The Effects of Nutritional Intervention on Metabolic Phenotype

A thesis submitted by
William Matthew Bell Edmands

For the degree of Doctor of Philosophy
Of Imperial College London

Biomolecular Medicine,
Department of Surgery and Cancer,
Imperial College London

2011

Abstract

This research aimed to examine the relationship between dietary interactions, health and disease and the effects of nutritional intervention on metabolic phenotype. High throughput proton nuclear magnetic resonance spectroscopy and liquid chromatography mass spectrometry analytical platforms coupled with multivariate statistical methods were utilised to identify metabolic changes in biological fluids related to nutritional status and intervention. These metabonomic approaches were applied initially to an animal study and subsequent human studies.

The aims of this research were tackled by examination of a preliminary animal study to assess the effects of methyl eugenol, a natural dietary flavour component that is genotoxic, on rat metabolic phenotype by analysis of urinary metabolic profiles. Methyl eugenol alterations of urinary metabotype at relatively low doses (1-50mg/kg bw/day) could not be discriminated, however discrimination based upon age, sex and administration method (orogastric gavage vs. dietary matrix) could be seen.

A second study examined the effects in humans of dietary fat and carbohydrate composition on the development of metabolic syndrome by analysis of urinary and faecal water metabolic profiles. Dietary intervention did not produce statistically significant alterations of human urinary and faecal water metabotype. Furthermore, regression analyses of bacterial population numbers to ^{1}H-NMR spectroscopic data of both urine and faecal water also did not produce statistically significant modelling of the datasets. This study highlighted the need to limit inter-subject variability particularly in human nutritional intervention studies.

The third metabonomic study investigated the effects of cruciferous vegetable, meat meal and caffeine dietary intervention on the human urinary metabolic phenotype by ^{1}H-NMR spectroscopy and UPLC-MS analysis. Chemometric analysis of the ^{1}H-NMR spectral dataset, use of two-dimensional NMR experimentation and synthetic standard urine spiking experiments identified *S*-methyl-L-cysteine-sulphoxide (SMCSO) and *N*-acetylated-SMCSO as urinary biomarkers of cruciferous vegetable consumption in man. Multivariate statistical analysis of the ^{1}H-NMR spectral dataset identified post-prandial biomarkers of meat meal consumption and bacterial co-metabolite excretion in human urine. Targeted interrogation and multivariate statistical modelling of the UPLC-MS database identified 36 tentative biomarkers of cruciferous vegetable, meat and caffeine dietary interventions.

This research demonstrated the successful application of metabonomics to the study of nutritional intervention and the identification of biomarkers of dietary exposure. Also highlighted were the challenges of biomarker identification particularly in UPLC-MS datasets and the need to limit both experimental and inter-subject variations in metabonomic studies of human biofluids.

Statement of originality

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

“Dedicated to my parents without whom none of this would have been possible”

"Yesterday we looked for little bits of a few things in some things; today we look for less of more things in anything; tomorrow we will look for nothing in everything."
- Dr. Francis Gunther

Acknowledgements

I would like to give many thanks to my supervisors Professors Nigel Gooderham and Elaine Holmes for their tremendous support over the years and for their kindness, understanding, patience and making this a thoroughly enjoyable and rewarding experience.

I would also like to thank my post-doctoral supervisors Dr. Jon Swann, Dr. Alison Campbell and Dr. Cinzia Stella for all of their help and for imbuing me with their enthusiasm.

Special thanks go to Dr. Olaf Beckonert for his tireless support as the NMR manager and Dr. Liz Want for her amazing advice and assistance on the Q-TOF.

I would also like to thank the Department of Surgery and Cancer for my generous bursary and the UK Food Standards Agency for their financial assistance.

Contents

List of Figures viii

List of tables xv

Abbreviations xviii

Chapter One – Introduction 1

1.1 Nutrition in health and disease 1

1.2 Diet and Cancer risk factors 3

1.3 Plant phytochemicals- Janus properties 5

1.4 Methods of assessing the relationship between diet and health 7

1.5 Biomarkers of nutritional intervention 10

1.6 Metabonomics 12

1.7 Application of metabonomics to nutritional intervention studies 13

1.8 The human gut microbiota 16

1.9 Aims and Objectives 19

Chapter Two: Analytical Methods 21

2.1 An Introduction to Spectroscopy 21

2.1.1 Biological Nuclear Magnetic Resonance (NMR) Spectroscopy 21
2.1.2 Principles of NMR Spectroscopy 21
2.1.3 Relaxation times 22
2.1.4 Chemical shift 24
2.1.5 Spin coupling 25

2.2 NMR pulse programs 25

2.2.1 Water pre-saturation 26
2.2.2 Standard 1D pre-saturation pulse sequence 26
2.2.3 Carr-Purcell-Meiboom-Gill (CPMG) 26

2.3 Two-dimensional NMR spectroscopy 27

2.3.1 J-Resolved ^{1}H-NMR spectroscopy 27
2.3.2 Correlation spectroscopy (COSY) 28
2.3.3 Total correlation spectroscopy (TOCSY) 28
2.3.4 Heteronuclear Multiple Bond Coherence (HMBC) and Heteronuclear Single Quantum Coherence (HSQC) 2D ^{13}C NMR experiments 28

2.4 Preparation of Biological Samples 29

2.4.1 Flow Injection Urine Sample Preparation 29
2.4.2 Solid Phase Extraction (SPE) 29
2.4.2.1 Solid Phase Extraction method 30

2.5 ^{1}H-NMR Spectroscopy Methodology 30

2.5.1 Flow Injection Experiment 30
2.5.2 Carbon 13 NMR spectroscopy 31
2.5.2.1 Heteronuclear Single Quantum Coherence (HSQC) 31
2.5.2.2 Heteronuclear Multiple Bond Coherence (HMBC) 31
2.5.3 Authentic standard compound spike in for NMR peak assignment 31
2.5.4 Structural assignments 32

2.6 Ultra High Performance Liquid Chromatography (UPLC) 32

2.7.3 Ultra High Performance Liquid Chromatography Methodology 33

2.7 Mass Spectrometry (MS) 34

2.7.1 Application of Mass Spectrometry to Biological Samples 34
2.7.2 Sample preparation for mass spectrometry analysis 36
2.7.3 Mass Spectrometry Methodology 36

Chapter Three: Chemometric Methods 38

3.1 ^{1}H-NMR Spectral Pre-Processing 38

3.1.1 Data Reduction and Pre-Processing 38
3.1.2 Recursive Segment-Wise Peak Alignment 38
3.1.3 Probabilistic Quotient Normalisation 38

3.2 Pattern Recognition 39

3.2.1 Principal Components Analysis (PCA) 39
3.2.2 R^2 and Q^2 values 40
3.2.3 Partial Least Squares Analysis (PLS) 40
3.2.4 Orthogonal Projection on Latent Structures Discriminant Analysis (O-PLS-DA) 40
3.2.5 STOCSY Statistical Total Correlation Spectroscopy 41

3.3 Statistical analysis of mass spectral data 42

3.3.1 MarkerLynx PCA analysis 42

3.3.2 XCMS[1] spectral feature identification 42
3.3.3 Application of XCMS[1] for Tentative Biomarker Identification. 43

Chapter four: The Effects of the Naturally Occurring Food-Derived Genotoxic Carcinogen Methyl Eugenol on Rat Metabolic Phenotype 44

4.1 Background 44

4.2 Aims and Objectives 45

4.3 Introduction 45

4.3.1. Methyl Eugenol-Chemical Properties 45
4.3.2 Toxicological evaluation of Methyl Eugenol 46
4.3.2 ME Metabolism 48
4.3.3 Histopathological, Biochemical and Molecular Biological Findings 51

4.4 Methods 53

4.4.1 Animal husbandry and twenty-eight day study outline 53
4.4.2 Urine sample collection 54
4.4.3 Urine preparation and ^{1}H-NMR spectroscopic analysis 55

4.5 Results 56

4.5.1 Principal Components Analysis (PCA) all samples 57
4.5.1.2 Orthogonal-Partial Least Squares Discriminant Analysis (O-PLS-DA) gender related metabolic differences 60
4.5.2 The effect of high dosage methyl eugenol dietary intervention 62
4.5.2.1 Pairwise PCA and PLS-DA high ME dietary exposure vs. vehicle controls 62
4.5.2.2 O-PLS-DA male rats day 28 high ME vs. Day 28 vehicle controls 65
4.5.3 Time-course effects of methyl eugenol dietary intervention 69
4.5.3.1 Pairwise timecourse PCA and PLS-DA high ME dietary exposure dose group 69
4.5.3.2 Pairwise timecourse O-PLS-DA high ME dietary exposure dose group 74
4.5.4 Gender related effects on methyl eugenol metabolism 78
4.5.4.1 Pairwise PCA and PLS-DA 78
4.5.4.2 Pairwise O-PLS-DA 80
4.5.5 Effect of methyl eugenol administration method on urinary metabolic phenotype 82
4.5.5.1 Pairwise PCA and PLS-DA high ME exposure dietary matrix vs. orogastric gavage 82
4.5.5.2 O-PLS-DA high ME exposure dietary matrix vs. orogastric gavage 84

4.6 Discussion 86

4.6.1 The effect of high dosage methyl eugenol dietary intervention 87
4.6.2 Gender related effects on methyl eugenol metabolism 90
4.6.3 Time-course effects of methyl eugenol dietary intervention 92
4.6.4 Effect of methyl eugenol administration method on urinary metabolic phenotype 94
4.6.5 Conclusion 97

Chapter Five: Changing Dietary Fat Composition and Carbohydrate Glycaemic Index in Subjects at Risk of Developing the Metabolic Syndrome – Mapping Metabolic Change Through Metabonomics 98

5.1 Background 98

5.2 Aims and Objectives 98

5.3 Introduction 99

5.3.1 Metabolic syndrome 99
5.3.2 Role of the Gut microbiota in Fatty Acid and Carbohydrate Metabolism 100
5.3.3 Application of Metabonomics to Human Studies 102
5.3.4 Reduction of Inter-subject Variability in Nutritional Intervention Studies 103

5.4 Methods 105

5.4.1 Study Design and Protocol 105
5.4.2 Diagnostic recruitment criteria 108
5.4.3 Faecal Water Sample Preparation 109
5.4.3 Faecal Water ^{1}H-NMR Spectroscopy acquisition 109
5.4.4 Urine ^{1}H-NMR Spectroscopy acquisition 110
5.4.5 Clinical markers and food diary analyses of study participants 111

5.5 Results 112

5.5.1 PCA analysis of urine and faecal water 114
5.5.2 PCA, PLS-DA and O-PLS-DA pair-wise comparisons of control and post dietary intervention urine and faecal water samples 116
5.5.3 O-PLS-DA comparisons of inter-dietary variation 124
5.5.4 Bacterial Regression analyses to dietary intervention ^{1}H-NMR spectra 127
5.5.5 Regression analyses of significant bacterial population changes to all ^{1}H-NMR spectra 133
5.5.6 Regression analyses of significant bacterial population changes to all ^{1}H-NMR spectra of urine and faecal water 135

5.6 Discussion 139

5.6.1 Correlation of clinical marker changes by regression analysis 140
5.6.2 Gut Microbial Population findings 143
5.6.3 Suitability of the RISCK study for Metabonomics..... 145
5.6.4 Conclusions 147

Chapter Six: The Effects of High Cruciferous Vegetable Consumption on Human Metabolic Phenotypes..... 149

6.1 Background 149

6.2 Aims and Objectives..... 149

6.3 Introduction..... 149

6.3.1 Cruciferous Vegetables, phytochemicals and cancer 149
6.3.2 Chemoprotective Mechanism..... 154
6.3.3 Carcinogenic heterocyclic amines in cooked meat..... 156
6.3.4 ITC Elimination..... 157
6.3.5 Evaluation of previous studies of anti-carcinogenic potential of CVs 159

6.4 Methods 160

6.4.1 Study Protocol 160
6.4.2 ^{1}H-NMR Spectroscopy and UPLC-MS Methods..... 162
6.4.3 Application of XCMS for Tentative Biomarker Identification. 162
6.4.4 Mercapturic acid urinary extraction..... 163
6.4.5 Synthesis of sulforaphane mercapturate 163
6.4.6 Brussels sprouts and broccoli aqueous extraction 164
6.4.7 Synthesis of S-methyl-L-cysteine sulphoxide and N-acetyl-S-methyl-L-cysteine-sulphoxide standards..... 166

6.5 Results..... 167

6.5.1 Effect of a cruciferous vegetable dietary intervention in humans, on the urinary metabolic profile. 167
6.5.1.1 Principal Components Analysis all ^{1}H-NMR spectra 168
6.5.1.2 O-PLS-DA..... 170
6.5.1.3 STOCSY..... 172
6.5.2.1 Mercapturate synthesis and extraction 173
6.5.2.2 Authentic Standard spiking experiments..... 175
6.5.2.3 Improved Peak Alignment and Normalisation 176
6.5.2.4 ^{1}H–^{1}H *J*-resolved spectra and COSY of Phase II urine 179
6.5.2.5 ^{13}C HSQC and HMBC of solid phase extract of Phase II urine ... 184
6.5.2.6 Cruciferous vegetable extraction: Analysis by 2D NMR experiment 187
6.5.3 Effect of meat consumption on human urinary metabolic profile. 200
6.5.3.1 PCA and PLS-DA analysis..... 200
6.5.3.2 Time course pair wise O-PLS-DA..... 203
6.5.3.3 Effect of cruciferous vegetable dietary intervention on meat and caffeine alterations of metabotype. 207

6.5.4 UPLC-MS biomarker identification 209
6.5.4.1 MarkerLynx PCA analysis 209
6.5.4.2 Masslynx targeted analysis 210
6.5.4.3 XCMS[1] statistical analysis 215
6.5.4.4 MS/MS fragmentation analysis of XCMS chromatographic features 221
6.5.4.5 XCMS[2] analysis 229

6.6 Discussion 231

6.6.1 S-Methyl-L-cysteine sulphoxide 231
6.6.2 Biosynthesis of SMCSO and role within the plant 232
6.6.3 Experimental studies of SMCSO and potential biological roles within humans. 233
6.6.5. Effect of meat consumption on human urinary metabolic profile. 241
6.6.6. Effect of cruciferous vegetable dietary intervention on meat and caffeine alterations of metabotype. 243
6.6.4 UPLC Mass Spectral Metabonomic analysis 245

6.7 Conclusions 248

Chapter Seven: Overall Conclusions and Future Work 250

7.1 The application of metabonomics to examine the effects of ME dietary intervention in rats 250

7.2 Application of the metabonomic approach to understanding human dietary intervention with varying composition of dietary fat and glycaemic index of carbohydrates in subjects at risk of developing metabolic syndrome 252

7.3 Metabonomic analysis of the effects of CV dietary intervention on the human metabolic phenotype 254

7.4 The future potential of metabonomics for nutrition studies 256

7.5 Summary 258

References 260

List of Figures

Fig. 1.1 Flowchart of thesis structure..... 20

Fig. 2.1 Bulk magnetisation vector (Mz) 22

Fig. 2.2 Effect of 90° pulse tilting magnetisation into the x-y plane..... 23

Fig. 2.3 T_1 spin relaxation 23

Fig. 2.4 Spin-spin (T_2) relaxation, time taken for all protons in the observed sample to decay in the x-y plane to the applied magnetic field B_0..... 24

Fig. 2.5 Schematic of a typical 2D NMR pulse sequence..... 27

Fig. 4.1 Molecular structures of ME, estragole and safrole. 46

Fig. 4.2 Metabolic scheme of ME in rodent liver**.** The primary route of ME excretion at low dosage is by *O*-demethylation and suphation or glucuronic acid conjugation of either eugenol or 5-allyl-2-methoxy phenol (pathway 1). Metabolic scheme adapted from (Smith, Adams et al. 2002) 50

Fig. 4.3 Representative 600 MHz ^{1}H NMR spectra of urine from a female Fischer 344/NH/sd strain rat..... 56

Fig. 4.4 PCA scores plot showing *t*[1]/*t*[2] principal components revealing an overview of the ^{1}H NMR spectra of rat urine samples from all dosage groups and dosage regimens showing no clustering of samples based on class. (Q^2:0.43 R^2:0.56)..... 57

Fig. 4.5 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.43 R^2: 0.56) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.28 R^2: 0.32) from the ^{1}H-NMR spectra of all rat urine samples showing clustering of samples based upon gender. Unit variance scaled. 59

Fig. 4.6 O-PLS-DA loadings plot showing urinary metabolic differences between male and female F344 strain rats all samples. 61

Fig. 4.7 PCA scores plot showing *t*[1]/*t*[2] principal components of urinary ^{1}H-NMR spectral data from the vehicle control diet and 50mg/kg bw/day ME dosage regimen samples in all time-points..... 62

Fig. 4.8 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2:0.36 R^2:0.73) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2:0.79 R^2:0.94) from the ^{1}H-NMR spectra of vehicle control diet and 50mg/kg bw/day ME dosage groups day 28 rat urine samples showing clustering of samples based on ME exposure. Unit variance scaled. 64

Fig. 4.9 O-PLS-DA loadings plot showing male day 28 vehicle control diet vs. 50mg/kg bw/day dietary dose group. ($Q^2\hat{Y}$: 0.21 and an $R^2\hat{Y}$:0.67). Normalised to total area, Unit Variance scaled. 65

Fig. 4.10 STOCSY analysis of singlet peak δ3.80 reveals correlation with creatine. This singlet peak was identified as guanidoacetate. 66

Fig. 4.11 STOCSY analysis of unknown triplet peak δ3.86 reveals strong correlation with the other unknown peaks identified by the O-PLS-DA analysis and also with *p*-cresol glucuronide. 68

Fig. 4.12 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.4 R^2: 0.59) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.62 R^2: 0.89) from the ^{1}H-NMR spectra of the 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based on the day 7 and day 28 timepoints. Unit Variance scaled. 71

Fig. 4.13 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.37 R^2: 0.46) and (B) *t*[2]/*t*[3] PLS-DA scores (Q^2: 0.21 R^2: 0.51) from the ^{1}H-NMR spectra of the 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based upon time-point. Unit Variance scaled. 72

Fig. 4.14 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.36 R^2: 0.53) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.16 R^2: 0.51) from the ^{1}H-NMR spectra of the dietary matrix control group of male rat urine samples some clustering of samples based upon time-point can be seen but there is also evidence of inter-subject variations. Unit Variance scaled. 73

Fig. 4.15 Aliphatic region of O-PLS-DA loadings plot demonstrating urinary metabolic differences between day 7 (7 weeks old) and day 28 (10 weeks old) male F344 strain rats of the 50mg/kg bw/day dose group. 75

Fig. 4.16 Aromatic region of O-PLS-DA loadings plot showing urinary metabolic differences between day 7 (7 weeks old) and day 28 (10 weeks old) male F344 strain rats of the 50mg/kg bw/day dose group. 76

Fig. 4.17 STOCSY analysis showing correlation of the unknown triplet peak (δ3.16) with taurine and another unknown peak (δ3.96). 77

Fig. 4.18 Scores maps showing (A) t[1]/sample no. PCA scores (Q^2: 0.46 R^2: 0.81) and (B) t[1]/t[2] PLS-DA scores (Q^2: 0.37 R^2: 0.73) from the 1H-NMR spectra of the 50mg/kg bw/day ME dietary intervention animals contrasting gender in order to determine any gender specific effects of ME exposure in the F344 rat strain. Clustering of samples based upon gender can be seen in the PLS-DA scores. Unit Variance scaled. 79

Fig. 4.19 O-PLS-DA loadings coefficients plot showing evidence of metabolic sexual dimorphism in the F344 rat strain. Robust modelling of the data was achieved suggesting the majority of urinary metabolic variation is gender specific in the first aligned O-PLS component ($Q^2\hat{Y}$: 0.93; $R^2\hat{Y}$: 0.98)..... 81

Fig. 4.20 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.44 R^2: 0.52) and (B) *t*[2]/*t*[3] PLS-DA scores (Q^2: 0.43 R^2: 0.56) from the ^{1}H-NMR spectra of the 50mg/kg bw/day gavage treatment and 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based upon treatment method in the PLS-DA model only. Unit Variance scaled. 83

Fig. 4.21 O-PLS-DA loadings coefficient plot showing (A) aliphatic region demonstrating urinary metabolic differences between 50mg/kg bw/day ME dietary matrix (downwards pointing peaks) and 50mg/kg bw/day ME gavage treatment (upwards pointing) male F344 strain rats. (B) aromatic region of the spectrum. Metabolic differences can be seen attributable to the orogastric gavage treatment. ($Q^2\hat{Y}$: 0.66 and an $R^2\hat{Y}$:0.96)..... 85

Fig. 5.1 Schematic showing study outline of the RISCK study dietary intervention ... 107

Fig. 5.2 Schematic showing the RISCK study diagnostic recruitment criteria. Intial targeted subject recruitment was based on an inclusion criteria. 108

Fig. 5.3 Typical 600 MHz ^{1}H-NMR spectra of Faecal Water obtained from a RISCK study participant..... 112

Fig. 5.4 Human urine PCA scores scatter plot of ^{1}H-NMR spectra from all subjects post dietary intervention (24 weeks). 115

Fig. 5.5 Scores maps showing *t*[1]/*t*[2] PCA scores from the ^{1}H-NMR spectra baseline vs. treatment urine samples of diets A (Q^2:0.017; R^2:0.26), B (Q^2:0.043; R^2:0.28), C (Q^2:0.092 ; R^2:0.25), D (Q^2:0.015; R^2:0.23) and E (Q^2:0.01; R^2:0.24). Models showed poor explanation of variance in the X matrix. Unit variance scaled..... 117

Fig. 5.6 Scores maps showing *t*[1]/*t*[2] PLS-DA scores from the ^{1}H-NMR spectra baseline vs. treatment urine samples of diets A (Q^2:-0.027; R^2:0.63), B (Q^2:0.17; R^2:0.63), C (Q^2:0.034; R^2:0.68), D (Q^2:-0.12; R^2:0.79) and E (Q^2:-0.056; R^2:0.701). Models showed poor discrimination of class indicating sources of variance in the dataset unrelated to dietary intervention. Unit variance scaled. 118

Fig. 5. 7 O-PLS-DA coefficient loadings plot showing dietary intervention B baseline (0 week) vs. Treatment (24 week) of urinary ^{1}H-NMR spectra normalised to total area. 121

Fig. 5.8 Model overview plots of dietary interventions A, B, C, D and E, showing bacterial regression analyses of total bacterial population changes (calculated by subtraction of treatment from baseline bacterial counts) to ^{1}H-NMR data of human urine. 131

Fig. 5.9 Example O-PLS-DA loadings plot of bacterial regression analysis of faecal water. Total bacterial population change of *E. rectale/C. coccoides gp* .within the Diet D intervention regressed to the ^{1}H-NMR spectra of faecal water. 132

Fig. 5.10 Regression analysis of faecal acetate treatment samples concentration values to TSP normalised faecal water ^{1}H-NMR spectra. Strong correlations can be seen between acetate concentrations measured by capture compound mass spectrometry (CCMS) by our collaborators at the University of Reading and faecal water ^{1}H-NMR signals corresponding to both acetate and butyrate (Q^2:0.41; R^2:0.77). 136

Fig. 6.1 Hydrolysis of glucosinolates by thioglucoside enzymes (myrosinases) to their isothiocyanate derivatives. 150

Fig.6.2 Schematic of metabolism of sinigrin the most abundant glucosinolate of Brussels sprouts, a typical aliphatic glucosinolate. Adapted from (Holst and Williamson 2004). 153

Fig. 6.3 Schematic of metabolism of indolic glucosinolates showing potential metabolic endpoints of hydrolysis. Variations are seen in the position of a methyl group in neoglucobrassicin and 4-methoxyglucobrassicin. Adapted from (Holst and Williamson 2004). 154

Fig. 6.4. The metabolic elimination of ITC showing the excreted dithiocarbamate (Mercapturic acid)-ITC conjugates..... 157

Fig. 6.5 Figure showing the study protocol and the nutritional intervention strategy. . 161

Fig. 6.6 Representative 600 MHz ^{1}H NMR spectra of human urine from restricted CV (bottom: Phase I pre-dose) and high CV diets (top: Phase II pre-dose). 167

Fig. 6.7 PCA scores plot demonstrating that variation within the NMR spectra X matrix multidimensional space is dominated by intersubject variation with no clear grouping based on sample class. Q^2: 1.3 relatively low cross validation of the model for biological matrices and the R^2: 0.32 shows relatively poor explanation of variance in the X matrix by this model. Outliers were removed prior to further chemometric analyses. 169

Fig. 6.8 O-PLS-DA coefficient plot; phase I predose (low CV) vs. Phase II predose (high CV) human urine samples showing the singlet resonances characteristic of Phase II. ($Q^2\hat{Y}$: 0.46; $R^2\hat{Y}$: 0.72). 171

Fig. 6.9 STOCSY analysis of peak 2.78ppm normalised to total area showing correlation between the singlet peaks of interest. Arrows indicates peak from which the analysis was driven from..... 172

Fig. 6.10 400MHz NMR spectrum of synthetic sulphoraphane-mercapturate conjugate. 174

Fig. 6.11 Following RSPA and normalisation to probabilistic quotient, O-PLS-DA identified a previously poorly aligned singlet peak at δ2.83 additionally characteristic of CV consumption. A: δ2.83; B: δ2.81; C: δ2.78; D: δ2.75. 177

Fig. 6.12 STOCSY analysis confirms statistical correlations between the four unknown singlets and with extra connectivity with peaks between δ3-3.5..... 178

Fig. 6.13 400MHz 2D ^{1}H-^{1}H *J*-Res NMR experiment. Deconvolution of the NMR spectral features identified as characteristic of CV consumption between δ3-3.50 provides evidence of an ABX spin system indicative of an unknown chiral metabolite. 181

Fig. 6.14 Comparison of the *J*-couplings demonstrates shared scalar coupling constants between the protons of the ABX spin system of the unknown metabolite diagnostic of CV consumption. 182

Fig. 6.15 2D COSY confirmation of the ABX spin system identified by *J*-res spectroscopy. 183

Fig. 6.16 Top: ^{1}H-^{13}C HSQC of 5% methanol SPE fraction of Phase II predose urine, the 4 singlet peaks share a similar ^{13}C shift of around δ39.6-40.2 suggestive of the chemically isolated S-methyl-sulphoxide group. 186

Fig. 6.17 Top: Aliphatic region of CPMG fresh Brocolli acetonitrile extraction. 190

Fig. 6.18 2D-TOCSY spectrum of fresh broccoli aqueous acetonitrile extraction confirming the presence of chlorogenic acid. 191

Fig. 6.19 Top: ^{1}H-^{13}C HSQC of fresh broccoli acetonitrile extraction; Bottom: ^{1}H-^{13}C HMBC of fresh broccoli acetonitrile extraction. 194

Fig. 6.20 400MHz 1D-NOESY of SMCSO synthetic standard spiked into Phase I predose urine. 197

Fig. 6.21 400MHz 1D-NOESY of NAc-SMCSO synthetic standard..... 198

Fig. 6.22 Top right: 400MHz ^{1}H-^{13}C HSQC of NAc-SMCSO synthetic standard; 199

Fig. 6.23 Scores maps showing *t*[1]/*t*[2] PCA scores (Q^2: 0.37 R^2: 0.46) from the ^{1}H-NMR spectra (Unit Variance scaled) of the phase I study dataset showing clustering of samples based upon time-point..... 201

Fig. 6.24 Scores maps showing *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.41 R^2: 0.49) from the ^{1}H-NMR spectra (Unit Variance scaled) of the phase I study dataset showing clustering of samples based upon time-point. 202

Fig. 6.25 Effect of red meat consumption on urinary metabolite profile..... 204

Fig. 6.26 Effect of meat consumption on urinary metabolite profile. 205

Fig. 6.27 STOCSY analysis of the triplet peak of phenylacetylglutamine δ7.36 demonstrating statistical relationships of bacterial co-metabolites in the phenolic region of the spectrum identified in the time-course O-PLS-DA. 206

Fig. 6.28 Expanded O-PLS-DA showing the characteristic splitting pattern of mannitol characteristic of the Phase I 0-10hrs time point. 208

Fig. 6.29 Markerlynx PCA scores plot of positive mode UPLC-MS data. No underlying clustering of subjects within the data can be seen when observing variance in the X matrix of the 0-6mins retention time subset. There is evidence of a cluster of outlying samples (ringed) which represent a range of different subjects and sample classes. ... 209

Fig. 6.30 Raw UPLC-MS total ion chromatograms (TICs) in positive mode for subject 12 for both phase I and II predose urine samples. 213

Fig. 6.31 Top: Semi-quantification of the relative integrals of 263.1*m/z* tentatively identified as allyl-isothiocyanate-mercapturate a known biomarker of CV consumption. Figure demonstrates a time-course excretory profile adjusted for urinary volume..... 214

Fig. 6.32 Example of an Extracted Ion Chromatograms (EIC) (A) and box and whisker plot (B) of significant feature; *m/z 449.1* retention time 173 seconds;..... 218

Fig. 6.33 Top: MS/MS negative mode fragmentation analysis of XCMS[1] significant feature 151.1*m/z* in human urine associated with CV consumption..... 224

Fig. 6.34 Metabolic schematic demonstrating enzymatic hydrolysis of SMCSO by cysteine lyase following CV tissue disruption. 233

Fig. 6.35 Metabolic scheme showing the human metabolism of SMCSO. 240

List of tables

Table 2.1 Reversed phase gradient conditions for UPLC of Human Urine samples. A=H_2O (0.1% formic acid), B=acetonitrile (0.1% formic acid). 33

Table 2.2 Summarising LCT Premier^{TM} parameters for both positive and negative ion modes. 37

Table 4.1 28 day ME F344 Rat Study animal numbers, treatment groups and dosage regimen..... 53

Table 5.1 Table showing RISCK study dietary interventions performed in human participants. (SFA: saturated fatty acid; MUFA: monounsaturated fatty acid; PUFA: polyunsaturated fatty acid; CHO: carbohydrate; GI: glycaemic index.) 106

Table 5.2 ^{1}H-NMR chemical shifts (δ) and multiplicity resonances observed from faecal water..... 113

Table 5.3 Summary of baseline vs. treatment O-PLS-DA models of urine and faecal water comparing effects of pre-processing on $Q^2\hat{Y}$ and $R^2\hat{Y}$ values. 122

Table 5.4 Table showing O-PLS-DA model strengths of experimental iterations of interest in urine and faecal water. 125

Table 5.5 Table showing change in bacterial population numbers (Log_{10} bacterial cells/g faeces) between baseline and treatment time-points and paired t-test values. Numbers highlighted in red represent statistically significant changes in bacterial populations attributable to diet according to *p*-statistics. (Bacterial population numbers enumerated by FISH analysis; provided by our collaborators at the University of Reading). 128

Table 5.6 Regression analyses of all treatment (24 week) urine and faecal water ^{1}H-NMR spectra to statistically significant bacterial population changes (assessed by *p*-statistics). ……..133

Table 5.7 Table showing statistically significant clinical markers (according to *p*-statistics) related to dietary intervention between baseline and treatment time-points. (Clinical measures provided by our collaborators at the University of Reading)…….136

Table 5.8 Table showing regression analyses of statistically significant clinical markers (according to *p*-statistics) to all treatment (week 24) urine and faecal water ^{1}H-NMR spectra. Models were characterised by poor strength reflected in low $Q^2\hat{Y}$ and high $R^2\hat{Y}$ values. Indicative biologically plausible metabolite associations were observed but could not be considered diagnostic……137

Table 6. 1 Total glucosinolate content in Brussels sprouts and Broccoli. Table adapted from (Murray S. et al. 2001)..... 151

Table 6.2 showing CV aqueous extractions experimental iterations..... 165

Table 6.3 Chemical shifts of potential sulphur/nitrogen containing compounds with biological relevance and a singlet splitting pattern..... 175

Table 6.4 Summary of ^{13}C shifts for both the HSQC and HMBC experiments..... 185

Table 6.5 Assignment of aqueous CV extraction metabolites confirmed by comparison of ^{13}C shifts with literature sources..... 189

Table. 6.6 Candidate metabolites of CV consumption generated by consideration of biological plausibility ranked by molecular weight including expected *m/z* of parent ion in both positive and negative modes..... 211

Table. 6.7 Experimental iterations addressing primary study aims. All pair-wise comparisons were examined using the XCMS software on both the positive and negative UPLC-MS datasets. 215

Table. 6.8 Effect of CV consumption on metabotype. Phase I predose vs. Phase II predose urine samples experimental iteration. 217

Table. 6.9 Showing statistically significant Extracted Ion Chromatograms (EICs) generated by the XCMS[1] algorithm from the UPLC-MS data in both positive (10) and negative modes (11) which discriminate the high CV from the restricted CV dietary phases. 220

Table. 6.10 Major MS/MS fragments of urinary XCMS[1] features discriminatory of meat and caffeine intervention in positive mode (10-50V voltage ramp). 222

Table. 6.11 Standard compounds subjected to MS/MS fragmentation analysis (10-50V voltage ramp) in positive mode. 226

Table. 6.12 Further candidate compounds of biological plausibility subjected to positive mode MS/MS fragmentation analysis (10-50V voltage ramp). 228

Table. 6.13 Subset of XCMS[2] similarity search report from the experimental iteration of Phase I predose vs. Phase II predose showing the two putative biomarkers of 277.1 *m/z* and 263.1*m/z*. 230

Abbreviations

δ-	Chemical shift unit
µM-	microMoles
1D-	One-dimensional
2D-	Two-dimensional
Ahr-	Aryl hydrocarbon receptor
AITC-	Allyl isothiocyanate
ARE-	Antioxidant response element
CH_3CN/ACN-	Acetonitrile
COSY-	Correlation spectroscopy
CPMG-	Carr-Purcell-Meiboom-Gill
CV –	Cruciferous Vegetable
CVD-	Cardiovascular disease
D_2O-	Deuterium oxide ^2H_2O
dd-	doublet of doublets
DIM-	Di-indolylmethane
EpRE-	Electrophile response element
ESI-	Electrospray ionisation
FID-	Free induction decay
FT-	Fourier transform
GI-	Glycaemic index
GPC-	Glycerophosphocholine
GSH-	Glutathione
GST-	Glutathione-S-transferase
h.p.d.-	hours post dose
HDL-	High density lipoprotein
HMBC-	Heteronuclear Multiple Bond Coherence
n-HPPA-	n-hydroxyphenylpropionic acid
HSQC-	Heteronuclear Single Quantum Coherence
I3C-	Indole-3-carbinol
ITC –	Isothiocyanate
KeapI-	Kelch-like ECH- associated protein I
LAB –	Lactic acid bacteria

LADME-	Liberation, absorption, distribution, metabolism and excretion
LDL-	Low density lipoprotein
m-	multiplet
Maf-	Musculoaponeurotic fibrosarcoma oncogene
MDH –	Mannitol-2-dehydrogenase
mg-	milligrams
mg/kg bw/day-	milligrams/kilogram body weight/day
MHz-	Megahertz
mM-	milliMoles
MMTSI-	Methane methylthiosulphinate
MMTSO-	Methane methylthiosulphenate
MS-	Mass spectrometry
MSI-	Methane sulfinate
MS-	Metabolic syndrome
MSO-	Methane sulfonate
MUFA-	Mono-unsaturated fatty acid
NMR-	Nuclear magnetic resonance
NQO1-	NAD(P)H : quinone oxidoreductase 1
Nrf2-	NF-E2 p45-related factor 2
O-PLS-DA-	Orthogonal projection to latent structures discriminant analysis
PAG-	Phenylacetylglycine
PCA-	Principal components analysis
PLS-	Partial least squares/ projection to latent structures
PLS-DA-	Partial least squares discriminant analysis
ppm-	Parts per million
PUFA-	Poly-unsaturated fatty acids
q-	quartet
RISCK-	**R**eading University, **I**mperial College London, **S**urrey University, MRC Human Nutrition Research **C**ambridge and **K**ings College London.
RSPA-	Recursive segment-wise peak alignment
SCFA-	Short chain fatty acid
SFA-	Saturated fatty acid
SMCS-	S-methyl-L-cysteine

SMCSO-	*S*-Methyl cysteine sulfoxide
STOCSY-	Statistical total correlation spectroscopy
t-	triplet
TMAO-	Trimethylamine-N-oxide
TOCSY-	Total correlation spectroscopy
TSP-	3-(Trimethylsilyl)- Propionic acid-D4, sodium salt
UGT-	Uridine 5’ diphospho-glucuronosyltransferase
UPLC-	Ultra performance liquid chromatography^{tm} (Waters)
UPLC-MS -	Ultra performance liquid chromatography- Mass spectrometry
VLDL-	Very low density lipoprotein
XRE-	Xenobiotic response elements

Chapter One – Introduction

1.1 Nutrition in health and disease

Diet and nutrition are known to be very important factors in the prevention of many chronic disease conditions and the maintenance of good health. The nature of the links and mechanisms between diet and health and as determinants of chronic disease are of prominent scientific and public health interest. Chronic diseases such as obesity, diabetes, cardiovascular diseases, cancer and many others are deemed preventable or at least ameliorated by alteration of diet and/or lifestyle. Chronic diseases are largely deemed preventable; however more research is required to elucidate the mechanisms which link diet to health and a reduction in disease incidence and the dietary components which contribute both detrimentally and beneficially to health. In the interests of the public, the favoured course of action is prevention by dietary intervention, which is seen as being more cost-effective and sustainable than the use of medical treatments.

Although age, sex and genetic factors can determine the incidence of chronic disease, many risks are due to environmental and lifestyle choices such as diet, physical inactivity, smoking, alcohol consumption and the impact of environmental toxins through inappropriate food choices. Also biological factors such as dyslipidaemia, hypertension, obesity and hyperinsulinaemia impact on chronic disease incidence. Traditional plant-based diets around the world have largely been replaced by processed high-fat, energy dense diets with substantial content of animal products. It is known that diets rich in fruits, vegetables and plant materials can prevent the onset of many chronic diseases in man, particularly cancer [1-4]. The mechanisms whereby bioactive plant components offer protection are a major focus for understanding the relationship between diet and disease.

The incidence of chronic diseases related to diet is rapidly increasing, accounting for 60% of the 56.5 million total reported deaths worldwide in 2001 [5]. Almost half of these are associated with cardiovascular disease (CVD), obesity and type II diabetes. Estimates suggest around half the women and a third of the men in UK are clinically obese (http://www.food.gov.uk/). Furthermore around 25% of the UK population demonstrate symptoms of metabolic syndrome (MS) [6, 7]. MS is a condition

characterised by an increased risk of coronary heart disease (CHD), type II diabetes and many forms of cancer. Symptoms of this metabolic condition are dyslipidaemia, hypertension, insulin resistance and increased bodily fat ratios. There is evidence that the amount and type of dietary fats and carbohydrates in the diet directly influence blood lipid profiles, insulin sensitivity, endothelial function and blood pressure [8, 9]. The quality and quantity of carbohydrates in the diet, particularly the glycaemic index (GI), are also known to have an impact on the development of many chronic diseases. The GI is the ranking system (low 0-100 high) of carbohydrates based upon how rapidly they are converted to simple monosaccharides and absorbed into the blood stream. High GI carbohydrates such as refined simple sugars typical of industrialised diets cause spikes and instabilities in the blood sugar level, leading to hyperinsulinaemia, type II diabetes and there is also clear evidence of increased risk of CVD [10, 11].

A diet containing substantial quantities of fruit, vegetable and fibre has been shown to be protective against many forms of cancer and conversely a diet rich in animal products causative of cancers such as colon, bowel, lung and prostate [12-17] and also the development of CVD [18, 19].

The relationship between diet and disease and particularly the chemoprotective properties of many compounds derived from fruits, vegetables and plants commonly consumed in the human diet are subject to substantial research focus.

1.2 Diet and Cancer risk factors

Humans are exposed to a vast array of environmental chemicals and behaviours that can affect the incidence of cancer, both in a beneficial and detrimental way.

The major environmental causes of cancer are thought to be smoking, dietary (carcinogenic food components, macro- and micronutrient imbalances), chronic bacterial and viral infections such as HPV (human papillomavirus) in cervical cancer and also hormonal factors [20]. The most important factor on an individual's risk of cancer is lifestyle choice. For example, smoking increases the risk of lung, larynx, nasal, bladder and oesophageal cancers. Tobacco smoke contains many chemical carcinogens, tumour promoters and co-carcinogens eg. Aromatic amines, phenols, aldehydes and polycyclic aromatic hydrocarbons. Other than exposure to tobacco smoke, the human diet is known to contain a vast array of compounds which increase the risk of cancer such as the cooked red meat heterocyclic amines PhiP (2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine) and MeIQX (2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline) [21], edible plant flavanoids, alkenylbenzene derivatives [22], food additives and contaminants such as the fungal mycotoxins aflatoxin B_1 and PAH [23, 24].

It is known from epidemiologic studies that an unhealthy diet can be the primary cause of between 10-30% of cancer deaths[25].

Although the chemical food components which can cause cancer are known, there are hardly any specific foods which convincingly increase or reduce cancer risk. This is thought to be attributable to the fact a person's diet consists of many different foods, nutrients and chemicals which can affect risk and the actual establishment of a person's diet is confounded by the problems of accurate reporting and recall[26].

The consumption of fibre has been demonstrated recently to be inversely correlated with bowel cancer in a large European epidemologic study the European Prospective Investigation into Cancer and Nutrition (EPIC) [27]. It was shown that the people who had the highest fibre consumption were 15-40% less likely to develop bowel cancer. There have been conflicting reports on the incidence of bowel cancer and fibre consumption in other investigated populations[28, 29], but this is confounded by studies in American populations where fibre consumption is very low by comparison would not detect this

correlation. It is thought that fibre protects against bowel cancer incidence by encouraging fermentative bacteria to produce short chain fatty acids[30] including butyrate which has been shown to inhibit *in vitro* cancer cell growth[31].

It is reported that increased fruit and vegetable consumption can prevent the risk of development of mouth, oesophageal, stomach and lung cancers[32, 33] but no association has been identified with prostate, ovarian or kidney cancers[34-36]. Specificially the highest consumers of fruits and vegetables could reduce their incidence of oesophageal[37] cancer by a third and lung cancer by a quarter[38].

Although, fruit and vegetables are undoubtedly a good source of vitamins, essential minerals and fibre the exact components responsible for the associated reduction of cancer risk is unclear and in isolation many micronutrients can actually exacerbate cancer risk[39]. It is most probable that variation in fruit and vegetable consumption and the synergistic effects of all micronutrient components are essential for the chemoprotective effect observed in large cohort epidemiologic studies[40].

The effect of eating red and processed meat and cancer risk is well documented and is known to influence bowel cancer risk and that high consumers, particularly of processed meats were particularly at a greater risk[4, 15]. Red meat consumption has also been linked to the development of pancreatic[41] and stomach cancers[42]. However, there is no strong evidence that eating white meats can increase risk in the same way as red and processed meats[43].

It has been suggested that haem present in red and processed meats is responsible for the difference in effect of red and white meats on cancer risk. Haem has been demonstrated to be metabolised by gut microbes into N-nitroso compounds which are known to be carcinogenic[44]. Haem may also cause irritation and inflammation of the gut mucosa promoting carcinogenic processes[45]. However, chlorophyll present in green vegetables may ameliorate the effects of haem in red and processed meats[46]. It is also believed that nitrite compounds used as preservatives in processed meats may also contribute to stomach cancer incidence[47]. The nitrites have also been shown to be converted into N-nitroso compounds in the gut[48].

Cancer risk is also known to increase due to deficiency of many essential micronutrients such as folic acid, vitamin B_{12}, iron, selenium, vitamin E, folic acid, vitamin C, vitamin D, and calcium [49].

1.3 Plant phytochemicals- Janus properties

Commonly consumed vegetables, fruit and plant materials contain a range of low-molecular weight secondary metabolites, which serve to protect the plant against attack by predators or pathogenesis of microbes, in the interaction of the plant with its environment, seed dispersal and flower pollination. Plants rely on their diversity of secondary metabolites, enzymatic defences and their cell walls, in the absence of the proteinaceous immune system found in animals, as a form of protection against physiological stress and attack [50].

Several hundred different plant species are regularly consumed by humans as food, which has placed selection pressure on varieties low in toxic compounds, and the promotion of the content of secondary metabolites with beneficial properties, such as reducing undesirable chemicals such as tannins, alkaloids, terpenoids and phenolics. In spite of this reduction of the plants endogenous defences due to cultivation practices, significant levels of biologically-active plant metabolites have evolved concurrently with humans. Many chemicals contained in the matrix of commonly consumed vegetables and plants are potentially toxic and can act as modulators of gene or protein expression [51], as genotoxins [24], as carcinogens [52], as metabolic inhibitors [53] or as enzyme inactivators [54]. Any diet in which plant foods are consumed will contain chemicals which have potentially negative effects on health. These compounds are primarily absorbed in the gut and metabolised in the liver where they may be activated by phase I and II enzymes. The compounds themselves or their metabolites can disrupt homeostatic regulation by interference with the biochemistry of human cells. In the past research focus was upon toxic secondary metabolites such as cyanogenic glycosides in cassava [55], the glycoalkaloids in potatoes and tomatoes [56] and goitrogens [57]. Overwhelming epidemiological evidence of an inverse relationship between plant material consumption and many forms of cancer and chronic disease has changed this research focus [3]. However, there are many plant secondary metabolites that are still of public health concern, particularly those that are manufactured and utilised as additives in food and cosmetics such as the flavanoids and alkenylbenzene compounds [22, 58].

Many bioactive food compounds have been found to have antinutritive value (ie. goitrin in livestock) [59, 60]. These plant compounds can possess outright toxic effects and in some instances carcinogenic potential [61]. Plant phytochemicals can possess 'Janus' properties of having both beneficial and detrimental effects to health dependent on concentration, cell type and secondary metabolite formation. Such phytochemicals can have procarcinogenic, anticarcinogenic and co-carcinogenic properties [62].

For example, following biotransformation the alkenylbenzene derivative methyl eugenol (commonly consumed as a food additive, flavouring agent and found in varying amounts in basil, mace, nutmeg, tarragon, bananas and clove oil) has been demonstrated to have genotoxic effects in animal models and is carcinogenic [63, 64]. The evaluation of the potential carcinogenic properties and toxicity of such commonly consumed plant phytochemicals is of great importance to public health.

However evidence overwhelmingly suggests a strong connection between vegetable and plant consumption and a reduction in incidence of many chronic diseases; such as cardiovascular disease [2], type II diabetes [65], Alzheimers [66] and an inverse relationship with many forms of cancer [3, 67].

It is generally agreed that diets in which there is a high consumption of a variety of fruits and vegetables is beneficial to human health, however the extent to which plant secondary metabolites contribute to this effect, as opposed to antioxidant vitamins and other essential micronutrient components, is the subject of debate [68-70].

The way in which plant phytochemicals reduce the incidence and severity of many chronic diseases is not fully understood. It is believed that the health benefits of consumption of vegetables and plant materials are additive and synergistic and involve combinations and interactions of plant phytochemicals in humans [39].

Amongst the most studied and potent phytochemicals identified are the hydrolysis products of the glucosinolates found in many varieties of the *cruciferae,* over 120 different vegetable glucosinolates have been identified [71]. It is believed that the abundance of glucosinolates is responsible for the epidemiological evidence of an

association of cruciferous vegetable consumption with a decreased risk of lung, stomach, colon and rectal cancer [17].

The active forms of many phytochemicals are limited by their bioavailability, concentration and ability to reach the relevant sites of the body. The concentration of dietary phytochemicals varies between cultivars and is highly dependent on storage and growing conditions. The liberation, absorption, distribution, metabolism and excretion (LADME) of bioactive compounds are all important considerations when interpreting epidemiological studies identifying a relationship between diet and health [72]. It is normally the metabolites of the parent compounds of dietary phytochemicals that are responsible for the observed effect in *in vitro, in vivo* and epidemiological studies. These active metabolites are either produced by action of vacuole bound enzymes within the plant material released following disruption of the cell wall ie. mastication/processing or by extensive phase I metabolism in the liver and intestine in the human body. In view of this many *in vitro* studies of plant phytochemicals need to be tentatively interpreted and use of parent compounds, plant extracts or metabolites in organ-specific cell cultures do not represent the complexity of the biological interaction *in vivo*.

Humans are exposed to an enormous diversity of plant phytochemicals in their diets with beneficial or sometimes detrimental effect on health and incidence of chronic diseases. The mechanisms of dietary food components are difficult to predict in the context of studies of dietary intervention but are of great importance to preventative medicine and reducing the incidence of many chronic diseases.

1.4 Methods of assessing the relationship between diet and health

The assessment of the relationship of specific dietary components and health outcomes and the use of experimental dietary interventions in human subjects is extremely challenging. The optimisation of specific foods and their bioactive compounds represents a sensible, non-invasive and cost-effective method of reducing the global burden of cancer and other chronic diseases [73]. However the identification of dietary compounds which are critical to human health is compounded by the thousands of compounds which are introduced into the human body by diet [39]. There also exists a lack of dietary exposure quantitation for many compounds commonly consumed in the

human diet, which makes extrapolation of actual contribution to health outcomes difficult to assess. For example humans consume over 5000 different flavanoids, of these only a few individual compounds have been assessed for their beneficial health properties [73]. Many plant phytochemicals remain entirely uninvestigated and the individual treatment of phytochemical components of food in isolation can lead to confusion of the role of diet and health. A holistic approach to the investigation of diet and beneficial/detrimental human health outcome is essential, particularly as it is thought that synergistic interactions among food components are crucial to their beneficial effects and many isolated components do not always result in the similar outcomes as in the intact plant material [71, 74].

In many cases *in vitro* empirical evidence is directly extrapolated from the laboratory bench to a supposed effect *in vivo* in humans. The observed effects should be used as an indication of potential *in vivo* effect and as a guide to further complete study within human subjects. Proposed mechanisms of action within cell culture is not necessarily true within the complex endogenous and microbiotic metabolism of a human "superorganism" [75]. Pleiotrophic factors influence the absorption, bioavailabilty, exposure and action within human subjects.

For example, the majority of evidence concerning the anticarcinogenicity of cruciferous vegetables and their glucosinolate hydrolysis products comes from *in vivo* animal studies and *in vitro* cell cultures in which the ITC or indole has been used in isolation [76]. In animal models and cell systems, combinations of ITCs have been shown to confer protection against genotoxic agents better than individual ITCs [77]. The focus on the individual components ignores the potential for chemical interactions and chemoprotective contributions from other compounds found in the normal matrix of cruciferous vegetables. Inter-species differences and the use of doses of the hydrolysed glucosinolate and carcinogenic compounds far in excess of the estimated levels in a normal human diet hinder extrapolation of the results of animal studies in relation to humans. Quantities of ITC's utilised in *in vitro* studies are usually far in excess of human exposure levels and cause acute cellular stress which may promote apoptosis and cell cycle arrest and may not be a relevant effect in human subjects [78].

However, *in vivo* human studies have strongly supported the beneficial effects of CV at high but realistic consumption levels [16, 79-82]. This is in accordance with the epidemiological evidence of the negative correlation between high cruciferous vegetable consumption and the incidence of diet-related cancers [71].

The design of scientific experiments assessing the effects of dietary interventions in humans and experimental animals is crucial to establish correct nutritional advocation, addressing the relationship of diet and disease and acquiring a holistic picture of the interactions of plant phytochemical and other dietary components [83]. The limiting factor remains the design of well defined nutritional interventions that assess the long-term effects of bioactive food components and foods, particularly the speed of discovery and cost of set-up.

Further problems arise in nutritional intervention studies in humans where dietary restrictions and incomplete or falsified food diaries/food questionnaires can confound data interpretation and produce false positive or negative results. Estimation of dietary intake is prone to error as many people's diets are extremely complex and consist of many different foods which are consumed intermittently. The two most commonly used form of dietary assessment are food frequency questionnaires (FFQ) and 24hr dietary recalls. The FFQ is a convenient and inexpensive way to measure an individual's dietary intake over a long period of time, however they are prone to incorrect dietary "self-reporting" [84] with an under-reporting of energy and protein intake by 20-50% [85]. This is even more detrimental when considering dietary components consumed at concentrations several orders of magnitude lower, such as plant phytochemicals and micronutrients. Twenty-four hour dietary recalls provide more in depth information but yet again when a dietary intervention is based on restriction or a reduction of consumption of certain foodstuffs, these can be falsified and intake on a single day is a poor predictor of long-term intakes. These methods of assessment are also costly both in time and financially [86].

The compositional analysis of food also presents analytical problems where a great variation in macro- and micronutrient contents exists. The specific cultivar and growing conditions of plants can determine a great variation in the content of many phytochemicals and micronutrients. For example the concentration of glucoraphanin in

broccoli is known to vary by as much as 25 fold depending on the cultivar [87]. Also the selenium content of many plants is affected by the selenium content of the soil [88].

The method of food processing is also very important in the absorption of dietary food components, as has been observed in increased lycopene availability in tomatoes following cooking and processing [89] and the decreased enzymatic activity reducing the cancer-protectant properties in heated garlic [90].

The quantification of dietary components is essential to a scientifically sound dietary intervention study, however the methods of processing and cooking should also be considered. These factors can all affect the absorption, metabolism, distribution and excretion of dietary food components and therefore exposure of the compound to the target organs and tissues of interest. It is of great interest to combine the assessment of dietary intake, the interplay of genetic factors, the influence of the gut microbiota, the pharmacokinetics, pattern of intakes and effects of other environmental factors in nutritional intervention studies. This represents one of the major challenges of modelling "mammalian biocomplexity" and the effects of dietary components in both health and disease [75].

1.5 Biomarkers of nutritional intervention

Biomarkers of exposure, effect and susceptibility can potentially circumvent the aforementioned problems associated with assessing the relationship between diet and detrimental/beneficial health outcomes and the efficacy of dietary interventions.

A biomarker is any substance which can be monitored in tissues or biofluids which either predicts or influences health, the exposure to dietary components or can influence the incidence or behaviour of a disease [91]. Biomarkers which are able to account for the variations in physiological state and which truly predict diseases states and outcomes have rarely been absolutely identified and usually represent biomarkers of an increased risk of detrimental/beneficial health outcomes [92]. There is little distinction between the requirement for biomarkers which identify the effects of beneficial phytochemicals and food components and biomarkers for environmental toxins and factors detrimental to health; in both cases these are equally elusive [74].

Biomarkers which are able to determine exposure to specific foods or components and also to inform on specific biological effects and inter-individual differences in effect, exposure and metabolism are needed for nutritional intervention studies. The selection and validation of potential biomarkers relies upon the ability to easily and reliably identify them in response to specific dietary intervention or disease state.

The use of molecular biomarkers to assess dietary exposures is highly desirable to identify predictors of dietary intake and as a means to quantify dietary exposure to a particular food stuff or dietary component. These could also potentially provide further insights into the mechanisms of action of beneficial/detrimental dietary components and be used to assess correlations between dietary exposure of certain foods and population epidemiological studies of disease incidence and risk and have important implications for disease prevention and public health.

As an example mercapturic acid conjugates of the hydrolysis products of glucosinolates are the most commonly studied biomarker of CV consumption. However, the excretion of these biomarkers is highly dependent on inter-individual genetic differences, metabolic differences, abundance of the particular glucosinolates in the vegetable consumed, cooking method, bioavailability, level of absorption, bacterial metabolism and the rates and specific metabolic endpoints of myrosinase hydrolysis in the plant material [93]. A more specific, predictable and stable biomarker of CV consumption is desirable for assessment of epidemiological data, dietary intervention study adherence and dietary assessments in human populations. There remains great need for non-invasive biomarkers in both humans and animal experimental models related to dietary intervention.

1.6 Metabonomics

Any alteration in physiological state in an organism can lead to perturbations of the levels of endogenous compounds involved in important biochemical pathways. This causes a disturbance in the cell; in order to return to a state of homeostasis a cell must alter its biochemical composition and this is reflected in the surrounding biofluid. Therefore dynamic changes in biofluid composition can indicate the underlying molecular mechanisms involved following physiological changes. These subtle perturbations can be examined by high-resolution spectral techniques such as ^{1}H-NMR spectroscopy and can yield information of particular metabolic pathways involved in disease states, toxicology, drug metabolism studies and nutritional interventions [94-96].

This study of dynamic changes in low molecular weight metabolites in biofluids such as urine, faecal water, blood serum, cerebral spinal fluid, saliva etc. and also in cells and tissues is known as Metabonomics. Metabonomics is defined as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification"[97] and it relies heavily on high resolution spectroscopy and on chemometric techniques to interpret the complex multivariate spectral data.

Individual patterns of metabolite changes represent the fluxes and alteration of the concentrations of the low molecular weight components of biofluids in response to homeostatic perturbation. Metabonomics analyses provide a time-related snap shot of the metabolic state of an individual. The metabotype is defined as the "probabilistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids or tissues" [98]. By taking serial measurements, the dynamics of such perturbations can be evaluated. Metabonomics can be applied to the process of biomarker discovery. Due to its sensitivity subtle perturbations of biofluids resulting from dietary interventions can be determined [96].

1.7 Application of metabonomics to nutritional intervention studies

The concept of metabonomics has largely been developed within toxicology and drug metabolism studies. However, the role of diet in metabolic outcomes in toxicology studies has been recognised as crucial for some time. The study of cadmium chloride utilising ^{1}H-NMR in rats observed that acute caloric restriction increased levels of plasma creatine [99].

Food-specific biomarkers that are present in the blood serum or urine can either be exogenous compounds whose mere presence indicates consumption of the food stuff or a significant change in an endogenous metabolite effected either by an increase or depletion by the dietary intervention concerned. In many cases, these compounds are phytochemicals or non-nutrients which are unmetabolised, or they represent secondary metabolites which can be eliminated by the same routes as many xenobiotic compounds by glucoronidation, sulphation or glycosylation. However, foodstuffs unlike drugs represent a complex matrix of exogenous constituents rather than a specific known compound. In spite of this complexity when considering dietary intervention Metabonomic techniques can be applied to the determination of the subtle metabolic consequence of dietary intervention and for the purposes of biomarker discovery. This has been demonstrated in a number of studies where low molecular weight biomarkers of dietary intervention were found to predict exposure to different dietary components.

Notably discrimination has been achieved in several human studies contrasting vegetarian, high and low meat diets with metabolites characteristic of both meat exposure and vegetarian metabotype reported [100-102] and in metabonomic studies of dietary standardisation in humans, intersubject variations were found to be most reduced within urine, in contrast to saliva and blood serum [103]. Endogenous metabolite concentration changes in lactate, alanine and citrate have also been reported in human urine samples attributable to a carbohydrate-rich diet[104]. Discriminatory urinary metabolites related to the effects of cooking protein[105] and blood serum metabolites of cooking of vegetables (alpha-carotene)[106] have also been identified. During the heat treatment of protein-containing foods, the amino acid lysine is most prone to undergo chemical reactions in the course of amino acid cross-linking or Maillard reactions. Among the reaction products formed, lysinoalanine, *N*-fructoselysine and *N*-carboxymethyllysine were found to be metabolites in increased concentration in rat

urine [105]. The effects of cooking onions in a dietary in intervention study in humans identified diagnostic urinary biomarkers (quercetin, methyl quercitin and quercitin glucuronides) by HPLC-tandem MS/MS analysis[107]. Urinary metabolic changes characteristic of consumption of other protein sources have also been identified in milk[108] (increased hippurate) and fish[109] (trimethylamine and trimethylamine-*N*-oxide).

Metabonomics has also successfully characterised the effect on metabolic phenotypes of chamomile [110], ethyl glucoside as a biomarker of alcohol dietary intervention [111], in soy isoflavone dietary intervention which was found to have effect on osmolyte and energy metabolic pathways [112]. The effect of epicatechin, a phytochemical founds in many foods such as green tea and dark chocolates, has also been examined using ^{1}H-NMR spectroscopy coupled to metabonomic chemometric analyses in Sprague-Dawley rats. Urinary profiles showed decreased concentrations of 2-oxoglutarate, citrate, dimethylamine and taurine attributable to epicatechin intake. These were concluded to be attributable to shifted energy metabolism from carbohydrate to lipid/amino acid metabolism and highlighted the ability of the metabonomic method to study the bioactivity of dietary plant phytochemicals [113].

Mennen *et. al.* [114] report a number of urinary metabolites which are characteristic of polyphenolic rich foods utilising HPLC-MS analysis such as wine (gallic acid and its metabolite 4-*O*-methylgallic acid), coffee (chlorogenic and caffeic acids), apple (phloretin), orange (hesperetin) and polyphenolic naringenin is reported to be diagnostic of both grapefruit[114] and cacoa comsumption[115].

There are a number of studies that have reported metabolites associated with black tea consumption specifically gallic acid and its metabolite 4-O-methylgallic acid [114]
A comparison of the effects of green tea and black tea consumption in humans using metabonomic analysis of urine and blood serum revealed gut-bacterially derived metabolic endpoints of tea consumption, hippurate and 1,3-dihydroxyphenyl-2-O-sulphate in urine. Most interestingly green and black teas were found to have a differential effect on the excretion of endogenous human metabolites. Green tea was shown to decrease concentrations of several urinary citric acid cycle intermediates and black tea blood serum levels of lipoproteins, glucose and acetate. This study revealed a

possible effect of tea consumption on energy metabolism and biosynthesis in humans [116].

Recently a novel metabolic profiling strategy for food biomarker discovery combined a nutritional intervention with citrus fruits and validation with a large-scale epidemiological study [96]. Proline betaine a biomarker of citrus fruit consumption was identified in a human citrus fruit dietary intervention study by ^{1}H-NMR spectroscopy. The plant origin of this osmolyte was then identified in citrus fruits by targeted ^{1}H-NMR spectroscopy. Finally, the putative biomarker was validated by profiling in ^{1}H-NMR spectra of urine in citrus fruit consuming and non-consuming UK participants of the INTERMAP study (International Collaborative Study of Macronutrients, Micronutrients, and Blood Pressure). Citrus fruit consumers in the study were shown to consume lower proportions of fats, have lower urinary sodium-potassium ratios, and consume higher proportions of vegetable protein, fibre, and most micronutrients than non-citrus consumers. This citrus fruit biomarker identification strategy was successful in correlating presentation of this biomarker in human urine in a large scale epidemiological study and healthier eating patterns.

It is known that, particularly within human subjects a high genetic, environmental and dietary diversity exists within a population. This is particularly confounding within multivariate statistical analyses of human biofluids. The limitation of this inherent inter-subject variability is essential to successful metabonomic analysis, particulary dietary intervention which in most instances does little to provoke a dramatic perturbation of homeostatic regulation. The biofluids of humans are also thought to be affected by pleiotrophic physiological factors unrelated to diet such as hormonal status, diurnal cycles and most importantly the inter-individual specificity of the gut microbial populations [75]. Metabonomic technologies represent a powerful tool for the elucidation of the effects of dietary intervention. However inter-subject variation particularly within human subjects requires consideration, particularly the gut microbiota which is thought to be one of the greatest determinants of inter-subject variability.

Diet-induced phenotypic responses are known to be highly individual-specific and are influenced by genetic factors (nutrigenetics), gene expression (epigenomics and transcriptomics), protein expression and activity (proteomics), fluxes and alterations of

small molecular weight compounds (metabolomics/metabonomics) and gut microbial contributions to metabolism (microbiological metabonomics) [117-120]. The effect of nutritional intervention and disease incidence is collectively described by the interplay of systems biology "-omics" and the identification of biomarkers is made in the context of the human "super-organism" [75].

A study examining caloric restriction in mice undertook a systems biology approach looking at and correlating blood plasma metabolic changes with genomic transcriptional changes in mice. It was found that caloric restriction induced a change in gene expression away from lipid biosynthesis towards fatty acid catabolism, β-oxidation and gluconeogenesis. Metabonomic changes revealed increasing lactate, 3-hydroxybutyrate, cholesterol and low-density lipoproteins with increased creatine and several glucogenic amino acids; methionine, glutamine, alanine and valine as a result of acute caloric restriction. These results indicated metabolic switching to energy conservation and gluconeogenesis due to caloric restriction and successfully illustrate the utility of metabonomics within a systems biology approach to metabolic consequences of dietary change in experimental animals [121]. Metabonomic analysis has also been used to examine the effects of controlled gene expression on metabolic profiles in a mouse model (mnd) of motor neuron degeneration and the effects of vitamin E supplementation on blood serum and solid-state MAS ^{1}H-NMR spectroscopic profiles of cerebral tissue [122].

Metabonomic analysis is still challenged by the difficulty of modelling responses to diets or foodstuffs in biologically complex organisms like humans and the ability to distinguish between the effects of nutritional intervention and normal physiological variation.

1.8 The human gut microbiota

There are believed to be between 10-100 trillion bacteria inhabiting the human gastro-intestinal tract with an estimated biomass of ~1.5kg, and they are thought to outnumber the individual mammalian cells in the human body [123].

There are 500-1000 different gut microbial species co-existing in stationary phase. These bacteria are ultimately responsible for physiological and pathological processes,

nutrient cycling and syntrophy of food components in the intestinal tract, trophic effects on the intestinal epithelia and the immune system and also compete against pathogenic bacteria in the intestinal lumen [124]. It is estimated that 99% of all active genes within the human body are microbially derived and as such understanding the symbiotic relationships between humans and the gut microbiota is crucial in systems biology [125].

The homeostasis of the gut microbiota is influenced by many factors including genetics, nutrition, age, stress and sex [126]. The GI tract is colonised immediately at birth by *Lactobacilli* and *Bifidobacteria* which assumes an adult colonisation pattern when solid food is introduced into the diet [127, 128]. The microbial composition of the GI tract is determined by the transitional decrease in the oxygen level between the upper and lower GI tract. Aerobic gut bacteria are thought to utilise oxygen in the upper GI tract which supports a favourable environment for facultative and obligate anaerobes further along the GI tract [129, 130]. The complex relationships of human commensual bacteria with other species and host have made it extremely challenging to culture individual anaerobic species in *in vitro* assays and will continue to be a barrier to further understanding of the proportions of the total component species. However assiduous studies are currently being performed to characterise the genetic and metabolic components of the human gut microbiota as part of the human microbiome project [131].

The human gut bacterial species which are found in the highest proportions have been characterised and include obligate anaerobes *Bacteroides spp. Provetella, Porphyromonas, Bifidobacterium spp., Peptostreptococcus spp., Clostridium spp., Eubacterium spp.* and facultative anaerobes *Enterobacteriaceae, Streptococcus spp., Enterococcus spp., Staphylococcus spp., Lactobacillus spp.,* and fungi [131].

It is known that many metabolites seen in human biofluids are exclusively or predominantly microbially derived such as *p*-cresol and its precursor *p*-hydroxyphenylacetate. *P*-cresol is formed from tyrosine *via* transamination to p-hydroxyphenylacetate by P.*vulgaris* which is then decarboxylated by C.*difficile* and *lactobacillus* species [132]. Interestingly, high proportions of anaerobic gut microbes such as C.*difficile* and a depletion of P.*vulgaris* can indicate poor diet and can lead to a build up of p-hydroxyphenylacetate in the urine. The increase in urinary concentration of bacterial associated metabolites such as *p*-cresol is related to amino acid absorption by

the gut mucosa, short-chain fatty acid concentration and bacterial activity in the gut by species Clostridium *difficile* [132] *lactobacilli sp.* [133] and Clostridium *scatologenes* [134]. Following *p*-cresol reabsorption in the intestinal lumen, sulphation occurs predominantly in the liver and subsequent excretion of *p*-cresol sulphate in the urine along with many other phenolic bacterial co-metabolites [135].

Bacterial metabolites are commonly observed in human biofluids and are related to tryptophan transamination, glycine benzoate conjugation (hippurate) and glutamine acetylation (phenylacetylglutamine). It has been suggested that the excretion of these metabolites serves as an alternative to urea for removal of waste nitrogen [136]. The human colonic microbiota plays a role in human lipid metabolism through the production of short chain fatty acids, deconjugation of bile acids and cholesterol assimilation in the colon [137]. Recent studies looking at the metabolic interactions between the gut microbiota of and the mammalian host have highlighted the role of the gut bacteria in the enterohepatic recirculation and in systemic lipid metabolism [138].

It is believed that diet influences the composition and activity of the gut microbiota [139]. In particular, a diet rich in both soluble and insoluble fibres promotes the growth and fermentation of some bacterial populations in the colon [140, 141].

Evidence suggests that humans and their microbiota have co-evolved to such an extent that we cannot survive without them and many commensuals produce essential metabolites which humans cannot produce endogenously [142]. The highly adaptive nature of the gut microbiota is highlighted by opportunistic bacteria such as (Entero Pathogenic *E.coli* EPEC) which can assume pathogenic characteristics under stress conditions [143]. Clostridium *difficile* is also known to become pathogenic in geriatric wards where long term low dose antibiotics are used and bacterial resistance emerges [144, 145].

Understanding how these symbiotic relationships contribute to metabolic variation in nutritional intervention studies is important to understanding the complex interplay of nutrition and human health.

1.9 Aims and Objectives

Diet is of crucial importance to health and the prevention of many chronic diseases of significant public health concern. Understanding the relationships between food components both detrimental and beneficial to health and disease is of vital importance to preventative medicine.

There exists a need for non-invasive biomarkers of dietary exposure in both humans and animal experimental models for interpretation of epidemiological studies, estimation of dietary intakes and exposures and for elucidating the pharmacokinetics of dietary food components.

This thesis addressed the hypothesis that application of metabonomic technologies to nutritional intervention studies can identify biomarkers of health and disease. Specifically I have examined cohort datasets as part of a systems biology approach to understanding dietary interactions in:

1. A dietary intervention study in rats examining the effect of methyl eugenol a genotoxic carcinogen on the rat urinary metabolic phenotype.
2. A dietary intervention study in humans examining the effect of alterations of dietary fat composition and GI of carbohydrates in subjects at risk of developing the MS on the urinary and faecal water metabolic phenotype.
3. A dietary intervention study in humans examining the effects on the urinary metabolic profile of a CV, meat meal and caffeine dietary interventions.

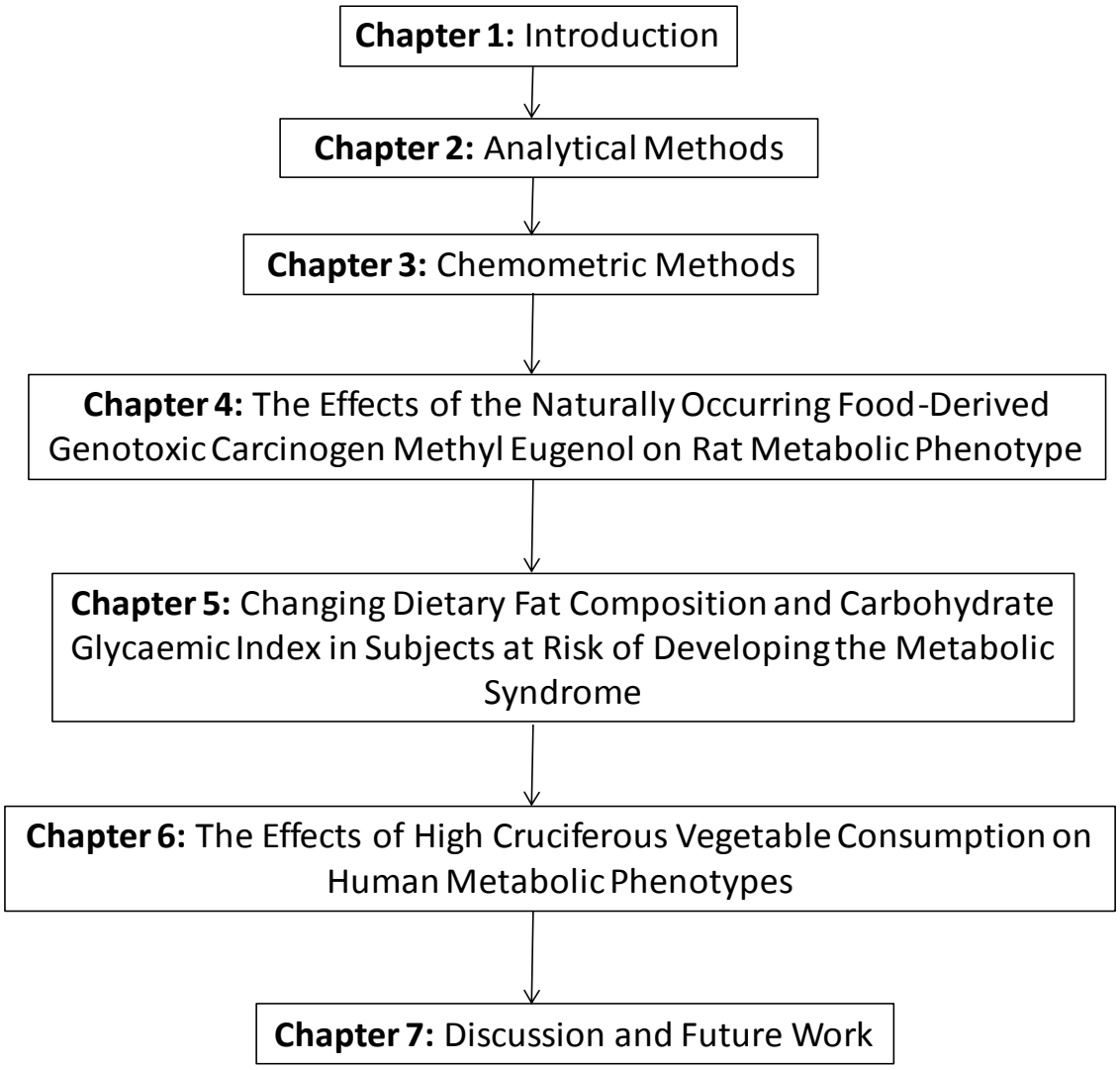


Fig. 1.1 Flowchart of thesis structure.

Chapter Two: Analytical Methods

2.1 An Introduction to Spectroscopy

2.1.1 Biological Nuclear Magnetic Resonance (NMR) Spectroscopy

The initial theoretical principle of nuclear magnetic resonance spectroscopy (NMR) was first proposed in 1924 by Austrian-Swiss physicist Wolfgang Pauli. He formulated the Pauli Exclusion Principle, which stated that no two electrons could exist in the same quantum state. This quantum degree of freedom resolved inconsistencies between observed molecular spectra and the developing theory of quantum mechanics. Kronig, Uhlenbeck and Goudsmit further developed this idea of degree of freedom in 1927 to the idea of spin. Certain nuclei possessing spin characteristics would exhibit split energy levels when exposed to an applied magnetic field. Pauli also derived the first observed spectrum of the hydrogen atom.

2.1.2 Principles of NMR Spectroscopy

Subatomic particles within an atom possess spin on their axes; in many atoms these spin states are paired such that the nucleus has no overall spin. If the neutrons and protons are even then the nucleus has no spin. However in ^{1}H and ^{13}C for example the nucleus possesses a net spin. For ^{1}H spin is created by the presence of a single proton and no neutron, the nucleus has a half integer spin possessing two possible orientations. Without an external magnetic field, these orientations have equal energy. When a magnetic field is applied, the energy levels of the nuclei in a population split into a lower and higher energy level. Initial populations are determined by thermodynamics described by the Boltzmann distribution. The application of electromagnetic radiation, in this case the magnetic field, will excite nuclei in the lower energy level to the higher level. The frequency of radiation needed is determined by the difference in energy between the two energy levels, the transitions between nuclear energy levels occur by absorption of electromagnetic radiation in the radiofrequency range.

$$\Delta E = h\nu$$

h is Planck's constant and ν is the frequency of the absorbed radiation. The Larmor frequency is defined as the frequency of radiation that causes a spin state transition (ΔE) in the nuclei and is dependent on the interaction between the applied magnetic field B_0 and the nuclei.

Either nuclei are aligned against the magnetic field in the low energy state, or with it in a high energy spin state. The precise amount of energy needed for this transition depends on the nucleus in question and strength of the external magnetic field.

$$v=\gamma B_0/2\pi$$

Where v is the frequency and γ is the gyromagnetic ratio a nuclide specific constant that provides a measure of the 'magnetic strength' of a nucleus. The intensity of the signal in an NMR spectrum is determined by area under the resonance line and proportional to the number of nuclei responsible for the NMR signal. This is useful in structural determination of the number of protons in each functional group contributing to a peak in a ^{1}H-NMR spectra.

2.1.3 Relaxation times

Signal is detected within the radio frequency range; this causes perturbations in the magnetisation field. In static magnetic fields nuclei precess about the field with random distribution of spin orientations; this is known as the bulk magnetisation vector (Mz), or the sum of the individual vectors (**Fig 2.1)**.

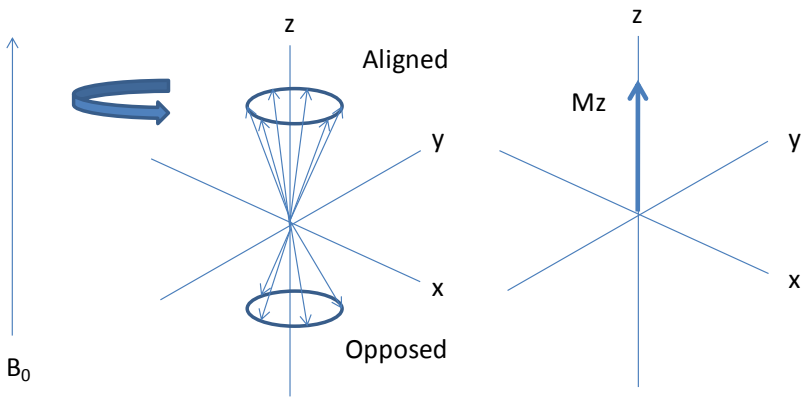


Fig. 2.1 Bulk magnetisation vector (Mz)

The application of a 90º radiofrequency pulse causes the magnetisation of the nucleus of interest to tilt perpendicular to the applied field B_0 in the x-y plane (**Fig. 2.2)**. Following this perturbation the nucleus goes through a relaxation process returning to the Z-axis in alignment with the applied field. There are two types of relaxation process, Spin-lattice transverse relaxation T_1 and spin-spin longitudinal relaxation T_2.

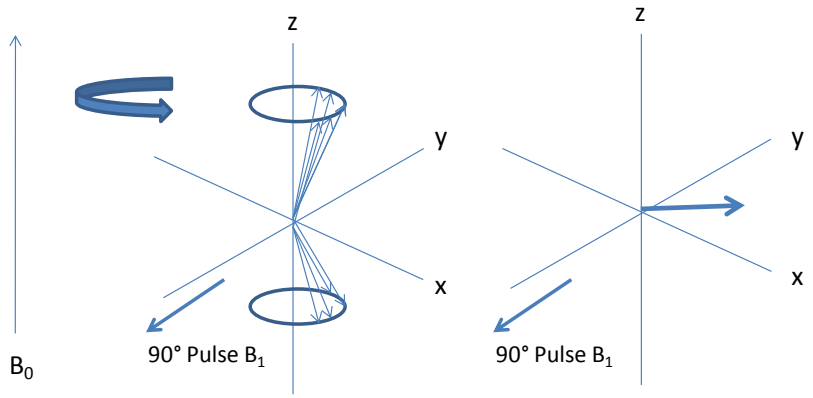


Fig. 2.2 Effect of 90° pulse tilting magnetisation into the x-y plane

The nuclei within an NMR sample are in vibrational and rotational motion creating a complex magnetic field. The sample in which the nuclei are held is known as the lattice. The relaxation time T_1 is dependent on the magnetogyric ratio of the nucleus and the mobility of the lattice (**Fig 2.3)**. This represents the average lifetime of nuclei in the higher energy state and a return to equilibrium population levels.

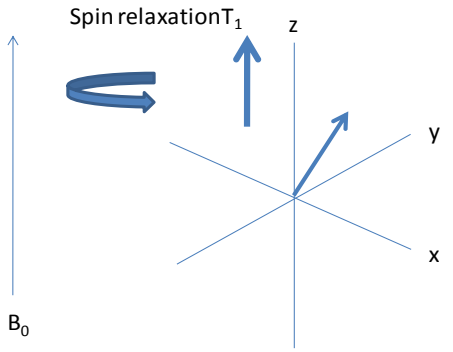


Fig. 2.3 T_1 spin relaxation

Spin-spin relaxation is the interaction between neighbouring nuclei with identical processional frequency but differing magnetic quantum states. Nuclei can freely exchange quantum states. This causes no net change in the populations in the different energy states but causes a decrease in the lifetime of a nucleus in the excited state. The T_2 is the time constant for the decay of magnetisation in the perpendicular plane (x-y) to the applied field B_0 (**Fig 2.4)**.

