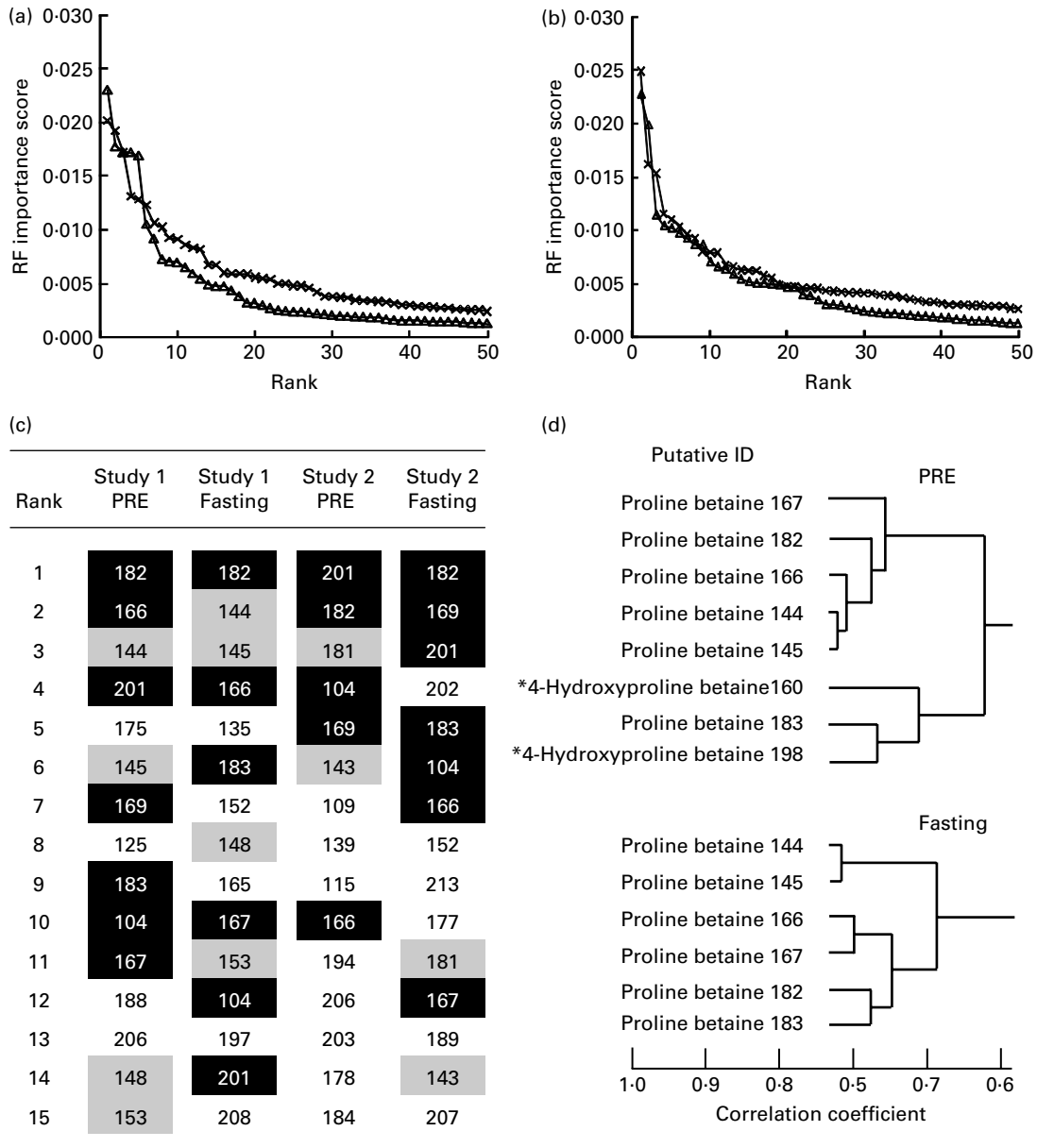


Fig. 2. Identification of signals explanatory of habitual dietary citrus exposure level following analysis of urine samples by positive ion mode flow injection electrospray-ionisation MS. Random forest (RF) importance scores of the top-ranked positive-mode discriminatory signals in a pairwise comparison between 'high' and 'low' citrus fruit consumption for study 1 (—△—) 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 classed 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 0.0015–0.003⁽²³⁾. 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, 167, 169, 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 MZedDB⁽²⁴⁾ 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 $[M+H]^{1+}$ (Fig. 3(d)). 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(e)) 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)

