Identification of Human Urinary Biomarkers of Cruciferous Vegetable Consumption by Metabonomic Profiling

William M. B. Edmands, Olaf P. Beckonert*, Cinzia Stella*, Alison Campbell*, Brian G. Lake**, John Lindon*, Elaine Holmes*, Nigel J. Gooderham*.*

*Biomolecular Medicine, Division of Surgery and Cancer, Faculty of Medicine, Imperial College London, South Kensington, London SW7 2AZ, UK; **Leatherhead Food International, Leatherhead, KT22 7RY, UK.

AUTHOR EMAIL ADDRESS: n.gooderham@imperial.ac.uk

RECEIVED DATE (to be automatically inserted after your manuscript is accepted if required according to the journal that you are submitting your paper to)

TITLE RUNNING HEAD: Human Urinary Biomarkers Cruciferous Vegetable Consumption

CORRESPONDING AUTHOR FOOTNOTE:

Please address correspondence to:

Professor NJ Gooderham,

Biomolecular Medicine,

Imperial College London,

Sir Alexander Fleming Building,

London SW7 2AZ, UK.

Email: n.gooderham@imperial.ac.uk

Tel: 020 7594 3188

ABSTRACT: Consumption of cruciferous vegetables (CVs) is inversely correlated to many human diseases including cancer (breast, lung, and bladder), diabetes, cardiovascular and neurological disease. Presently, there are no readily measureable biomarkers of CV consumption and intake of CVs has relied on dietary recall. Here, biomarkers of CV intake were identified in the urine of 20 healthy Caucasian adult males using ^{1}H-NMR spectroscopy with multivariate statistical modeling. The study was separated into three phases of 14 days: a run-in period with restricted CV consumption (phase I); a high CV phase where participants consumed 250g/day of both broccoli and Brussels sprouts (phase II); a wash-out phase with a return to restricted CV consumption (phase III). Each study participant provided a complete cumulative urine collection over 48 hours at the end of each phase; a spot urine (U0), 0-10 hours (U0-10), 10-24 hours (U10-24) and 24-48 hours (U24-48) urine samples. Urine samples obtained after consumption of CV were differentiated from low CV diet samples by 4 singlet ^{1}H-NMR spectroscopic peaks, two of which were identified as S-methyl-L-cysteine sulphoxide (SMCSO) and *N*-acetyl-S-methyl-L-cysteine sulphoxide (NAc-SMCSO). These stable urinary biomarkers of CV consumption will facilitate future assessment of CV in nutritional population screening and dietary intervention studies and may correlate to population health outcomes.

Introduction

Epidemiological studies have consistently demonstrated an inverse association between cruciferous vegetable (CV) consumption and risk of cancer [1,2] particularly lung [3,4], stomach and neoplasms of the lower intestinal tract [5], bladder [6] and breast [7]. Similarly, a protective effect has also been reported for cardiovascular disease[8], diabetes[9] and neurological disease[10]. The family *cruciferae* is composed of foods of the species *Brassica oleracea* (cabbage, broccoli, cauliflower, Brussels sprouts, kale) and >350 other genera. The chemoprotective power of CV against neoplastic disease is attributed to their abundant levels of glucosinolates (-thioglycoside-N-hydroxysulfates) in the human diet. Glucosinolates are hydrolysed by thioglucoside enzymes known as myrosinases which are released upon mastication of the plant material, this produces, dependent on reaction conditions, the biologically active isothiocyanates (ITC), indoles, nitriles and epithionitriles[11].

The ITCs are reported to form glutathione (GSH) conjugates and albumin adducts *in vivo* [12,13]. ITC-GSH conjugates are known to be primarily excreted as ITC-mercapturates by the mercapturic acid pathway and have subsequently been used as urinary biomarkers of glucosinolate exposure[14]. However, the efficiency of the ITC mercapturates as measures of CV exposure is limited by a multitude of factors, such as abundance of the particular glucosinolates in the vegetable cultivar and species consumed, liberation and rate of formation from the plant material, cooking method, bioavailability, absorption from the digestive system, distribution around the body, endogenous metabolism and excretion[15]. Effective correlation between measurement of urinary ITC mercapturates and assessment of CV consumption is therefore reliant on many factors. Biomarkers of exposure, effect and susceptibility can potentially circumvent the difficulties associated with assessing the relationship between diet and detrimental/beneficial health outcomes and the efficacy of dietary interventions. Novel dietary biomarkers could also potentially provide further insights into the mechanisms of action of dietary components.

Metabolic profiling using high resolution spectroscopy, mainly nuclear magnetic resonance (NMR), or mass spectrometry with either liquid-chromatography (LC) or gas-chromatography technology, can be used to characterize the dynamic metabolic response of organisms to experimental interventions[16]. High resolution analytical platforms together with the aid of multivariate statistical modeling have been successfully applied to determine the metabolic response to a range of nutritional interventions in humans, for example, the effects of vegetarian and meat diets[17] and recently by cluster analysis the identification of metabolic phenotypes associated with habitual dietary patterns; highlighting potentially robust biomarkers of meat and vegetable consumption[18]. Furthermore, metabolic phenotypic change has been reported attributable to soy isoflavone[19], almond[20], chamomile (*Matricaria recutita L.*)[21], alcohol[22], citrus fruit[23], polyphenolics[24] and green and black tea dietary interventions[25]. The use of metabonomic technologies for the identification of putative biomarkers of foodstuffs of significant public health interest has recently been reviewed by the Food Standards Agency (UK)[26]. It was concluded that it is possible to successfully identify potentially useful biomarkers of nutritional intervention utilising metabonomic methods. However, these putative biomarkers must be validated in population cohort studies and dietary surveys.

Here we applied a metabolic profiling strategy using ^{1}H-NMR spectroscopy and chemometric modeling to assess the changes in the human metabolic phenotype following nutritional intervention with CV intake over a two week period and have identified novel and robust urinary biomarkers of CV consumption.

Experimental Section

Study protocol

Twenty healthy male Caucasian subjects were recruited to a three period dietary intervention study that was controlled and planned by a clinical nutritionist[27]. Each of the three periods consisted of 14 days duration. In phases I and III a low cruciferous vegetable diet was consumed for 12 days and for day's 11-14 of each phase dietary restrictions were extended to pharmaceutical agents and foods such as *alliums*, which might confound dietary intervention analysis (Figure 1). Phase I was immediately followed by Phase II where subjects were maintained under supervision on a high cruciferous vegetable diet for 12 days. This consisted of a 250g/day of both broccoli and Brussels sprouts incorporated into a varying menu consumed under supervision at the BIBRA international clinical unit (Carshalton) at breakfast and dinner. Phase III, a return to the low CV diet immediately followed Phase II. On day 13 of each phase, after an overnight fast, a time zero spot urine sample was obtained from each study participant. Participants were then maintained on a standardized diet and urine sample collections were obtained for the periods 0-10 hr, 10-24 hr and 24-48 hr. Therefore each subject provided 4 samples for each of the three dietary intervention periods. Urine samples were stored at -80˚C until required for analysis.