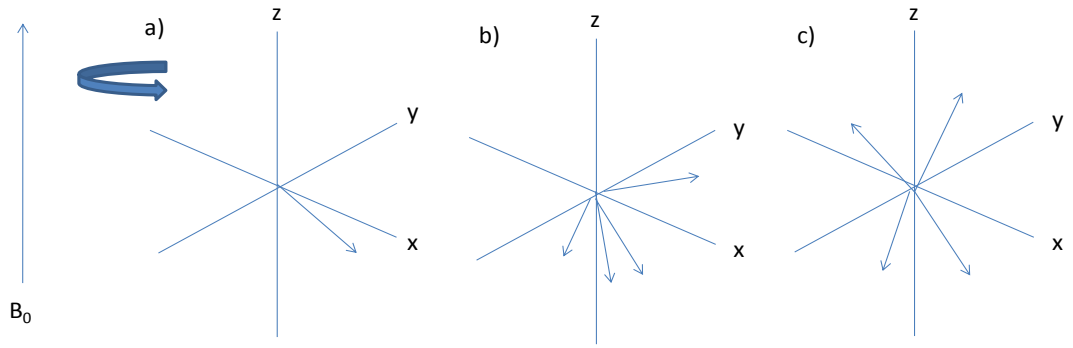


Fig. 2.4 Spin-spin (T_2) relaxation, time taken for all protons in the observed sample to decay in the x-y plane to the applied magnetic field B_0.

The variations between T_1 and T_2 that exist between protons of different molecules are very important for NMR spectroscopy as they can be manipulated by the radiofrequency pulse programs used.

2.1.4 Chemical shift

Chemical shift is a function of the level of nuclear shielding in proportion to the applied magnetic field *B* and is a function of the nucleus and its environment. The chemical shift is measured relative to a reference compound which is tetramethylsilane in the case of ^{1}H-NMR.

Electrons around the nucleus act as a shield to the applied field. The difference between the applied magnetic field and the magnetic field of the nucleus is termed nuclear shielding. Circulating electrons in S-orbitals produce a magnetic field opposing the applied field. A diamagnetic shift is required for the nucleus to absorb at its transition frequency. Electrons in p-orbitals however have no symmetry, which in turn produces a paramagnetic shift due to large magnetic fields at the nucleus. In ^{1}H-NMR, there are no p-orbitals, therefore there is a small range of observable chemical shifts as a molecule becomes increasingly electronegative, electron density around the proton decreases, and resonance occurs at lower field strengths.

Following radiofrequency pulse application, a time domain decay signal is detected by the spectrometer between the pulse intervals. This represents the emission of energy as the nuclei gradually relax between the two energy transitions and process back to population equilibrium. This presents as oscillations that gradually decay and is known

as the *free induction decay* (FID). The FID represents a series of frequency components that are manipulated mathematically by fourier transform to a frequency domain signal. This gives the NMR spectrum showing protons in all of the different chemical environments.

The process of shimming is performed to correct any inhomogeneity in the applied electromagnetic field. This involves the adjustment of small electromagnets located around the sample within the superconducting magnet. This is very important as inhomogeneity of the magnetic field leads to line broadening of the resonance peaks within the spectrum.

2.1.5 Spin coupling

Spin-spin or scalar coupling occurs when neighbouring nuclei within a compound cause a splitting of spin states into a fine structure. The influence of a methyl group containing three protons upon a neighbouring single proton is to split the NMR resonance into a quartet of four peaks. As a rule of thumb, resonances are split by the number of neighbouring protons within three bonds distance +1. However exceptions occur with chiral molecules where AB and ABX spin systems can occur resulting in a more complex splitting pattern. Spin-spin coupling is very useful for structure elucidation and assignment in NMR spectra.

2.2 NMR pulse programs

The basic pulse program involves a simple system of improving the signal to noise ratio. The basic program consists of pulsing the sample acquiring the FID and allowing a delay to avoid saturation and allow relaxation. Dummy scans are also performed at the beginning of the acquisition to allow the applied magnetic field to stabilise to a steady state or in decoupling experiments where the high power irradiation increases the sample temperature. They are also performed to reach steady state conditions for T_1 relaxation if the recycle delay is shorter than four times the T_1 value of the nucleus measured. The serial collection of FIDs gives a signal to noise improvement equivalent to the square root of the number of FIDs acquired. Therefore the acquisition of 100 FID scans provides a 10 fold improvement in the signal to noise ratio.

2.2.1 Water pre-saturation

Water pre-saturation is obtained using a two pulse NMR programme first with a low energy pulse from the RF coil which selectively saturates the single large water signal in the NMR spectra using a highly specific frequency. This can also be applied to other solvents within the field of the NMR spectrum; this is followed by a non-specific 90º pulse to excite all of the other desired resonances. The result of this pulse sequence is a spectrum with a reduced water signal and improved signal to noise ratio for the metabolite resonances of interest and a reduction in the distortion of the spectrum baseline. Suppression of the water peak also allows an increase in the available gain.

2.2.2 Standard 1D pre-saturation pulse sequence

The standard 1D pulse sequence consists of three consecutive 90° pulses with water presaturation pulses applied between each 90° pulse. The magnetisation vector flips in the x-y plane, into the *z*-axis and finally back into the x-y plane. This has the effect of improving the signal to noise ratio and amplifying the returning signal by suppressing all signals that are not in the 90° orientation before acquisition of the FID. Field inhomogeneity in B_1 results in the edges of the sample volume not experiencing the applied field in totality. These regions normally contain residual solvent, this leads to improved solvent suppression in aqueous samples.

2.2.3 Carr-Purcell-Meiboom-Gill (CPMG)

The purpose of the Carr-Purcell-Meiboom-Gill (CPMG) sequence (Meiboom and Gill 1958) is to manipulate the T_2 (time constant for the decay of magnetisation in the x-y plane perpendicular to the applied field B_0) of protons to reduce contributions from proteins, lipids and other macromolecules found within tissue extractions and normal matrices of biofluids such as blood serum. The CPMG pulse sequence is a modification of the Hahn spin-echo sequence generating signals weighted according to T_2. The sequence begins with a 90° pulse which tilts the proton magnetisation into the xy-plane; different nuclei possess B_0 heterogeneity and process at different frequencies. This results in 'fanning-out' of the magnetisations. In contrast to the Hahn sequence, following the delay time T2 (the spin-echo delay), the CPMG pulse sequence applies a second 180° pulse to the *y*'-axis rather than the *x*-axis, the magnetised components subsequently invert echoes forming along the same direction improve 180° pulse accuracy. The influence of the applied magnetic field B_o causes the components to

precess at the same frequency and arrive at the Iy'I-axis. This produces an echo signal at time T_2 after the pulse sequence. To minimise signal loss mechanisms a series of 180° pulses is applied to improve detection of the signal.

Within the context of this thesis the CPMG pulse sequence was applied to aqueous vegetable extracts to clearly observe the water soluble components and to reduce the broadening seen within the ^{1}H-NMR spectra due to macromolecular components and to give clear representation of the lower molecular weight components of interest.

2.3 Two-dimensional NMR spectroscopy

The application of two-dimensional (2D) NMR spectroscopy is to elucidate the structure of complex unknown molecules by dispersion of the intensity of NMR signals plotted along two frequency axes or dimensions. 2D NMR spectroscopy provides information on the correlations/relationships between different spins within the same molecule. 2D pulse programs consist of a series of preparation and mixing pulses the induced magnetisation precesses within the subsequent evolution and detection time phases. The detection period provides the 1D time-domain data from the Fourier transformed FID, the evolution phase consists of repeated NMR pulses with increasingly longer time periods. This creates the second time-domain (**Fig. 2.5**).

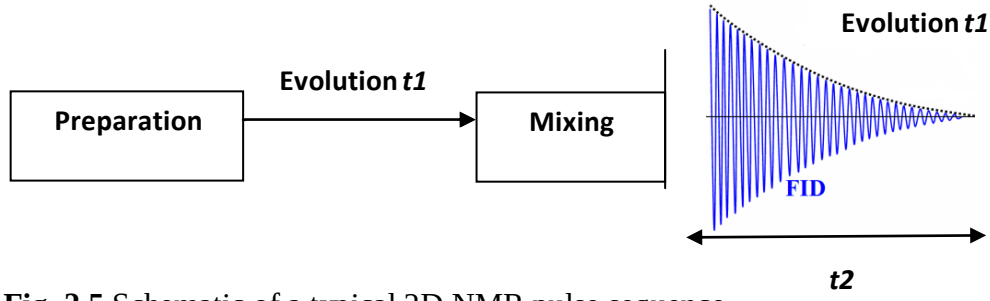


Fig. 2.5 Schematic of a typical 2D NMR pulse sequence

2.3.1 J-Resolved ^{1}H-NMR spectroscopy

In standard 1D ^{1}H-NMR spectra chemical shift and spin-spin overlapping multiplets can obscure coupling information. In a 2D *J*-resolved experiment the chemical shift and coupling information are separated and displayed on different axes of the 2D matrix. On the F_2 axis only chemical shift information is presented and on the F_1 axis only spin-spin coupling information. A 2D ^{1}H *J*-resolved spectrum, when projected onto the F_2,

gives a proton decoupled spectrum, each appearing simply as a singlet. 2D *J*-resolved spectroscopy is based on a repetition of the spin echo sequence varying the delay period to allow *J* modulation effects to evolve. This experiment is particularly useful for deconvolution of crowded areas of a ^{1}H-NMR spectra [146].

2.3.2 Correlation spectroscopy (COSY)

The homonuclear COSY experiment utilises coherence or magnetisation transfer within the COSY90°. The second 90° pulse causes magnetisation transfer through adjacent directly *J*-coupled bonds. Frequency axes both contain chemical shift information and the 1D intensity peaks are seen along the diagonal, cross-peaks in the spectrum represent nuclei which are spin-spin coupled.

2.3.3 Total correlation spectroscopy (TOCSY)

In homonuclear TOCSY spectra correlations are seen between directly coupled nuclei within a molecule as in COSY spectra. However, correlation is also seen within nuclei within an unbroken spin system. The diagonal on TOCSY spectra represents the intense peaks which match the 1D spectrum; the cross peak multiplets represent correlations of nuclei in an unbroken spin system of up to 5 or sometimes more bonds. AB spin systems or a carbonyl for example would constitute an interruption of a chain of TOCSY correlated nuclei. This form of 2D NMR spectroscopy is complimentary to COSY spectra in elucidating correlations between protons within unknown metabolites

2.3.4 Heteronuclear Multiple Bond Coherence (HMBC) and Heteronuclear Single Quantum Coherence (HSQC) 2D ^{13}C NMR experiments

These heteronuclear ^{1}H-^{13}C 2D NMR experiments are very useful and complimentary experiments for identifying unknown compounds, particularly those for which there is little structural information based on the 1D-NMR spectrum ie. unknown singlet peaks. Heteronuclear refers to the use of both ^{1}H and another nucleus possessing spin usually ^{13}C, but correlation with other isotopes is also possible such as ^{31}P. The HMBC experiment utilises the stable ^{13}C isotope which constitutes approximately 1.1% natural abundance of carbon. This experiment shows correlations between protons in a molecule and carbons over long range *J*-coupling of 2,3 or sometimes more bonds by an inverse detection experiment. Peaks from protons attached to ^{12}C are removed by the pulse sequence implemented. HSQC on the other hand utilises ^{1}H protons to observe

^{13}C which are directly coupled and usually of no more than one bond length in distance. HSQC provides the chemical shift of directly bonded ^{13}C to the proton observed in the ^{1}H NMR spectrum. The identification of metabolites can be made based on comparison of the carbon 13 and proton chemical shifts to literature sources or authentic standards. This is because the ^{13}C chemical shift is relatively insensitive to pH and sample concentration, and it can be used to compare different biological matrices. Due to the low abundance of ^{13}C high resolution NMR spectroscopy and upwards of 200 scans are required for low concentration samples. A chromatographic concentration of the unknown is desirable prior to conducting this experiment to aid structure elucidation eg. Solid phase extraction.

2.4 Preparation of Biological Samples

2.4.1 Flow Injection Urine Sample Preparation

A 1ml aliquot of each of the urine samples was taken for ^{1}H-NMR spectroscopy. To reduce pH induced chemical shift variability in the urine samples a 400µl aliquot of urine samples was combined with 200 µl of 100 mM phosphate buffer solution (pH 7.4). The samples were allowed to stand for 10mins before centrifugation at 13,000 g for 10 mins to remove insoluble materials. Aliquots (500 µl) were transferred in a randomised order and with controls inserted every fifteenth sample to a 96 deep-well plate and the samples 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 ($\delta^{1}H$ 0.0) and D_2O provides a lock signal for the NMR spectrometer.

The completed plates were covered with adhesive film and mixed carefully before being stored at 5˚C. Immediately prior to NMR analysis the plates were centrifuged at 3,000 g $_{av}$ for 5 mins. Push solvent (1g NaN_3 in 1000 ml dH_2O) was used for NMR automation.

2.4.2 Solid Phase Extraction (SPE)

Solid phase extraction is a chromatographic technique that is used for the concentration and purification of samples. This is particularly useful for the elucidation of unknowns from low concentration/large sample volume biological samples such as urine. SPE normally utilises a C_{18} reversed phase column containing a prepacked sorbent which consists of chemically bonded *n*-octadecylsilane, an alkane with an 18 carbon chain to a

silica matrix. The column is first conditioned and washed with the appropriate solvent and the sample is then loaded onto the column. Analytes within the urine are Van der Waals bounded to the sorbent bed and are eluted by disrupting the interaction of analyte and solid phase by sequential washes of increasingly concentrated solvent. Analytes are eluted or retained on the column based on their polarity and therefore the corresponding fraction of elution provides information on the polarity of an unknown constituent of urine. In a reverse phase the polarity of the stationary phase is less than the mobile solvent phase which usually consists of methanol equilibrated with water. Increasing concentration of methanol decreases polarity of the solvent phase across the column.

2.4.2.1 Solid Phase Extraction method

Urine (2 ml) was acidified to pH 2 with 1 M HCl and extracted with a C_{18} bonded cartridge (Isolute, 200 mg, supplied by International Sorbent Technology Ltd, Hengoed, UK). The cartridges had previously been activated by washing twice with 5ml methanol. The fractions were eluted with, 5, 10, 20, 40, 60, 80 and 100% methanol/H_2O solutions (pH 2). All samples were freeze-dried and reconstituted in 600 µl LC-MS grade H_2O (Sigma-Aldrich, UK). An aliquot (500 µl) was transferred into a 5 mm NMR tube prior to NMR spectroscopic analysis and 100µl was utilised for MS/MS analysis.

2.5 ^{1}H-NMR Spectroscopy Methodology

2.5.1 Flow Injection Experiment

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 pipetted into the transfer line and separated from subsequent samples by 500µl of a push buffer containing HPLC grade water containing 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 utilising a standard one-dimensional pulse sequence, the first increment of which is the NOESY pulse sequence which suppresses the large ^{1}H-NMR signal of the solvent water. 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, phase and baseline correction. The magnetic field was shimmed for the first control sample only and then held throughout all of the subsequent sample data collections.

2.5.2 Carbon 13 NMR spectroscopy

2.5.2.1 Heteronuclear Single Quantum Coherence (HSQC)

^{1}H-^{13}C 2D HSQC NMR spectra were recorded in echo-antiecho mode using the gradient selection. The ^{1}H and ^{13}C spectral widths were 10.5 and 200 ppm respectively with the recycle delay of 2 s. A total of 200 scans were collected into 2k data points for the acquisition dimension and 128 increments for the second dimension (^{13}C). The δ-shifted sine bell window functions were applied for both dimensions before Fourier transforming into data with the standard Bruker linear-prediction function applied for the second dimension.

2.5.2.2 Heteronuclear Multiple Bond Coherence (HMBC)

^{1}H-^{13}C 2D HSQC NMR spectra were performed using a standard pulse sequence from the library supplied by the vendor. 900 t1 increments with 128 transients and 4096 complex data points were recorded with a spectral width of 6613 Hz in f2 and 30183 Hz in f1, resulting in a total acquisition time of about 57 h and 36 min. The delay for the 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.

2.5.3 Authentic standard compound spike in for NMR peak assignment

The identities of unknown NMR spectral features were confirmed for metabolite assignment by successive addition of 1µl of 100mM authentic standard compound and 1D ^{1}H NMR spectra acquisition on a Bruker 400MHz spectrometer maintaining consistency in the acquisition parameters for the purposes of comparison. Standard compound was added in a step-wise manner to attempt to increase the resonances of the unknown metabolite(s). Increasing intensity of the resonance and similarity in splitting pattern at the frequency of the unassigned unknown resonance of interest was taken as positive confirmation.

2.5.4 Structural assignments

Both ^{1}H and ^{13}C NMR assignments were made using both online NMR database searches (particularly human metabolome database Hmdb, Genome Alberta), NMR prediction software (ACD labs) and previously assigned metabolites, taken within the group and where available assignments were made using published assignments at neutral pH [147, 148]. Any ambiguity in assignment was surmounted by statistical total correlation spectroscopy (STOCSY), 2D homo- and hetero- nuclear NMR experiments and spiking of authentic (sigma-aldrich) or synthetic standards.

2.6 Ultra High Performance Liquid Chromatography (UPLC)

UPLC chromatography is a chromatographic technique utilised in high-throughput biological analyses to provide high resolution separation of complex sample mixtures for quantitative analysis. UPLC utilises high pressures to force samples through a column packed with a sorbent solid medium, usually the separation consists of a mobile liquid phase which passes the analyte samples through a stationary phase which separates components in the solution based upon polarity. In the context of this metabonomics study C_{18} columns were used in a reversed phase, this is where the stationary phase is less polar than the mobile liquid phase whereby increasing the polarity of the mobile phase non-polar metabolites are repelled onto the surface of the column stationary phase where they remain attracted until the mobile phase becomes increasingly less polar over time. The desired result of this effect is that metabolites are retained upon the column entirely dependent on their hydrophobicity and this serves to serially separate metabolites in urine this serves to greatly increase the resolution of subsequent mass spectral analyses. The resolution and peak efficiency of liquid chromatograpy is highly reliant on the particle size of the stationary phase which necessarily requires high pressures to force the mobile phase through the column. However, the flow rate through the column is limited by the considerable back pressures in the system when small particle sizes in the stationary phase are used. UPLC chromatography utilises sub 2μm stationary phase particle sizes operating at pressures of ca. 12,000 psi, this provides increased speed, sensitivity and peak resolution. The advantages over larger stationary phases (~3.5μm) in high performance liquid chromatography (HPLC) which in a reversed phase is a doubling of peak capacity, a 10-fold increase in speed and a 3-5 fold increase in sensitivity [149]. The qualities of UPLC are highly desirable in high-throughput metabonomic biofluid studies where large sample numbers need to be time-efficiently and accurately quantified particularly when

analysing urine which is the most chemically complex biofluid produced by the body containing tens of thousands of low molecular weight components, consisting of many by products of metabolism, excess metabolites, bacterial co-metabolites, waste metabolites and environmental xenobiotics. [150]

2.7.3 Ultra High Performance Liquid Chromatography Methodology

Chromatographic separations were performed on a Waters ACQUITY^{TM} UPLC BEH 2.1x 100 mm 1.7 µm C_{18} column (Waters Corp, Milford, USA) using an ACQUITY ^{TM} Ultra Performance Liquid Chromatography system (Waters). UV spectra were acquired with a Waters PDA UV detector. The column was maintained at 40 °C and eluted with a reversed phase gradient (See **Table 2.2**) of 100- 5% A (0–10.1 mins) with 0–95% (0–10.1 mins) B where A = H_2O (UPLC grade, SigmaAldrich, UK) with 0.1% formic acid and B = acetonitrile (UPLC grade, VWR, UK) with 0.1% formic acid. The gradient duration was 12 min at a flow rate of 0.5 ml min^{-1}. A 10 µL injection of each sample was made onto the column.

Table 2.1 Reversed phase gradient conditions for UPLC of Human Urine samples. **A**=H_2O (0.1% formic acid), **B**=acetonitrile (0.1% formic acid).

Time (mins)	**%A**	**%B**
0	100	0
1	100	0
3	85	15
6	50	50
9	5	95
10	5	95
10.1	100	0
12	100	0

2.7 Mass Spectrometry (MS)

The basic principal of mass spectrometry is the separation of compounds based on their mass to charge ratio or *m/z*. Mass spectrometers can provide accurate masses, structural information and dependent on the resolution of the spectrometer exact mass which can be used for the identification of unknown compounds within a biological matrix. Samples are ionised producing either a positive (+ve) or negative (-ve) ion which are directed electrostatically in a vacuum through a mass analyser based on *m/z* and are subsequently detected in the ion detector.

2.7.1 Application of Mass Spectrometry to Biological Samples

Metabolites in biological samples such as urine can be ionised using a variety of different sources, however within this study the most suitable was electrospray ionisation (ESI), as ESI provides superior mass range (>50 *m/z,* however this is often close to the noise threshold) and sensitivity (in the fentimolar range, dependent on metabolite). ESI ionises metabolites within a biofluid by creating a fine spray of highly charged droplets of the sample and solvent in a strong applied voltage across the ionisation cone (cone voltage). The applied voltage separates the charges within the gaseous ionised droplets on the liquid surface, when ionic repulsive forces and surface tension of the droplet are equal the droplets change into a cone shape known as a Taylor cone [151]. Following this ionisation process the electrosprayed sample droplets are electrostatically attracted into the mass analyser. The significant advantage of ESI for biological sample analysis lies in the ease of use with liquid chromatography (LC) and compatibility with MS/MS tandem mass analysers. However, ESI suffers from a reduction in sensitivity due to ion suppression in complex mixtures (such as urine) and the presence of salts, a particular problem in urine, although this can be surmounted by careful sample preparation and the gradient selection. Reproducibility can also be affected by carry over between samples with high-through put analysis for metabonomic applications.

The choice of mass analyser is another important consideration when applying mass spectrometry to biological samples. For the purposes of high throughput LC-MS metabonomic analysis and MS/MS fragmentation analysis, an orthogonal acceleration time of flight mass analyser was found to be most suitable for this application. Quadrupole mass analysers are robust, reliable and have a good dynamic range.

Quadrupoles separate ions based on their *m/z* ratio by passing the ions through an oscillating electric field. The linear quadrupole consists of four rods across which a voltage is applied by electrodes at either end of the four rods, through which the ions produced by ESI pass. The ions are confined between the rods by a 2D radio frequency field and by varying the voltage across the quadrupole rods brings the full range of ions based on their *m/z* to the detector.

Time of flight (TOF) mass analysers are higher resolution than quadrupole mass analysers and are also commonly used for many biological applications and in biofluid analyses. TOF mass analysers produce ions with the same kinetic energy. The ions are then separated based on the inverse relationship between the velocity of each ion to the square root of its mass and the time of flight of each ion will be proportional to its *m/z* ratio. The UPLC-MS data acquired and the MS/MS fragmentation of both urine and authentic standards utilised an orthogonal acceleration time of flight mass spectrometer. ESI is coupled to the TOF mass analyser by orthogonal extraction whereby the ions introduced are accelerated along an axis perpendicular to the initial direction of motion. This orthogonal acceleration provides a very high resolution. Prior to orthogonal acceleration the ESI produced ions are focused into a narrower beam by first passing through electrostatic lenses in high vacuum to minimise divergence. This is followed by collision with residual gas in multipole mass guides. The combination of collisional cooling and orthogonal acceleration allows a significant increase in resolution without a detrimental reduction in mass sensitivity [152].

When the separated ions in the mass analyser reach a detector, a small current signal is generated. This determines the associated *m/z* ratio of the ion reaching the detector. The mass analyser in a TOF mass spectrometer consists of a microchannel plate detector. The electrical signal generated is recorded by a time to digital converter (TDC). These register the arrival of single ions at discrete time bins which are able to discriminate between noise and ion arrivals. The Gaussian-shaped electrical pulses generated are converted into common-shape pulses. The TDC detector can count down to picoseconds resolution. However it suffers from a reduction in dynamic range, therefore TDC detectors are utilised in a multiple system where each single ion-hit event is placed in a time bin and the large number of single ion detection events in each time bin is summed up in a histogram providing a value for each peak.

2.7.2 Sample preparation for mass spectrometry analysis

For human urine UPLC-MS analysis a 50µl aliquot of the urine samples was centrifuged at 13,000rpm for 5mins in order to remove particulates. However, human urine in healthy individuals should contain no proteins or other macromolecules unlike rats which are known to be proteinuric. Centrifuged urine samples were then mixed with 50µl UPLC grade H_2O containing 0.1% formic acid and subsequently transferred to 96-well plates with cap mats in a random order with UPLC-grade H_2O blanks in position 1 and 96 of each plate with additional blanks every 7^{th} sample. Prepared 96-well plates were maintained at 0°C prior to transference to the autosampler for UPLC-MS analysis.

2.7.3 Mass Spectrometry Methodology

Mass spectrometry was performed on a LCT Premier™ (Waters MS Technologies, Manchester, UK) orthogonal acceleration time of flight (oa-ToF) mass spectrometer which was operated in both positive and negative ion electrospray modes. The desolvation gas was set to 900 L h^{-1} for positive mode and 500 L h^{-1} for negative mode at a temperature of 350 °C, the cone gas set to 0 L h^{-1} and the source temperature set to 120 °C. The capillary voltage and the cone voltage were set to 3200 V (positive mode)/ 2500 V (negative mode) and 30 V, respectively. Data from the LCT-Premier™ was collected in continuum rather than centroid mode for increased resolution. All analyses were acquired using the lock spray to ensure accuracy and reproducibility; leucine-enkephalin was used as the lock mass (*m*/*z* 556.2771) at a concentration of 250 pg/ul and flow rate 50 µl min^{-1}. Parameters are summarised in (**Table 2.1**).

Table 2.2 Summarising LCT Premier^{TM} parameters for both positive and negative ion modes.

Parameters	**Positive mode**	**Negative mode**
Capillary voltage	3200	2500
Sample cone voltage	30	30
Desolvation temperature (˚C)	350	350
Source Temperature (˚C)	120	120
Desolvation set at litres hr^{-1}	900	500
MCP	2100	2100

Chapter Three: Chemometric Methods

3.1 ^{1}H-NMR Spectral Pre-Processing

3.1.1 Data Reduction and Pre-Processing

Full resolution ^{1}H NMR spectroscopic data (δ10.0-0.24) were zero filled by a factor of 2 and all FIDs for the acquired spectra were multiplied by an exponential weighting function to a line broadening of 0.3 Hz before fourier transformation. Data processing was initially performed using NMRProc (in house software) which corrected the NMR spectra for phase shift, peak alignment and baseline distortion, referenced to TSP at 0.0δ and bucketed into reduced 0.001ppm spectral buckets by polynomial curve fit (using a Matlab Version 7.0 program developed in house). The raw spectra were then imported into Matlab for multivariate data analysis using MetaSpectra software. Spectra were then Spline cubic interpolated from 32k to 42k data points and the region corresponding to water (4.9-4.5ppm) was removed prior to peak alignment. Spectra were then normalised to total spectral area and alternatively to the probabilistic quotient for improved biomarker recovery. Matlab imported NMR data was also exported to SIMCA-P^{+} for comparative analysis

3.1.2 Recursive Segment-Wise Peak Alignment

Chemical shift variation in ^{1}H-NMR signals of biofluids reduces biomarker information recovery following multivariate statistical techniques [153]. Most peak alignment methods fail to maintain minor peak shift variations, however recursive segment-wise peak alignment (RSPA) reduces the peak shift variability across all peaks and ^{1}H NMR spectra by top down segmentation of reference and test spectra. Sequential sub-division of larger segments improves local spectral alignments across all spectra in a dataset. RSPA improves alignment accuracy by aligning NMR spin coupled multiplets as a whole instead of separate singlet peaks by local recursion.

3.1.3 Probabilistic Quotient Normalisation

Normalisation accounts for different sample dilutions by scaling ^{1}H NMR spectra to the same overall concentration; this produces more robust multivariate models. A major problem in human biofluid samples is reduction of inter-individual differences and maximising the contribution of subtle dietary changes. Integral normalisation is the preferred method of data normalisation and works well for more highly controlled

studies such as in genetically and environmentally homogenous animal studies. Large increases in metabolites confound integral normalisation particularly which can produce poor chemometric models and incorrectly scaled spectra. Probabilistic quotient normalisation calculates a most probable dilution factor by factoring in the distribution of quotients of the amplitudes of test spectrum when compared to a reference spectrum. Spectra normalised to probabilistic quotient have been found to be more successfully pre-processed for multivariate modelling than both integral and vector length normalisation methods [154].

3.2 Pattern Recognition

3.2.1 Principal Components Analysis (PCA)

PCA is an orthogonal linear projection which calculates eigenvectors with the data set, the first principal component (or eigenvector with the largest eigenvalue) within a data set is a coordinate onto with all data points are projected down corresponding to the maximum source of variation within the X matrix. PCA is used with the context of Metabonomic investigation as an initial dimensionality reduction of the data and is the initial linear transformation providing the subspace with greatest variation.

The principal components are a set of scores that highlight clustering, trends or outliers in the sample direction in the data matrix, which gives a quick visual analysis of the data in the form of a 2D scores plot, where each point on the plot represents an individual sample. The 'loadings' plot focuses on the variation in the variable direction and gives an indication of the influence of the variables on the scores. Each coordinate on the loadings plot represents one NMR spectral region. Non-systematic variations that cannot be explained by the model gives the residual or unexplained variation matrix.

PCA analyses were used primarily before any other data treatment to overview and explain clustering of sample and trends within the multivariate data and to identify any outlier samples. PCA is an 'unsupervised' chemometric method which uses no information other than the spectral data (X matrix) without any knowledge of variables in the Y data matrix such as class membership (i.e. dietary intervention).

3.2.2 R^2 and Q^2 values

It is important to distinguish in any multivariate model the difference between a models fit to the data and its ability to predict class membership. Typically in statistics the wellness of fit is provided by the parameter R^2, this Figure represents the amount of variation within the X matrix that is explained by the model. Unfortunately with a sufficient number of components within a model the R^2 can anomalously appear close to 1 the optimum value. The Q^2 on the other hand measures the predictive ability of the model and how well the X matrix can be predicted, this is measured using either the existing data set or an independent validation set.

The strength of the Q^2 value is that if many intrinsic variables are introduced into the data set, lack of correlation of these with patterns in the dataset deteriorates its predictive power. The number of components in a model is selected based on the combination of both the R^2 and Q^2 parameters. The Q^2 is calculated using cross validation within the model, a component's significance is determined compared to the rest of the components in the model. A test set and training set are generated and using an iterative process removed samples are predicted against the remaining samples in a model until every sample has been removed and tested against the remainder.

3.2.3 Partial Least Squares Analysis (PLS)

The PLS method provides a model of the maximum separation between defined class samples in the data; it differs from PCA in that variables in the Y data matrix (class, dietary intervention) are known and used to optimise the model. After calculation and validation of the PLS model it can be used for prediction of the class membership of unknown samples. The sources of variation in the X and Y matrix may disturb the PLS model causing the imprecise prediction of new samples and can affect the robustness of the model.

3.2.4 Orthogonal Projection on Latent Structures Discriminant Analysis (O-PLS-DA)

A more favourable method has been developed which avoids the two-step process of OSC followed by PLS called O-PLS-DA [155]. This statistical method incorporates the orthogonal variation partitioning in the algorithm. This method produces a far more

robust model that models any 'structured noise' separate from the data variation common to both the X and Y data matrices.

The O-PLS-DA model is based on unit variance scaled data and back-transformed by multiplying the O-PLS coefficients by their standard deviations. The O-PLS coefficients relate to the relative influence of the each coordinate of the loadings and its relative influence on the class discrimination. The O-PLS-DA provides an excellent model in which the loadings are plotted in a similar fashion to an NMR spectrum for ease of interpretation. On the same plot the O-PLS coefficients are colour coded, the range of which are set to the maximum and minimum of the unit variance scaled model weight. The colour of the coefficient in the back-transformed loadings plot represents the correlation of the X variables with Y and gives an indication of the influence of an individual metabolite in class discrimination.

3.2.5 STOCSY Statistical Total Correlation Spectroscopy

STOCSY offers a powerful chemometric method for identification of highly correlated peak intensities in this case within a set of ^{1}H-NMR spectra. STOCSY also confers an advantage over the comparable NMR experiment of Total Correlation Spectroscopy, which is unable to correlate resonances between protons in a molecule of interest, which are magnetically independent of one another, for example where resonances are a singlet. STOCSY operates by generating a pseudo-2D NMR spectrum that shows correlation among intensity variables across an entire sample of the various peaks in a set of spectra. As a rule of thumb, the method necessitates a minimum of 25 spectra in any one sample to class to be statistically significant [156]. Intermolecular correlations due to biological covariance or negative correlation can provide information on a metabolite of interest, with regards to homology of molecular structure or common metabolic pathway inter-dependency or subject to a common regulatory mechanism.

O-PLS-DA combined with STOCSY provides an excellent combination of techniques for metabonomic analysis. O-PLS-DA first identifies class discriminating spectral peaks; this information combined with STOCSY analysis helps to discern the statistically significant correlations within the sample class.

3.3 Statistical analysis of mass spectral data

3.3.1 MarkerLynx PCA analysis

UPLC-MS data in both negative and positive modes was aligned within mass and retention time windows and then normalised using the Waters MarkerLynx 4.1 software. This generated a matrix of *m/z* and retention time pairs with associated peak areas. This data was then processed using an unsupervised statistical approach, principal components analysis (PCA) initially with pareto scaling. Chromatograms were reduced to the (0-6mins) regions to include the more polar urinary compounds of urine in the analysis and to reduce data size and the required computing power. The scores plot produced gives an overview of potential class separation in the raw data and the loadings plot allows extraction of potential biomarker candidates which are statistically significant and within defined confidence limits. This also allows easier visualisation of the raw UPLC-MS data and detection of any outlying samples.

The Waters MassLynx 4.1 software was also used to extract chromatograms of interest based on the expected *m/z* of candidate biomarkers using a targeted approach. Chromatographic peaks exclusive to sample classes were considered as putative biomarkers.

3.3.2 XCMS[1] spectral feature identification

XCMS[1] incorporates nonlinear retention time alignment, matched filtration, peak detection, and peak matching; this software operates without the need to use internal standards. The method identifies hundreds of endogenous metabolites for use as 'internal standards', calculating a nonlinear retention time correction profile for each sample. This retention time correction allows the relative metabolite ion intensities to be directly compared to identify changes in specific metabolites such as potential biomarkers within the mass spectral dataset. The step by step XCMS[1] data processing method begins with filtering and identification of peaks within the raw mass spectral data, extraction of ion chromatograms, application of the model peak matched filter and then identification and integration of the peak areas. This is followed by matching the peaks across samples by segmentation of global peak lists by mass, identification of retention times with high peak density and the creation of groups in these dense regions. Simultaneously the retention times are corrected identifying groups for use as standards, calculation of retention time deviations and finally alignment of peaks using non-linear

warping. The missing peak data is subsequently filled identifying groups missing samples, calculation of the mass/time peak regions and integration of intensity from the raw data. These results are then statistically analysed revealing a visualisation of the statistically important peaks in the form of an extracted ion chromatogram and box and corresponding box and whisker plot.

3.3.3 Application of XCMS[1] for Tentative Biomarker Identification.

A significant challenge of interpretation of UPLC-Mass Spectral data is its sheer size which makes manual analysis of raw data almost impossible. The use of chemometric techniques particularly in the form of software packages from research groups, such as XCMS[1] [157] are useful for the discovery of biomarkers.

Data-mining programs such as XCMS[1] use special algorithms that filter and bin raw mass spectral data then extract chromatographic features identified by *m/z* ratio and retention time. The features are then aligned, normalized and a peak table created which can be subjected to further statistical analysis. XCMS[1] uses a peak picking algorithm that identifies chromatographic features that represent the differences between sample classes and produces a list of Extracted Ion Chromatograms (EICs) and box and whisker plots and sorted in ascending order of confidence based on *p*-value and other statistical measures. Tentative biomarkers of cruciferous vegetable consumption were selected based on their class exclusivity within the mass spectral data and by interpretation of the EIC, box and whisker plot, *p*-value and presence within the raw data. The purpose of this process is to identify ions of interest on which the investigations can focus attention as a possible source of biomarker information.

Initial biomarker identification is based on knowledge of potential exogenous and endogenous candidate metabolites, database mining, (i.e. Human Metabolome Database, Hmdb, *Genome Alberta* www.hmdb.ca) fragmentation analysis and accurate mass atomic formulae identification or a combination of these. Final confirmation is achieved using authentic standard.

Chapter four: The Effects of the Naturally Occurring Food-Derived Genotoxic Carcinogen Methyl Eugenol on Rat Metabolic Phenotype

4.1 Background

Methyl Eugenol (ME) is a member of the alkenylbenzene derivatives family and a natural component of many plants, fruits and spices. It is most abundantly found within the herb basil which is widely consumed in the Mediterranean. It is widely used in the food additives and cosmetics industries as a flavour enhancer and fragrance component respectively (ICEM 2005). The ubiquity of ME's use and its similarity in structure to known carcinogens estragole and safrole has generated considerable interest in its potential carcinogenicity and is the subject of great environmental toxicological interest. ME has been identified as a genotoxin and hepatocarcinogen and been shown to form DNA adducts at high doses in rats and as such may be a potential human carcinogen. A previous 2 year NTP (national toxicology programme) study demonstrated significant gastric and hepatic carcinogenicity of ME (NTP 2000). However, the study may have overestimated the carcinogenic potential of ME by using a dosage far in excess of normal human levels of exposure and administration by orogastric gavage rather than using a food matrix method.

This study utilised urine samples for metabonomic analysis from a 28 day feeding study with F344 strain rats at relatively low ME doses of 1, 5, 50 and 50 (administered by gavage) mg/kg bw/day with appropriate controls to assess effects of a low ME feeding regimen on the global rat metabolic phenotype. This dietary regimen was designed to reflect more realistic levels of human ME exposure (~1mg/kg bw/day) and by encouraging a more natural dietary administration.

4.2 Aims and Objectives

- To determine the effect of ME dietary exposure at exposure levels (1-50mg/kg bw/day) relevant to human dietary exposure (~1mg/kg bw/day) on the rat urinary metabolic phenotype.
- To examine dose related responses (1, 5 and 50mg/kg bw/day) of the rat metabolic phenotype in a 28 day repeat dose oral ME study.
- Identify potential urinary biomarkers of ME exposure and any metabolic pathways modified.
- To investigate gender-specific differences on ME metabolism.
- To assess any effect of route of administration by either orogastric gavage or dietary matrix on the rat metabolic phenotype.

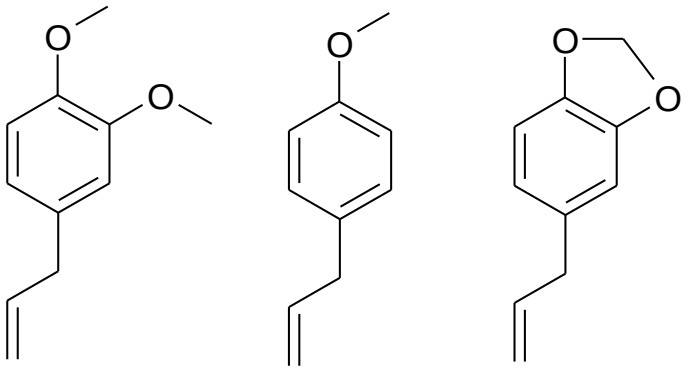
4.3 Introduction

4.3.1. Methyl Eugenol-Chemical Properties

ME is also found in many commonly consumed fruits, vegetables and spices eg. basil, bananas and nutmeg and is also present as an environmental contaminant. ME is extensively used as a flavouring agent in many processed foods such as soft drinks, chewing gum and ice creams where it imparts a cooling and slightly bitter flavour, which is remiscent of the flavour of basil. It is also used as a fragrance in perfumes, laundry detergents and soaps, and as an insect attractant and rodent anaesthetic [158].

ME was identified as a toxicological concern and investigated in a 2 year NTP toxicology study due to its structural similarity to safrole, isosafrole and estragole which are known carcinogens. The study confirmed that ME is a rodent carcinogen inducing tumours of the fore stomach and liver (NTP 2000).

ME is an alkenylbenzene derivative, which consists of a benzene ring with a 1-allyl side chain and methoxy groups at *meta*- and *para*- ring positions. It is a hydrophobic compound and is highly insoluble in water but is miscible in organic solvents such as chloroform and DMSO. In common with the known carcinogens estragole and safrole, ME contains the allyl side chain with a terminal carbon – carbon double bond (**Fig. 4.1**). It is this side chain that is thought to be responsible for their genotoxicity



methyl eugenol estragole safrole

Fig. 4.1 Molecular structures of ME, estragole and safrole.

4.3.2 Toxicological evaluation of Methyl Eugenol

Median lethal oral doses of ME are reported to be 810 to 1560 mg/kg bw in rats (NTP 2000), therefore ME is considered to possess low toxicity as such doses are far beyond doses relevant to normal dietary exposure in humans (estimated to be ~1mg/kg bw/day). The primary issue of concern relates to carcinogenicity rather than acute toxicity.

In vitro genetic toxicology assays show little to no activity with ME, however positive results were obtained using unscheduled DNA synthesis (UDS) assays [58]. Negative results were observed for ME, estragole and safrole in the Ames Salmonella reversion test and the *E.coli* WP2 uvrA reversion assay [159]. However, ME has been shown to demonstrate intrachromosomal and interchromosomal recombination events in *Saccharomyces cerevisiae* [160].

ME presents negative results in most *in vitro* genotoxicity models but in *in vivo* models ME appears to have a more pronounced effect. In the Big Blue transgenic female rat and male mouse ME significantly increases mutation frequency of the *lacl* gene when dosed with high doses (300-1000mg/kg bw/day) for a period of 90 days by gavage [161].

It is believed that ME produces a genotoxic effect in both *in vivo* and *in vitro* experimental models in a similar fashion to that of safrole and estragole by DNA adduct formation. Safrole and estragole are known to form 3 different adducts at the exocyclic position N^2 of guanine and also at the N^6 of adenine [162]. ME is also known to form adducts with hepatic proteins *in vivo* in F344 rats and *in vitro* hepatocyte cell

cultures[163]. ME has been shown to form DNA adducts in the livers of treated rats [164]. This was observed in a dose-dependent manner using ELISA, however the biological significance of the ME-protein adducts is unclear as ME has little cytotoxicity or apoptotic effect and the proteins affected are therefore unlikely to be involved in cancer-related processes such as cell cycle, cell growth or apoptotic mechanisms.

The toxicity and carcinogenic potential of ME has been examined in a 2 year NTP bioassay (NTP 2000). This study examined high doses of ME (37mg-300mg/kg bw/day) *via* an orogastric gavage delivery system. This dosage regimen produced extensive gastric and hepatic genotoxicity, hyperplasia, neoplasia and tumourigenesis for all of the doses examined. The dosing regimen was selected according to accepted bioassay guidelines and to ensure toxicity. In general human exposure is estimated to be much lower (~1mg/kg bw/day) than the doses used in the bioassay and according to [64] the lowest dose examined is 370 times higher than the estimated dose for "specialised" or super eaters of ME. Human exposure occurs mainly through naturally occurring ME within the matrix of many fruits and vegetables and through exposure to food and cosmetic additives and collectively would likely never achieve acutely carcinogenic levels in the normal human diet. In the NTP study animals were also dosed by gavage which would have undoubtedly increased the rate of ME absorption and potentially its site-specific genotoxicity in the liver following rapid clearance of this lipophilic compound from the stomach. Non-neoplastic treatment-related changes seen in the liver such as inflammation and hyperplasia and extensive damage of the stomach would also certainly have exacerbated hepatic and gastric tumourigenesis by altering ME absorption and metabolism [165].

Here I have utilised a more conservative dosing regimen than the NTP study with a lower range of ME doses between 1-50mg/kg bw/day over a 28day repeat dose period in an attempt to achieve more realistic human levels of exposure in order to evaluate the true potential risk of ME and to probe further potential toxic mechanisms.

4.3.2 ME Metabolism

It was initially hypothesised that the high structural similarity of the alkenylbenzene derivatives ME, estragole, safrole and isosafrole would imply shared metabolic, pharmacokinetic and toxicological properties [64].

Following oral administration of ME by gavage, absorption is rapid; with peak serum concentration within 5mins and ~95% of the dose recovered in urine after 24 hours. Therefore urinary excretion represents the predominant route of elimination of ME from the rat with some evidence of conversion to CO_2 and no unchanged ME is found within rat urine [166].

Metabolism of ME is primarily hepatic and occurs by three main routes and these are believed to be common amongst the alkenylbenzene derivatives (**Fig.4.2**). Two of these pathways of metabolism are not thought to produce procarcinogenic metabolites. At low doses the primary route of ME metabolism is *via O*-demethylation with *m*- or *p*-hydroxylations of the benzene ring. This is followed by excretion by sulphation or glucuronic acid conjugation. Further reduction of the allyl side chain of demethylated ME to quinine methide followed by glutathione conjugation is also known to occur along the same pathway [167].

It is known that only at higher doses of ME are the other two routes of metabolism the epoxidation and 1'-hydroxylation pathways significant [168]. It is anticipated that with the lower doses used within this study (1-50mg/kg bw/day) that the main urinary metabolites of ME would be the sulphate or glucuronic acid conjugate of eugenol or 5-allyl-2-methoxy phenol. As such these metabolites represent potential candidates for metabonomic biomarkers of ME exposure within rat urine.

Eugenol is less genotoxic than ME but is more cytotoxic. It is rapidly eliminated by glucuronidation and sulphonation at the *para*-hydroxy group, and has also been shown to conjugate with glutathione by formation of the highly reactive quinine methide [169].

At higher doses ME is reported to be metabolised by two further routes of metabolism within the rat liver [170].

The second route of metabolism of ME is by epoxidation of the allyl side chain forming the ME-2',3'-epoxide which is eliminated by conjugation with glutathione (**Fig. 4.2**).
The third route of ME biotransformation is by 1'-hydroxylation by a member of the phase I cytochrome P450 superfamily, which has been shown to be the primary intoxication pathway (**Fig. 4.2** pathway 3). The 1'-hydroxymethyl ME is sulphated by a sulphotransferase at the hydroxyl group and the sulphated 1'-hydroxymethyl ME forms a reactive intermediate carbonium ion. It is this reactive intermediate that is thought to be responsible for the genotoxic and cytotoxic effects of methyl eugenol by reacting with DNA/RNA and protein [64].

This current study examines whether the exposure of rats to ME at doses that are more representative of human exposure alters the urinary metabotype.

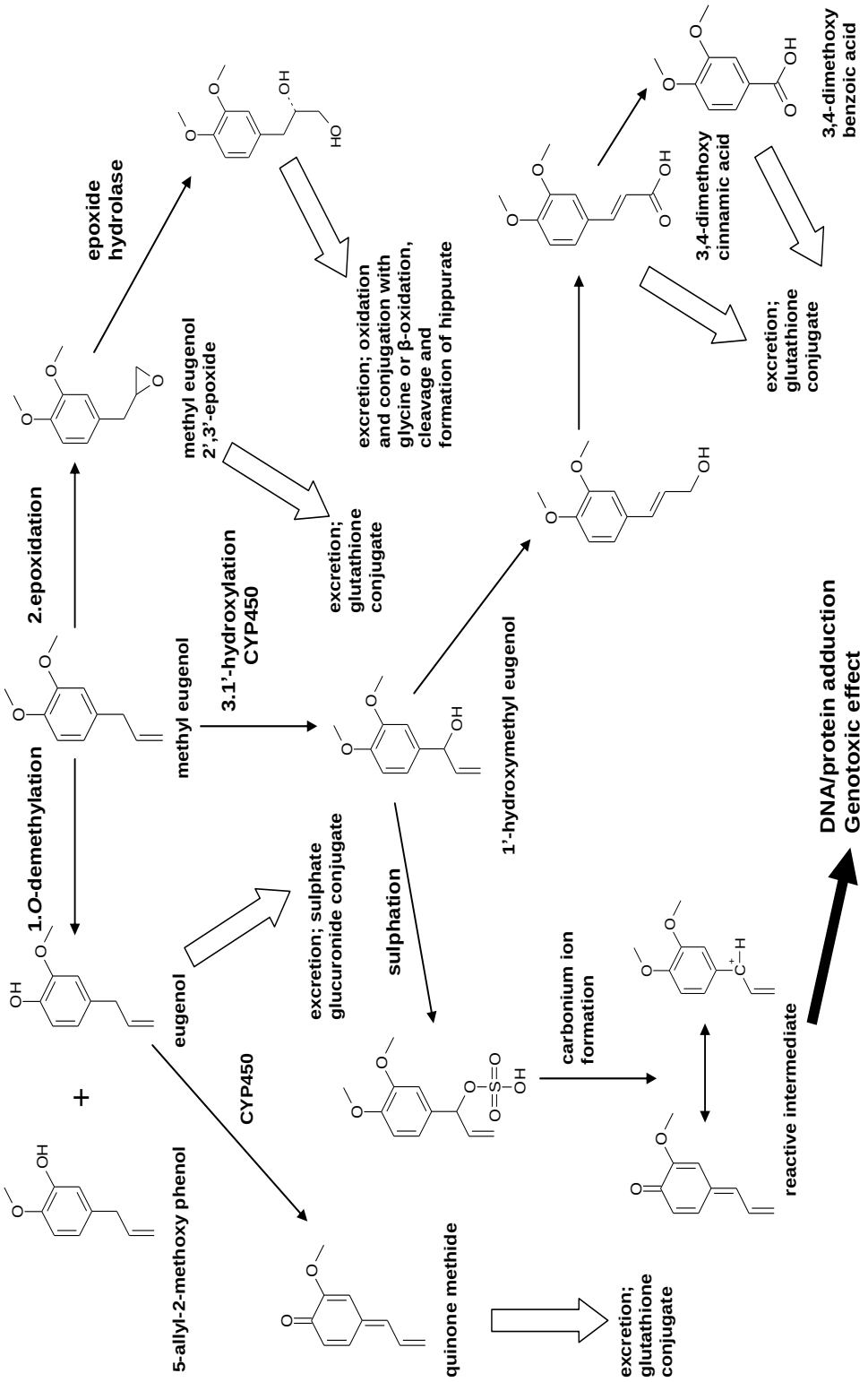


Fig. 4.2 Metabolic scheme of ME in rodent liver**.** The primary route of ME excretion at low dosage is by *O*-demethylation and suphation or glucuronic acid conjugation of either eugenol or 5-allyl-2-methoxy phenol (pathway 1). Metabolic scheme adapted from [64]

4.3.3 Histopathological, Biochemical and Molecular Biological Findings

Toxicological properties of ME using a low dosage regimen were assessed by Dr J. Ellis, Dept. Of Biomolecular Medicine (BMM), Imperial College (IC) [165]. Body weight gain, liver weight and blood/urine biochemistry were all recorded to identify overt evidence of toxicity, however little evidence of this was observed. Proliferative, atrophic and inflammatory effects of ME on the liver and gastric mucosa were assessed with immunoblot assay/proliferating cell nuclear antigen (PCNA) enzyme-linked immunoabsorbent assay (ELISA) and a competitive binding radio immunoassay (RIA) for gastrin.

ME dosage between 1-50mg/kg bw/day *via* dietary dosage and the 50mg/kg bw/day *via* gavage over 28 days did not produce any overt signs of toxicity, with no treatment related effects on body or liver weight [165]. Histopathological findings showed no macro- or microscopic lesions in any organs. The blood and urine biochemical analysis did not identify any toxicological effects of ME with this dosing regimen. ELISA and competitive binding RIA failed to identify any proliferative effect in the liver or inflammatory or atrophic effect on the gastric mucosa. Thus no gross toxicological aberrations were observed using a dosing regimen more in line with realistic human exposure levels [165].

Analysis of the expression levels of hepatic cytochrome P450 enzymes (1A2, 2E1 and 2C6) found no significant alteration in the dose range 1-50mg ME /kg bw/day of this 28 day study dietary matrix or gavage [165]. However it is known that at higher dose ranges (100-300mg/kg bw/day) by *intra peritoneal* injection CYP1A2 protein expression is increased [171]. This is thought to be one of the major enzymes involved in 1'-hydroxylation intoxication pathway of ME in the rat and human liver. The 28 day dietary regimen therefore did not markedly induce any hepatic detoxification enzymes, suggesting that the effect of ME was subtle.

The molecular biology findings did however identify a dose-dependent quantifiable effect on DNA adduct formation utilising a ^{32}P-postlabelling assay in both the liver and stomach. This genotoxic effect was evident at doses of 5mg/kg bw/day and greater in

the liver and at 50mg/kg bw/day in the stomach at a detection limit of 1 DNA adduct in 10^7 bases [165]. The higher level of DNA-ME adducts in the liver than in the gastrum was expected as the liver is a more biologically active organ and CYP450 enzymes and a sulphotransferase are required for the intoxication pathway [171] and these are found in greater concentration in the liver. It was found the dietary dosing regimen produced more pronounced effects at the high dose (50mg/kg bw/day) in the stomach than did the gavage treatment. The constant dosing *via* feeding of ME was thought to provide a continual exposure as against a one-shot gavage dosage in which the highly lipophilic nature of ME would mean rapid clearance from the stomach and a saturation of the intoxication pathway in the liver.

Gene fishing-polymerase chain reaction (GF-PCR) analysis identified differentially expressed genes in the livers of rats dosed with the high dosage feeding regimen (50mg/kg bw/day). This molecular biology technique uses primers to amplify and quantify differentially expressed genes [165]. There was also an increased expression of specific genes within the male rats compared to females, and suggests gender difference in the response to ME. Unfortunately the gene expression findings were not confirmed by protein expression analysis. Systemic change at the genomic, transcriptomic or proteomic level can in theory cause potentially significant perturbation of the metabolic phenotype.

In summary the genotoxicity and molecular biological findings associated with the ME treatment provided impetus for further complimentary investigation by metabonomic analytical techniques. The histopathological and biochemical findings did not demonstrate any gross toxicological effects in the dosage range (1-50mg/kg bw/day) and thus strong perturbations of homeostatic control in the experimental animals was unlikely. However, the subtle genotoxic effects seen in DNA adduct formation in the liver and stomach and differential gene expression in the liver, particularly in male animals, is potentially detectable using the highly sensitive analytical platforms of metabonomic analyses.

4.4 Methods

4.4.1 Animal husbandry and twenty-eight day study outline

One hundred and ten Fischer 344/NH/sd strain rats were supplied by Harlan UK ltd (Oxford) (55 male and 55 female) were acclimatised for 6 days prior to commencement of the ME dosing regimen and were approximately 6 weeks old at the start of the experiment. The animals were randomly allocated to control or the 1mg, 5mg or 50mg/kg bw/ day dosage groups and the control or 50mg/kg bw/day gavage groups (**table 4.1**). Animals were housed with 5 per cage of the same sex in a single air-conditioned room. The study was performed by Safepharm in a Barrier-Maintained Rodent Facility.

ME microencapsulated into a maltodextrin matrix was supplied by FEMA (Flavour and Extracts Manufacturers Association). Food was prepared by mixing with the microencapsulated ME and stored at 4ºC. Analysis of the microencapsulated matrix revealed a load of 3.76% ME oil.

All animals dosed with ME or vehicle controls were treated once a day for each of the 28 days of the study until the necropsy date on the 29th day.

Table 4.1 28 day ME F344 Rat Study animal numbers, treatment groups and dosage regimen

Treatment Group	Dose Route	ME Dosage	No. of animals	
		mg/kg bw/day	male	female
Dietary untreated control	Dietary	0	5	5
Dietary vehicle control*	Dietary	0	5	5
Dietary 1mg/kg bw	Dietary	1	10	10
Dietary 5mg/kg bw	Dietary	5	10	10
Dietary 50mg/kg bw	Dietary	50	10	10
Gavage vehicle control	Gavage	0	5	5
Gavage 50mg/kg bw	Gavage	50	10	10

*Vehicle-maltodextrin matrix (no ME) mixed with food.

4.4.2 Urine sample collection

All animal work was carried out by SafePharm laboratories (Derby,UK). Urine samples were taken from all animals in each control and dose group on the day indicated (**table. 4.2**). Urine samples were collected in metabolism cages overnight from all dose groups and control animals on a weekly basis with water but restricted food access. Four hundred and fifty urine samples were snap frozen and sent to the department of Biomolecular Medicine, Imperial College London on dry ice at a temperature of -20ºC. Upon receipt the urine samples were maintained at -80ºC.

Table 4.2 28 day ME F344 Rat Study showing urine sample collection time-points and numbers.

Treatment Group	Urine collection time point							
	day 7 (no. of samples)		day 14 (no. of samples)		day 21 (no. of samples)		day 28 (no. of samples)	
	male	*female*	*male*	*female*	*male*	*female*	*male*	*female*
Dietary untreated control	5	5	5	5	5	5	5	5
Dietary vehicle control*	5	5	5	5	5	5	5	5
Dietary 1mg/kg bw	10	10	10	10	10	10	10	10
Dietary 5mg/kg bw	10	10	10	10	10	10	10	10
Dietary 50mg/kg bw	10	10	10	10	10	10	10	10
Gavage vehicle control	5	5	5	5	5	5	5	5
Gavage 50mg/kg bw	10	10	10	10	10	10	10	10

Total: **440**

*Vehicle-maltodextrin matrix (no ME) mixed with food.

4.4.3 Urine preparation and ^{1}H-NMR spectroscopic analysis

Rat urine samples were prepared for ^{1}H-NMR analysis. Aliquots of 200 μL of phosphate buffer solution (0.2 M Na_2HPO_4/0.2 M NaH_2PO_4, pH 7.4) were added to 400 μL of rat urine. The resulting solutions were left to stand for 15 min before centrifuging at 10,000 g for 15 min, to separate solids. Aliquots of 500 μL of the resulting supernatants were added to a 96-well plate containing 50 μL of a solution of 0.1% (w/v) sodium 3-trimethylsilyl-[2,2,3,3-$^{2}H_4$]-1-propionate (TSP) chemical shift reference (δ = 0)/1% (w/v) sodium azide (bactericide) in deuterium oxide ($^{2}H_2O$) for a field frequency lock.

Ninety-six well plates were placed into a Gilson sample handling system for automated injection of urine samples into the spectrometer. Single-pulse ^{1}H NMR spectra of urine samples were measured at 600.13 MHz on a Bruker DRX 600-MHz spectrometer at 300 K using a standard excitation sculpting pulse programme to remove the water peak [172]. Spectra were acquired using 64 free-induction decays (FIDs) into 64 K data points, with a spectral width of 9615.4 Hz, an acquisition time (t_1) of 3.41 s, a mixing time (t_m) of 100 ms, and a total pulse recycle delay of 6.41 s. Prior to applying Fourier transform (FT), the FIDs were multiplied by an exponential function corresponding to a line broadening of 0.3 Hz.

4.5 Results

A representative ^{1}H-NMR urine spectrum for a Fischer 344/NH/sd strain rat is shown in (**Fig. 4.3)**. Typically, control urine spectra were dominated by resonances from normal urinary constituents including glucose, taurine, creatinine, hippurate, citrate, 2-oxoglutarate, succinate and trimethylamine-*N*-oxide. A complex region of the spectrum revealed resonances deriving from amino acids, and organic acids. Creatine, acetate, acetoacetate, dimethylglycine, *N*-acetyl glycoprotein fragments and *m*-hydroxyphenylpropionate (*m*-HPPA) were also identified in the urine spectra. Although it was expected that in the high dosage (50mg/kg bw/day) animals urinary metabolites of ME may have potentially been identified, namely the sulphate or glucuronide of eugenol or 5-allyl-2-methoxy phenol, visual comparison of control and the 50mg/kg bw/day dose animals failed to identify significant difference between groups in the phenolic region of the spectra.

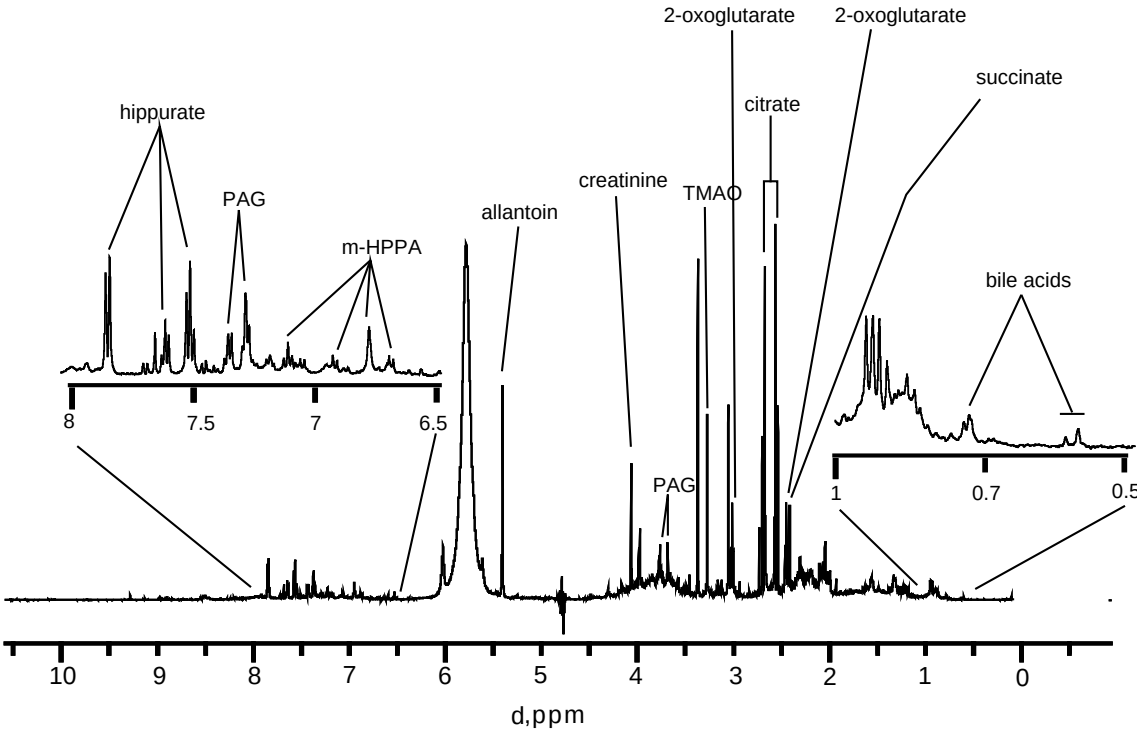


Fig. 4.3 Representative 600 MHz ^{1}H NMR spectra of urine from a female Fischer 344/NH/sd strain rat.
TMAO: trimethylamine-*N*-oxide; *m*-HPPA: *meta*-hydroxyphenylpropionic acid; PAG: phenylacetylglycine.

4.5.1 Principal Components Analysis (PCA) all samples

An unsupervised overview of the ^{1}H-NMR spectra of all samples from all time-points and dose groups revealed homogeneity in the data set with no obvious grouping of data based on sample class, with variation in all components dominated by inter-subject variations (**Fig.4.4**). The model demonstrated a high cross validation parameter Q^2:0.43 but relatively low R^2:0.56 by the second principal component and only reaching a cumulative value of 0.7 by the 7th principal component with a very gradual rise toward 1. The R^2 value suggests that total variation in the data set is described by many principal components, however the fit of data to the model is satisfactory for urinary data. It can be inferred that the variance due to dose group and dosage regimen is underlying inter-subject variations.

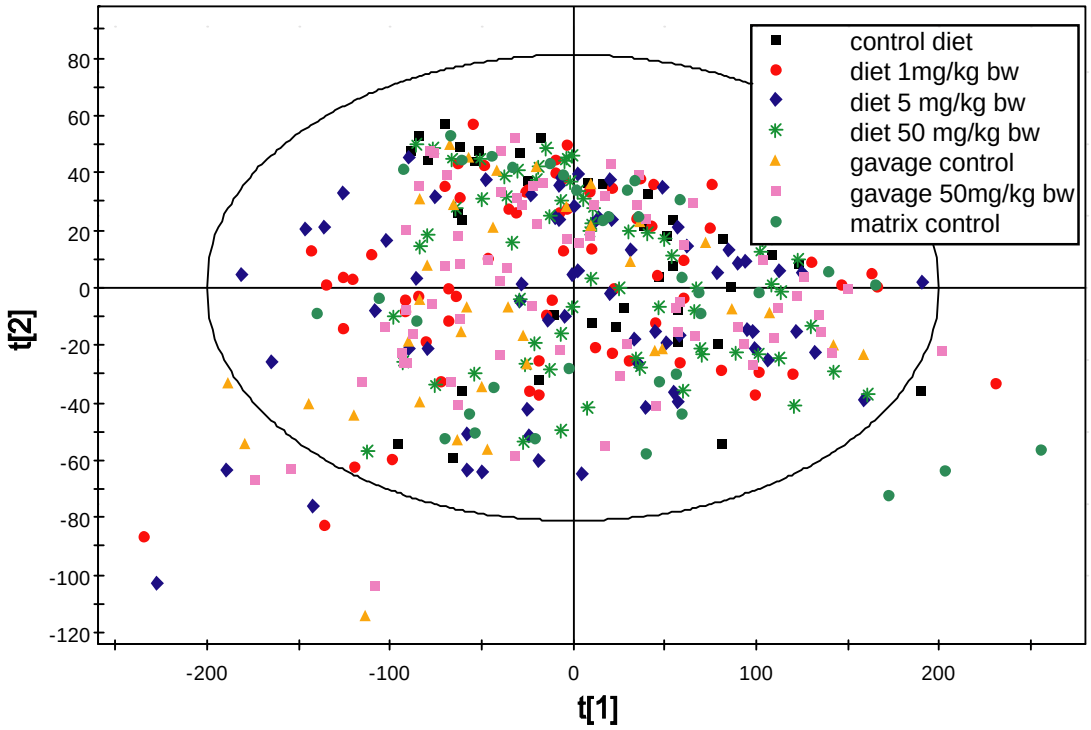


Fig. 4.4 PCA scores plot showing *t*[1]/*t*[2] principal components revealing an overview of the ^{1}H NMR spectra of rat urine samples from all dosage groups and dosage regimens showing no clustering of samples based on class. (Q^2:0.43 R^2:0.56).

4.5.1.1 Pairwise PCA and Partial Least Squares Discriminant Analysis (PLS-DA)

PCA calculated on the entire dataset revealed gender based clustering in the first two principal components, however there was evidence of overlap of male and female samples in the scores plot. This suggests that there are underlying sources of variation related to inter-subject variation in the dataset. To augment separation PLS-DA modelling improved gender based separation but the influence of inter-subject variations could still be seen (**Fig. 4.5**). The female rat urine samples showed evidence of tighter clustering and therefore greater homogeneity.

Analysis of the PLS-DA loadings plot revealed variables corresponding to urinary metabolites identified as 2'-deoxycytidine and glycerophosphocholine (GPC) discriminatory of male samples and loadings corresponding to TMAO and axial C_{18} protons of bile acids discriminatory of female samples.

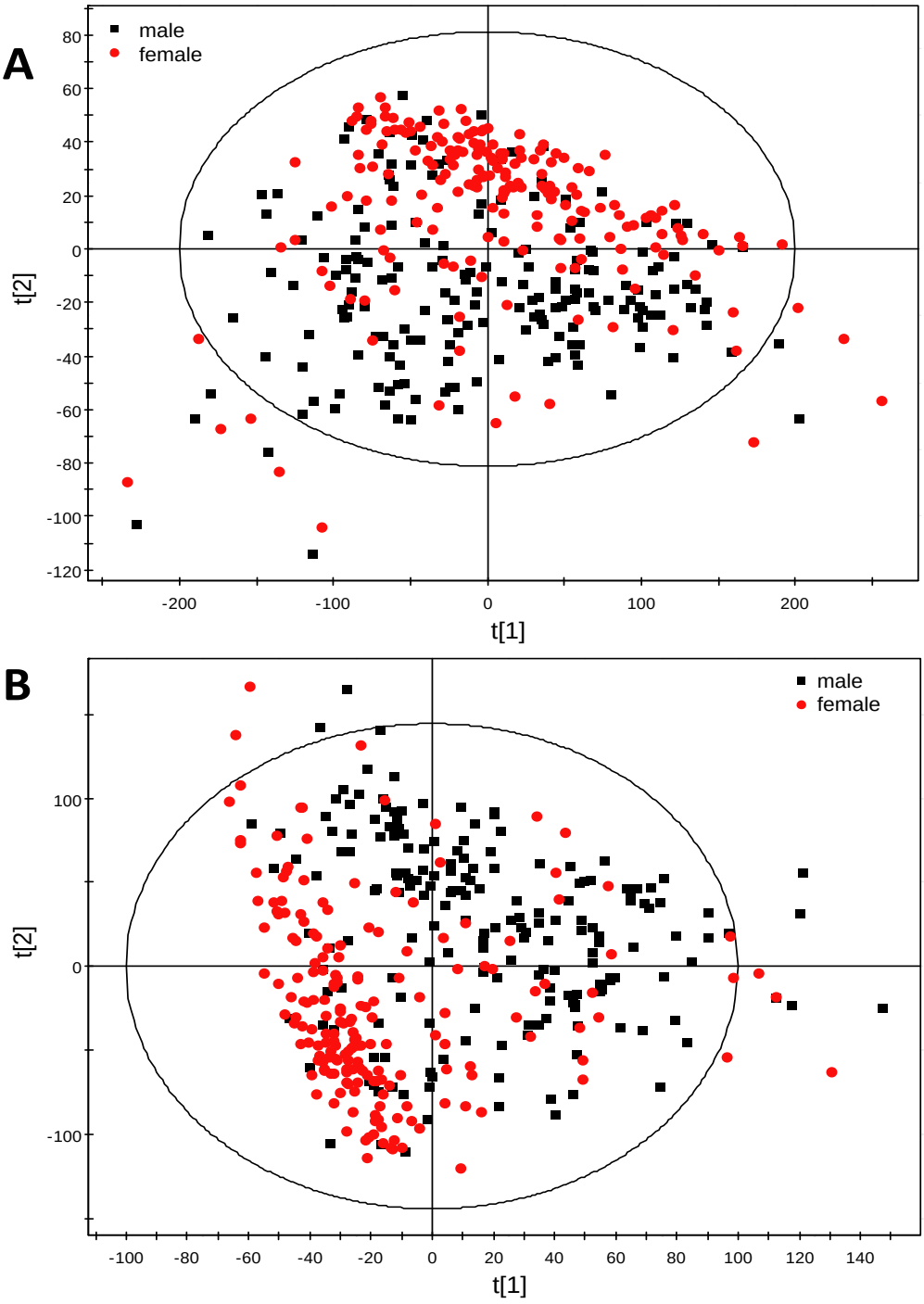
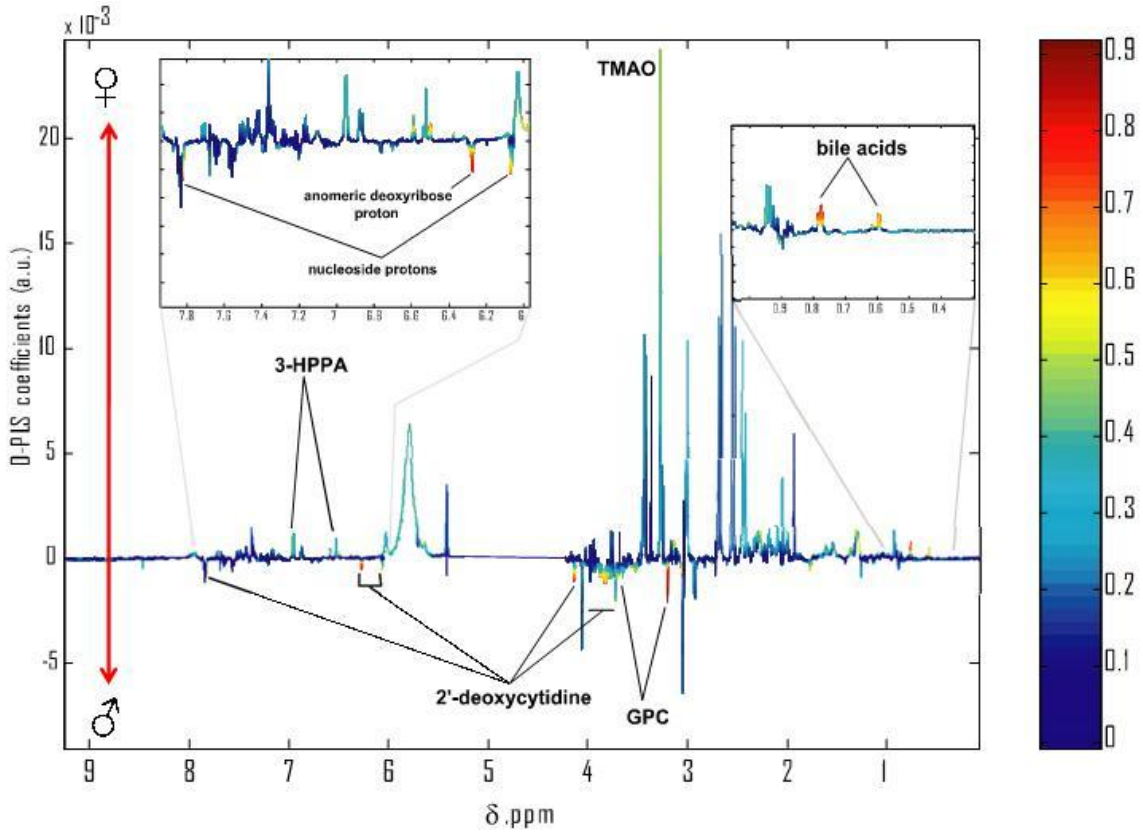


Fig. 4.5 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.43 R^2: 0.56) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.28 R^2: 0.32) from the ^{1}H-NMR spectra of all rat urine samples showing clustering of samples based upon gender. Unit variance scaled.

4.5.1.2 Orthogonal-Partial Least Squares Discriminant Analysis (O-PLS-DA) gender related metabolic differences

Analysis of all samples in all dose groups and time-points provided evidence of sexual dimorphism within the data set. Examination of the O-PLS-DA loadings plot identified statistically significant biomarkers related to gender. Modelling of gender related differences was particularly robust for biological matrices reflected in high $R^2\hat{Y}$: 0.87 (explanation of total variance) and $Q^2\hat{Y}$: 0.8 (cross validated predictive ability) values. Metabolites diagnostic of male rats were increased excretion of GPC and a deoxyribonucleoside, 2'-deoxycytidine identity was confirmed by STOCSY analysis and comparison to database reference spectra. Metabolites diagnostic of female rats were TMAO, the chlorogenic acid metabolite 3-hydroxyphenylpropionic acid and the axial C_{18} methyl protons of bile acids (**Fig. 4.6**).



-

Fig. 4.6 O-PLS-DA loadings plot showing urinary metabolic differences between male and female F344 strain rats all samples.

Clear differences are seen between the metabolic phenotypes of male and female rats with upward pointing metabolites diagnostic of females and downwards pointing metabolites diagnostic of males. ($Q^2\hat{Y}$: 0.73; $R^2\hat{Y}$: 0.96).

Statistical significance of the O-PLS-DA coefficients is indicated by the colour with red: strong correlation and blue: weak correlation.

4.5.2 The effect of high dosage methyl eugenol dietary intervention

In order to ascertain any effect of ME exposure on the rat urinary metabolic phenotype following the 28 day dietary regimen, chemometric methods were used to compare the high 50mg/kg bw/day dosage group with the vehicle control diet group at all timepoints.

4.5.2.1 Pairwise PCA and PLS-DA high ME dietary exposure vs. vehicle controls

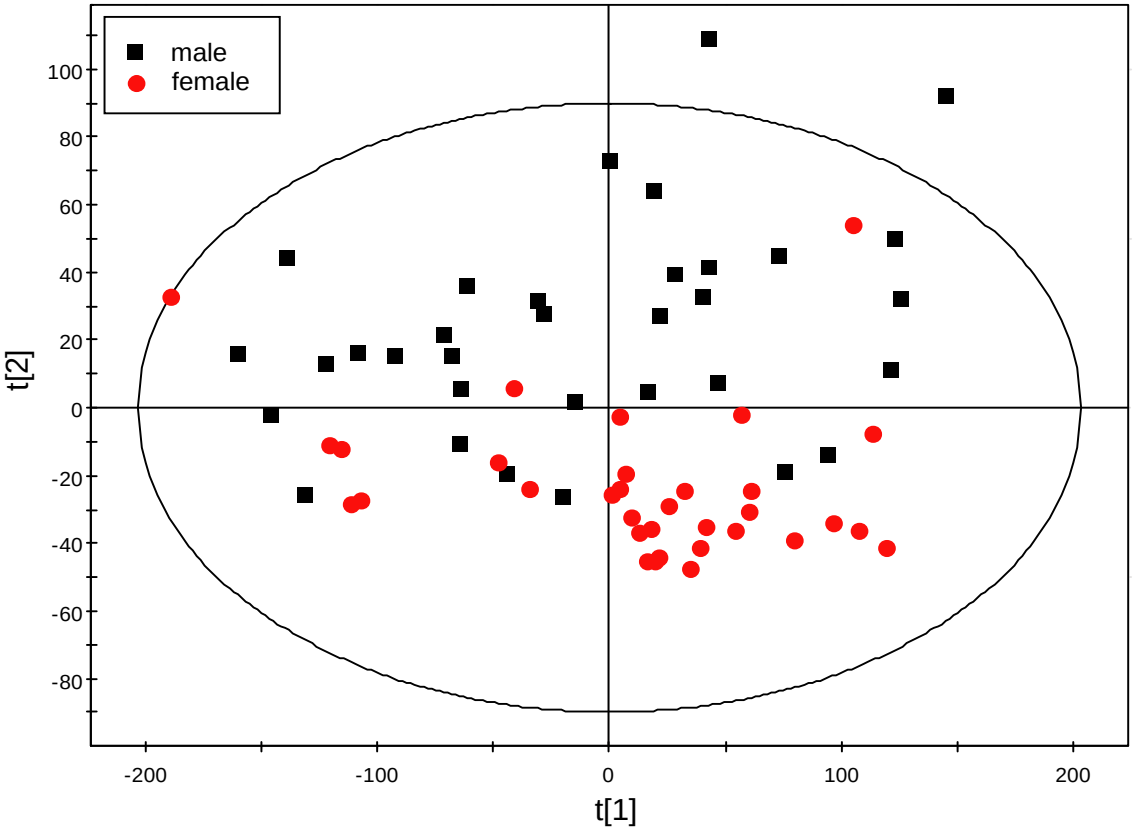


Fig. 4.7 PCA scores plot showing *t*[1]/*t*[2] principal components of urinary ^{1}H-NMR spectral data from the vehicle control diet and 50mg/kg bw/day ME dosage regimen samples in all time-points.

Clustering can be seen based on gender rather than dosage group indicating that the metabolic effects of ME, if any, are subtle. Unit variance scaled (Q^2: 0.39 R^2: 0.46).

Pairwise PCA performed on unit variance scaled data of the vehicle control diet and 50mg/kg bw/day dosage group samples at all time-points showed that variation in the first two principal components (PC1–PC2) was gender-based predominantly along the second principal component (**Fig.4.7**). Clustering could also be seen based on time point along the first and second principal component. However there was no evidence of clustering of samples based on dietary class. Analysis of the loadings plot identified specific loadings attributable to TMAO and GPC which discriminated female from male rats respectively along the first principal component. Loadings attributable to dietary time-point were widely scattered and therefore it was difficult to identify urinary metabolite changes responsible for the grouping seen in the scores plot along the second principal component.

In order to prevent incorporation of variation in the model due to the known sexual dimorphism in experimental animals, statistical modelling was performed on each sex separately [173]. In order to remove variation related to gender and time-point bias, PCA was recalculated on a smaller subset of the dataset. Male day 28 samples of the vehicle control diet and 50mg/kg bw/day ME diet groups were first analysed by PCA which revealed some evidence of clustering based upon ME dietary exposure along the first principal component (**Fig 4.8**).

This separation was enhanced using PLS-DA modelling which achieved improved separation between the 50mg/kg bw/day ME dietary exposure regimen and the vehicle control diet in male F344 strain rats in the first two PLS-DA components (**Fig 4.8)**. Analysis of the PLS-DA loadings plot revealed variables corresponding to unidentified urinary metabolites between δ3.8-3.9 responsible for discrimination between the control and ME dose groups.