Nominal mass	Accurate mass in FT-ICR-MS	Putative identification	Ionisation product
144	144.10187	Proline betaine	$[M+H]^{1+}$
145	145.10518	Proline betaine	^{13}C isotope of $[M+H]^{1+}$
166	166.08377	Proline betaine	$[M+Na]^{1+}$
167	167.08713	Proline betaine	^{13}C isotope of $[M+Na]^{1+}$
182	182.05772	Pro betaine	$[M+K]^{1+}$
183	183.06109	Proline betaine	^{13}C isotope of $[M+K]^{1+}$
160	160.09674	4-Hydroxyproline betaine	$[M+H]^{1+}$
198	198.05267	4-Hydroxyproline betaine	$[M+K]^{1+}$

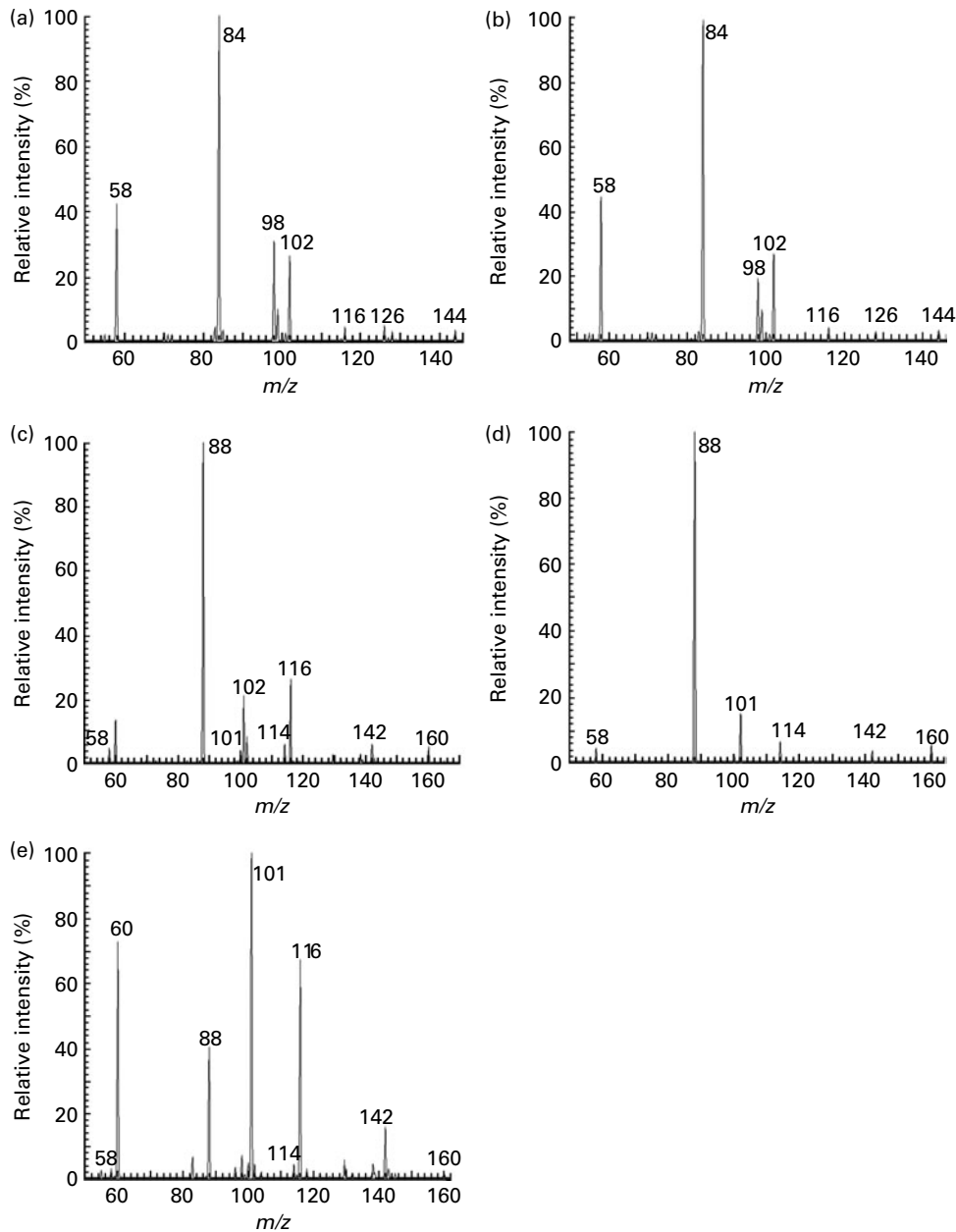


Fig. 3. Confirmation of signals explanatory of habitual dietary citrus exposure level following: analysis of urine samples by flow injection electrospray-ionisation MS (FIE-MS/MS). FIE-MS/MS spectra of the nominal mass bin contain the following: (a) FIE-MS/MS pooled m/z 144, putative proline betaine $[M+H]^{1+}$ (m/z 144) from a pool of four randomised fasting volunteers; (b) proline betaine $[M+H]^{1+}$ standard, the FIE-MS/MS of an authentic sample of synthetic proline betaine $[M+H]^{1+}$; (c) FIE-MS/MS individual m/z 160, putative hydroxyproline betaine $[M+H]^{1+}$ (m/z 160) from a single individual; (d) 4-hydroxyproline betaine $[M+H]^{1+}$, the FIE-MS/MS of an authentic sample of synthetic 4-hydroxyproline betaine $[M+H]^{1+}$; (e) FIE-MS/MS pooled m/z 160, putative hydroxyproline betaine $[M+H]^{1+}$ (m/z 160) from a pool of four randomised fasting volunteers.

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 discriminating 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

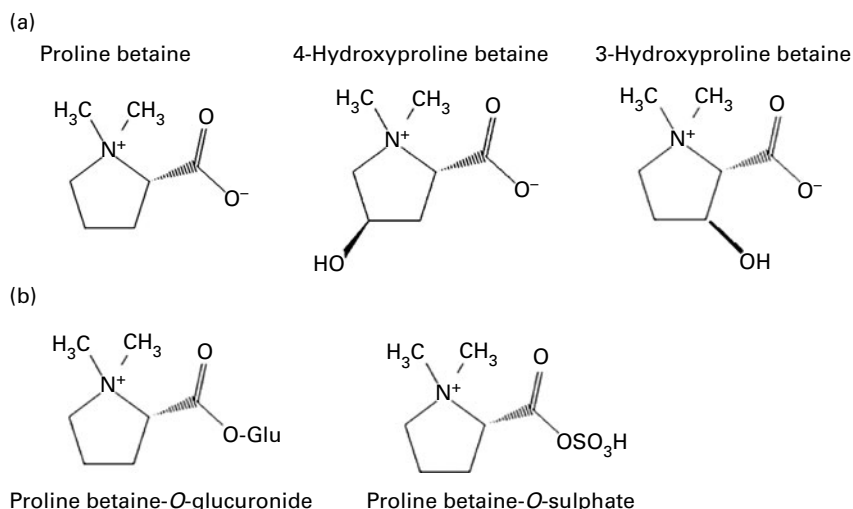