Urine Sample Preparation

A 1ml aliquot of each of the urine samples was taken for ^{1}H-NMR spectroscopic analysis. To reduce pH induced chemical shift variability, a 400μl aliquot of urine sample was combined with 200 μl of 100 mM phosphate buffer solution (pH 7.4), before centrifugation at 13,000 g $_{av}$ for 10 mins to remove particulate materials. Aliquots (500 μl) were transferred to a 96 deep-well plate in a randomized order with quality control samples (comprised of a pooled urine sample from 6 study participants) inserted every fifteenth sample and were mixed with 50μl of TSP/D_2O solution (3-trimethylsilyl 2,2,3,3-tetradeuteropropionate in deuterium oxide, 1 mM final concentration). TSP is an internal chemical shift reference ($δ^1$H 0.0) and D_2O provides a lock signal for the NMR spectrometer. The plates were covered

with adhesive film, mixed carefully and immediately prior to NMR analysis, were centrifuged at 3,000 g $_{av}$ for 5 mins.

Flow Injection 600MHz ^{1}H-NMR Spectroscopic analysis of Urine Samples

NMR spectra were acquired in the stop-flow mode using a Bruker AMX600 spectrometer (operating at 600.13MHz ^{1}H frequency at ambient probe temperature (298 ± 1 K)) with a 5mm single cell ^{1}H/^{13}C inverse detection flow probe with an active volume of 250µl. Sample transfer from the 96-well plate to the probe was automated by a Gilson (Middleton, WI, USA) XL233 sample-handling system interfaced to the XWinNMR data system for control and timing. For each sample, 450µl was aspirated from the well plate by sampling needle into the transfer line and separated from subsequent samples by 500µl of a push buffer containing HPLC grade water with 0.1% Sodium Azide (NaN_3). The transfer time for each sample was approximately 180s to allow time for the sample to adjust to the probe temperature. Spectra were acquired, a standard one dimensional pulse sequence using the first increment of a NOE pulse sequence to suppress the large NMR signal of the solvent water[28]. For each sample 64 free induction decays were collected into 32768 time domain data points with a spectral width of 3140.7Hz, an acquisition time of 5.22s and a total recycle time of 6.22s. The FIDs were multiplied by a line-broadening function of 1.0Hz to improve the signal–to-noise ratio prior to Fourier transformation, phasing and baseline correction.

^{1}H-NMR Urine spectra pre-processing

Full resolution spectra (δ10.0-0.24) were used for chemometric pattern recognition. Variation in the effect of pre-saturation of the water signal was removed by zeroing intensities in the region of the spectra δ 4.5-4.9. All spectra were referenced to the TSP internal standard reference (δ 0.0). Baseline correction of all of the spectra was achieved by polynomial curve fit (using a Matlab Version 7.0). Spline cubic interpolation was used to give all the spectra (42,000 data points) the same chemical shift. All urinary ^{1}H-NMR spectra were spectrally aligned utilizing the recursive segment-wise peak

alignment (RSPA) algorithm[29]. To partially compensate for urine concentration differences between different samples all spectra were normalized using probabilistic quotient on full resolution data sets[30].

Principal Component Analysis of the ¹H NMR Urine Spectra.

In order to provide a data overview, explore intrinsic structure in the data and to identify any outlier samples, NMR spectral data were reduced to latent variables using principal components analysis (PCA)[31].

Partial Least Squares-Discriminant Analysis of the ¹H NMR Urine Spectra.

Partial Least Squares-Discriminant Analysis (PLS-DA), which models the maximum separation between defined class samples in the data using a dummy class variable as the Y data matrix, was used to optimize the extraction of candidate biomarkers. Subsequently, an orthogonal (O)-PLS-DA algorithm based on the method of orthogonal signal correction by Wold *et. al* [32] but incorporated directly into the PLS model[33], was used to remove sources of systematic variation in the X matrix, unrelated to the class of interest (in this case cruciferous vegetable consumption). The O-PLS-DA model is based on unit variance 'uv' scaled data, which is subsequently back-transformed by multiplying the O-PLS coefficients by their standard deviations. The O-PLS coefficients are colour coded according to correlation with class, the range of which are set to the maximum and minimum of the unit variance scaled model weight and gives an indication of the relative influence of an individual metabolite in class discrimination.

Structural Identification of Candidate Biomarkers.

Signals noted as being key discriminators of the high CV consumption class were identified using a combination of 2-dimensional NMR spectroscopy and statistical total correlation spectroscopy (STOCSY)[34]. The 2D *J*-resolved (J-RES) ¹H homonuclear urinary spectrum was recorded using a 90° pulse of 8.9μs, 48 transients of 4096 complex points, 64 increments, relaxation delay of 2 s, and spectral

widths of 4 kHz in the ^{1}H (f2) and 65 Hz in the J (f1) dimensions, respectively. ^{1}H-^{13}C 2D HSQC NMR spectra were recorded in echo-anti-echo mode using the gradient selection. The ^{1}H and ^{13}C spectral widths were 10.5 and 200 ppm respectively using a standard pulse sequence with recycle delay of 2 s. The δ-shifted sine bell window functions were applied for both dimensions before Fourier transforming into data with the standard Bruker linear-prediction. The delay for the ^{1}H-^{13}C 2D HMBC proton carbon long-range coupling was optimised for 10 Hz. For processing, a (π/2) squared sine bell window function in both dimensions was applied. Data comprising 4096 data points were processed in f2 and 2048 in f1, respectively.

STOCSY driven from the key signals differentiating the CV class δ2.83 (s) and δ2.78 (s) was also used to extract further information regarding correlations between signals. Both ^{1}H and ^{13}C NMR assignments were made using online NMR database searches (eg. human metabolome database Hmdb, Genome Alberta), NMR prediction software (ACD labs) and published spectral data at neutral pH [35, 36]. The candidate metabolites SMCSO and NAc-SMCSO were synthesised following previously published methods [37-39]. The synthetic standards were subjected to ^{1}H-NMR, 2D HSQC and HMBC analysis, and spiked into human urine in order to confirm identity.