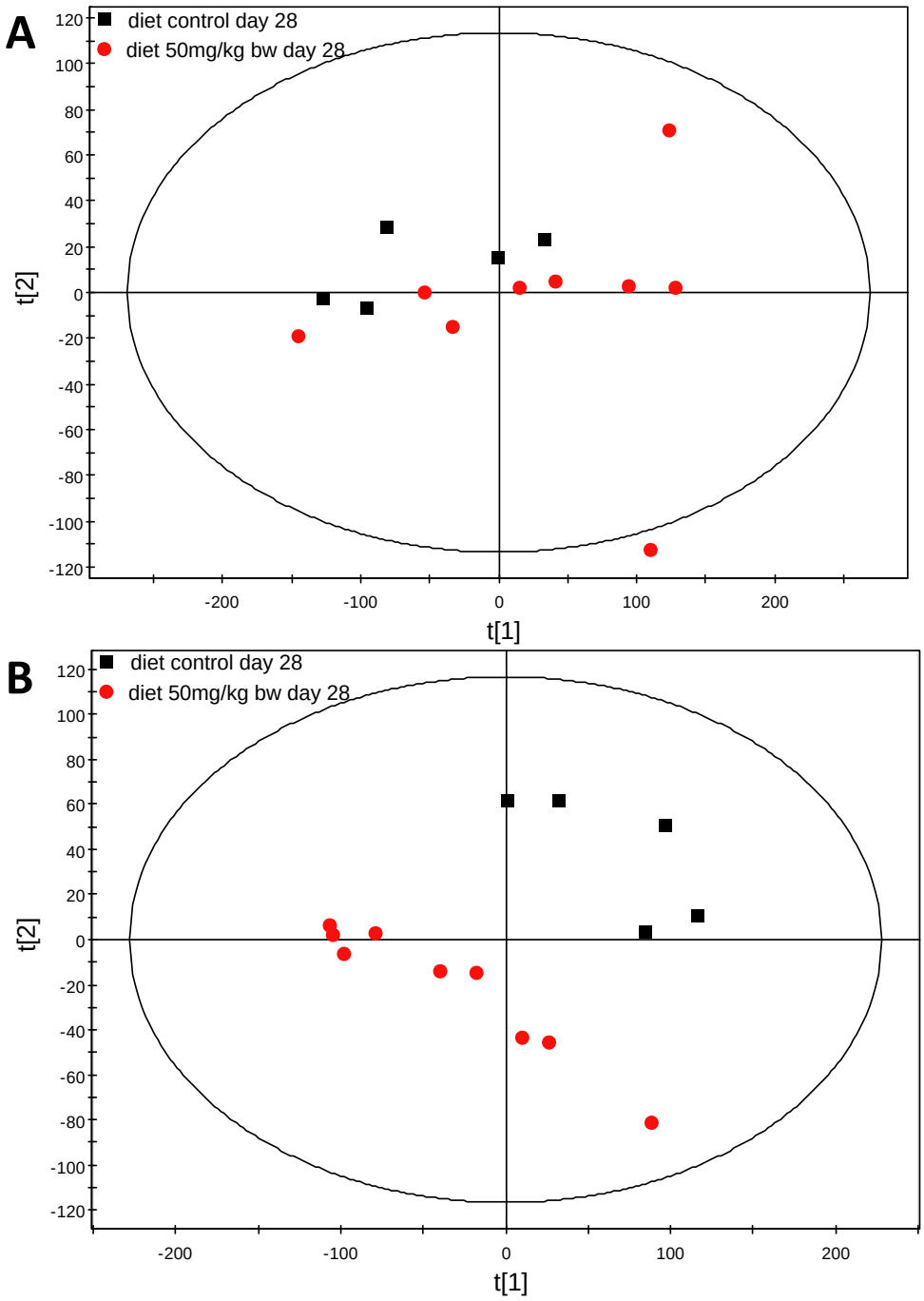


Fig. 4.8 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2:0.36 R^2:0.73) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2:0.79 R^2:0.94) from the ^{1}H-NMR spectra of vehicle control diet and 50mg/kg bw/day ME dosage groups day 28 rat urine samples showing clustering of samples based on ME exposure. Unit variance scaled.

4.5.2.2 O-PLS-DA male rats day 28 high ME vs. Day 28 vehicle controls

An O-PLS-DA model was constructed to maximise the separation identified by PCA and to more readily identify the spectral features identified in the loadings of the PLS-DA analysis. The cross validated scores and loadings coefficient plot of the supervised model is shown in (**Fig. 4.9**).

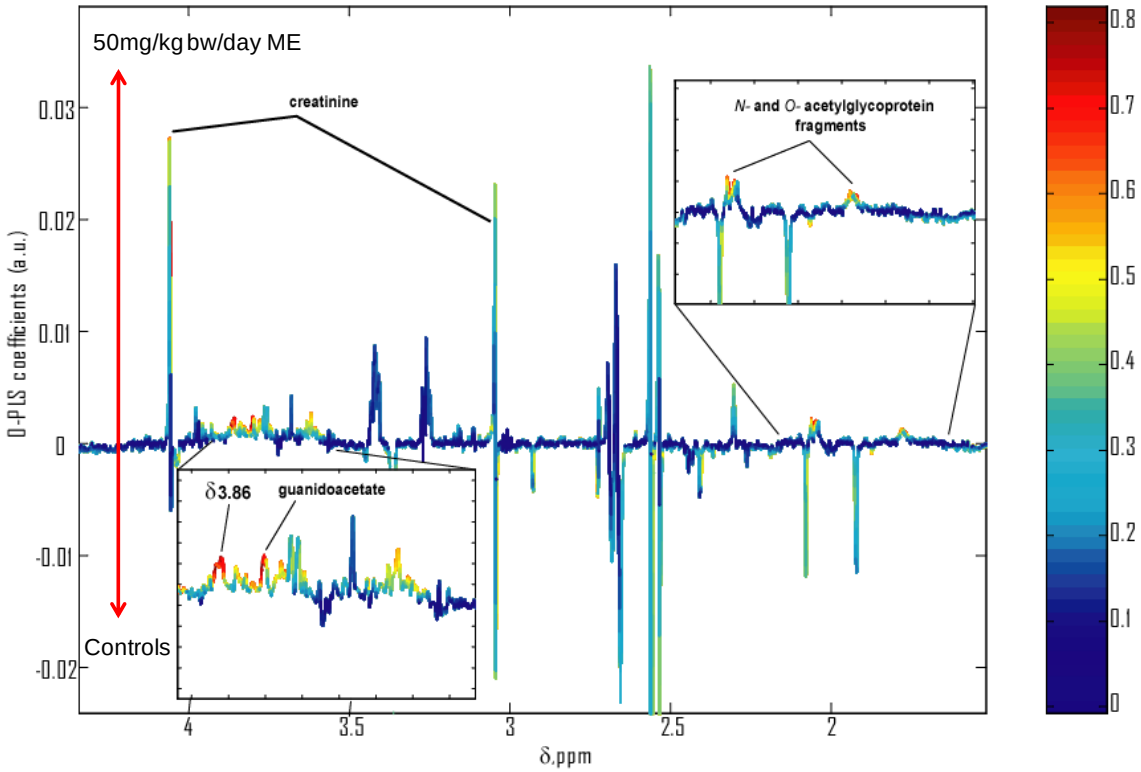


Fig. 4.9 O-PLS-DA loadings plot showing male day 28 vehicle control diet vs. 50mg/kg bw/day dietary dose group. ($Q^2\hat{Y}$: 0.21 and an $R^2\hat{Y}$:0.67). Normalised to total area, Unit Variance scaled.

A number of peaks were identified as diagnostic of ME exposure in rats; however the splitting patterns were necessarily identified by comparison with the raw spectral traces. The model suffered from a relatively low $Q^2\hat{Y}$ value: 0.21. The low excretory concentration of the metabolites identified may be a potential reason for the poor model strength.

O-PLS-DA analysis revealed a triplet peak (δ3.86), three singlet peaks (δ3.80, δ2.06, δ2.04) and a multiplet (δ1.78) in the day 28 50mg/kg bw/day dose samples in both male and female rats (**Fig. 4.9**). Based upon chemical shift of the singlet peak (δ3.80) and correlation of this peak with creatine revealed by STOCSY analysis this peak was identified as guanidoacetate (**Fig. 4.10**). Analysis of the raw spectra revealed that these peaks were at a very low excretory concentration level this would also limit the use of sensitive 2D-NMR experiments to aid structural identification. This suggests that the metabolites associated with ME exposure are being excreted at a very low concentration or that effect of ME dietary on the urinary metabotype is subtle. Analysis of the raw ^{1}H-NMR spectra revealed that these peaks were also found to be present in the day 7, 14 and 21 urine samples but not in the lower concentration dosage groups 1 or 5mg/kg bw/day samples. This is possibly due to a lower urinary concentration of the metabolite(s) at lower ME exposure levels. The O-PLS-DA loadings plot also showed an increased urinary concentration of creatinine in the 50mg/kg bw/day dose group animals.

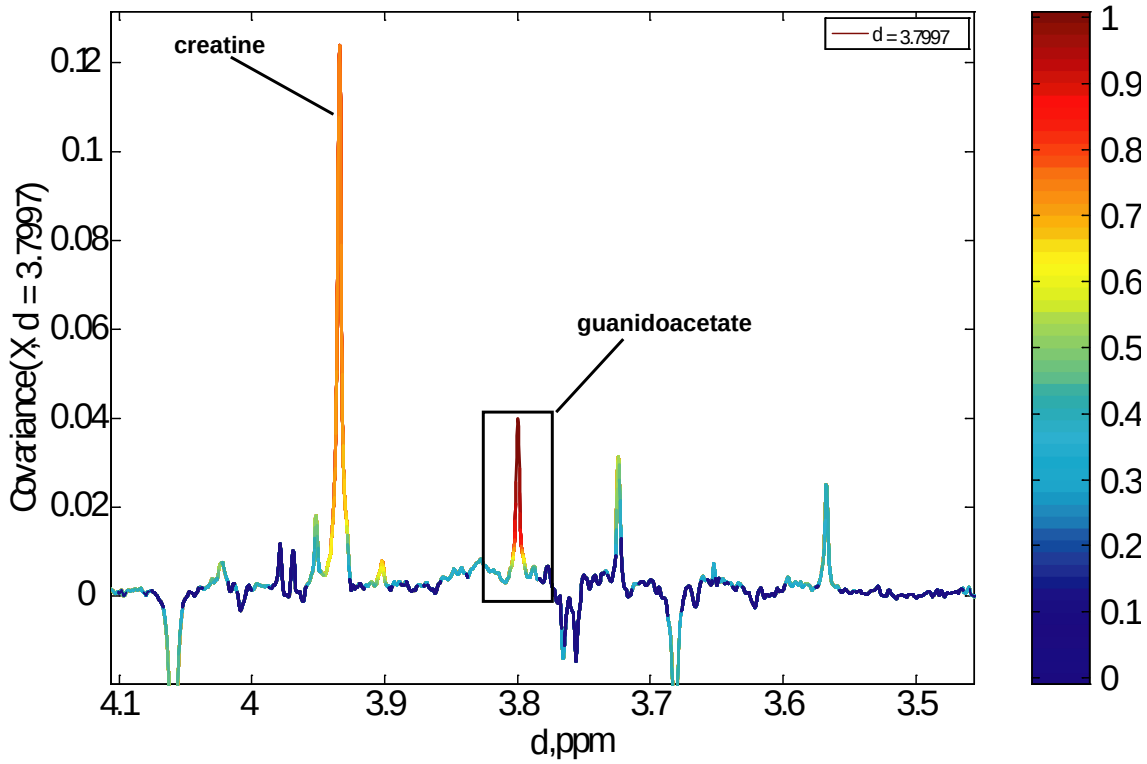


Fig. 4.10 STOCSY analysis of singlet peak δ3.80 reveals correlation with creatine. This singlet peak was identified as guanidoacetate.

STOCSY analysis of the triplet peak at δ3.86 revealed a clear correlation with the other unknown peaks identified by the O-PLS-DA analysis but also strong correlation with a singlet at δ2.30 and two doublets at δ7.06 and δ7.23. This was identified as *p*-cresol glucuronide (**Fig. 4.11**). However this biological covariance could be attributable to spectral overlap of the sugar protons of glucuronic acid in this area of the spectrum with the triplet peak (δ3.86) and does not necessarily provide further evidence of structure or biological origin for this spectral feature. The spectral features the two singlet peaks (δ2.06, δ2.04) and the multiplet (δ1.78) were tentatively identified as *N*- and *O*-acetylglycoprotein fragments by STOCSY analysis and comparison to reference spectra. However, low urinary concentration of these spectral features prevented absolute structural identification.

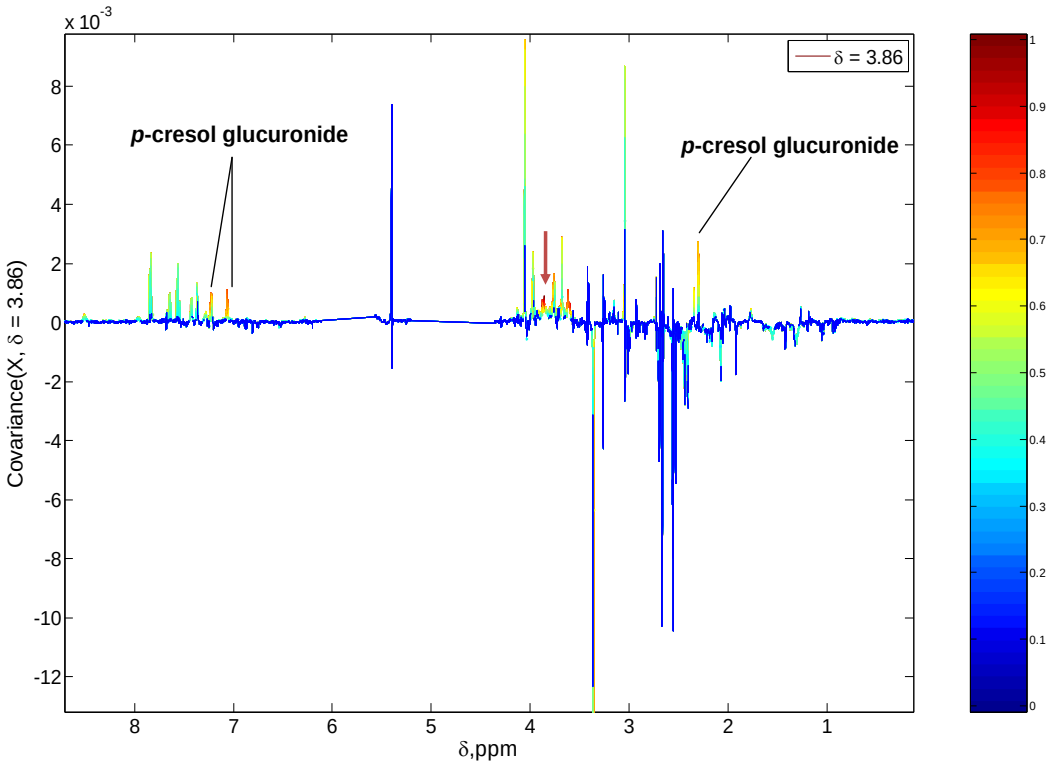


Fig. 4.11 STOCSY analysis of unknown triplet peak δ3.86 reveals strong correlation with the other unknown peaks identified by the O-PLS-DA analysis and also with *p*-cresol glucuronide.

Further analysis of the 1mg/kg bw/day and 5mg/kg bw/day dietary exposure dose groups did not identify any discrimination between control and dose groups in either of the sexes. Chemometric models suffered from a poor fit ($R^2\hat{Y}$) and predictive strength ($Q^2\hat{Y}$) this was due to a predominance of underlying sources of variation. Attempts to chromatographically separate these spectral features were made using a concentrated urine sample from a 50mg/kg bw/day dietary dosage group rat in a SPE. SPE proved unsuccessful as the peaks of interest eluted in the highly polar fraction with many overlapping spectral features in higher abundance.

4.5.3 Time-course effects of methyl eugenol dietary intervention

In an attempt to improve model strength and to determine whether metabolic changes could be observed throughout the course of the 28 day dietary regimen related to ME exposure, ^{1}H-NMR spectra of urines of days 7, 14, 21 and 28 were compared by multivariate analytical methods. By comparison of the same experimental animals, variance due to inter-subject variability was limited and the animals acted as their own controls eg. Comparison of day 7 with day 28. Cumulative effects of ME were assessed by various chemometric modelling techniques.

4.5.3.1 Pairwise timecourse PCA and PLS-DA high ME dietary exposure dose group

PCA was performed to model the time-course cumulative effect of ME exposure in the high (50mg/kg bw/day) dosage group and to identify any age-related effects of ME exposure.

To remove variation related to sexual dimorphisms, genders were modelled separately. Day 28 and day 7 urine samples from the 50mg/kg bw/day dietary intervention group were initially subjected to PCA analysis. Consideration of animals over time course of the study removed variation due to inter-subject variations and was found to improve model strengths. In both the PCA and PLS-DA scores plots clear separation could be seen between the day 7 and day 28 urine samples (**Fig. 4.12)**. Loadings attributable to either the day 7 or day 28 time-points were widely scattered and therefore it was difficult to extract urinary metabolite changes responsible for the grouping seen in the scores plot along the second principal component.

PCA analysis of all time-points days 7, 14, 21 and 28 of male rats in the 50mg/kg bw/day dose group revealed evidence of a time-point dependent clustering of samples. This separation was further augmented using PLS-DA modelling of the data subset (**Fig. 4.13**). Most variation in the PLS-DA model was found to be non-systematic inter-subject variations, however in the second and third PLS components clustering can be seen discriminating samples over the time course of the dietary intervention. PCA and PLS-DA analyses were repeated using the dietary matrix control animals to attempt to

identify any separation based upon ME dietary exposure. The PCA scores plot of the male dietary matrix control samples demonstrated limited evidence of clustering based upon time-point however it offered satisfactory cross validation and fit of data (Q^2:0.36 ; R^2:0.52) (**Fig.4.14**). The PLS-DA scores plot satisfactorily separated the dietary matrix control urine samples based upon time-point, however evidence of inter-subject variations still existed where samples appeared to cluster more strongly and demonstrate class membership with a different time-point. These models suffered from weaker Q^2 and R^2 values. This could possibly have been brought about by a greater degree of inter-subject variation or the elucidation of a true dose response pattern in the 50mg/kg bw/day ME dietary intervention. It could also be due to the lower sample number in the dietary matrix control group (n=19) compared to the 50mg/kg bw/day ME dietary intervention (n=32) animals which could potentially bias the data. It could also indicate a dose response of the animals in the 50mg/kg bw/day dietary intervention group based upon ME dietary exposure, however analysis of the loadings revealed a similar ambiguity of ^{1}H-NMR spectral features responsible for the time-point based clustering observed in both the dietary matrix control and 50mg/kg bw/day ME animals. These results suggested that the time-point based separation seen in the 50mg/kg bw/day ME dietary intervention samples were most likely based upon age of the animals rather than a specific effect of the 50mg/kg bw/day ME administered.

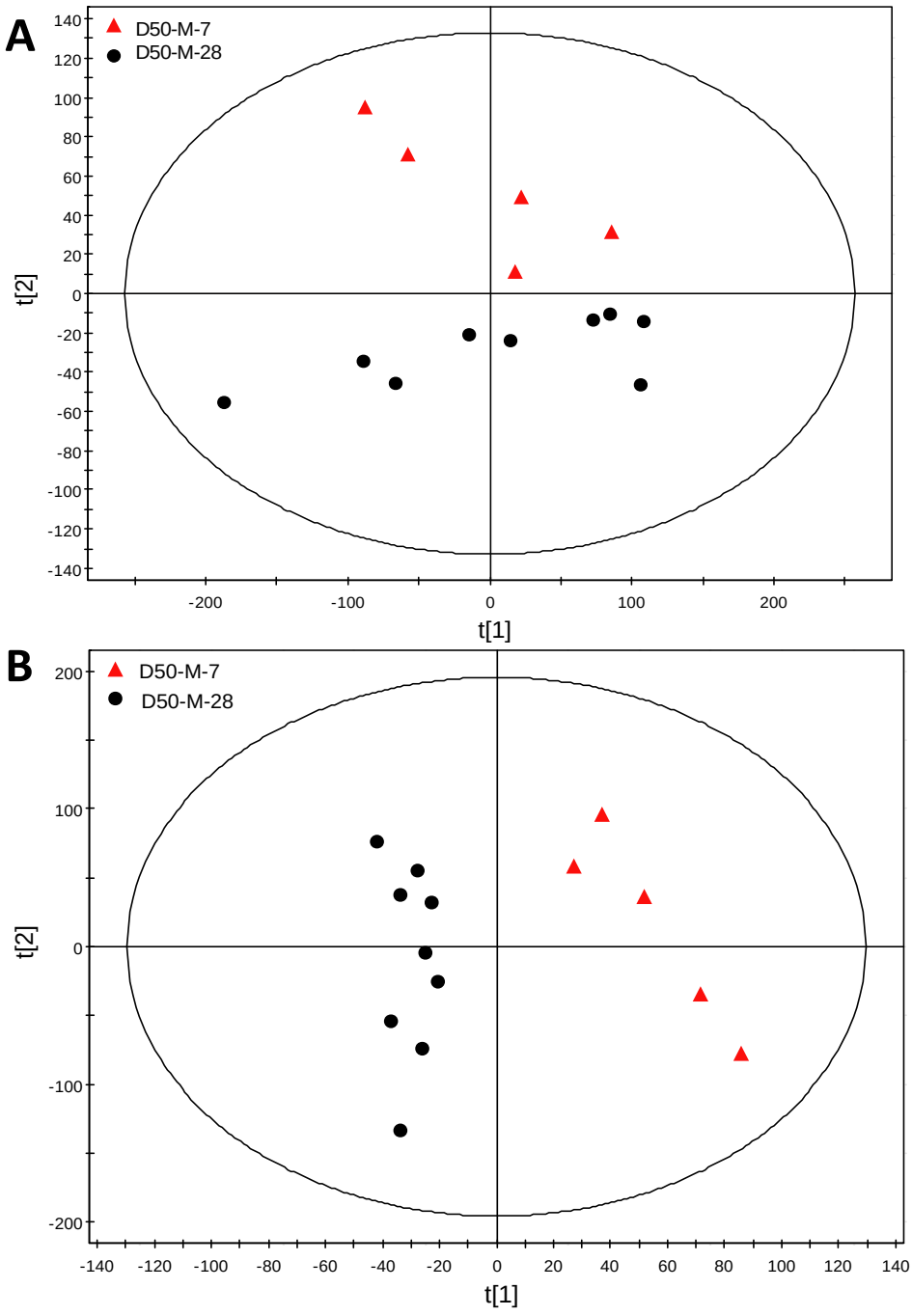


Fig. 4.12 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.4 R^2: 0.59) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.62 R^2: 0.89) from the ^{1}H-NMR spectra of the 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based on the day 7 and day 28 timepoints. Unit Variance scaled.

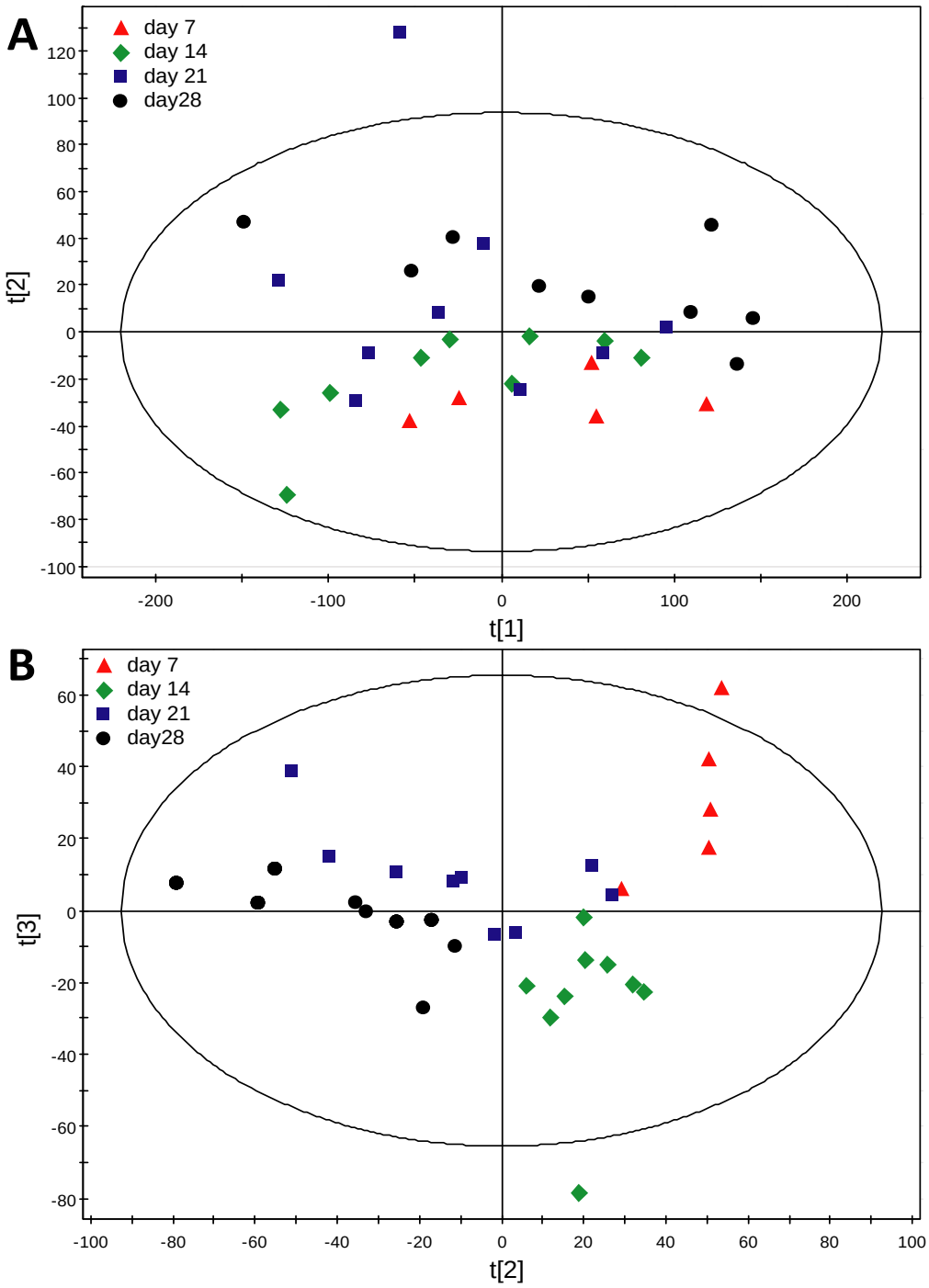


Fig. 4.13 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.37 R^2: 0.46) and (B) *t*[2]/*t*[3] PLS-DA scores (Q^2: 0.21 R^2: 0.51) from the ^{1}H-NMR spectra of the 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based upon time-point. Unit Variance scaled.

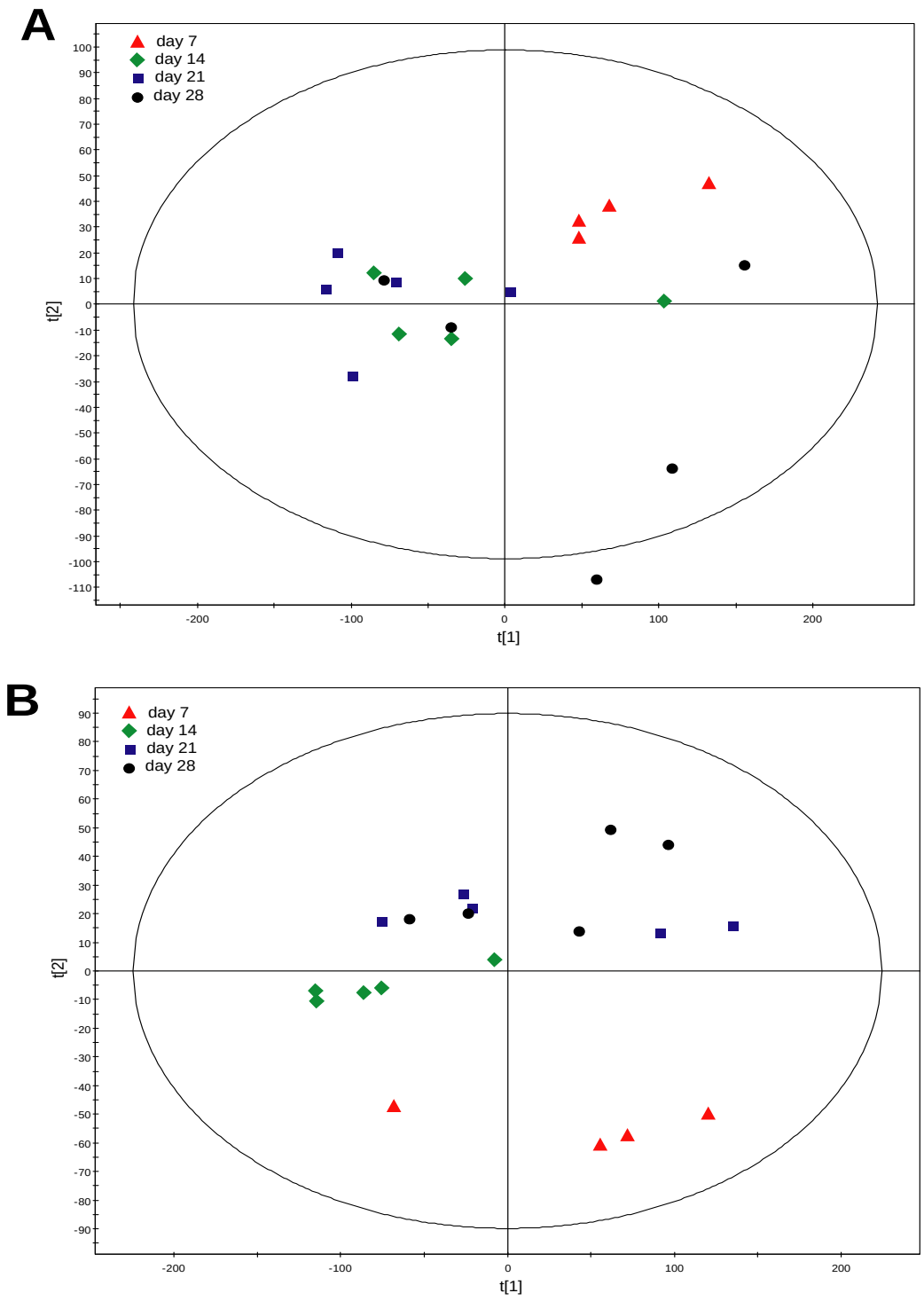


Fig. 4.14 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.36 R^2: 0.53) and (B) *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.16 R^2: 0.51) from the ^{1}H-NMR spectra of the dietary matrix control group of male rat urine samples some clustering of samples based upon time-point can be seen but there is also evidence of inter-subject variations. Unit Variance scaled.

4.5.3.2 Pairwise timecourse O-PLS-DA high ME dietary exposure dose group

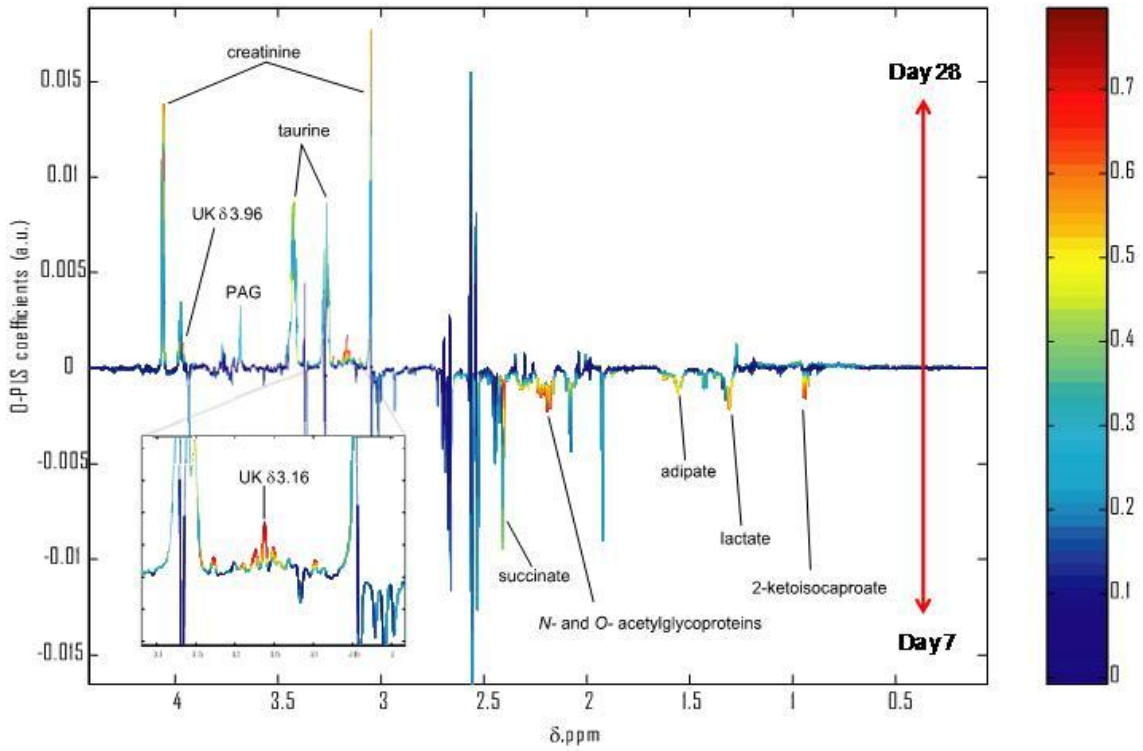
Pair-wise time-course analysis of day 7 to day 14, day 14 to day 21 and day 21 to day 28 produced robust O-PLS-DA models with excellent cumulative $R^2\hat{Y}$ and $Q^2\hat{Y}$ values that demonstrated clear discrimination based upon time-point. The result suggests the modelling of the maturation of the rats metabolism and alterations identified are therefore likely attributable to the age of the rat. At the time of the commencement of the dietary regimen rats were 6 weeks old and considered juvenile rats. The statistically significant metabolite changes observed in the O-PLS-DA loadings plots were attributable to transition of the experimental animals into adulthood. To identify metabolites attributable to ME comparison of the control vehicle diet group O-PLS-DA loadings to the high ME dose group O-PLS-DA loadings, found that metabolites contributing to discrimination between time-points did not contain any signals unique to ME exposure.

Comparison of the 7 day urine to the 28 day urine of the 50mg/kg bw/day dosage group produced the statistically strongest modelling of age related metabolic changes. A number of statistically significant changes in metabolic phenotype between the 7 week old (day 7) rats and the 10 week old (day 28) rats were seen. The model was particularly robust at discriminating the age difference between the rats with a $Q^2\hat{Y}$: 0.86 and an $R^2\hat{Y}$:0.98 by the second O-PLS-DA component.

Ten week old (day 28) animals were shown to excrete increased concentrations of creatinine, taurine, an unknown triplet (δ3.16), hippurate, PAG and *p*-cresol glucuronide. There were also various unidentified ^{1}H-NMR features discriminatory of the 10 weeks old animals. STOCSY analysis unknown triplet (δ3.16) showed strong correlation with taurine, a peak at δ3.96. The strong correlation of this unknown with taurine suggested the unknown may have been a taurine derived metabolite however potential taurine metabolites were discounted by comparisons with database references spectra (**Fig.4.15-17**).

The 7 week old rats were shown to excrete increased concentrations of 2-ketoisocaproate and the dicarboxylates; succinate, adipate and lactate and also

resonances corresponding to *N*- and *O*-acetylglycoprotein fragments. Identifications were made utilising STOCSY analysis, literature sources and database mining [138, 147].



-

Fig. 4.15 Aliphatic region of O-PLS-DA loadings plot demonstrating urinary metabolic differences between day 7 (7 weeks old) and day 28 (10 weeks old) male F344 strain rats of the 50mg/kg bw/day dose group.

Differences in metabolic phenotype can be seen between juvenile (7 week old) and adult (10 week old) rats. A similar pattern of metabolic differences can be seen in the control dose group with no changes indicated based on ME exposure. ($Q^2\hat{Y}$: 0.86 and an $R^2\hat{Y}$:0.98).

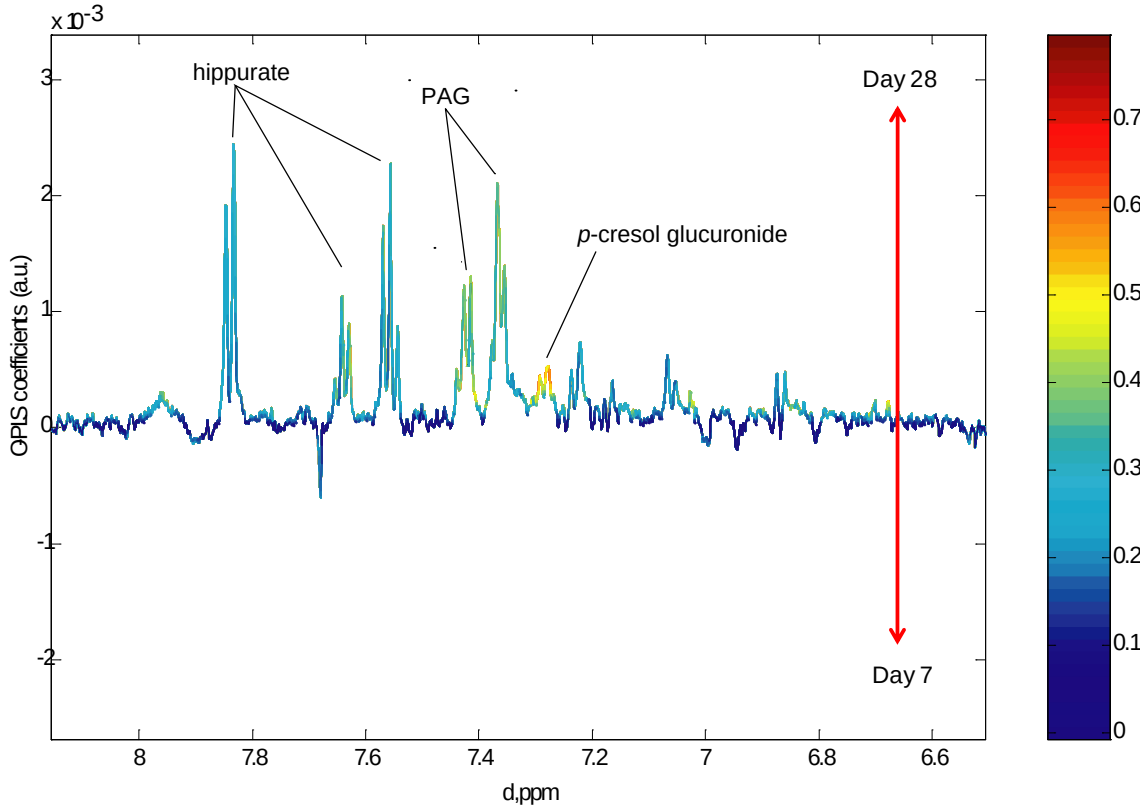


Fig. 4.16 Aromatic region of O-PLS-DA loadings plot showing urinary metabolic differences between day 7 (7 weeks old) and day 28 (10 weeks old) male F344 strain rats of the 50mg/kg bw/day dose group.

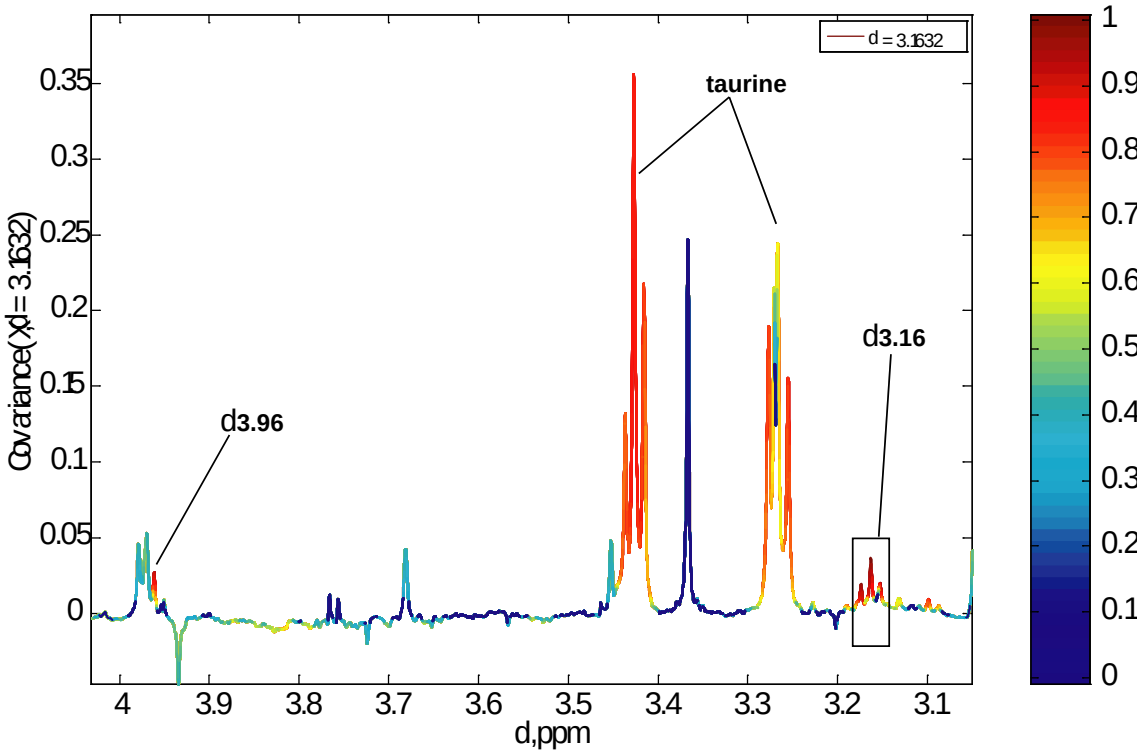


Fig. 4.17 STOCSY analysis showing correlation of the unknown triplet peak (δ3.16) with taurine and another unknown peak (δ3.96).

4.5.4 Gender related effects on methyl eugenol metabolism

Sexual dimorphism and age-related effects have been shown to dominate variations between animals in chemometric models of this dietary intervention urinary dataset. In order to ascertain whether there was an under-lying gender-specific effect on the metabolism of ME, comparison was made between the 50mg/kg bw/day ME urine of male rats to the 50mg/kg bw/day ME urine of female rats. To limit variation due to age and maximise cumulative effects of the high ME dosage groups, comparison was made between the 28 day samples. It has been shown that male rats have a greater degree of differential hepatic gene expression to female rats due to ME exposure, also males are thought to be more susceptible to the genotoxic effects of ME [165].

4.5.4.1 Pairwise PCA and PLS-DA

A subset of the 50mg/kg bw/day ME dietary intervention dataset was initially analysed using PCA to determine the source of greatest variation in the ^{1}H-NMR X matrix. It was found that a one PC model was sufficient to determine variation in the 28day subset (Q^2: 0.46; R^2: 0.81). Clustering in the scores plot along the first PC was not related to gender as female and male samples (♀: 9; ♂: 7) fell equally above the 0 standard deviation line, however the data points were between the +/- 2.50 standard deviations confidence interval and this did not indicate outlying samples within the subset (**Fig.4.18**). This finding was further augmented using supervised PLS-DA which suggested a gender biasing of the variation in the urinary ^{1}H-NMR data. Analysis of the loadings plot showed separation attributable to various signals however TMAO and GPC could be identified as separating female and male rats respectively.

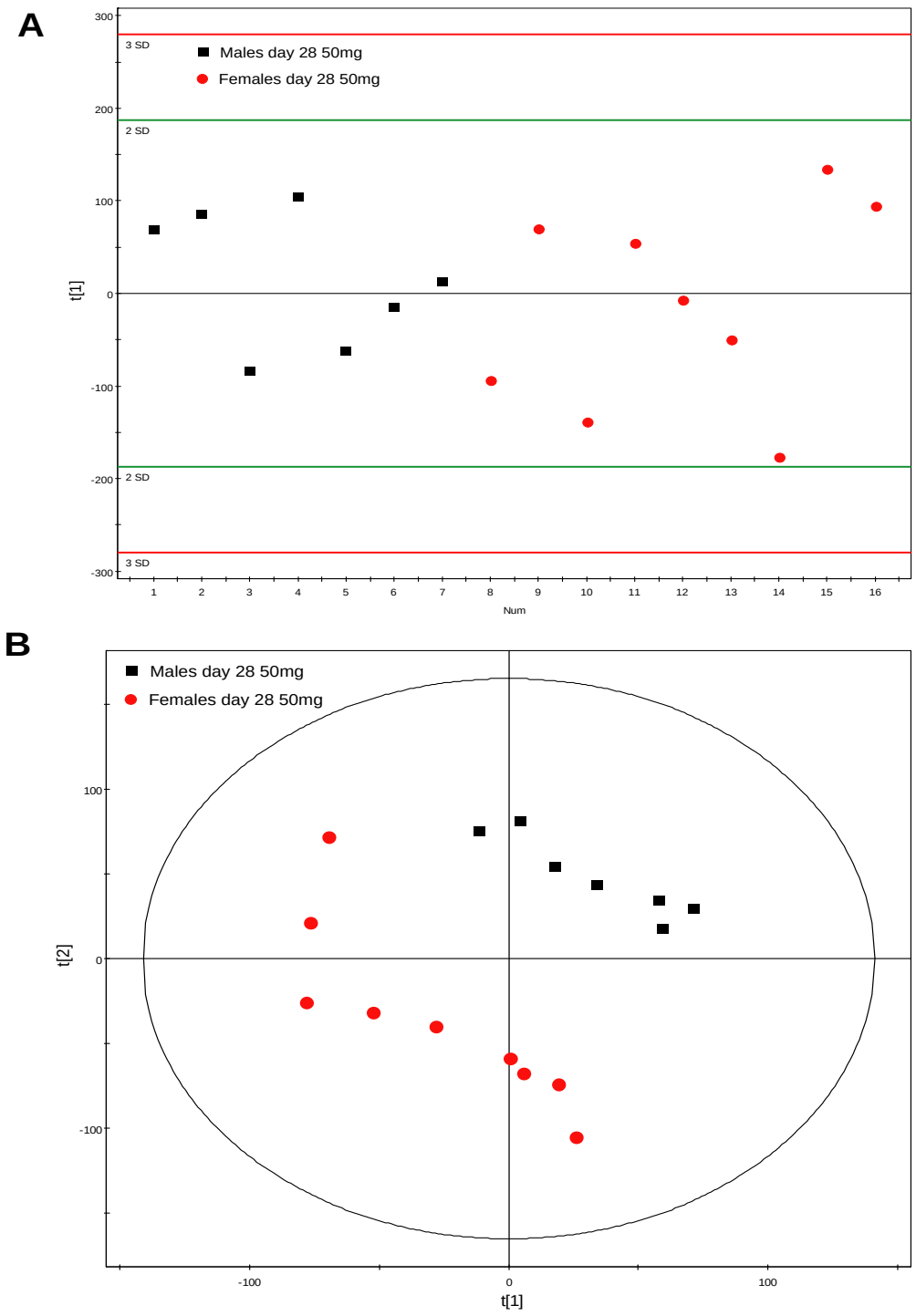


Fig. 4.18 Scores maps showing (A) t[1]/sample no. PCA scores (Q^2: 0.46 R^2: 0.81) and (B) t[1]/t[2] PLS-DA scores (Q^2: 0.37 R^2: 0.73) from the ^{1}H-NMR spectra of the 50mg/kg bw/day ME dietary intervention animals contrasting gender in order to determine any gender specific effects of ME exposure in the F344 rat strain. Clustering of samples based upon gender can be seen in the PLS-DA scores. Unit Variance scaled.

4.5.4.2 Pairwise O-PLS-DA

Calculation of the O-PLS-DA model of the day 28 50mg/kg bw/day ME dietary intervention samples revealed robust statistical discrimination between male and female animals. This model utilised one aligned component to give particularly strong modelling of the variation in Y ($Q^2\hat{Y}$: 0.93; $R^2\hat{Y}$: 0.98). This model succeeded in confirming the sexual dimorphism seen in the F344 rat strain previously identified by the comparison of all male and female samples but with a more robust cross validation and explanation of variance (**Fig.4.19**). Key metabolic differences in females included evidence of bile acid metabolites, *N*-acetylglutamate, alpha-ketoglutarate and TMAO, and in male rats evidence of increased GPC and 2'-deoxycytidine excretion. This finding was largely consistent with the previously known metabonomic sexual dimorphism in experimental animals [173] but was not suggestive of any gender related metabolic change related to ME exposure.

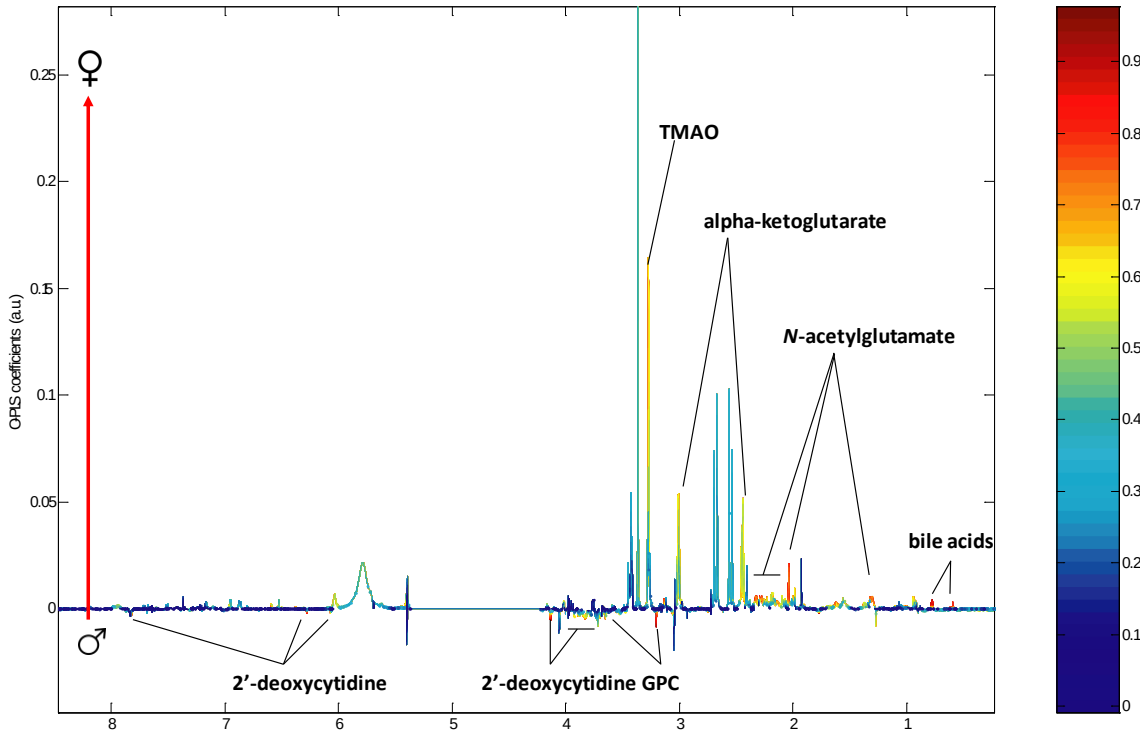


Fig. 4.19 O-PLS-DA loadings coefficients plot showing evidence of metabolic sexual dimorphism in the F344 rat strain. Robust modelling of the data was achieved suggesting the majority of urinary metabolic variation is gender specific in the first aligned O-PLS component ($Q^2\hat{Y}$: 0.93; $R^2\hat{Y}$: 0.98).

4.5.5 Effect of methyl eugenol administration method on urinary metabolic phenotype

Comparison of the orogastric gavage administration method to the dietary matrix intervention were made to examine any effect on the metabolism of ME or any homeostatic effect on the metabolic phenotype due to gavage administration. Oral gavage is known to cause rapid absorption of ME from the stomach and a rapid accumulation in the liver, it also can potentially damage the GI tract and cause confounding metabolic changes due to an acute stress response in experimental animals. Chemometric modelling of the 50mg/kg bw/ day gavage treatment against the dietary matrix and gavage controls against vehicle dietary controls were made to establish any effect of ME exposure on the urinary metabolic phenotype in F344 strain rats.

4.5.5.1 Pairwise PCA and PLS-DA high ME exposure dietary matrix vs. orogastric gavage

In order to model the effect of orogastric gavage treatment on ME dietary exposure a PCA model was calculated comparing the 50mg/kg bw/day dietary matrix group to the 50mg/kg bw/day gavage treatment group rats at all time-points. To remove variation related to sexual dimorphism only male samples were included in the model. Variation in the first four principal components was dominated by inter-subject variation; however cross-validation of the model showed an ability to correctly predict 52% of samples with 44% of the predictions being within the 99% confidence interval. No clear clustering was seen attributable to gavage treatment. PLS-DA modelling succeeded in identifying variation attributable to gavage treatment and dietary matrix intervention in the second and third PLS-DA components (**Fig. 4.20**).

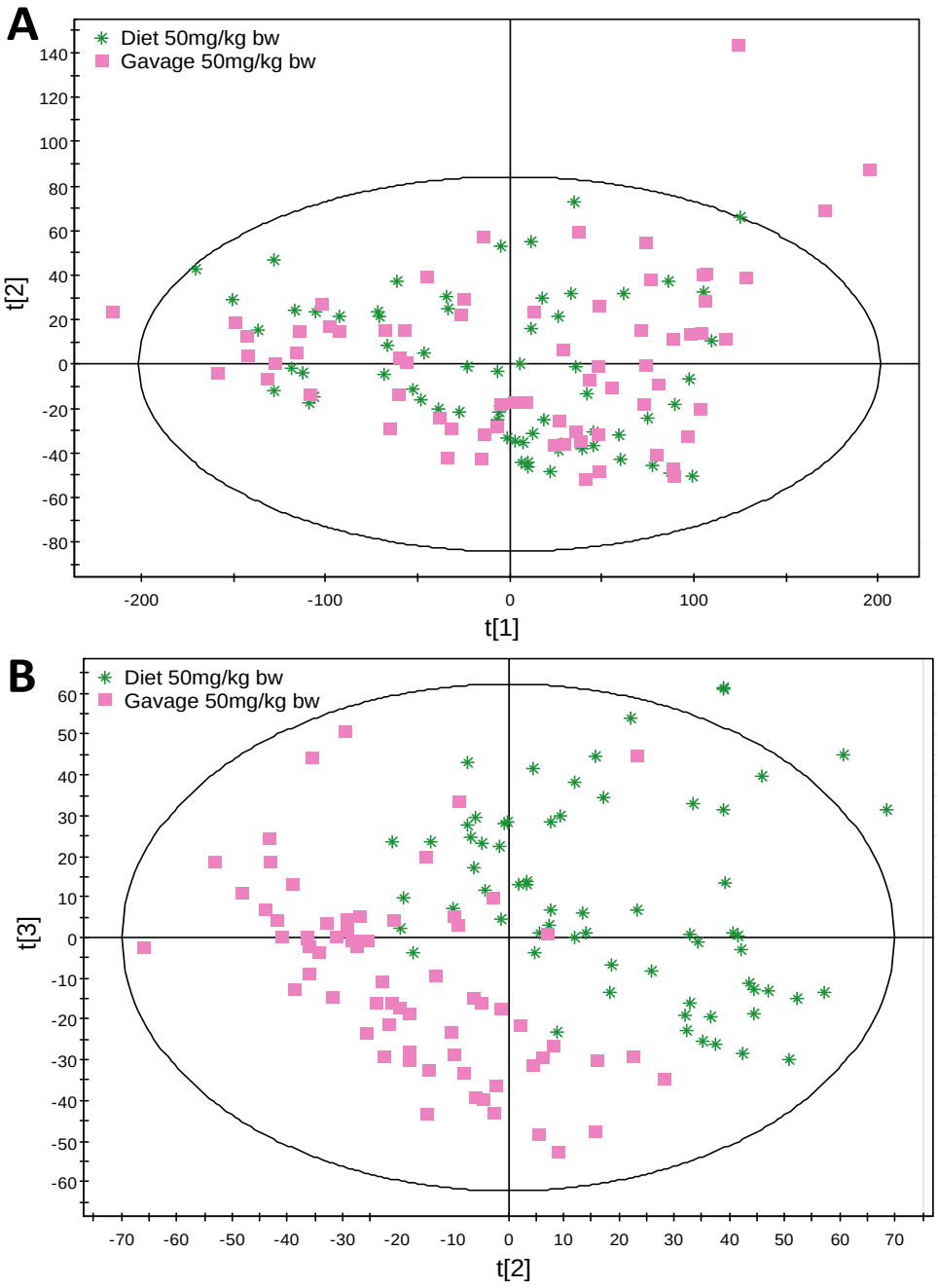


Fig. 4.20 Scores maps showing (A) *t*[1]/*t*[2] PCA scores (Q^2: 0.44 R^2: 0.52) and (B) *t*[2]/*t*[3] PLS-DA scores (Q^2: 0.43 R^2: 0.56) from the ^{1}H-NMR spectra of the 50mg/kg bw/day gavage treatment and 50mg/kg bw/day ME dosage group of male rat urine samples showing clustering of samples based upon treatment method in the PLS-DA model only. Unit Variance scaled.

4.5.5.2 O-PLS-DA high ME exposure dietary matrix vs. orogastric gavage

Evidence of an acute stress-related response was observed in male F344 strain rats following repeated orally gavaged high dose ME animals when compared to the high dose ME dietary matrix animals. O-PLS-DA revealed statistically significant increases in hippurate, methylmalonate, lactate, succinate, methylamine, *scyllo*-inositol and an unknown singlet peak at δ4.03 in the high ME dose (50mg/kg bw/day) gavage animals. The high ME dietary matrix (50mg/kg bw/day) animals showed increased urinary excretion of creatinine, taurine, phenylacetylglycine, *p*-hydroxyphenylacetate and *m*-hydroxymandelate (**Fig. 4.21**).

The unknown at δ4.03 which was found to discriminate the gavage treatment from the dietary dosing regimen was found to show correlations with creatine, dimethylamine, succinate and *N*-acetylglyprotein by STOCSY analysis.

Calculation of an O-PLS-DA model of the gavage control group samples against the vehicle control samples revealed a similar pattern of metabolic changes and no evidence of a ME specific effect.

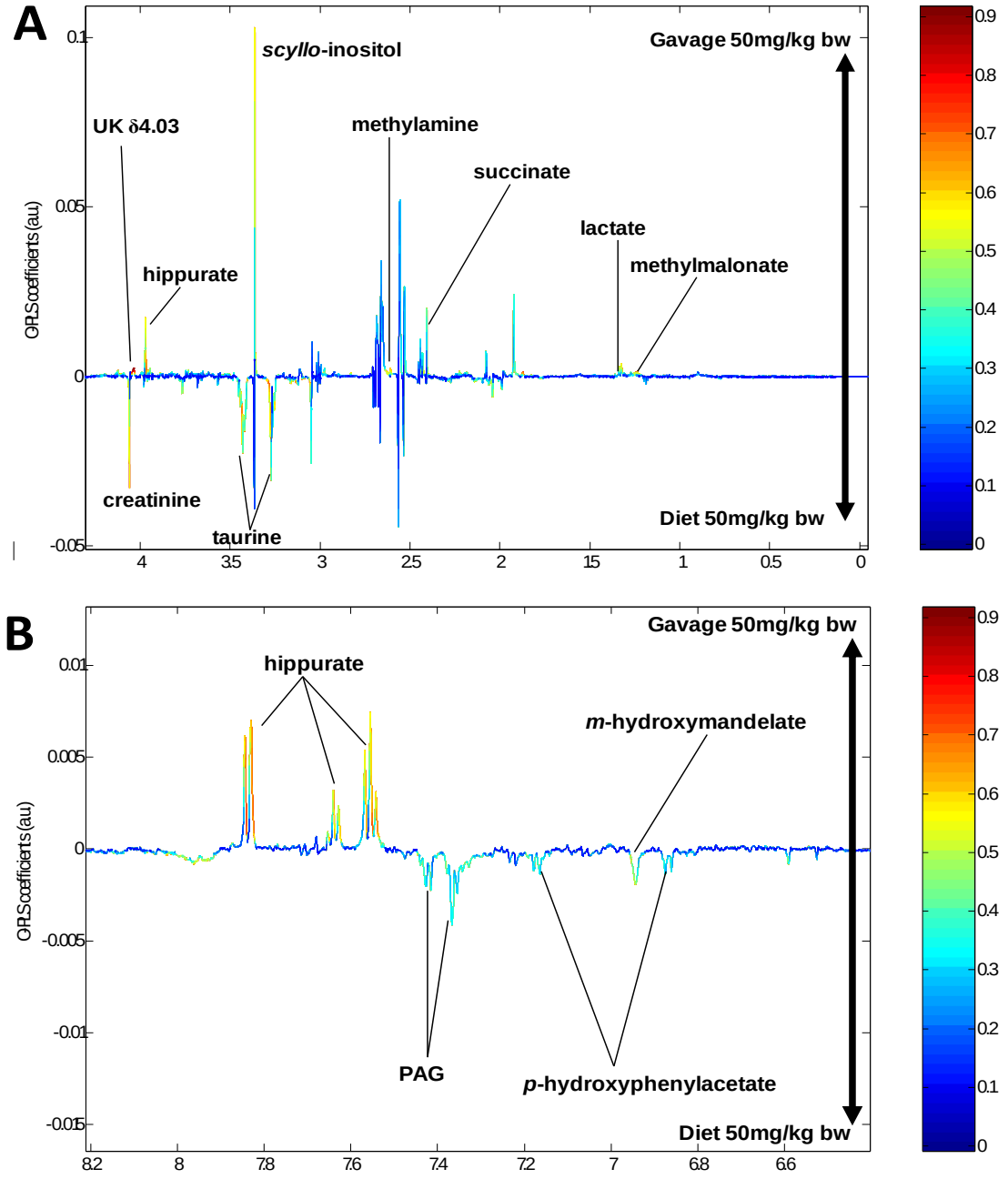


Fig. 4.21 O-PLS-DA loadings coefficient plot showing (A) aliphatic region demonstrating urinary metabolic differences between 50mg/kg bw/day ME dietary matrix (downwards pointing peaks) and 50mg/kg bw/day ME gavage treatment (upwards pointing) male F344 strain rats. (B) aromatic region of the spectrum. Metabolic differences can be seen attributable to the orogastric gavage treatment. ($Q^2\hat{Y}$: 0.66 and an $R^2\hat{Y}$:0.96).

4.6 Discussion

ME has been identified as a multiple site carcinogen when using high dosages in both the rat and mouse (NTP, 2000). It is of concern to environmental toxicology due to its ubiquity in many commonly consumed foodstuffs and its use as a food and cosmetics additive. Genotoxic DNA/RNA adduct formation has been seen in both *in vitro* and *in vivo* animal models [167]. Mutagenicity, however, has been shown to be largely negative in standard *in vitro* cell cultures of both human and rat/mouse tissues and positive in *in vivo* rat and mouse models [159, 163]. The high dosages administered by gavage within the 2 year NTP bioassay (37-300mg/kg bw/day) most likely overestimated the carcinogenicity of ME and its potential effects in humans (NTP, 2000). This is exemplified by the fact that the serum concentration of ME in human populations due to dietary exposure is several orders of magnitude lower in humans than the lowest doses utilised in the NTP bioassay [167].

Assessment of carcinogenicity is further confounded by the dose dependent metabolism of ME where *O*-demethylation is most likely to be the primary route of metabolism and elimination from the body in humans where the estimated dietary exposure is (~10-100µg/kg bw/day) [174]. It is only at higher dosage levels where the genotoxic metabolite 1'-hydroxymethyl ME is formed which is metabolised to the reactive carbocation [64]. This study utilised a more realistic 28 day dosage regimen in accordance with the known levels of human exposure to resolve the issues highlighted by the 2 year NTP bioassay. It also utilised a more realistic dosing strategy, comparing administration using a dietary matrix with the commonly used but potentially confounding oral gavage method.

4.6.1 The effect of high dosage methyl eugenol dietary intervention

It was anticipated with the lower doses used within this study (1-50mg/kg bw/day) the most abundant metabolites of ME in the urine would be the sulphate or glucuronic acid conjugate of eugenol or 5-allyl-2-methoxy phenol formed by *O*-demethylation of ME, based on the known metabolism of ME in the rat [167].

As such these metabolites represent candidates for metabonomic biomarkers of ME exposure within rat urine, although with human dietary relevant doses of ME the identification of these metabolites by ^{1}H-NMR spectroscopy proved problematic due to low urinary concentration and spectral overlap of constitutive urinary metabolites in higher concentration. Metabolites tentatively identified as discriminatory of ME exposure were guanidoacetate and acetylglycoproteins fragments, however absolute identification was confounded by intensity of the spectral features close to the threshold of noise. The STOCSY correlation of the singlet peak δ3.80 to creatine was highly indicative of the assignment of the methylene protons of guanidoacetate. Guanidoacetate is an essential component of the *de novo* synthesis of creatine in the body. Guanidoacetate is synthesised in the kidneys from the amino acids arginine by the enzyme arginine:glycine amidinotransferase where it also serves the function as a osmolyte. In the rat liver hepatocytes the enzyme guanidoacetate methyltransferase converts guanidoacetate into creatine by methylation [175].

The identification of guanidoacetate in the urine of the 50mg/kg bw/day dose group animals suggests an effect of ME dietary exposure on the energy metabolism and a potential blockage of creatine synthesis and would be consistent with the known hepatic genotoxicity of ME (NTP 2000). However the low urinary concentration and tentative assignment of this peak is inconclusive.

Acetylglycoprotein fragments in the urine are indicative of an inflammatory response, acute phase acetylglycoproteins are released by hepatocytes as an anti-inflammatory response to cytokine release following tissue damage, cellular disruption or infections. Acetylglycoprotein fragments are also known to be excreted during normal growth processes and are not necessarily pathologically indicative [176].

The presence of signals attributable to acetylglycoprotein fragments in the experimental animals exposed to ME, although tentative may possibly indicate an inflammatory response in the rat liver.

However discrimination was obtained between the high ME dietary exposure dosage group and the vehicle control group animals, yet confirmation of the identity of the metabolites is problematic as the regions of the spectrum in which the potential biomarker of ME exposure appeared, revealed a very low urinary concentration. This would make the identification of this unknown difficult by 2D-NMR experiments such as J-resolved, COSY, TOCSY and carbon 13 isotope experiments such as HSQC and HMBC where minimum urinary concentrations are necessary for successful metabolite identification [177].

Experimental ^{1}H-NMR spectra of eugenol were found to be available in the literature [178]. The splitting pattern of eugenol was found to be inappropriate to account for the discriminatory spectral peaks identified by O-PLS-DA. The experimental ^{1}H-NMR of eugenol from the spectral database for organic compounds (SDBS) showed aromatic proton peaks at δ6.6 and δ6.75; an aromatic alcoholic proton peak at δ5.0; allyl protons between δ4.9 to δ6.3 and methoxy protons at δ3.8 and benzylic protons at δ3.2. The unknown triplet peak at δ3.8 could potentially represent the methoxy protons of eugenol, however other peaks of eugenol which would present at higher intensity in the spectrum are notably absent. It is also important to note that the sulphation or glucuronidation of eugenol would cause a shift of the peaks upfield in the spectrum identified by O-PLS-DA and also result in a more complicated proton NMR spectrum. It can also be reasonably anticipated that due to the great structural similarity of 5-allyl-2-methoxy phenol to eugenol that the splitting pattern and peak positions of the ^{1}H-NMR spectrum would be similar.

The STOCSY correlation of the peak at δ3.80 with *p*-cresol glucuronide suggested that the unknown metabolite may be associated with a similar metabolic pathway to *p*-cresol glucuronide or demonstrate biological covariance. However it is more likely that spectral overlap may have occurred in this region with the proton signals of glucuronic

acid causing the STOCSY correlations observed with *p*-cresol glucuronide. It is also possible the peak at δ3.80 could also be a glucuronide conjugate of a ME metabolite or perhaps a similar microbially-derived metabolite associated with ME exposure. However, there were no statistically significant features in the aromatic region of the ^{1}H-NMR spectra associated with ME exposure.

The low urinary concentration of the spectral features associated with the 50mg/kg bw/day ME dietary exposure group animals confounded absolute metabolite identification. It is conceivable that if the spectral features are urinary metabolites of ME detoxication these would be expected to be in relatively low concentration in the urine with a 50mg/kg bw/day dosage regimen and the use of higher ME dosages may improve the ^{1}H-NMR signal from these spectral features. The low urinary concentration of the spectral features associated with ME exposure could also be related to a subtle perturbation of endogenous homeostasis which also may be more readily observed utilising a higher concentration ME dosage regimen. The use of high-throughput UPLC-MS analysis of the rat urine samples combined with a targeted mass spectral analysis could be used to characterise the urinary metabolites of ME. Further UPLC-MS analysis was unfortunately beyond the financial scope of this study.

4.6.2 Gender related effects on methyl eugenol metabolism

No urinary biomarkers of ME were identified in this study and therefore sex-related differences related to ME exposure could not be fully investigated. However sex-related differences in endogenous urinary metabolites were identified between male and female F344 strain rats.

As proof of principal metabonomic investigation succeeded in identifying sexual dimorphism in the ^{1}H-NMR urinary metabolic profiles of the F344 rats in this study. Many of these metabolites have previously been identified within our group in Han Wistar strain rats [173].

The higher concentration of TMAO found in female rats is potentially due to the influence of the oestrus cycle. TMAO is known to form from oxidation of trimethylamine by flavin-containing monooxgenases [179], a degradation product of choline which is derived from catabolism of phospholipids or from the diet [180]. Testosterone is known to inhibit expression of flavin-containing monooxygenases which would account for the higher excretion of TMAO in the female and a higher concentration of GPC in the male rat urine samples [181]. Rats are known to excrete aromatic chlorogenic acid metabolites which are processed by and dependent on the composition of the gut microbiota [182]. The excretion of the chlorogenic acid metabolite 3-hydroxyphenylpropionic acid is known to occur in both male and female rats [183]. However male rats predominantly have a higher sulphation rate for this metabolite which inactivates its inhibitory effect on platelet aggregation. Decreased platelet aggregation seen in female rats is possibly due to this gender-specific metabolic difference [184]. It is also known that urinary concentrations of *N*-acetylglycoproteins fluctuate dependent on the stage of the oestrus cycle and this could explain its increased concentration in female rat urine [185]. Female rats were also shown to be characterised by increased excretion of alpha-ketoglutarate, an important keto acid formed from the deamination of glutamine and an important intermediate in the krebs cycle and also an important element in the removal of excess nitrogen in the body. Also evidence was seen of the C_{18} axial methyl protons of bile acids in the female F344 strain rat urine [173].

In male rats an interesting discriminatory metabolite was identified in the aromatic region of the ^{1}H-NMR spectra attributable to the deoxyribonucleoside, 2'-deoxycytidine. Deoxycytidine is formed by the attachment of the pyrimidine cytosine to a deoxyribose by a β-glycosidic linkage. Metabonomic studies have previously identified this deoxyribonucleoside in urine in F344 strain male rats treated with the model hepatotoxin galactosamine. 2'-deoxycytidine was found to be depleted in galactosamine treated rats and therefore present in higher concentrations in untreated control F344 strain male rats [186].

The region in the aliphatic region of the spectrum was attributed to the –CH protons of deoxyribose. STOCSY analysis also confirmed statistical covariance of the peaks in the aromatic region and these peaks in the aliphatic region. It is thought that male F344 strain rats are susceptible to the genotoxic effects of ME and are subject to increased differential hepatic gene expression [165]. Potentially the increased urinary concentration of a deoxyribonucleoside such as deoxycytidine could be representative of the increased ME genotoxicity in male F344 strain rats. It is known that rats excrete higher concentrations of deoxycytidine in response to DNA damage caused by X-irradiation [187] and in animals bearing ascites sarcoma [188] and as such it represents a potential urinary biomarker of oxidative DNA damage and a potentially important metabolic difference of male experimental F344 strain rats. However, it is more likely that this deoxyribonucleoside represents a normal constitutive urinary metabolite associated with male F344 strain rats and represents a gender-related differential deoxyribonucleoside turnover in these experimental animals.

Male F344 strain rats were also shown to be characterised by an increased excretion of GPC. GPC is a ubiquitous choline derived metabolite which acts along with phosphocholine as the major source of choline storage in the cytosol, it is also a major organic osmolyte in renal medullary cells. GPC functions to counteracts the effect of urea on proteins and other macromolecules in the renal medullar, renal cells are exposed to greatly fluctuating extracellular solute concentrations and respond to hypertonic stress by accumulating GPC and other osmolytes such as betaine and myo-inositol. The presence of GPC in rat urine attributable to male rats is unexpected as GPC would

normally be utilised as a storage source of choline and retained within the renal medullary cells.

The findings of this study reaffirm previously demonstrated gender specific urinary metabolic change already elucidated by the metabonomic approach. However the identification of unknown discriminatory metabolites such as the unknown aromatic spectral features associated with male F344 strain rats is still of great interest to understanding the relationship between experimental animal sex, urinary metabotype and differential drug metabolism/toxicology.

4.6.3 Time-course effects of methyl eugenol dietary intervention

Chemometric evaluation of ^{1}H-NMR data by PCA, PLS-DA and O-PLS-DA provided robust models demonstrating evidence of age-related differences in metabolic phenotype between the 7 day, 14 day, 21 day and 28 day time-points. The F344 rats were 6 weeks old and considered juveniles at commencement of the dietary intervention study. Time course pair wise examination of animals in each of the separate dosage groups provided strong evidence of changes in metabolic phenotype due to age differences. This was shown not to be related to an influence of ME in the higher 50mg/kg bw dosage group as a very similar fingerprint of metabolic changes was observed in the control animals dosage group.

Evaluation of age-related differences in the urinary metabolic composition of rats has been previously reported and is in agreement with the findings of this study [189, 190]. Ten week old animals were shown to excrete increased concentrations of creatinine, taurine, hippurate, PAG and *p*-cresol glucuronide and also an unidentified triplet peak at δ3.16 which when submitted to STOCSY analysis correlated strongly with taurine.

Increased creatinine urinary concentration with age is associated with increased muscle mass and concentration is directly correlated with body mass [191].

The increased concentration of metabolites strongly associated with the gut microbiota in the 10 week old (day 28) rats such as hippurate, PAG and *p*-cresol glucuronide

indicates that the gut microbiota is still in the process of establishment and stabilisation in the 6 week old (day 7) rats as they continue to adapt to an adult diet [97, 192].

The increased excretion of the amino acids taurine in the 10 week old adult rats is almost certainly attributable to increased muscle mass and also more specific maturation of organs in the rat [189]. Increased taurine excretion is related to decreased liver function and is known to be protective against reactive oxygen species [193].

The increased urinary excretion of amino acids and creatinine is also associated with changes in renal function in older rats [194].
There was also an unidentified metabolite discriminatory of the 10 week old (day 28) animals a triplet peak at δ3.16 which was shown to correlate strongly with taurine when subjected to STOCSY analysis. Statistical correlations were also observed with the peak at δ3.98. Potential analogues and metabolites of taurine such as hypotaurine, amidino-taurine and taurine carbamate were discounted as candidate biomarkers as the expected peaks were not identified in the spectra.

The 7 week old rats were shown to excrete increased concentrations of 2-ketoisocaproate and the dicarboxylates; succinate, adipate and lactate and also resonances corresponding to *N*- and *O*-acetylglycoprotein fragments. Metabonomic investigation has shown decreased concentrations of metabolic end-points of the mitochondrial respiratory chain, dicarboxylates and Kreb's cycle intermediates in the urine of older rats with increased concentrations in younger animals [189, 190]. This is consistent with the increased energy requirements of growth and development in adolescence and a diminishing mitochondrial function in older animals. It is known that the ability to metabolise ME decreases with age in rats [166]. The excretion of acetylglycoprotein fragments in adolescent animals is also known to be associated with growth processes and has been shown to be discriminatory of adolescence in a long-term study of aging in dogs [195].

The complexity of characterizing the response to dietary intervention superimposed on the background of dynamic biological processes such as aging presents a substantial

challenge and confounding biological information associated with aging needs to be taken into consideration when analysing metabonomic datasets.

4.6.4 Effect of methyl eugenol administration method on urinary metabolic phenotype

A similar pattern of metabolic change was seen by chemometric comparisons of the vehicle control diet dosage group to the control gavage dosage group as in comparison of the high dosage gavage to the high dosage dietary intervention. This suggests there were no observable effects of acute ME exposure by gavage treatment in the experimental animals and metabolic phenotypic change was due to stress-related responses only.

It was found the dietary dosing regimen produced more pronounced genotoxic effects at the acute high dose (50mg/kg bw/day) in the stomach with increased DNA adduct formation than the gavage treatment [165]. The constant feeding of ME was thought to provide a continual exposure, whereas the one-shot gavage dosage and the highly lipophilic nature of ME would mean rapid clearance from the stomach and a saturation of the intoxication pathway in the liver with this latter route of administration.

The oral gavage is a commonly used method of drug administration in toxicology and can potentially introduce confounding effects in experimental models due to acute stress. Oral gastric gavaging involves physical stressors of handling and restraining experimental animals, insertion of a rigid needle from the mouth to stomach and can lead to respiratory interference and stomach distension. This method can also cause oesophageal/tracheal perforations, aspiration pneumonia, haemothorax and increased morbidity [196]

Repeated daily oral gavage can also lead to tissue damage of the pharyngea with development of granulation tissue. This granulomatous inflammation has been linked to higher mortality in the F344 strain rat in a carcinogenicity study [197], although this was found to be strain specific [198]. It has important implications for the observed metabonomic effect in this study as it can lead to oral intake aversion, acute stress and a

reduction in body weight and other confounding histopathological findings unrelated to ME exposure.

Furthermore acute stress can cause induction of the corticotrophin-releasing hormone system potentially causing a catabolic impact in the experimental animals [199]. There exists a need for less invasive measurement of the acute stress response in experimental models as the principal measurement of stress is by analysis of plasma concentrations of corticosterone, the main glucocorticoid produced by stress, which is in itself an invasive method. Measurement of faecal corticosterone concentrations is also insufficient as the oral gavage event itself is brief [200]. Metabonomic analysis offers a potential method of gauging drug administration stress accurately and non-invasively in urine of experimental animals. Metabonomic analysis of blood plasma of rats has already been shown to be effective in gauging the extent and the amelioration by diet of the severity of the stress response [201].

The biochemical changes due to repeated oral gavage treatment are potentially related to a disturbance in energy metabolism, amino acid metabolism and gut microbial changes related to stress. The accumulation of short-chain organic acids in the urines of oral gavage dosed rats is suggestive of a disruption of energy metabolism and components of the respiratory chain and krebs cycle in the mitochondria. Increased urinary concentration of succinate, lactate and methylmalonate indicate impairment of mitochondrial function the detrimental effects of which result from accumulation of acyl-coA groups, which reduce the availability of coenzyme A for important metabolic enzymes [202, 203]. These findings provide evidence of oral gavage induced organic aciduria. Succinate is an important component of the citric acid cycle which donates electrons to the electron transfer chain and methylmalonate urinary excretion is known to increase in rats with vitamin B_{12} deficiency [204].

The higher urinary concentration of an osmolyte *scyllo*-inositol in the urine of the gavage treated rats was not anticipated and could potentially be an introduced contaminant, however the gavage vehicle consisted simply of maltodextrin and distilled water and was thoroughly tested for its stability over a period of 14 days. A single

singlet peak for *scyllo*-inositol at δ3.35 is also shared by methanol, however the presence of significant quantities of methanol in the gavage treated urine samples seems unlikely. Scyllo-inositol is a stereo isomer of inositol and is a naturally occurring plant sugar alcohol. In Alzheimers disease research *scyllo*-inositol has been shown to block the formation of β-amyloid plaques in mice brain and reduce memory deficits associated with this condition [205]. The presence of *scyllo*-inositol in these urine samples is not readily explainable.