Fig. 4. Proposed structure of proline betaine and its biotransformation products in human urine samples identified with Fourier-transform ion cyclotron resonance mass spectroscopy ultra and flow injection electrospray-ionisation tandem MS with standards. (a) Positive-ion mode signals: proline betaine, 4-hydroxyproline betaine and 3-hydroxyproline betaine; (b) negative-ion mode signals: proline betaine-O-glucuronide and proline betaine-O-sulphate.

data (italicised in Table 4) discriminated strongly between fasting and postprandial urine samples at both 2 and 8 h. These two signals were highly correlated both with each other (Fig. 5(a)) and with the positive-ion signals representative of proline betaine (Fig. 5(b)). Detailed FIE-MS/MS analysis of m/z 223 showed a loss of m/z 80, suggesting that it was a sulphonated derivative of proline betaine (Table 5). Analysis of m/z 319 showed fragment ions at m/z 175 and 113 corresponding to glucuronate 'fingerprint' ions⁽²⁹⁾. Both m/z 223

and m/z 319 produced a fragment ion at m/z 143, of which the second-generation fragment ions (m/z 125, 115, 113 and 99) matched the fragmentation spectra of an authentic sample of synthetic proline betaine m/z 143 (Table 5). In addition, FIE-MS/MS analysis of m/z 161, which also correlated with m/z 223 and m/z 319 (Fig. 5(a) and (b)), yielded a fragment ion at m/z 143, with identical second-generation fragment ions (data not shown). In addition, two other correlated negative-ion features, m/z 345 and 343, are most probably glucuronides of a further, as yet unidentified, compound (data not shown).