Biochemical Characterisation of Broccoli and Brussels Sprout Aqueous extractions

^{1}H-NMR spectra of broccoli and Brussels sprouts were acquired to ascertain whether any of the discriminatory signals in the urine profiles of participants on the high CV diet directly originated from these vegetables. One ml of CH_3CN:D_2O (50:50; Acetonitrile:Deuterium Oxide) was combined with approximately 600mg of microwave cooked Brussels sprouts or broccoli and homogenised using a Quiagen tissue lyser (Retsch, ltd.) for 10 mins at the highest setting. An internal ^{1}H-NMR reference standard TSP (δ~0ppm) was added to a final concentration of 0.1mM. Sodium azide (1% w/v) was added as an antimicrobial. The resulting homogenates were mixed in a 1:1 ratio with human saliva and allowed to ferment for 4 hours at physiological temperature (37°C) at low pH (1-3) to simulate the

hydrolysis conditions found in the human stomach. Following fermentation the homogenates were centrifuged at 10,000 g for 10 mins and the supernatant was filtered using 25mm Micropore syringe filters and transferred to a standard 5mm outer diameter NMR tube. The Carr-Purcell-Meiboom-Gill (CPMG) was applied to aqueous vegetable extracts to clearly observe the water soluble components and to reduce the line-broadening effect of macromolecular components and to give clear representation of the lower molecular weight components of interest. The CPMG sequence [40-42] manipulates the T_2 (time constant for the decay of magnetization in the x-y plane perpendicular to the applied field B_0) of protons to reduce contributions from proteins, lipids and other macromolecules.

Results and Discussion

Analysis of the raw spectra showed clear differentiation of Phase II (high CV) with Phases I and III (low CV) by the presence of 4 singlet peaks δ2.75 (i), δ2.78 (ii), δ2.80 (iii) and δ2.83 (iv) (Figure 2A). Even though the diet was strictly controlled over the 42 day study duration, inter-sample differences were still apparent in the ^{1}H-NMR urine spectra obtained for the 20 volunteers over the 3 dietary phases making the metabolic response to dietary intervention difficult to identify. PCA analysis of this entire dataset supported this observation showing the dataset was characterized by inter-sample variations rather than a significant variation due to the dietary interventions applied (Figure 2B). Outliers identified in the PCA scores plot due to poor water suppression or analytical artifact (n=13) were eliminated prior to subsequent modeling. Similarly, outliers due to the presence of ethanol, indicating non compliance of the participant with dietary protocol were also excluded (n=3).

Both global and pair-wise comparison of ^{1}H NMR spectra for each study phase at all collection times was performed using O-PLS-DA. Phases I and III, corresponding to the baseline diet and the washout period (both low CV) could not be separated on the basis of the metabolic profiles (Figure 2C).

The O-PLS-DA coefficient plots of phase I vs. phase II and Phase II vs. Phase III urine samples normalised to the probability quotient indicated that the strongest discriminatory signals for CV intervention were the four singlet peaks identified in the raw spectra at δ2.75 (i), δ2.78 (ii), δ2.80 (iii) and δ2.83 (iv) (Figure 2D). The chemical shifts of these resonances were typical of bis-allylic compounds. These four singlets were found by comparison of any of the phase II time-points with any of the phase I or phase III time-point urine samples and were highly discriminant of the high CVs period. These peaks, therefore, were considered statistically significant biomarkers of CV intake. The O-PLS-DA regression coefficient plot provided evidence of lower signal intensity peaks located further up field of the 4 singlets in the region δ3-3.5, which were overlapped and prevented facile determination of splitting pattern (Figure 2D). In order to derive further structural information for each of these candidate

CV biomarkers, a correlation matrix was calculated from the apex of each of the peaks i-ix using the STOSCY method [34]. Each individual singlet peak showed strong correlation with all of the other singlet peaks identified by O-PLS-DA. The lower intensity signals in the region δ3-3.5 which were discriminatory in the O-PLS-DA model also showed relatively strong correlation to peaks i, ii, iii and iv, however the lower intensity of the peaks v-ix reduced the correlation coefficients, this was explicable due to peak overlap and potentially the effects of urinary components such as proteins and peptide fragments affecting the zero baseline values. The STOCSY coefficients representing strength of correlation between these signals are summarized in Table 1.

Intensity of the 4 singlet peaks was measured by integration of the normalized ^{1}H-NMR spectra to ascertain a semi-quantitative urinary excretion profile of the spectral features in Phases I, II and III. The four singlet peaks were exclusive to the high CV consumption phase II (Figure 3), the highest concentration being excreted in the U0 spot urine sample. This is entirely consistent with the termination of the high CV diet on the evening of the day prior to urine collection. Analysis of the intensity plots also revealed evidence of inter-individual differences in the excretion of the 4 singlet peaks across the urinary time-points (reflected in the standard deviation). Although depleted from the first collection point post cessation of the CV diet, candidate biomarkers i-iv were still significantly elevated above baseline levels at 48 h post consumption.

The structural identity of the NMR spectral features identified as diagnostic of CV consumption between ~δ3-3.5 was difficult to ascertain by interrogation of the standard one-dimensional spectrum. The splitting pattern of these spectral features was confounded by spectral overlap of metabolites in greater abundance within the urine. Thus a 2D ^{1}H-^{1}H J-Res NMR experiment was conducted to increase signal dispersion and provide visualisation of the underlying J-couplings of these peaks in a two dimensional plot.

The *J*-Res experiment confirmed that these unknown peaks in this crowded area of the spectra to be part of an ABX spin system due to the similarity in *J*-coupling between an A (δ3.10) B (δ3.27) and X representing a doublet of doublets further up field at (δ4.38) (Figure 4A). An ABX spin system occurs when two methylene protons A and B are close to a proton on a chiral centre and are said to be *diastereoscopic* and are likely to have differing chemical shifts and *J*-couplings. The *J*-coupling patterns does not arise due to any restricted C-C bond rotation but from an absence of any bisecting plane of symmetry in the methylene group. Essentially, each proton appears as a doublet of doublets as it is *J*-coupled to two different protons in differing chemical environments in the chiral molecule. However, the X doublet of doublets at δ4.38 was not identified in the O-PLS-DA loadings coefficient plot due to the proximity of this peak to that of water, resulting in its removal due to varying solvent suppression. Analysis of the raw NMR spectra identified this doublet of doublets at δ4.38 confirming its presence in the phase II urine samples. The *J*-Res spectrum provided *J*-couplings of 13.43Hz between the A and B of the spin system and a value of 2.87Hz between the A and X and 10.56Hz between the B and X. This pattern of *J*-couplings was shared between all three spectral features suggesting membership of all three spectral features to the same ABX spin system.

The 2D homonuclear *J*-res experiment suggesting a chirally centred metabolite combined with the strong association of the singlet resonances δ2.83, δ2.81, δ2.78 and δ2.75 with high cruciferous vegetable consumption were consistent with low molecular weight sulphur or nitrogen containing chirally centred compounds. These unknown metabolites contained magnetically isolated functional groups giving rise to the singlet peaks of interest, which were likely derived from the vegetables themselves or a metabolite of vegetable-derived material. It is known from the literature that CVs ubiquitously contain a methylated cysteine sulphoxide; SMCSO[37]. The 2D *J*-res NMR experiment and the spectral region of the singlet peaks was compatible with the structure of this sulphur containing amino acid.