The increased excretion of methylamine in the gavage treated animals is likely to be derived from the degradation of dietary choline by the gut microbiota [206]. Concurrent changes in excretion of aromatic metabolites such as hippurate, phenylacetylglycine, hydroxyphenylacetate and *m*-hydroxymandelate are suggestive of fluctuations in the activities or populations of members of the gut microbiota. Changes in gut microbial metabolites are known to occur in rats subjected to acute and chronic stress and fluctuations in these metabolites are perhaps related to the stress response or gastric dysbiosis induced by the gavage treatment method [207].

4.6.5 Conclusion

The findings of this study confirm that at doses relevant to human levels ME has no detectable acute toxicity in the rat urinary metabolic profile. As proof of principle the metabonomic approach has successfully identified metabolic differences related to gender, age and a stress-response related to the use of the orogastric gavage. The evidence of genotoxicity in the dose range 1-50mg ME/kg bw/day confirmed by DNA-ME adduct formation and differential gene expression by Dr. J Ellis BMM, IC [165] produced no pronounced perturbation of homeostasis in the F344 strain rat, however low urinary concentration NMR spectral features were implicated as attributable to ME exposure but absolute identification was confounded by signals close to the threshold of noise. Tentative assignments were made of guanidoacetate and acetylglycoprotein fragments to the spectral features associated with ME exposure. These were biologically plausible and were indicative of a potential disruption of energy metabolism and inflammatory response. It can also be reasonably surmised that the peaks attributable to ME exposure could derive from the *O*-demethylation of ME and its subsequent excretion as either the sulphate or glucuronic acid conjugate of either eugenol or 5-allyl-2-methoxy phenol.

Chapter Five: Changing Dietary Fat Composition and Carbohydrate Glycaemic Index in Subjects at Risk of Developing the Metabolic Syndrome – Mapping Metabolic Change Through Metabonomics

5.1 Background

The RISCK study (an acronym derived from the participating institutions; Reading University, Imperial College London, Surrey University, MRC Human Nutrition Research Cambridge and Kings College London) was designed to examine a significant issue in public health; the effects of different amounts and types of dietary fat and carbohydrate on metabolic syndrome, which is diagnosed by a set of criteria associated with an increased risk of morbidity; coronary heart disease (CHD), type 2 diabetes and other chronic diseases and increased mortality. This cohort study aimed to address the health impact of changes in dietary fat and carbohydrate amount and composition, to inform public health policy. Understanding metabolic syndrome and associated risks will aid management and therapy and help alleviate the growing social and financial burden associated with heart disease and type 2 diabetes. Within the context of this study metabonomic techniques were applied to a 130 participant subset of the study, which was performed to attempt to identify global metabolic changes associated with dietary change in humans.

5.2 Aims and Objectives

- The overarching aim of the study was to identify metabolic changes in both urine and faecal water samples attributable to dietary modulation of the type and amount of fat and carbohydrate, utilising ^{1}H-NMR spectroscopy combined with multivariate statistical analyses.
- To monitor metabolic responses caused by dietary change in human biofluids attributable to gut microbial metabolism
- To investigate co-variation of ^{1}H-NMR spectroscopic data of both urine and faecal water with faecal bacterial population numbers and clinical/anthropometric measures in obese volunteers at risk of or suffering from MS.

5.3 Introduction

5.3.1 Metabolic syndrome

The increasing incidence of obesity, the pleiotrophic diagnostic criteria of metabolic syndrome (MS) and the associated health risks is an area of serious public health concern and study. Obesity is highly prevalent with estimations of half of the female population and around a third of the male population considered obese in the UK (Food Standards Agency data www.food.gov.uk). MS is strongly correlated with obesity and many of the diagnostic criteria of these conditions are shared, it is estimated of obese individuals in the UK ~25% are symptomatic of MS [7, 208]. Higher incidence of MS has also been noted in ethnic subgroups such as Asians and African-Caribbeans [209].

MS is a cornucopia of characteristics, which confer increased risk of coronary heart disease (CHD), diabetes, cancer incidence and increased morbidity and mortality. The diagnostic criteria characteristic of MS include hypertension, dyslipidaemia (elevated plasma triacylglyceride (TAG) levels, low High Density Lipoprotein (HDL)-cholesterol levels), central adiposity and insulin resistance. MS is thought to confer greater risk to the development of diabetes and CHD than any of the individual diagnostic components of the syndrome [7]. Dyslipidaemia is a strong risk factor for CHD and is defined by a resistance of insulin to lipid metabolism [210, 211]. Insulin resistance is known to be causative of adiposity and a primary cause of many of the metabolic perturbations linked with MS (European Group for Study of Insulin Resistance, 2002).

Diet is known to be strongly associated with the aetiology and severity of CHD. Highly processed, westernised diets containing a high proportion of saturated fats and refined simple carbohydrates with low intakes of fruits, vegetables and fibre have been demonstrated to be linked to increased risk of CHD [212], while conversely reductions in calorific intake and increased consumption of fruit and vegetables have been shown to be preventative of CHD [18, 19]. The complexity and quantity of dietary carbohydrates in the human diet is known to affect the risk factors and development of MS and CHD. Increased consumption of complex carbohydrates with a low GI (classification to quantify the relative blood glucose response to carbohydrate-containing foods) can reduce the severity of insulin insensitivity by dietary modulation [10].

The proportions and type of dietary fatty acids are known to affect many of the diagnostic risk factors associated with MS which include blood pressure, blood serum lipid profiles, insulin desensitisation and function of the endothelial system [212]. In particular, the most favourable alteration of dietary fatty acid intake is substitution of SFA especially from animal sources with monounsaturated fatty acids (MUFA) and a subsequent reduction in calorific intake. This has been shown to alter in blood serum lipid profiles in human hyperlipidaemic study participants; increasing HDL-cholesterol levels and reducing TAG [213].

It is important to note that MS does not have the status of a disease and is a syndrome because there are many factors that can be considered diagnostic of MS. This means that the classification of MS is open to a degree of subjectivity and is not absolute. Practically, this can cause problems when considering study participants as a homogeneous set, as would be the case in a more defined disease state such as type II diabetes or CVD.

5.3.2 Role of the Gut microbiota in Fatty Acid and Carbohydrate Metabolism

It is important to consider the influence of the human gut microbiota in metabolic processes in metabonomic analyses. The gut microbiota is recognised to be essential to proper lipid and carbohydrate metabolism in humans. It is known that populations of the microbiota are responsible for deconjugation of bile acids and the production of secondary bile acids, the synthesis of short chain fatty acids such as acetate, propionate and butyrate and the absorption of cholesterol [137, 214].

Dietary intake levels of different food components are known to affect the composition and metabolic rates of specific gut microbial populations [215]. It has been shown that diets containing high levels of soluble and insoluble fibres encourage the growth of fermentative gut microbiota populations in the GI tract such as *Lactobacillus* and *Bifidobacteria* which are known to be beneficial to health [126, 139, 141]. Investigations with prebiotics (fermentable carbohydrates which encourage growth of beneficial bacteria in the gastro-intestinal tract) in both animal and human studies, have shown subsequent

changes in gut microbial metabolism can affect blood serum fatty acid composition. Supplementation with inulin or oligofructose has been shown to reduce the blood serum fatty acids in hypercholesterolaemic and hyperlipidaemic study participants [216].

Several studies, including that by Martin, Wang et al. 2008 have demonstrated the critical role of the gut microbiota in germ free mice where a human gut microbiota population had been introduced. This study demonstrated strong influence of the gut microbiota in the recycling of bile acids between the liver and gut and in fatty acid metabolic processes and blood serum lipid profiles, particularly in the synthesis of very low density lipoprotein (VLDL)-TAG. Treatment of this humanised colonic microbe mouse model with either the probiotic bacteria *Lactobacillus paracasei* or *Lactobacillus rhamnosus* was shown to produce reduced serum VLDL and low density lipoprotein concentration, evidence of a stimulation of glycolysis and increased lipid concentrations in the liver [138].

Few studies have elucidated the effect of high fat composition diets on the human gut microbiota. It is thought the increased fat intake typical of western diets would increase the concentrations of both fat and bile acids metabolised by the gut microbiota. Gut microbial populations are known to convert mammalian primary bile acids (synthesised in the liver) into microbially derived secondary bile acids and are also responsible for the synthesis of diacylglycerols, which can directly affect the population levels of specific gut microbial species [217, 218]. Several studies have demonstrated the effect of high fat dietary interventions on the most abundant species of the gut microbiota in mice particularly reductions in *Bifidobacterium* spp., *Bacteroides* spp. and *E.rectale/C.coccides* populations [219]. A high fat diet in mice, has also been shown to produce inflammation in the gut, alter insulin sensitivity and increase blood serum concentrations of lipopolysaccharides. Conversely, introduction of the prebiotic oligofructose into the diet of these high fat diet mice led to reduction in the alterations of gut microbial populations with an increase in the levels of *Bifidobacterium* spp. and a reduction of intestinal inflammation [220]. The important role of the gut microbiota in the metabolism of both fatty acids and carbohydrates is suggestive of a strong role in obesity, the development of MS and the subsequent onset of type II diabetes and CVD.

5.3.3 Application of Metabonomics to Human Studies

It is known that the analysis of data from human subjects is confounded due to high genetic and environmental diversity of populations. The metabolic profiles of human participants are also affected by pleiotrophic physiological factors unrelated to diet such as hormonal status, diurnal cycles and additionally, variations in the gut microbial population of an individual. These factors have been shown to exert a powerful effect on the metabolic phenotype. In animal studies where a higher degree of genetic and environmental certainty is possible, such variation is significantly reduced. Convolution of the data in human studies is reduced primarily by limiting inter-subject variability, ensuring dietary compliance and by use of mathematical filtering tools for removing and interpreting "irrelevant" data in complex multivariate datasets.

Techniques such as metabonomics can determine accurately the subtle metabolic effects related to dietary and lifestyle changes. Spectral pre-treatment such as OSC (orthogonal signal correction) has been used to remove disrupting variation in human studies related to non relevant sources of variation to reveal the relevant features in ^{1}H-NMR spectra associated with the subtle effects of nutritional intervention. This has notably been used to characterise the metabotype in vegetarian, and high and low meat consumption diets [221]. OSC has also been used successfully in facilitating the diagnosis of coronary heart disease [222, 223], the interaction of stress and dietary intervention [201], in soy isoflavone [112], chamomile [110, 224] and in alcohol dietary interventions [111].

OSC ensures that systematic variation in the spectroscopic data unrelated to the biological response of interest is removed so that separation between different sample classes caused by more subtle sources of variation may be observed. When OSC is incorporated into supervised pattern recognition algorithms such as PLS-DA (Partial Least Squares-Discriminant Analysis), it is a powerful tool for identifying potential biomarkers and for understanding the metabolic effects of dietary change.

5.3.4 Reduction of Inter-subject Variability in Nutritional Intervention Studies

Long-term effects of different dietary components and dietary population epidemiological studies have investigated the beneficial/detrimental effects of certain diets and food components. However, there is little understanding of the underlying mechanisms of nutritional protection against disease.

Evaluation of dietary intake is crucial to nutritional studies to assure dietary compliance. Prosaic methods of dietary assessment include quantitative food diaries, dietary recall, or simple blood biochemistry measures. These methods are simplistic and in some cases flawed and are open to subjectivity and falsification by study participants. Thus, these methods can cause false positives/negatives within a dataset and can misinform datasets when the aim of a study is to examine subtle short-term dietary interventions and underlying mechanistic information.

Dietary recall and compliance is difficult to assess by the researcher and relies upon subject trust, therefore biochemical confirmation of dietary compliance is a major goal of nutritional studies. A recent metabonomic study has successfully identified proline betaine as a putative biomarker of citrus fruit consumption with further validation in citrus fruit consumers within the ^{1}H-NMR urinary dataset of the INTERMAP epidemiological study [96]. Likewise, in Chapter 6 of this thesis, S-methyl-L-cysteine sulphoxide and its *N*-acetylated mammalian derived metabolite were shown to be stable urinary biomarkers of cruciferous vegetable intake. Identifying anomalous outliers in chemometric models, which would otherwise bias the data and removing them from the model, can reduce the influence of non-compliance. For example, this has been observed previously in studies where an outlier in the data was due to the presence of trimethylamine-*N*-oxide (TMAO) [225]. TMAO is used to increase osmotic concentration and thus depress the freezing point of body fluids, in cold-water fish. The finding of TMAO indicated that the subject in question had consumed fish and thereby did not adhere to the dietary restrictions. Similarly it is also possible to identify subjects who have not adhered to alcohol and prescription/over-the-counter drug restrictions, due to the characteristic ^{1}H-NMR splitting patterns of ethanol and presence of drug metabolites in the urine [111, 226].

The importance of limiting inter-subject variability to produce meaningful, statistically relevant results is of paramount importance in metabonomics. The ability to reduce variation within study populations is extremely important, for example differences in age, gender, race, smoking status, alcohol consumption, prescription drug use and consumption of biologically active dietary components, such as constituents of many fruit and vegetables, can all affect the statistical strength and predictive power of a model of human data.

Assessing the effects of subtle perturbations of metabolic phenotype in nutritional intervention studies is potentially confounded by many environmental, genetic and commensal gut microbial variations in human populations. These considerations are of great importance in design of experiment in human nutritional intervention studies and in interpretation of the statistical significance of results.

5.4 Methods

5.4.1 Study Design and Protocol

Urine and faecal water samples were obtained for a subset of 130 volunteers taking part in the RISCK study at the University of Reading. Men and women (age range: 30–70 y), who were recruited from the general population, attended a clinic at the participating centers in a fasting state for measurement of height, weight, waist, seated blood pressure (BP), liver function test, glucose and lipid concentrations, and haematology. RISCK was a Food Standards Agency funded project involving 5 UK research institutions [227]. Free-living volunteers (650) were recruited to a randomised, controlled, single blind, parallel design dietary intervention study involving the manipulation of total dietary fat (~38% vs. ~28%) and its composition (i.e. saturated fatty acids SFA vs. monounsaturated fatty acids MUFA) and total dietary carbohydrate (~45% vs. ~55%) and its composition (i.e. high vs. low GI) for a six month period.

Exclusion criteria for this study were as follows:

- medical history of ischaemic heart disease;
- >30% 10-y risk of CVD;
- diabetes mellitus;
- cancer;
- pancreatitis;
- cholestatic liver disease;
- renal disease;
- use of lipid-lowering drugs;
- systemic corticosteroids;
- androgens;
- phenytoin;
- erythromycin or drugs for regulating haemostasis (excluding aspirin);
- exposure to any investigational agent ≤30 d before the study;
- presence of gastrointestinal disorder or use of a drug that is likely to alter gastrointestinal motility or nutrient absorption;
- a history of substance misuse or alcoholism;
- a current pregnancy, planned pregnancy, or given birth in the past 12 mo;

- an allergy or intolerance to intervention foods;
- an unwillingness to follow the protocol or to give informed consent;
- a weight change of >3 kg in the 2 months before the study;
- intake of >1 g eicosapentaenoic and docosahexaenoic acids/d;
- smoking >20 cigarettes/d.

The 130 volunteers were randomly assigned to one of five different dietary intervention groups, four experimental diets (Diets B-E) and a control diet (Diet A). All experimental groups were age matched (average age = 54 ± 9.5 years). (**Table 5.1**)

Table 5.1 Table showing RISCK study dietary interventions performed in human participants. (SFA: saturated fatty acid; MUFA: monounsaturated fatty acid; PUFA: polyunsaturated fatty acid; CHO: carbohydrate; GI: glycaemic index.)

	% energy from fat	% types of fatty acids			% energy from CHO	GI points reduction
Dietary Intervention		SFA %	MUFA %	PUFA %		
A	38	18	12	6	45	-
B	38	10	20	6	45	-
C	38	10	20	6	45	11
D	28	10	11	6	55	-
E	28	10	11	6	55	13

All the diets were isoenergetic and the protein intake remained constant. The n-6:n-3 PUFA ratio in the study foods was kept constant at 5:1. The volunteers were provided with spreads and oils for cooking and other foods with modified fat composition, in order to facilitate the dietary change. To reduce the overall GI of the diets, appropriate complex carbohydrate substitutes for simple carbohydrate-rich foods were also provided. Dietary calculations were based on estimates of habitual intake derived from the latest National Diet and Nutrition Survey (NDNS) and National Food Survey [228]. Dietary manipulation was achieved by replacing all major fats and carbohydrates in volunteers' baseline diets with products specifically chosen or formulated to provide the required fat/carbohydrate intake and composition in each of the intervention diets (spreads, oils, dairy products, cakes, biscuits, breakfast cereals, pasta, rice, bread, potatoes).

All volunteers underwent a 4-week run-in period following the control diet (Diet A). Subsequently, the volunteers were randomly allocated to one of the five dietary interventions for a period of 24 weeks, urine and faecal water sample collections were performed at 0 and 24 weeks (**Fig. 5.1**). One hundred and fourteen volunteers provided urine and faecal water samples at the baseline, of which eighty-eight provided samples at the end of the 24 weeks dietary intervention.

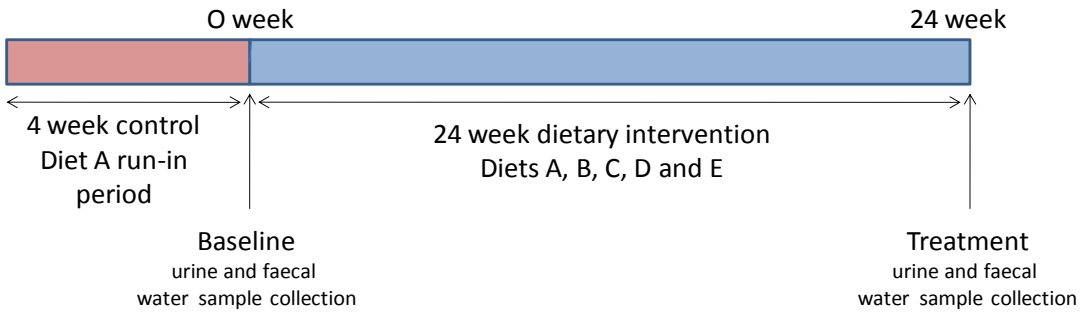


Fig. 5.1 Schematic showing study outline of the RISCK study dietary intervention

5.4.2 Diagnostic recruitment criteria

One hundred and thirty volunteers were recruited at the Hugh Sinclair Unit of Human Nutrition, Department of Food Biosciences at the University of Reading on the basis of their increased risk for developing the MS and its associated cardiovascular risk factors. The volunteers were recruited in two cohorts of 65 volunteers. The recruited volunteers were male and female, non-smoking and aged between 30-65 years. In order to be included in the study, the volunteers were screened for their risk of developing MS. The entry criteria for the volunteers were insulin resistance, dyslipidaemia, hypertension, obesity and central adiposity. Screening of the volunteers consisted of single measurements of fasting blood glucose and insulin, fasting serum lipids (e.g. HDL-cholesterol, LDL-cholesterol, TAG), seated diastolic and systolic blood pressure, waist circumference, weight and height. The recruitment of the volunteers was based on a scoring system which ensured that volunteers expressed a minimum of two features of the metabolic syndrome (score ≥ 4) (**Fig. 5.2**). Scores were weighted according to their relevance to the development of the metabolic syndrome, and/or technical reliability of the measurement with respect to biological and analytical variation.

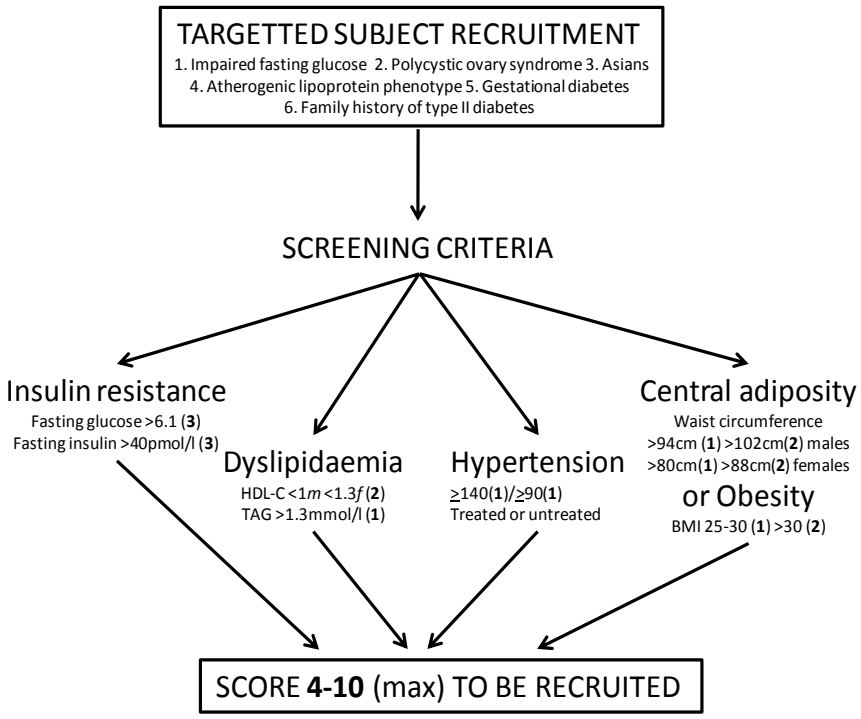


Fig. 5.2 Schematic showing the RISCK study diagnostic recruitment criteria**.** Intial targeted subject recruitment was based on an inclusion criteria.

5.4.3 Faecal Water Sample Preparation

Faeces for all five dietary interventions were collected for both a baseline (0 week) and a treatment (24 week) time point and frozen at -40˚C. Faeces (1g) were mixed with 5ml of PBS agitated for 10mins and then centrifuged for 30mins at 10,000 g. The supernatant was finally filtered using Whatman filter paper to remove cells and cell debris. The completed samples were received from our collaborators at the Hugh Sinclair Unit of Human Nutrition, Department of Food Biosciences at the University of Reading, where faecal samples were also prepared for bacterial enumeration by fluorescent *in situ* hybridisation (FISH) and faecal short chain fatty acids (SCFA) analysis

Samples for ^{1}H-NMR spectroscopic analysis were prepared by adding between 200-400µl of faecal water (dependent on volume available) to UPLC standard water containing 50% D_2O (which acts as a field frequency lock) and 0.01% TSP (sodium 3-(trimethylsilyl) propionate-2,2,3,3-d_4 for chemical shift reference) to a final volume of 600µl. The samples were centrifuged at 10,000 g for 10mins to remove any precipitates and 580µl aliquot of the supernatant transferred to a 5mm outer diameter NMR tube (Goss Scientific Instruments, Great Baddow, Essex, UK) for analysis.

5.4.3 Faecal Water ^{1}H-NMR Spectroscopy acquisition

One hundred and fourteen volunteers provided faecal samples at the baseline 0 week time point, with a total of 88 participants providing faecal samples at the end of the 24 weeks dietary intervention. The analysis was performed on samples from volunteers who completed one of the five dietary intervention regimens and provided faecal samples at both time points.

Faecal water samples were analysed using a Bruker DRX 600 NMR spectrometer (Bruker Analytische GmbH, Rheinstetten, Germany) operating at 600.13 MHz for ^{1}H equipped with 5mm triple resonance probe with inverse detection. The pulse program used was RD-90^{o}-t_1-90^{o}-t_m-90^{o}-acquire free induction decay (FID). The RD represents a relaxation delay of 1.5s; t_1 refers to an interval of 4µ seconds. The water signal was then

re-irradiated during t_m for 150ms the mixing and relaxation delay. FIDs (64) were collected for each sample into 32k data points with a spectral width of 8000Hz. The FIDs were then multiplied by an exponential weighting function with a line broadening of 0.3Hz prior to Fourier transformation. Acquired NMR spectra were phase and baseline corrected using a MATLAB script (Dr. Tim Ebbels, Imperial College) and referenced to TSP at δ0.0.

5.4.4 Urine ^{1}H-NMR Spectroscopy acquisition

The same participants provided urine samples at weeks 0 and 24 of the dietary intervention study. The analysis was performed on samples from volunteers who completed one of the five dietary intervention regimens and provided urine samples at both time points. All urine analyses were performed at 600.13MHz on a Bruker DRX 600 Spectrometer (Bruker Analytische GmbH, Rheinstetten, Germany) coupled to a sample-handling system Gilson (Middleton, WI, USA) XL233 interfaced to the XWinNMR data system for control and timing.

5.4.5 Clinical markers and food diary analyses of study participants

Clinical assays to assess efficacy of the dietary interventions were performed by collaborators at the Hugh Sinclair Unit of Nutrition at the University of Reading. These were expected to correlate with and inform the further metabonomic analysis carried out in this study. Four day food diaries at the baseline and treatment timepoints as well as an interim analysis at 12 weeks were also analysed by the Nutrition Epidemiology Group at Human Nutrition Research, Department of Nutrition and Health Research, Cambridge to monitor the macronutrient content of participants dietary regimens to ensure isoenergetic interventions and the correct proportions of PUFAs, MUFAs and SFAs in each dietary intervention. All study participants were assessed for the following clinical markers of interest at the end of the run-in period (0 week baseline) and after 24 weeks of dietary intervention (treatment) for all five diets:-

1. Insulin sensitivity measures of glucose and insulin by intravenous glucose tolerance tests (IVGTT).
2. Fasting lipid profile; total blood serum levels of low density lipoprotein (LDL), high density lipoprotein (HDL) cholesterol, triglycerides and non-esterified fatty acids (NEFA).
3. Anthropometric measures including, weight, waist circumference and body fat percentage.
4. Seated blood pressure measurements.
5. Haemostatic factors; PAI-1, factor VIIc and fibrinogen.
6. C-reactive protein (CRP) as an indication of inflammatory response.
7. Plasma leptin.
8. Urinary microalbumin:creatinine ratio.
9. Plasma fatty acid composition.

These clinical measures were performed as markers of risk of cardiovascular disease (CVD) to determine if the dietary interventions carried out resulted in any significant changes. These data are not included in this study but results pertaining to the effects of the dietary regimens are referred to in order to facilitate interpretation of the metabonomic analysis.

5.5 Results

A representative ^{1}H-NMR spectrum of a faecal water sample from a RISCK study participant is given in (**Fig. 5.3**). Also provided is a table of typical constituents of faecal water which includes the short chain fatty acids; acetate, propionate, butyrate, valerate and isovalerate, amino acids such as alanine, lysine, glutamine, tyrosine and threonine, also urocanate, β-glucose and the pyrimidine uracil (**Table 5.1**).

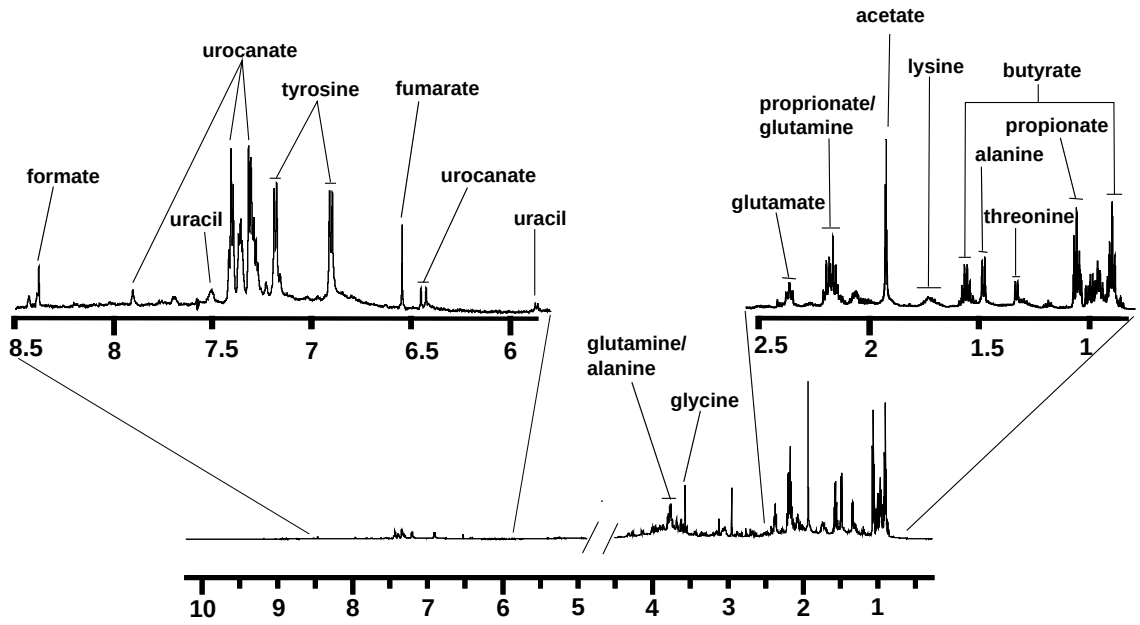


Fig. 5.3 Typical 600 MHz ^{1}H-NMR spectra of Faecal Water obtained from a RISCK study participant.

Assignments were made based on literature and ^{1}H-NMR peak assignment databases available publically and within the section of Biomolecular Medicine.

Table 5.2 ^{1}H-NMR chemical shifts (δ) and multiplicity resonances observed from faecal water

^{1}H-NMR Shift (δ)	Molecule	Multiplicity
0.9	Butyrate (CH_3)	Triplet
1.06	Propionate (CH_3)	Triplet
1.33	Threonine (CH_3)	Doublet
1.48	Alanine (CH_3)	Multiplet
1.56	Butyrate (CH_2)	Doublet
1.73	Lysine (βCH_2)	Multiplet
1.91	Lysine (γCH_2)	Quartet
2.11	Glutamate (αCH_2)	Multiplet
1.92	Acetate (CH_3)	Singlet
2.16	Glutamine (αCH_2)	Triplet
2.19	Propionate (CH_2)	Quartet
2.36	Glutamate (βCH_2)	Doublet of triplets
2.41	Succinate (CH_3)	Singlet
3.03	Lysine (αCH_2)	Triplet
3.06	Tyrosine (CH_2)	Doublet
3.2	Tyrosine (CH_2)	Doublet
3.57	Glycine (CH_3)	Singlet
3.59	Threonine (αCH)	Doublet
3.76	Lysine (CH)	Triplet
3.77	Glutamine/glutamate (-CH)	Triplet
3.79	Alanine (CH)	Quartet
3.94	Tyrosine (CH)	Doublet
4.26	Threonine (βCH)	Multiplet
5.23	β-glucose (CH)	Doublet
5.81	Uracil (αCH)	Doublet
6.4	Urocanate (βCH)	Doublet
6.53	Fumarate (CH)	Singlet
6.91	Tyrosine (phenolic CH)	Doublet
7.2	Tyrosine (phenolic CH)	Doublet
7.31	Urocanate (αCH)	Doublet
7.41	Urocanate (imidazole CH)	Singlet
7.54	Uracil (βCH)	Doublet
7.89	Urocanate (imidazole CH)	Singlet
8.46	Formate (CH)	Singlet

5.5.1 PCA analysis of urine and faecal water

Initial unsupervised chemometric analysis of the NMR data was performed to obtain an overview of variance in the ^{1}H-NMR data X matrix. All subjects post treatment samples were included within the model to obtain an overview of the sources of variation in the data set and for the purposes of detecting outlying samples. It was also performed to identify any evidence of clustering associated with dietary intervention in the urinary and faecal water X matrices. PCA models of both the urinary and faecal water ^{1}H-NMR spectra produced very poor predictive abilities and modelling of variance as reflected in the Q^2 (0.01) and R^2 (0.12) values indicating a high degree of variation within the data. Analyses of the scores plots revealed no evidence of grouping of samples based upon dietary intervention (**Fig. 5.4**). It is unusual to observe clustering of samples in an unsupervised analysis based on dietary changes in human studies in the PCA scores plot due to a predominance of variance unrelated to dietary intervention. This suggests the greatest source of variation based on PCA is due to inter-subject variations between the study participants rather than the effects of dietary change in the X matrix. Four strong outlying (outside of the hotellings ellipse) samples were identified in the urine dataset and 3 strong outliers in the faecal water dataset and were subsequently removed from all subsequent modelling of the datasets. Analysis of the raw ^{1}H-NMR spectra revealed poor spectral acquisition and evidence of poor water suppression in these samples.

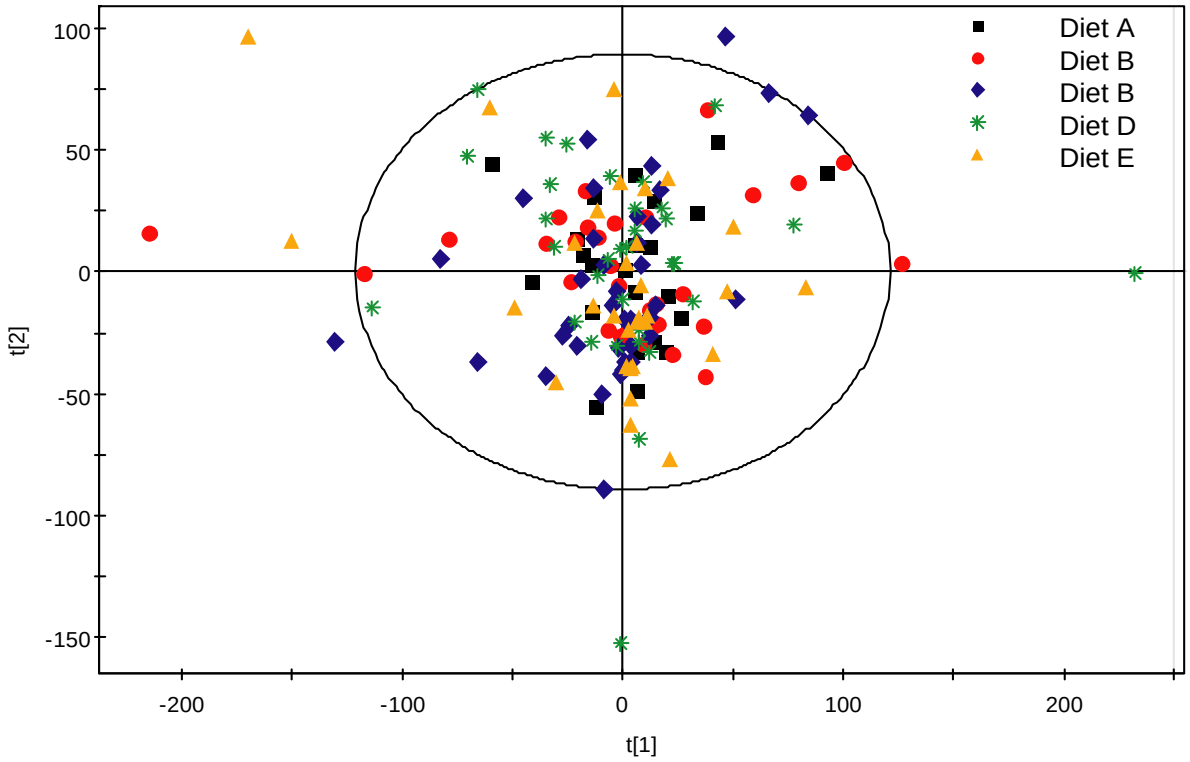


Fig. 5.4 Human urine PCA scores scatter plot of ^{1}H-NMR spectra from all subjects post dietary intervention (24 weeks).

The plot does not identify any separation or grouping of subjects based on dietary intervention, variation is accounted for largely by intersubject variability.

5.5.2 PCA, PLS-DA and O-PLS-DA pair-wise comparisons of control and post dietary intervention urine and faecal water samples

Pair-wise comparisons of baseline (0 week) to treatment (24 week) urine and faecal water samples were performed initially using PCA (**Fig. 5.5**) followed by supervised PLS-DA (**Fig. 5.6**). This was in order to obtain evidence of any statistically significant effect of the 24 week dietary interventions (A-E) on either the urinary or faecal water metabolic phenotype altering with respect to the composition and proportions of energy from fats and the GI of carbohydrates.

Initial PCA analysis of diets A-E demonstrated low cross validation values (Q^2) and relatively low explanation of variance (R^2). As a result the scores plots could not be considered an accurate representation of the sources of variance in the X matrix, clustering based upon dietary intervention could not seen within the scores, however it was anticipated that subtle metabolic variations related to dietary intervention would not be observed in the unsupervised PCA.

Supervised PLS-DA analysis of diets A-E was not suggestive of class discrimination between the baseline and treatment samples due to poor modelling strengths (**Fig. 5.6**) Therefore, evidence of clustering within the PLS-DA scores plots could not be considered a true representation of class membership. Diet B represented the most statistically significant PLS-DA model of the effects of dietary intervention. Analysis of the spectral loadings of diet B (where alterations to total energy derived from PUFAs (~20%) and SFAs (~10%) were made), identified alanine, creatinine, glycine and TMAO as discriminatory of the baseline samples. The modelling strength of this dietary intervention based upon Q^2 (0.17) and R^2 (0.63) indicates a degree of over-fitting of the data and spectral assignments of candidate biomarkers may be misleading (**Fig.5.6**).

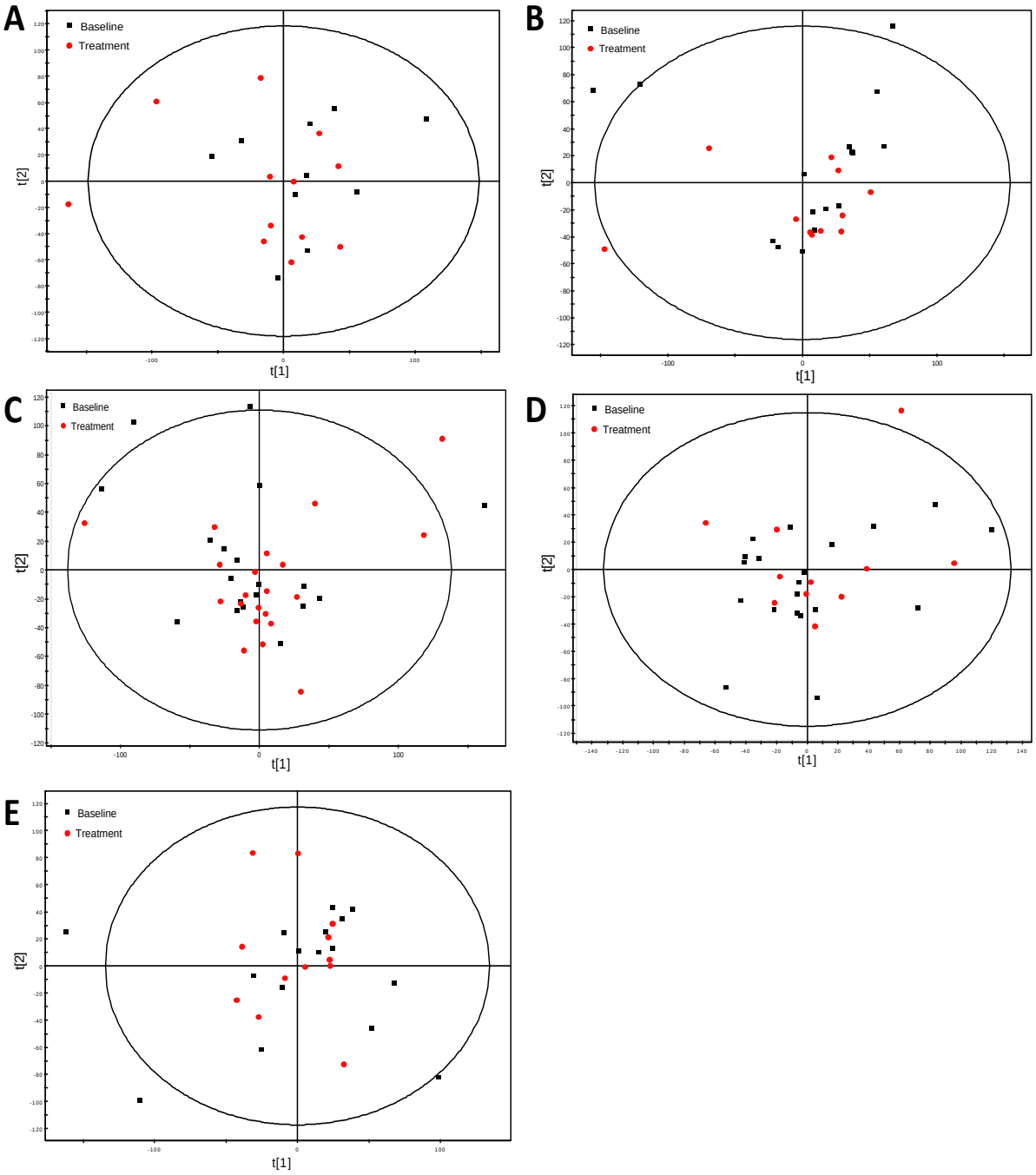


Fig. 5.5 Scores maps showing *t*[1]/*t*[2] PCA scores from the ^{1}H-NMR spectra baseline vs. treatment urine samples of diets **A** (Q^2:0.017; R^2:0.26), **B** (Q^2:0.043; R^2:0.28), **C** (Q^2:0.092 ; R^2:0.25), **D** (Q^2:0.015; R^2:0.23) and **E** (Q^2:0.01; R^2:0.24). Models showed poor explanation of variance in the X matrix. Unit variance scaled.

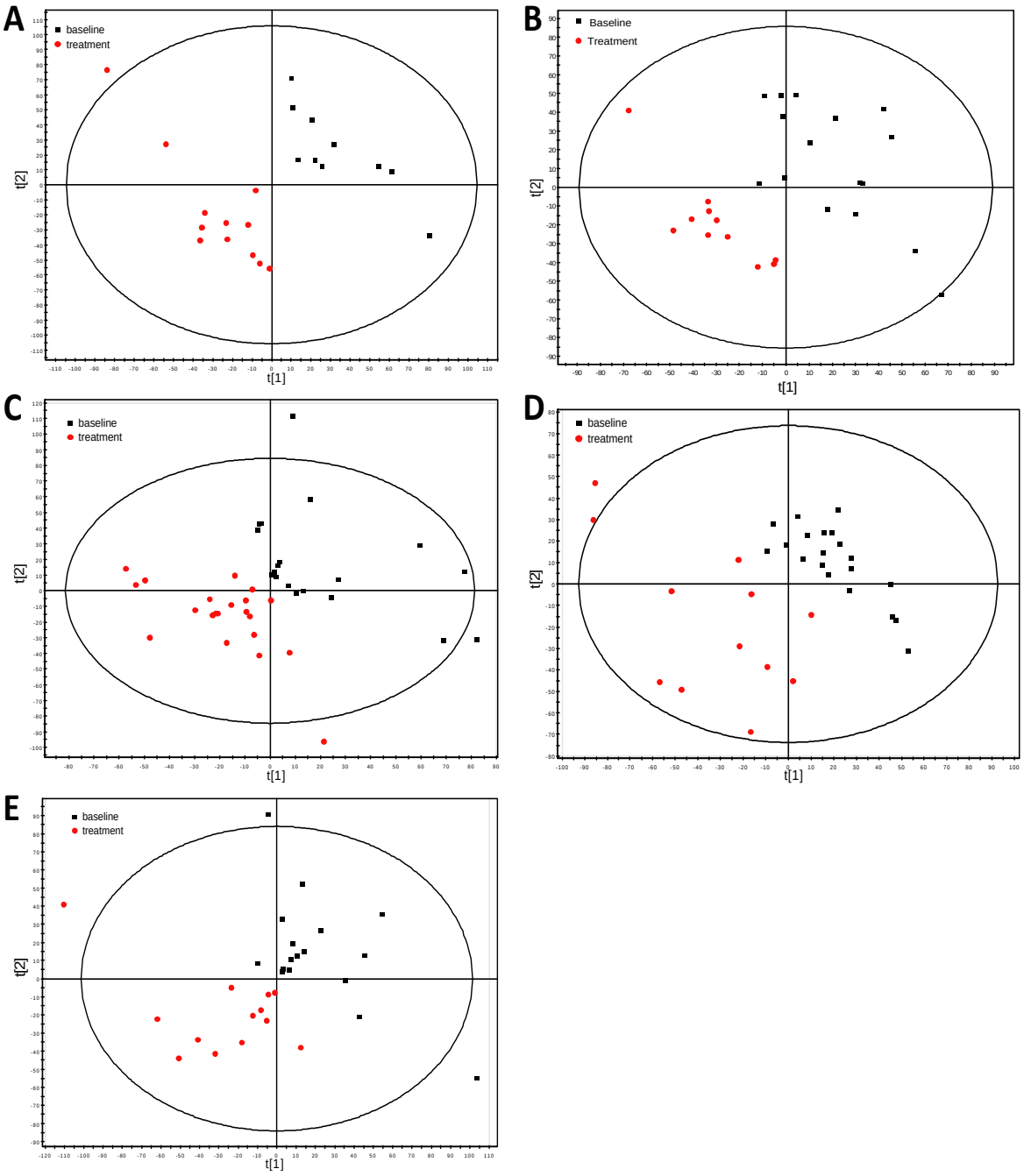


Fig. 5.6 Scores maps showing *t*[1]/*t*[2] PLS-DA scores from the ^{1}H-NMR spectra baseline vs. treatment urine samples of diets **A** (Q^2:-0.027; R^2:0.63), **B** (Q^2:0.17; R^2:0.63), **C** (Q^2:0.034; R^2:0.68), **D** (Q^2:-0.12; R^2:0.79) and **E** (Q^2:-0.056; R^2:0.701). Models showed poor discrimination of class indicating sources of variance in the dataset unrelated to dietary intervention. Unit variance scaled.

Pair-wise comparisons of ^{1}H-NMR spectra of the baseline (0 week) and treatment (24 week) time points for each dietary intervention (A-E) (see section 5.4.1) using O-PLS-DA were performed. Initially, baseline and treatment samples were compared with each other using "Unit Variance" (1/standard deviation) scaling, pre-processed data were bucketed into 0.001ppm buckets and normalised to total spectral area to control for variable dilution of samples.

The O-PLS-DA models were constructed using the ^{1}H-NMR spectral data as X variables and the dietary group as the Y variables. The number of orthogonal components calculated for each model varied based on the values of the cross-validation parameter ($Q^2\hat{Y}$) and the proportion of explained variance of the model parameter ($R^2\hat{Y}$).

During the 4 week run-in period all subjects followed Diet A; a diet typical of the UK where approximately ~38% of energy is obtained from fat with a high (~18%) proportion from saturated fats and 45% of energy from largely simple carbohydrates with a high GI. Comparison of the 0 week to 24 week time points for the group following the control Diet A should therefore not detect any statistically significant differences in the O-PLS-DA models. This null hypothesis was proved correct by a negative predictive ability of the models in the first O-PLS component ($Q^2\hat{Y}$: -0.26) and with poor explanation of sample class to variation within the (X) matrix ($R^2\hat{Y}$: 0.29). However, O-PLS-DA loadings plots of pair-wise comparisons within Diets B, C, D and E also suffered from poor model strength for both urine and faecal water, loadings indicative of discrimination between the baseline and treatment samples for diets A, C, D and E were found to be in very low concentration and close to the threshold of noise in the urine and faecal water samples. It was therefore not possible to construct statistically robust models and identify metabolites in such low concentration in the urine and faecal water samples.

The O-PLS coefficients indicated by the applied colour code with red being the highest correlation with diet, in subjects following Diet B suggested increased excretion of alanine, creatinine, glycine and TMAO as discriminatory of the baseline samples. However poor model strength prevented absolute identification of these metabolites as diagnostic of diet B ($Q^2\hat{Y}$: 0.18; $R^2\hat{Y}$: 0.59). However, this represented the most robust modelling of the dietary interventions (**Fig. 5.7**). The subtle effect of changing dietary fat composition, proportions of energy from fat and glycaemic indexes of carbohydrates in human subjects produced no observable effects within the ^{1}H-NMR spectra of either urine or faecal water. These results suggested inter-subject variation in metabotype dominated the modelling with greater variation between subjects than due to following the dietary interventions assigned.

In an attempt to improve the O-PLS-DA models, alternative data pre-processing steps were utilised; peak alignment, normalisation and scaling. A superior data pre-processing step RSPA or Recursive Segment-wise Peak Alignment was utilised to improve peak alignment within the ^{1}H-NMR spectra [153]. This was followed by changing the normalisation step method to probabilistic quotient normalisation. However improvement of ^{1}H-NMR pre-processing, followed by re-calculation of the O-PLS-DA also failed to improve model strengths sufficiently for robust modelling of metabolic phenotypic change attributable to the dietary interventions the results of which are summarised in (**Table 5.3**).

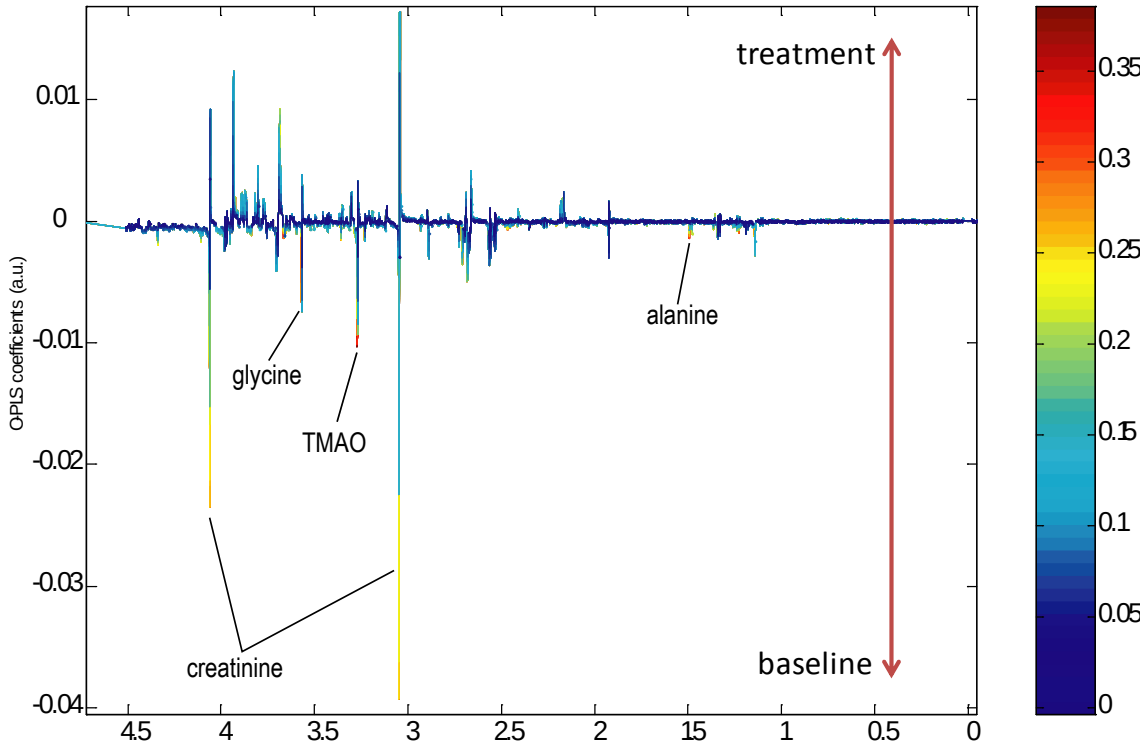


Fig. 5. 7 O-PLS-DA coefficient loadings plot showing dietary intervention B baseline (0 week) vs. Treatment (24 week) of urinary ^{1}H-NMR spectra normalised to total area.

The red colour scaling of alanine, creatinine, glycine and TMAO indicates a reduction in excretion in the treatment group with an increased concentration in the baseline ^{1}H-NMR spectra.

However due to the poor predictive strength of the model reflected in the low values of the $Q^{2}\hat{Y}$ (0.18) and $R^{2}\hat{Y}$ (0.59) statistical parameters these metabolites cannot be considered absolutely diagnostic but indicative of dietary change.

Table 5.3 Summary of baseline vs. treatment O-PLS-DA models of urine and faecal water comparing effects of pre-processing on $Q^2\hat{Y}$ and $R^2\hat{Y}$ values.

RSPA: recursive segment-wise peak alignment; TA: total area; PQ: probabilistic quotient; UV: unit variance; MC: mean centred.

	(Predictive component + Orthogonal component)			**1 + 0**				**1 + 1**				**1 + 2**			
	Pre-processing			*Urine*		*Faecal water*		*Urine*		*Faecal water*		*Urine*		*Faecal water*	
	Peak Alignment	Normalisation	Scaling	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat
Diet A	-	TA	UV	-0.31	0.28	-0.35	0.43	-	-	-	-	-	-	-	-
	-	TA	MC	-0.22	0.37	-0.02	0.11	-	-	-	-	-	-	-	-
	-	PQ	UV	-0.18	0.41	-0.24	0.52	-	-	-0.20	0.87	-	-	-0.12	0.96
	RSPA	TA	UV	-0.27	0.39	-0.12	0.19	-	-	-	-	-	-	-	-
	RSPA	PQ	UV	-0.06	0.20	-0.26	0.53	-	-	-0.19	0.91	-	-	-0.08	0.97
Diet B	-	TA	UV	0.12	0.36	0.11	0.32	0.18	0.59	-	-	-	-	-	-
	-	TA	MC	0.01	0.24	-0.03	0.43	-	-	-0.01	0.67	-	-	-	-
	-	PQ	UV	0.16	0.41	-0.21	0.24	-	-	-	-	-	-	-	-
	RSPA	TA	UV	0.19	0.38	0.03	0.21	-	-	-	-	-	-	-	-
	RSPA	PQ	UV	0.23	0.58	-0.17	0.12	-	-	0.03	0.46	-	-	0.15	0.85
Diet C	-	TA	UV	-0.26	0.64	-0.43	0.58	-	-	-0.31	0.84	-	-	-0.11	0.98
	-	TA	MC	-0.35	0.48	-0.33	0.41	-0.22	0.71	-	-	-0.13	0.92	-	-
	-	PQ	UV	-0.05	0.52	-0.24	0.38	-	-	-0.09	0.59	-	-	-	-
	RSPA	TA	UV	-0.12	0.35	-0.13	0.34	-0.01	0.64	-	-	0.00	0.93	-	-
	RSPA	PQ	UV	-0.18	0.21	-0.24	0.46	-	-	-	-	-	-	-	-
Diet D	-	TA	UV	0.03	0.26	-0.37	0.34	-	-	-	-	-	-	-	-
	-	TA	MC	0.10	0.43	-0.23	0.37	-	-	-	-	-	-	-	-
	-	PQ	UV	-0.02	0.50	-0.09	0.26	0.07	0.67	-	-	0.12	0.81	-	-
	RSPA	TA	UV	-0.13	0.34	-0.02	0.17	-	-	-	-	-	-	-	-
	RSPA	PQ	UV	-0.28	0.47	-0.42	0.18	-	-	-	-	-	-	-	-
Diet E	-	TA	UV	0.18	0.30	0.04	0.57	-	-	-	-	-	-	-	-
	-	TA	MC	0.09	0.24	0.12	0.36	-	-	0.15	0.62	-	-	-	-
	-	PQ	UV	0.11	0.31	0.01	0.28	-	-	-	-	-	-	-	-
	RSPA	TA	UV	-0.02	0.22	-0.23	0.45	0.13	0.53	-	-	0.21	0.86	-	-
	RSPA	PQ	UV	0.09	0.12	-0.03	0.24	-	-	-	-	-	-	-	-

The inter-subject variations in global metabolic phenotype were not surpassed by dietary intervention. Disappointingly, the O-PLS-DA model for diet E failed to discriminate between the 0 week and 24 week time-points in spite of attempts to improve data pre-processing. However the model fitting of diet E was slightly superior to many of the other diets with a one predictive component model in the urine and a one orthogonal component model in the faecal water samples (**Table 5.3**). Associated ^{1}H-NMR variables in the O-PLS-DA loadings plot appeared to be widely dispersed, close to the threshold of noise and subsequently difficult to identify. Diet E represented the greatest change in SFA%, proportion of energy from fat, and reduction in carbohydrate GI from the control diet A and it was anticipated that the greatest change in metabolic phenotype would be observed in study participants following this dietary intervention. The negative $Q^2\hat{Y}$ of many of the models regardless of alteration of pre-processing and data treatment methods is indicative of poor cross validation of the model and a poor predictive ability to distinguish sample classes. Sources of variance in the urinary and faecal water datasets were therefore found to be unrelated to dietary intervention.

5.5.3 O-PLS-DA comparisons of inter-dietary variation

Step-wise comparisons of treatment (24 week) urine and faecal water samples using O-PLS-DA were performed to identify global metabolic changes between the different dietary regimens. 24 week time-point samples were compared between diets A, B, C, D and E in order to address aims of the study. The following experimental iterations of interest were generated and for which O-PLS-DA were performed:-

1. **Diet A vs. B** (HIGH SFA vs. HIGH MUFA)
2. **Diet A vs. C** (HIGH SFA vs. LOW GI, HIGH MUFA)
3. **Diet A vs. D** (HIGH SFA vs. LOW FAT)
4. **Diet A vs. E** (HIGH SFA vs. LOW GI, LOW FAT)
5. **Diet B vs. D** (HIGH MUFA vs. LOW FAT)
6. **Diet D vs. E** (HIGH GI vs. LOW GI)
7. **Diet C vs. E** (HIGH MUFA vs. LOW FAT)
8. **Diet B vs. C** (HIGH GI vs. LOW GI)
9. **Diet D vs. C** (LOW FAT vs. HIGH MUFA, LOW GI)

Unfortunately, analysis of the O-PLS-DA loadings plots for all of the experimental iterations revealed no evidence of metabolic variation due to dietary intervention in either the urinary or faecal water spectral data (**Table 5.4**). Models demonstrated poor predictive power and low O-PLS-DA coefficient values. Superficial changes in the loadings plots were seen in acetate, creatinine and creatine in urinary ^{1}H-NMR spectra. However these were statistically insignificant and could not be considered diagnostic of dietary class due to a very low and often negative value for $Q^2\hat{Y}$. In most cases it was found that only one orthogonal O-PLS component was necessary as the calculation of successive orthogonal components lead to a reduction in the $Q^2\hat{Y}$ parameter and an introduction of noise into the O-PLS-DA models (**Table 5.4**).

Table 5.4 Table showing O-PLS-DA model strengths of experimental iterations of interest in urine and faecal water.

The number of O-PLS components included for each model was based on the value of $Q^2\hat{Y}$ and $R^2\hat{Y}$ to avoid over-fitting the data.

As can be seen each model suffered from poor predictive value with a negative/low $Q^2\hat{Y}$ cross-validation parameter.

(Predictive component + Orthogonal component)	**1 + 0**				**1 + 1**				**1 + 2**			
Biofluid	*Urine*		*Faecal water*		*Urine*		*Faecal water*		*Urine*		*Faecal water*	
Experimental iteration	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat
Diet A vs. B (HIGH SFA vs HIGH MUFA)	-0.39	0.69	-0.12	0.33	-0.12	0.96	-	-	-0.08	0.99	-	-
Diet A vs. C (HIGH SFA vs LOW GI, HIGH MUFA)	-0.02	0.40	-0.09	0.39	-	-	0.01	0.73	-	-	0.03	0.92
Diet A vs. D (HIGH SFA to LOW FAT)	0.09	0.15	-0.19	0.21	-	-	-0.12	0.68	-	-	-	-
Diet A vs. E (HIGH SFA vs LOW GI, LOW FAT)	0.11	0.26	-0.31	0.50	-	-	-0.18	0.80	-	-	-	-
Diet B vs. D (HIGH MUFA vs LOW FAT)	0.08	0.18	-0.54	0.49	-	-	-0.39	0.61	-	-	-	-
Diet D vs. E (HIGH GI vs LOW GI)	-0.25	0.23	-0.07	0.28	0.15	0.50	-	-	-	-	-	-
Diet C vs. E (HIGH MUFA vs LOW FAT)	0.07	0.19	-0.08	0.32	-	-	-	-	-	-	-	-
Diet B vs. C (HIGH GI vs LOW GI)	0.06	0.41	-0.01	0.12	-	-	-	-	-	-	-	-
Diet D vs. C (LOW FAT vs HIGH MUFA, LOW GI)	0.00	0.21	-0.01	0.22	-	-	-	-	-	-	-	-

Further attempts to improve data pre-processing using RSPA and probabilistic quotient normalization followed by recalculation of the O-PLS-DA models also failed to improve detection of significant correlations between diet classes. The dietary interventions set out within this study involved changes in macronutrient component of the diet, the effects of which were not seen to be reflected in the urinary or faecal water metabolic phenotypes.

5.5.4 Bacterial Regression analyses to dietary intervention ^{1}H-NMR spectra

Changes between the baseline (0 week) and treatment (24 week) in the total gut microbial numbers enumerated by fluorescence *in situ* hybridisation (FISH) by our collaborators at the University of Reading were regressed to the ^{1}H-NMR spectra of both urine and faecal water. The changes in total bacterial population numbers as well as of 7 major groups of bacteria of the human gut microbiota were assessed, namely *E rectale/C.coccoides, Bacteriodes spp., Bifidobacterium spp, Lactobacillus/Enterococcus spp., Atopobium gp., F. Prasnutzii* and *C. Histolyticum gp.*

This was performed in order to identify metabolic indices statistically associated with changes in gut microbial profiles due to the dietary interventions A-E and to determine if gut microbial alterations would concurrently produce any change in urinary or faecal water metabolic phenotype. It was anticipated that any correlation between bacterial population changes and global metabolic phenotype would result in identification of statistically significant metabolites within the ^{1}H-NMR data characteristic of a particular bacterial population.

The difference in the bacterial population data from the baseline and treatment groups (**T-B**) was calculated for each of the study subjects (88) for which FISH data was obtained to produce a Y-matrix of bacterial population changes where the dummy matrix was substituted for bacterial population changes (**Table 5.3**).

Table 5.5 Table showing change in bacterial population numbers (Log_{10} bacterial cells/g faeces) between baseline and treatment time-points and paired t-test values. Numbers highlighted in red represent statistically significant changes in bacterial populations attributable to diet according to *p*-statistics. (Bacterial population numbers enumerated by FISH analysis; provided by our collaborators at the University of Reading).

Diet	Log_{10} (bacterial cells/g faeces)	Total bacteria	*E. rectale/ C. coc.*	*Bacteroides spp.*	*Atopobium gp.*	*F. prasnutzii*	*Bifidobac.*	*Lactobac./Enterococc.*	*C. histolyticum gp.*
A	Baseline (B)	10.667	10.042	9.658	9.196	9.522	8.716	8.614	8.375
	Treatment (T)	10.576	10.096	9.778	9.172	9.779	8.678	8.664	8.406
	T-B	-0.090	0.054	0.120	-0.023	0.323	-0.038	0.050	0.031
	Paired t-test (*p*-value)	0.105	0.620	0.251	0.813	**0.018**	0.857	0.739	0.852
B	Baseline (B)	10.049	9.438	9.175	8.681	8.961	8.153	8.064	7.689
	Treatment (T)	9.879	9.406	9.202	8.700	9.005	8.418	8.000	7.748
	T-B	-0.180	-0.034	0.028	0.020	0.046	0.281	-0.068	0.063
	Paired t-test (*p*-value)	**0.023**	0.651	0.756	0.817	0.691	0.124	0.583	0.691
C	Baseline (B)	10.711	10.180	9.702	9.308	9.765	8.773	8.454	8.493
	Treatment (T)	10.539	10.085	9.826	9.186	9.729	8.926	8.491	8.246
	T-B	-0.172	-0.095	0.125	-0.122	-0.036	0.153	0.037	-0.248
	Paired t-test (*p*-value)	**0.005**	0.119	0.132	0.074	0.636	0.376	0.653	0.141
D	Baseline (B)	10.702	10.089	9.722	9.287	9.723	8.823	8.441	8.384
	Treatment (T)	10.683	10.147	9.903	9.423	9.816	9.170	8.596	8.424
	T-B	-0.019	0.058	0.181	0.136	-0.374	0.346	0.155	0.040
	Paired t-test (*p*-value)	0.659	0.531	**0.038**	0.063	0.239	**0.005**	0.063	0.796
E	Baseline (B)	10.610	10.040	9.624	9.180	9.586	8.968	8.465	8.186
	Treatment (T)	10.599	10.042	9.766	9.189	9.723	9.232	8.453	8.304
	T-B	-0.011	0.001	0.142	0.009	0.136	0.264	-0.012	0.118
	Paired t-test	0.863	0.987	0.328	0.933	**0.022**	**0.052**	0.924	0.378

Y-matrices for total bacterial population change and for each of the 7 major gut bacterial groups were used to calculate regression models for the dietary intervention (A-E) regressing to the urinary and faecal water ^{1}H-NMR treatment (week 24) data. The ^{1}H-NMR spectra were optimally pre-processed using RSPA and probabilistic quotient normalisation to improve signal to noise ratio and enhance statistical discrimination.

Regression analyses failed to produce good model strengths and low predictive ability for any of the regression models of diets A, B, C, D, or E calculated against the FISH enumerations of bacterial populations. Typically samples outside of normal F-distributions above the F-critical line in each regression model were not found to be driving the separation or causative of the poor model statistics (**Fig. 5.7**). Removal of these outlying samples did not improve model strengths and often resulted in further samples appearing above the F-critical line following recalculation of the model. Our collaborators found statistically significant changes in some bacterial populations numbers attributable to the dietary intervention by t-test however these were not reflected in the multivariate statistical analyses of the urine and faecal water ^{1}H-NMR spectral datasets (**Table. 5.5**).

In terms of the bacterial population number to ^{1}H-NMR spectral data regression models the upwards and downwards direction of metabolites in the regression models indicates positive and negative correlation of metabolites with bacterial population numbers respectively (**Fig. 5.8**). As can be seen in the regression models generated here, initial observation showed evidence of strongly correlated metabolites, with red coloured regression coefficients in the loadings plot. However, even red coloured features in the loadings plots were considered not to be statistically significant and to have low discriminatory value to the changes seen in members of the gut microbiota. Low and often negative values for the $Q^{2}\hat{Y}$ parameter for each O-PLS-DA model demonstrates low predictive values and cross validation and any interpretation of red coloured O-PLS-DA loadings coefficients could be misleading.

Analysis of the raw ^{1}H-NMR spectra of urine and faecal water for metabolites potentially discriminatory of bacterial population change revealed that in many cases

the metabolite in question was only present in increased concentration for one individual. Succinate in faecal water of Diet D study participants revealed that this peak was only in substantially elevated concentration in only one study participant (**Fig. 5.8**). This perhaps explains the low discriminatory value of the regression models and that most variation in the urine and faecal water are unrelated to bacterial population change (Y) and are more likely based upon unique host-microbe associations for each study participant.

The poor model strengths of the bacterial regression analysis suggested the subtle changes in bacterial population numbers identified by FISH due to dietary intervention did not assert a statistically significant influence on the urinary or faecal water ^{1}H-NMR metabolic profiles. Furthermore the regression analyses did not identify an influence of specific individual bacterial populations upon the global metabolic phenotype. Relatively small but statistically significant changes in bacterial species attributable to dietary intervention would not necessarily effect global metabolic change reflected in the urine and faecal water of study participants and sources of variation unrelated to bacterial population changes were shown to predominate.

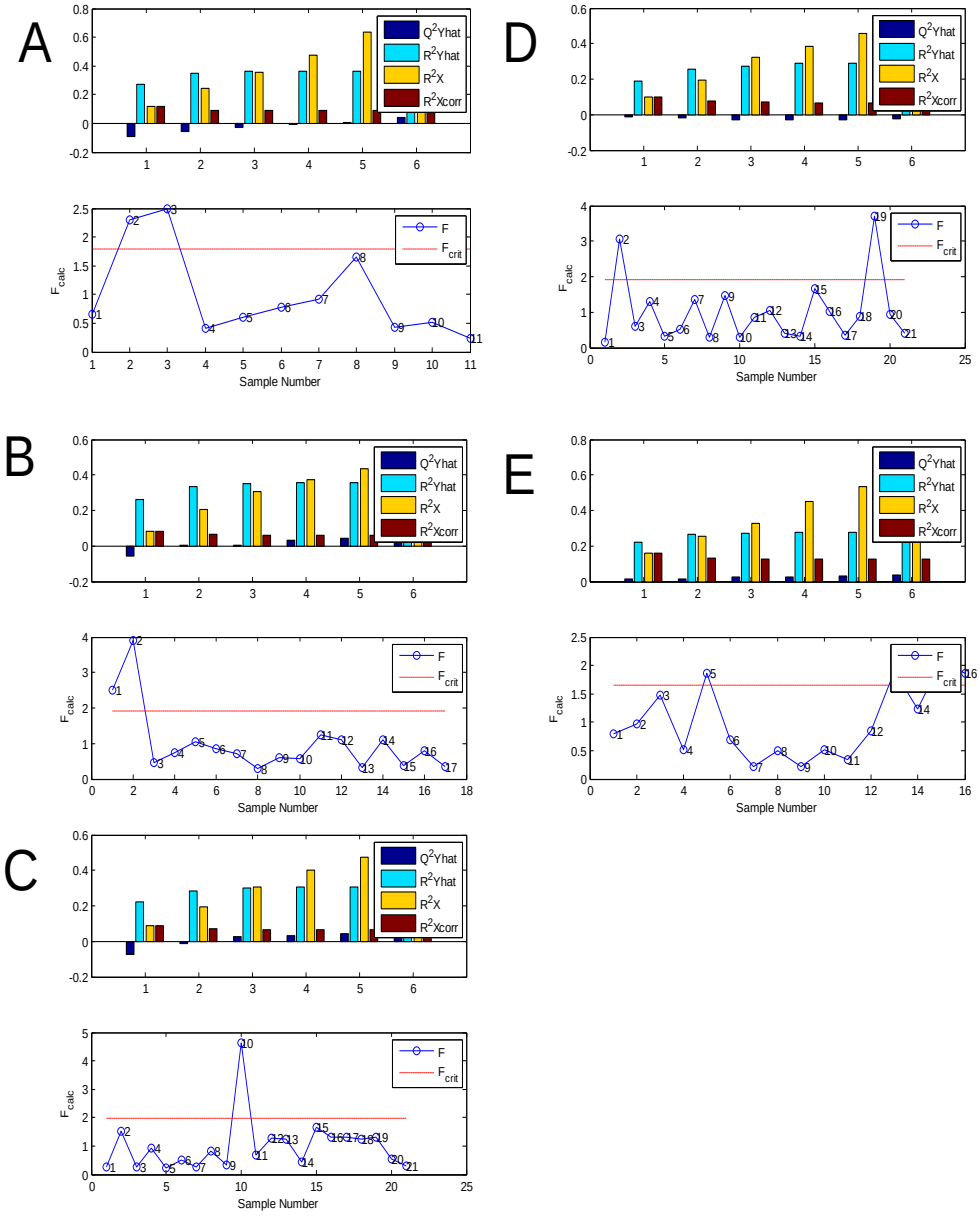


Fig. 5.8 Model overview plots of dietary interventions A, B, C, D and E, showing bacterial regression analyses of total bacterial population changes (calculated by subtraction of treatment from baseline bacterial counts) to ^{1}H-NMR data of human urine.

The models demonstrate poor cross validation and predictive value reflected in the low and negative values of the $Q^2\hat{Y}$ statistical parameter. This suggests a poor correlation of bacterial population changes to urinary metabolic phenotype due to dietary intervention. The calculation of 6 model components gradually improved the model strength. However, this was at the potential cost of overfitting the regression analyses.

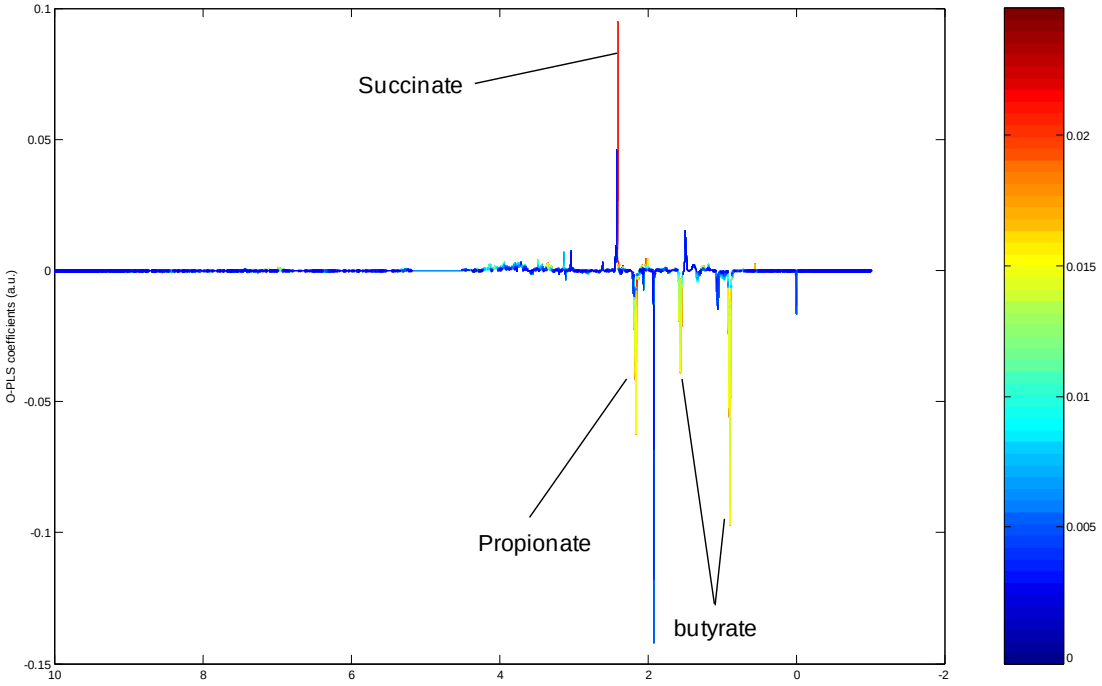


Fig. 5.9 Example O-PLS-DA loadings plot of bacterial regression analysis of faecal water. Total bacterial population change of *E. rectale/C. coccoides gp* .within the Diet D intervention regressed to the ^{1}H-NMR spectra of faecal water.

The direction of features in the loadings plot represent upwards positive correlation of metabolites and downwards negative correlation of metabolites with bacterial population change.

Prima facie evidence suggests strong correlations with gut microbial change. However correlation strength is weak, which is reflected in the low values of the O-PLS-DA coefficients of 0.025. Analysis of the raw ^{1}H-NMR spectra revealed the increased concentration of succinate was attributable to one study participant.

5.5.5 Regression analyses of significant bacterial population changes to all ^{1}H-NMR spectra

In an attempt to improve elucidation of discriminatory metabolites related to bacterial population numbers, regression analyses were performed using all treatment (24 week) ^{1}H-NMR spectra of both urine and faecal water irrespective of dietary class. All treatment ^{1}H-NMR spectra were regressed to statistically significant changes in bacterial population numbers (T-B) as assessed by *p*-statistics by our collaborators at Reading University (**Table. 5.5**). This approach also failed to produce any evidence of metabolite associations with the bacterial population number changes observed (**Table 5.6**). This indicated the changes in major bacterial populations due to dietary intervention could not in this case be related to alteration of metabolic phenotype. Similarly poor values of the $Q^{2}\hat{Y}$ parameter in relation to the $R^{2}\hat{Y}$ parameter was evidential of poor correlation of changes in bacterial population numbers from baseline to treatment values in relation to metabolic phenotype in both urine and faecal water samples. This indicated sources of variance in the dataset which were unrelated to bacterial population changes resulting from dietary interventions.

Table 5.6 Regression analyses of all treatment (24 week) urine and faecal water ^{1}H-NMR spectra to statistically significant bacterial population changes (assessed by *p*-statistics).

(Predictive component + Orthogonal component)	**1 + 0**				**1 + 1**			
Biofluid	*Urine*		*Faecal water*		*Urine*		*Faecal water*	
Bacterial population	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat
Total Bacteria	-0.18	0.24	-0.18	0.32	-0.31	0.62	-0.15	0.62
Bacteroides spp.	0.03	0.28	-0.05	0.29	0.08	0.75	-0.21	0.59
F. prasnutzii	-0.09	0.19	0.09	0.48	-0.18	0.57	-0.01	0.77
Bifidobacterium spp.	-0.01	0.20	-0.33	0.22	-0.03	0.61	-0.57	0.63

5.5.6 Regression analyses of significant bacterial population changes to all ^{1}H-NMR spectra of urine and faecal water

A number of statistically significant blood, faecal, urinary and anthropometric clinical markers were found to be related to dietary intervention by our collaborators at the University of Reading (**Table 5.7**). The changes observed between the baseline and treatment samples in these clinical markers were regressed to the urinary and faecal water ^{1}H-NMR spectral data in order to identify any correlations with urinary or faecal water metabolites. Regression analyses to the observed changes in faecal short chain fatty acids (SCFA) acetate, propionate, isobutyrate and butyrate failed to produce correlations to any metabolites within either the urinary or faecal water ^{1}H-NMR spectra. In an attempt to reconcile the concentrations of these SCFAs measured by our collaborators with the corresponding signals of these SCFAs in the faecal water ^{1}H-NMR spectra, TSP normalised raw spectra were regressed to measured concentrations in the treatment (24 weeks) samples. These regression analyses produced strong correlations of the measured concentrations of these SCFAs to their respective ^{1}H-NMR spectral peaks in the faecal water (**Fig. 5.8**). This result suggested that in principle regression analyses could be made to clinical markers successfully however the biological significance of changes in concentrations and measurements of the clinical markers between the baseline and treatment sample time-points could not be correlated to any metabolic changes in the ^{1}H-NMR spectral profiles (**table 5.8**).

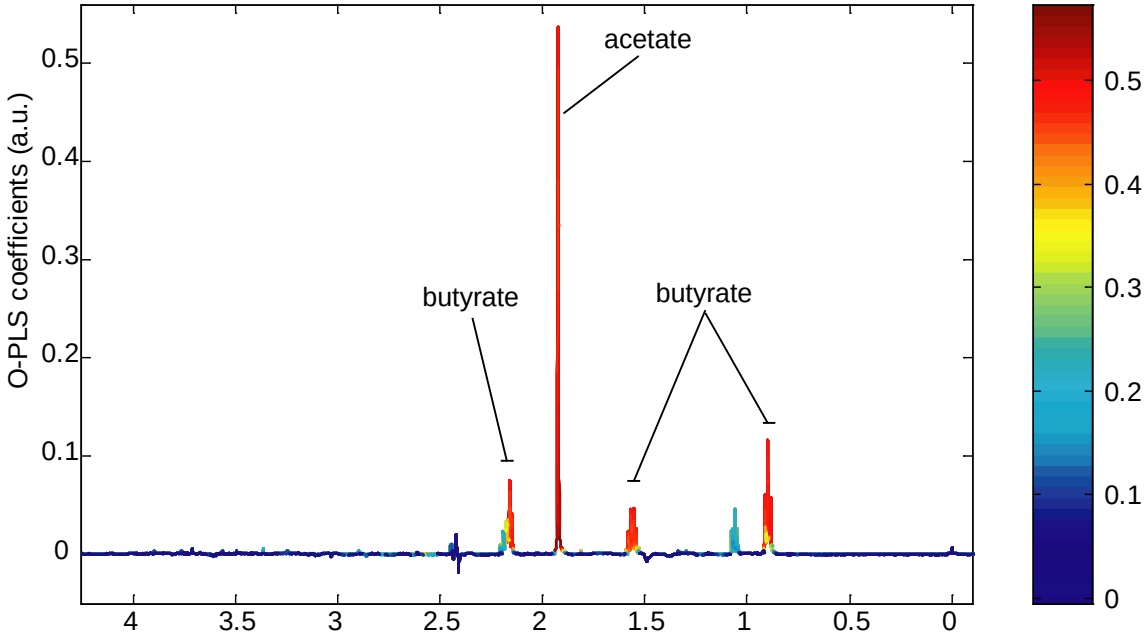


Fig. 5.10 Regression analysis of faecal acetate treatment samples concentration values to TSP normalised faecal water ^{1}H-NMR spectra. Strong correlations can be seen between acetate concentrations measured by capture compound mass spectrometry (CCMS) by our collaborators at the University of Reading and faecal water ^{1}H-NMR signals corresponding to both acetate and butyrate (Q^2:0.41; R^2:0.77).

Interestingly, despite poor modelling strength in the soluble intercellular adhesion molecule (ICAM) regression analysis (Q^2:0.01; R^2:0.19) an indicative association was seen with acetaminophen metabolites in the urine. It is biologically plausible as ICAM is a blood biochemical measure of inflammation and acetaminophen is a very common anti-inflammatory drug. There also appeared to be an indication of an inverse correlation between total cholesterol measured in blood with hippurate in the urine, suggestive of a shift in gut microbial metabolism related to cholesterol levels (Q^2:-0.32; R^2:0.59). However the significance of these findings is extremely low and possibly related to a subset or individual within the study participants. These findings are most probably misleading due to low modelling strengths and predictive ability of the regression analyse in spite of a suggestion of biological plausibility.