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

Rank*	Positive-ion signals		Negative-ion signals	
	Fasting v. 2 h	Fasting v. 8 h	Fasting v. 2 h	Fasting v. 8 h
1	<i>166</i>	<i>160</i>	<i>223</i>	<i>223</i>
2	<i>160</i>	<i>301</i>	<i>345</i>	<i>319</i>
3	<i>167</i>	<i>166</i>	<i>161</i>	<i>262</i>
4	<i>391</i>	<i>167</i>	<i>319</i>	<i>422</i>
5	<i>184</i>	<i>171</i>	<i>273</i>	<i>138</i>
6	<i>407</i>	<i>134</i>	<i>207</i>	<i>199</i>
7	<i>145</i>	<i>198</i>	<i>141</i>	<i>202</i>
8	<i>144</i>	<i>170</i>	<i>136</i>	<i>119</i>
9	<i>312</i>	<i>144</i>	<i>215</i>	<i>277</i>
10	<i>381</i>	<i>145</i>	<i>346</i>	<i>142</i>
11	<i>198</i>	<i>149</i>	<i>193</i>	477
12	<i>182</i>	<i>279</i>	<i>209</i>	<i>159</i>
13	<i>183</i>	<i>102</i>	<i>343</i>	<i>427</i>
14	<i>279</i>	<i>273</i>	<i>427</i>	<i>115</i>
15	<i>270</i>	<i>208</i>	<i>333</i>	381
16	<i>365</i>	<i>165</i>	<i>205</i>	<i>380</i>
17	<i>115</i>	<i>109</i>	<i>335</i>	<i>204</i>
18	<i>163</i>	<i>254</i>	<i>199</i>	<i>424</i>
19	<i>282</i>	<i>263</i>	<i>184</i>	447
20	<i>257</i>	<i>478</i>	<i>115</i>	<i>358</i>

* Rank in a random forest classification of fasting v. postprandial urine samples. Fasting, spot urine samples collected after a 12 h (minimum) fast. Signals associated with proline betaine are italicised; flavonone conjugate signals are bold, m/z 381, hesperetin sulphate $[M-H]^{-1}$; m/z 447, naringenin glucuronide $[M-H]^{-1}$; m/z 477, hesperetin glucuronide $[M-H]^{-1}$; twelve individuals, twenty-four samples per time point.

The flavonoid glycosides hesperidin (hesperetin-7-rutinoside) and narirutin (naringenin-7-rutinoside) are abundant in oranges and have previously been suggested as possible biomarkers for citrus exposure^(7,13–15,25,26). As expected⁽¹⁴⁾, conjugates of the flavonoid aglycones (hesperetin sulphate, hesperetin monoglucuronide and naringenin monoglucuronide; m/z 381, 477 and 447, respectively) only appeared in urine 6–8 h after consumption of the standard breakfast (Fig. 6) when monitored by targeted analysis of specific m/z in negative-ion mode data (confirmed by FIE-MS/MS, data not shown). Although these ions appeared in urine 8 h after consumption of orange juice in the breakfast (signals bold in negative-ion data; Table 4), these signals were not explanatory of habitual citrus exposure levels in either 'PRE' or 'fasting' urine samples (see Table S5 of the supplementary material, available online at <http://www.journals.cambridge.org/bjn>).