Two dimensional heteronuclear NMR experiments ^{13}C HSQC and HMBC were used to provide further structural information on the tentative singlet biomarkers and ABX spin system(s) identified within the high CV consumption study phase (Figure 4). The ^{13}C HSQC experiment provided a carbon shift of the carbon isotope directly bonded to proton(s) of the unknown singlet peaks and the protons of the ABX spin system identified in the ^{1}H-NMR spectra of the urine. This was further augmented by acquisition of multiple bond coherence values of the spectral features using ^{13}C HMBC, summarized in Table 2. Based on the shared HSQC ^{13}C shifts of the 4 singlet peaks this suggested closely related or very similar compounds sharing a common origin or metabolic pathway.

The variation of the HMBC values for the four singlet peaks suggests they are diagnostic of CV consumption and represent four separate metabolites all with a similar structure and shared metabolic origin. This evidence suggests that the functional group responsible for the four singlet peaks are likely methyl protons from an S-methyl-sulphoxide functional group, supporting the proposal that one of the singlet peaks was the vegetable derived product SMCSO and the other 3 singlet peaks are metabolites of this. The two heteronuclear HSQC and HMBC NMR experiments were repeated for an aqueous extraction of broccoli and compared to the human urinary spectra (Figure 4B).

The presence in high concentration of the singlet of interest δ2.83 in the aqueous broccoli and Brussel sprouts extract and the similarity in both the directly coupled and long distance ^{13}C shifts with the product in urine confirmed the excretion of SMCSO in human urine and a major metabolite of SMCSO; the most predominant urinary metabolite of which is reported to be NAc-SMCSO [38, 43]. CVs are known to contain the sulphur containing phytochemical SMCSO derived from amino acid biosynthesis[46].Within the plant tissues it forms 1-2% of the dry weight of many varieties of CVs and is therefore ubiquitous and abundant in these vegetables[47].

Perturbations in the global metabolic phenotype of the study subjects provided a clear separation between a dietary regime high in cruciferous vegetables and a diet without cruciferous vegetable consumption. Application of the O-PLS-DA chemometric method to the ^{1}H-NMR spectra of human urine samples identified four significant singlet peaks characteristic of CV consumption, which showed decreasing urinary concentration with time after cessation of CV consumption. Application of two-dimensional NMR spectroscopic experiments, homonuclear *J*-res, HSQC and HMBC to human urine samples indicated a low molecular weight chiral molecule with isolated functional groups of protons attached to either a sulphur or nitrogen. Repetition of these experiments on aqueous extracts of CVs confirmed the vegetable origin of one of the biomarkers in human urine. Based on the analytical evidence, two of the unknown metabolites were identified as a methylated cysteine sulphoxide and its *N*-acetylated product. This is consistent with the known abundant occurrence of SMCSO in CVs[37,46,47]. Based on the similarity of their spectral properties, the remaining unknown singlet peaks were hypothesized to be metabolites of deamination of SMCSO (Figure 5).

This present study confirms that urinary excretion of SMCSO and its *N*-acetylated mammalian metabolite is diagnostic of subjects consuming CV. SMCSO (also known as methiin) is reported to occur widely within both CVs and the *allium* family as a plant secondary metabolite and component of the sulphur containing amino acid pool[46]. The fact that alliums were restricted during the study period means that the compound was almost certainly obtained from the high CV diet in Phase II. SMCSO is found in abundance in both Brussels sprouts (~68mg/100g) and broccoli (~19.1mg/100g) and dependent on species and cultivar, can form between 0.05-0.86% of the dry weight[47]; it is ubiquitously present in vegetables of the family *cruciferae*[47] as a sulphur containing amino acid.

Conclusions

The bulk of research on CVs has focused upon the glucosinolates and their hydrolysis products and the role of SMCSO has largely been ignored, despite reports in the literature of the compound's biological activity including effects on hyperlipidemia, as an antidiabetic, antimicrobial and evidence of anti-genotoxicity[48-51]. The level of targeted research of SMCSO as an anti-carcinogenic compound appears to be limited and it has not been studied as a compound of biological significance in human *in vivo* studies within the context of diet and disease. The use and limitations of glucosinolate hydrolysis products as biomarkers of CV consumption has been examined thoroughly by Holst and Williamson[15]. These authors discuss the issues of glucosinolate hydrolysis, bioavailability, metabolism, the multitude and diversity of glucosinolates and their hydrolysis products derived from cruciferous vegetables and the difficult analytical challenges they present and conclude that better biomarkers of exposure are required.

The use of SMCSO as a biomarker of CV consumption is less affected by such variables and therefore has potential utility in targeted epidemiological studies of diet, cancer risk, diabetes, neurological and cardiovascular disease and population metabonomics. It could also be used to monitor adherence to a proscribed diet of CVs and/or *alliums* in human dietary intervention studies; this would certainly be preferable to the use of food diaries. Moreover, the use of this NMR-directed metabolic profiling approach enabled the detection of four statistically correlated metabolites, whose excretion covaries with CV consumption. A panel of four, rather than a single candidate may provide better sensitivity as an indicator of CV consumption. The identification of SMCSO and its N-acetylated metabolite as urinary biomarkers of CV consumption provides opportunities to examine the role of CV in human health, and also the impetus for further study of the biological role and anti-carcinogenic potential of this plant secondary metabolite within humans.

Figure 1 Study protocol and the nutritional intervention strategy.

U0: initial spot urine sample; U0-10; U10-24; U24-48 hours post spot urine sample cumulative collections.