Table 5.7 Table showing statistically significant clinical markers (according to *p*-statistics) related to dietary intervention between baseline and treatment time-points. (Clinical measures provided by our collaborators at the University of Reading).

Clinical marker	**Av. Treatment - Av. Baseline**	**Paired t-test (*p* -value)**
Faecal acetate (*m* mol/l)	5.4029	0.0105
Faecal proponiate (*m* mol/l)	1.2933	0.0045
Faecal isobutyrate (*m* mol/l)	0.1876	0.0478
Faecal butyrate (*m* mol/l)	1.7394	0.0093
Waist measurement (cms)	2.0364	0.0008
Body fat %	-1.1524	0.0103
Total cholesterol (*m* mol/l)	-0.2722	0.0494
LDL cholesterol (*m* mol/l)	-0.3182	0.0115
NEFA* (*m* mol/l)	78.8130	0.0130
Av. blood glucose (*m* mol/l)	-0.1905	0.0149
Av. blood insulin (*p* mol/l)	-11.0919	0.0284
ICAM** (*u* /ml)	35.1111	0.0452

*non-esterified fatty acids

**Soluble Inter-cellular adhesion molecule

Table 5.8 Table showing regression analyses of statistically significant clinical markers (according to *p*-statistics) to all treatment (week 24) urine and faecal water ^{1}H-NMR spectra. Models were characterised by poor strength reflected in low $Q^2\hat{Y}$ and high $R^2\hat{Y}$ values. Indicative biologically plausible metabolite associations were observed but could not be considered diagnostic.

(Predictive component + Orthogonal component)	**1 + 0**				**1 + 1**			
Biofluid	*Urine*		*Faecal water*		*Urine*		*Faecal water*	
Clinical marker	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat	Q^2Yhat	R^2Yhat
Faecal acetate (*m* mol/l)	-0.25	0.43	-0.25	0.38	-0.13	0.61	-0.21	0.65
Faecal proponiate (*m* mol/l)	-0.27	0.57	-0.19	0.23	-0.19	0.86	-0.42	0.66
Faecal isobutyrate (*m* mol/l)	-0.18	0.19	-0.10	0.33	-0.55	0.79	-0.23	0.76
Faecal butyrate (*m* mol/l)	-0.17	0.22	-0.18	0.50	-0.33	0.61	-0.08	0.63
Waist measurement (cms)	-0.03	0.20	-0.34	0.24	-0.30	0.68	-0.30	0.68
Body fat %	-0.18	0.19	-0.48	0.23	-0.09	0.55	-0.55	0.51
Total cholesterol (*m* mol/l)	-0.44	0.47	-0.51	0.30	-0.32	0.59	-0.72	0.69
LDL cholesterol (*m* mol/l)	-0.24	0.20	-0.53	0.24	-0.19	0.60	-0.77	0.61
NEFA* (*m* mol/l)	-0.42	0.46	-0.22	0.26	-0.37	0.55	-0.17	0.74
Av. blood glucose (*m* mol/l)	-0.23	0.32	-0.32	0.22	-0.16	0.59	-0.52	0.59
Av. blood insulin (*p* mol/l)	-0.32	0.41	-0.58	0.44	-0.17	0.54	-0.43	0.62
ICAM** (*u* /ml)	0.01	0.19	-0.21	0.21	-0.18	0.58	-0.48	0.53

*non-esterified fatty acids

**Inter-cellular adhesion molecule

5.6 Discussion

In this study dietary interventions in human participants considered at risk of developing MS were carried out over a 24 week period. These dietary interventions differed in the type and quantity of fat and carbohydrates. Manipulation of the diet included either a replacement of SFA with MUFA in a high fat-diet or a substitution of energy from fat for energy from carbohydrate in a low fat diet. Moreover, a simultaneous variation of the type of carbohydrates within each intervention group was performed to alter the GI, comprising low GI- and high GI- diets. Unfortunately, no effects were seen in the urinary or faecal water metabolic phenotypes attributable to dietary intervention by the chemometric methods utilised within this metabonomic study. The changing of proportions of fatty acids and energy from fats and carbohydrate GI would be unlikely to cause any acute homeostatic perturbation detectable within the urine, particularly over a study period of 24 weeks. However, alteration of excretory concentration of endogenous metabolites associated with changes in fatty acid metabolism, energy metabolism and gluconeogenesis may have reasonably been anticipated to be detectable in urine or faecal water samples. The alteration of normal constituents of the diet, fatty acids and GI of carbohydrates, may have required a longer term adjustment to the dietary interventions to produce observable effects in the urinary or faecal water metabolic phenotypes of study participants. Alteration of the gut microbiota as a consequence of the diet may potentially have been reflected in bacterial co-metabolites in the urine, but these were also not detected. ^{1}H-NMR spectroscopic analysis of blood serum may have been useful for identification of changes in lipid profiles and glycoprotein levels. However this was not possible to assess within this study

5.6.1 Correlation of clinical marker changes by regression analysis

Original findings examining CVD risk factors and gut microbial ecology were performed by our collaborators at the University of Reading and are summarised here.

The dietary intervention was found to be successful as analysis of 4-day food diaries supported dietary compliance within the study cohort. Even though diets were designed to be isoenergetic, a significantly lower total energy intake was observed compared to baseline within diets D and E by paired t-test of the food diary analysis (1920.1 ± 569.5 vs 1644.7 ± 479.2 and 2212.5 ± 425.9 vs 1853.7 ± 313.7 kcal, mean ± SD, $p = 0.015$ and $p = 0.002$ respectively for diet D and diet E) with a resultant reduction in body weight and BMI. Blood plasma analysis found a reduction of total cholesterol and triacylglyceride TAG following the consumption of the high MUFA diets and a decrease in LDL-cholesterol after diets C, D and E. There was also no observed increase in HDL-cholesterol compared to baseline in any of the dietary intervention groups. Elevated levels of LDL cholesterol, lower levels of HDL cholesterol and raised TAG is known as the 'atherogenic lipoprotein phenotype-ALP' a risk factor for CHD [229]. A decrease in non-essential fatty acids NEFA was observed in diet E compared to both the control diet A and diet D. Lower fasting NEFA levels are associated with improved insulin sensitivity and post-prandial lipaemia and raised levels with increased insulin resistance. [230, 231]. Furthermore, no significant change was seen in vascular or haemostatic risk factors or anthropometric measures except for reduced body fat percentage in diets D and E, which was almost certainly due to reduced energy intake. Attempts were made to regress the changes seen in serum levels of cholesterol and particularly the elevation of LDL-cholesterol to the ^{1}H-NMR urinary and faecal water datasets but no statistically significant correlations were observed although an indicative inverse correlation with hippurate levels in the urine could be seen. The poor modelling of this regression analysis makes any potential biological significance of this association highly misleading despite biological plausibility

The findings demonstrated a healthy alteration of diet reflected in the CVD risk factors examined. This was not supported by the metabonomic findings due to poor modelling ability of both the urinary and faecal water ^{1}H-NMR spectral data. Nevertheless,

superficial evidence of metabonomic changes in urine were observed particularly in Diet B and changes in Diets A, C, D and E, which were found to be attributable to metabolites in very low concentration in the ^{1}H-NMR spectra, however the poor modelling strengths associated with these tentatively discriminatory spectral features, meant any biological interpretation would be misleading. In spite of rigorous attempts to improve data analysis by use of alternative pre-processing, normalisation and peak-alignment algorithms, models of the ^{1}H-NMR spectral datasets were confounded by low statistical significance. It can be concluded that the alterations of dietary fats consumed, GI of carbohydrates, and the proportions of energy obtained from these produced a subtle effect on both the urinary and faecal water metabolic profile. The analysis of which was dominated by inter-subject variation and it was not anticipated that changes in the CVD risk factors observed would not necessarily produce statistically significant effects in the urinary or faecal water metabolic phenotypes of study participants.

An important consideration in the construction and interpretation of multivariate statistical models is the avoidance of over-fitting of the data and maintaining a good relationship between the Q^2 predictive ability and R^2 explained variation model parameters. Unfortunately within many chemometric models in this study the Q^2 was found to be negative and the R^2 close to 1, in the case of supervised chemometric models this suggests very low dietary class predictive ability and a predominance of variation unrelated to the experimental diets. Inadequate statistical modelling suggested that the individual research questions being posed within the dataset could not be answered and that sources of variation unrelated to dietary intervention were entirely responsible for the variation in the X matrices of both the urinary and faecal water spectral data.

Effects of dietary changes have previously been identified in human urine using metabonomic methods [112, 221, 232]. Metabonomic studies have proven successful when homeostasis is disrupted such as by disease or the introduction of toxins or xenobiotic compounds from food sources, which are more readily detectable by ^{1}H-NMR spectroscopy [193, 233, 234]. The findings suggest that there was no statistically significant effect upon the urinary or faecal water metabolic phenotype due to the dietary

interventions, however a longer term study and a reduction of inter-subject variability may have allowed more reliable identification of metabolic changes attributable to the diets.

5.6.2 Gut Microbial Population findings

Fluorescence *In Situ* Hybridisation (FISH) analysis using 16S rRNA oligonucleotidic probes of faecal samples were performed by our collaborators at the University of Reading to assess population changes due to the five dietary interventions A-E, the findings of which are summarised with regard to the aims of the metabonomic analyses.

An overall decrease in total bacterial numbers were seen in diets A, B and C and were found to be derived from populations not enumerated by FISH analysis, reduction of dietary fat in diets D and E did not produce the same effect. However, decreases in total colonic microbiota and more specifically in Bacteroides related bacteria are known to occur following a high fat diet [219]. Metabonomic regression analyses specifically with bacterial population changes found to be significant by pairwise t-testing with urinary and faecal water ^{1}H-NMR spectral datasets did not support any correlation of metabolic phenotype with diet induced decreases in total bacterial population observed. This is perhaps attributable to an inability to elucidate the effects of the bacterial population changes observed on the global metabolic phenotypes in urine and faecal water of study participants. Sources of variance in the spectral datasets appeared to confound elucidation of statistically representative markers of the bacterial population changes. The bacterial population changes although significant would not necessarily result in sufficiently increased concentrations of urinary or faecal water bacterial co- metabolites to overcome inter-individual variations in gut microbial populations.

Diets D and E produced decreases in body fat percentage and fasted insulin levels and increases in *bifidobacterium*, and *atopobium* populations (known to be carbohydrate degraders in the gut). This finding was possibly due to increased substrate availability in these diets. Bifidobacteria, are known to be beneficial to human health [220] and reduce CVD risk factors and development of type 2 diabetes. Diet D also showed significant increase in Bacteroides population numbers compared to baseline, reductions of which are associated with animal models of obesity and obese humans [235]. Unsupervised chemometric analysis, O-PLS-DA and bacterial regression analyses did not provide any evidence for subsequent changes in metabotype, attributable to the beneficial alteration of gut microbes observed in diets D and E study participants.

Levels of bacterial fermentation products in faecal samples from volunteers revealed increases in the SCFAs; acetate, propionate and butyrate. This is in accordance with known increases in SCFA levels in faeces in obese individuals, despite lower prevalence of polysaccharide degrading bacteria in the gut microbiota [235], and this is perhaps due to decreased intestinal absorption or modulation of gut microbial activity. Evidence of subtle increases in urinary levels of SCFAs was observed in control diet A but this trend was not statistically significant.

No differences in bacterial population levels were observed in high vs. low GI diets, however fasting NEFA levels were shown to decrease following the low GI high carbohydrate dietary intervention (diet E). This study did not include compositional analysis of soluble and insoluble sources of fibre within the dietary interventions which is known to have a GI altering effect; increases of which are recognised to decrease the GI of dietary carbohydrates. The fibre containing dietary constituents may also have had an influence on the gut microbial composition.

Regression of the gut microbial findings to the urinary and faecal water ^{1}H-NMR spectral data was performed in an attempt to identify biomarkers associated with gut microbial population changes and to potentially identify gut microbial co-metabolites that may be associated with response to dietary intervention. Unfortunately, the correlations of these data sets produced poor model strengths for both the urinary and faecal water ^{1}H-NMR spectral data. The very low statistical significance of the bacterial regression analyses suggested that perhaps more dramatic alterations of diet, a longer period of dietary adjustment, an increased number of time-point urine and faecal water collections and a limitation of variation between the study subjects could have produced a greater degree of statistical relevance.

Little is known concerning the impact of both type and quantity of dietary fats on the colonic microbiota although it is of great interest. There is also little understanding of how diets designed to reduce the risk of CVD and the metabolic syndrome impact on the gut microbiota. However, recent studies have shown that the gut microbiota might

be a significant extragenomic contribution to obesity in animal models and in humans [235, 236]. It is of great interest to elucidate the extragenomic contribution of the gut microbiota to obesity and its impact on the global metabolic phenotype in humans, but this was found to be beyond the scope of this study. For the purposes of further metabonomic analysis the use of an additional cohort of this multi-centre study would be useful to augment the urinary and faecal water ^{1}H-NMR datasets already examined. This would allow the selection of a more homogeneous subset, which is age, sex and racial background matched and with a similar range of diagnostic recruitment criteria for more robust multivariate modelling.

5.6.3 Suitability of the RISCK study for Metabonomics

Human studies cannot ever be as highly controlled as animal experimental models; however limitation of experimental variables can be achieved. The subjects involved within this study were selected based on their risk of developing the MS and were not necessarily considered to be metabolic syndrome sufferers when participating in the study. The study participants therefore represented a heterogeneous group in the context of metabonomic analysis and not a discrete disease group. The 130 volunteers recruited at the University of Reading were split into two cohorts of 65 volunteers. This number of subjects which participated in this study represented a statistically significant dataset for the purposes of metabonomic analysis.

The diagnostic recruitment criteria allowed for a minimum expression of two features of the metabolic syndrome based on a weighted scoring system. The scores were weighted according to relevance to development of the metabolic syndrome and reliability of the measurement with respect to biological and analytical variation. However this meant the volunteer group represented a heterogeneous subset of the population which could not easily be classified into a distinct sample or disease group.

The diagnostic criteria included measures of insulin resistance, dyslipidaemia, hypertension, obesity and central adiposity. Screening of the volunteers consisted of single measurements of fasting blood glucose and insulin, fasting serum lipids (e.g. HDL-cholesterol, LDL-cholesterol, TAG), seated diastolic and systolic blood pressure,

waist circumference, weight, height. Although these measurements are excellent for determination of risk of CHD and MS, they were not discrete variables for a controlled metabonomic study. More importantly the volunteers were both male and female, with varying racial group and aged between 30-65 years. These demographic factors are known to affect metabonomic analysis and cause variance within chemometric models [185, 237]. Therefore the study may well have been underpowered.

In order for a model of metabonomic data to demonstrate good statistical separation, differences in metabotype need to transcend inter-subject variations within a sample group for true metabolic differences to be identified. For example for biological matrices such as urine a good O-PLS-DA model should have a $Q_2\hat{Y}$ of around 0.4 if metabolite changes are to be accepted as being a statistically relevant pattern associated with the perturbations of metabolic phenotype related to dietary intervention. The low values of model strengths in this study was highly indicative of a predominance of inter-subject variations overpowering any effects related to changes in the fatty acid composition and GI of carbohydrates in the experimental dietary interventions.

Notable metabonomic differences have been detected in rats between common experimental strains and also demonstration of sexual dimorphism in rat urinary metabolic profiles [97, 173]. Therefore it would be expected that even more profound differences would also be seen in human studies. Indeed urinary metabolic differences were identified between healthy British and Swedish populations of both sexes and in the age range 21-65 [225]. Increased trimethylamine oxide (TMAO) excretion was found to be characteristic of the Swedish population due to increased fish consumption and increasing excretion of taurine was found to be characteristic of study subjects following the Atkins diet.

5.6.4 Conclusions

Ideally in initial metabonomic experimental design, human study participants can be limited in all controllable extraneous variations particularly sex, race and age-matching to limit inter-subject variability. Dietary compliance and limitations on prescription and over the counter drugs is also of great importance in focusing dietary intervention groups, reductions in outlying samples and the creation of a more homogeneous group of study participants. Appropriate variable limitation in this way, can have a favorable effect on biomarker elucidation in metabolically heterogeneous free-living human populations. However, it is important to note that dietary responses even in human studies should transcend experimentally narrowed studies. In spite of rigorous chemometric optimisation, the selection criteria for the RISCK study were found to be too varied between individual participants for accurate discrimination between the dietary interventions carried out. Correlations of faecal water ^{1}H-NMR spectra to SCFA CCMS concentration evaluation carried out by our collaborators suggested that regression analyses were possible in principle in the dataset but the regression analyses to changes in the significant clinical markers and bacterial population numbers were not possible.

The analysis of a larger sample set utilising one or more additional clinical cohorts of the RISCK study may have been sufficient to improve the statistical strength of the chemometric models elucidating the effect of the RISCK study dietary interventions. Also a larger sample set would also have allowed the selection of a more homogeneous group of study participants based upon age, sex, ethnic background and in specific diagnostic criteria which would have perhaps been more suitable for metabonomic analysis. The 24 week time course of the study may not have allowed sufficient time for participants to adjust to the dietary interventions and a longer term study may have been necessary to determine changes in metabolic phenotype, however, Diet D and E participants did show evidence of weight loss however this was found to be an effect of an overall reduction in caloric intake in spite of measures to ensure the diets were isoenergetic. The metabonomic analysis may also possibly have benefitted from the increased sensitivity of UPLC-MS and provided a broader scope for the elucidation of

the effect of the dietary interventions on the human metabotype. Additional urine and faecal water samples for each study participant over the 24 week study period may also have been useful to augment the ^{1}H-NMR datasets and for the purposes of monitoring time-course effects of the dietary intervention.

The optimal design of experiment in human studies represents one of the challenges of modeling mammalian biocomplexity [75]. An optimised design of experiment would perhaps be necessary to create statistically relevant biomarkers of dietary change that could then be validated in large scale epidemiological studies.

Chapter Six: The Effects of High Cruciferous Vegetable Consumption on Human Metabolic Phenotypes

6.1 Background

This study investigated the effects of cruciferous vegetable (CV) consumption on the handling of chemical carcinogens formed when red meat is cooked and on the kinetics of caffeine. Spectroscopic profiling coupled to chemometric pattern recognition was performed to elucidate the effect of CV and meat meal/caffeine dietary intervention on the human metabotype.

6.2 Aims and Objectives

- Determine the effect of a CV dietary intervention in humans, on the urinary metabolite profile.
- From the metabotype, identify urinary biomarkers associated with the CV intervention.
- Determine the effect of meat consumption and caffeine intake on metabotype.
- Investigate the effect of CVs on the metabolic response to meat and caffeine.
- Identify metabolic pathways modified by CV intervention.

6.3 Introduction

6.3.1 Cruciferous Vegetables, phytochemicals and cancer

Epidemiological studies have consistently demonstrated an inverse association between cruciferous vegetable (CV) consumption and risk of cancer [71].

The family *cruciferae* is composed of foods of the species Brassica *oleracea* (cabbage, broccoli, cauliflower, Brussels sprouts, Kale) and >350 other genera (rocket, radish, watercress, horseradish, wasabi). Consumption of common CVs are strongly associated by epidemiological evidence with a decreased risk of lung, stomach, colon and rectal cancer [17].

The chemoprotective power of CV against neoplastic disease is attributed to their abundant levels of glucosinolates (-thioglycoside-N-hydroxysulfates) in the human diet.

For a comprehensive review of CV species and their respective glucosinolates see [238]. Glucosinolates are hydrolysed by thioglucoside enzymes known as myrosinases which are released upon damage of the plant material, e.g. by mastication. This produces the biologically active isothiocyanates (ITC) and indoles [239] (**Fig 6.1**). If myrosinase has been inactivated (e.g., with cooking), consumption of CVs can still result in exposure to the hydrolysis products of the glucosolinate released by members of the intestinal microbiota; Enterobacter *cloacae* and Paracolobactrum *aerogenoides* which have been identified with myrosinase-like activity and which undoubtedly also contribute to ITC exposure, albeit at a lower level. For example, the metabolism of the glucosinolate Sinigrin has been demonstrated in an *in vitro* large intestinal model where ~4% of Sinigrin was converted to allyl isothiocyanate (AITC) by cultured human gut microbes [240].

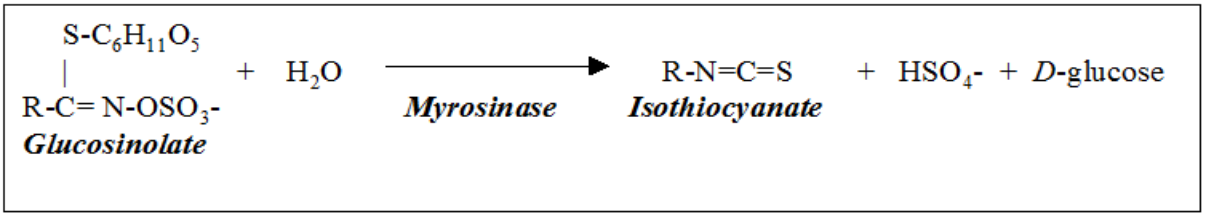


Fig. 6.1 Hydrolysis of glucosinolates by thioglucoside enzymes (myrosinases) to their isothiocyanate derivatives.

Within Brassica species different glucosinolates predominate and yield distinct ITCs. Although Brussels sprouts and broccoli both contain similar glucosinolates they are distinguished by their prevailing glucosinolate species and a complete absence of the dominant species in the other Brassica (**Table 6.1**). In Brussels sprouts Sinigrin represents 35% of total glucosinolate content (radical name Propenyl according to Chemical Abstracts) and Glucoraphanin (4-methylsulphinylbutyl) in Broccoli accounts for 35-60% of the total glucosinolate content [241]. The glucosinolate content in the CVs used in this study was analysed using the methods described in [242] at The Institute of Food Research (Norwich, UK).

Table 6.1 Total glucosinolate content in Brussels sprouts and Broccoli. Table adapted from (Murray S. et al. 2001).

Glucosinolate	**Brussels sprouts***	**Broccoli***
Methylsulphinylpropyl	1.33	0.86
Hydroxybutenyl	1.82	0.1
Methylsulphinylbutyl (sulforaphane)	0	5.02
Propenyl (sinigrin)	4.74	0
Butenyl	0.03	0
Methylthiopropyl	2.46	1.15
Indolyl-3-methyl	2.94	1.65
4-Methoxyindolyl-3-methyl	0.7	0.51
1-Methoxyindolyl-3-methyl	0	0.6
Phenethyl	0	0.07
Total *Content (µmol/g dry weight)*	14.02	9.96

*Fresh vegetables were used to prepare the meals consumed in this study.

Mechanistically plant phytochemicals such as the isothiocyanates have been shown to inhibit the carcinogenic process by modulation of the cellular antioxidant potential, upregulation of phase I and phase II detoxification mechanisms, apoptosis of cancerous cells, alteration of estrogen status and binding and elimination of procarcinogens/carcinogens in *in vivo* and *in vitro* studies [78, 93].

In the case of cruciferous vegetables important phytochemicals include the glucosinolates and their numerous hydrolysis products, flavonoids, kaempfrol and quercetin, and various anthocyanins [83]. Broccoli is known to contain relatively high levels of the glucosinolate glucoraphanin the parent compound of the isothiocyanate sulforaphane [243] Sinigrin the most abundant glucosinolate in Brussels sprouts is primarily converted to allyl isothiocyanate (**Fig. 6.2**).

Glucosinolates are thermally stable and hydrolysis products are formed by enzymatic hydrolysis by the β-thioglucosidase myrosinases. When myrosinase and glucosinolate are released following cell wall disruption the thioglycosidic bond is hydrolysed releasing glucose and forming an unstable thiohydroxymate-*O*-sulphone. Rearrangement of the unstable intermediate leads to formation in the vast majority of cases to isothiocyanate (ITC), nitrile, thiocyanate, epithionitrile, indolic or goitrin-type compounds. The formation of the hydrolysis products is highly dependent on reaction conditions particularly *pH* and at a neutral *pH* ITCs are normally formed.

Glucobrassicin (Radical Indolyl-3-Methyl) is an indolic glucosinolate and forms a significant component of the glucosinolate content in both Broccoli and Brussels sprouts. However in the case of the indolic glucosinolates the metabolic endpoints of hydrolysis are numerous; the indolic ITCs are unstable and form a corresponding alcohol eg. indole-3-carbinol which can condense into dimers (diindolylmethane DIM), trimers or tetramers. Plant factors known as nitrile-forming factors are suspected but not conclusively identified and are thought to contribute to this metabolic endpoint especially in low *pH* conditions [244]. Indolic glucosinolate hydrolysis products are also known to form ascorbigens by the condensation of ascorbic acid and indole-3-carbinol at low *pH* (4-7) conditions in the stomach; these are believed to possess

immunomodulatory properties [245]. Epithionitrile hydrolysis products from terminally unsaturated glucosinolate side chains are thought to form in the presence of the epithiospecifer protein and Fe^{2+} ions [246] (see **Fig 6.3**).

There are many endogenous mammalian and plant derived factors influencing the formation of glucosinolate hydrolysis products and the levels of each possible metabolic endpoint are almost certainly highly variable within the complex human biological matrix. For example, it is known that myrosinase can be destroyed by cooking, particularly microwave cooking [247].

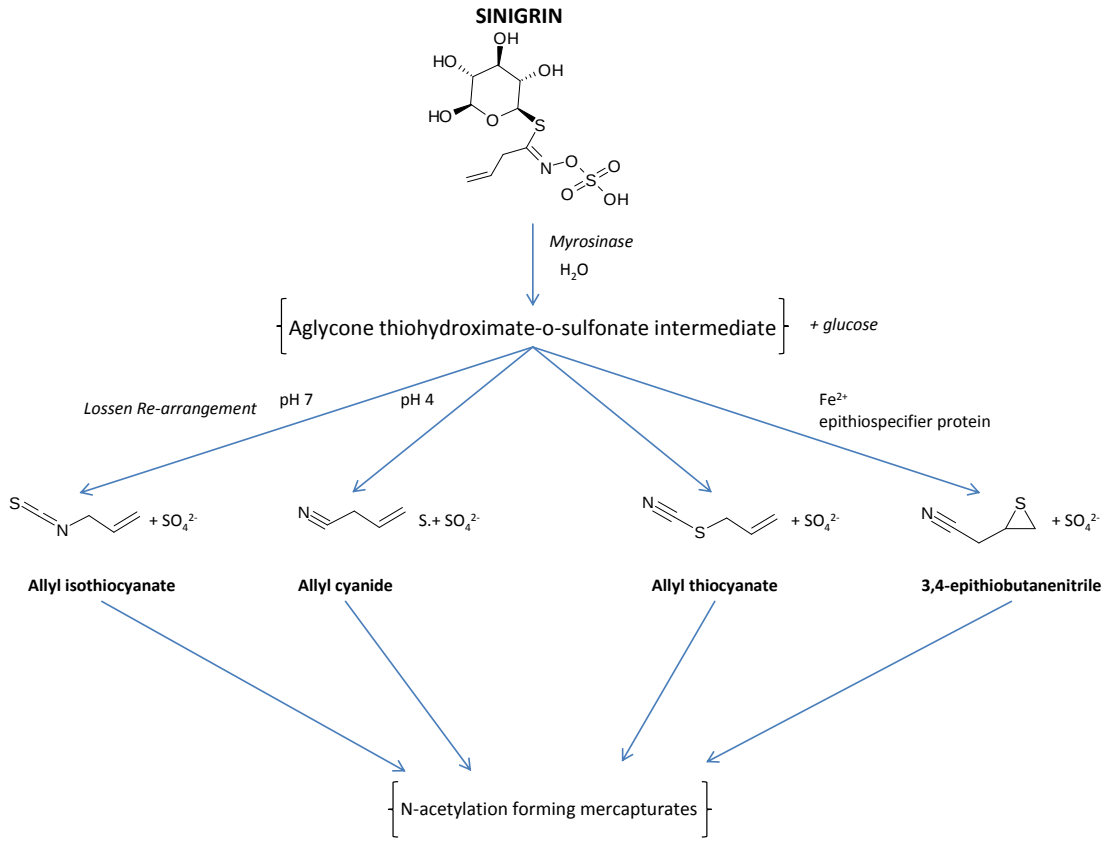


Fig.6.2 Schematic of metabolism of sinigrin the most abundant glucosinolate of Brussels sprouts, a typical aliphatic glucosinolate. Adapted from [72].

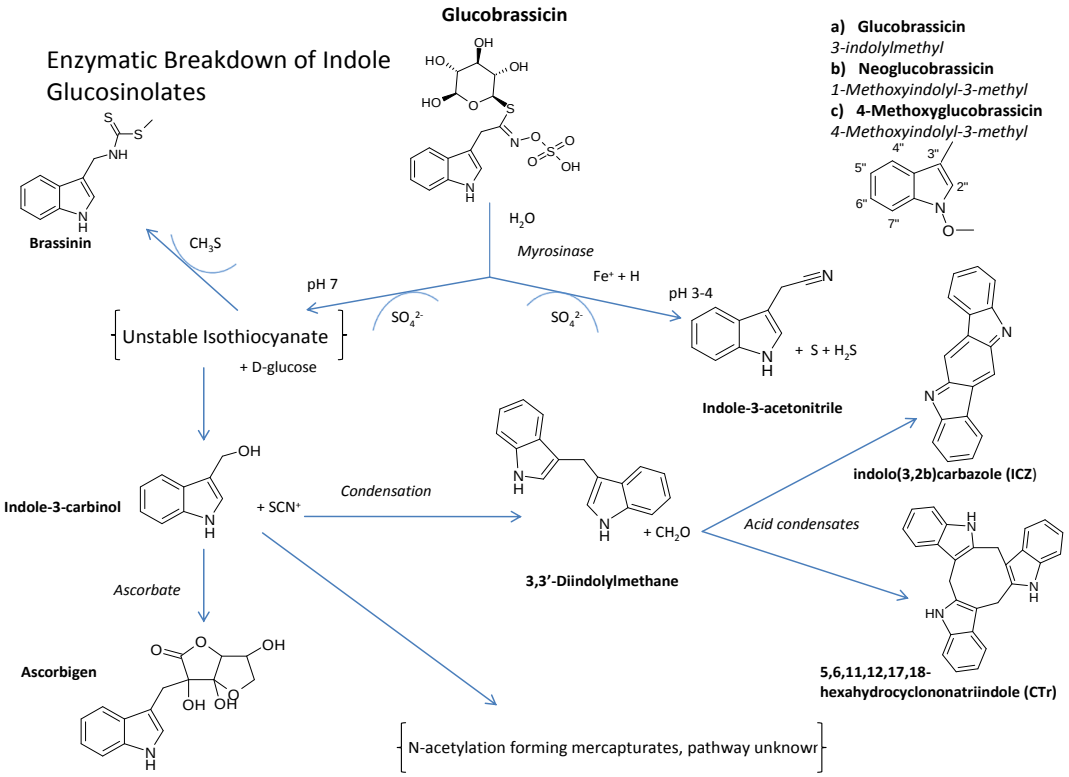


Fig. 6.3 Schematic of metabolism of indolic glucosinolates showing potential metabolic endpoints of hydrolysis. Variations are seen in the position of a methyl group in neoglucobrassicin and 4-methoxyglucobrassicin. Adapted from [72].

6.3.2 Chemoprotective Mechanism

The active principal ITCs and indoles are known to influence several steps in the carcinogenic process by: - (i) Preventing DNA damage by inhibiting the activity of phase 1 enzymes (the cytochrome P450 family and flavin dependent monooxygenases) and inducing phase 2 biotransformation enzymes such as quinone reductase and glutathione S-transferases (GST); (ii) Reducing the DNA – binding, mutagenicity and metabolism of procarcinogens by augmenting the endogenous intracellular antioxidant potential by inducing GSH synthesis; (iii) The induction of apoptosis in malignant and premaligmant cells and the cell cycle arrest of transformed cells [248-251]. Sulforaphane has recently been shown to have potential tumour suppression activity by inhibiting Histone Deacetylase (HDAC) [252].

GSH is the most abundant sulfhydryl agent in cells and plays a key role in cellular defence against a range of exogenous and endogenous toxins. The relationship between

GSH and ITC's is well characterised [253, 254]. GSH abundance is directly linked to the level of uptake of ITC's and the formation of ITC-GSH conjugates, the levels of which are known to be fundamental to the anticarcinogenic activity of ITC's.

Following absorption of the principle active thiol, conjugates of ITC are formed by conjugation with glutathione, a reaction catalysed by GST. Benzyl-ITC, PEITC, allyl ITC and sulforaphane are the most common glucosinolate hydrolysis products and are all substrates for the four major human GSTs- GST A1-1, GST P1-1, GST M1-1 and GST M2-2. However the reaction velocities differ by as much as 700-fold so there is inherent individual variation in the extent to which ITCs are conjugated [251]. Therefore, human ITC exposure is modulated by the type and amount of CV consumed, how the food was prepared, level of mastication and individual differences in GST isozyme levels.

ITCs can directly inhibit the P450 family and also induce expression of either or both of the phase I and II enzyme systems. The mode of induction by the CV compounds is dependent on the distinct structure effects of the ITCs and indole derivatives. Commonly consumed CVs and their isolated constituents are known to modulate two major phase II enzyme systems: GST and Uridine 5' diphospho-glucuronosyltransferase (UGT) [255] and also to alter steroid hormone metabolism [82, 256]. For example, watercress consumption was found to increase UGT glucuronidation of nicotine and tobacco carcinogen metabolites [257].

ITCs like sulforaphane and AITC activate genes *via* the antioxidant/electrophile response element (ARE/EpRE). Sulforaphane induces expression by dissociating the cytoplasmic-anchoring protein Kelch-like ECH associated protein 1 (KeapI) from a transcription factor Nrf2 (NF-E2 p45-related factor 2). Nrf2 translocates to the nucleus forming Nrf2/Maf (musculoaponeurotic fibrosarcoma oncogene) heterodimers, which then activate transcription of Phase II enzymes through ARE/EpRE [258]. This forms part of the MAP-kinase signalling pathway. Indolic derivatives, on the other hand, largely induce phase I enzyme systems *via* xenobiotic response elements (XRE) in the gene promoter. Indoles induce expression of enzymes such as CYP1A, CYP1B, GSTA,

NQO1 (NAD(P)H:quinone oxidoreductase) and UGT by activation of AhR (aryl hydrocarbon receptor) [259]. It is thought that the glucosinolate hydrolysis products that induce both Phase I (XRE) and Phase II (ARE/EpRE) have a chemoprotective effect by increasing the speed of metabolic elimination of carcinogens/procarcinogens.

6.3.3 Carcinogenic heterocyclic amines in cooked meat

Heterocyclic amines (HA's) are common carcinogens formed in cooked meat. Their carcinogenicity has been demonstrated in rodent bioassays [260].

Murray *et al* (2001) found that a high cruciferous vegetable diet had a physiological effect on the carcogenicity of the carcinogen PhIP (2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine) by inducing levels of the xenobiotic metabolising enzyme CYP1A2 with an associated reduction in the levels of excreted red meat heterocyclic aromatic amines. There was also an increase in N^2-glucuronidation of the pro-carcinogenic N-hydroxy-PhIP [81]. The reduction in amine excretion remained even after the withdrawal of the dietary CV and indicated a prolonged physiological response to a high cruciferous vegetable diet.

6.3.4 ITC Elimination

The primary route of *in vivo* metabolism of ITC is by the mercapturic acid pathway, a major pathway for elimination of many xenobiotics [261]. Subsequent stepwise cleavage of glutamine and glycine yields L-cysteine-ITC, which is *N*-acetylated to produce *N*-acetyl-L-cysteine ITC conjugates (mercapturic acids); these are excreted in urine (see **Fig 6.4**).

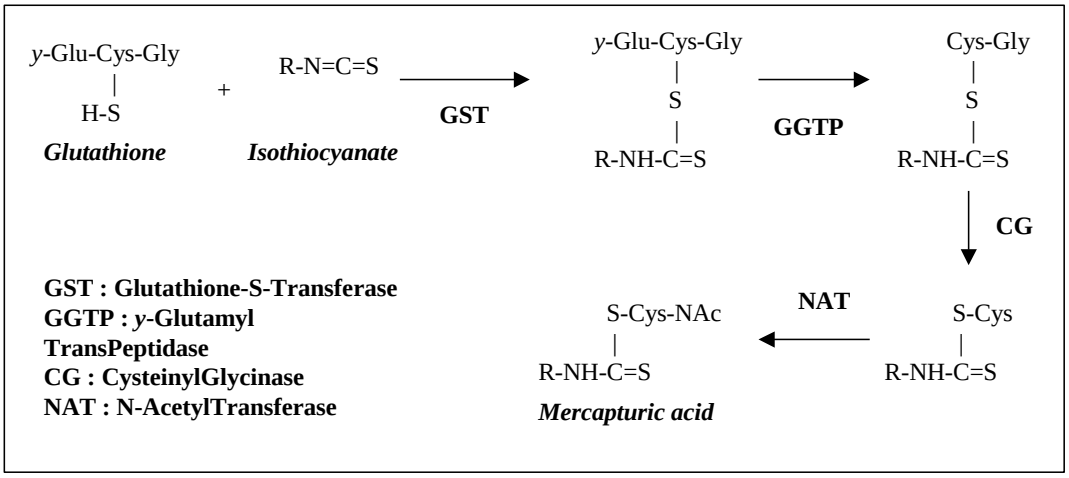


Fig. 6.4. The metabolic elimination of ITC showing the excreted dithiocarbamate (Mercapturic acid)-ITC conjugates.

A detailed examination of the fates of isothiocyanate and glucosinolate hydrolysis products of dietary crucifers has been performed in humans [93]. Human metabolism and urinary excretion of chemoprotective ITCs and glucosinolate hydrolysis products were analysed in this study by paired ion chromatography examination of urine following oral administrations of broccoli containing known levels of sulforaphane. This led to excretion of dithiocarbamates. The observed urinary profiles related to the glucosinolate profiles of the cruciferous vegetables administered. Graded doses of ITCs led to rapid excretion of proportionate amounts with 1st order kinetics [93]. Urine was analysed using paired-ion chromatography. Excretory products were quantified using a cyclo-condensation reaction of isothiocyanates with 1,2-benzenedithiol which yields a product with highly favourable properties for spectroscopic quantification. Humans were shown to convert substantial amounts of isothiocyanates to dithiocarbamates [262]. The initial glutathione adduct and four sequential metabolic products were excreted in the urine (See **Fig 6.4.**). *N*-acetylcysteine mercapturic acids were predominant amongst the

metabolites. Excretion of sulforaphane excretory products was substantial, amounting to <67% of the administered dose 8-24 hours post dose. [262].

London *et al.* (2000) showed that genetic polymorphisms in human GSTs reduced their activity causing a longer circulating half-life of ITC and greater chemoprotective effect. Urinary biomarkers of CV exposure were assessed by Seow, Shi et al (1998). Detectable Dithiocarbamate-ITC levels were shown to be inversely associated with lung cancer risk in men with homozygous deletions in GSTM1 and GSTT1 [263].

Mercapturates are the most commonly studied biomarker of cruciferous vegetable consumption. However the excretion of these biomarkers is highly dependent on inter individual differences in metabolism, abundance of glucosinolates in the vegetable consumed, cooking method, bioavailability, absorption, bacterial metabolism and rate and specific metabolic endpoints of myrosinase hydrolysis. A more stable biomarker of CV consumption is therefore desirable for assessment of epidemiological data, dietary intervention study adherence, dietary assessments in human populations and also new knowledge on the mechanisms of the beneficial effects of CVs.

6.3.5 Evaluation of previous studies of anti-carcinogenic potential of CVs

The majority of evidence concerning the anti-carcinogenicity of cruciferous vegetables and their glucosinolate hydrolysis products comes from *in vivo* animal studies and *in vitro* cell cultures in which the ITC or indole has been used in isolation (for a review see [76]. In animal models and cell systems combinations of ITCs have been shown to confer protection against genotoxic agents better than individual ITCs [264]. However, the focus on the individual components ignores the potential for chemical interactions and synergistic chemoprotective contributions from other compounds found in the normal matrix of cruciferous vegetables. Inter-species differences and the use of doses of the hydrolysed glucosinolate and carcinogenic compounds far in excess of the estimated levels in a normal human diet hinder extrapolation of the results of animal studies in relation to humans. The high levels of ITC's used for *in vitro* studies causes acute cellular stress which may induce apoptosis and cell cycle arrest and may not be a relevant effect in human subjects.

However, *In vivo* human studies have confirmed the beneficial effects of CV at high but realistic consumption levels [79, 80, 82, 265-271] and this is in accordance with the epidemiological evidence of the negative correlation between high cruciferous vegetable consumption and the incidence of diet-related cancers [71].

6.4 Methods

6.4.1 Study Protocol

Twenty healthy male Caucasian subjects were subjected to a three period dietary intervention study (see Murray *et al.* (2001) for details). This study was based at a single centre and planned with the advice of a clinical nutritionist. To eliminate metabolic variables due to ethnic origin, sex and contraceptive steroids, the subjects were male, non-smoking Caucasians that were chosen from the TNOBIBRA Clinical Unit volunteer panel. Potential recruits were excluded if they had any history of drug abuse, chronic alcohol consumption (>28 units/week, 1 unit = 0.5 pint of beer or 1 glass of wine or 1 measure of spirits) or had taken prescribed drugs in a 1 month period prior to commencement of the study. All subjects underwent a medical examination including serum chemistry measurements before the start of the study. The mean age of the subjects was 33 years with an age range of 22–46 years. Mean body weight was 82.0 kg (range 68.6–103.2 kg) and all subjects were within 15% of the desirable weight range for their height.

Each of the three periods comprised 14 days duration (**Fig 6.5**). In periods 1 and 3 a low CV diet was consumed for 12 days and dietary restrictions were applied designed to avoid exposure to known modifiers of enzyme activity. Excluded dietary components were cruciferous vegetables and alliums such as Brussels sprouts, broccoli, calabrese, cabbage, cauliflower, swede, turnip, mustard, cress, onions, leeks, parsley, kale, radish, horseradish, garlic and spinach. On days 11–14 of all three periods, additional dietary restrictions were imposed which included avoidance of cooked and smoked meat and fish products, toasted items and drinks containing caffeine, chocolate, coffee, alcohol and grapefruit juice. The subjects were recommended to consume vegetarian meals based on eggs or cheese. The drugs paracetamol, aspirin and ibuprofen were also prohibited during these periods and to monitor compliance, subjects were asked to keep a written record of food intake during days 10, 11 and 12 of each period of the study. All meals were consumed under supervision in the TNOBIBRA Clinical Unit and subjects were provided with packed lunches for consumption outside the unit.

Following an overnight fast, subjects reported to the TNOBIBRA Clinical Unit on day 13 of each period of the study. After providing a urine sample, each subject was given a meal of 275 g well-cooked lean minced steak, which was consumed with vegetable stock and water. A 10 h urine collection was then commenced. A 2 mg/kg body weight dose of caffeine was administered, dissolved in a cup of decaffeinated coffee. The subjects were provided with a vegetarian lunch and afternoon snack and were released from the unit after completing the 10 h urine collection. All subjects were supplied with containers for a 10–24 h urine collection. On day 14, subjects returned to the TNOBIBRA Clinical Unit with their urine collections. The subjects were then provided with containers for a 24–48 h urine collection that were brought to the unit the following day.

All samples were maintained at -80˚C.

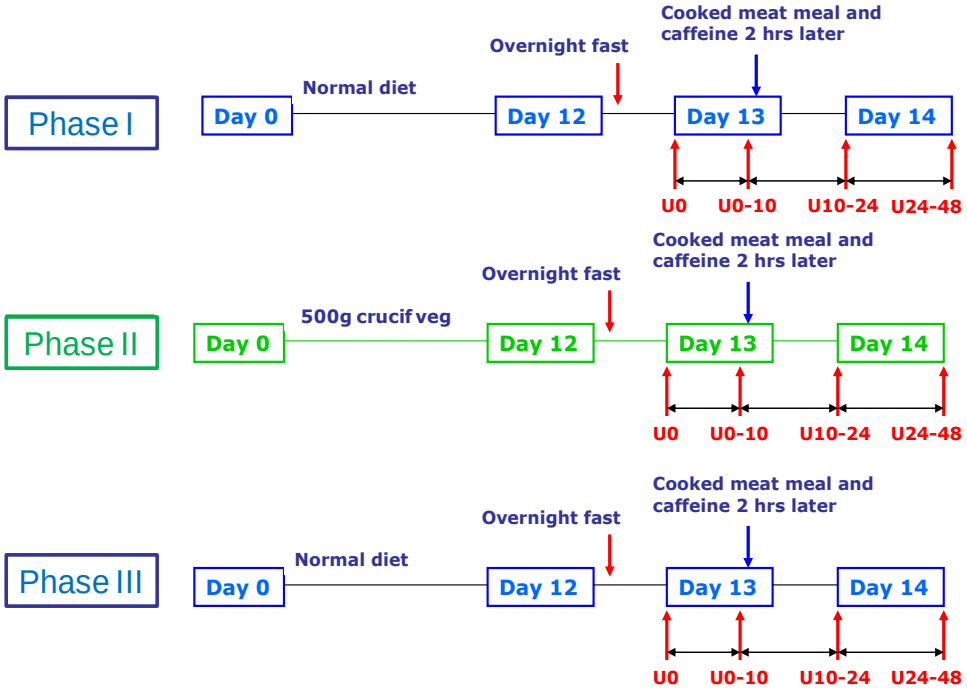


Fig. 6.5 Figure showing the study protocol and the nutritional intervention strategy.

6.4.2 ^{1}H-NMR Spectroscopy and UPLC-MS Methods

All ^{1}H-NMR urine analyses were performed at 600.13MHz on a Bruker DRX 600 Spectrometer (Bruker Analytische GmbH, Rheinstetten, Germany) coupled to a sample-handling system Gilson (Middleton, WI, USA) XL233 interfaced to the XWinNMR data system using parameters described in chapter 2 section 2.5.1. UPLC-MS urine analyses were performed on a Waters LCT premier (Waters MS Technologies, Manchester, UK) orthogonal acceleration time of flight mass spectrometer which was operated in both positive and negative ion mode. Chromatographic separations were performed on a Waters ACQUITY^{TM} UPLC BEH 2.1x 100 mm 1.7 µm C_{18} column (Waters Corp, Milford, USA) using an ACQUITY ^{TM} Ultra Performance Liquid Chromatography system (Waters). UPLC-MS was conducted using standard parameters detailed in chapter 2 section 2.7.

6.4.3 Application of XCMS for Tentative Biomarker Identification.

A significant challenge of interpretation of UPLC-Mass Spectral data is its sheer size which makes manual analysis of raw data almost impossible. The use of chemometric techniques particularly in the form of software packages, such as XCMS[1] [157] are useful for the deconvolution of complex data sets.

Data-mining programs such as XCMS[1] use special algorithms that filter and bin raw mass spectral data then extract chromatographic features identified by *m/z* ratio and retention time. The features are then aligned, normalized and a peak table created which can be subjected to further statistical analysis. XCMS[1] uses a peak picking algorithm that identifies chromatographic features that represent the differences between sample classes and produces a list of Extracted Ion Chromatograms (EICs) and their corresponding box and whisker plots which are sorted in ascending order of confidence based on *p*-value and other statistical measures. Tentative biomarkers of CV consumption were selected based on their class exclusivity within the mass spectral data and by interpretation of the EIC, box and whisker plot, *p*-value and presence within the raw data. The purpose of this process is to identify ions of interest on which the investigations can focus as a possible source of biomarker information.

Initial biomarker identification is based on knowledge of potential exogenous and endogenous candidate metabolites, database mining, (i.e. Human Metabolome Database, Hmdb, *Genome Alberta*) fragmentation analysis and accurate mass atomic formulae identification or a combination of these. Final confirmation of structural identity is achieved using authentic standards. The process of identification is confounded by the limitations of slowly developing metabonomic databases and is an ongoing challenge. Using approaches such as this, HPLC-MS analysis has been used to identify the neutral loss of 129 m/z mercapturates from acetaminophen in human urine [272].

6.4.4 Mercapturic acid urinary extraction

Mercapturate conjugates from human urine samples were extracted using ethyl acetate [273]. Urine samples (5 ml) determined to be highest in concentration of the unknown signatures of interest by PCA analysis were acidified with 0.25ml 4 M HCl. The acidified urine was vortexed for 15 mins with 8 ml of ethyl acetate. The layers were then separated by centrifugation (3000 g_{av}, 5 mins). The ethyl acetate layer was then collected and evaporated to dryness using a rotary evaporator (Savant SpeedVac SPD111V). The extraction procedure was then repeated using an additional 8 ml of ethyl acetate. The reduced extraction was reconstituted with 500µl D_2O and corrected to pH 7.2 with 1M NaOH prior to 400MHz ^{1}H-NMR analysis.

6.4.5 Synthesis of sulforaphane mercapturate

For the synthesis reaction equimolar concentrations of 1mM Sulforaphane and 1mM *N*-acetyl cysteine (both purchased from Sigma-Aldrich, UK) were reacted overnight in methanol and then evaporated to dryness using a rotary evaporator (Savant SpeedVac SPD111V). The reduced synthesis was reconstituted with 500µl D_2O and corrected to pH 7.2 with 1M NaOH prior to 400MHz ^{1}H-NMR analysis.

6.4.6 Brussels sprouts and broccoli aqueous extraction

Eighteen experimental iterations were performed in order to assess different extraction procedures [274], observe the effect of low and neutral pH on enzymatic activity and hydrolysis product formation, simulation of mastication by addition of saliva and the effect of temperature, and raw versus microwave cooked vegetables, as conducted within the original study [81] see (**Table 6.2**)**.**

One ml of either CH_3CN:D_2O (50:50 ; Acetonitrile:Deuterium Oxide) or a D_2O sodium phosphate buffer (0.2M Na_2HPO_4/ 0.2M NaH_2PO_4 pH7.4, 1% (w/v) NaN_3) was used as the aqueous extraction media and combined with approximately 600mg of either Brussels sprouts, broccoli or a 50:50 combination. It was homogenised using a Quiagen tissue lyser (Retsch, ltd.) for 10 mins at the highest setting. An internal ^{1}H-NMR reference standard 3-(trimethylsilyl)propionic-2,2,3,3-$_{d4}$ (TSP) (~0ppm) was added to a final concentration of 0.1mM. Sodium azide (NaN_3) (1% w/v) was added as an antimicrobial to prevent microbial growth during the plant material fermentation and prevent confounding peaks due to bacterial fermentation in the ^{1}H-NMR spectra. The resulting homogenates were allowed to ferment for 4 hours at either physiological temperature (37°C) or room temperature (RT) and neutral pH or low pH (1-3). 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.

Table 6.2 showing CV aqueous extractions experimental iterations.

Experiment no.	Vegetable	Cooking procedure	Homogenisation Solvent	lyophilisation	Temperature	pH
1	Broccoli	Fresh	NaPO4 buffer NaN3	X	RT	7.4
2	Broccoli	Fresh	NaPO4 buffer NaN3	X	37°C	7.4
3	Broccoli	Fresh	NaPO4 buffer NaN3	X	37°C	7.4
4	Broccoli	Fresh	CH3CN:D2O	Freeze drier	RT	7.4
5	Broccoli	Microwaved	NaPO4 buffer NaN3	X	RT	7.4
6	Broccoli	Microwaved	NaPO4 buffer NaN3	X	37°C	7.4
7	Broccoli	Microwaved	NaPO4 buffer NaN3	X	37°C	7.4
8	Broccoli	Microwaved	CH3CN:D2O	Freeze drier	RT	7.4
9	Brussels' Sprout	Fresh	NaPO4 buffer NaN3	X	RT	7.4
10	Brussels' Sprout	Fresh	NaPO4 buffer NaN3	X	37°C	7.4
11	Brussels' Sprout	Fresh	NaPO4 buffer NaN3	X	37°C	7.4
12	Brussels' Sprout	Fresh	CH3CN:D2O	Freeze drier	RT	7.4
13	Brussels' Sprout	Microwaved	NaPO4 buffer NaN3	X	RT	7.4
14	Brussels' Sprout	Microwaved	NaPO4 buffer NaN3	X	37°C	7.4
15	Brussels' Sprout	Microwaved	NaPO4 buffer NaN3	X	37°C	7.4
16	Brussels' Sprout	Microwaved	CH3CN:D2O	Freeze drier	RT	7.4
17	Broccoli/Brussels' Sprout	Fresh	NaPO4 buffer NaN3	X	37°C	2.3
18	Broccoli/Brussels' Sprout	Fresh	NaPO4 buffer NaN3	X	37°C	2.3

6.4.7 Synthesis of S-methyl-L-cysteine sulphoxide and N-acetyl-S-methyl-L-cysteine-sulphoxide standards.

Authentic S-methyl-L-cysteine (SMCS) was purchased (Sigma-Aldrich) and utilised for the synthesis of both S-methyl-L-cysteine sulphoxide (SMCSO) and *N*-acetyl-S-methyl-L-cysteine sulphoxide (NAc-SMCSO). For synthesis of SMCSO, 0.53g of SMCS was dissolved in 5 ml H_2O by use of magnetic stirring, to give a final concentration of 100mM. To this solution was added 200µl of H_2O_2 (according to the protocol of Synge RLM *et. al.*1954). The sulphoxidation reaction was left to occur overnight at room temperature. Complete destruction of the starting product was observed by ^{1}H-NMR and evidence of sulphone formation was also observed. Preparation of NAc-SMCSO (S-methyl mercapturic acid sulphoxide) was conducted with slight modification according to the protocol of Thomson AER *et.al.* (1963).

One gram of SMCS was dissolved in 28ml of ice-cold NaOH kept on ice and with magnetic stirring 1.75ml of acetyl chloride was added cautiously drop-wise. Then a further 14ml of ice-cold NaOH was added followed by another 1.75ml of acetyl chloride also added in a drop-wise fashion. After a 5min mixing time and when all of the acetyl chloride droplets had disappeared the resultant solution was left at 5ºC overnight and then shaken 3 times with 10ml of chloroform. The chloroform extracts were then combined and left to evaporate to dryness under nitrogen. The resultant gummy residue was redissolved in 2ml of ethanol, 5ml H_2O and 10ml warm hexane. The mixture was shaken vigorously and the aqueous layer was aspirated and filtered. The aqueous filtrate was then evaporated to dryness using a speedvac at 60ºC under reduced pressure for 5 hours to remove any residual solvent that may interfere with a clear ^{1}H-NMR spectral signal. The resulting residue was then redissolved in 2ml acetonitrile for solid phase extraction using an acetonitrile:water mobile phase. A 2g C-18 SPE column was first conditioned twice with 5ml acetonitrile. Following this 1ml of the N-Ac-SMCSO solution was carefully loaded onto the column. After an initial 1ml H_2O elution, serial 1 ml elutions of 5, 10, 20, 40, 60, 80, 100% acetonitrile were performed and the eluents collected and lyophilised by freeze-drying. The residues of each eluent fraction were redissolved in 90% H_2O: 10% D_2O containing an internal reference standard TSP prior to ^{1}H-NMR spectroscopic analysis.

6.5 Results

6.5.1 Effect of a cruciferous vegetable dietary intervention in humans, on the urinary metabolic profile.

Representative ^{1}H-NMR urine spectra for one subject (number 12) for both the phase I predose and phase II time-points clearly revealed the presence of 4 singlet peaks (δ2.83, δ2.81, δ2.78 and δ2.75) which discriminate between the restricted CV (phase I) and high CV (phase II) study phases (**Fig.6.6**). These peaks are found in all subjects at varying urinary concentrations post ingestion of CVs. There also appeared to be a time-course decline in signal intensity of these peaks at later sample time-points.

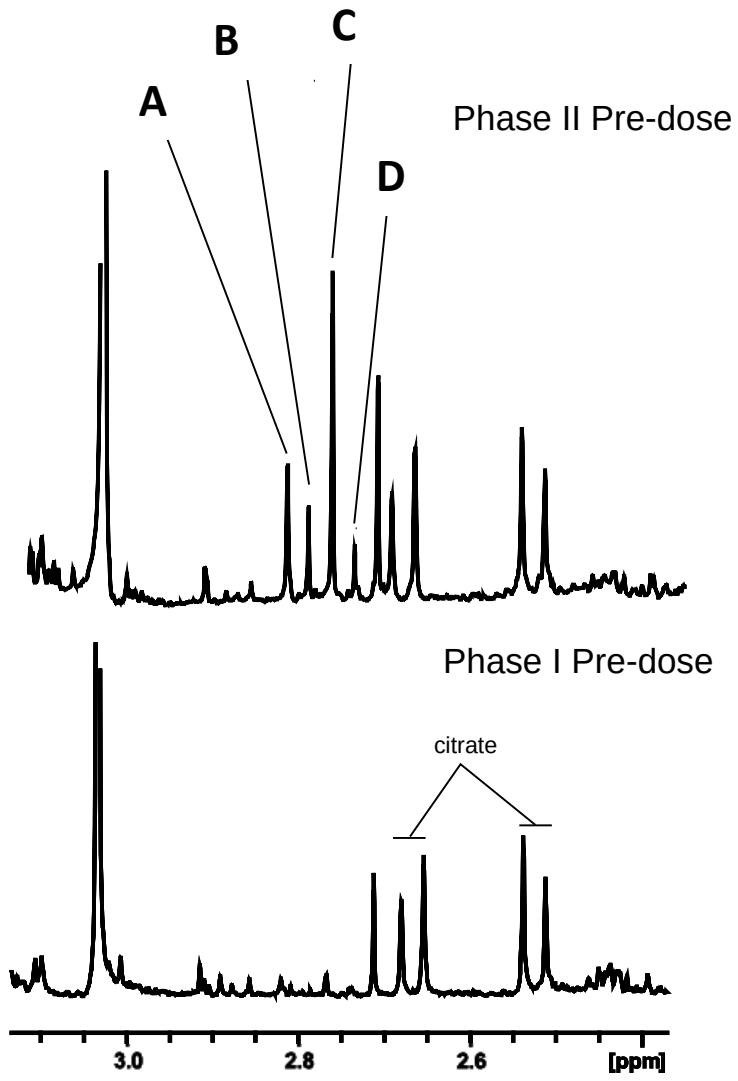


Fig. 6.6 Representative 600 MHz ^{1}H NMR spectra of human urine from restricted CV (bottom: Phase I pre-dose) and high CV diets (top: Phase II pre-dose).
A: δ2.83; **B**: δ2.81; **C**: δ2.78; **D**: δ2.75.

6.5.1.1 Principal Components Analysis all ^{1}H-NMR spectra

Following pre-processing, baseline correction, phasing, correction to the internal standard TSP, pre-processed data was bucketed into 0.001ppm buckets and normalised to total spectral area. Data points in the region δ 4.5-5.5 were removed to eliminate artefacts associated with varying water suppression.

PCA analysis of this entire dataset revealed no clustering of samples based upon class and the dataset was characterised by intersubject variations rather than a significant variation due to the dietary interventions applied (**Fig. 6.7**). This is to be expected in human datasets where dietary interventions produce largely subtle effects. The next step of investigation is to perform supervised discriminant analysis.

Outliers identified in the PCA scores plot were shown to be due to the presence of ethanol in two different pre-dose samples by analysis of the corresponding loadings plot. Subjects were asked to abstain from alcohol during the study period. This suggests a certain level of non-compliance of these subjects to the dietary restrictions. Outliers were removed from all subsequent chemometric analyses.

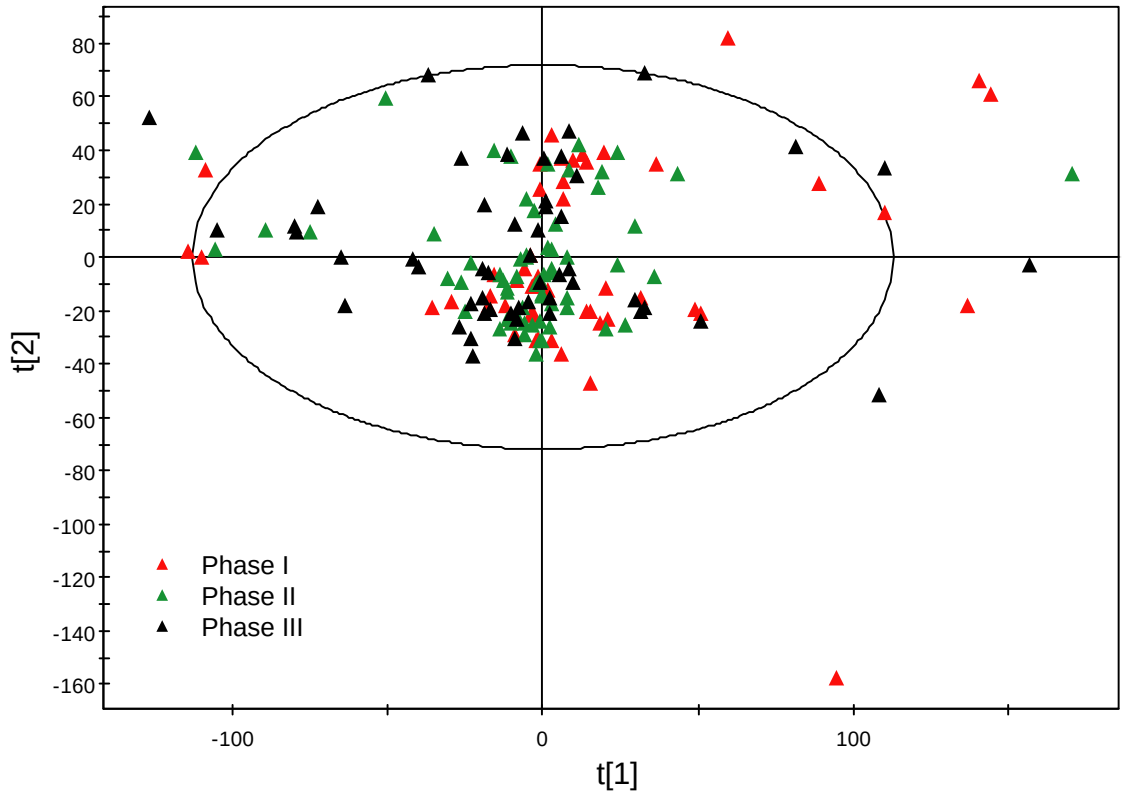


Fig. 6.7 PCA scores plot demonstrating that variation within the NMR spectra X matrix multidimensional space is dominated by intersubject variation with no clear grouping based on sample class. Q^2: 1.3 relatively low cross validation of the model for biological matrices and the R^2: 0.32 shows relatively poor explanation of variance in the X matrix by this model. Outliers were removed prior to further chemometric analyses.

6.5.1.2 O-PLS-DA

Proton NMR spectra for each study phase in all time points were pair-wise compared using O-PLS-DA analysis to determine any significant changes between the run-in normal restricted CV diet (Phase I), the high CV diet (Phase II) and the final washout period of restricted CV intake (Phase III). Every phase was compared with each other using "Unit Variance" (1/standard deviation) scaling. Differentiation and variation in metabolites was observed in the O-PLS-DA regression coefficient plot.

O-PLS-DA coefficient plots of phase I vs. phase II urine samples normalised to total area consistently revealed a discriminatory set of three singlet peaks at δ2.755, δ2.781 and δ2.809, these resonances being typical of bis-allylic compounds (see **Fig. 6.8**). These three singlets were found by comparison of any of the phase II time-points with any of the phase I or phase III time-points urine samples and were highly discriminant of the high CVs period and were therefore considered statistically significant ($Q^2\hat{Y}$: 0.46; $R^2\hat{Y}$: 0.72) biomarkers of CV intake.

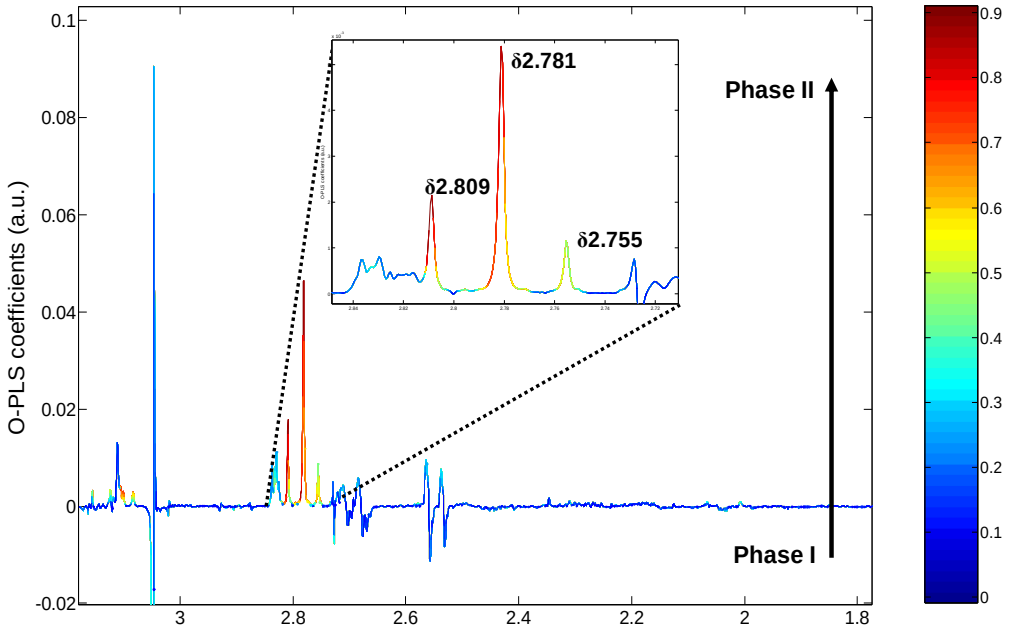


Fig. 6.8 O-PLS-DA coefficient plot; phase I predose (low CV) vs. Phase II predose (high CV) human urine samples showing the singlet resonances characteristic of Phase II. ($Q^2\hat{Y}$: 0.46; $R^2\hat{Y}$: 0.72).

The signal orientation indicates the higher (positive orientation) or lower (negative orientation) concentration of characteristic metabolites in phase II. The colour of the signal reflects the significance of the correlation with diet (red indicates highest significance and blue indicates low significance).

6.5.1.3 STOCSY

In an attempt to further elucidate the molecular structure of these discriminatory signals, a STOCSY analysis was performed on each of these three peaks. This revealed correlations between the three singlet peaks identified by O-PLS-DA particularly between the singlet peak at δ2.781 and δ2.809, with less significant correlation with δ2.755 (**Fig. 6.9**). There was also some evidence of significant correlations with small intensity peaks further up field at δ3.083, δ3.104 and δ3.288. This suggested a structural similarity between the individual singlet peaks or a shared/similar metabolic origin/pathway. The weaker correlations were perhaps indicative of less abundant metabolites derived from a similar metabolic origin or perhaps structural elements related to the diagnostic singlet peaks at lower field. However, low peak intensity and significant overlap of peaks of other urinary constituents in the area of the spectrum prevented correct identification of splitting pattern by STOCSY analysis alone.

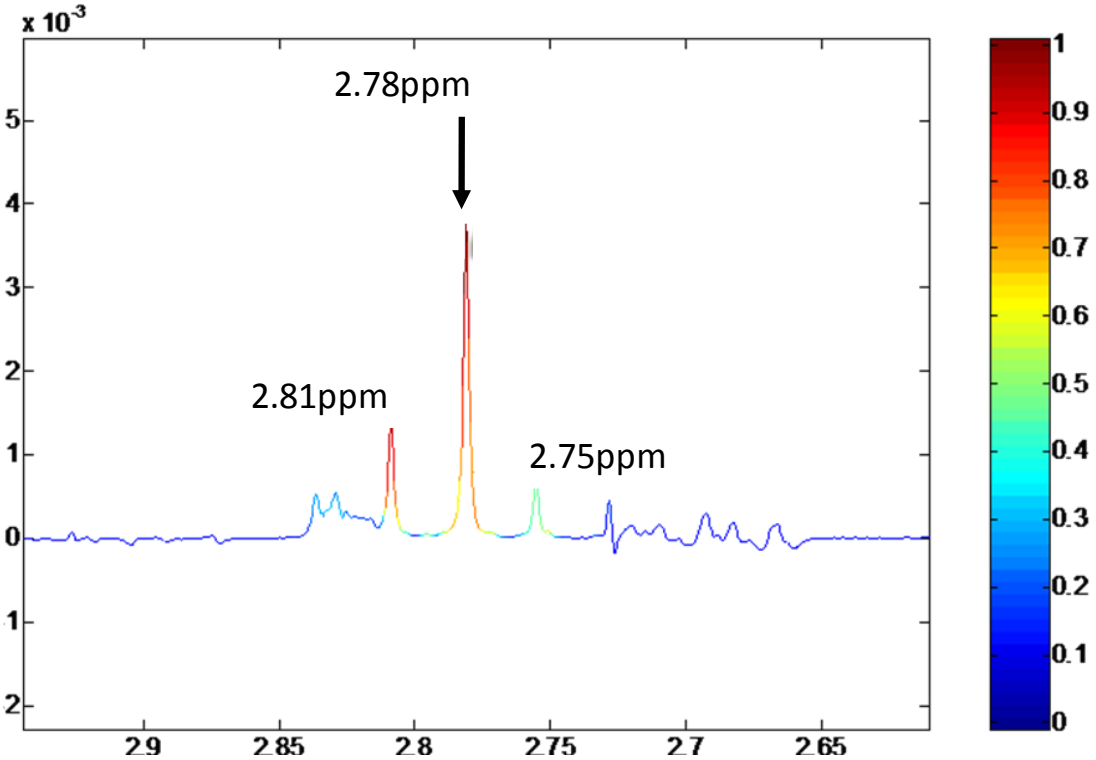


Fig. 6.9 STOCSY analysis of peak 2.78ppm normalised to total area showing correlation between the singlet peaks of interest. Arrows indicates peak from which the analysis was driven from.

6.5.2 NMR spectral biomarker identification

To identify the unknown singlet resonances associated with CV consumption, ^{1}H-NMR databases, literature and in-house sources were initially explored for metabolites with similar resonances, peak intensities, and splitting patterns. This failed to identify a potential candidate with a singlet splitting pattern and chemical shift in this area of the NMR spectra. Pair-wise time course O-PLS-DA analysis of phase II samples revealed a decrease over-time of the singlet resonances following cessation of the CV consumption, consistent with the notion that they are diagnostic biomarkers of CV intake.

6.5.2.1 Mercapturate synthesis and extraction

It is known from previous studies [93] that urinary biomarkers characteristic of CV consumption in humans are the mercapturic acid conjugates of the isothiocyanates sulforaphane and sinigrin, derived from glucosinolates found in broccoli and Brussels sprouts respectively. In order to eliminate these candidate compounds as the discriminating signals, a mercapturate synthesis reaction was performed. Following synthesis the samples were reduced to dryness and the samples reconstituted in D_2O prior to 400MHz ^{1}H-NMR analysis. The samples showed the chemical shift of the acetyl group of the acetyl-cysteine, a singlet at δ2.03. Complex multiplets arising from the magnetically non-equivalent protons surrounding the chiral carbon of the cysteinyl conjugate were also detected. The β-CH_2 representing the AB of the ABX of an ABX spin-system around the dissymmetric chiral centre could be seen at δ3.31 and the α-CH (the X of the ABX) could be observed at δ4.47 (**Fig. 6.10**).

Additionally urine samples were subjected to a mercapturate extraction protocol as described in section 6.4.5, however it was found the extracted mercapturates from the urine were below the concentration limits of detection and structural assignment capability of ^{1}H-NMR analysis under the current conditions.

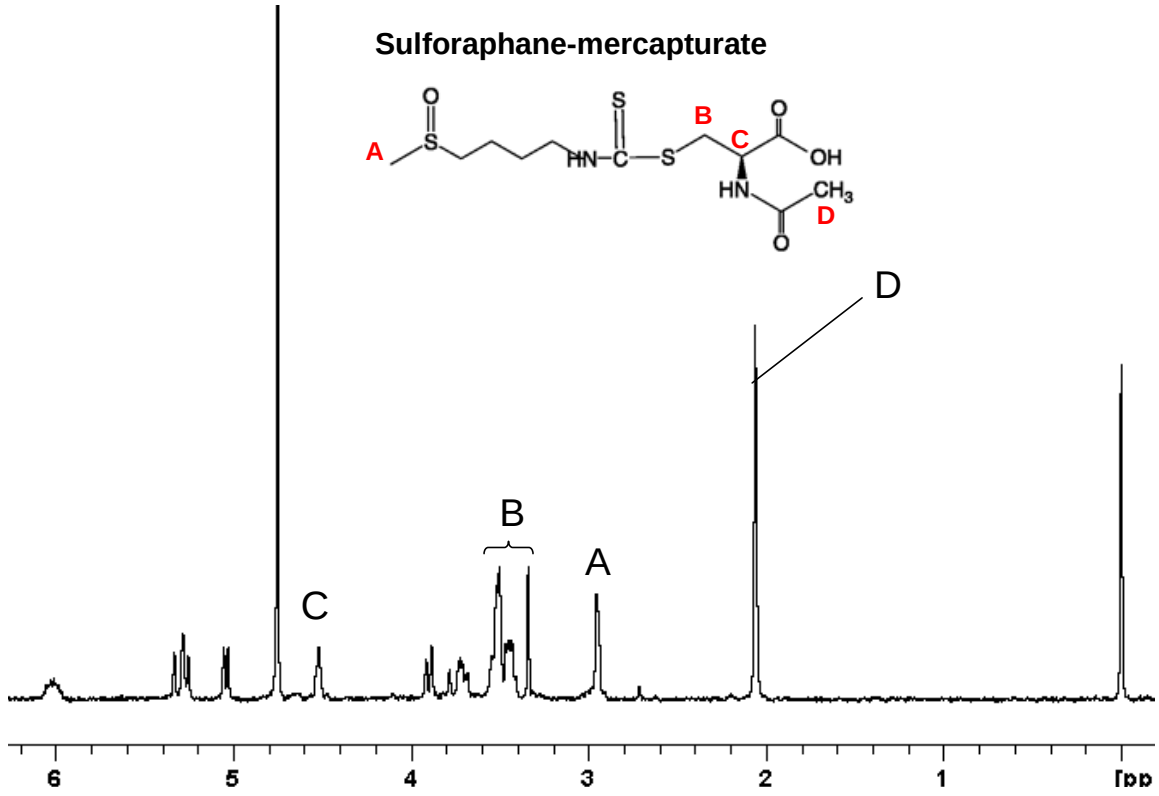


Fig. 6.10 400MHz NMR spectrum of synthetic sulphoraphane-mercapturate conjugate. A: Terminal methyl group of sulforaphane; B: AB of ABX chiral spin system; C: X constituent of ABX spin system; D: Terminal acetyl group of *N*-acetyl-cysteine. Remaining signals within the spectra correspond to the structure of sulforaphane. A well resolved peak of the terminal methyl group can be seen at δ2.98.

The mercapturate conjugate extractions, examined with reference to the sulforaphane-mercapturic acid conjugate (SFMA) synthesis, failed to show any evidence of the singlet peaks of interest at δ2.81, δ2.78 and δ2.75 and these candidate biomarkers were eliminated. Detailed analysis of the raw urine spectra failed to show any evidence of the expected mercapturate signatures.

6.5.2.2 Authentic Standard spiking experiments

The three singlet peak biomarkers of CV consumption were at first assumed to be characteristic of independent small molecular weight metabolites. The relative position in the ^{1}H-NMR spectra between δ2.75-2.81 and the absence of any further structural information suggested chemically isolated methyl groups producing a singlet splitting pattern. On this basis, a list of small molecular weight metabolites with biological plausibility was generated and authentic standard compounds were spiked into phase I control urine to attempt to identify singlet peaks. These analyses failed to identify any similarity in chemical shift between the unknowns and standards. However, the chemical shifts of both methanesulfinic acid and methanesulphonic acid suggested an S-methyl-sulphoxide functional group may be a possibility for the unknown singlet peaks associated with CV consumption (**Table. 6.3**).

Table 6.3 Chemical shifts of potential sulphur/nitrogen containing compounds with biological relevance and a singlet splitting pattern.

The similarity in chemical shift of both methanesulfinic acid and methanesulphonic acid suggested an S-methyl-sulphoxide functional group for the unknown singlet peaks associated with CV consumption.

Standard compound	Chemical shift (ppmδ)
Dimethyl disulphide	2.43
Dimethyl sulphide	2.16
Methyl sulphoxide	3.02
Dimethyl sulphoxide	2.71
Methyl amine	2.57
Dimethyl amine	2.34
Trimethyl amine	2.87
Methanesulfinic acid	2.69
Methyl thiocyanate	2.65
Methanesulphonic acid	2.812
Dimethylsulphone	2.93
Dimethyl sulphate	3.69
N-Methylthiourea	2.961
Methylisothiocyanate	3.278
Methyldithiocarbamate	2.998
S-Methyl methanethiosulfonate MMTS	2.764, 3.31
Dimethylglycine	2.912, 3.723

6.5.2.3 Improved Peak Alignment and Normalisation

Interrogation of the raw NMR spectra identified a singlet peak in the phase II urines which showed evidence of pH shifting or a poor alignment of spectra. This peak was exclusive to the CV consumption sample class and it was therefore important to attempt to correctly align this peak prior to chemometric analysis. A superior data preprocessing step RSPA or recursive segment-wise peak alignment [153] was utilised to improve peak alignment within the NMR spectra. This was further improved by using a superior normalisation step by normalising to the probabilistic quotient [154]. Improved pre-processing, followed by re-calculation of the O-PLS-DA (**Fig. 6.11)** and STOCSY models identified a further singlet biomarker diagnostic of CV consumption at δ2.83. This indicated the presence of 4 distinct metabolites with potentially a shared metabolic origin and shared structural elements associated with CV consumption in the range δ2.75-2.83. The statistical correlation of these 4 peaks was further confirmed by STOCSY analysis (**Fig. 6.12**). Excision of the 4 unknown singlet peaks from the NMR spectra and recalculation of the O-PLS-DA model produced a strong correlation of peaks located further up field of the 4 singlets. This indicated that peaks in the region δ3-3.5 were also significant discriminators of CV consumption and could provide further structural information on the unknown metabolites identified at δ2.75-2.83.

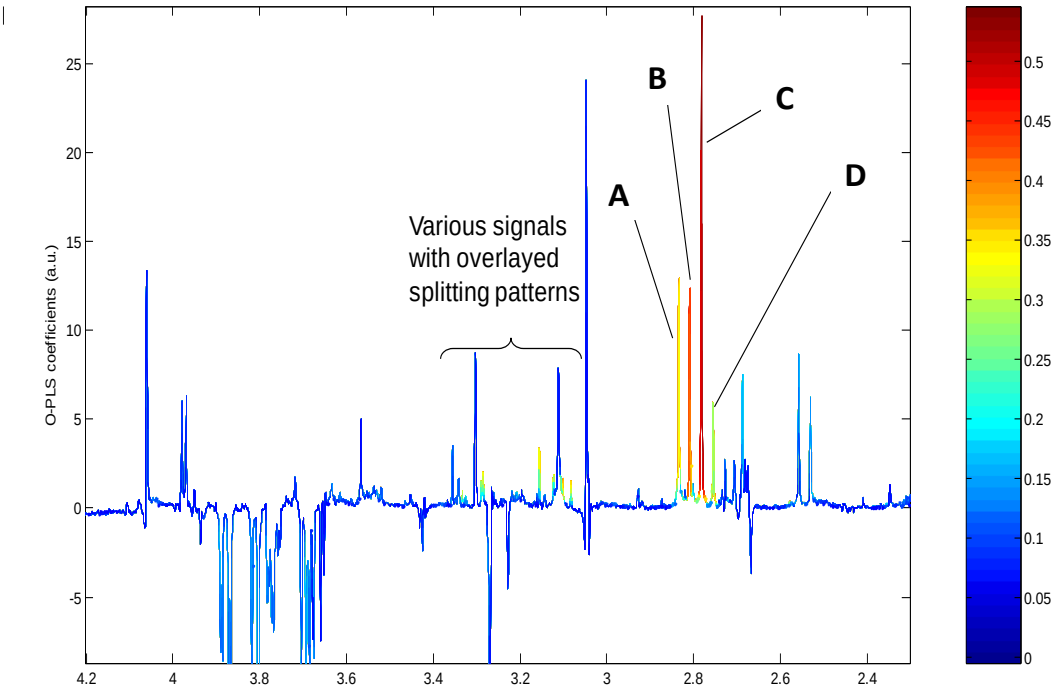


Fig. 6.11 Following RSPA and normalisation to probabilistic quotient, O-PLS-DA identified a previously poorly aligned singlet peak at δ2.83 additionally characteristic of CV consumption. A: δ2.83; B: δ2.81; C: δ2.78; D: δ2.75.