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

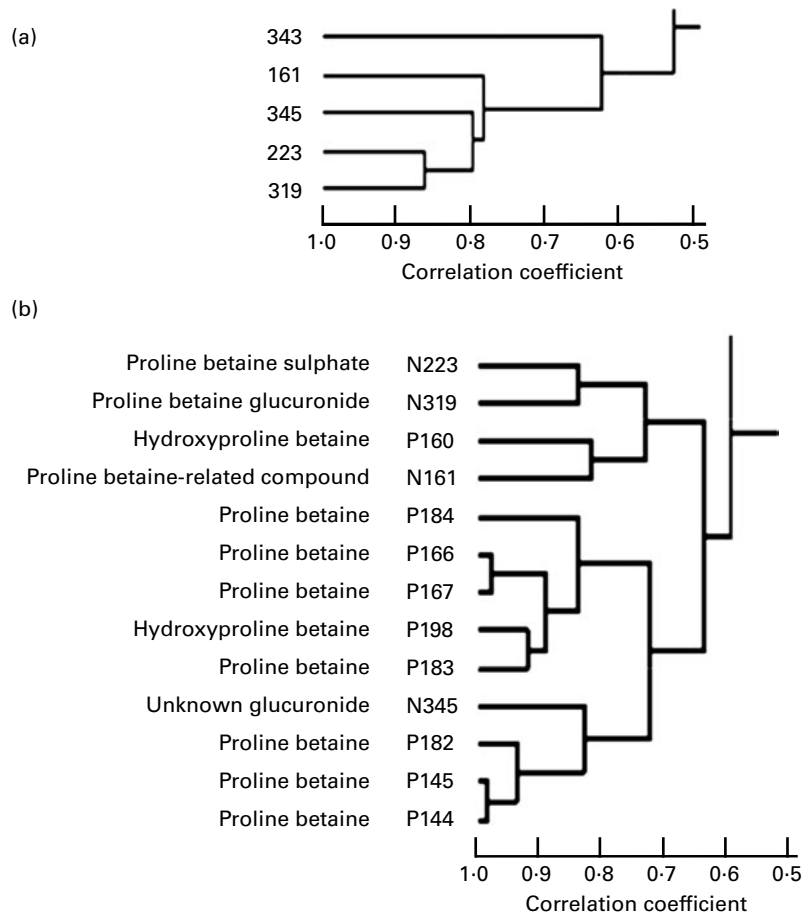


Fig. 5. Analysis of flow injection electrospray-ionisation MS fingerprinting signals discriminating fasting from postprandial urine samples after exposure to a test breakfast containing orange juice. Hierarchical cluster analysis based on the correlation coefficient (Pearson's correlation method) of (a) top negative-ion signals discriminating postprandial urine samples; (b) the combined negative and positive-ion signals in postprandial urine samples, where 'N' represents negative and 'P' represents positive-ion mode signals. Information relating to putative individuals is presented in Fig. 3 and Tables 3 and 5.

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⁽⁶⁾ of replicate urine samples collected at intervals over a 14-month period. We observed that proline betaine, an abundant component of citrus fruits^(30–32), 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⁽³³⁾ or orange juice^(31,33). A recent investigation of proline betaine excretion

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

<i>m/z</i> 223		<i>m/z</i> 319		Proline betaine standard [M-H] ¹⁻ (<i>m/z</i> 143*)	
MS ² ions	MS ² fragment identification	MS ² ions	MS ² fragment identification	MS ² ions	MS ² fragment identification
208	[M-CH ₃]	-	-	125	[M-H ₂ O]
205	[M-H ₂ O]	301	[M-H ₂ O]	115	[M-C ₂ H ₄]
-	-	-	-	113	[M-CH ₃] ₂
193	[M-(CH ₃) ₂]	289	[M-(CH ₃) ₂]	99	[M-N ⁺ (CH ₃) ₂]
179	[M-N ⁺ (CH ₃) ₂]	275	[M-N ⁺ (CH ₃) ₂]	-	-
165	[M-CH ₂ N ⁺ (CH ₃) ₂]	261	[M-CH ₂ N ⁺ (CH ₃) ₂]	-	-
-	-	193	Glucuronide + H ₂ O	-	-
-	-	175	Glucuronide - H	-	-
143†	[M-SO ₃]	143†	[M-Glucuronide]	-	-
141	[M-SO ₃ H ₂]	113	Glucuronide - H - CO ₂ - H ₂ O	-	-
115	[M-SO ₃ -CO]	-	-	-	-
113	[M-SO ₃ H ₂ -CO] or [M-SO ₃ -(CH ₃) ₂]	-	-	-	-
97	[M-SO ₃ -HO]	-	-	-	-
81	[M-SO ₃ -CH ₃ -H ₂ O]	-	-	-	-

* Proline betaine standard with abundant signals at *m/z* 125, 115, 113 and 99.