Figure 2 Analysis of 600 MHz ^{1}H NMR human urine data from participants before (Phase I), during (Phase II) and after (Phase III) consumption of a high concentration cruciferous vegetable diet (500g/day): A) Partial ^{1}H spectra at the end of high CV diet Phase II (top) and low CV diet Phase I (bottom) **i**: $\delta 2.75$; **ii**: $\delta 2.78$; **iii**: $\delta 2.81$; **iv**: $\delta 2.83$. B) PCA scores plot demonstrating that metabolic variation is dominated by inter-subject variation rather than dietary effect; C) O-PLS-DA score plot analysis showing discrimination of high CV (Phase II) and restricted CV consumption (Phases I and III) and is based on a model with good explanatory ($R^2\hat{Y}$:0.783) and predictive ($Q^2\hat{Y}$:0.562) ability; D) O-PLS-DA regression coefficient loadings plot of high CV (Phase II) vs. restricted CV (phases I and III) confirming the four discriminatory singlet peaks of interest identified by analysis of the raw spectra **i**: $\delta 2.75$; **ii**: $\delta 2.78$; **iii**: $\delta 2.81$; **iv**: $\delta 2.83$. **inset:** expanded region $\delta 3$-3.5, provides evidence of further discriminatory ^{1}H-NMR signals at higher field with a lower discriminatory power due to peak overlap and lower relative intensity (**v**: $\delta 3.10$; **vi**: $\delta 3.12$; **vii**: $\delta 3.15$; **viii**: $\delta 3.28$; **ix**: $\delta 3.32$). The signal orientation indicates the higher (positive orientation) or lower (negative orientation) concentration of characteristic metabolites with CV consumption. The colour of the signal reflects the significance of the correlation with diet (red indicates highest significance and blue indicates no significance). Coefficient values for the peaks are as follows **i**: $\delta 2.75$ (R^2 0.51). **ii**: $\delta 2.78$(R^2 0.85); **iii**: $\delta 2.81$ (R^2 0.72); **iv**: $\delta 2.83$ (R^2 0.76).

Figure 3 Average individual urinary excretion kinetics of the 4 singlet peaks associated with CV consumption. **i**: $\delta 2.75$; **ii**: $\delta 2.78$; **iii**: $\delta 2.81$; **iv**: $\delta 2.83$

Integration of the singlet peaks from the normalized spectra revealed the ^{1}H-NMR spectral features were exclusive to the high CV consumption phase II. Values are mean and SD of the peak integrals.

Figure 4 2D [1]H-^{1}H *J*-Res and Carbon 13 NMR experiments (400MHz).

A. Deconvolution of the NMR spectral features identified as characteristic of CV consumption between δ3-3.50 provide evidence of an ABX spin system of a chiral molecule

B. ^{1}H-^{13}C HSQC (upper spectra): ^{1}H-^{13}C HMBC (lower spectra); Left: Human Urine; Right: aqueous broccoli extract; ^{1}H and ^{13}C shifts in human urine diagnostic of CV consumption match those of the peak at δ2.83 in brocolli. ^{1}H-NMR: A: δ2.83; B: δ2.81; C: δ2.78; D: δ2.75. ^{13}C-NMR: 1: δ40.2; 2: δ55.7; 3: 60.5; 4: 60.9; 5: 63.0.

Figure 5 The human metabolism of SMCSO based on the report of Waring *et. al* [39].

The pathways were deduced after radioactive SMCSO was administered to humans (*po*) and identification of urea and inorganic sulphate formation as the primary route of excretion (~50%), followed by N-acetylation (~30%), unmetabolised excretion (~12%) and finally by deamination (~8%). Products in solid-boxes were detected in human urine in the present study by ^{1}H-NMR spectroscopy.

TABLES

Table 1. Pearson correlation coefficient values for the covariant ^{1}H-NMR spectral features identified by O-PLS-DA and STOCSY.

Pearson correlation coefficients (R) of covarying peaks identified by STOCSY									
Covariant peaks	2.75 (i)	2.81 (ii)	2.78 (iii)	2.83 (iv)	*3.10 (v)	*3.12 (vi)	*3.15 (vii)	*3.28 (viii)	*3.32 (ix)
2.75 (i)	1	0.673	0.735	0.683	0.531	0.131	0.431	0.446	0.506
2.81 (ii)	0.673	1	0.898	0.954	0.535	0.128	0.369	0.517	0.456
2.78 (iii)	0.735	0.898	1	0.932	0.630	0.112	0.426	0.461	0.457
2.83 (iv)	0.683	0.954	0.932	1	0.536	0.064	0.330	0.500	0.457
*3.10 (v)	0.531	0.535	0.630	0.536	1	0.411	0.481	0.644	0.630
*3.12 (vi)	0.131	0.128	0.112	0.064	0.411	1	0.416	0.338	0.300
*3.15 (vii)	0.431	0.369	0.426	0.330	0.481	0.416	1	0.299	0.257
*3.28 (viii)	0.446	0.517	0.461	0.500	0.644	0.338	0.299	1	0.827
*3.32 (ix)	0.506	0.456	0.457	0.457	0.630	0.300	0.257	0.827	1

*reduced correlation coefficients for lower intensity spectral peaks are methodologically explicable by spectral overlapping and zero baseline effects potentially caused by urinary proteins and peptides fragments.

Table 2. Summary of ^{13}C shifts for both the HSQC and HMBC experiments.

^{1}H-NMR unknowns O-PLS-DA			
^{1}H-NMR (δ PPM)	splitting pattern	HSQC (δ^{13}C)	HMBC (δ ^{13}C)
2.83	singlet	40.7	55.7
2.81	singlet	40.1	60.5
2.78	singlet	40.2	61.1
2.75	singlet	39.6	63.1
3.32	ABX	60.5	39.5; 57.1
3.28	ABX	62.1	40.2; 69.6
3.1	ABX	61.7	69.2; 180.1
3.12	ABX	47.8	55.1
3.15	ABX	44.4	59.8
4.38	doublet of doublets	69.1	180.7

REFERENCES

1. Verhoeven DT, Goldbohm RA, van Poppel G, Verhagen H, van den Brandt PA. Epidemiological studies on brassica vegetables and cancer risk. Cancer Epidemiol. Biomarkers Prev. 1996;5(9): 733-48.

2. Gullett NP, Ruhul Amin ARM, Bayraktar S, et al. Cancer Prevention With Natural Compounds. Semin. Oncol. 2010;37(3): 258-81.

3. Tang L, Zirpoli G, Jayaprakash V, et al. Cruciferous vegetable intake is inversely associated with lung cancer risk among smokers: a case-control study. BMC Cancer. 2010;10(1): 162.

4. Lam TK, Ruczinski I, Helzlsouer KJ, Shugart YY, Caulfield LE, Alberg AJ. Cruciferous Vegetable Intake and Lung Cancer Risk: A Nested Case-Control Study Matched on Cigarette Smoking. Cancer Epidemiol. Biomarkers Prev. 2010; 19(10): 1–7..

5. van Poppel G, Verhoeven DT, Verhagen H, Goldbohm RA. Brassica vegetables and cancer prevention, Epidemiology and mechanisms. Adv. Exp. Med. Biol. 1999;472: 159-68.

6. Tang L, Zirpoli GR, Guru K, et al. Intake of Cruciferous Vegetables Modifies Bladder Cancer Survival. Cancer Epidemiol. Biomarkers Prev. 2010;19(7): 1806-11.

7. Boggs DA, Palmer JR, Wise LA, et al. Fruit and Vegetable Intake in Relation to Risk of Breast Cancer in the Black Women's Health Study. Am. J. Epidemiol. 2010;172(11): 1268-79.