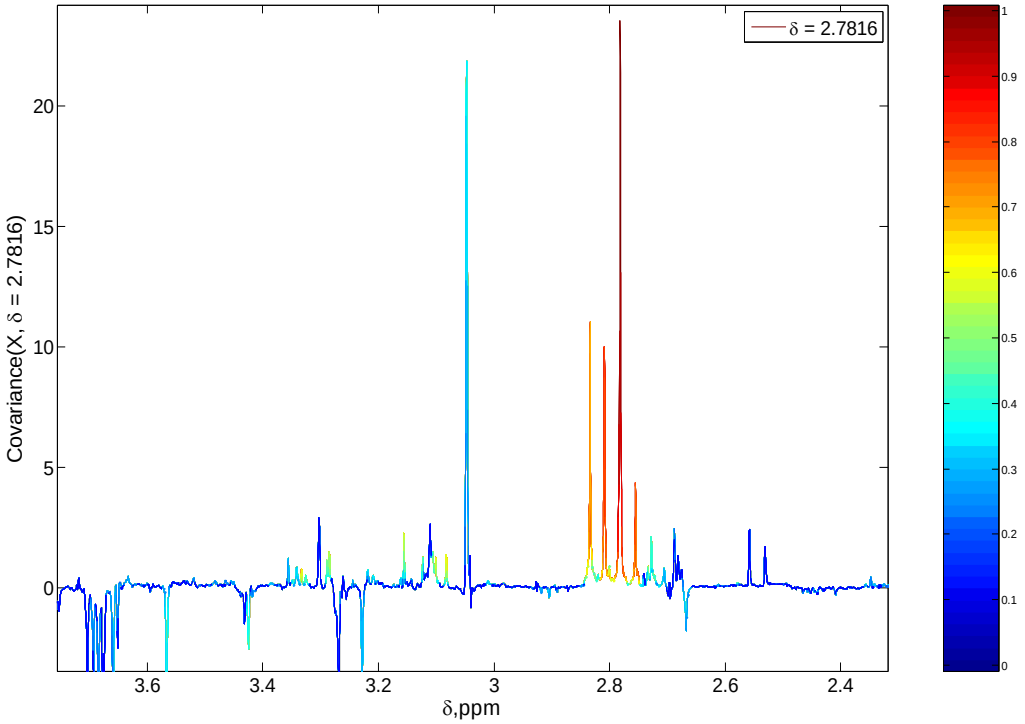


Fig. 6.12 STOCSY analysis confirms statistical correlations between the four unknown singlets and with extra connectivity with peaks between δ3-3.5.

6.5.2.4 ^{1}H–^{1}H *J*-resolved spectra and COSY of Phase II urine

The 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 1-dimensional spectrum. The splitting pattern of these spectral features was confounded by spectral overlap of metabolites in greater abundance within the urine. However, the identity of these peaks is essential for correct structural identification of the unknown metabolite(s). The problem of structural identity and identification of splitting pattern in this crowded area of the spectrum was approached using a 2D ^{1}H-^{1}H *J*-Res NMR experiment. This two dimensional NMR experiment separates the chemical shift in one dimension from the scalar couplings on the second, providing a visualisation of the underlying *J*-couplings of ^{1}D peaks in a two dimensional plot.

The membership of convoluted peaks to a functional group can be ascertained and the splitting patterns visualised along with the potential to observe *J*-couplings of an unknown set of spectral features. In order to concentrate the biomarker identified as characteristic of CV consumption, the scores plot of a simple PCA plot incorporating only the X matrix data for the 4 unknown singlet peaks was performed to identify the subject and urine sample containing the greatest concentration of the unknown singlets. This was found to be the predose sample of participant number 12. Ten ml of this sample was lyophilised by freeze-drying and reconstituted in 1ml of H_2O. The sample was then loaded onto an SPE column, and eluted with a methanol gradient. The majority of the unknown singlet peaks were found to be in the 5% methanol fraction. This SPE sample was subsequently utilised for improved two dimensional NMR experimentation.

The *J*-Res experiment confirmed the identity of the unknown peaks in this crowded area of the spectra to be the peaks of an ABX spin system and also 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) (**Fig. 6.13**). However the X doublet of doublets at δ4.38 was not identified in the O-PLS-DA loadings coefficient plot the proximity of this peak to that of water resulted 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 (**Fig. 6.14**). 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.

Following the evidence of a similarity in chemical shift to the S-methyl-sulphoxide groups of methanesulphinic acid and methanesulphonic acid, the *J*-res NMR experiment was suggestive of a sulphur containing amino acid. It is known from the literature that CVs ubiquitously contain a methylated cysteine; SMCSO [275].

NMR prediction using the ACD labs software confirmed that the splitting patterns and presence of an a chiral centre, and therefore an ABX spin system in both SMCSO and its principal mammalian metabolite NAc-SMCSO are supported by the findings of the *J*-Res experiment [276, 277]. Correlations between the A, B and X of the spin system were subsequently confirmed using a COSY experiment (**Fig. 6.15**).

Evidence of correlations exists between the NMR spectral features at δ3.10 and δ3.30 and the doublet of doublets at δ4.38. However there appears to be further correlations of the signals at δ3.3 with another spectral feature at δ3.50 and between the NMR signals at δ3.10 and δ3.85 confounding a facile interpretation. This could possibly be explained by the presence of structural features of one of the other unknown metabolites associated with CV consumption with similar structure.

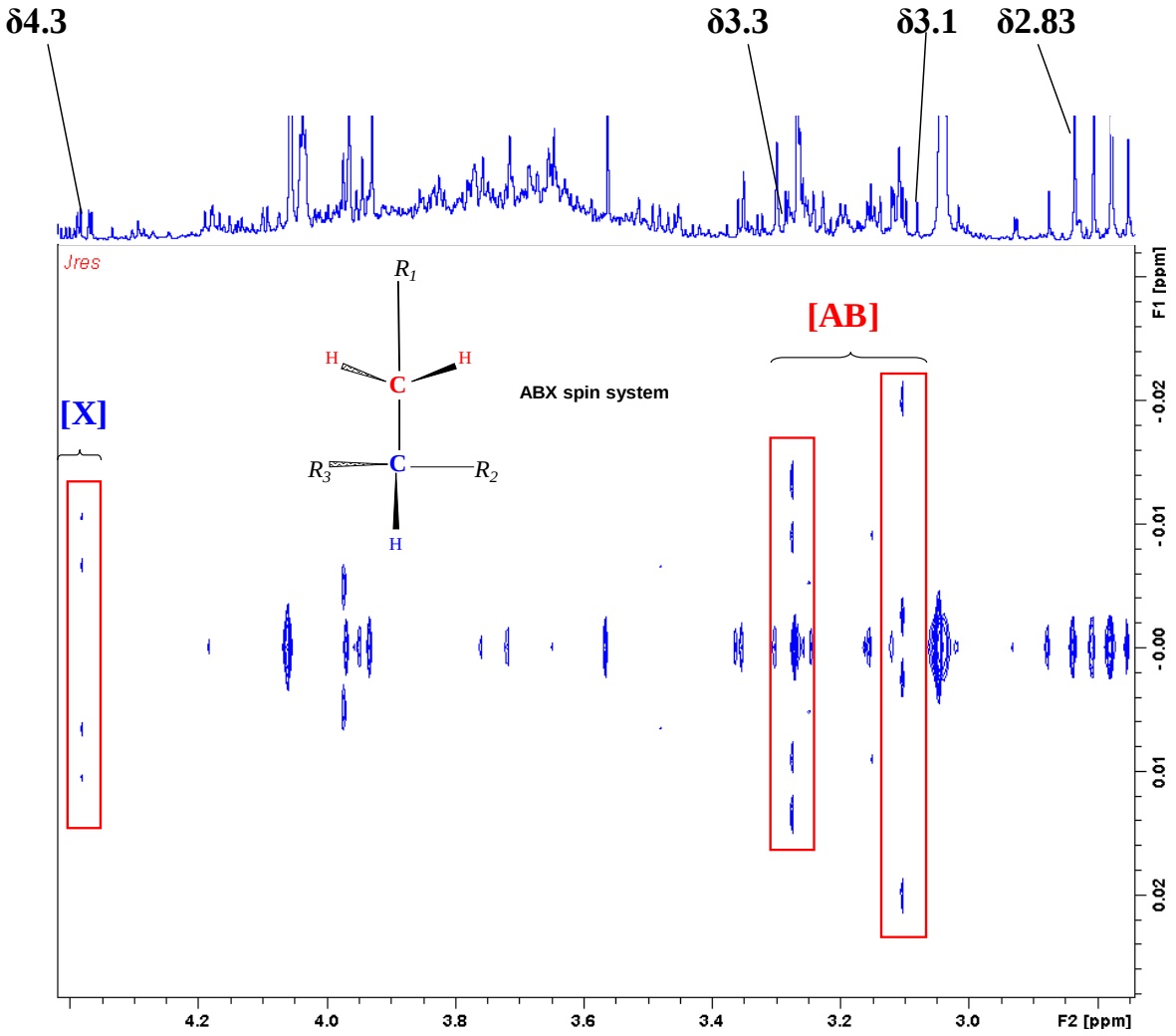


Fig. 6.13 400MHz 2D ^{1}H-^{1}H *J*-Res NMR experiment. Deconvolution of the NMR spectral features identified as characteristic of CV consumption between δ3-3.50 provides evidence of an ABX spin system indicative of an unknown chiral metabolite.

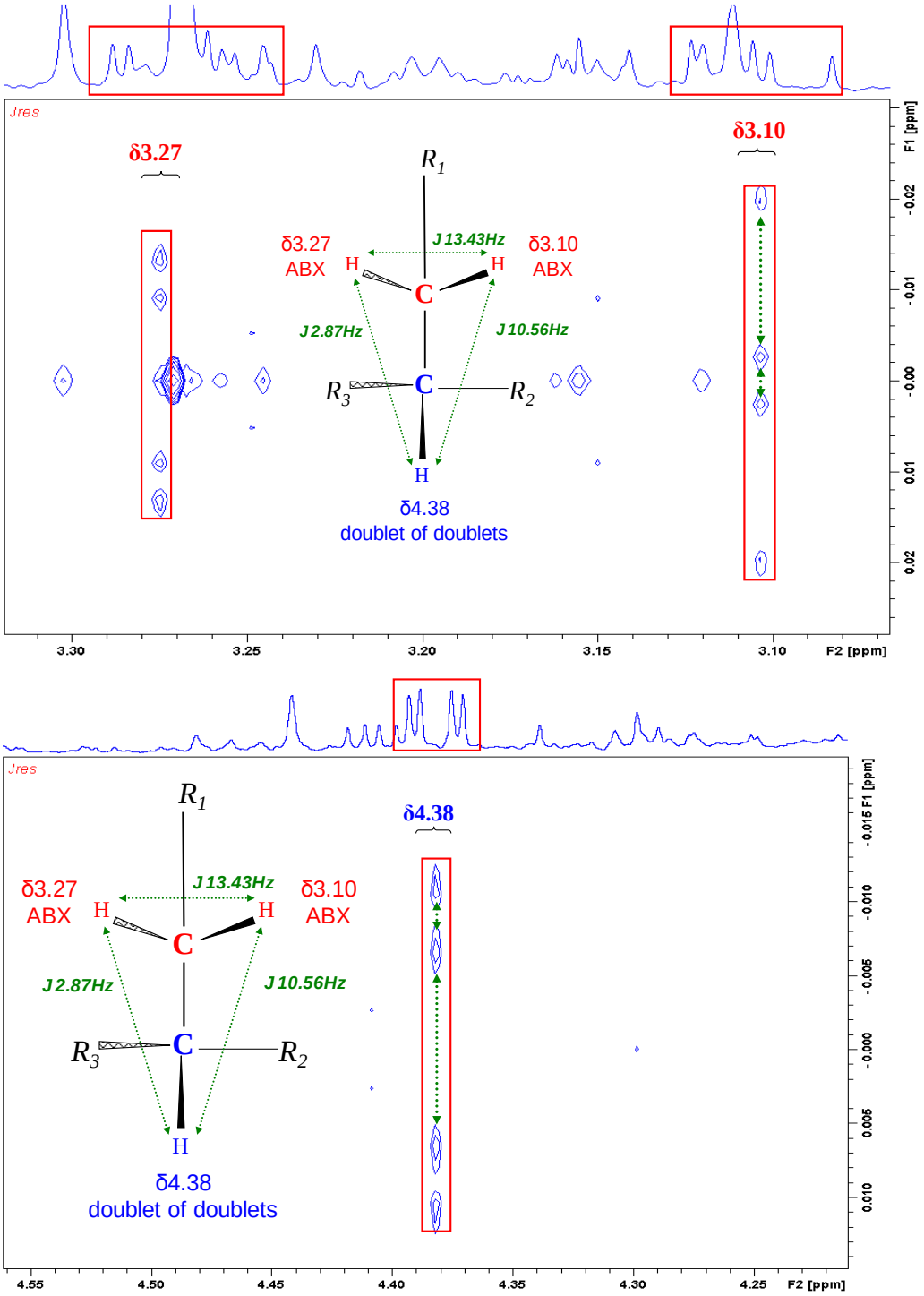


Fig. 6.14 Comparison of the *J*-couplings demonstrates shared scalar coupling constants between the protons of the ABX spin system of the unknown metabolite diagnostic of CV consumption.

This confirms the membership of the spectral features identified by O-PLS-DA to the same metabolite. 13.43Hz between the A and B of the spin system, 2.87Hz between the A and X and 10.56Hz between the B and X.

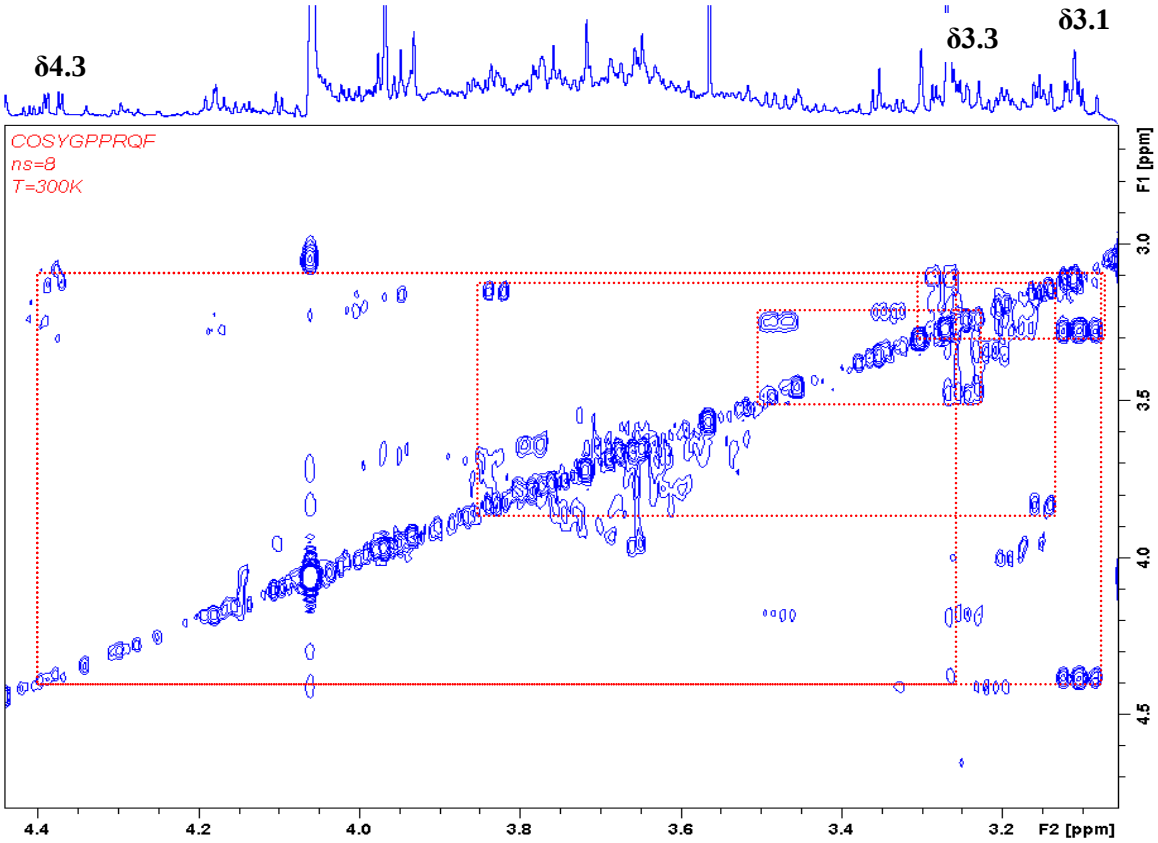


Fig. 6.15 2D COSY confirmation of the ABX spin system identified by *J*-res spectroscopy.

Evidence of correlations exists between the NMR spectral features at δ3.10 and δ3.30 and the doublet of doublets at δ4.38, however there appears to be further correlations of the signals at δ3.30 with another spectral feature at δ3.50 and between δ3.10 and δ3.85 confounding facile interpretation. However this could be possibly explained by the presence of structural features of one of the other unknown metabolites associated with CV consumption with similar structure.

6.5.2.5 ^{13}C HSQC and HMBC of solid phase extract of Phase II urine

Two dimensional heteronuclear NMR experiments 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. The ^{13}C Heteronuclear Single Quantum Coherence (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 analysis can provide further information on the chemical environment of the proton(s) in question and by comparison of the carbon shifts with ^{13}C NMR databases and comparison of ^{13}C chemical shifts of authentic standards. ^{13}C HSQC of the four singlet peaks provided ^{1}H-^{13}C shifts of δ2.75: δ39.6, δ2.78: δ40.2, δ2.81: δ40.1 and δ2.83: δ40.7 (**Table 6.4**). Furthermore the ^{13}C shifts of around ~δ 40 provided by the HSQC experiment are suggestive of a close proximity of the protons to either a sulphur or nitrogen.

Based on the HSQC ^{13}C shifts the 4 singlet peaks are suggestive of closely related or very similar compounds sharing a common origin or metabolic pathway. This is supported by STOCSY analysis. The variation of the HMBC values for the other four singlet peaks suggests the four singlets diagnostic of CV consumption represent four separate metabolites all with a similar structure and shared metabolic origin. As preliminary evidence suggests that the functional group responsible for the four singlet peaks are methyl protons from an S-methyl-sulphoxide functional group, this lends weight to the idea that one of the singlets is a vegetable derived metabolite and the other 3 singlet peaks are metabolites of this as yet unknown compound. This variation in ^{13}C around δ55.7 suggests that the *R*-group in proximity to the distant carbon are varying subtly (**Fig. 6.16**). The ^{13}C shifts of the signals thought to be responsible for the ABX spin system were of a lower intensity due to the complex splitting pattern and reduction in number of protons responsible for the signal and were therefore just above the threshold of noise. However a similarity of the HMBC ^{13}C shifts for the unknown singlets and the HSQC of the ABX signals exists. In particular the HMBC for δ3.1 of δ180.1 suggests proximity to a ^{13}C of a carboxylic acid group supporting evidence of an amino-acid like metabolite(s).

Table 6.4 Summary of ^{13}C shifts for both the HSQC and HMBC experiments.

Due to the low sensitivity of HSQC and HMBC the strongest heteronuclear signals were observed for the methyl groups of the unknown singlets. The ^{13}C shifts of the signals thought to be responsible for the ABX spin system were of a lower intensity due to the complex splitting pattern and reduction in number of protons responsible for the signal and were therefore found to be just above the threshold of noise.

^{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

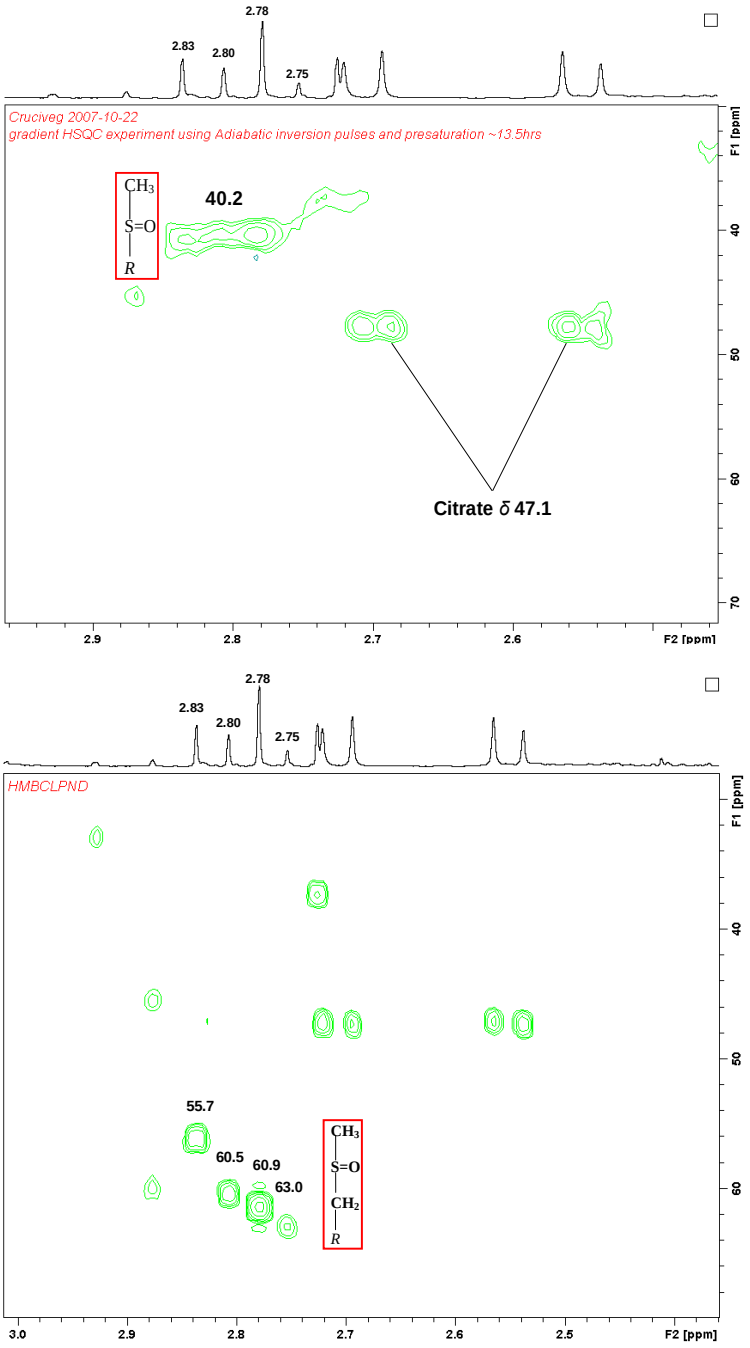


Fig. 6.16 Top: ^{1}H-^{13}C HSQC of 5% methanol SPE fraction of Phase II predose urine, the 4 singlet peaks share a similar ^{13}C shift of around δ39.6-40.2 suggestive of the chemically isolated S-methyl-sulphoxide group.

Bottom: ^{1}H-^{13}C HMBC of 5% methanol SPE fraction of Phase II predose urine, variations in ^{13}C shift can be seen for the four singlet peaks δ63.0, δ60.9, δ60.5, δ55.7; hypothesised to be a variation in *R* group adjacent to the long range carbon shift. Citrate is shown for reference purposes.

6.5.2.6 Cruciferous vegetable extraction: Analysis by 2D NMR experiment

Extractions of both broccoli and Brussels sprouts were performed to assess a potential plant origin of the ^{1}H-NMR biomarkers associated with CV consumption and also to attempt to identify hydrolysis products from mastication of plant tissue, effect of low pH on plant enzyme activity, the effect of cooking on hydrolysis product formation, comparison of different extraction strategies and to observe the formation of volatile compounds. Both broccoli and Brussels' sprouts samples were homogenised in either phosphate buffer or a 50:50 mixture of acetonitrile:D_2O and left to ferment at either physiological (37.5°C) or room temperature for 2 hours. Human saliva or HCl was also added to recreate the conditions in the human mouth and gut respectively.

Vegetable extracts were allowed to ferment to encourage the formation of enzymatic degradation products particularly of glucosinolates and SMCSO. Following fermentation of the vegetable extracts, phosphate buffered extractions were filtered and D_2O and TSP an internal standard were added prior to ^{1}H NMR analysis. Acetonitrile extractions were filtered and then lyophilised after snap-freezing prior to resolubilising in D_2O.

Vegetable extract ^{1}H-NMR spectra were significantly improved by the application of the CPMG pulse sequence which reduced the broadening of the baseline due to proteinaceous and other macromolecular contaminants and improved signal strength of the low molecular weight components.

There was an enormous wealth of information provided by the ^{1}H-NMR analysis of the vegetable extractions due to the abundance of carbohydrates, amino acids, organic acids and phenolic compounds found within the aqueous extract of the vegetable matrix (**Fig. 6.17, Table 6.5**). Identification of NMR spectral features was initially made using the plant metabolomics assignments available within the literature and online databases [147, 148, 274, 278]. The use of acetonitrile as an extraction solvent was found to be more stable than phosphate buffered extractions. This was possibly due to the infiltration of soluble enzymes and was in accordance with previous reports of vegetable extraction strategies [279]. No obvious changes could be seen in the raw spectra due to changing pH or addition

of saliva. However differences were seen between the different vegetables in the phenolic region and between fresh and microwaved vegetables with a reduction in unmetabolised SMCSO and glucobrassinin in fresh vegetables subjected to phosphate buffer extraction.

Due to the abundance of signals throughout the ^{1}H-NMR spectra, there was significant overlap between signals, particularly within the region associated with carbohydrate resonances. This initially hampered peak identification but was clarified by the application of several two-dimensional NMR experiments which greatly increased resolution and de-convoluted the ^{1}H NMR spectra. A combination of ^{1}H–^{1}H *J*-resolved, correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), heteronuclear multiple bond correlation (HMBC) and heteronuclear single quantum coherence spectroscopy (HSQC) were applied to accurately identify less well known or unknown resonances within the one dimensional spectra (**Table. 6.5**). For example 2D-TOCSY was used to confirm the presence of chlorogenic acid. Correlations could be seen between the two doublet peaks at δ6.93 and δ7.20 in this spectrum. Furthermore the shifts ^{13}C by HSQC analysis were found to match those found in the literature [148] (**Fig. 6.18)**

Table 6.5 Assignment of aqueous CV extraction metabolites confirmed by comparison of ^{13}C shifts with literature sources.

^{1}H-NMR Assignments	^{1}H-NMR δ (splitting pattern)	^{13}C δ
Isoleucine	δ 0.94 (t)	24.41
	δ 1.02 (d)	19.77
	δ 1.98 (m)	26.7
Valine	δ 1.00 (d)	17.4
	δ 1.05 (d)	20.7
Lactate	δ 1.34 (d)	22.4
Alanine	δ 1.48(d)	18.9
Arginine	δ 1.65 (m)	26.7
	δ 1.72 (m)	26.7
	δ 3.54 (t)	48.7
γ-aminobutyrate	δ 1.90 (m)	30.3
	δ 2.30 (t)	37.4
	δ 3.01 (dd)	41.8
Acetic acid	δ 1.93 (s)	26.3
Acetoacetate	δ 2.27 (s)	32.2
Citrate	δ 2.74 (d)	48.3
	δ 2.56 (d)	48.3
Succinate	δ 2.41 (s)	36.8
Sulforaphane	δ 2.98 (s)	39.5
	δ 4.7 (s)	
	δ 3.2(s)	
	δ 3.3(m)	
	δ 4.98(d)	
	δ5.23(t)	
Methyl-methanethiosulphonate	δ 2.76 (s)	39.31
	δ 3.51 (s)	53.69
S-methyl-L-cysteine sulfoxide	δ 2.83 (s)	40.8
	δ 3.32(m)	57.1
	δ 4.1(m)	78.2
Choline	δ 3.24 (s)	56.9
β-glucose	δ 4.58 (d)	98.8
Fructopyranose	δ 3.73 (d)	66.6
Sucrose	δ 5.40 (d)	95
	δ 4.17 (d)	79.8
Malic acid	δ 4.34 (dd)	73.4
	δ 2.74 (dd)	44.7
	δ 2.68 (dd)	44.7
α-glucose	δ 5.18 (d)	94.9
Fumaric acid	δ 6.56 (s)	138.2
trans-sinapoyl malate	δ 7.01 (s)	119.9
chlorogenic acid	δ6.97(d)	118.6
	δ7.22(d)	122.4
glucobrassinin	δ7.33(d)	132.2
	δ7.38(t)	130.8
	δ7.43(t)	132.2
phenylpropanoid	δ 7.74 (d)	121.6
kaempferol	δ 7.87 (t)	144.7
adenine	δ 8.2 (s)	144.1
	δ 8.21 (s)	145.2

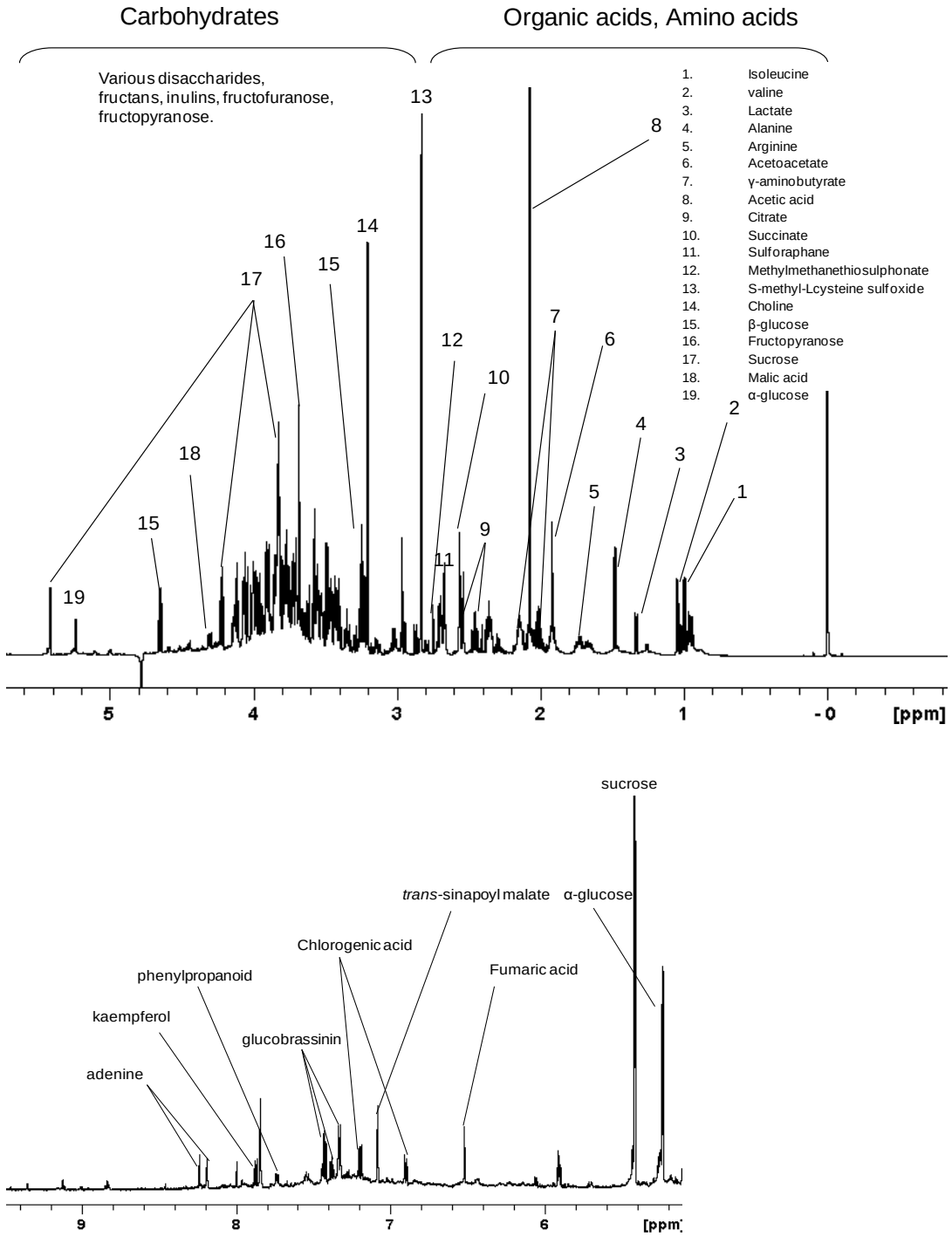


Fig. 6.17 Top: Aliphatic region of CPMG fresh Brocolli acetonitrile extraction. Assignments were made using a combination of 2D-NMR experiments and comparison with literature sources and authentic standards. A significant peak of SMCSO can be seen at δ2.83. Bottom: Phenolic region common plant phenolics including unmetabolised glucobrassinin. Unknown phenolics possibly include anthocyanins, quercitin compounds, and unknown phenylpropanoids.

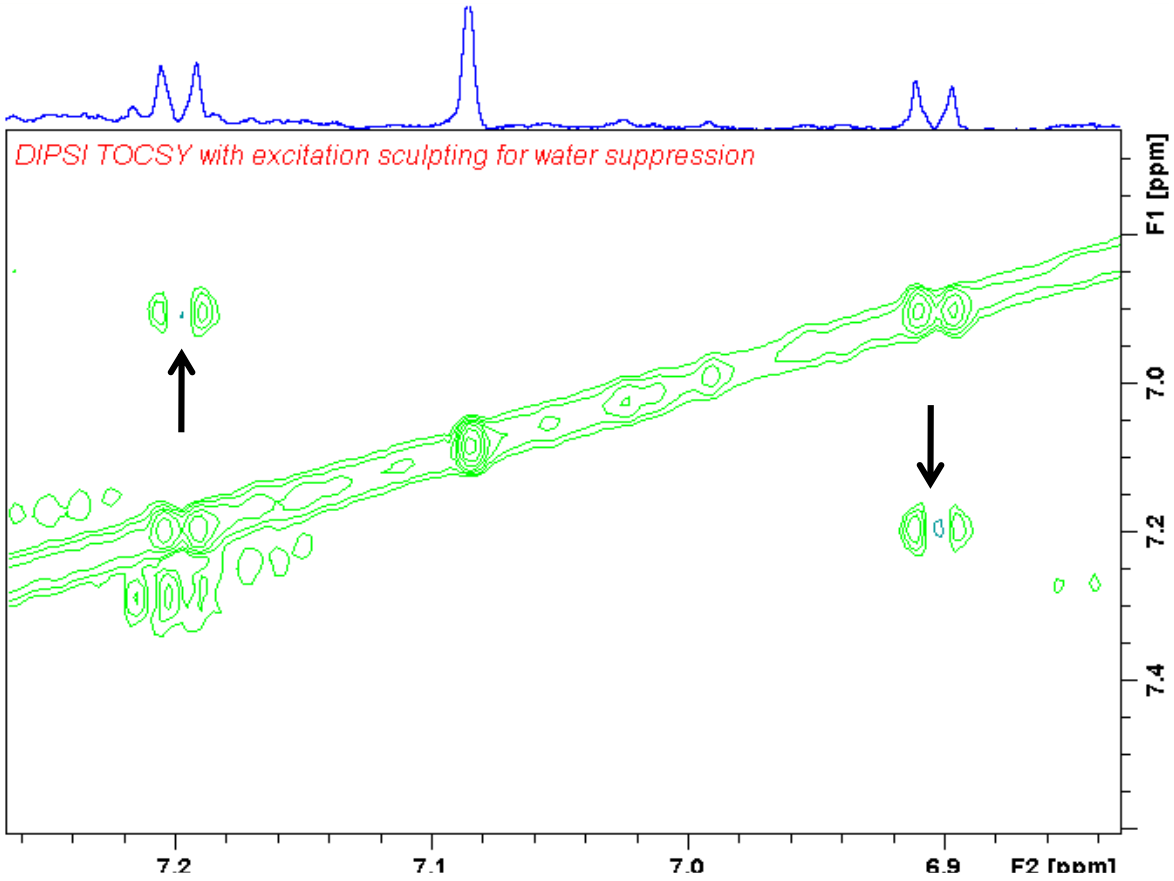


Fig. 6.18 2D-TOCSY spectrum of fresh broccoli aqueous acetonitrile extraction confirming the presence of chlorogenic acid.

Correlation can be seen between the two doublet peaks at δ6.93 and δ7.2 (arrows). Identifications were augmented with ^{13}C confirmation.

^{1}H–^{1}H *J*-resolved spectra proved a useful experiment for deconvolution of the crowded areas of the spectrum, reducing complex multiplets to single resonances and separating chemical shifts from their respective coupling constants in a two-dimensional plot [280].

Complimentary to ^{1}H–^{1}H *J*-resolved spectra, the COSY experiment was utilised to understand the relationship between mutual spin-spin couplings and confirm the membership of spin systems to a particular molecule. The COSY experiment was further augmented by the TOCSY pulse sequence which showed spin systems of coupled protons within the same molecule over a greater distance.

The combination of HMBC and HSQC heteronuclear spectroscopy proved very useful for both identification of unknowns within the phenolic region and more importantly for the identification of SMCSO within the plant extracts (**table 6.5)**. The typical singlet resonance of the methyl group of SMCSO at δ2.83 was clearly observed in the ^{1}H-NMR spectra of both the broccoli and Brussels' sprouts vegetable extracts. This corresponded to one of the four similarly shifted unknown singlet resonances observed in the human urine following CV consumption.

Comparison of HSQC spectra of both the vegetable extractions and human urine demonstrated the similarity in ^{13}C of δ40.2 of the singlet at δ2.83 in the vegetable extract, not only to the unknown singlet in urine but also to the other three unknown singlets in the human urine (**Fig. 6.19**). This finding was enhanced using HMBC experimentation which gave long range correlations between the ^{13}C identified by HSQC with another ^{13}C within the same molecule at δ55.7. Comparison of the HMBC experiments of the vegetable extracts and human urine revealed the same ^{13}C chemical shift of δ55.7 for the singlet peak δ2.83 in human urine (**Fig. 6.19**).

However the other four unknown singlet resonances observed in the high CV consumption urine showed slight chemical shift variation in this long range ^{13}C resonance δ2.81 (δ60.5), δ2.78 (δ60.9), δ2.78 (δ63.0). These results suggested that the ^{1}H NMR singlet peak δ2.83 which statistically discriminates the high CV consumption study phase was of exogenous CV derived origin and that structural similarity in the

functional group for the ^{13}C resonance δ40.2 exists between the four unknown singlet resonances. This is most likely due to a strong similarity in the functional group responsible for all of the four unknown singlet resonances. It could also be due to the presence and varying abundance of diastereoisomers of distinct metabolites. However it is known that the *L*-isomer of SMCSO predominates in nature and not the racemate [275]. The difference in ^{13}C shift in the HMBC experiment suggests that the neighbouring carbons of each singlet peak is being altered by different *R* groups adjacent to the carbon identified by this long range correlation.

The ^{13}C-^{1}H heteronuclear experiments provided clear evidence that the four singlet resonances are most probably four distinct molecules which are closely related in structure. The singlet resonance at δ2.83 is present in both the aqueous vegetable extraction and high CV consumption human urine and likely represents unmetabolised excretion of a vegetable derived compound. The differences seen in the HMBC ^{13}C resonances of the four singlet peaks associated with CV consumption could therefore potentially be mammalian derived metabolites of the CV derived compound and the relatively small differences in both the ^{1}H and ^{13}C shifts is due to as yet unidentified *R* groups on each of the other four biomarkers observed in urine. Assuming that δ2.83 does indeed represent the unmetabolised SMCSO, the four singlet resonances are most likely to represent methyl protons adjacent to the sulphoxide group of SMCSO or a closely related molecule. Variation in chemical shift of the remaining unknown singlets is probably due to different *R* groups arising from common routes of mammalian elimination of methylated sulphur containing amino acids. Principally SMCSO is cleaved and enters the urea pool (~50%), but as is common with many amino acids, it is also *N*-acetylated (~30%), and converted to thiopyruvates and thioacetates by the process of deamination (~8%) [276, 277].

The HMBC experiment confirmed that the unknown singlet at δ2.83 δ40.7 was the same compound seen in the aqueous extract of broccoli as the multiple bond coherence value was of the same value (δ55.7) in both urine and broccoli. This suggested that the metabolite in urine was indeed of plant origin.

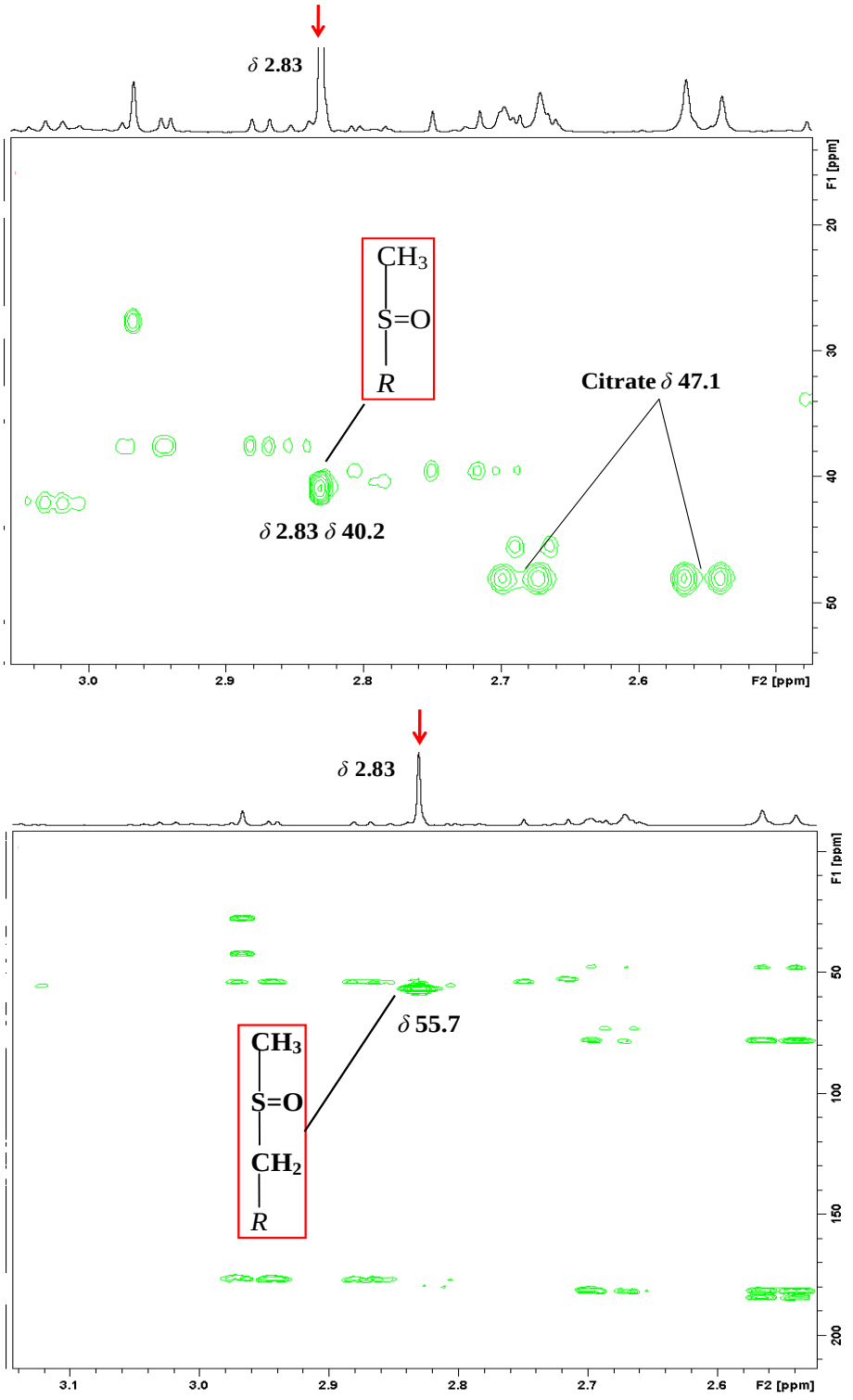


Fig. 6.19 Top: ^{1}H-^{13}C HSQC of fresh broccoli acetonitrile extraction; Bottom: ^{1}H-^{13}C HMBC of fresh broccoli acetonitrile extraction.

^{1}H and ^{13}C shifts match those seen in human urine diagnostic of CV consumption. This suggests a CV origin of one of the unknown NMR singlets at δ2.83. Citrate is shown for reference purposes

6.5.2.7 Synthesis of SMCSO and mammalian metabolites

SMCSO was synthesised by selective oxidation of the sulphur of SMCS authentic standard (according to the protocol of Synge RLM *et. al.* 1954). In order to fully characterise this synthetic standard the same 2D NMR experiments were performed on the synthetic standard to confirm the identification made in the urine samples. Initially a 1D standard experiment was performed to monitor the synthesis and the relative positions of the singlet peak and ABX spin systems within the spectra. This result was augmented using the homonuclear *J*-resolved spectrum to separate chemical shift from scalar couplings, an HSQC experiment to confirm the carbon 13 chemical shifts of the carbons from the methyl group singlet, and the two carbons of the ABX spin system. This confirmation was further augmented by performing an HMBC experiment to identify the long range proton-^{13}C couplings and to match these to the NMR experiments performed in urine and vegetable extract experiments (**Fig. 6.20**).

The *N*-acetylation of SMCSO with acetyl chloride and ice cold NaOH resulted in a number of contaminants which complicated the spectra (**Fig. 6.21**). The reaction of acetyl chloride with water results in a highly exothermic reaction producing HCl, acetic acid and sodium acetate. There was also the formation of NaCl and evidence of the presence of some of the solvents used within the reaction such as ethanol, chloroform and hexane in spite of lyophilising the synthesis by freeze drying prior to ^{1}H NMR acquisition. The salt content of the synthesis proved challenging for the acquisition of good quality spectra as a high concentration of monovalent ions can affect probe tuning and sensitivity, an increased pulse length is required leading to a smaller bandwidth of excitation and can lead to a reduction in the homogeneity of the applied pulse with molecules in the centre of the sample experiencing a weaker RF-field.

An attempt to lyophilise again under high temperature, low pressure conditions (60ºC) in a speedvac to remove solvent contaminants, followed by reconstitution in 50:50 H_2O:acetonitrile and an SPE using an acetonitrile gradient, succeeded in removing solvent but failed to desalt. Following the SPE, the concentration of *N*-acetyl-SMCSO in all fractions was unfortunately reduced below NMR detection limits required for two dimensional NMR experimentation.

The presence of contaminants was surmounted by performing the two dimensional NMR experiments, *J*-res, COSY, HSQC and HMBC. The presence of the singlet peak at δ2.78 was confirmed with the correct ^{13}C shift of δ 40.2 and the ^{13}C shift of the ABX spin system at δ3.1 of δ61.7 (**Fig. 6.22**). The HMBC experiment also confirmed the ^{13}C shift of the NAc-SMCSO methyl group of δ61.1. In the absence of an authentic NAc-SMCSO standard, the two dimensional NMR evidence presents confirmation of identity but is by no means conclusive. According to [277], the majority of SMCSO is cleaved to form urea and inorganic sulphate (~50%) the majority of which occurs in the first 24hrs (~59%) with a complete recovery within 14 days (~96%). Since the human volunteers in the CVs study consumed 500 g of CV per day, their intake of SMCSO was estimated to in the order of 218 mg per day [4]. A change in excretion of urea in the human volunteer study would be difficult to detect assuming the subjects were consuming only ~218mg of SMCSO a day.

The next major route of metabolism is by *N*-acetylation (~30%), followed by unmetabolised excretion (~12%) and by deamination (8%). The height and significance of the peak for NAc-SMCSO at δ2.78, when compared to the peak at δ2.83 for unmetabolised SMCSO, is compatible with *N*-acetylation being the primary route of excretion of SMCSO.

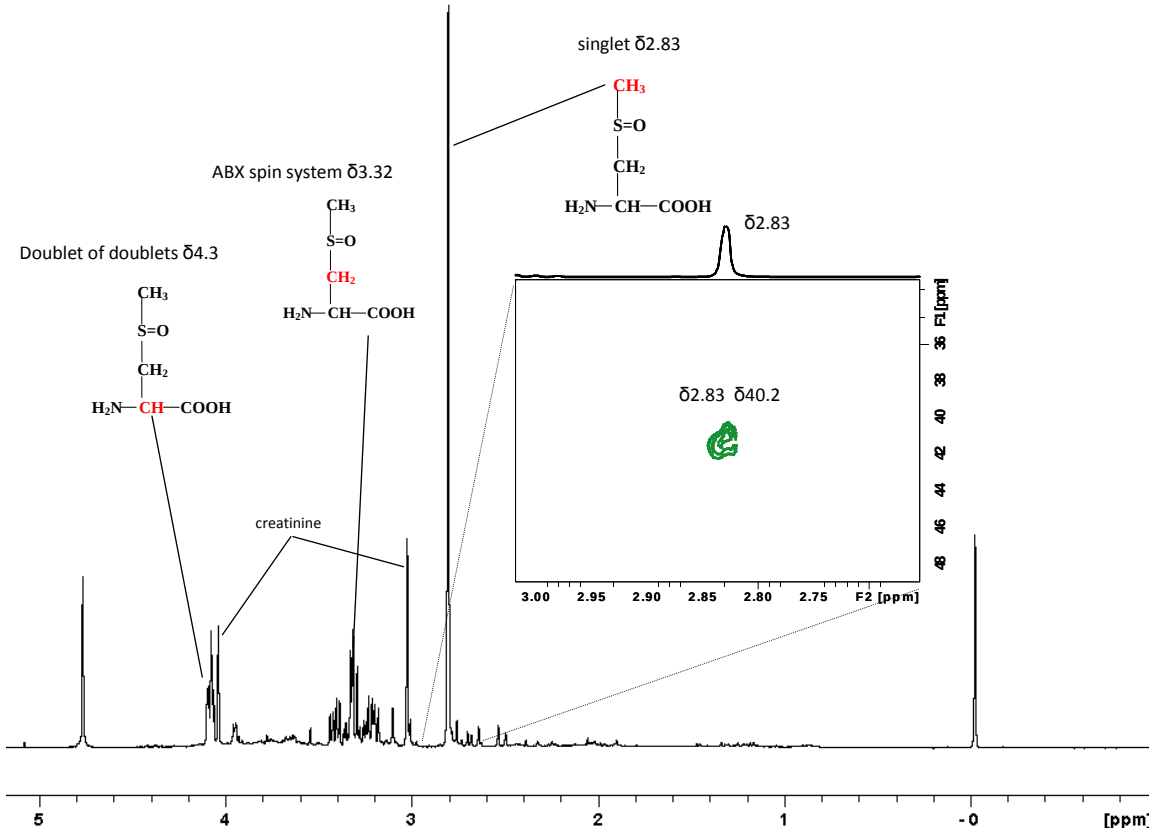


Fig. 6.20 400MHz 1D-NOESY of SMCSO synthetic standard spiked into Phase I predose urine.

AB protons can be seen at δ3.32 and the corresponding X proton at δ4.3.

Inset: 400MHz ^{1}H-^{13}C HSQC of SMCSO synthetic standard the methyl proton singlet at δ2.83 ^{13}C δ40.2 matches the chemical shifts observed in both CV aqueous extracts and within human urine as a biomarker of CV consumption.

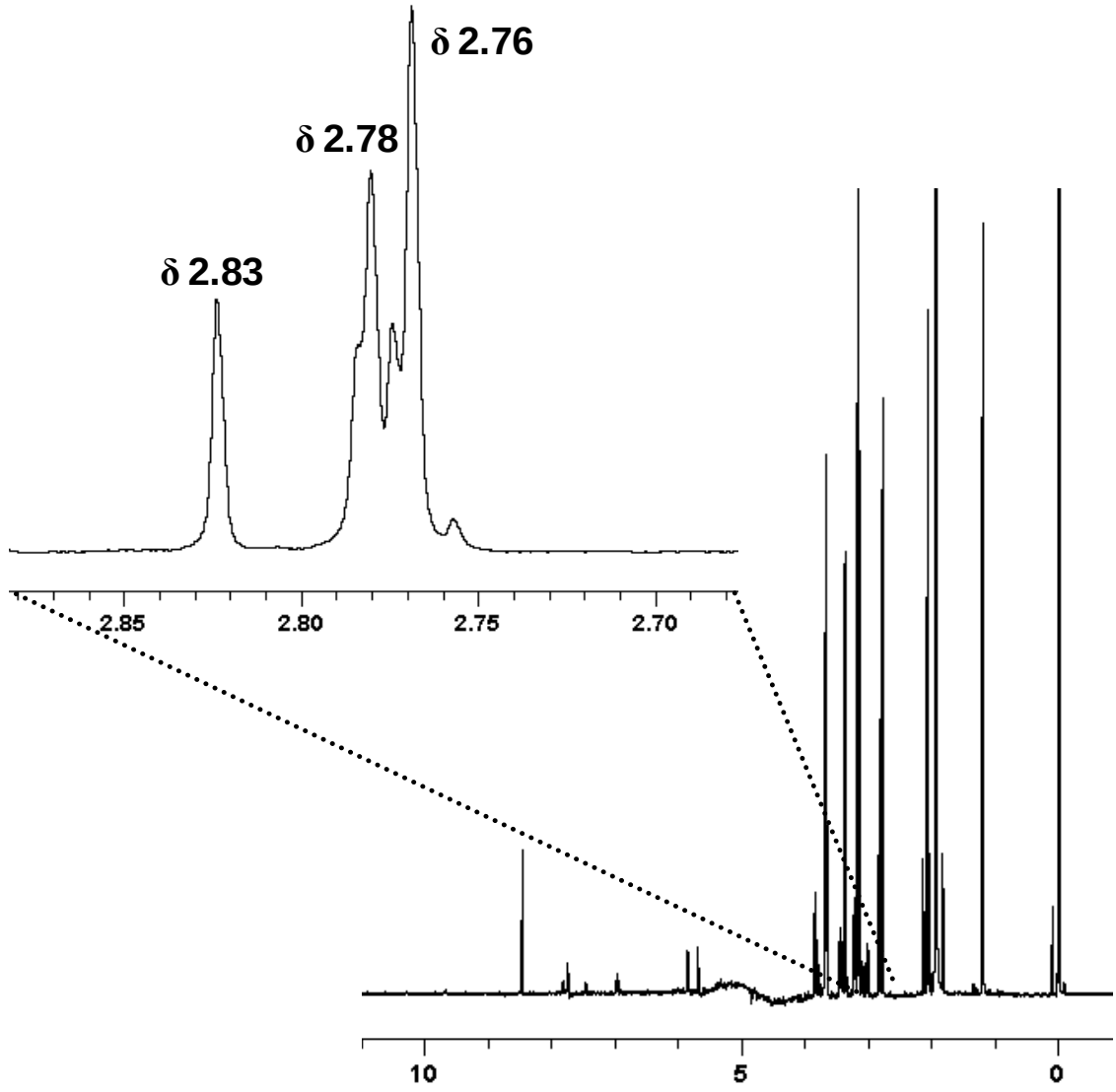


Fig. 6.21 400MHz 1D-NOESY of NAc-SMCSO synthetic standard.

Levels of salts affected acquisition and other solvent contaminants confounded interpretation of the NMR spectrum. Unreacted SMCSO can be seen as can an overlap of the singlet at δ2.78.

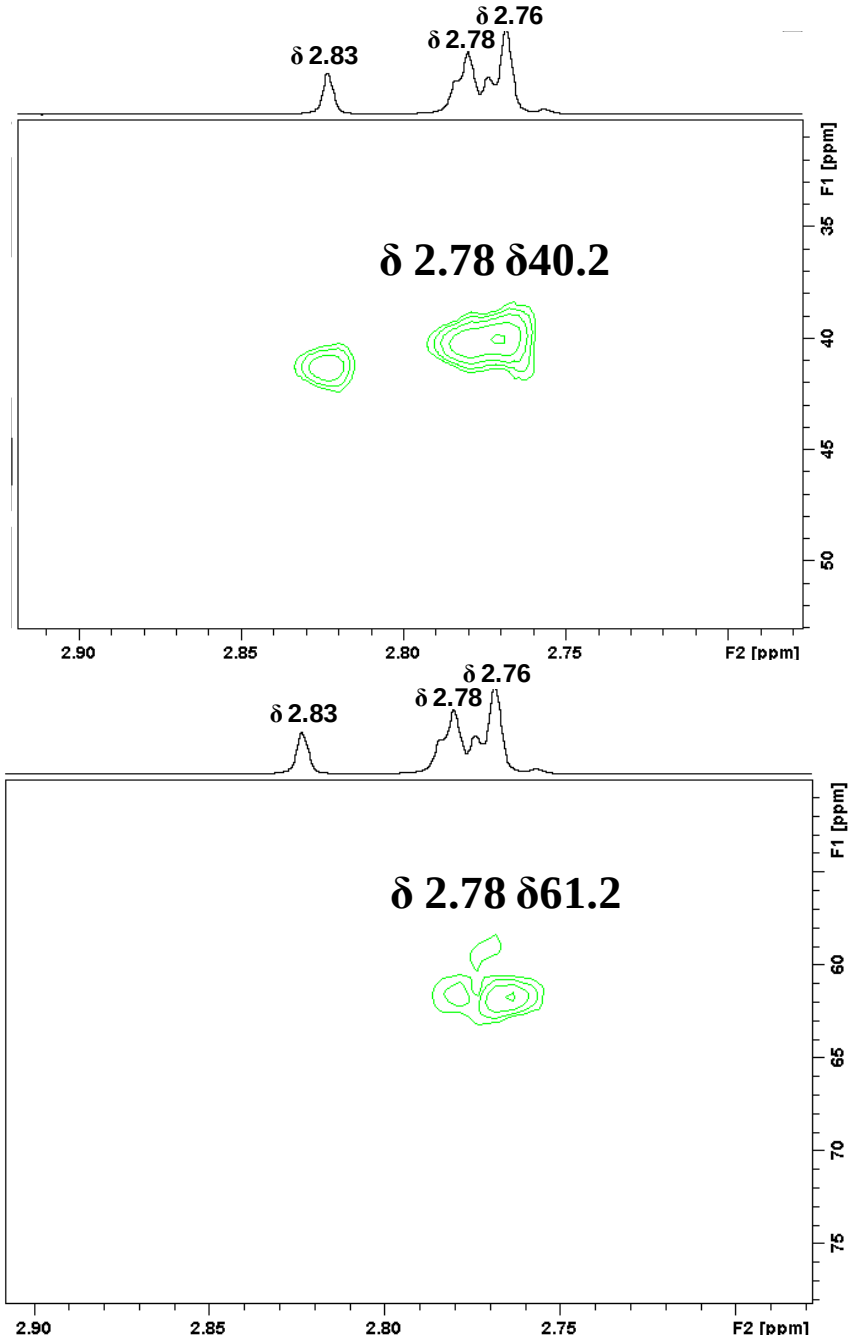


Fig. 6.22 Top right: 400MHz ^{1}H-^{13}C HSQC of NAc-SMCSO synthetic standard; the methyl proton singlet at δ2.78 ^{13}C δ40.2 matches the chemical shifts observed within the CV biomarker in human urine.

Bottom right: 400MHz ^{1}H-^{13}C HMBC of NAc-SMCSO synthetic standard. The methyl proton singlet at δ2.78 ^{13}C δ61.2 matches the chemical shifts of the long range ^{13}C of the unknown singlet δ2.78 in human urine.

6.5.3 Effect of meat consumption on human urinary metabolic profile.

6.5.3.1 PCA and PLS-DA analysis

To attempt to model a metabolic trajectory following meat meal and caffeine consumption PCA modelling was performed on unit variance scaled data of all study phases. This included all time-points pre-dose, 0-10 h.p.d, 10-24 h.p.d and 24 h.p.d. urine samples. Analysis of the scores plot revealed a clustering of the 0-10 h.p.d. dose group and some clustering could be seen based upon other timepoints (**Fig. 6.23)**. Loadings attributable to dietary time-point were widely scattered and therefore it was difficult to extract urinary metabolite changes responsible for the grouping seen in the scores plot along the second principal component.

Enhancement of separation was attempted using PLS-DA modelling which achieved improved separation between all dietary time-points in the first two PLS-DA components (**Fig 6.24)**. Validation of the PLS-DA model showed that using two PLS-DA components it was possible to predict red meat meal consumption with 41% of these predictions being significant at the 99% confidence interval. Analysis of the PLS-DA loadings plot revealed variables corresponding to the urinary metabolites, carnitine, taurine, glycine, alanine, and histidine responsible for discrimination between the 0-10 h.p.d and the other time points and hippurate and TMAO discriminatory of the 10-24 h.p.d. timepoint. This was suggestive of a metabolic trajectory of the meat meal over the 48 hour time course with evidence of bacterial co-metabolites in the 10-24 h.p.d.

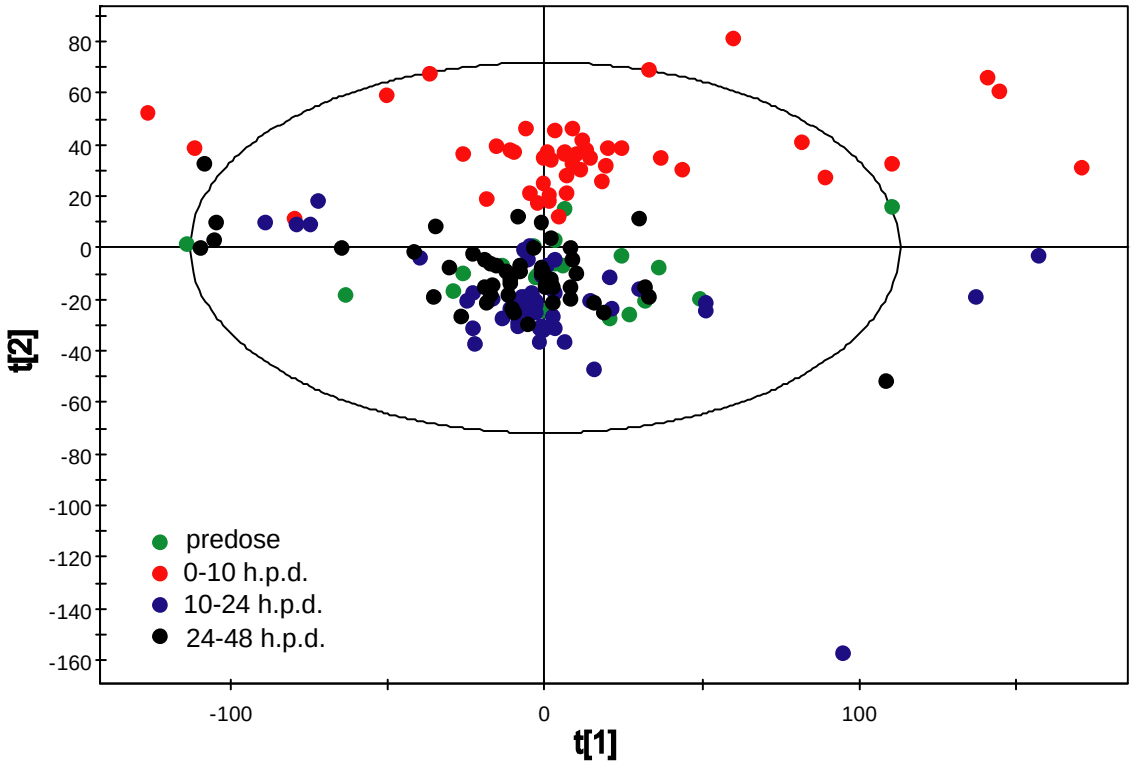


Fig. 6.23 Scores maps showing *t*[1]/*t*[2] PCA scores (Q^2: 0.37 R^2: 0.46) from the ^{1}H-NMR spectra (Unit Variance scaled) of all study phases showing clustering of samples based upon time-point.

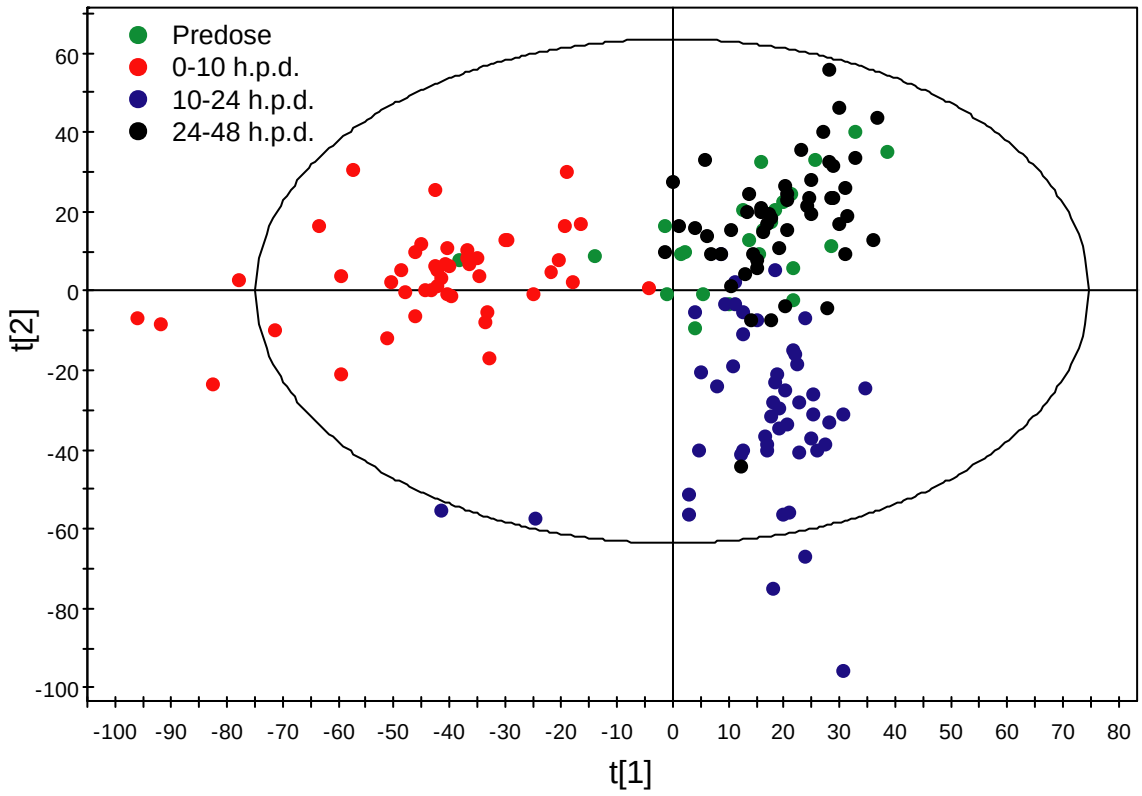


Fig. 6.24 Scores maps showing *t*[1]/*t*[2] PLS-DA scores (Q^2: 0.41 R^2: 0.49) from the ^{1}H-NMR spectra (Unit Variance scaled) of all study phases showing clustering of samples based upon time-point.

6.5.3.2 Time course pair wise O-PLS-DA

In order to further improve the observed separation in the PCA and PLS-DA models of all phases, O-PLS-DA was applied to produce a regression model to distinguish sequential timepoints within each individual study phase. This described the maximum separation between each timepoint. The loadings revealed marked time-course differences between pre-dose urine samples and all time points following meat consumption and caffeine intake. Many metabolites could be very clearly established as being postprandial markers of meat consumption (**Fig. 6.25**). Pre-dose urine samples were characterised by higher urinary concentrations of hippurate, DMA and *N*-acetylcysteine. Major differences following meat consumption in the 0-10 hours urine collection were increased excretion of creatine, creatinine, carnitine, taurine, *N*-acetyl-5-hydroxytryptamine, 1- 3- methylhistidine, acetylcarnitine, carnosine and several other amino acids; glycine, glutamine, alanine, threonine and histidine. Ten to twenty four hours following meat consumption and caffeine intake, urinary excretion of bacterial co-metabolites TMAO, *p*-cresol sulphate, phenylacetylglutamine and to a lesser extent hippurate increased (**Fig. 6.26**). However the excretory levels of the metabolites identified in the 0-10 hours urine collection are all significantly decreased in the O-PLS-DA model. After twenty four to forty eight hours post meat consumption and caffeine intake, urinary excretion of phenylacetylglutamine and TMAO decreased.

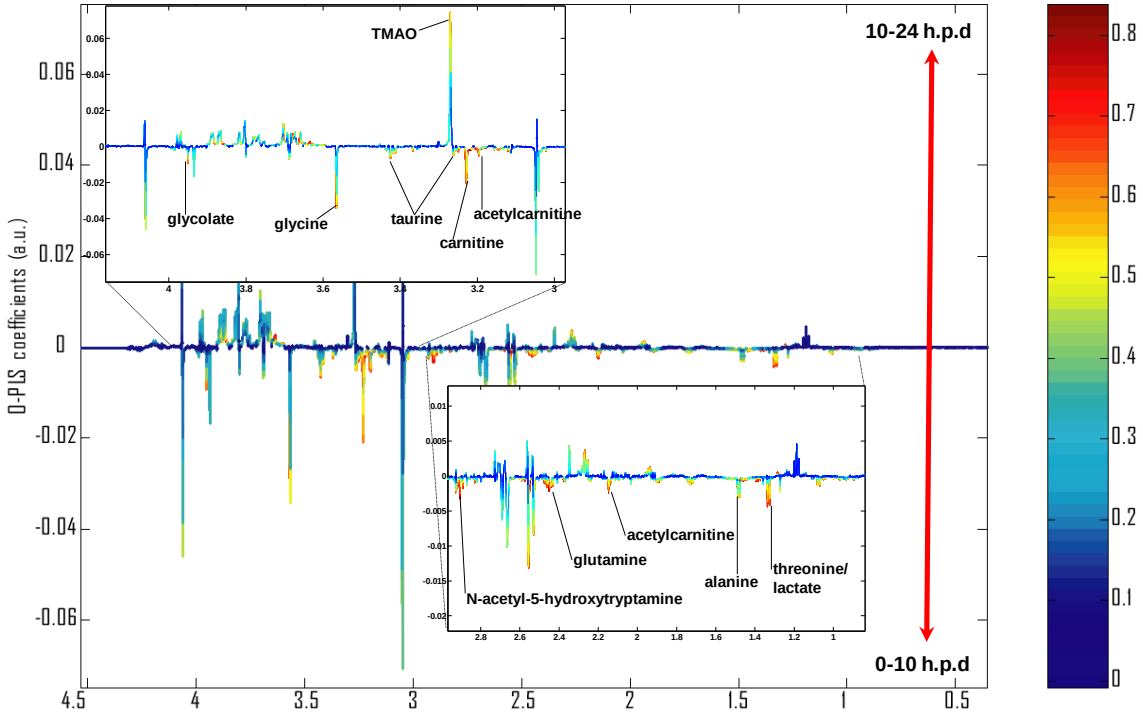


Fig. 6.25 Effect of red meat consumption on urinary metabolite profile.

Time course pair-wise O-PLS-DA Loadings plot in the aliphatic region of the spectrum showing clear statistical significance of amino acid metabolites characteristic of the 0-10 h.p.d time point following red meat consumption. The model shows excellent predictive strength and fit with a $Q^2\hat{Y}$: 0.8 and $R^2\hat{Y}$: 0.95.

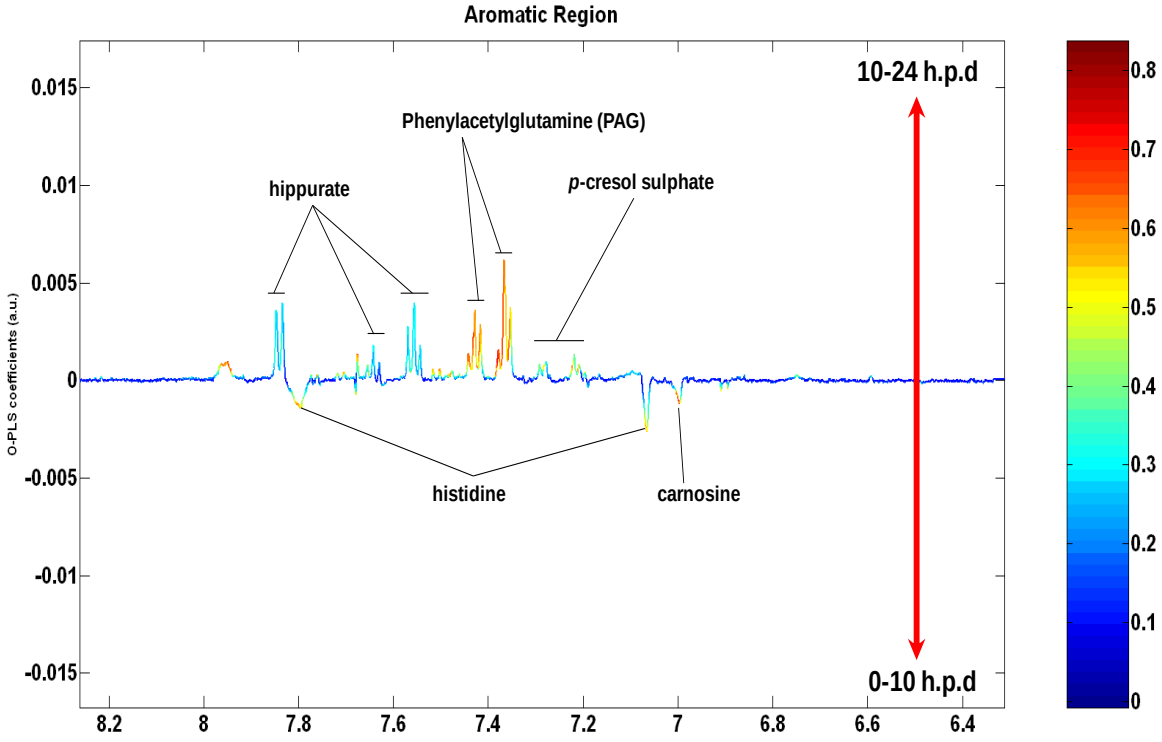


Fig. 6.26 Effect of meat consumption on urinary metabolite profile.

Time course pair-wise O-PLS-DA Loadings plot in the aromatic region of the spectrum showing significance of bacterially derived co-metabolites in the 10-24 h.p.d time point following red meat consumption.

STOCSY analysis of the triplet of phenylacetylglutamine at δ7.36 revealed a correlation with not only itself but also with *p*-cresol sulphate, and to a lesser extent *p*-hydroxyphenylacetate and hippurate. This statistical correlation supports the gut microbial origin of these aromatic compounds based on the strong statistical correlations observed. (**Fig. 6.27**).

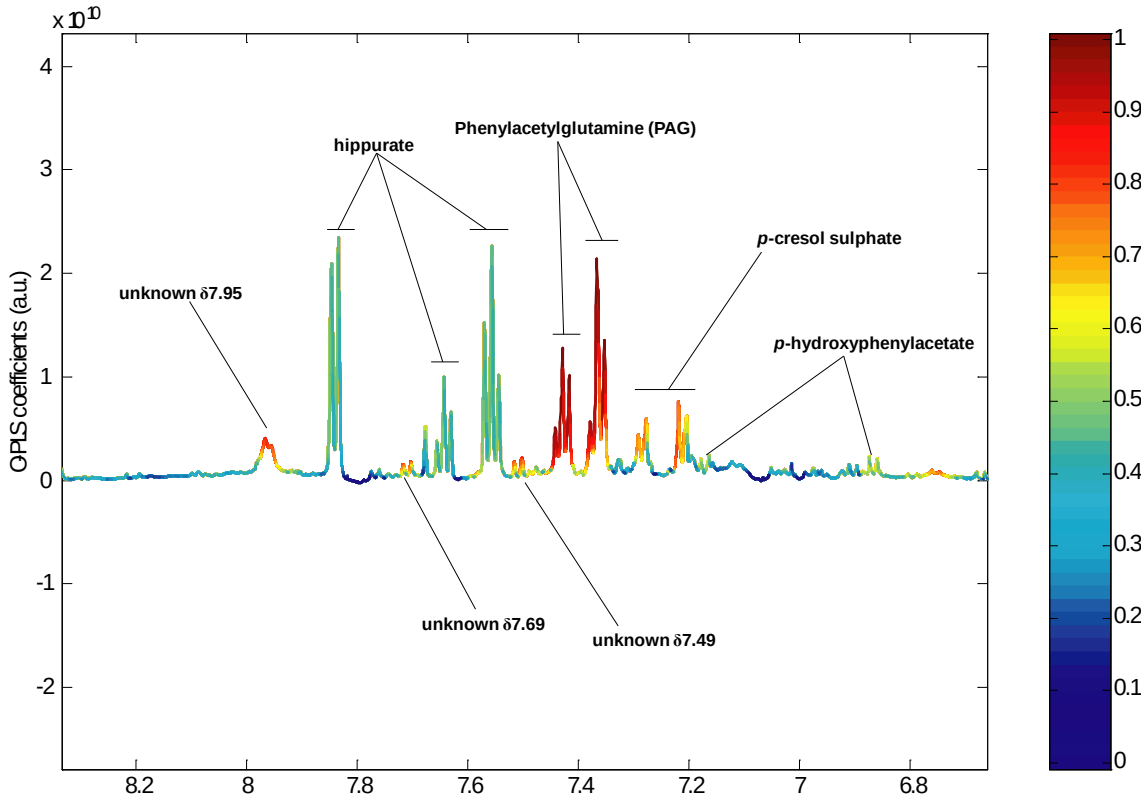


Fig. 6.27 STOCSY analysis of the triplet peak of phenylacetylglutamine δ7.36 demonstrating statistical relationships of bacterial co-metabolites in the phenolic region of the spectrum identified in the time-course O-PLS-DA.

6.5.3.3 Effect of cruciferous vegetable dietary intervention on meat and caffeine alterations of metabotype.

To examine the effect of CV consumption on meat and caffeine alterations of metabotype, pair-wise O-PLS-DA of each time point in each phase was performed. The dominant metabolite influencing separation of phase II with both phase I and III at all time points was consistently the characteristic singlet resonances associated with CV consumption (previously described). However following meat and caffeine consumption, increased urinary excretion of mannitol was shown to discriminate phase I 0-10 h.p.d from both phase II 0-10 h.p.d. and phase III 0-10 h.p.d.. This was also seen to a lesser extent to discriminate later time points in phase I from both phase II and III. Mannitol was confirmed by STOCSY analysis and from NMR databases of standard compounds. The splitting pattern, as observed in the O-PLS-DA coefficient plot and the raw spectra, was clearly attributable to mannitol (**Fig. 6.28**). Time course overlay NMR spectra identified the presence of mannitol in all phases and time points. The appearance of mannitol in the spectra was not anticipated, but has been previously reported as attributable to evidence of non-compliance due to consumption of chewing gum containing this sweetener [96].

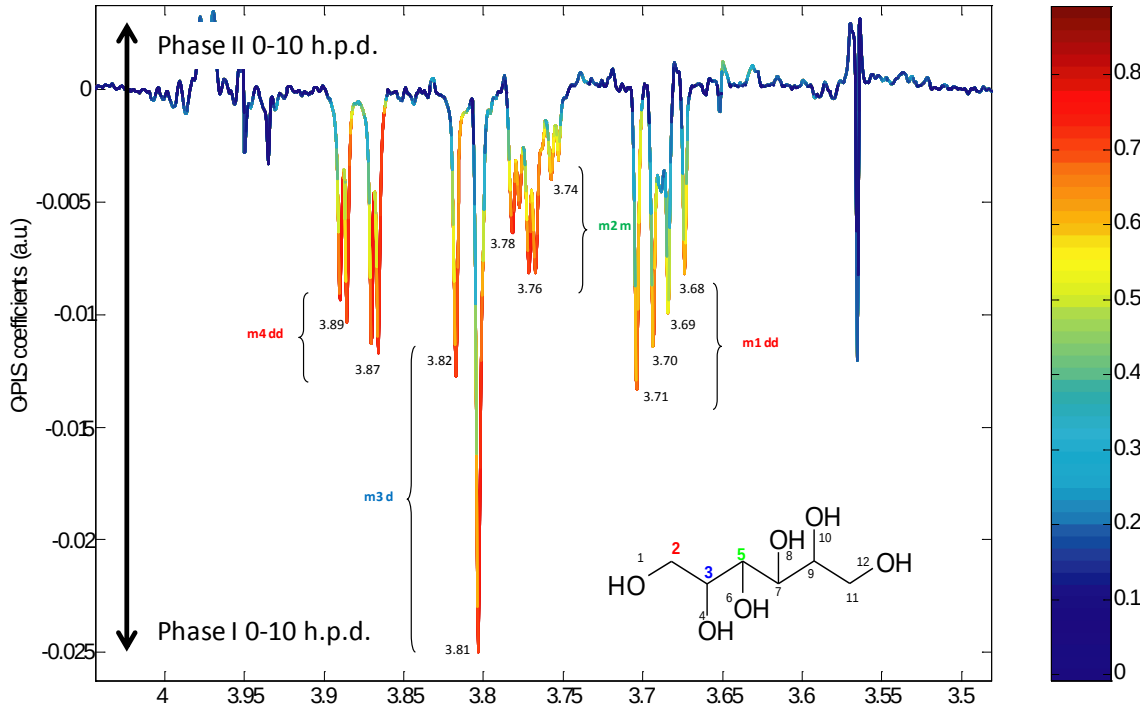


Fig. 6.28 Expanded O-PLS-DA showing the characteristic splitting pattern of mannitol characteristic of the Phase I 0-10hrs time point.