† MS³ analysis of the MS² product ion *m/z* 143 produced the same fragmentation as the proline betaine standard.

using NMR analysis of postprandial urine samples suggested that it was cleared from the body rapidly and could not be detected easily 14h 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 24h dietary recalls repeated after 3 weeks and analysis of two 24h urine sample collections made concurrently⁽³¹⁾. The previous report concluded that proline betaine was an effective biomarker of citrus exposure, where 24h dietary recall data indicated that citrus products had been consumed within the previous 24h. 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 8h postprandial urine samples after acute exposure to orange juice, we demonstrated that this metabolite is present at elevated levels in overnight void ('PRE') and 'fasting' urine samples in individuals reporting habitually 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 sample, 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⁽³¹⁾ 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

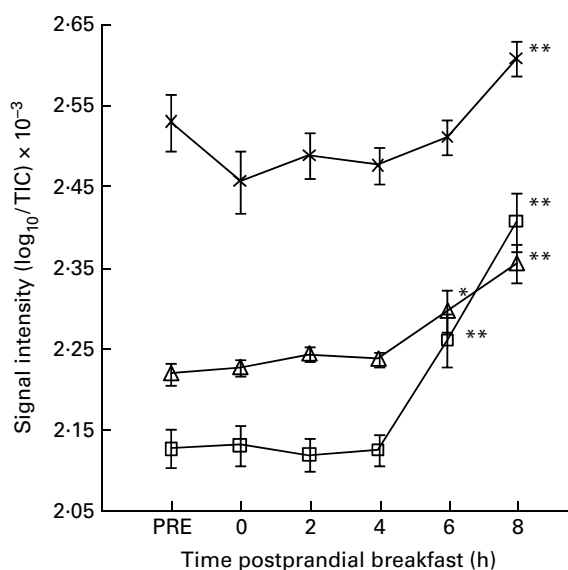


Fig. 6. Orange juice flavonoid aglycone conjugate signals in postprandial urine samples. Flow injection electrospray-ionisation MS signal intensity plots of ions indicative of hesperetin sulphate [M-H]¹⁻ (*m/z* 381, -x-), naringenin monoglucuronide [M-H]¹⁻ (*m/z* 447, -Δ-) and hesperetin monoglucuronide [M-H]¹⁻ (*m/z* 477, -□-). Log/total ion count (TIC), normalisation using sample TIC and logarithmic transformation of data; 'PRE', pre-test day overnight urine voids; '0', spot urine sample after a 12h (minimum) fast; 2, 4, 6 and 8h postprandial breakfast. *Values were significantly different compared with fasting urine samples ($P < 0.05$). **Values were significantly different compared with fasting urine samples ($P < 0.005$).