8. Joshipura KJ, Ascherio A, Manson JE, et al. Fruit and Vegetable Intake in Relation to Risk of Ischemic Stroke. JAMA: 1999;282(13): 1233-39.

9. Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am. J. Clin. Nutr. 2005;82(3): 675-84.

10. Gu Y, Nieves JW, Stern Y, Luchsinger JA, Scarmeas N. Food Combination and Alzheimer Disease Risk: A Protective Diet. Arch. Neuro. 2010;67(6): 699-706.

11. Getahun SM, Chung FL. Conversion of glucosinolates to isothiocyanates in humans after ingestion of cooked watercress. Cancer Epidemiol. Biomarkers Prev. 1999;8(5): 447-51.

12. Conaway CC, Yang YM, Chung FL. Isothiocyanates as cancer chemopreventive agents: their biological activities and metabolism in rodents and humans. Curr. Drug Metab. 2002;3(3): 233-55.

13. Kumar A, Vineis P, Sacerdote C, Fiorini L, Sabbioni G. Determination of new biomarkers to monitor the dietary consumption of isothiocyanates. Biomarkers. 2010;15(8): 739-45.

14. Zhang Y. Cancer-preventive isothiocyanates: measurement of human exposure and mechanism of action. Mutat. Res. Fundam. Mol. Mech. Mugag. 2004;555(1-2): 173-90.

15. Holst B, Williamson G. A critical review of the bioavailability of glucosinolates and related compounds. Nat. Prod. Rep. 2004;21(3): 425-47.

16. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11): 1181-89.

17. Stella C, Beckwith-Hall B, Cloarec O, et al. Susceptibility of Human Metabolic Phenotypes to Dietary Modulation. J.Proteome Res. 2006;5(10): 2780-88.

18. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am. J. Clin. Nutr. 2011; 93(2):314-21.

19. Solanky KS, Bailey NJ, Beckwith-Hall BM, et al. Biofluid ^{1}H NMR-based metabonomic techniques in nutrition research - metabolic effects of dietary isoflavones in humans. J. Nutr. Biochem. 2005; 16(4): 236-44.

20. Llorach R, Garrido I, Monagas M, Urpi-Sarda M, Tulipani S, Bartolome B, Andres-Lacueva C. Metabolomics study of human urinary metabolome modifications after intake of almond (Prunus dulcis (Mill.) D.A. Webb) skin polyphenols. J. Proteome Res. 2010 5; 9(11):5859-67

21. Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Holmes E. A metabonomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion. J. Agri. Food Chem. 2005;53(2): 191-96.

22. Teague C, Holmes E, Maibaum E, et al. Ethyl glucoside in human urine following dietary exposure: detection by 1H NMR spectroscopy as a result of metabonomic screening of humans. The Analyst. 2004;129(3): 259-64.

23. Heinzmann SS, Brown IJ, Chan Q, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. Am. J. Clin. Nutr. 2010;92(2): 436-43.

24. van Duynhoven J, Vaughan EE, Jacobs DM, Kemperman RA, van Velzen EJ, Gross G, Roger LC, Possemiers S, Smilde AK, Doré J, Westerhuis JA, Van de Wiele T. Metabolic fate of polyphenols in the human superorganism. Proc. Natl. Acad. Sci. USA. 2011;108(1):4531-38.

25. Van Dorsten FA, Daykin CA, Mulder TPJ, Van Duynhoven JPM. Metabonomics Approach To Determine Metabolic Differences between Green Tea and Black Tea Consumption. J. Agri. Food Chem. 2006;54(18): 6929-38.

26. Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C, Beckmann M, Haldar S, Beattie JH, Lodge JK, Jenab M, Keun H, Scalbert A. Metabolomics and human nutrition. Br. J. Nutr. 2011:105(8):1277-83.

27. Murray S, Lake BG, Gray S, et al. Effect of cruciferous vegetable consumption on heterocyclic aromatic amine metabolism in man. Carcinogenesis; 2001;22(9): 1413-20.

28. Beckonert O, Keun HC, Ebbels TMD, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat. Protocols 2007;2(11): 2692-703.

29. Veselkov KA, Lindon JC, Ebbels TMD, et al. Recursive Segment-Wise Peak Alignment of Biological 1H NMR Spectra for Improved Metabolic Biomarker Recovery. Anal. Chem. 2008;81(1): 56-66.

30. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ^{1}H NMR Metabonomics. Anal. Chem. 2006;78(13): 4281-90.

31. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. Philosophical magazine. A, Physics of condensed matter, defects and mechanical properties 1901;2(11): 559.

32. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. Chemom. Intell. Lab. Syst. 1998;44(1-2): 175-85.

33. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J. Chemom. 2002;16(3): 119-28.

34. Cloarec O, Dumas ME, Craig A, et al. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. Anal. Chem. 2005;77(5): 1282-89.

35. Fan TWM. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. Prog. Nucl. Mag. Res. Sp. 1996;28(2): 161-219.

36. Sobolev AP, Brosio E, Gianferri R, Segre AL. Metabolic profile of lettuce leaves by high-field NMR spectra. Mag. Res. Chem. 2005;43(8): 625-38.

37. Synge RL, Wood JC. (+)-(S-methyl-L-cysteine S-oxide) in cabbage. Biochem J. 1956;64(2): 252-59.

38. Mitchell SC, Smith RL, Waring RH, Aldington GF. The metabolism of S-methyl-L-cysteine in man. Xenobiotica 1984;14(10): 767-79.

39. Ostermayer F. Products of Acidic Hydrolysis of S-Methyl-L-cysteine Sulfoxide; the Isolation of Methyl Methanethiolsulfonate, and Mechanism of the Hydrolysis. J. Am. Chem. Soc. 1960;82(14): 3752.

40. Meiboom S. Modified spin-echo method for measurement of relaxation times. Rev. Sci. Instr. 1958;29: 688.

41. Loria JP, Rance M, Palmer AG. A Relaxation-Compensated Carr−Purcell−Meiboom−Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. J. Am. Chem. Soc. 1999;121(10): 2331-32.

42. Kim HK, Choi YH, Verpoorte R. NMR-based metabolomic analysis of plants. Nat. Protocols 2010;5(3): 536-49.

43. Waring RH, Harris RM, Steventon GB, Mitchell SC. Degradation to sulphate of S-methyl-L-cysteine sulphoxide and S-carboxymethyl-L-cysteine sulphoxide in man. Drug Metabol. Drug Interact. 2003;19(4): 241-55.

44. Stoewsand GS. Bioactive organosulfur phytochemicals in Brassica oleracea vegetables--A review. Food Chem. Tox. 1995;33(6): 537-43.

45. Sangwan NS, Shanker S, Sangwan RS, Kumar S. Plant-derived products as antimutagens. Phytotherapy Res. 1998;12(6): 389-99.