The presence of this metabolite was not anticipated and its significance in characterising the post meat meal/caffeine time point between the study phases is difficult to account for.

Red: atom 2, blue: atom 3, green:atom 5. Two doublets attributable to atom 2 can be seen, a multiplet attributable to atom 5 and a doublet attributable to atom 3. Protons on atoms 7, 9 and 11 are symmetrically magnetically equivalent to 2, 3 and 5.

6.5.4 UPLC-MS biomarker identification

6.5.4.1 MarkerLynx PCA analysis

Evaluation of the UPLC-Mass Spectrometric data set was initially undertaken with a chemometric approach using the MarkerLynx software (Waters MS Technologies, Manchester, UK). MarkerLynx generates a PCA plot of the MS data where the intensities of peaks within a spectrum are reduced from multidimensional space to a two dimensional plot in order to visualise any clustering of points which may help to explain the variation within the data and to hopefully identify differences between the sample classes (**Fig. 6.29**). The data set was also reduced to a more manageable size, specifically to the 0-6 min fraction of the UPLC gradient, which usually contains more pertinent metabolic information than later fractions. The PCA analyses of the positive and negative ion mode data sets utilising in turn, pareto, mean centred and unit variance scaling unfortunately failed to identify any unsupervised chemometric clustering of the data and inter-sample variability proved to be more significant than any separations based on class.

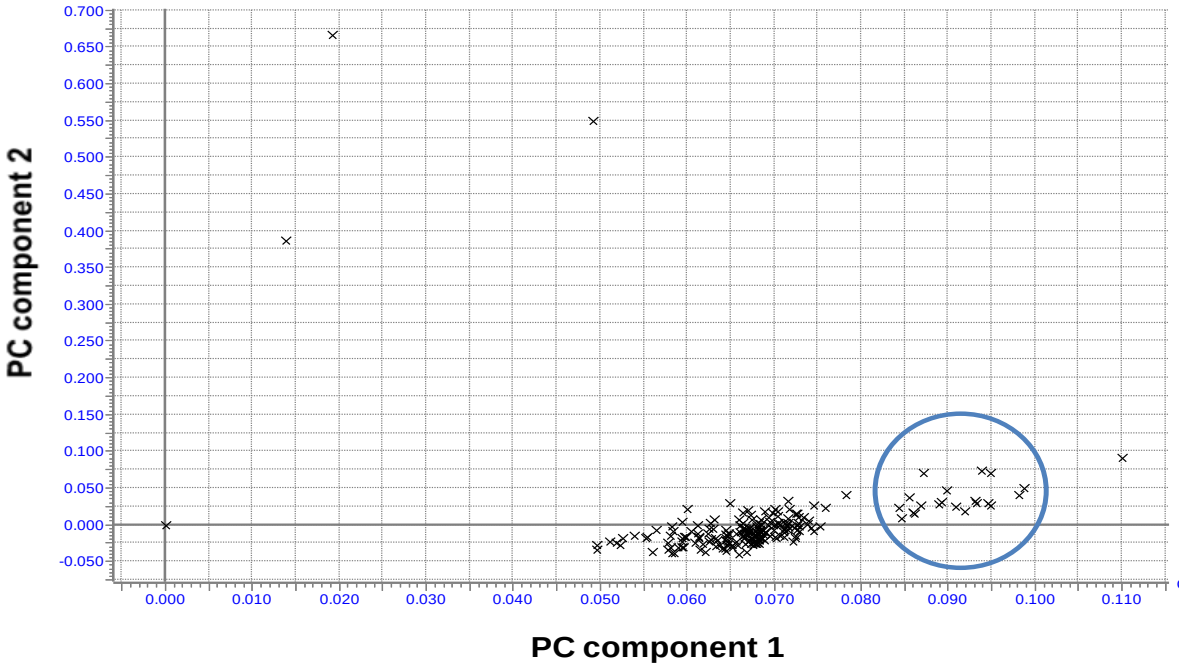


Fig. 6.29 Markerlynx PCA scores plot of positive mode UPLC-MS data. No underlying clustering of subjects within the data can be seen when observing variance in the X matrix of the 0-6mins retention time subset. There is evidence of a cluster of outlying samples (ringed) which represent a range of different subjects and sample classes.

6.5.4.2 Masslynx targeted analysis

Further to the statistical analysis a targeted approach was also conducted which has been used previously in toxicology/ drug metabolism studies in rats [281-283]. A list of potential metabolites of interest was generated (**Table. 6.6.**). CV candidate molecules were selected based on the glucosinolates, isothiocyanates and their associated metabolites (inc. the mercapturates) from the literature. Post prandial biomarkers were selected based on the NMR findings and included meat meal metabolites (amino acids, bacterial co-metabolites), caffeine metabolites and the metabolites of the red meat carcinogens PhIP and MeIQx. All of the potential metabolites of interest were profiled in both positive and negative modes.

Table. 6.6 Candidate metabolites of CV consumption generated by consideration of biological plausibility ranked by molecular weight including expected *m/z* of parent ion in both positive and negative modes.

Candidate compound	***MW***	***Neg mode***	***Pos mode***
4-Methoxyglucobrassicin	478.49	477.49	479.49
Neoglucobrassicin	478.49	477.49	479.49
Glucobrassicin	448.46	447.46	449.46
Glucoraphanin	436.49	435.49	437.49
Gluconasturtiin	423.45	422.45	424.45
Glucoiberin	422.47	421.47	423.47
Glucoibervirin	406.48	405.48	407.48
Progoitrin	388.39	387.39	389.39
5,6,11,12,17,18-hexahydrocyclononatriindole (CTr)	387.47	386.47	388.47
Gluconapin	372.39	371.39	373.39
Sinigrin	358.36	357.36	359.36
SF-NAc	341.28	340.28	342.28
4-Methoxyascorbigen	335.28	334.28	336.28
1-Methoxyascorbigen	335.28	334.28	336.28
IB-NAc	327.26	326.26	328.26
PEITC-Nac	327.24	326.24	328.24
MPITC-Nac	311.26	310.26	312.26
SFN-NAc	309.22	308.22	310.22
Ascorbigen	305.28	304.28	306.28
IBN-Nac	295.19	294.19	296.19
PEN-Nac	295.17	294.17	296.17
MPN-Nac	278.26	277.26	279.26
BITC-Nac	277.18	276.18	278.18
BTC-Nac	277.18	276.18	278.18
BEN-Nac	277.18	276.18	278.18
3,3'-Di-4-methoxyindolylmethane	276.30	275.30	277.30
3,3'-Di-1-methoxyindolylmethane	276.30	275.30	277.30
4-Methoxybrassinin	266.35	265.35	267.35
1-Methoxybrassinin	266.35	265.35	267.35
AITC-Nac	263.15	262.15	264.15
ATC-Nac	263.15	262.15	264.15
AEN-Nac	263.15	262.15	264.15
indolo(3,2b)carbazole (ICZ)	254.28	253.28	255.28
3,3'-Diindolylmethane	246.30	245.30	247.30
BC-Nac	245.11	244.11	246.11
Brassinin 3-S-methyldithiocarbamoyl aminomethyl indole	236.35	235.35	237.35
AC-Nac	231.08	230.08	232.08
4-Methoxyindole-3-acetonitrile	186.18	185.18	187.18
1-Methoxyindole-3-acetonitrile	186.18	185.18	187.18
sulforaphane	177.28	176.28	178.28
4-Methoxyindole-3-carbinol	177.17	176.17	178.17
1-Methoxyindole-3-carbinol	177.17	176.17	178.17
Iberin	163.26	162.26	164.26
Phenethyl isothiocyanate (PEITC)	163.24	162.24	164.24
Indole-3-acetonitrile	156.18	155.18	157.18
3-(methylthio)propyl-isothiocyanate (radish isothiocyanate)	147.26	146.26	148.26
Indole-3-carbinol	147.17	146.17	148.17
sulforaphane nitrile (5-methylsulfinylpentanenitrile)	145.22	144.22	146.22
Iberin nitrile	131.19	130.19	132.19
3-phenylpropionitrile	131.17	130.17	132.17
Goitrin ((R)-5-vinyl-2-oxazolidinethione)	129.18	128.18	130.18
Iberverin Nitrile 1-cyano-3-(methylthio)propane	114.26	113.26	115.26
Butenyl Isothiocyanate	113.18	112.18	114.18
butenyl thiocyanate	113.18	112.18	114.18
1-cyano-3,4-epithiobutane	113.18	112.18	114.18
Allyl Isothiocyanate	99.15	98.15	100.15
allyl thiocyanate	99.15	98.15	100.15
3,4-epithiobutanenitrile	99.15	98.15	100.15
Crambene (1-cyano-2-hydroxy-3-butene)	97.12	96.12	98.12
Alanine	89.09	88.09	90.09
butenyl cyanide	81.11	80.11	82.11
Allyl cyanide	67.08	66.08	68.08

To determine the presence/absence of these various classes of metabolites, ions of the correct m/z were extracted from pre- and post-dose samples (where dose refers to CV or meat or caffeine). For example a comparison of extracted ion chromatograms (EICs) for sulforaphane-mercapturic acid would compare phase I pre-dose urine to phase II pre-dose urine samples in an individual. Peaks occurring only in phase II pre-dose urine samples could be considered to be relevant to the consumption of CV. Peaks that consistently appeared in the post dose samples were considered to potentially contain metabolites of interest (**Fig. 6.30**). EIC peaks were integrated for each sample in every time-point and phase. The relative integrals for each peak were then averaged and used to plot a bar chart showing a semi-quantitative overview of the excretion of the metabolite (**Fig. 6.31**). The relative integrals for each subject phase and time-point were adjusted for urinary volume. The standard deviation for each metabolite was also calculated for every time-point and phase. It should be stressed that the identification of these potential metabolites is based upon appropriate molecular ion and biological plausibility.

Potential metabolites for which there was a consistent extracted ion peak in the post dose time point/ phase included sulforaphane-NAc, allyl Isothiocyanate-NAc, caffeine, paraxanthine/theobromine/theophylline, 5-acetylamino-6-formylamino-3-methyluracil (AFMU). Signals indicative of tyrosine/mannitol/1-methylxanthine/ 7-methylxanthine, had similar MW and retention time and were therefore indistinguishable. The putative biomarkers also demonstrated evidence of time-course excretion profiles. The excretion of one of the potential metabolites of caffeine paraxanthine/theobromine/theophylline in particular demonstrated this over the 48 hours of urine sample collection throughout all phases of the study.

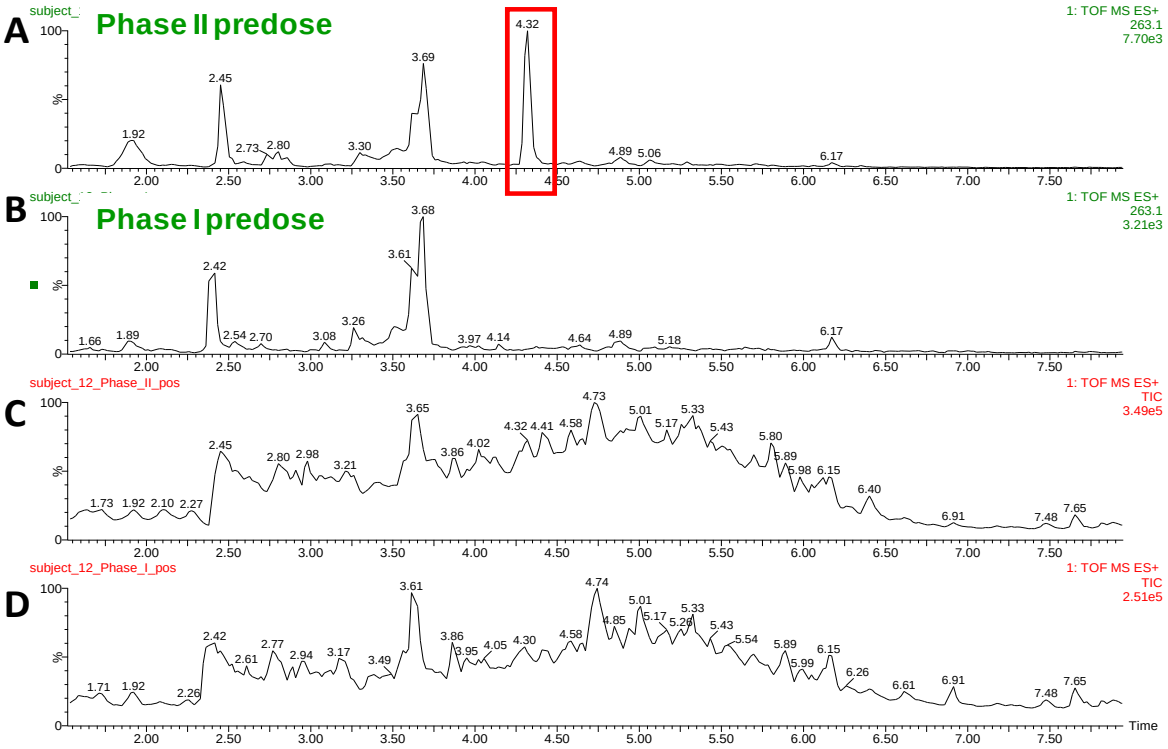


Fig. 6.30 Raw UPLC-MS total ion chromatograms (TICs) in positive mode for subject 12 for both phase I and II predose urine samples.

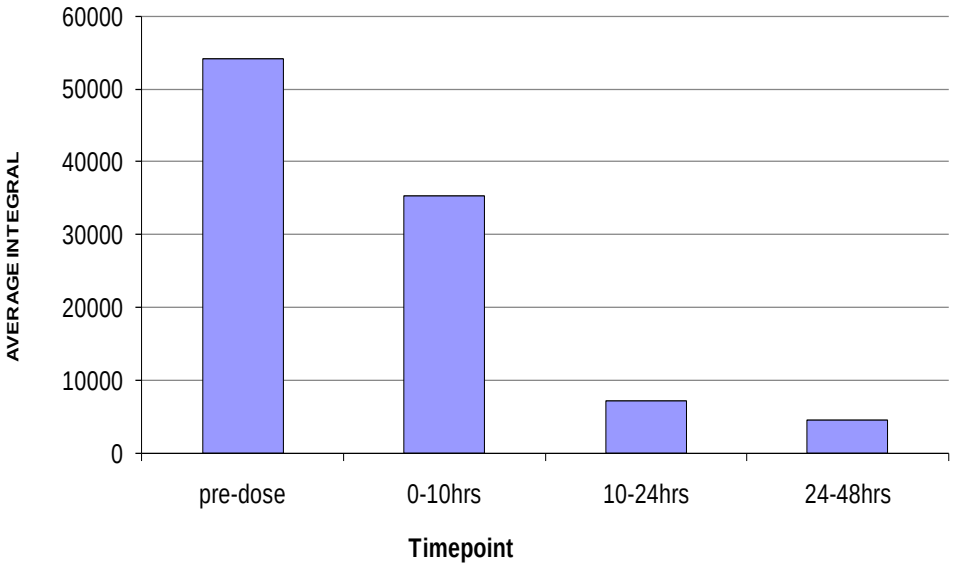
Comparison of extracted ion chromatograms in both Phase I and II predose samples showed a consistent peak at 4.32mins retention time. This can be considered a chromatographic peak diagnostic of Phase II predose urine samples. In this case the *m/z* 263.1 (**A** & **B**) is the expected parent ion of allyl isothiocyanate-NAc cysteine conjugate.

A=Extracted ion chromatogram for *m/z* 263.1 subject 12 phase II pre-dose urine.

B=Extracted ion chromatogram for *m/z* 263.1 subject 12 phase I pre-dose urine;

C=Chromatogram: Total ion count subject 12 phase II predose urine (high CV);

D=Chromatogram: Total ion count subject 12 phase I predose urine (restricted CV);



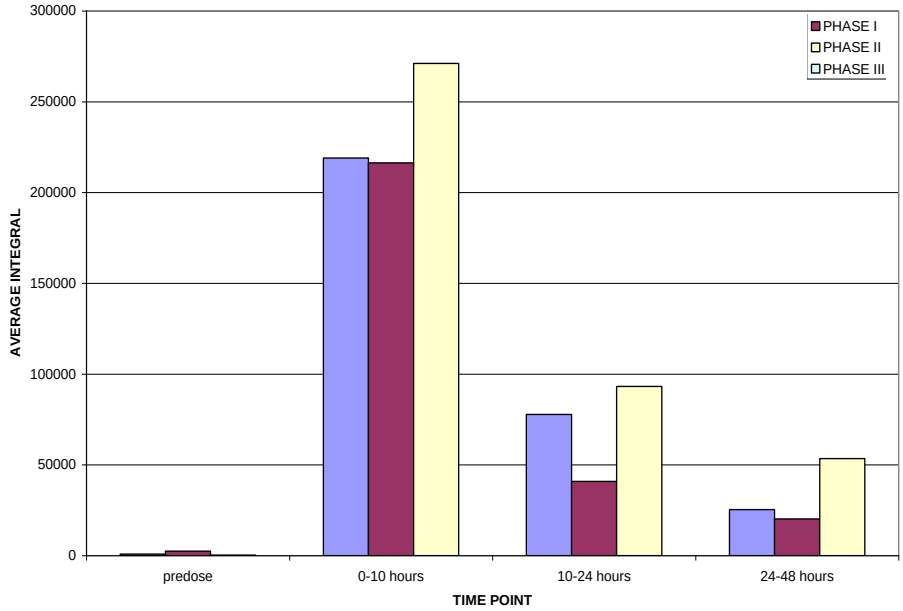


Fig. 6.31 Top: Semi-quantification of the relative integrals of 263.1*m/z* tentatively identified as allyl-isothiocyanate-mercapturate a known biomarker of CV consumption. Figure demonstrates a time-course excretory profile adjusted for urinary volume.
Bottom: Time course excretory profile for of one of the potential metabolites of caffeine paraxanthine/theobromine/theophylline over the 48 hours of urine sample collection throughout all phases of the study.

6.5.4.3 XCMS[1] statistical analysis

The acquired UPLC-MS chromatograms were additionally evaluated using XCMS[1] [157]. This programme performs peak picking in an automated and unbiased way. The programme uses the common MS netCDF file format that enables data export from any instrument platform. In this case the Waters LCT premier data (*.raw) files were converted to netCDF (*.cdf) using the databridge programme. Chromatograms were separated into their 12 separate sample classes for XCMS[1] processing, *via* Phase I predose, 0-10hrs, 10-24hrs, 24-48hrs, Phase II predose, 0-10hrs, 10-24hrs, 24-48hrs, Phase III predose, 0-10hrs, 10-24hrs, 24-48hrs. Following file conversion the XCMS[1] parameters for the R language were entered into an automation script. The XCMS[1] group bandwidth was set to 60, the Extracted Ion Chromatogram maximum was set to 250 and the width of overlapping m/z values was set to 0.5. All experimental iterations were performed on both the UPLC-MS positive and negative mode data sets (experimental iterations summarised in **Table. 6.7**.). XCMS[1] utilises hundreds of endogenous metabolites as standards and calculates a non-linear retention time correction profile for each sample.

Table. 6.7 Experimental iterations addressing primary study aims. All pair-wise comparisons were examined using the XCMS software on both the positive and negative UPLC-MS datasets.

Experimental Iteration		
Effect of CV consumption on metabotype		
Phase I predose	vs.	*Phase II predose*
Phase II predose	vs.	*Phase III predose*
Effect of meat meal and caffeine on metabotype		
Phase I predose	vs.	*Phase I 0-10 h.p.d.*
Phase III predose	vs.	*Phase III 0-10 h.p.d.*
Effect of CV on meat meal and caffeine alterations of metabotype		
Phase I 0-10 h.p.d.	vs.	*Phase II 0-10 h.p.d.*
Phase II 0-10 h.p.d.	vs.	*Phase III 0-10 h.p.d.*
Phase II predose	vs.	*Phase II 0-10 h.p.d.*

Following processing and peak picking, mass spectral features were retrieved as a report in a text file which was imported into excel. The report provides a list of extracted ion chromatograms from lowest to highest *p*-value; it also provides ANOVA, t statistics, fold change between m/z intensities and the number of peaks in the raw chromatograms containing the feature of interest.

The XCMS[1] programme has proven successful in identifying significant mass spectral biomarkers of CV consumption. As part of the experimental report XCMS[1] provides extracted ion chromatograms (EIC) showing the significant discriminators between classes of comparisons which are ordered according to *P* statistics (from t-test) and their corresponding box and whisker plots visually describe the standard deviation, error, mean and median values for each EIC. The *p*-values indicate the statistical significance and the fold change of this feature from the baseline.

Comparison of the Phase I predose vs. Phase II predose sample classes was processed in XCMS[1] in positive mode. Chromatograms of 11 urine samples from the Phase I predose and 13 urine samples from the Phase II predose were obtained from the experimental comparison in XCMS[1]; six significant features were identified and of these M411T155, M263T260 and M357T229 (where M refers to *m/z* ratio and T to retention time in seconds) can be considered unique to Phase II and therefore CV consumption (**Table. 6.8**). The other 3 represent significant statistical up-regulation within Phase II. Depending on the performance of the XCMS[1] peak picking algorithm, and assuming there was relatively low chromatographic drift between samples within the limits of the XCMS[1] alignment algorithm, features identified represent *m/z* values and retention times that differ from background noise and are therefore potential biomarkers of CV consumption. Peak selection problems can occur during the XCMS[1] peak alignment procedure, due to retention drifts between samples and overlapping or wrongly aligned peaks with similar *m/z* and can cause the peak picking algorithm to fail. This was problematical with the Phase II predose vs. Phase II 0-10hrs and Phase I 0-10 hrs vs. Phase II 0-10 hrs experiments in the positive mode data and comparison of Phase I predose vs. 0-10hrs and Phase III predose vs. 0-10hrs in negative mode. Unfortunately

in both positive and negative modes the XCMS[1] algorithm failed to identify any significant features associated with CV alterations of meat and caffeine metabolism.

Table. 6.8 Effect of CV consumption on metabotype. Phase I predose vs. Phase II predose urine samples experimental iteration.
Table shows the top 6 most statistically significant XCMS[1] features (*m/z* and retention time in seconds) in positive mode ranked by *p*-value and showing fold change from baseline.

UPLC-MS XCMS component

statistical significance (p value)	*m/z*	retention time (seconds)	fold change from baseline
1.31E-06	277.1	*77*	*8.2*
1.84E-05	337.1	*148*	*12.1*
4.71E-04	411.1	155	8.1
0.00226	263.1	260	70
0.00239	357.1	*229*	*3.5*
0.0164	488.1	*318*	*6.6*

An example of an EIC and box and whisker plot can be seen in **Fig. 6.32**. This significant feature; *m/z* 449.1 and retention time 173 seconds; represents the most statistically significant (based on *p*-value: 4.51E-07) biomarker of CV consumption in the negative mode dataset.
The EIC presents the alignment and peak intensity of the raw chromatograms. There was a fold increase of 8.18 from the baseline in the Phase II pre-dose urine samples and this is highlighted by a significant upregulation of this feature in the phase II predose urine samples. There is also evidence of time-dependent reduction in excretion of this feature in the phase II urine samples with evidence of excretion in the 0-10 h.p.d. and 10-24 h.p.d. urine samples which can be seen in the box and whisker plot with a visual representation of the standard deviation, error, mean and median values.

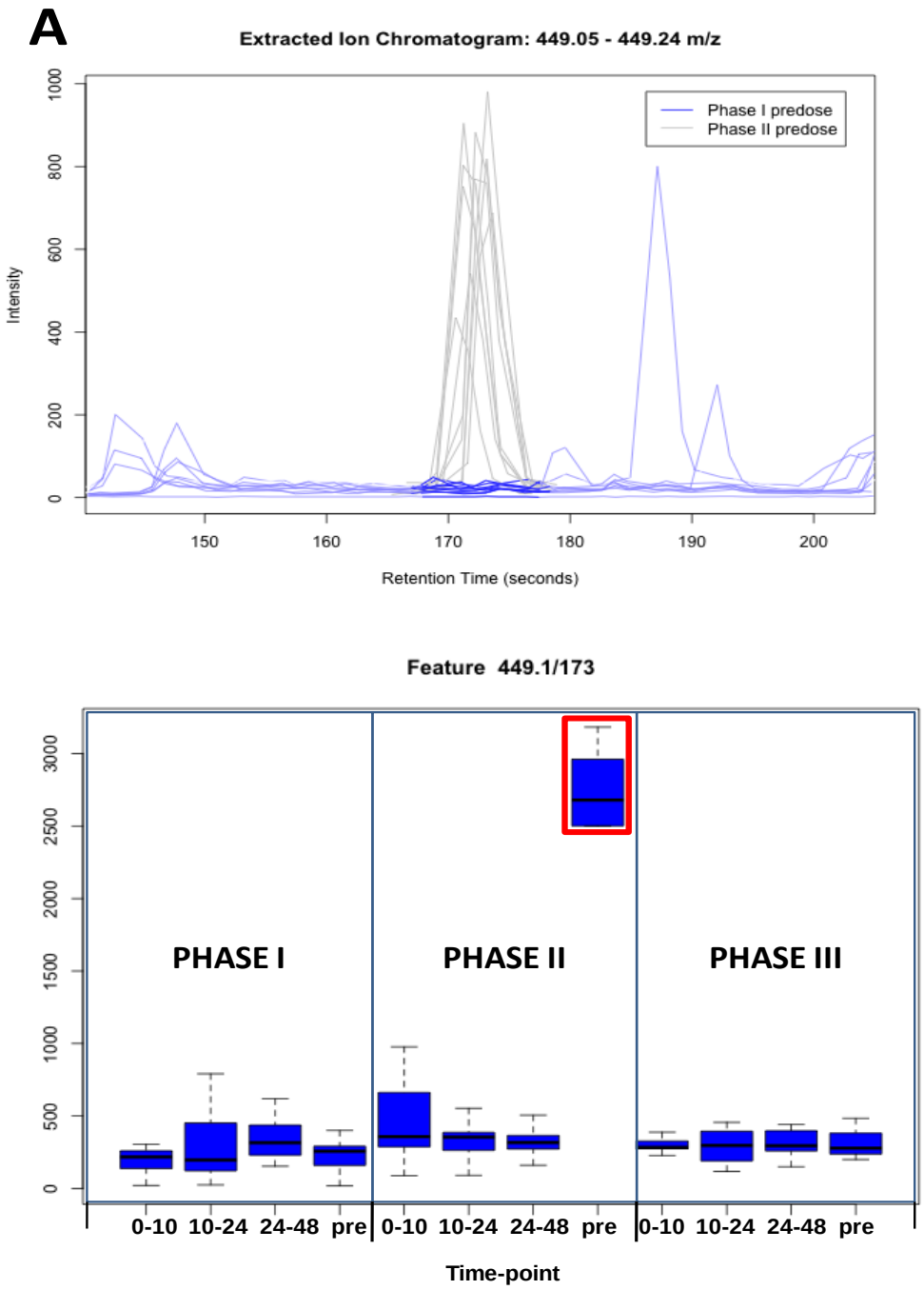


Fig. 6.32 Example of an Extracted Ion Chromatograms (EIC) (A) and box and whisker plot (B) of significant feature; *m/z 449.1* retention time 173 seconds;
taken from the XCMS[1] comparison of Phase I pre-dose vs. Phase II pre-dose urine samples in negative mode. This represents the most statistically significant biomarker based on *p*-value: 4.51E-07) of CV consumption. Red ringed box demonstrates significant up-regulation of this feature in the phase II predose samples

The presence and correct retention time of these extracted molecular ions was therefore manually confirmed within the raw chromatograms. The relative ion intensities of each metabolite were compared to reveal differences and highlight potential biomarkers.

Analysis of the positive and negative mode chromatograms revealed a total of 10 statistically relevant putative biomarkers in the positive mode and 11 putative biomarkers in the negative mode data associated with CV dietary intervention based on their *m/z* ratio and retention time (**Table. 6.9**). The features represent either complete exclusivity to the CV vegetable consumption sample class or significant statistical up regulation within the class.

Table. 6.9 Showing statistically significant Extracted Ion Chromatograms (EICs) generated by the XCMS[1] algorithm from the UPLC-MS data in both positive (10) and negative modes (11) which discriminate the high CV from the restricted CV dietary phases.

Features are ordered by statistical significance (*p*-value).

Positive Mode XCMS significant features				
M/z	***Retention time (sec)***	***Fold Change***	***P -value***	***Experimental Iteration***
277.1	77	8.26	1.31E-06	PIpre vs PIIpre
175.1	34	5.60	7.95E-06	PIIpre vs PIIIpre
337.1	148	12.14	1.84E-05	PIpre vs PIIpre
411.1	155	8.19	4.72E-04	PIpre vs PIIpre
210.1	55	1.90	5.39E-04	PIIpre vs PIIIpre
420.1	224	5.00	8.82E-04	PIIpre vs PIIIpre
132.1	33	5.40	1.10E-03	PIIpre vs PIIIpre
263.1	260	70.07	2.26E-03	PIpre vs PIIpre
357.1	229	3.57	2.39E-03	PIIpre vs PIIIpre
315.1	146	7.30	3.84E-03	PIIpre vs PIIIpre
488.1	318	6.63	0.0164	PIpre vs PIIIpre

Negative Mode XCMS significant features				
M/z	***Retention time (sec)***	***Fold Change***	***P -value***	***Experimental iteration***
449.1	173	8.18	4.51E-07	PIpre vs PIIpre
259.1	172	22.66	1.07E-06	PIpre vs PIIpre
303.1	37	9.15	3.91E-06	PIpre vs PIIpre
275.1	77	17.44	6.34E-06	PIpre vs PIIpre
165.1	98	12.86	1.22E-05	PIpre vs PIIpre
211.1	98	5.79	5.27E-05	PIIpre vs PIIIpre
285.1	115	15.37	1.75E-04	PIpre vs PIIpre
151.1	37	8.93	2.32E-04	PI 0-10 vs PII 0-10
251.1	226	15.77	8.51E-04	PIpre vs PIIpre
261.1	262	8.79	2.71E-03	PIpre vs PIIpre
267.1	162	8.20	8.47E-03	PIpre vs PIIpre

However, similarities have been observed in expected *m/z* ratio of the parent ions of known constituents/metabolites of CV vegetables and of putative biomarkers. Standard compounds acquisition was limited by financial considerations but many biologically plausible metabolites were acquired for each significant MS feature. *De novo* metabolite identification is challenging and was unfortunately beyond the scope of this thesis,

however these routines have discovered spectral features associated with CV consumption.

6.5.4.4 MS/MS fragmentation analysis of XCMS chromatographic features

The most significant features identified by XCMS[1] in human urine were characterised using MS/MS fragmentation analysis and initially the online MS/MS fragmentation databases such as Metlin (Scripps) and the human metabolome database (Hmdb, Genome Alberta) were utilised to attempt to identify the observed fragmentation patterns.

MS/MS fragmentation analysis of the significant features identified by XCMS[1] was obtained at a variety of collision energies to give high, medium and low energy fragmentation patterns, whilst retaining the parent molecular ion at lower ionisation energies. MS/MS fragmentation was performed using voltage ramping between 10-50v for each of the 19 significant MS features putatively identified in human urine in positive mode and for the 11 significant features positively identified in negative mode. The major fragments were tabulated and common mass losses were identified. An example is shown of XCMS[1] spectral features diagnostic of meat and caffeine intervention in the positive mode subjected to tandem MS/MS fragmentation analysis (10-50V voltage ramp) (**Table 6.10**). The major fragments produced at 10V, 30V and 50V ionisation energies were tabulated; these were used for comparison to online tandem MS/MS metabonomic databases and standard compound fragmentation patterns.

Table. 6.10 Major MS/MS fragments of urinary XCMS[1] features discriminatory of meat and caffeine intervention in positive mode (10-50V voltage ramp).

Significant XCMS features positive mode							
Meat and caffeine intervention					**ms/ms major fragments**		
m/z	**Retention time (sec)**	**Fold Change**	***P* -value**	**XCMS experiment**	**10v**	**30V**	**50V**
299.1	56	5.82	2.00E-06	PI pre vs PI 0-10h	281.99 262.96 253.1 182.05 156.07 126.05	247.94 237.01 217.09 156.06 126.05 110.07	156.04 126.05 110.08
309.1	82	3.32	4.85E-06	PI pre vs PI 0-10h	292 277.2 271.18 263.09 201.95	277.2 271.25 263.1 223.97 201.95 156.07 142.95 122.1 110.08	263.11 154.05 110.08
413.1	192	3.55	3.09E-05	PI pre vs PI 0-10h	281.09	281.09	285.12 239.09 162.04 136.05 120.2
302.1	62	4.29	4.10E-05	PI pre vs PI 0-10h	232.9 164.92 143.03 126.06 106.95	269.08 227.1 186.9 171.06 164.92 143.05 126.05 114.06 106.95	186.93 164.92 143.05 126.05 86.05
565.1	155	5.11	5.00E-05	PI pre vs PI 0-10h	549.22 539.21 404.01 302.02 279.13	520.05 485.05 390.02	low ion count
558.1	317	3.74	1.94E-04	PI pre vs PI 0-10h	463.17 392.06 349.2 303.15 274.13 79	543.32 476.88 425.28 363.15 347.15 202.58 127.09 102.02	383.18 343.12
186.1	50	12.88	2.32E-06	PIII pre vs PIII 0-10h	169.03 144.02 130.05 126.07 99.97 84.95	169.04 144.06 140.07 130.05 126.07 114.01 101.97 99.97 84.08	84.08 98.97
205.1	42	6.00	2.86E-06	PIII pre vs PIII 0-10h	145.04 125.03 85.04	191.04 145.08 125.05 107.08 85.02	144.09 85.05
168.1	201	31.04	2.92E-06	PIII pre vs PIII 0-10h	153.09 128.96 96.97 90.21 86.05	153.09 128.95 96.96 86 84.1	153.09 138.06 113.01 110.96 98.02 96.96 89.93 85.98 70.95

The most promising XCMS[1] feature suggestive of the ^{1}H-NMR findings at 151.1*m/z* negative mode was the correct *m/z* to potentially be the deanimated thiolactate of SMCSO; 3-methylthiolactic acid sulphoxide. MS/MS fragmentation of this unknown biomarker produced a fragmentation pattern similar to SMCS according to an experimental low energy fragmentation on the human metabolome database (**Fig. 6.33**). This XCMS[1] significant feature also shows evidence of increasing excretory concentration in the negative mode when comparing phase I to phase II. Therefore perhaps there is an effect of CV consumption on level of excretion of this metabolite which could possibly be 3-methylthiolactic acid. Furthermore XCMS[1] peak picked a significant feature 303.1*m/z* which, following fragmentation using a 10-30v voltage ramp, proved to be a dimer formation of 151.1*m/z* within the ion source. However, in the absence of an authentic standard or ease of synthesis, these identities were not confirmed.

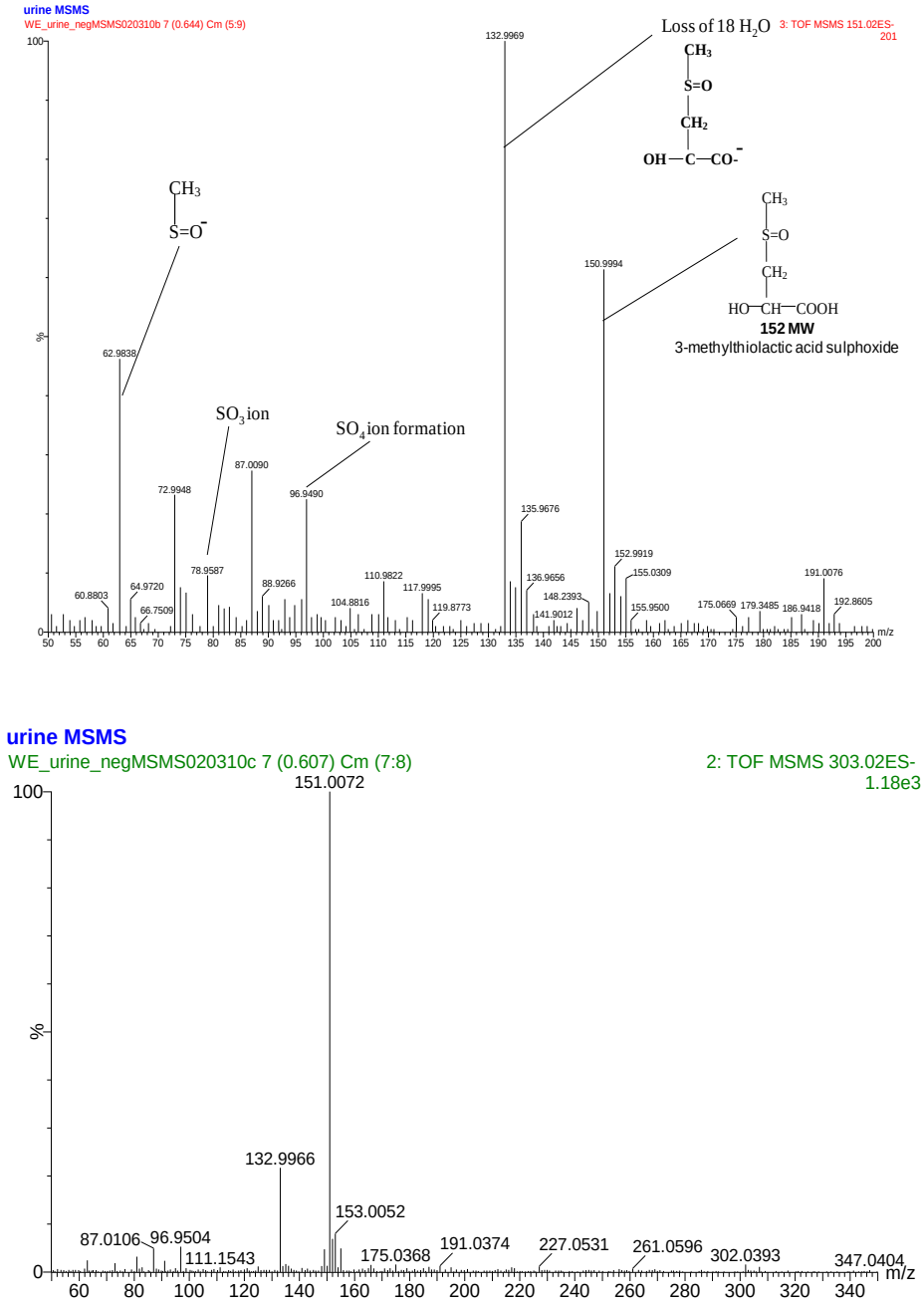


Fig. 6.33 Top: MS/MS negative mode fragmentation analysis of XCMS[1] significant feature 151.1*m/z* in human urine associated with CV consumption.

A 10-40v voltage ramp was used for the fragmentation of this biomarker in urine. A putative fragmentation pattern is attempted to address the observed pattern in the absence of an authentic standard.

Bottom: MS/MS negative mode fragmentation analysis of XCMS[1] significant feature 303.1*m/z*. A 10-50v voltage ramp was used for the fragmentation. This is believed to represent dimer formation of 151.1*m/z* in the ionisation source.

For example, the feature at 263.1 m/z, retention time 260 secs with a fold change of 70.07, was associated exclusively with the CV dietary intervention predose urine sample. Online database searches on the Metlin database provided 24 matches but these were found to have little biological relevance. The Human Metabolome database (Hmdb, Genome, Alberta) provided 6 *m/z* matches which were biologically plausible, these included the following adducts; homovanillic acid sulphate ($M+H^{1+}$); 3-methylindole ($2M+H^{1+}$); sorbitol-6-phosphate ($M+H^{1+}$); L-propyl-L-phenylalanine ($M+H^{1+}$); L-phenylalanyl-L-proline ($M+H^{1+}$) and creatine ($2M+H^{1+}$).

Where database experimental fragmentation patterns were available, candidate metabolites were eliminated by comparison of fragmentation patterns to the MS/MS fragmentation patterns of the features identified in human urine. Based on database searches and the self generated candidate metabolites list available authentic standard compounds were acquired and dissolved in UPLC-MS grade H_2O to a final concentration of 100mM. In total 34 authentic standard compounds in positive mode (**Table. 6.11**) and 31 authentic standard compounds in negative mode were subjected to MS/MS fragmentation using the same experimental conditions, elution gradients and ionisation parameters (10-50V voltage ramp) as for the metabolites in urine for ease of comparison and reproducibility.

Table. 6.11 Standard compounds subjected to MS/MS fragmentation analysis (10-50V voltage ramp) in positive mode.

The corresponding XCMS[1] significant features in urine, fold change from baseline and *p*-values are included. Candidate compounds were generated based on biologically plausibility and database searches. Putative adduct formations were obtained from Human Metabolome database (Hmdb, Genome, Alberta) are also included.

XCMS feature	Fold change	*p*- value	XCMS exp.	Candidates	Putative adduct formation (Hmdb)
M277T77	8.26	1.31E-06	PIpreVsPIIpre	nicotinamide n-oxide	2M+H(1+)
				2-hydroxybenzoic acid	2M+H(1+)
				3-hydroxybenzoic acid	2M+H(1+)
M337T148	12.14	1.84E-05	PIpreVsPIIpre	3-hydroxymandelic acid	2M+H(1+)
				homogentisic acid	2M+H(1+)
				5-methoxysalicylic acid	2M+H(1+)
				p-hydroxymandelic acid	2M+H(1+)
				pyridoxamine	2M+H(1+)
M263T260	70.07	2.26E-03	PIpreVsPIIpre	Homovanillic acid sulfate	M+H(1+)
				creatine	2M+H (1+)
M299T56	5.82	2.00E-06	PIpreVsPI0-10	homoanserine	M+2Na-H(1+)
				L-methionine	2M+H(1+)
				urocanic acid	2M+Na(1+)
				3-hydroxybenzoic acid	2M+Na(1+)
				4-hydroxybenzoic acid	2M+Na(1+)
				Enkephalin L	M+ACN+2H(2+)
				n-acetyl putrescine	2M+K(1+)
M310T82	3.32	4.85E-06	PIpreVsPI0-10	picolinic acid	2M+ACN+Na(1+)
				nicotinic acid	2M+ACN+Na(1+)
M413T192	3.55	3.09E-05	PIpreVsPI0-10	3-hydroxyhippuric acid	2M+Na(1+)
				alpha-hydroxyhippuric acid	2M+Na(1+)
				methylisocitric acid	2M+H(1+)
				L-Glutamine	2M+NH4(1+)
M302T62	4.29	4.10E-05	PIpreVsPI0-10	1-Methylnicotinamide	2M+3H2O+2H(2+)
				Pantothenic acid	2M+2ACN+H(1+)
				Tyramine	2M+3H2O+2H(2+)
M558T317	3.74	1.94E-04	PIpreVsPI0-10	Pyridoxal 5'-phosphate	2M+ACN+Na(1+)
M175T34	5.61	7.95E-06	PIIpreVsPIIIpre	N-acetylputrescine	M+2Na-H(1+)
				Arginine	M+H(1+)
				N-Methyltryptamine	M+H(1+)
				3-hydroxyphenylacetic acid	M+Na(1+)
M210T55	1.93	5.39E-04	PIIpreVsPIIIpre	acetylcholine	M+ACN+Na(1+)
				Sulforaphane	M+CH3OH+H(1+)
M420T224	5.83	8.82E-04	PIIpreVsPIIIpre	N-acetylglutamic acid	2M+CAN+H(1+)

Putative biomarkers of CV and meat and caffeine consumption were eliminated by the process of tandem MS/MS fragmentation analysis. Unfortunately even the more biologically relevant standard compounds were not comparable to the urinary mass spectral features by MS/MS fragmentation. The spectral feature *m/z* 263.1 identified by targeted analysis and XCMS[1] statistical analysis in the positive mode dataset was thought to potentially be the *N*-acetyl-L-cysteine conjugate of the glucosinolate hydrolysis product allyl isothiocyanate due to a similarity in expected molecular ion, however this was found not to be the case as the standard compound produced a different fragmentation to the metabolite identified in urine. The MS/MS fragmentation in urine using a 10-50V produced characteristic fragments (*m/z* 147, 136, 130, 91 and 84), however the AITC-NAc standard produced characteristic fragments corresponding to the NAc and allyl isothiocyanate respectively (*m/z* 164 and 100).

With exhaustive resources and more comprehensive metabonomics mass spectral databases, the identity of the biomarkers identified could be confirmed; however within the confines of this study it was not possible. In order to augment the fragmentation analysis further standard compounds of interest were also subjected to MS/MS fragmentation analysis in both the positive and negative modes. These were compounds that did not necessarily match the expected molecular ion *m/z* of any of the significant spectral features identified by XCMS[1] but were biologically plausible and available within the group (**Table. 6.12**). Furthermore, ion source adduct formation can be difficult to predict in urine and significant XCMS[1] features could have potentially represented fragments of urinary metabolites.

Literature searches for molecular ions and common fragments were also conducted and the results compiled to aid elucidation of the unknown mass spectral features.

Table. 6.12 Further candidate compounds of biological plausibility subjected to positive mode MS/MS fragmentation analysis (10-50V voltage ramp).
Major tandem MS/MS fragments were recorded to facilitate similarity comparison with urinary spectral features identified by XCMS1 statistical analysis.

Other compounds of interest	**major MS/MS fragments** voltage ramp 10-50V
Glutamine	84,130,147
N-acetylcysteine	76,122,146,164
carnitine	162
Mannitol	211,183,162,147,129,111,85,69
Caffeine 1,3,7-trimethylxanthine	181, 157,140,116,111,100,77
Paraxanthine	162,144,126,116,103,79
5-Acetylamino-6-formylamino-3-methyluracil	185,148,130,118,100,77
xanthine	137
1-methylxanthine	162,144,126,116,103,85
1-methyluric acid	157,140,118,116,111,100,91
chlorogenic acid	195,163,147
1,3,9-trimethylxanthine	185,162,144,126,85
indole-2-methylalcohol	163,148,130,118,111
indole-2-carbonyl acetate	189,144,111,72
indole-2-acetonitrile	185,157,130,111,100
catechin	291,139
Sulforaphane	140,114,79
AITC-Nac conjugate	164, 100

6.5.4.5 XCMS^2 analysis

Whilst XCMS^1 statistically assesses and ranks features in metabonomic UPLC-MS data, no information is provided on structural identity. XCMS^2 provides an automatic search of tandem MS/MS fragmentation data against experimental MS/MS data from metabolites contained in the Metlin online database. Search hits are scored based on shared peak counts and provide information on shared fragment ions between the experimental MS/MS data and reference fragmentation patterns thereby identifying unknowns based on their fragmentation patterns. The greatest utility of the software is identifying structural similarity of unknowns with reference fragmentation patterns using a 'similarity search' algorithm. This algorithm detects possible structural motifs in unknown metabolites which might produce characteristic fragment ions and neutral losses [284].

Comparison of MS/MS data of the UPLC-MS biomarkers identified in urine did not result in any 100% search hits on the Metlin database however similarity was seen in some fragments and neutral losses (**Table. 6.13**) Unfortunately, the candidate compounds deduced from similarity search of the Metlin database are dominated by drug molecules, fatty acids, etc and are unlikely to be relevant here. The Metlin database is still in development, with 2175 MS/MS fragmentation patterns available presently. This is a time intense process and is slowly being augmented.

The concept of XCMS^2 is excellent for discovering structural similarity of unknowns and holds great future potential. All chromatographic features identified by XCMS^1 in both positive and negative modes were submitted to tandem MS/MS fragmentation analysis and then compared to the Metlin database using XCMS^2.

Table. 6.13 Subset of XCMS^2 similarity search report from the experimental iteration of Phase I predose vs. Phase II predose showing the two putative biomarkers of 277.1 *m/z* and 263.1*m/z*.

The report compares the experimental MS/MS fragmentation data to experimental MS/MS data within the Metlin MS/MS database. Similarity is calculated based on *m/z*, retention time and similarity of both common fragments and common neutral losses. As can be seen similarity scores of common fragments and neutral losses are given along with the name of the compound on the Metlin database where an experimental MS/MS fragmentation exists. This experiment can provide more information on potential structures for unknown mass spectral biomarkers.

m/z	rtmin	rtmax	Experiment Collision Energy	Fragment Score	Neutral Score	Common Neutral loss	Common Fragment	Compound Name	Metlin Mass	Collision Energy
277.0788	33.586	47.346	30	50	0	42.0457	84.0441	Tolbutamide	277.058639	40
277.1025	80.132	98.07	30	50	0	80.0623	136.076	Neostigmine	222.1368208	40
277.1025	80.132	98.07	30	50	0	27.011	91.054	Oleamide	117.057849	0
277.1025	80.132	98.07	30	50	0	43.971	136.076	Omeprazole	345.114714	0
277.1025	80.132	98.07	30	50	0	110.987	136.076	Sulfadoxine	310.073578	40
277.1025	80.132	98.07	30	50	0	42.0457	84.0441	Tolbutamide	277.058639	40
263.0175	150.555	164.081	30	50	0	0.0005	85.0301	Canrenoate	174.111676	10
263.0783	177.601	191.127	30	50	0	32.9882	91.0542	L-Thyronine	301.167794	10
263.0783	177.601	191.127	30	50	0	27.011	91.054	Oleamide	117.057849	0
263.0379	206.859	222.585	30	0	50	1.007	117.058	1-Acetylindole	159.0684	0
263.0379	206.859	222.585	30	0	100	1.007	131.037	alpha-Estradiol	133.0528	0
263.0379	206.859	222.585	30	50	0	17.026	101.071	Glutamine	146.069143	0
263.0379	206.859	222.585	30	50	0	0.984	101.071	Glutamine	146.069143	10
263.0379	206.859	222.585	30	50	0	32.9882	91.0542	L-Thyronine	301.167794	10
263.0379	206.859	222.585	30	50	0	48.0855	83.0549	Myristic Acid	280.24023	40
263.0379	206.859	222.585	30	50	0	75.9818	95.0132	N-Acetyl-D-galactosamine	194.042655	20
263.0379	206.859	222.585	30	0	50	1.007	43.0187	Norpseudoephedrine	151.099714	20
263.0379	206.859	222.585	30	50	0	27.011	91.054	Oleamide	117.057849	0
263.0379	206.859	222.585	30	0	50	1.007	43.0187	Phenylpropanolamine	151.099714	20

6.6 Discussion

6.6.1 S-Methyl-L-cysteine sulphoxide

CVs are known to contain the sulphur containing phytochemical SMCSO derived from amino acid biosynthesis [285]. 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 [286].

The finding of SMCSO and its *N*-acetylated mammalian metabolite being diagnostic of subjects consuming 500g of CV per day over a period of 12 days is significant as a biomarker of consumption of CV. Unfortunately SMCSO is known to occur widely within the *allium* family (also known as methiin) as a plant secondary metabolite and component of the sulphur containing amino acid pool [287] and one of the reasons garlic and other alliums are the subject of investigation as functional foods. It is therefore not a plant metabolite exclusively obtained from the broccoli and Brussels sprouts. The fact that alliums were restricted during the study period suggests that the compound was indeed obtained from the high CV diet in Phase II as it is also found in abundance in both Brussels sprouts (~68mg/100g) and broccoli (~19.1mg/100g) and is ubiquitously present in vegetables of the family *cruciferae*.

Evidence exists within the literature of biological activities of this sulphur containing amino acid within both *in vitro* and *in vivo* animal models. However 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 statistical significance of SMCSO and its *N*-acetylated metabolite as urinary biomarkers of CV consumption provides impetus for further study of the biological role and anti-carcinogenic potential of this plant secondary metabolite within humans.

6.6.2 Biosynthesis of SMCSO and role within the plant

The biosynthetic pathway of sulphur containing cysteines within plants has been elucidated by chemical analysis and radiolabel feeding studies. These studies reveal an origin of SMCSO in plant sulphur metabolism, derived from SO_4^{2-} which is the primary source of inorganic sulphur in plants, for the biosynthesis of cysteine, glutathione and subsequently SMCSO [287]. The accumulation of SO_4^{2-} in plant cells occurs in the root, where it is initially converted to 5-adenylylsulphate (APS) catalysed by the enzyme ATP sulphurylase in plastids. APS is enzymatically converted to an intermediate sulphite by APS reductase and finally to a sulphide by sulphide reductase.

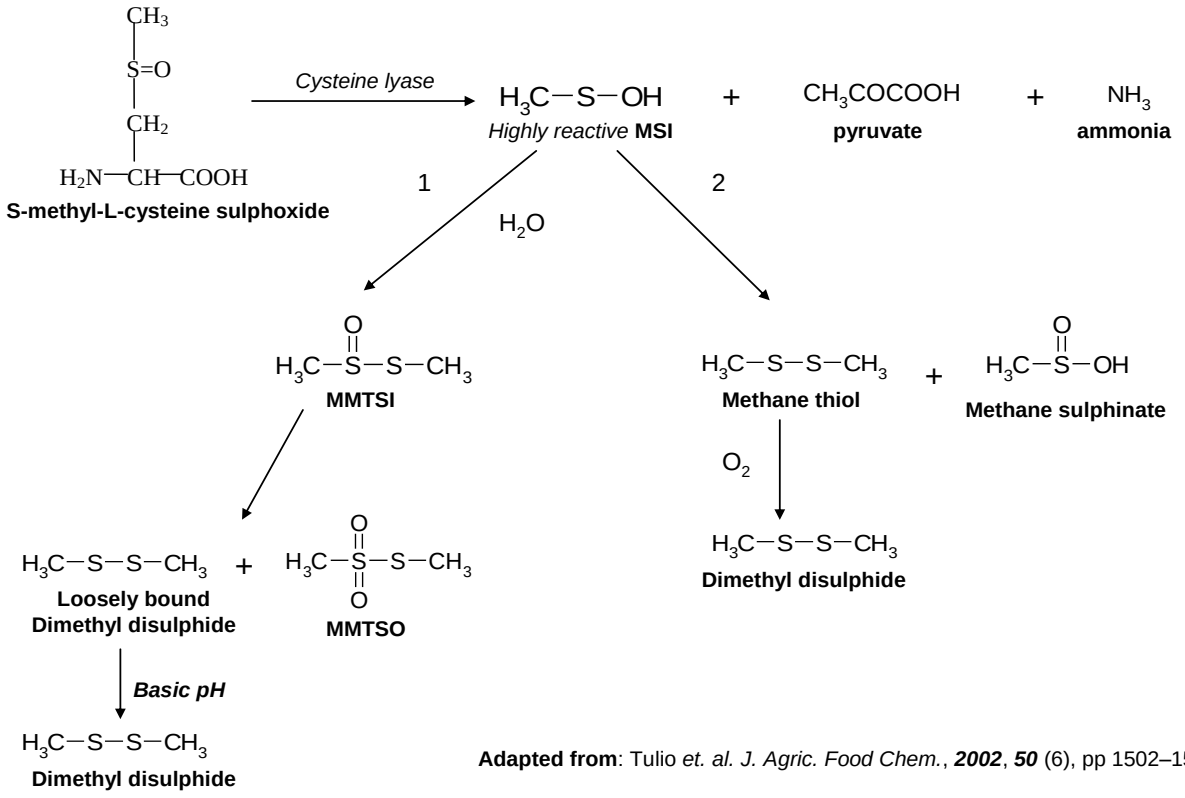
Cysteine is then formed by reaction of the sulphide with *O*-acetylserine by OAS-thiol-lyase. The cysteine formed is then channelled into protein synthesis pathways and numerous plant secondary metabolic pathways. In the case of SMCSO, a methylation and oxidation step takes place.

It is known that only the L stereoisomer of SMCSO occurs in nature. SMCSO represents the most ubiquitously and in some cases, such as Brussels sprouts, the most abundantly found sulphur containing amino acid in both vegetables of the *cruciferae* and *alliums*.

The biological role of SMCSO and its cysteine-lyase mediated breakdown products following tissue disruption, is almost certainly as an antimicrobial, antifungal and antiparasitic. Its reactive intermediate products most probably act as enzyme inducers and inhibitors following herbivorous attack and mechanical damage. Interestingly, these same products appear to be beneficial to human health [288].

6.6.3 Experimental studies of SMCSO and potential biological roles within humans.

Following mechanical disruption of CV tissue, cysteine sulfoxide *lyases* break down SMCSO into pyruvate, ammonia and methanesulfenic acid (MSI). Dependent on conditions the highly reactive methane sulfenic acid intermediate chemically disproportionates to a variety of metabolic endpoints, S-methyl methanethiosulfenate MMTSO, S-methyl methanethiosulfinate (MMTSI), MSI, dimethyldisulfide and methanethiol [289]. MMTSO formed from dimerization of methane sulfenic acid and the disproportionation of methyl methanethiosulphinate, is regarded as a form of phytoalexin which on wounding of the plant tissue prevents bacterial or yeast infections (**Fig. 6.34.**) [290].



Adapted from: Tulio *et. al. J. Agric. Food Chem.*, ***2002***, ***50*** (6), pp 1502–1507

Fig. 6.34 Metabolic schematic demonstrating enzymatic hydrolysis of SMCSO by cysteine lyase following CV tissue disruption.

Dependent on conditions such as pH and vegetable treatment the volatile dimethyl disulphide represents the metabolic endpoint by reaction of methane thiol or disproportionation of methane methylthiosulphinate (MMTSI).

MMTSI contains the thiosulphinate functional group *R*-S(O)-S-*R*, a commonality with the allicin compound found in garlic which is thoroughly researched and implicated in many of the health benefits of *alliums* found in animal studies. It is thought to be responsible for reducing atherosclerosis and fat deposition, normalising the lipoprotein balance, decreasing blood pressure, anti-thrombotic effects, anti-inflammatory activities and functions as an antioxidant [287]. It is expected that MMTSI and MMTSO would share many of these biological effects and this is reflected by studies in the literature that confirm similar beneficial effects.

Oral administration of SMCSO was found to mimic the effects of cabbage extract in suppressing hypercholesterolemia by upregulating cholesterol catabolism, namely cholesterol 7α-hydroxylase in hepatoma-bearing rats [291].

Studies of the antimicrobial properties of SMCSO were performed against 15 species of bacteria and 4 species of yeasts using both cabbage and fermented cabbage (sauerkraut) [292]. It was found that SMCSO itself and both dimethyl sulphide/dimethyl disulphide had a minimum inhibitory concentration (MIC) of 500ppm. It was also found that the breakdown products of SMCSO had antimicrobial activity. MMTSI had an MIC between 50-200ppm for all bacteria and between 6-10ppm for all species of yeast and MMTSO also had a similar antimicrobial activity with an MIC of 20-100ppm for bacteria and between 50-200ppm for species of yeast [292]. This suggests an antimicrobial role of this secondary plant metabolite for both disrupted and pathogenically invaded plant tissues and may also be a contributory factor to the beneficial properties of CV within the mammalian GI tract.

The nutritional role of SMCSO seems to have been largely studied within Asian countries and in the developing world where CVs represent a significant component of diet [291, 293]. These studies focus on upon both CVs and *alliums* and are motivated by a desire to utilise both CV's and *alliums* as functional foods. There is accumulated evidence to suggest a significant nutritional role of SMCSO as chemoprotective agent in

various aetiologies however these studies are largely in rat models and no *in vivo* human studies exist to confirm the observed beneficial effects [294, 295].

Rats maintained on a hypercholesterolaemic diet which produced high cholesterol levels in both the blood plasma and liver showed a marked depression of both plasma and liver cholesterol levels particularly LDL and VLDL following addition of SMCSO to this diet [296]. However this was not shown to be the case for other sulphur containing amino acids such as SMCS, cysteine, S-methylcysteine and S-methylmethionine which had little to no effect on hypercholesterolaemia.

Antidiabetic and antihyperlipidaemic effects attributable to SMCSO have also been observed in experimental rats and are comparable to those of antidiabetic drugs glibenclamide and insulin and the antilipidaemic drug gugulipid [293]. Sprague-Dawley rats made diabetic with the toxic glucose analogue Alloxan were orally administered with a 200mg/kg body weight dose of SMCSO for a 45 day period. This significantly reduced blood glucose and lipid levels in both serum and tissues. Also observed was an altered activity of liver hexokinase, glucose 6-phosphatase and HMG CoA reductase to normal levels.

Lipid profiles in serum and tissues showed a reduced concentration of total cholesterol, triglycerides and phospholipids [294]. Total lipoprotein lipase activity in the adipose tissues was also decreased with a subsequent decrease in the free fatty acid levels in serum and tissues. Increased excretion of bile acids and sterols was also observed following the SMCSO dosing regimen. The results of this study suggested that SMCSO causes reduction of endogenous lipogenesis, increased lipid catabolism and excretion of bile acids. This is thought to occur by interaction of the cysteine moiety of SMCSO inhibiting lipogenic enzymes within the liver as well as increasing activity of cholesterol 7α-hydroxylase [293, 294].

It is also known that the reduction of SMCSO contributes to oxidation of the lipogenic coenzyme NADPH which may be contributing to the hypolipidaemia observed [297].

As much of the research focus has been upon the anti-carcinogenic activity of the glucosinolates and their hydrolysis products, the contribution of SMCSO to the beneficial effects of CV has largely been overlooked. The anti-mutagenic activity of SMCSO in isolation has been demonstrated utilising the mouse bone marrow micronucleus assay [295]. ICR mice were treated with the tobacco carcinogen benzo[*a*]pyrene concurrently with SMCSO and its metabolite MMTSO known to form as a result of cysteine lyase activity following disruption of the plant cell wall. Observation of micronucleated polychromatic erythrocytes (MCPEs) revealed a 0.5mmol of SMCSO and a 0.05mmol per kg bodyweight dose produced a reduction in micronucleus formation of 31% and 33% respectively [295]. Higher doses of MMTSO (0.5mmol and 1.0mmol / kg body weight) however produced severe acute toxicity. This suggests that these two plant secondary metabolites of CV could contribute to the anticarcinogenic activity along with the hydrolysis products of glucosinolates.

SMCSO has a propensity to form MMTSO under many conditions and would most likely occur in mammals. MMTSO can be formed by the action of the CV enzyme cysteine lyase in the presence of pyridoxal-5'-phosphate as a cofactor. There is evidence to suggest MMTSO may also be formed by the gut microbiota, as evidence exists in ruminants where it was identified as the Kale anaemia cofactor in kale poisoning [59]. Thermal hydrolysis of SMCSO in HCl to MMTSO and DMDS has also been demonstrated [298]. Since SCMSO can readily form the active metabolite MMTSO, even without the action of the CV cysteine lyase, it suggests that in the context of this study (where the vegetables were microwaved, a cooking process known to disrupt tertiary protein structure and inactivate enzymatic activity) the beneficial health effects of SCMSO identified in the literature would still be likely to occur.

The degree of unsaturation of the sulphur atom in MMTSO may confer the ability to affect hepatic phase II enzyme protein expression although this is unclear. The nucleophilic centre of the thiosulphinate may also have the ability to scavenge electrophilic carcinogenic intermediates and inhibit genotoxic initiation [295].

Despite extensive research on the organic sulphur-containing compounds (OSCs) in *allium,* focus has been on whole extracts of garlic and onions and upon the diallyl sulphides [288]. Little research has been performed on SMCSO in isolation but it is thought to contribute to the health benefits of both *alliums* and *brassica*. However, SMCSO is also the only OSC in common between these two groups of vegetables. As *allium* vegetables were restricted in this study, the presence of SMCSO and its *N*-acetylated metabolite can be attributable to CV consumption.

On the basis of the statistically discriminating effect of SMCSO and its *N*-acetylated metabolite in human urine on the consumption of CVs and numerous links to beneficial health effects of this compound in animal models, it should certainly be considered for further study within the context of human *in vivo* and *in vitro* cell studies. The known health benefits of CVs are almost certainly in part due to this plant secondary metabolite. It is surprising to note that in spite of the evidence of anticarcinogenic, antioxidant capacity, antihypercholesterolaemic, hypolipidaemic, antimicrobial and antidiabetic effects of SCMSO, this compound has not been more extensively studied in isolation especially in humans. This is even more remarkable when considering its abundance in both CVs and *alliums* and comparing the extensive number of publications on the glucosinolates and their hydrolysis products and of OSCs in *alliums.*

MMTSO has also been shown to have a suppressive effect on the mutagen aflatoxin B_1 an indirect-acting carcinogen which induces clastogenic and aneugenic changes in mammalian cells [23]. Aflatoxin B_1 chromosome aberrations were observed in rat bone marrow cells and were found to be potently suppressed by MMTSO in the range 1-20mg/kg body weight particularly if injected in a window of two hours before and after treatment with aflatoxin $B_{1.}$ MMTSO is thought to exhibit this anti-genotoxic effect by modulation of enzymatic sulfhydryl groups. This proposed mechanism of action was

identified by a similar effect on aflatoxin B_1 induced chromosome aberrations by diphenyl disulphide [23]; which is known to modify sulphydryl groups in proteins. Juices of both cabbages and onion were also shown to produce the same effect further strengthening the idea that SMCSO is a chemoprotective compound in the human diet [23].

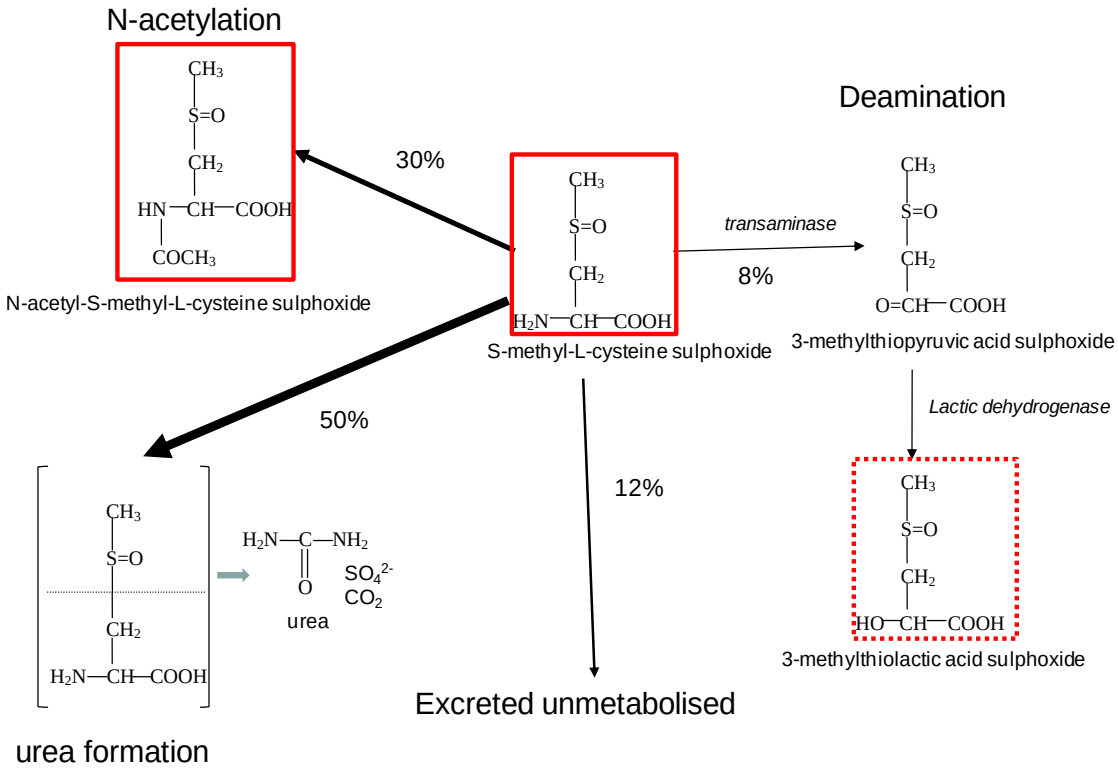
Furthermore, MMTSO isolated from cauliflower were also shown to modify the effect of azoxymethane on rat colon carcinogenesis upon dietary exposure [299]. This effect was observable on biomarkers of cell proliferation in the rat model. Over the course of five weeks, rats were given a 15mg/kg body weight dose of azoxymethane and concurrently fed with 5mg/kg body weight MMTSO. This was shown to reduce the occurrence of aberrant cryptic foci in the rat colon [299]. The biomarkers of colonic mucosal ornithine decarboxylase activity, silver stained nucleolar organiser regions per nucleus in colonic epithelium and polyamine levels in blood were significantly reduced by MMTSO inclusion in the diet of the rats. The incidence of intestinal neoplasms in a longer term study with azoxymethane was also shown to decrease in a dose dependent manner with MMTSO. Further to this study the same investigators co-administered MMTSO with sulindac a non-steroidal anti-inflammatory drug in rats [300]. A synergistic effect of this combination was demonstrated in further suppressing aflatoxin B1 induced intestinal neoplasms in a long term 52 week study. This further strengthens the potential significance of SCMSO and MMTSO in the human diet.

Radiolabelling studies of both SMSC and SMCSO carried out by [276, 277] provided evidence that no demethylation of either SMCS or SMCSO was observed when humans were dosed with known concentrations of either these plant derived amino acids. The primary route of excretion was found to be by cleavage to urea and inorganic sulphate (~50%), followed by *N*-acetylation (~30%), excretion of the unmetabolised SMCSO (~12%), and by deamination to the pyruvic acid derivative and finally 3-methylthiolactic acid sulphoxide (~8%) (**Fig. 6.35**). In the studies presented here, ^{1}H-NMR spectroscopy succeeded in identifying both NAc-SMCSO and unmetabolised SMCSO in human urine. It can be reasonably hypothesised that the remaining unknown NMR singlets detected in this study may be the products of deamination; 3-

methythiopyruvic acid sulphoxide and 3-methylthiolactic acid sulphoxide. Without absolute quantification, the proportional excretion of the metabolites of SMCSO are reflected in the peak heights in the ^{1}H-NMR spectra; the dominant singlet peak at δ2.78 shown to be NAc-SMCSO representing ~30% of SMCSO metabolism and the smaller peak at δ2.83 shown to be SMCSO representing ~12% of excreted SMCSO.

Consumption of naturally occurring SMCSO in CV's may have an effect on the proportions of SMCSO metabolites in human urine compared to ingestion of SMCSO in isolation. The dosing of SMCSO or SMCS in a one-shot dosage by gelatine caps [276, 277], may lead to a rapid absorption and potential saturation of one of the metabolic pathways, the relatively slower release of dietary SMCSO obtained by CV consumption may have altered the proportions of metabolic endpoints and the rates of excretion. This could only be confirmed by quantification of the metabolites in human urine. Whether endogenous mammalian biosynthesis of SMCS and SMCSO occurs is unknown. SMCS and SMCSO are also not known to be recycled into the free cysteine amino acid pool and do not represent a dietary source of cysteine and are purely plant derived metabolites.

Only the *L*-isomer of SMCSO is known to occur in nature, therefore the other two unknown peaks identified are unlikely to represent stereo isomers of SMCSO; these are most likely related metabolites of SMCSO.



based on S.C. Mitchell et. al. Xenobiotica 1984 The metabolism of S-methyl-L-cysteine in man

Fig. 6.35 Metabolic scheme showing the human metabolism of SMCSO.

Radiolabelling studies identified 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 ringed in solid-red were detected in human urine in the present study by ^{1}H-NMR spectroscopy. Product ringed in dashed-red putatively identified by UPLC-MS analysis.

6.6.5. Effect of meat consumption on human urinary metabolic profile.

Subtle variations in metabotype in human subjects following either a high meat, low meat or vegetarian diet have been previously described [221]. Of particular interest are the metabolites, which were shown to be associated with a high meat diet. Many of these same metabolites were also observed in this study following consumption of a high-meat content meal. The subjects in the present study abstained from meat for at least 24 hrs prior to the meat meal on day 13 and post-prandial biomarkers of meat consumption were readily identified. Alteration of urinary metabolic phenotype therefore occurred soon after the meat meal. They included increased urinary excretion of creatine, creatinine, carnitine, taurine, *N*-acetyl-5-hydroxytryptamine, 1- 3- methylhistidine, acetylcarnitine, carnosine and several other amino acids; glycine, glutamine, alanine, threonine and histidine. These meat-derived biomarkers declined in the subsequent 10-24 hours urine collection. Most of these metabolites including many non-essential amino acids and some essential such as histidine and threonine are found in high abundance in meat and derive from a variety of metabolic pathways; for example, *N*-acetyl-5-hydroxytryptamine produced during tryptophan metabolism. This is likely to be representative of the digestive transit of meat from the higher duodenum and small intestine.

Ten to twenty four hours following meat consumption, urinary excretion of TMAO, *p*-cresol sulphate, phenylacetylglutamine and to a lesser extent hippurate increased. These are all known to be metabolites derived from the human gut microbiota. *p*-Cresol is formed from tyrosine *via* transamination to p-hydroxyphenylacetate by P.*vulgaris* then decarboxylation by *C.difficile* and *lactobacillus* species [132]. High proportions of anaerobic gut microbes such as C.*difficile* and a depletion of P.*vulgaris* can indicate poor diet and can lead to a build up of *p*-hydroxyphenylacetate. The increase in urinary concentration of bacterial associated metabolites 10-24 hours post meat consumption is therefore consistent with gut physiology. The increase in bacterial co-metabolites seen in the 10-24 hour urine collection is reflective of tryptophan transamination, glycine benzoate conjugation (hippurate) and glutamine acetylation (phenylacetylglutamine) further along the digestive transit in the lumen of the lower small intestine and large intestine. The 24-48 hour urine collection shows a depletion of these phenolic bacterial

co-metabolites. It has been suggested that the excretion of these metabolites serves as an alternative to urea for removal of waste nitrogen [136].

In the 10-24 hours urine collection the concentration of TMAO is increased. TMAO is known to increase following meat and fish consumption. However TMAO is also derived from enteric bacteria from the parent molecule L-carnitine [301]. The increased concentration of TMAO in the 10-24 hours post meat consumption urine suggests that it originates from gut microbes especially as it is associated with other microbial co-metabolites seen in the O-PLS-DA coefficients plot. The temporal dependence of the appearance of these metabolites after meat consumption appears to demonstrate the series of metabolic events taking place when a meat meal passes through the digestive tract. High abundance components of meat are absorbed and excess is eliminated from the body by endogenous metabolism. Then after 10-24 hours bacterial co-metabolites derived from amino acid metabolism are observed arising from the lower small intestine and large intestine.

Hippurate, phenylacetylglutamine and TMAO, all appear to be associated by time-course O-PLS-DA analysis with 10-24 hours post meat consumption. STOCSY analysis of phenylacetylglutamine not only reveals a correlation with self but also with p-cresol sulphate and strongly with three unknown doublets at δ 7.69, δ 7.48 and δ 7.95 and to a lesser extent *p*-hydroxyphenylacetate and hippurate. This lends weight to the idea that these phenolic metabolites originate from particular species of the human gut microbiota.

6.6.6. Effect of cruciferous vegetable dietary intervention on meat and caffeine alterations of metabotype.

A most perplexing finding distinguishing post meat meal urine sample classes was the presence of mannitol. Mannitol is an osmotic diuretic and weak renal vasodilator. It is a stereoisomer of sorbitol and is an exogenous metabolite in humans, and its presence in the urine of these subjects is quite unusual. Mannitol is the most highly abundant sugar alcohol or polyol found in nature and is found in a wide variety of commonly consumed fruits and vegetables. Mannitol is primarily used as an artificial sweetener but also in the pharmaceutical industry as filler and to aid drug delivery across the blood-brain barrier. It is manufactured commercially by industrial hydrogenation or by fermentation of fructose, glucose or maltose by *Lactobacillus intermedius*. The meat meal was prepared using only high quality pure beef steak without any 'fillers' such as mannitol which is commonly found as an additive in cheap meat patties. The subjects were also not provided with any artificial sweeteners or sugar with their coffee.

Lactobacillus intermedius is not known to be part of the human gut microbiota, also none of this bacteria's substrates (for fermentation of mannitol) were present in the meat meal or the coffee given to the study subjects. The presence of mannitol in the urine is an enigma after discounting all potential sources from the study

The fact that mannitol appears to be present at all time points suggests that it was deliberately introduced perhaps as a saliva inducer for the purposes of the caffeine kinetics analysis in the original study [81], in chewing gum or perhaps as a coffee sweetener for example. As the excretory half life of mannitol is only 100 minutes, it appears that mannitol was administered throughout the duration of the study. It could suggest that its presence is a spurious contaminant arising from the study. Mannitol is almost completely unmetabolised in humans, it has a bioavailability of only 7% and hepatic metabolism is negligible [302]. Consumption of mannitol in sufficient quantity is known to produce a laxative effect by drawing water into the large intestine by osmosis. Mannitol is commonly used in clinical medicine as an indicator of renal function with a 90% renal excretion. O-PLS-DA revealed an increased urinary excretion of mannitol in

phase I compared to both phase II and phase III. This indicates an unexplained effect on mannitol excretion following CV consumption.

Several heterofermentative lactic acid bacteria (LAB) belonging to the genera *Lactobacillus* are known to produce mannitol from fructose, some of which are known to be human gut microbes. Mannitol is produced by the enzyme mannitol-2-dehydrogenase (MDH) which catalyses the pyridine nucleotide dependent reduction of fructose. In particular the enzyme MDH has been purified and characterized from *L. Brevis* [303], *L. Sanfranciscensis* [304], *Leuconostoc mesenteroides* [305], and *Lc. Pseudomesenteroides* [306].

STOCSY analysis of mannitol reveals correlations with other sugars such as glucose in the spectra. This suggests mannitol is perhaps less likely to be a contaminant as it correlates with other endogenous sugar metabolites.

Interestingly outliers in chemometric models were found to be due to the characteristic ^{1}H-NMR splitting pattern of ethanol in pre-dose urine samples. These outlying samples were removed to prevent poor modelling of the data. Study subjects were requested to abstain from alcohol during the study period however this finding suggests that study subjects did not adhere to dietary restrictions. This could perhaps shed light on the presence of mannitol as a contaminant.

6.6.4 UPLC Mass Spectral Metabonomic analysis

UPLC-MS as a metabonomic analytical platform succeeded in detecting 19 statistically significant mass spectral features by *m/z* and retention time in positive mode, 11 significant features in negative mode by use of the XCMS[1] peak picking programme. Targeted analysis also succeeded in semi-quantifying a further 6 potential biomarkers in positive mode but failed to do so within the negative mode dataset. Identification of tentative biomarkers was attempted by online database searches which provided a wealth of potential candidates many of which were biologically plausible and a self-generated list of potential metabolites based on metabolites of CVs, caffeine, meat and metabolites of the red-meat carcinogens. Authentic standards were subjected to MS/MS fragmentation analysis in positive mode (34) and in negative mode (31) and also further compounds of interest (17) at a range of ionisation voltages and compared to MS/MS fragmentation patterns of the unknown spectral features in human urine. Unfortunately this failed to fully confirm identity of the mass spectral biomarkers or to suggest any significant commonality in fragmentation pattern.

Similarity of the tentative biomarker fragmentation patterns were also compared to experimental fragmentation patterns on both the Metlin (using the XCMS[2] programme) and the HMDB. A metabolite of SMCSO; 3-methylthiolactic acid sulphoxide was a potential biomarker within the negative mode dataset, diagnostic of CV consumption. However in the absence of an authentic standard this was unconfirmed. In the absence of absolute confirmation of tentative biomarkers, the pattern of mass spectral features can be seen as 'finger-print' of changes associated with both CV consumption and red-meat/caffeine dietary intervention.