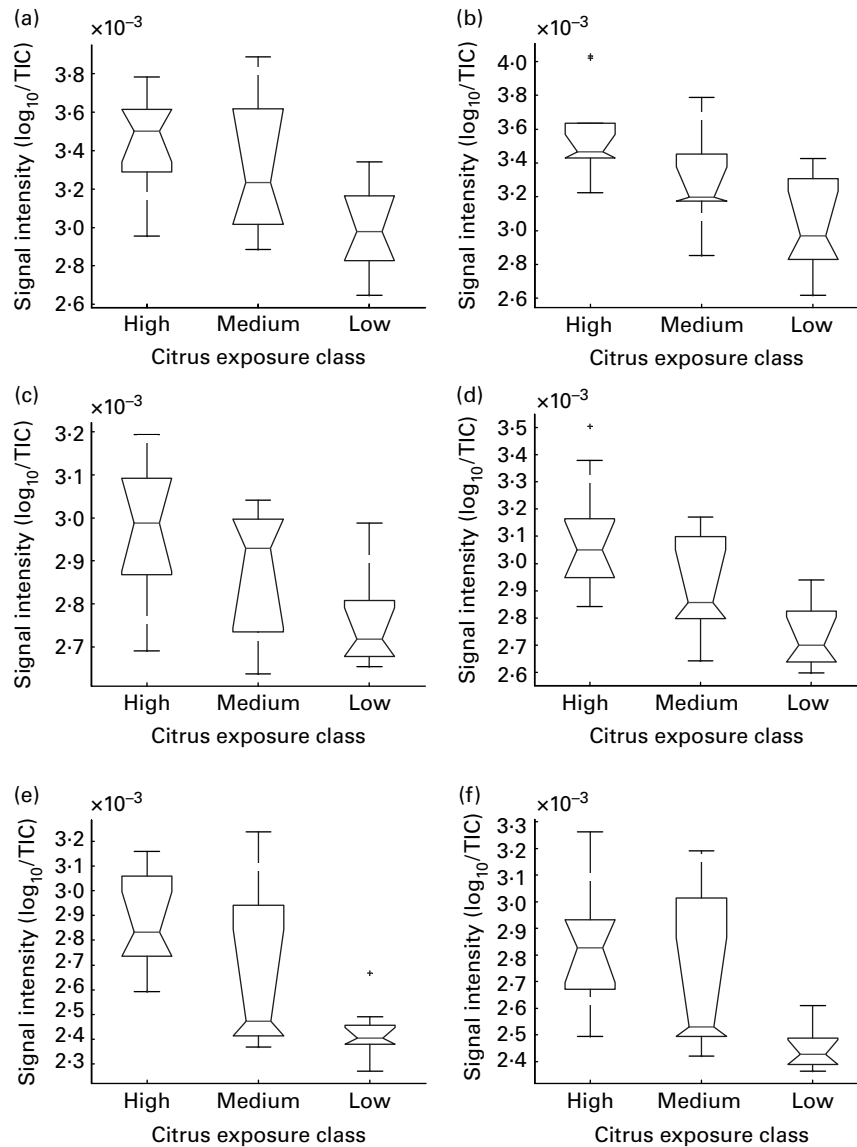


Fig. 7. Box plots of the top positive-ion mode explanatory metabolite signals between 'high', 'medium' and 'low' habitual citrus fruit consumers. (a) *m/z* 144, fasting urine; (b) *m/z* 144, PRE urine; (c) *m/z* 166, fasting urine; (d) *m/z* 166, PRE urine; (e) *m/z* 182, fasting urine; (f) *m/z* 182, 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 *t* 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 × 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.

spectra of an authentic sample of synthetic 3-hydroxyproline betaine [M+H]¹⁺, but the origin of the fragment ion at *m/z* 116 is currently unknown⁽³⁴⁾. 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⁽³¹⁾ 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^(14,15). This is in agreement with a recent review⁽³⁵⁾, which also

concludes that the biotransformation products of the flavone 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 observation 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 individuals 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 participants who reported high citrus consumption (in both pooled overnight urine and spot fasting urine samples), thus further validating the potential biomarker status of this metabolite. 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 provides 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 relationships 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 support, and Dr Wanchang Lin (Manchester School of Biomedicine) 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 fingerprinting experiments, supervised MS support staff, pre-processed data for analysis and edited the manuscript; G. F. undertook the volunteer recruitment, coordinated the volunteer 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

1. Jenab M, Slimani N, Bictash M, *et al.* (2009) Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet* **125**, 507–525.
2. Kristal AR, Peters U & Potter JD (2005) Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* **14**, 2826–2828.
3. Willett WC & Hu FB (2006) Not the time to abandon the food frequency questionnaire: point. *Cancer Epidemiol Biomarkers Prev* **15**, 1757–1758.
4. Bingham S (2002) Biomarkers in nutritional epidemiology. *Public Health Nutr* **5**, 821–827.
5. Bingham SA, Gill C, Welch A, *et al.* (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records *v.* 24-h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* **72**, 619–643.
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. *Eur J Clin Nutr* **61**, 1328–1336.
7. Mennen LI, Sapinho D, Ito H, *et al.* (2006) Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. *Br J Nutr* **96**, 191–198.
8. Porrini M, Gentile MG & Fidanza F (1995) Biochemical validation of a self-administered semiquantitative food-frequency questionnaire. *Br J Nutr* **74**, 323–333.
9. Signorello LB, Buchowski MS, Cai QY, *et al.* (2010) Biochemical validation of food frequency questionnaire-estimated carotenoid, alpha-tocopherol, and folate intakes among African Americans and non-Hispanic Whites in the