46. Rose P, Whiteman M, Moore PK, Zhu YZ. Bioactive S-Alk(en)yl Cysteine Sulfoxide Metabolites in the Genus Allium: The Chemistry of Potential Therapeutic Agents. Chem. Inform. 2005;36(35).

47. Gómez-Campo C. Biology of Brassica Coenospecies. 1999.

48. Komatsu W, Miura Y, Yagasaki K. Suppression of hypercholesterolemia in hepatoma-bearing rats by cabbage extract and its component, S-methyl-l-cysteine sulfoxide. Lipids 1998;33(5): 499-503.

49. Kumari K, Mathew BC, Augusti KT. Antidiabetic and hypolipidemic effects of S-methyl cysteine sulfoxide isolated from Allium cepa Linn. Indian. J. Biochem. Biophys. 1995;32(1): 49-54.

50. Kumari K, Augusti KT. Lipid lowering effect of S-methyl cysteine sulfoxide from Allium cepa Linn in high cholesterol diet fed rats. J. Ethnopharm. 2007;109(3): 367-71.

51. Marks HS, Anderson JA, Stoewsand GS. Effect of S-methyl cysteine sulphoxide and its metabolite methyl methane thiosulphinate, both occurring naturally in Brassica vegetables, on mouse genotoxicity. Food Chem. Toxicol. 1993;31(7): 491-5.

^{1}H-NMR spectroscopy combined with multivariate statistical modeling successfully identified 4 singlet spectroscopic peaks found to discriminate a 14 day high cruciferous vegetable (CV) consumption period (500g/day) from 14 day restricted CV dietary interventions in urine samples from 20 healthy Caucasian males, two of which were identified as S-methyl-L-cysteine sulphoxide and *N*-acetyl-S-methyl-L-cysteine sulphoxide. These stable urinary biomarkers of CV consumption will facilitate future assessment of CV in population screening and dietary intervention studies.
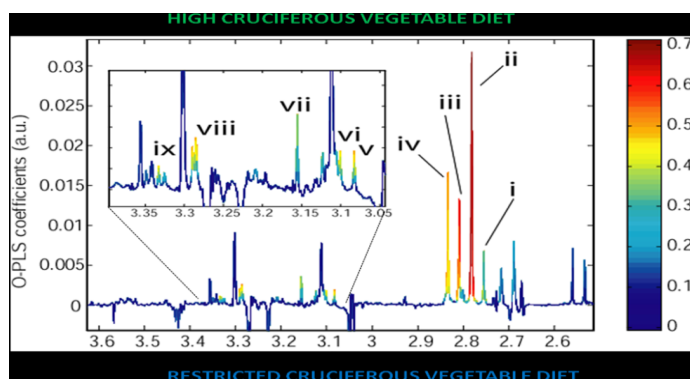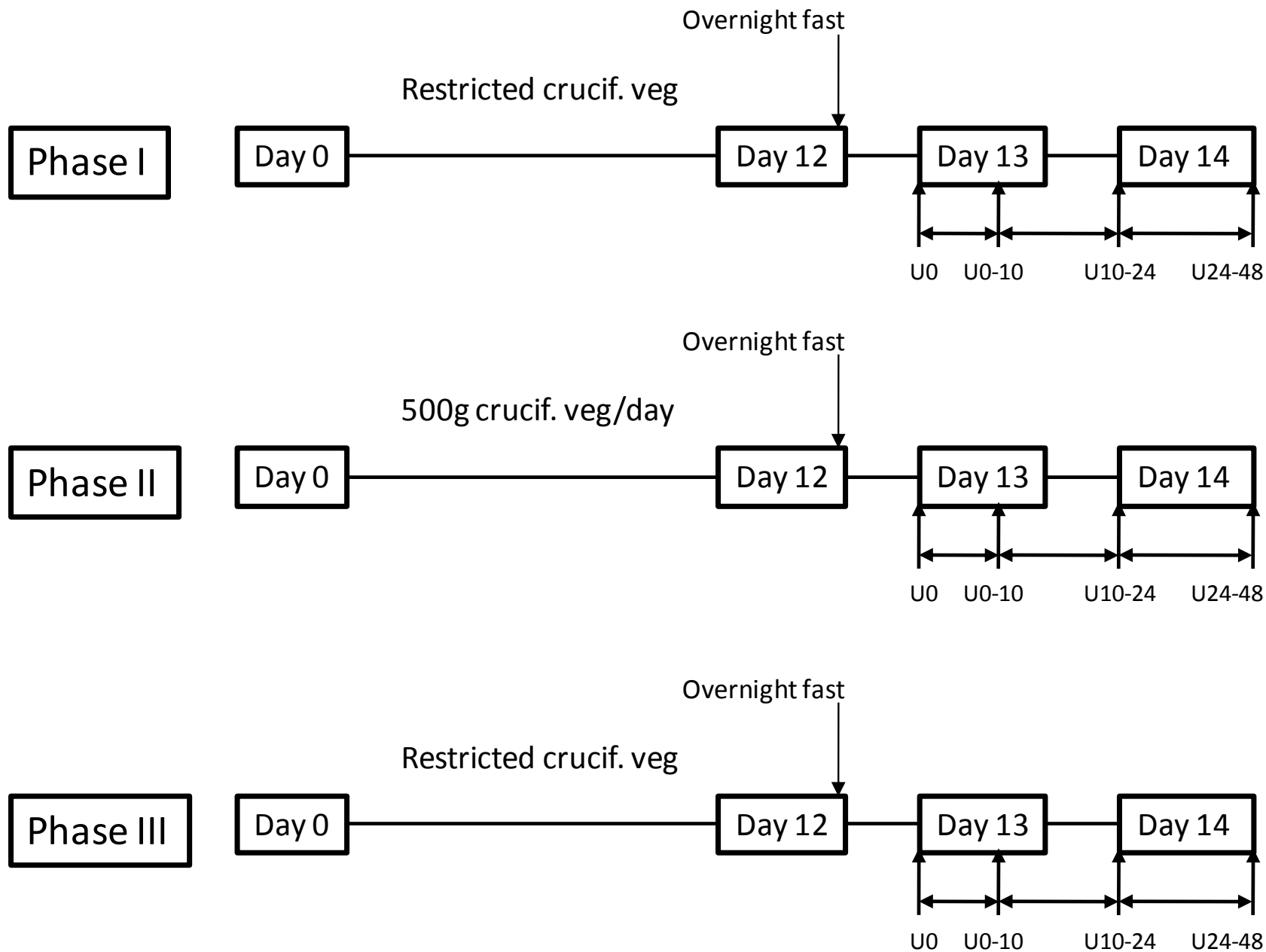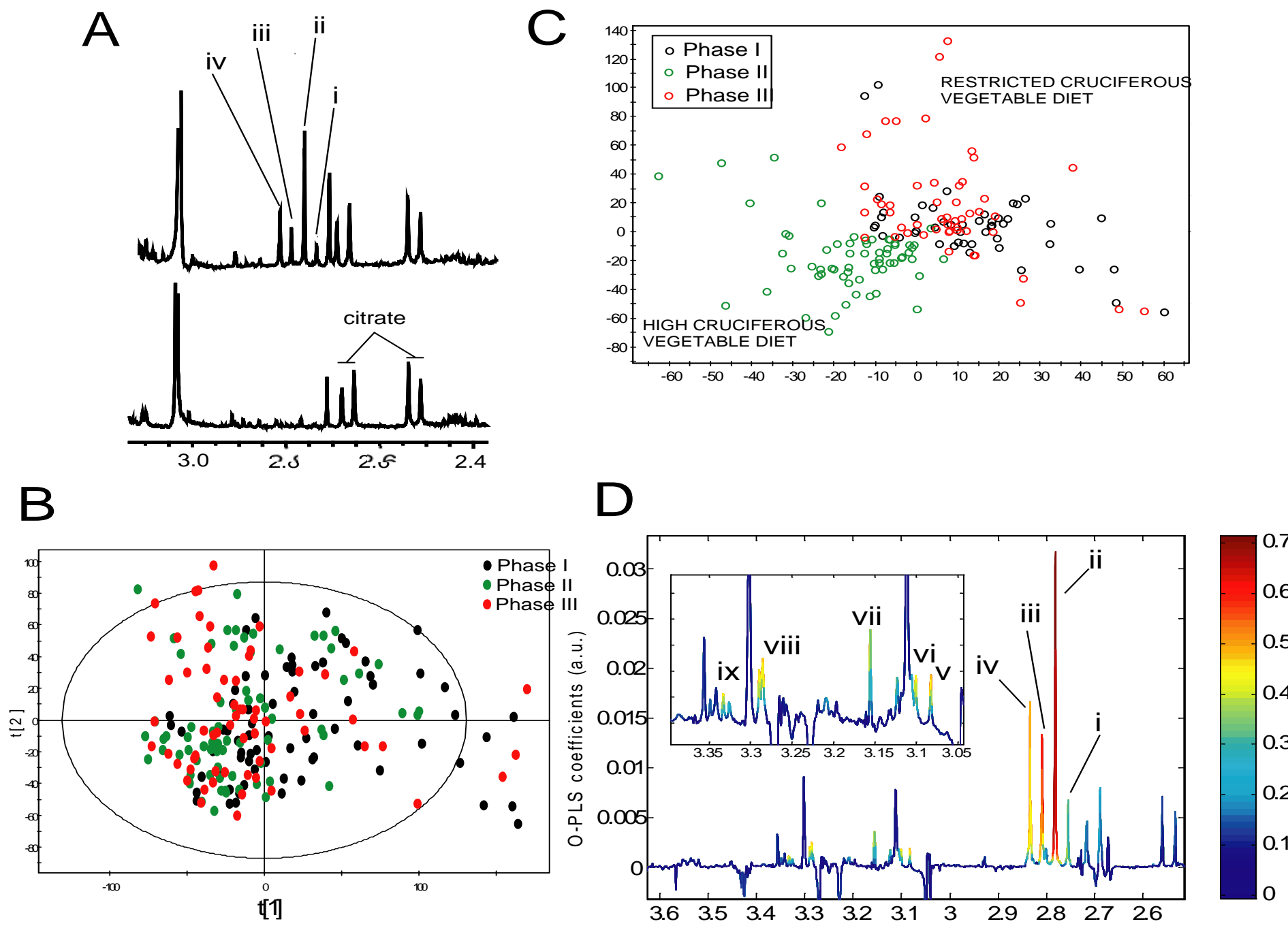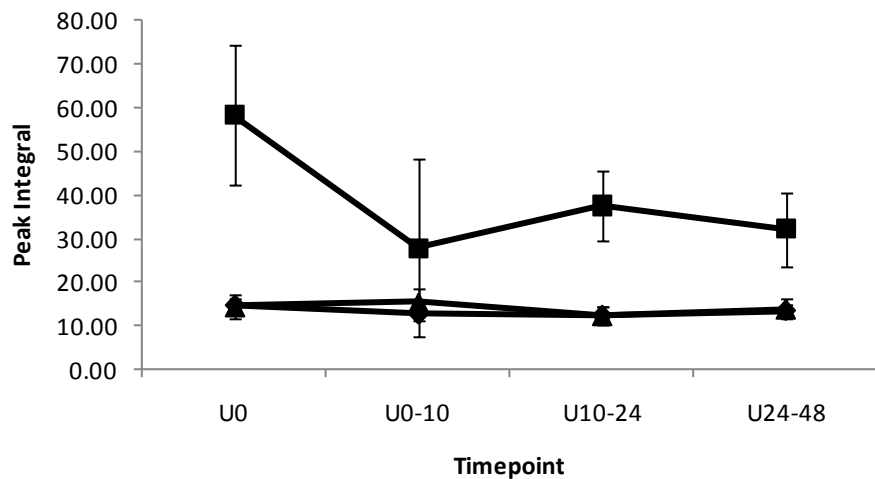
Figure 1

Figure 2

Figure 3

δ2.83

δ2.81

δ2.78

δ2.75

Phase I
Phase II
Phase III

Figure 4

A

δ4.38   δ3.3   δ3.1   δ2.83

R$_1$

H — C — H

ABX spin system

[X]

R$_3$ — C — R$_3$

H

[AB]

δ ^{1}H

B

HUMAN URINE   BROCCOLI EXTRACT

HSQC

A B C D   A

1   1

HMBC

C   A
A B D

2   2
3 4 5

Figure 5

N-acetylation

CH$_3$
|
S=O
|
CH$_2$
|
HN—CH—COOH
|
COCH$_3$

N-acetyl-S-methyl-L-cysteine sulphoxide

30%

CH$_3$
|
S=O
|
CH$_2$
|
H$_2$N—CH—COOH

S-methyl-L-cysteine sulphoxide

Deamination

transaminase

8%

CH$_3$
|
S=O
|
CH$_2$
|
O=CH—COOH

3-methylthiopyruvic acid sulphoxide

Lactic dehydrogenase

CH$_3$
|
S=O
|
CH$_2$
|
HO—CH—COOH

3-methylthiolactic acid sulphoxide

50%

CH$_3$
|
S=O
|
....
|
CH$_2$
|
H$_2$N—CH—COOH

H$_2$N—C—NH$_2$
||
O

urea

SO$_4^{2-}$
CO$_2$

urea formation

12%

Excreted unmetabolised