- Southern Community Cohort Study. *Am J Epidemiol* **171**, 488–497.
10. Talegawkar SA, Johnson EJ, Caritbers TC, *et al.* (2008) Carotenoid intakes, assessed by food-frequency questionnaires (FFQs), are associated with serum carotenoid concentrations in the Jackson Heart Study: validation of the Jackson Heart Study Delta NRI Adult FFQs. *Public Health Nutr* **11**, 989–997.
 11. Cade J, Thompson R, Burley V, *et al.* (2002) Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutr* **5**, 567–587.
 12. Jia X, Craig LCA, Aucott LS, *et al.* (2008) Repeatability and validity of a food frequency questionnaire in free-living older people in relation to cognitive function. *J Nutr Health Aging* **12**, 735–741.
 13. Spencer JPE, Mohsen MMA, Minihane AM, *et al.* (2008) Biomarkers of the intake of dietary polyphenols: strengths, limitations and application in nutrition research. *Br J Nutr* **99**, 12–22.
 14. Bredsdorff L, Nielsen ILF, Rasmussen SE, *et al.* (2010) Absorption, conjugation and excretion of the flavanones, naringenin and hesperetin from alpha-rhamnosidase-treated orange juice in human subjects. *Br J Nutr* **103**, 1602–1609.
 15. Del Rio D, Costa LG, Lean MEJ, *et al.* (2010) Polyphenols and health: what compounds are involved? *Nutr Metab Cardiovasc Dis* **20**, 1–6.
 16. Favé G, Beckmann ME, Draper JH, *et al.* (2009) Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes Nutr* **4**, 135–141.
 17. Holmes E, Loo RL, Stalder J, *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400.
 18. Llorach R, Urpi-Sarda M, Jauregui O, *et al.* (2009) An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. *J Proteome Res* **8**, 5060–5068.
 19. Scalbert A, Brennan L, Fiehn O, *et al.* (2009) Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435–458.
 20. Walsh MC, Brennan L, Malthouse JPG, *et al.* (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am J Clin Nutr* **84**, 531–539.
 21. Beckmann M, Parker D, Enot D, *et al.* (2008) High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat Protoc* **3**, 486–504.
 22. Favé G, Beckmann M, Lloyd AJ, *et al.* (2011) Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics* (epublication ahead of print version 22 February 2011).
 23. Enot DP, Lin W, Beckmann M, *et al.* (2008) Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat Protoc* **3**, 446–470.
 24. Draper J, Enot DP, Parker D, *et al.* (2009) Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive *m/z* annotation tool utilising predicted ionisation behaviour ‘rules’. *BMC Bioinformatics* **10**, 227.
 25. Brett GM, Hollands W, Needs PW, *et al.* (2009) Absorption, metabolism and excretion of flavanones from single portions of orange fruit and juice and effects of anthropometric variables and contraceptive pill use on flavanone excretion. *Br J Nutr* **101**, 664–675.
 26. Kawaii S, Tomono Y, Katase E, *et al.* (1999) Quantitation of flavonoid constituents in citrus fruits. *J Agric Food Chem* **47**, 3565–3571.
 27. Major HJ, Williams R, Wilson AJ, *et al.* (2006) A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition. *Rapid Commun Mass Spectrom* **20**, 3295–3302.
 28. Bingham S, Gill C, Welch A, *et al.* (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* **26**, S137–S151.
 29. Gu JK, Zhong DF & Chen XY (1999) Analysis of *O*-glucuronide conjugates in urine by electrospray ion trap mass spectrometry. *Fresenius J Anal Chem* **365**, 553–558.
 30. de Zwart FJ, Slow S, Payne RJ, *et al.* (2003) Glycine betaine and glycine betaine analogues in common foods. *Food Chem* **83**, 197–204.
 31. Heinzmann SS, Brown IJ, Chan Q, *et al.* (2010) Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**, 436–443.
 32. Slow S, Donaggio M, Cressey PJ, *et al.* (2005) The betaine content of New Zealand foods and estimated intake in the New Zealand diet. *J Food Compos Anal* **18**, 473–485.
 33. Atkinson W, Downer P, Lever M, *et al.* (2007) Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. *Eur J Nutr* **46**, 446–452.
 34. Chen H-X, Shen S-L, Han F-M, *et al.* (2006) HPLC-ESI/MS analysis of stachydrine and its metabolites in rat urine. *Acta Pharm Sin* **41**, 467–470.
 35. Pérez-Jiménez J, Hubert J, Hooper L, *et al.* (2010) Urinary metabolites as biomarkers of polyphenol intake in humans: a systematic review. *Am J Clin Nutr* **92**, 801–809.
 36. Urpi-Sarda M, Garrido I, Monagas M, *et al.* (2009) Profile of plasma and urine metabolites after the intake of almond [*Prunus dulcis* (Mill.) DA Webb] polyphenols in humans. *J Agric Food Chem* **57**, 10134–10142.
 37. Urpi-Sarda M, Monagas M, Khan N, *et al.* (2009) Targeted metabolic profiling of phenolics in urine and plasma after regular consumption of cocoa by liquid chromatography–tandem mass spectrometry. *J Chromatogr A* **1216**, 7258–7267.