The use of UPLC-MS data for global metabolic profiling has only been applied relatively recently and though perhaps less advanced than ^{1}H-NMR spectroscopy, it has significant advantages and has met with success, notably within rat urine and plasma metabonomic datasets [307, 308] [150]. UPLC-MS utilizes high pressures combined with small capillary particle size for increased chromatographic resolution and boasts significant advantages; rapid screening, relatively high resolution and high throughput. However

UPLC-MS for metabonomic analysis suffers from some disadvantages; chromatographic and mass spectral drift, ion enhancement/suppression and reproducibility compared to NMR data.

A significant challenge of interpretation of UPLC-MS data is its sheer size which makes manual analysis of raw data almost impossible. The use of chemometric techniques particularly in the form of software packages such as XCMS[1] [157] and from instrumentation manufacturers (eg. Masslynx/MarkerLynx Waters) have proven useful for the discovery of biomarkers. HPLC-MS analysis was used recently to identify the neutral loss of 129 m/z mercapturates from acetaminophen in human urine [272] and holds promise for future human metabonomic studies.

Recently new methods of UPLC-MS data treatment have been introduced which advocate the use of quality control samples placed within the 96-well plates of high throughput sample sets. The quality control sample would consist of a pooled urine sample from samples across the study to broadly reflect the different urinary components of participants throughout the study. These samples can then be treated as a quality control sample set that can be submitted to PCA analysis and further scrutiny in an effort to reduce any problem associated with reproducibility, chromatographic and spectral drift in large datasets. The run of quality control samples can be checked for any irregularity in retention time of peaks and intensities from the start of the run to the finish and the ion intensities for individual peaks can also be compared to ensure that analysis has performed properly throughout the run and validate the results obtained from the experimental dataset.
Unfortunately when the CV urinary sample set was submitted to high-throughput UPLC-MS analysis this methodology was not implemented. The data analysis would most certainly have been improved by using this method and would have improved result assessment and validity [309] .

Data-mining programs such as XCMS[1] use special algorithms that filter and bin raw mass spectral data then extract chromatographic features identified by *m/z* ratio and retention time. The features are then aligned, normalized and a peak table created,

which can be subjected to further statistical analysis. XCMS uses a peak picking algorithm that identifies chromatographic features that represent the differences between sample classes and produces a list of Extracted Ion Chromatograms (EICs) and box and whisker plots and sorted in ascending order of confidence based on *p*-value and other statistical measures. However the XCMS[1] can fail to work when peak alignment is poor due to spectral drift, this occurred within the positive mode dataset when examining the effect of CVs on meat and caffeine alteration of metabotype and failed to obtain adequate peak alignment. This could also be due to essentially no detectable difference between the Phase II 0-10hrs timepoint and the Phase I and III 0-10hrs timepoints other than the markers identified in the predose urine samples. The tentative biomarkers of CV consumption were selected based on their class exclusivity within the mass spectral data and by interpretation of the EIC, box and whisker plot, *p*-value and presence within the raw data. This relies on a degree of subjectivity and selection of significance, alignment and peak height within the EIC and exclusivity with the box and whisker plot. The purpose of this process was to identify ions of interest on which the investigations could focus attention as a possible source of biomarker information.

Initial biomarker identification was based on knowledge of potential exogenous and endogenous candidate metabolites, database mining (ie. Human Metabolome Database, Hmdb, *Genome Alberta*) and fragmentation analysis. A further technique that could be attempted is use of a high resolution orbitrap mass spectrometer to obtain accurate mass atomic formulae of the unknown biomarkers in urine. Final confirmation is achieved using authentic standards but fragmentation of standard compounds failed to identify any of the unknown urinary biomarkers. The lack of absolute identification of unknown metabolites obtained from MS leads was disappointing. However identification was confounded by the number of standard compounds available within the scope of this thesis and the limitations of slowly developing metabonomic databases and is an ongoing challenge.

6.7 Conclusions

In conclusion this human dietary intervention study succeeded in identifying 4 distinct urinary biomarkers of CV dietary intervention by ^{1}H-NMR spectroscopy and a total of 21 statistically significant urinary biomarkers and 2 putative biomarkers using a targeted approach by UPLC-MS analysis of both positive and negative mode datasets utilising various chemometric modelling techniques.

^{1}H-NMR spectroscopy combined with the O-PLS-DA modelling proved particularly successful in identifying four distinct biomarkers of CV consumption.
A combination of ^{1}H-^{1}H *J*-Res, COSY, ^{13}C-^{1}H HSQC, ^{13}C-^{1}H HMBC and TOCSY two dimensional NMR spectroscopy was indicative of two biomarkers of CV consumption; SMCSO and NAc-SMCSO. The vegetable origin of this biomarker of CV consumption was confirmed by application of the same two dimensional NMR experiments to aqueous extracts of CVs. The identification of *N*-Acetylated-SMCSO was as the most significant biomarker of CV vegetable consumption, however due to contamination of the synthetic reaction this was highly indicative of the unknown singlet peak 2.78 in human urine but not truly conclusive. This problem could have been surmounted by the purchase of a custom synthetic standard compound of NAc-SMCSO, but this was not financially possible. The two remaining biomarkers of CV consumption could reasonably be hypothesised to be the thiopyruvate and thiolactate of SMCSO derived from deamination, although this was not confirmed.

Targeted biomarker identification and the XCMS[1] programme were successful in identifying 2 potential targeted biomarkers and 21 statistically significant tentative biomarkers of CV consumption in both the positive and negative mode UPLC-MS datasets. Problems arose in identification of these unknowns however, despite rigorous online database searches and the comparison of tandem MS/MS fragmentation patterns to that of promising biological candidate compounds. The availability and expense of standard compounds and the development of metabonomic mass spectral databases was highlighted as problematic to absolute identification of these unknowns.

The effect of red meat dietary intervention was demonstrated by ^{1}H-NMR spectroscopy combined with supervised multivariate analyses. This corresponded to what appeared to be a digestive transit of the meat meal over the 48 hour urinary time course. These included post-prandial biomarkers of meat consumption such as excess amino acids with evidence of increased concentrations of bacterial co-metabolites which are associated with nitrogen removal in the gut. UPLC-MS analysis produced 9 statistically significant urinary biomarkers of meat consumption and 4 targeted approach putative biomarkers which bore expected molecular ion similarity to known metabolites of caffeine, however absolute identification proved problematic.

The effect of CVs on meat and caffeine alterations metabolism was not satisfactorily addressed as the presence of mannitol a discriminatory of the post meal/ caffeine was entirely unanticipated and difficult to account for. Furthermore due to peak-picking difficulties in the XCMS[1] programme only one biomarker addressing this experimental aim was discovered in the negative mode dataset. This feature had the expected *m/z* of the parent ion of deaminated SMCSO; 3-methylthiolactic acid however this could not be confirmed due to the lack of an authentic standard.

Chapter Seven: Overall Conclusions and Future Work

The purpose of this thesis was to apply analytical techniques, including high-throughput ^{1}H-NMR spectroscopy and UPLC-MS of biofluids coupled to multivariate statistical modelling, to examine the effects of nutritional intervention in three different experimental studies. Initially a study in experimental animals (rats) examined the effect of a dietary genotoxic carcinogen methyl eugenol on the metabolic phenotype. This was followed by a human dietary study examining the effect on metabolic phenotype of changing dietary fat composition and carbohydrate GI in subjects at risk of developing the metabolic syndrome; a condition considered a precursor to cardiovascular disease (CVD), type II diabetes and increased morbidity. Finally a further human study to determine the effects of CVs, meat meal and caffeine dietary interventions on the human metabolic phenotype, this represented a study examining dietary components reputed to be beneficial to human health.

7.1 The application of metabonomics to examine the effects of ME dietary intervention in rats

As a preliminary metabonomic study the effect of ME a known genotoxic carcinogen on the rat metabolic phenotype was investigated. This compound is of significant public health interest as it is widely consumed in many fruits, vegetables and spices and is used extensively as an additive in many foods, cosmetics and household items but has been shown to cause hepatic and gastric carcinogenesis in experimental animals. The study described in this thesis examined doses in F344 strain rats which were designed to more realistically reflect estimated human exposure levels (~1mg/kg bw/day), using doses in the range 1-50mg/kg bw/day in a 28 day dietary regimen; previous laboratory animal studies had used doses far in excess of this.

The primary aim of this study was to assess any effect of methyl eugenol on the global rat metabolic phenotypes. It was hypothesised that any disruptions of normal homeostatic control in the rats suggesting a toxic, genotoxic or carcinogenic effect would be reflected in the urinary metabolite profile. The findings of this study confirmed that a dosing regimen of ME of human relevance over twenty-eight days with weekly sample collection did not produce any dramatic effect on the rat metabolic

phenotype. However there were subtle differences in urinary metabolic profile in the high dose group rats which were attributed to a putative urinary metabolite of ME, perhaps the sulphate or glucuronic acid conjugate of *O*-demethylated ME. The low urinary concentration of these discriminatory peaks in rats exposed to 50mg/kg bw/day made further structural elucidation by two-dimensional NMR experimentation problematic. However, attempts could be made in the future to concentrate urine samples from rats in this dosage group and combined with liquid chromatographic techniques and solid phase extraction may help to absolutely identify this putative urinary biomarker of ME exposure.

There is considerable interest in the accurate biomonitoring of ME exposure in experimental animal toxicology studies and in human populations, in environmental toxicology and public health epidemiological studies. Good urinary biomarkers of ME could potentially be useful in high-throughput metabonomic studies of human populations. Average human exposure levels to ME (~1mg/kg bw/day), are likely to be below the detection limits of high-throughput urinary ^{1}H-NMR spectroscopic analysis. The current study supports this position since the urinary concentration of potential biomarkers of ME in the rat urine samples was very low and the overlapping of constitutive metabolites in higher concentration confounded the data. However the use of targeted UPLC-MS analysis, which has higher resolution and sensitivity than ^{1}H-NMR spectroscopy may allow accurate identification of urinary biomarkers/metabolites of ME.

As proof of principle, the application of metabonomic multivariate modelling of rat urinary metabotype successfully identified metabolic differences related to gender, age and a stress-response related to the use of the orogastric gavage. Clear evidence of sexual dimorphism in the urinary metabolic profile was identified in the F344 strain rats. Also evidence of age-related metabolic change in endogenous and bacterial co-metabolites suggested increases in body weight, muscle mass, development of gut microbial populations and post-weaning adjustment of metabolism to an adult diet. Interestingly metabonomic approaches identified metabolic changes related to an induced acute stress-response in the experimental animals attributable to the orogastric gavage technique commonly used in toxicology studies. These findings demonstrated

the robustness of metabonomics for modelling changes in urinary metabolic phenotypes. However it also highlighted the potential influence of gender, age and method of administration in confounding results in dietary intervention studies. This has important implications in limiting inter-subject and experimental variation in animal studies despite using a relatively homogeneous population that is genetically and environmentally identical (F344). This must be considered when applying metabonomic methods to human dietary interventions where inter-subject and environmental variations are potentially more confounding.

7.2 Application of the metabonomic approach to understanding human dietary intervention with varying composition of dietary fat and glycaemic index of carbohydrates in subjects at risk of developing metabolic syndrome

The metabonomic strategies successfully applied in the preliminary animal study were utilised to examine a human dietary intervention urinary dataset. This study concerned a significant issue in public health, specifically the effects of human dietary intervention on subjects at risk of developing metabolic syndrome, a condition characterised by obesity, glucose intolerance, dyslipidaemia and hypertension. Also of interest was the influence of the gut microbiota on the development of this syndrome. Subjects diagnosed with MS are considered at increased risk of developing subsequent associated aetiologies of CVD, type II diabetes, cancer and increased morbidity and mortality.

This study analysed urine and faecal water samples from a cohort of subjects (n=130) at risk of developing the MS. Two urine samples were collected from each subject; a 0 week following a 4 week “typical western” dietary regimen run-in and a 28 weeks collection. Subjects were divided into 5 dietary intervention groups; a control diet and 4 other dietary groups where the composition and proportion of dietary fats and GI of carbohydrates had been altered.

Application of unsupervised and supervised multivariate statistical analyses to this human dataset failed to identify any detectable effect of this dietary intervention on the human metabolic phenotype. This was reflected in poor modelling strengths of

chemometric models and a dominance of inter-subject variation in the dataset. No variation attributable to the effect of altering dietary fat composition and carbohydrate in human study subjects or discriminated between the different dietary classes was found. Collaborators at the University of Reading identified subtle changes in bacterial population numbers in the study subjects using FISH analysis. These bacterial population findings were regressed to the both the urinary and faecal water ^{1}H-NMR spectroscopy datasets in an attempt to identify metabolic correlation particularly in the bacterial co-metabolites with the bacterial population findings. Unfortunately, this regression analysis also proved unsuccessful as only subtle variations were seen between the dietary classes in the bacterial numbers enumerated by FISH and this was confirmed by the spectral data regression analyses.

In spite of attempts to improve data pre-processing, peak alignment and removal of outliers from the datasets no effect of dietary intervention was observed in this human study.

This study was unsuccessful in terms of elucidating the effects of changing dietary fat and carbohydrate GI on human metabotype. However, the results of this study highlighted the inherent inter-subject variability in human populations and the subsequent confounding of metabonomic modelling of subtle dietary interventions. The subjects in this study were a particularly heterogeneous cohort. Both male and female subjects were recruited, with multi-racial backgrounds (although predominantly Asians) and variable age. Furthermore, the diagnostic criteria for recruitment were based on subjects at risk of developing MS and not a clinical diagnosis of the syndrome. The criteria consisted of a scoring system based on measures of blood pressure, glucose intolerance, central adipodisity, BMI and dyslipidaemia, family histories of type II diabetes, polycystic ovary syndrome, gestational diabetes and artherogenic lipid profiles. Dietary compliance was monitored by use of 4-week food diaries, therefore subjects may not have adhered to the dietary restrictions or recommended foodstuffs over the course of a 24 week dietary intervention. For the purpose of metabonomic analysis the experimental, environmental and inter-subject variability introduced particularly by the recruitment criteria proved problematic for robust modelling of the subtle dietary intervention effects.

7.3 Metabonomic analysis of the effects of CV dietary intervention on the human metabolic phenotype

It was hoped that more rigorous control of inter-subject, environmental and experimental variation in human study subjects, which proved problematic in the previous human study, would enable the effects of dietary intervention to be observed in the metabolic profile. This human study examined the effect of CVs, known to be beneficial to human health and inversely associated with many forms of cancer, on the human metabolic profile. This study also examined the metabolic consequences of meat consumption and attempted to address any effect of CVs on alterations of meat and caffeine metabolism.

In this study inter-subject variation was limited by recruiting male, Caucasian, non-smoking individuals, of comparable age and also by restriction of prescription drugs, alcohol and other foodstuffs believed to possess strong biological properties such as *alliums*. In this instance, the metabonomic approach proved successful when applied to this human nutritional intervention data set. ^{1}H-NMR spectroscopy combined with the O-PLS-DA chemometric technique proved particularly successful in identifying four distinct biomarkers of CV consumption and characterising time-course effects on metabolic phenotype following consumption of a red meat meal. The use of improved data pre-processing techniques RSPA and normalisation to the probabilistic quotient and a combination of ^{1}H-^{1}H *J*-Res, COSY, ^{13}C-^{1}H HSQC, ^{13}C-^{1}H HMBC and TOCSY two dimensional NMR spectroscopy confirmed the identity of two biomarkers of CV consumption namely SMCSO and NAc-SMCSO. Aqueous extracts of CVs demonstrated the dietary origin of SMCSO. The *N*-acetylated derivative of SMCSO, a common route of amino acid metabolism was also shown to be present in human urine as the most significant biomarker of CV vegetable consumption. This identification was not definitive, but was strongly indicative. The presence of NAc-SMCSO as the dominant biomarker of CV consumption is in accordance with the known excretory pathways of SMCSO; the two remaining biomarkers of CV consumption were hypothesised to be the thiopyruvate and thiolactate of SMCSO derived from deamination. Although not confirmed, such metabolites are predicted based upon this study and previous reported studies [276, 277].

The identification of NAc-SMCSO and unmetabolised SMCSO within human urine as biomarkers of CV consumption has potential utility in targetted epidemiological studies of diet, and perhaps cancer risk and population metabonomics. It could also be conceivably 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. The findings of this study should provide impetus for further study of SMCSO particularly within humans. The bulk of research on CVs has focused upon the glucosinolates and their hydrolysis products however SMCSO is reputed to be partly responsible for the health benefits of CVs [295].

The effect of consumption of a red meat meal was clearly elucidated by ^{1}H-NMR spectroscopy revealing an apparent digestive transit of the meat meal over the 48 hour time course with post-prandial biomarkers of meat consumption, excretion of excess amino acids, followed by evidence of bacterial co-metabolites associated with nitrogen removal in the gut. Each of these phases of digestive transit was identifiable in the time-dependent urine samples. The effect of CVs on meat and caffeine metabolism was not satisfactorily addressed due in part to the presence of mannitol, which was entirely unanticipated and difficult to account for.

The use of UPLC-MS global analysis in metabonomics is a relatively recent analytical platform when compared to ^{1}H-NMR spectroscopy, particularly in human studies. However the use of targeted biomarker identification and the XCMS[1] programme was successful in identifying 6 potential targeted biomarkers, 19 statistically significant tentative biomarkers in the positive mode and 11 tentative biomarkers in the negative mode datasets. The sensitivity of UPLC-MS allows identification of considerable numbers of class discriminating mass spectral features however significant problems remain in absolute identification of unknowns. The biomarkers identified by a mixture of targeted and statistical approaches can be considered a ‘fingerprint’ of metabolic changes that characterise consumption of CVs, red meat and caffeine. In spite of identifying promising candidates based upon biological expectation and online database searches, none of the biomarkers were chemically confirmed. The ongoing development of metabonomic mass spectral databases will be a key limiting factor in the discovery process.

7.4 The future potential of metabonomics for nutrition studies

There are currently concerted efforts being made to create standardised methods and databases focused on nutritional metabolite biomarker discovery and nutrigenomics [310]. The limiting factor in utilisation of metabonomic technologies in biomarker discovery is the interpretation of datasets which can contain 2000+ metabolites [311] and the cataloguing and standardisation of methods between research groups is of the utmost importance particularly in the case of UPLC-MS analysis [312]. These particular challenges were highlighted by the difficulty encountered in identifying ^{1}H-NMR spectral biomarkers and even more so tentative biomarkers elucidated by UPLC-MS analysis.

The use of stable non-invasive, robustly validated biomarkers of nutritional interventions particularly of foodstuffs of significant public health interest could be of enormous utility to the study of nutrition and its contribution to beneficial/detrimental health outcomes. The ability to gauge actual exposure of study participants to dietary components holds great future potential within human epidemiologic studies, to quantify absorbtion and exposure levels of dietary components in *in vivo* animal studies and to ensure dietary compliance in human nutritional intervention studies. The use of a robust stable urinary/serum biomarker has potential to improve or perhaps even negate the reliance upon inferior methods of dietary analysis such as food diaries, 24 hour diary recalls and food frequency questionnaires which are prone to inaccuracies [84-55].

There are many studies in the scientific canon regarding the purported health benefits of components of the diet, particularly in the case of vegetables and fruits which are the subject of enormous research focus [2,3,33-35]. However, the findings of these studies are potentially confounded by the difficulties and inaccuracies associated with dietary self-reporting. The use of non-invasive biomarkers in biofluids which either predict or influences health, the exposure to dietary components or can influence the incidence or behaviour of a disease of dietary interventions could be used to improve the accuracy and validation of current dietary self-reporting methods. The use of dietary intervention biomarkers could be used in future studies and also in theory be performed post-humously, the quantification of exposure levels to beneficial foodstuffs could lend

added weight to their findings and give an enhanced interpretation of true physiological effects of a dietary intervention [313].

In future studies examining CVs an ideal study design would initially analyse the concentration of SMCSO in aqueous extracts of the CVs which are to be consumed, followed by subsequent blood serum and urine analysis from serially collected samples. This would provide a quantifiable account of the LADME of CVs and potentially *alliums* in human study participants or equally in animal studies.

Currently, the biomarkers of CV consumption (SMCSO and NAc-SMCSO) discovered in the course of this thesis is being validated in the INTERMAP (INTERnational collaborative study of MAcronutrients, micronutrients and blood Pressure) epidemiology study [314]. INTERMAP is a multi-centre cross-sectional epidemiologic investigation designed to help clarify unanswered questions regarding the role of dietary factors in the development of unfavourable blood pressure (BP) levels in adults. Preliminary exploration of this large urinary dataset (n=4860) of individuals from the UK, US, Japan and China. Based upon 24hour dietary recalls UK participants consuming CVs/*alliums* and non-consumers have been discriminated robustly by the excretion of SMCSO and NAc-SMCSO in all species of the *cruciferae* and *alliums* examined. This investigation will be further extended to all populations and by comparison of reported weight and type of vegetables of the *cruciferae* and *alliums* consumed to the excretory levels of these biomarkers in urine, to address accurate levels of LADME and an improved picture of the pharmacokinetics of bioactive food components. Quantified urinary concentrations of these biomarkers will also be regressed to systolic and diastolic blood pressure measurements, CVD and types II diabetes disease status and BMI.

7.5 Summary

In summary, metabonomic approaches proved largely efficacious in the characterisation of nutritional intervention on the metabolic phenotype in both human and experimental animal studies. This study successfully identified ^{1}H-NMR spectral biomarkers of both CV and meat consumption in human urine, tentative evidence of a urinary biomarker of ME exposure at a low dosage regimen (1-50mg/kg bw/day) and a discrimination of age, sex and dietary administration method in rat urinary metabolic profiles. In the preliminary human study, the limitation of experimental and intersubject variations was demonstrated to be of great importance in the robust modelling of metabonomic datasets, especially in human nutritional intervention studies where great environmental and genetic variations exist and therefore appropriate design of experiments is of prime concern.

UPLC-MS analysis proved successful in the identification of biomarkers (36) of both CV and meat meal dietary interventions in human urine. The superior resolution and sensitivity of this analytical platform compared to ^{1}H-NMR spectroscopy offers great potential for the identification of low molecular weight biomarkers in low concentration in human and experimental animal biofluids. Further statistical analyses may have been attempted to enhance interpretation of the mass spectral data such as statistical heterospectroscopy (SHY) correlation. SHY correlation coanalyses NMR and MS spectral data based upon intrinsic covariance correlations and can provide additional biological or structural information on unknown spectral features [315]. However, the absolute chemical identification of UPLC-MS biomarkers of nutritional intervention proved problematic, highlighting the ongoing development of mass spectral databases as a limiting factor in the discovery process. Further mass spectral fragmentation analysis would be necessary for absolute chemical identification of the unknown biomarkers identified.

The analysis of metabolic consequences of exposure to dietary components is of great public health interest to the understanding of the complex relationship between diet and disease, particularly the identification of robust biomarkers of exposure and response in human and experimental animal biofluids. Biomarkers of nutritional intervention are of interest as they allow the biomonitoring of dietary interventions, absorption and exposure levels, the interrogation of epidemiological datasets of disease risk and outcomes and the identification of mechanisms of the beneficial/detrimental effects of components of the diet.

References

1. Ames BN, Gold LS, Willett WC. The causes and prevention of cancer. Proc Natl Acad Sci U S A 1995;92(12): 5258-65.

2. Ness AR, Powles JW. Fruit and vegetables, and cardiovascular disease: a review. International Journal of Epidemiology 1997;26(1): 1-13.

3. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: A review of the epidemiological evidence. Nutrition and Cancer 1992;18(1): 1-29.

4. Gonzalez CA. The European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition 2006;9(1A): 124-26.

5. Waxman A. Prevention of chronic diseases: WHO global strategy on diet, physical activity and health. Food Nutr Bull 2003;24(3): 281-4.

6. Tonkin A. The metabolic syndrome a growing problem. European Heart Journal Supplements 2004;6(suppl A): A37-A42.

7. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabetic Medicine 2006;23(5): 469-80.

8. Gardner CD, Kraemer HC. Monounsaturated Versus Polyunsaturated Dietary Fat and Serum Lipids : A Meta-analysis. Arteriosclerosis, Thrombosis, and Vascular Biology 1995;15(11): 1917-27.

9. Laaksonen D, Niskanen L, Lakka HM, Lakka T, Uusitupa M. Epidemiology and treatment of the metabolic syndrome. Annals of Medicine 2004;36(5): 332-46.

10. Frost PH, Davis BR, Burlando AJ, et al. Serum Lipids and Incidence of Coronary Heart Disease: Findings From the Systolic Hypertension in the Elderly Program (SHEP). Circulation 1996;94(10): 2381-88.

11. Frost G, Leeds AA, Doré CJ, Madeiros S, Brading S, Dornhorst A. Glycaemic index as a determinant of serum HDL-cholesterol concentration. The Lancet 1999;353(9158): 1045-48.

12. Bostick RM. Nutrition and colon cancer prevention. Nestle Nutr.Workshop Ser.Clin.Perform.Programme 2000;4(Journal Article): 67-85; discussion 85-6.

13. Alabaster O, Tang Z, Shivapurkar N. Dietary fiber and the chemopreventive modelation of colon carcinogenesis. Mutation research 1996;350(1): 185-97.

14. Davis CD, Hord NG. Nutritional "omics" technologies for elucidating the role(s) of bioactive food components in colon cancer prevention. The Journal of nutrition 2005;135(11): 2694-97.

15. Butler LM, Sinha R, Millikan RC, et al. Heterocyclic amines, meat intake, and association with colon cancer in a population-based study. American Journal of Epidemiology 2003;157(5): 434-45.

16. Traka M, Gasper AV, Melchini A, et al. Broccoli consumption interacts with GSTM1 to perturb oncogenic signalling pathways in the prostate. PLoS ONE 2008;3(7): e2568.

17. van Poppel G, Verhoeven DT, Verhagen H, Goldbohm RA. Brassica vegetables and cancer prevention. Epidemiology and mechanisms. Advances in Experimental Medicine and Biology 1999;472(Journal Article): 159-68.

18. Feldeisen SE, Tucker KL. Nutritional strategies in the prevention and treatment of metabolic syndrome. Applied Physiology, Nutrition, and Metabolism 2007;32: 46-60.

19. Lutsey PL, Steffen LM, Stevens J. Dietary Intake and the Development of the Metabolic Syndrome: The Atherosclerosis Risk in Communities Study. Circulation 2008;117(6): 754-61.

20. Ames BN, Gold LS.

21. Gooderham NJ, Murray S, Rice JC, Boobis AR, Davies DS. Dietary heterocyclic amines and human cancer. Drug metabolism reviews; Drug metabolism reviews 1989;20(2-4): 285-305.

22. Rietjens IMCM, Boersma MG, van der Woude H, Jeurissen SMF, Schutte ME, Alink GM. Flavonoids and alkenylbenzenes: Mechanisms of mutagenic action and carcinogenic risk. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2005;574(1-2): 124-38.

23. Ito Y, Nakamura Y, Nakamura Y. Suppression of aflatoxin B1- or methyl methanesulfonate-induced chromosome aberrations in rat bone marrow cells after treatment with S-methyl methanethiosulfonate. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 1997;393(3): 307-16.

24. Goldman R, Shields PG. Food Mutagens. Journal of Nutrition 2003;133(3): 965S-73.

25. Peto J. Cancer epidemiology in the last century and the next decade. Nature 2001;411(6835): 390-95.

26. Boffetta P, Couto E, Wichmann J, et al. Fruit and Vegetable Intake and Overall Cancer Risk in the European Prospective Investigation Into Cancer and Nutrition (EPIC). Journal of the National Cancer Institute 2010;102(8): 529-37.

27. Bingham SA, Day NE, Luben R, et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. The Lancet 2003;361(9368): 1496-501.

28. Fuchs CS, Giovannucci EL, Colditz GA, et al. Dietary Fiber and the Risk of Colorectal Cancer and Adenoma in Women. New England Journal of Medicine 1999;340(3): 169-76.

29. Terry P, Giovannucci E, Michels KB, et al. Fruit, Vegetables, Dietary Fiber, and Risk of Colorectal Cancer. Journal of the National Cancer Institute 2001;93(7): 525-33.

30. Bingham SA. Mechanisms and experimental and epidemiological evidence relating dietary fibre (non-starch polysaccharides) and starch to protection against large bowel cancer. Proceedings of the Nutrition Society 1990;49(02): 153-71.

31. Boffa LC, Lupton JR, Mariani MR, et al. Modulation of Colonic Epithelial Cell Proliferation, Histone Acetylation, and Luminal Short Chain Fatty Acids by Variation of Dietary Fiber (Wheat Bran) in Rats. Cancer Research 1992;52(21): 5906-12.

32. Miller AB, Altenburg H-P, Bueno-de-Mesquita B, et al. Fruits and vegetables and lung cancer: Findings from the European prospective investigation into cancer and nutrition. International Journal of Cancer 2004;108(2): 269-76.

33. Boeing H, Dietrich T, Hoffmann K, et al. Intake of fruits and vegetables and risk of cancer of the upper aero-digestive tract: the prospective EPIC-study. Cancer Causes and Control 2006;17(7): 957-69-69.

34. Key TJ, Allen N, Appleby P, et al. Fruits and vegetables and prostate cancer: No association among 1,104 cases in a prospective study of 130,544 men in the European Prospective Investigation into Cancer and Nutrition (EPIC). International Journal of Cancer 2004;109(1): 119-24.

35. Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. The American Journal of Clinical Nutrition 2005;82(3): 675-84.

36. Schulz M, Lahmann PH, Boeing H, et al. Fruit and Vegetable Consumption and Risk of Epithelial Ovarian Cancer: The European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiology Biomarkers & Prevention 2005;14(11): 2531-35.

37. Freedman ND, Park Y, Subar AF, et al. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. International Journal of Cancer 2008;122(10): 2330-36.

38. Linseisen J, Rohrmann S, Miller AB, et al. Fruit and vegetable consumption and lung cancer risk: Updated information from the European Prospective Investigation into Cancer and Nutrition (EPIC). International Journal of Cancer 2007;121(5): 1103-14.

39. Liu RH. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. American Journal of Clinical Nutrition 2003;78(3): 517S-20.

40. Hajime O, Akira M. Cancer prevention with food factors: Alone and in combination. BioFactors 2004;22(1): 49-55.

41. Stolzenberg-Solomon RZ, Cross AJ, Silverman DT, et al. Meat and Meat-Mutagen Intake and Pancreatic Cancer Risk in the NIH-AARP Cohort. Cancer Epidemiology Biomarkers & Prevention 2007;16(12): 2664-75.

42. González CA, Jakszyn P, Pera G, et al. Meat Intake and Risk of Stomach and Esophageal Adenocarcinoma Within the European Prospective Investigation Into Cancer and Nutrition (EPIC). Journal of the National Cancer Institute 2006;98(5): 345-54.

43. Bingham SA, Hughes R, Cross AJ. Effect of white versus red meat on endogenous N-nitrosation in the human colon and further evidence of a dose response. The Journal of nutrition 2002;132(11 Suppl): 3522S-25S.

44. Hughes R, Cross AJ, Pollock JR, Bingham S. Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. Carcinogenesis 2001;22(1): 199-202.

45. Cross AJ, Pollock JRA, Bingham SA. Haem, not Protein or Inorganic Iron, Is Responsible for Endogenous Intestinal N-Nitrosation Arising from Red Meat. Cancer Research 2003;63(10): 2358-60.

46. Hughes R, Pollock JR, Bingham S. Effect of vegetables, tea, and soy on endogenous N-nitrosation, fecal ammonia, and fecal water genotoxicity during a high red meat diet in humans. Nutrition and cancer 2002;42(1): 70-77.

47. Sinha R, Peters U, Cross AJ, et al. Meat, Meat Cooking Methods and Preservation, and Risk for Colorectal Adenoma. Cancer Research 2005;65(17): 8034-41.

48. Jakszyn P, Gonzalez C-A. Nitrosamine and related food intake and gastric and oesophageal cancer risk: a systematic review of the epidemiological evidence. World journal of gastroenterology : WJG 2006;12(27): 4296-303.

49. Greenwald P, Milner JA, Anderson DE, McDonald SS. Micronutrients in Cancer Chemoprevention. Cancer and Metastasis Reviews 2002;21(3): 217-30.

50. Rhodes MJC. Physiological roles for secondary metabolites in plants: some progress, many outstanding problems. Plant Molecular Biology 1994;24(1): 1-20.

51. Tsuyoshi G, Nobuyuki T, Shizuka H, Teruo K. Various Terpenoids Derived from Herbal and Dietary Plants Function as PPAR Modulators and Regulate Carbohydrate and Lipid Metabolism: Hindawi Publishing Corporation, 2010.

52. Hatch FT, MacGregor JT, Zeiger E. Review: Putative mutagens and carcinogens in food. VII. Genetic toxicology of the diet. A summary of a satellite symposium of the Fourth International Conference on Environmental Mutagens. Copenhagen, June 19-22, 1985. Abstracts. Environ Mutagen 1986;8(3): 467-84.

53. Ross JA, Kasum CM. DIETARY FLAVONOIDS: Bioavailability, Metabolic Effects, and Safety. Annual Review of Nutrition 2003;22(1): 19-34.

54. Ryan CA. Protease Inhibitors in Plants: Genes for Improving Defenses Against Insects and Pathogens. Annual Review of Phytopathology 2003;28(1): 425-49.

55. Conn EE. Cyanogenic glycosides. Journal of Agricultural and Food Chemistry 1969;17(3): 519-26.

56. Maga JA. Glycoalkaloids in solanaceae. Food Reviews International 1994;10(4): 385-418.

57. Gaitan E. Goitrogens in Food and Water. Annual Review of Nutrition 2003;10(1): 21-37.

58. Howes AJ, Chan VSW, Caldwell J. Structure-specificity of the genotoxicity of some naturally occurring alkenylbenzenes determined by the unscheduled DNA synthesis assay in rat hepatocytes. Food and Chemical Toxicology 1990;28(8): 537-42.

59. STEVEN FS, GRIFFIN MM, SMITH RH. Disulphide Exchange Reactions in the Control of Enzymic Activity. European Journal of Biochemistry 1981;119(1): 75-78.

60. Mason RW, Lucas P. Acute poisoning in cattle after eating old non-viable seed of chou moellier (Brassica oleracea convar. acephala). Aust Vet J 1983;60(9): 272-3.

61. Heaney RK, Fenwick GR. Natural toxins and protective factors in brassica species, including rapeseed. Natural Toxins 1995;3(4): 233-37.

62. von Borstel RC, Higgins JA. Janus carcinogens and mutagens. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1998;402(1-2): 321-29.

63. NTP. NTP Toxicology and Carcinogenesis Studies of Methyleugenol (CAS NO. 93-15-2) in F344/N Rats and B6C3F1 Mice (Gavage Studies). Natl Toxicol Program Tech Rep Ser 2000;491: 1-412.

64. Smith RL, Adams TB, Doull J, et al. Safety assessment of allylalkoxybenzene derivatives used as flavouring substances -- methyl eugenol and estragole. Food and Chemical Toxicology 2002;40(7): 851-70.

65. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. The New England Journal of Medicine 2001;344(18): 1343-50.

66. Dai Q, Borenstein AR, Wu Y, Jackson JC, Larson EB. Fruit and Vegetable Juices and Alzheimer's Disease: The Kame Project. The American Journal of Medicine 2006;119(9): 751-59.

67. Surh Y-J. Cancer chemoprevention with dietary phytochemicals. Nat Rev Cancer 2003;3(10): 768-80.

68. Scalbert A, Johnson IT, Saltmarsh M. Polyphenols: antioxidants and beyond. American Journal of Clinical Nutrition 2005;81(1): 215S-17.

69. Greenberg ER, Baron JA, Tosteson TD, et al. A Clinical Trial of Antioxidant Vitamins to Prevent Colorectal Adenoma. The New England Journal of Medicine 1994;331(3): 141-47.

70. Greenberg ER, Sporn MB. Antioxidant Vitamins, Cancer, and Cardiovascular Disease. The New England Journal of Medicine 1996;334(18): 1189-90.

71. Verhoeven DT, Goldbohm RA, van Poppel G, Verhagen H, van den Brandt PA. Epidemiological studies on brassica vegetables and cancer risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 1996;5(9): 733-48.

72. Holst B, Williamson G. A critical review of the bioavailability of glucosinolates and related compounds. Natural product reports 2004;21(3): 425-47.

73. Milner JA. Molecular Targets for Bioactive Food Components. Journal of Nutrition 2004;134(9): 2492S-98.

74. Davis CD. Nutritional Interactions: Credentialing of Molecular Targets for Cancer Prevention. Experimental Biology and Medicine 2007;232(2): 176-83.

75. Nicholson JK, Holmes E, Lindon JC, Wilson ID. The challenges of modeling mammalian biocomplexity. Nature biotechnology 2004;22(10): 1268-74.

76. Zhang Y, Yao S, Li J. Vegetable-derived isothiocyanates: anti-proliferative activity and mechanism of action. The Proceedings of the Nutrition Society 2006;65(1): 68-75.

77. Smith TJ. Mechanisms of carcinogenesis inhibition by isothiocyanates. Expert opinion on investigational drugs 2001;10(12): 2167-74.

78. Verhoeven DT, Verhagen H, Goldbohm RA, van den Brandt PA, van Poppel G. A review of mechanisms underlying anticarcinogenicity by brassica vegetables. Chemico-biological interactions; Chemico-biological interactions 1997;103(2): 79-129.

79. Bogaards JJ, Verhagen H, Willems MI, van Poppel G, van Bladeren PJ. Consumption of Brussels sprouts results in elevated alpha-class glutathione S-transferase levels in human blood plasma. Carcinogenesis; Carcinogenesis 1994;15(5): 1073-75.

80. Nijhoff WA, Grubben MJ, Nagengast FM, et al. Effects of consumption of Brussels sprouts on intestinal and lymphocytic glutathione S-transferases in humans. Carcinogenesis; Carcinogenesis 1995;16(9): 2125-28.

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

82. Kall MA, Vang O, Clausen J. Effects of dietary broccoli on human in vivo drug metabolizing enzymes: evaluation of caffeine, oestrone and chlorzoxazone metabolism. Carcinogenesis; Carcinogenesis 1996;17(4): 793-99.

83. Slattery ML, Potter JD, Coates A, et al. Plant foods and colon cancer: an assessment of specific foods and their related nutrients (United States). Cancer Causes and Control 1997;8(5): 575-90.

84. Dodd KW, Guenther PM, Freedman LS, et al. Statistical Methods for Estimating Usual Intake of Nutrients and Foods: A Review of the Theory. Journal of the American Dietetic Association 2006;106(10): 1640-50.

85. Olafsdottir AS, Thorsdottir I, Gunnarsdottir Ir, Thorgeirsdottir H, Steingrimsdottir L. Comparison of women's diet assessed by FFQs and 24-hour recalls with and without underreporters: associations with biomarkers: Karger, 2006.

86. Michels KB, Bingham SA, Luben R, Welch AA, Day NE. The Effect of Correlated Measurement Error in Multivariate Models of Diet. American Journal of Epidemiology 2004;160(1): 59-67.

87. Jeffery EH, Brown AF, Kurilich AC, et al. Variation in content of bioactive components in broccoli. Journal of Food Composition and Analysis 2003;16(3): 323-30.

88. Cary EE, Allaway WH. Selenium Content of Field Crops Grown on Selenite-Treated Soils. Agronomy Journal 1973;65(6): 922-25.

89. Stahl W, Sies H. Uptake of Lycopene and Its Geometrical Isomers Is Greater from Heat-Processed than from Unprocessed Tomato Juice in Humans. Journal of Nutrition 1992;122(11): 2161-66.

90. Fleischauer AT, Poole C, Arab L. Garlic consumption and cancer prevention: meta-analyses of colorectal and stomach cancers. American Journal of Clinical Nutrition 2000;72(4): 1047-52.

91. Timbrell JA. Biomarkers in toxicology. Toxicology 1998;129(1): 1-12.

92. Vainio H. Use of biomarkers -- new frontiers in occupational toxicology and epidemiology. Toxicology Letters 1998;102-103: 581-89.

93. Shapiro TA, Fau - Fahey JW, Fahey JW, et al. Human metabolism and excretion of cancer chemoprotective glucosinolates and isothiocyanates of cruciferous vegetables. - Cancer Epidemiol Biomarkers Prev.1998 Dec;7(12):1091-100.(1055-9965 (Print)).

94. Lindon JC, Nicholson JK, Holmes E, et al. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicology and applied pharmacology 2003;187(3): 137-46.

95. Clayton TA, Lindon JC, Cloarec O, et al. Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 2006;440(7087): 1073-77.

96. Heinzmann SS, Brown IJ, Chan Q, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. American Journal of Clinical Nutrition 2010;92(2): 436-43.

97. Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK. An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. FEBS letters 2000;484(3): 169-74.

98. Spraul M, Neidig P, Klauck U, et al. Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples. Journal of Pharmaceutical and Biomedical Analysis 1994;12(10): 1215-25.

99. Gray J, Nicholson JK, Creasy DM, Timbrell JA. Studies on the relationship between acute testicular damage and urinary and plasma creatine concentration. Arch Toxicol 1990;64(6): 443-50.

100. Stella C, Beckwith-Hall B, Cloarec O, et al. Susceptibility of human metabolic phenotypes to dietary modulation. Journal of proteome research 2006;5(10): 2780-88.

101. Myint T, Fraser GE, Lindsted KD, Knutsen SF, Hubbard RW, Bennett HW. Urinary 1-Methylhistidine Is a Marker of Meat Consumption in Black and in White California Seventh-day Adventists. American Journal of Epidemiology 2000;152(8): 752-55.

102. Lawrence JR, Peter R, Baxter GJ, Robson J, Graham AB, Paterson JR. Urinary excretion of salicyluric and salicylic acids by non-vegetarians, vegetarians, and patients taking low dose aspirin. Journal of Clinical Pathology 2003;56(9): 651-53.

103. Walsh MC, Brennan L, Malthouse JPG, Roche HM, Gibney MJ. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. American Journal of Clinical Nutrition 2006;84(3): 531-39.

104. Zuppi C, Messana I, Forni F, Ferrari F, Cristina R, Giardina B. Influence of feeding on metabolite excretion evidenced by urine 1H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard). Clinica Chimica Acta 1998;278(1): 75-79.

105. Somoza V, Wenzel E, Weiß C, Clawin-Rädecker I, Grübel N, Erbersdobler HF. Dose-dependent utilisation of casein-linked lysinoalanine, N(epsilon)-fructoselysine and N(epsilon)-carboxymethyllysine in rats. Molecular Nutrition & Food Research 2006;50(9): 833-41.

106. Brantsæter AL, Haugen M, Rasmussen SE, Alexander J, Samuelsen SO, Meltzer HM. Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). Public Health Nutrition 2007;10(08): 838-47.

107. Hong Y-J, Mitchell AE. Metabolic Profiling of Flavonol Metabolites in Human Urine by Liquid Chromatography and Tandem Mass Spectrometry. Journal of Agricultural and Food Chemistry 2004;52(22): 6794-801.

108. Bertram HC, Hoppe C, Petersen BO, Duus JØ, Mølgaard C, Michaelsen KF. An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys. British Journal of Nutrition 2007;97(04): 758-63.

109. Lee MB, Storer MK, Blunt JW, Lever M. Validation of 1H NMR spectroscopy as an analytical tool for methylamine metabolites in urine. Clinica Chimica Acta 2006;365(1-2): 264-69.

110. 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. Journal of Agricultural and Food Chemistry; Journal of Agricultural and Food Chemistry 2005;53(2): 191-96.

111. 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; The Analyst 2004;129(3): 259-64.

112. Solanky KS, Bailey NJ, Beckwith-Hall BM, et al. Biofluid 1H NMR-based metabonomic techniques in nutrition research - metabolic effects of dietary isoflavones in humans. The Journal of nutritional biochemistry 2005;16(4): 236-44.

113. Solanky KS, Bailey NJC, Holmes E, et al. NMR-Based Metabonomic Studies on the Biochemical Effects of Epicatechin in the Rat. Journal of Agricultural and Food Chemistry 2003;51(14): 4139-45.

114. Mennen LI, Sapinho D, Ito H, et al. Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. British Journal of Nutrition 2006;96(01): 191-98.

115. Ito H, Gonthier M-P, Manach C, et al. Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. British Journal of Nutrition 2005;94(04): 500-09.

116. Van Dorsten FA, Daykin CA, Mulder TPJ, Van Duynhoven JPM. Metabonomics Approach To Determine Metabolic Differences between Green Tea and Black Tea Consumption. Journal of Agricultural and Food Chemistry 2006;54(18): 6929-38.

117. Trujillo E, Davis C, Milner J. Nutrigenomics, Proteomics, Metabolomics, and the Practice of Dietetics. Journal of the American Dietetic Association 2006;106(3): 403-13.

118. Lindon JC, Holmes E, Nicholson JK. Metabonomics and its role in drug development and disease diagnosis. Expert Rev.Mol.Diagn. 2004;4(2): 189-99.

119. Lindon JC, Holmes E, Nicholson JK. Metabonomics: systems biology in pharmaceutical research and development. Current opinion in molecular therapeutics 2004;6(3): 265-72.

120. Eriksson L, Antti H, Gottfries J, et al. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). Anal.Bioanal Chem. 2004;380(3): 419-29.

121. Selman C, Kerrison ND, Cooray A, et al. Coordinated multitissue transcriptional and plasma metabonomic profiles following acute caloric restriction in mice. Physiological Genomics 2006;27(3): 187-200.

122. Griffin JL, Muller D, Woograsingh R, et al. Vitamin E deficiency and metabolic deficits in neuronal ceroid lipofuscinosis described by bioinformatics. Physiological Genomics 2002;11(3): 195-203.

123. Gill SR, Pop M, DeBoy RT, et al. Metagenomic Analysis of the Human Distal Gut Microbiome. Science 2006;312(5778): 1355-59.

124. Gordon JI, Hooper LV, McNevin MS, Wong M, Bry L. Epithelial cell growth and differentiation. III. Promoting diversity in the intestine: conversations between the microflora, epithelium, and diffuse GALT. The American Journal of Physiology 1997;273(3 Pt 1): G565-70.

125. Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat.Rev.Microbiol. 2005;3(5): 431-38.

126. Guarner F, Malagelada JR. Gut flora in health and disease. Lancet 2003;361(9356): 512-19.

127. Long SS, Swenson RM. Development of anaerobic fecal flora in healthy newborn infants. The Journal of pediatrics 1977;91(2): 298-301.

128. Long SS, Swenson RM. Determinants of the developing oral flora in normal newborns. Applied and Environmental Microbiology 1976;32(4): 494-97.

129. Tannock GW. New perceptions of the gut microbiota: implications for future research. Gastroenterology clinics of North America 2005;34(3): 361-82, vii.

130. Tannock GW. Analysis of the intestinal microflora using molecular methods. European journal of clinical nutrition 2002;56 Suppl 4(Journal Article): S44-9.

131. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature 2007;449(7164): 804-10.

132. Selmer T, Andrei PI. p-Hydroxyphenylacetate decarboxylase from Clostridium difficile. A novel glycyl radical enzyme catalysing the formation of p-cresol. European journal of biochemistry / FEBS 2001;268(5): 1363-72.

133. Yokoyama MT, Carlson JR. Production of Skatole and para-Cresol by a Rumen Lactobacillus sp. Applied and Environmental Microbiology 1981;41(1): 71-76.

134. Yu L, Blaser M, Andrei PI, Pierik AJ, Selmer T. 4-Hydroxyphenylacetate Decarboxylases: Properties of a Novel Subclass of Glycyl Radical Enzyme Systems†. Biochemistry 2006;45(31): 9584-92.

135. Smith EA, Macfarlane GT. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. Microbial Ecology 1997;33(3): 180-88.

136. Shih VE. Alternative-pathway therapy for hyperammonemia. The New England journal of medicine 2007;356(22): 2321-22.

137. Pereira DIA, Gibson GR. Cholesterol Assimilation by Lactic Acid Bacteria and Bifidobacteria Isolated from the Human Gut. Applied and Environmental Microbiology 2002;68(9): 4689-93.

138. Martin F-PJ, Wang Y, Sprenger N, et al. Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Mol Syst Biol 2008;4.

139. Collins MD, Gibson GR. Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. American Journal of Clinical Nutrition 1999;69(5): 1052S-57.

140. Costabile A, Klinder A, Fava F, et al. Whole-grain wheat breakfast cereal has a prebiotic effect on the human gut microbiota: a double-blind, placebo-controlled, crossover study. British Journal of Nutrition 2008;99(01): 110-20.

141. Wong JMW, de Souza R, Kendall CWC, Emam A, Jenkins DJA. Colonic Health: Fermentation and Short Chain Fatty Acids. Journal of Clinical Gastroenterology 2006;40(3).

142. Xu J, Gordon JI. Inaugural Article: Honor thy symbionts. Proceedings of the National Academy of Sciences of the United States of America 2003;100(18): 10452-59.

143. Frankel G, Phillips AD, Rosenshine I, Dougan G, Kaper JB, Knutton S. Enteropathogenic and enterohaemorrhagic Escherichia coli: more subversive elements. Molecular microbiology 1998;30(5): 911-21.

144. Iehl-Robert M, Monnier F, North C, Bailly P. Epidemiology. Epidemics of Clostridium difficile in geriatrics. Soins. Gerontologie 2001;(28)(28): 17-19.

145. Wistrom J, Norrby SR, Myhre EB, et al. Frequency of antibiotic-associated diarrhoea in 2462 antibiotic-treated hospitalized patients: a prospective study. The Journal of antimicrobial chemotherapy 2001;47(1): 43-50.

146. Aue WP, Karhan J, Ernst RR. Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy. The Journal of Chemical Physics 1976;64(10): 4226-27.

147. Fan TWM. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. Progress in Nuclear Magnetic Resonance Spectroscopy 1996;28(2): 161-219.

148. Sobolev AP, Brosio E, Gianferri R, Segre AL. Metabolic profile of lettuce leaves by high-field NMR spectra. Magnetic Resonance in Chemistry 2005;43(8): 625-38.

149. Plumb R, Castro-Perez J, Granger J, Beattie I, Joncour K, Wright A. Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 2004;18(19): 2331-37.

150. Wilson ID, Nicholson JK, Castro-Perez J, et al. High Resolution "Ultra Performance" Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies. Journal of proteome research 2005;4(2): 591-98.

151. Taylor G. Disintegration of Water Drops in an Electric Field. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 1964;280(1382): 383-97.

152. Verentchikov AN, Ens W, Standing KG. Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. Analytical Chemistry 1994;66(1): 126-33.

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

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

155. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics 2002;16(3): 119-28.

156. 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. Analytical Chemistry 2005;77(5): 1282-89.

157. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical Chemistry 2006;78(3): 779-87.

158. U.S. Department of Health and Human Services PHS, National Toxicology Program. Report on Carcinogens, Eleventh Edition; U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program, 2009.

159. Sekizawa J, Shibamoto T. Genotoxicity of safrole-related chemicals in microbial test systems. Mutation Research/Genetic Toxicology 1982;101(2): 127-40.

160. Schiestl RH, Chan WS, Daniel Gietz R, Mehta RD, Hastings PJ. Safrole, eugenol and methyleugenol induce intrachromosomal recombination in yeast. Mutation Research/Genetic Toxicology 1989;224(4): 427-36.

161. 9th International Conference on Environmental Mutagens Abstracts. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2005;577(Supplement 1): e1-e256.

162. Phillips DH, Miller JA, Miller EC, Adams B. Structures of the DNA Adducts Formed in Mouse Liver after Administration of the Proximate Hepatocarcinogen 1â€ ²-Hydroxyestragole. Cancer Research 1981;41(1): 176-86.

163. Gardner I, Bergin P, Stening P, Kenna JG, Caldwell J. Immunochemical Detection of Covalently Modified Protein Adducts in Livers of Rats Treated with Methyleugenol. Chemical Research in Toxicology 1996;9(4): 713-21.

164. Ellis J, Carmichael P, Gooderham N. **DNA adduct levels** in the **liver** of the **F344 rat treated** with the **natural flavour methyl eugenol**. Toxicology: Elsevier, 2006:73.

165. Ellis J. Thesis: Studies on the Molecular Basis of Methyeugenol Carcinogenicity, 2008.

166. Graves SW, Runyon S. Determination of methyleugenol in rodent plasma by high-performance liquid chromatography. Journal of Chromatography B: Biomedical Sciences and Applications 1995;663(2): 255-62.

167. Robison SH, Barr DB. Use of biomonitoring data to evaluate methyl eugenol exposure. Environ Health Perspect 2006;114(11): 1797-801.

168. Solheim E, Scheline RR. Metabolism of Alkenebenzene Derivatives in the Rat. II. Eugenol and Isoeugenol Methyl Ethers. Xenobiotica 1976;6(3): 137-50.

169. Burkey JL, Sauer J-M, McQueen CA, Glenn Sipes I. Cytotoxicity and genotoxicity of methyleugenol and related congeners -- a mechanism of activation for methyleugenol. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2000;453(1): 25-33.

170. Thompson D, Constantin-Teodosiu D, Egestad B, Mickos H, Moldéus P. Formation of glutathione conjugates during oxidation of eugenol by microsomal fractions of rat liver and lung. Biochemical Pharmacology 1990;39(10): 1587-95.

171. Gardner I, Wakazono H, Bergin P, et al. Cytochrome P450 mediated bioactivation of methyleugenol to 1'- hydroxymethyleugenol in Fischer 344 rat and human liver microsomes. Carcinogenesis 1997;18(9): 1775-83.

172. Stott K, Stonehouse J, Keeler J, Hwang T-L, Shaka AJ. Excitation Sculpting in High-Resolution Nuclear Magnetic Resonance Spectroscopy: Application to Selective NOE Experiments. Journal of the American Chemical Society 1995;117(14): 4199-200.

173. Stanley EG, Bailey NJ, Bollard ME, et al. Sexual dimorphism in urinary metabolite profiles of Han Wistar rats revealed by nuclear-magnetic-resonance-based metabonomics. Analytical Biochemistry 2005;343(2): 195-202.

174. Schecter A, Lucier GW, Cunningham ML, et al. Human consumption of methyleugenol and its elimination from serum. Environ Health Perspect 2004;112(6): 678-80.

175. da Silva RP, Nissim I, Brosnan ME, Brosnan JT. Creatine synthesis: hepatic metabolism of guanidinoacetate and creatine in the rat in vitro and in vivo. AJP - Endocrinology and Metabolism 2009;296(2): E256-61.

176. Saric J. Interactions between immunity and metabolism ? contributions from the metabolic profiling of parasite-rodent models. Parasitology 2010;137(Special Issue 09): 1451-66.

177. Rai RK, Tripathi P, Sinha N. Quantification of Metabolites from Two-Dimensional Nuclear Magnetic Resonance Spectroscopy: Application to Human Urine Samples. Analytical Chemistry 2009;81(24): 10232-38.

178. T **Saito** KH, M Yanagisawa. **Integrated spectral data base system** for **organic compounds**, 2004.

179. Lang DH, Yeung CK, Peter RM, et al. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes : Selective catalysis by fmo3. Biochemical Pharmacology 1998;56(8): 1005-12.

180. Flynn CJ, Wecker L. Concomitant Increases in the Levels of Choline and Free Fatty Acids in Rat Brain: Evidence Supporting the Seizure-Induced Hydrolysis of Phosphatidylcholine. Journal of Neurochemistry 1987;48(4): 1178-84.

181. Ayesh R, Mitchell SC, Smith RL. Dysfunctional N-oxidation of trimethylamine and the influence of testosterone treatment in man. Pharmacogenetics 1995;5(4): 244-6.

182. Gavaghan CL, Nicholson JK, Connor SC, Wilson ID, Wright B, Holmes E. Directly Coupled High-Performance Liquid Chromatography and Nuclear Magnetic Resonance Spectroscopic with Chemometric Studies on Metabolic Variation in Sprague-Dawley Rats. Analytical Biochemistry 2001;291(2): 245-52.

183. AN Phipps BW, J Stewart. **Use** of **proton NMR** for **determining changes** in **metabolite excretion profiles induced** by **dietary changes** in the **rat**: The Royal Pharmaceutical Society, 1997.

184. Yasuda T, Takasawa A, Nakazawa T, Ueda J, Ohsawa K. Inhibitory effects of urinary metabolites on platelet aggregation after orally administering Shimotsu-To, a traditional Chinese medicine, to rats. Journal of Pharmacy and Pharmacology 2003;55(2): 239-44.

185. Bollard ME, Holmes E, Lindon JC, et al. Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1)H NMR spectroscopy of urine and pattern recognition. Analytical Biochemistry 2001;295(2): 194-202.

186. Coen M, Hong YS, Clayton TA, et al. The Mechanism of Galactosamine Toxicity Revisited; A Metabonomic Study. Journal of Proteome Research 2007;6(7): 2711-19.

187. Guri CD, Swingle KF, Cole LJ. Urinary Excretion of Deoxycytidine in Rats after X-irradiation: Dose-response and Effect of Age. International Journal of Radiation Biology 1967;12(4): 355-65.

188. Shimizu M, Fujimura S. Origin of Increased Deoxycytidine Excretion into Urine of Rats Bearing Yoshida Ascites Sarcoma. Cancer Research 1984;44(6): 2387-92.

189. Williams RE, Lenz EM, Lowden JS, Rantalainen M, Wilson ID. The metabonomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1H NMR and HPLC-TOF MS. Mol Biosyst 2005;1(2): 166-75.

190. Schnackenberg LK, Sun J, Espandiari P, Holland RD, Hanig J, Beger RD. Metabonomics evaluations of age-related changes in the urinary compositions of male Sprague Dawley rats and effects of data normalization methods on statistical and quantitative analysis. BMC Bioinformatics 2007;8 Suppl 7: S3.

191. Davies KM, Heaney RP, Rafferty K. Decline in muscle mass with age in women: A longitudinal study using an indirect measure. Metabolism 2002;51(7): 935-39.

192. Nicholls AW, Mortishire-Smith RJ, Nicholson JK. NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. Chemical research in toxicology 2003;16(11): 1395-404.

193. Holmes E, Nicholls AW, Lindon JC, et al. Chemometric Models for Toxicity Classification Based on NMR Spectra of Biofluids. Chemical Research in Toxicology 2000;13(6): 471-78.

194. Sanchez-Bernal C, Martin-Gonzalez A, Cabezas JA, Villar E, Perez-Gonzalez N. Age-related excretion of six glycosidases in rat urine. Mech Ageing Dev 1992;66(2): 121-9.

195. Wang Y, Lawler D, Larson B, et al. Metabonomic Investigations of Aging and Caloric Restriction in a Life-Long Dog Study. Journal of Proteome Research 2007;6(5): 1846-54.

196. Balcombe JP, Barnard ND, Sandusky C. Laboratory routines cause animal stress. Contemp Top Lab Anim Sci 2004;43(6): 42-51.

197. Germann PG, Ockert D. Granulomatous inflammation of the oropharyngeal cavity as a possible cause for unexpected high mortality in a Fischer 344 rat carcinogenicity study. Lab Anim Sci 1994;44(4): 338-43.

198. Germann PG, Ockert D, Heinrichs M. Pathology of the oropharyngeal cavity in six strains of rats: predisposition of Fischer 344 rats for inflammatory and degenerative changes. Toxicol Pathol 1998;26(2): 283-9.

199. Krahn DD, Gosnell BA, Majchrzak MJ. The anorectic effects of CRH and restraint stress decrease with repeated exposures. Biological Psychiatry 1990;27(10): 1094-102.

200. Good T, Khan MZ, Lynch JW. Biochemical and physiological validation of a corticosteroid radioimmunoassay for plasma and fecal samples in oldfield mice (Peromyscus polionotus). Physiology & Behavior 2003;80(2-3): 405-11.

201. Wang Y, Holmes E, Tang H, et al. Experimental metabonomic model of dietary variation and stress interactions. Journal of proteome research 2006;5(7): 1535-42.

202. Lücke T, Pérez-Cerdá C, Baumgartner M, et al. Propionic acidemia: unusual course with late onset and fatal outcome. Metabolism 2004;53(6): 809-10.

203. Calabrese V, Calvani M, Butterfield DA. Increased formation of short-chain organic acids after chronic ethanol administration and its interaction with the carnitine pool in rat. Archives of Biochemistry and Biophysics 2004;431(2): 271-78.

204. Barness LA, Young DG, Nocho R. Methylmalonate Excretion in Vitamin B12 Deficiency. Science 1963;140(3562): 76-77.

205. Fenili D, Brown M, Rappaport R, McLaurin J. Properties of scyllo-inositol as a therapeutic treatment of AD-like pathology. J Mol Med 2007;85(6): 603-11.

206. Zhang AQ, Mitchell SC, Smith RL. Dietary Precursors of Trimethylamine in Man: A Pilot Study. Food and Chemical Toxicology 1999;37(5): 515-20.

207. Wang X, Zhao T, Qiu Y, et al. Metabonomics Approach to Understanding Acute and Chronic Stress in Rat Modelsâ€ . Journal of Proteome Research 2009;8(5): 2511-18.

208. Tonkin A. The metabolic syndrome â□ " a growing problem. European Heart Journal Supplements 2004;6(suppl A): A37-A42.

209. McKeigue PM, Pierpoint T, Ferrie JE, Marmot MG. Relationship of glucose intolerance and hyperinsulinaemia to body fat pattern in South Asians and Europeans. Diabetologia 1992;35(8): 785-91.

210. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. Diabetes 1988;37(12): 1595-607.

211. Hanley AJG, Williams K, Stern MP, Haffner SM. Homeostasis Model Assessment of Insulin Resistance in Relation to the Incidence of Cardiovascular Disease. Diabetes Care 2002;25(7): 1177-84.

212. Laaksonen DE, Niskanen L, NyyssÃ¶nen K, et al. C-reactive protein and the development of the metabolic syndrome and diabetes in middle-aged men. Diabetologia 2004;47(8): 1403-10.

213. Hu FB, Willett WC. Optimal Diets for Prevention of Coronary Heart Disease. JAMA: The Journal of the American Medical Association 2002;288(20): 2569-78.

214. Pereira DIA, Gibson GR. Effects of Consumption of Probiotics and Prebiotics on Serum Lipid Levels in Humans. Critical Reviews in Biochemistry and Molecular Biology 2008;37(4): 259-81.

215. Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annual Review of Nutrition 2002;22(Journal Article): 283-307.

216. Williams CM, Jackson KG. Inulin and oligofructose: effects on lipid metabolism from human studies. British Journal of Nutrition 2002;87(Supplement S2): S261-S64.

217. Reddy BS, Simi B, Patel N, Aliaga C, Rao CV. Effect of Amount and Types of Dietary Fat on Intestinal Bacterial 7{alpha}-Dehydroxylase and Phosphatidylinositol-specific Phospholipase C and Colonic Mucosal Diayclglycerol Kinase and PKC Activities during Different Stages of Colon Tumor Promotion. Cancer Research 1996;56(10): 2314-20.

218. Morotomi M, Guillem JG, LoGerfo P, Weinstein IB. Production of Diacylglycerol, an Activator of Protein Kinase C, by Human Intestinal Microflora. Cancer Research 1990;50(12): 3595-99.

219. Cani P, Neyrinck A, Fava F, et al. Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. Diabetologia 2007;50(11): 2374-83.

220. Cani PD, Bibiloni R, Knauf C, et al. Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Dietâ “Induced Obesity and Diabetes in Mice. Diabetes 2008;57(6): 1470-81.

221. 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.

222. Brindle JT, Antti H, Holmes E, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nature medicine 2002;8(12): 1439-44.

223. Brindle JT, Nicholson JK, Schofield PM, Grainger DJ, Holmes E. Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. The Analyst 2003;128(1): 32-36.

224. Wang Y, Tang H, Nicholson JK, et al. Metabolomic strategy for the classification and quality control of phytomedicine: a case study of chamomile flower (Matricaria recutita L.). Planta Medica; Planta Medica 2004;70(3): 250-55.

225. Lenz EM, Bright J, Wilson ID, et al. Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. Journal of Pharmaceutical and Biomedical Analysis 2004;36(4): 841-49.

226. Shockcor JP, Unger SE, Wilson ID, Foxall PJ, Nicholson JK, Lindon JC. Combined HPLC, NMR spectroscopy, and ion-trap mass spectrometry with application to the detection and characterization of xenobiotic and endogenous metabolites in human urine. Analytical Chemistry 1996;68(24): 4431-35.

227. Jebb SA, Frost G, Griffin B, et al. The RISCK study: testing the impact of the amount and type of dietary fat and carbohydrate on metabolic risk. Nutrition Bulletin 2007;32(2): 154-56.

228. Swan G. Findings from the latest National Diet and Nutrition Survey. Proceedings of the Nutrition Society 2004;63(04): 505-12.

229. Austin MA, King MC, Vranizan KM, Krauss RM. Atherogenic lipoprotein phenotype. A proposed genetic marker for coronary heart disease risk. Circulation 1990;82(2): 495-506.

230. Roche HM, Zampelas A, Knapper JM, et al. Effect of long-term olive oil dietary intervention on postprandial triacylglycerol and factor VII metabolism. American Journal of Clinical Nutrition 1998;68(3): 552-60.

231. Dresner A, Laurent D, Marcucci M, et al. Effects of free fatty acids on glucose transport and IRS-1â "associated phosphatidylinositol 3-kinase activity. The Journal of Clinical Investigation 1999;103(2): 253-59.

232. Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR in biomedicine; NMR in biomedicine 2005;18(3): 143-62.

233. Scarfe GB, Wright B, Clayton E, et al. 19F-NMR and directly coupled HPLC-NMR-MS investigations into the metabolism of 2-bromo-4-trifluoromethylaniline in rat: a urinary excretion balance study without the use of radiolabelling. Xenobiotica; the fate of foreign compounds in biological systems 1998;28(4): 373-88.

234. Foxall PJ, Bending MR, Gartland KP, Nicholson JK. Acute renal failure following accidental cutaneous absorption of phenol: application of NMR urinalysis to monitor the disease process. Human toxicology 1989;8(6): 491-96.

235. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: Human gut microbes associated with obesity. Nature 2006;444(7122): 1022-23.

236. Dumas ME, Barton RH, Toye A, et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proceedings of the National Academy of Sciences of the United States of America; Proceedings of the National Academy of Sciences of the United States of America 2006;103(33): 12511-16.

237. Ebbels TM, Holmes E, Lindon JC, Nicholson JK. Evaluation of metabolic variation in normal rat strains from a statistical analysis of 1H NMR spectra of urine. Journal of pharmaceutical and biomedical analysis 2004;36(4): 823-33.

238. Jongen WM. Glucosinolates in Brassica: occurrence and significance as cancer-modulating agents. The Proceedings of the Nutrition Society; The Proceedings of the Nutrition Society 1996;55(1B): 433-46.

239. Getahun SM, Chung FL. Conversion of glucosinolates to isothiocyanates in humans after ingestion of cooked watercress. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 1999;8(5): 447-51.

240. Krul C, Humblot C, Philippe C, et al. Metabolism of sinigrin (2-propenyl glucosinolate) by the human colonic microflora in a dynamic in vitro large-intestinal model. Carcinogenesis 2002;23(6): 1009-16.

241. Fenwick GR, Heaney RK. Glucosinolates and their breakdown products in cruciferous crops, foods and feedingstuffs. Food Chemistry 1983;11(4): 249-71.

242. Faulkner K, Mithen R, Williamson G. Selective increase of the potential anticarcinogen 4-methylsulphinylbutyl glucosinolate in broccoli. Carcinogenesis; Carcinogenesis 1998;19(4): 605-09.

243. Zhang Y, Talalay P, Cho CG, Posner GH. A major inducer of anticarcinogenic protective enzymes from broccoli: isolation and elucidation of structure. Proceedings of the National Academy of Sciences of the United States of America; Proceedings of the National Academy of Sciences of the United States of America 1992;89(6): 2399-403.

244. Kissen R, Bones AM. Nitrile-specifier Proteins Involved in Glucosinolate Hydrolysis in Arabidopsis thaliana. Journal of Biological Chemistry 2009;284(18): 12057-70.

245. Preobrazhenskaya MN, Korolev AM, Rozhkov II, et al. From ascorbigens to indolocarbazoles. Il Farmaco 1999;54(5): 265-74.

246. Foo HL, Grønning LM, Goodenough L, et al. Purification and characterisation of epithiospecifier protein from Brassica napus: enzymic intramolecular sulphur addition within alkenyl thiohydroximates derived from alkenyl glucosinolate hydrolysis. FEBS letters 2000;468(2-3): 243-46.

247. Song L, Thornalley PJ. Effect of storage, processing and cooking on glucosinolate content of Brassica vegetables. Food and Chemical Toxicology 2007;45(2): 216-24.

248. Gamet-Payrastre L, Li P, Lumeau S, et al. Sulforaphane, a naturally occurring isothiocyanate, induces cell cycle arrest and apoptosis in HT29 human colon cancer cells. Cancer research 2000;60(5): 1426-33.

249. Hasegawa T, Nishino H, Iwashima A. Isothiocyanates inhibit cell cycle progression of HeLa cells at G2/M phase. Anti-Cancer Drugs 1993;4(2): 273-79.

250. Yu R, Mandlekar S, Harvey KJ, Ucker DS, Kong AN. Chemopreventive isothiocyanates induce apoptosis and caspase-3-like protease activity. Cancer research 1998;58(3): 402-08.

251. Zhang Y, Talalay P. Anticarcinogenic activities of organic isothiocyanates: chemistry and mechanisms. Cancer research 1994;54(7 Suppl): 1976s-81s.

252. Myzak MC, Karplus PA, Chung FL, Dashwood RH. A novel mechanism of chemoprotection by sulforaphane: inhibition of histone deacetylase. Cancer research 2004;64(16): 5767-74.

253. Zhang Y. Molecular mechanism of rapid cellular accumulation of anticarcinogenic isothiocyanates. Carcinogenesis 2001;22(3): 425-31.

254. Kosower NS, Kosower EM. The glutathione status of cells. International review of cytology 1978;54(Journal Article): 109-60.

255. Steinkellner H, Rabot S, Freywald C, et al. Effects of cruciferous vegetables and their constituents on drug metabolizing enzymes involved in the bioactivation of DNA-reactive dietary carcinogens. Mutation research 2001;480-481(Journal Article): 285-97.

256. Michnovicz JJ, Adlercreutz H, Bradlow HL. Changes in levels of urinary estrogen metabolites after oral indole-3-carbinol treatment in humans. Journal of the National Cancer Institute 1997;89(10): 718-23.

257. Hecht SS, Carmella SG, Murphy SE. Effects of watercress consumption on urinary metabolites of nicotine in smokers. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 1999;8(10): 907-13.

258. Kong AN, Owuor E, Yu R, et al. Induction of xenobiotic enzymes by the MAP kinase pathway and the antioxidant or electrophile response element (ARE/EpRE). Drug metabolism reviews 2001;33(3-4): 255-71.

259. Nioi P, Hayes JD. Contribution of NAD(P)H:quinone oxidoreductase 1 to protection against carcinogenesis, and regulation of its gene by the Nrf2 basic-region leucine zipper and the arylhydrocarbon receptor basic helix-loop-helix transcription factors. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2004;555(1-2): 149-71.

260. Layton DW, Bogen KT, Knize MG, Hatch FT, Johnson VM, Felton JS. Cancer risk of heterocyclic amines in cooked foods: an analysis and implications for research. Carcinogenesis 1995;16(1): 39-52.

261. Conaway CC, Jiao D, Kohri T, Liebes L, Chung FL. Disposition and pharmacokinetics of phenethyl isothiocyanate and 6-phenylhexyl isothiocyanate in F344 rats. Drug metabolism and disposition: the biological fate of chemicals 1999;27(1): 13-20.

262. Shapiro TA, Fau - Fahey JW, Fahey JW, et al. Human metabolism and excretion of cancer chemoprotective glucosinolates and isothiocyanates of cruciferous vegetables. - Cancer Epidemiol Biomarkers Prev.1998 Dec;7(12):1091-100. 1998(1055-9965 (Print)).

263. London SJ, Yuan JM, Chung FL, et al. Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet 2000;356(9231): 724-29.

264. Nho CW, Jeffery E. The synergistic upregulation of phase II detoxification enzymes by glucosinolate breakdown products in cruciferous vegetables. Toxicology and applied pharmacology 2001;174(2): 146-52.

265. Bradlow HL, Michnovicz JJ, Halper M, Miller DG, Wong GY, Osborne MP. Long-term responses of women to indole-3-carbinol or a high fiber diet. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 1994;3(7): 591-95.

266. Michnovicz JJ, Bradlow HL. Induction of estradiol metabolism by dietary indole-3-carbinol in humans. Journal of the National Cancer Institute 1990;82(11): 947-49.

267. Michnovicz JJ, Bradlow HL. Altered estrogen metabolism and excretion in humans following consumption of indole-3-carbinol. Nutrition and cancer 1991;16(1): 59-66.

268. Nijhoff WA, Mulder TP, Verhagen H, van Poppel G, Peters WH. Effects of consumption of brussels sprouts on plasma and urinary glutathione S-transferase class-alpha and -pi in humans. Carcinogenesis; Carcinogenesis 1995;16(4): 955-57.

269. Pantuck EJ, Pantuck CB, Garland WA, et al. Stimulatory effect of brussels sprouts and cabbage on human drug metabolism. Clinical pharmacology and therapeutics; Clinical pharmacology and therapeutics 1979;25(1): 88-95.

270. Pantuck EJ, Pantuck CB, Anderson KE, Wattenberg LW, Conney AH, Kappas A. Effect of brussels sprouts and cabbage on drug conjugation. Clinical pharmacology and therapeutics; Clinical pharmacology and therapeutics 1984;35(2): 161-69.

271. Verhagen H, Poulsen HE, Loft S, van Poppel G, Willems MI, van Bladeren PJ. Reduction of oxidative DNA-damage in humans by brussels sprouts. Carcinogenesis 1995;16(4): 969-70.

272. Wagner S, Scholz K, Donegan M, Burton L, Wingate J, Völkel W. Metabonomics and Biomarker Discovery: LC−MS Metabolic Profiling and Constant Neutral Loss

Scanning Combined with Multivariate Data Analysis for Mercapturic Acid Analysis. Analytical Chemistry 2005;78(4): 1296-305.

273. van Doorn R, Borm PJ, Leijdekkers CM, Henderson PT, Reuvers J, van Bergen TJ. Detection and identification of S-methylcysteine in urine of workers exposed to methyl chloride. International archives of occupational and environmental health 1980;46(2): 99-109.

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

275. Synge RL, Wood JC. (+)-(S-methyl-L-cysteine S-oxide) in cabbage. Biochem J 1956;64(2): 252-9.

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

277. 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.

278. Widarto H, Van Der Meijden E, Lefeber A, et al. Metabolomic Differentiation of Brassica rapa Following Herbivory by Different Insect Instars using Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. Journal of Chemical Ecology 2006;32(11): 2417-28.

279. Kaiser KA, Jr GAB, Larive CK. A comparison of metabolite extraction strategies for ^{<FONT SIZE='-1'>1</FONT>}H-NMR-based metabolic profiling using mature leaf tissue from the model plant <I>Arabidopsis thaliana</I>. Magnetic Resonance in Chemistry 2009;47(S1): S147-S56.

280. Tiziani S, Lodi A, Ludwig C, Parsons HM, Viant MR. Effects of the application of different window functions and projection methods on processing of 1H J-resolved nuclear magnetic resonance spectra for metabolomics. Analytica Chimica Acta 2008;610(1): 80-88.

281. Kumar BS, Lee Y-J, Yi HJ, Chung BC, Jung BH. Discovery of safety biomarkers for atorvastatin in rat urine using mass spectrometry based metabolomics combined with global and targeted approach. Analytica Chimica Acta 2010;661(1): 47-59.

282. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat.Rev.Drug Discov. 2002;1(2): 153-61.

283. Oliveira EJ, Watson DG. Liquid chromatography-mass spectrometry in the study of the metabolism of drugs and other xenobiotics. Biomed Chromatogr 2000;14(6): 351-72.

284. Benton HP, Wong DM, Trauger SA, Siuzdak G. XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. Analytical Chemistry 2008;80(16): 6382-89.

285. Stoewsand GS. Bioactive organosulfur phytochemicals in Brassica oleracea vegetables--a review. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 1995;33(6): 537-43.

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

287. 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. ChemInform 2005;36(35).

288. 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. Nat Prod Rep 2005;22(3): 351-68.

289. Chin H-W, Lindsay RC. Mechanisms of formation of volatile sulfur compounds following the action of cysteine sulfoxide lyases. Journal of Agricultural and Food Chemistry 1994;42(7): 1529-36.

290. Tulio AZ, Jr., Yamanaka H, Ueda Y, Imahori Y. Formation of methanethiol and dimethyl disulfide in crushed tissues of broccoli florets and their inhibition by freeze-thawing. Journal of Agricultural and Food Chemistry 2002;50(6): 1502-07.

291. 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.

292. Kyung KH, Fleming HP. Antimicrobial activity of sulfur compounds derived from cabbage. J Food Prot 1997;60(1): 67-71.

293. 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.

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

295. 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.

296. Itokawa Y, Inoue K, Sasagawa S, Fujiwara M. Effect of S-methylcysteine sulfoxide, S-allylcysteine sulfoxide and related sulfur-containing amino acids on lipid metabolism of experimental hypercholesterolemic rats. J Nutr 1973;103(1): 88-92.

297. Black S. [133b] Reduction of sulfoxide and disulfides. In: Sidney PCaNOK, editor. Methods in Enzymology: Academic Press, 1962:992-98.

298. Ostermayer F. Products of Acidic Hydrolysis of S-Methyl-L-cysteine Sulfoxide; the Isolation of Methyl Methanethiolsulfonate, and Mechanism of the Hydrolysis1. Journal of the American Chemical Society 1960;82(14): 3752.

299. Kawamori T, Tanaka T, Ohnishi M, et al. Chemoprevention of azoxymethane-induced colon carcinogenesis by dietary feeding of S-methyl methane thiosulfonate in male F344 rats. Cancer Res 1995;55(18): 4053-8.

300. Reddy BS, Kawamori T, Lubet R, Steele V, Kelloff G, Rao CV. Chemopreventive effect of S-methylmethane thiosulfonate and sulindac administered together during the promotion/progression stages of colon carcinogenesis. Carcinogenesis 1999;20(8): 1645-48.

301. Rebouche CJ, Seim H. Carnitine metabolism and its regulation in microorganisms and mammals. Annual Review of Nutrition 1998;18(Journal Article): 39-61.

302. Roberts I, Fau - Smith R, Smith R, Fau - Evans S, Evans S. Doubts over head injury studies. - BMJ.2007 Feb 24;334(7590):392-4.(1468-5833 (Electronic)).

303. Liu S, Saha B, Cotta M. Cloning, expression, purification, and analysis of mannitol dehydrogenase gene mtlK from Lactobacillus brevis. Applied Biochemistry and Biotechnology 2005;121-124(Journal Article): 391-401.

304. Korakli M, Vogel RF. Purification and characterisation of mannitol dehydrogenase from Lactobacillus sanfranciscensis. FEMS microbiology letters 2003;220(2): 281-86.

305. Aarnikunnas J, Ronnholm K, Palva A. The mannitol dehydrogenase gene (mdh) from Leuconostoc mesenteroides is distinct from other known bacterial mdh genes. Applied Microbiology and Biotechnology 2002;59(6): 665-71.

306. Hahn G, Kaup B, Bringer-Meyer S, Sahm H. A zinc-containing mannitol-2-dehydrogenase from Leuconostoc pseudomesenteroides ATCC 12291: purification of the enzyme and cloning of the gene. Archives of Microbiology 2003;179(2): 101-07.

307. Plumb RS, Johnson KA, Rainville P, et al. The detection of phenotypic differences in the metabolic plasma profile of three strains of Zucker rats at 20 weeks of age using ultra-performance liquid chromatography/orthogonal acceleration time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 2006;20(19): 2800-06.

308. Granger JH, Williams R, Lenz EM, Plumb RS, Stumpf CL, Wilson ID. A metabonomic study of strain- and age-related differences in the Zucker rat. Rapid Communications in Mass Spectrometry 2007;21(13): 2039-45.

309. Want EJ, Wilson ID, Gika H, et al. Global metabolic profiling procedures for urine using UPLC-MS. Nat. Protocols 2010;5(6): 1005-18.

310. Zeisel SH. Nutrigenomics and metabolomics will change clinical nutrition and public health practice: insights from studies on dietary requirements for choline. The American Journal of Clinical Nutrition 2007;86(3): 542-48.

311. Kind T, Scholz M, Fiehn O. How Large Is the Metabolome? A Critical Analysis of Data Exchange Practices in Chemistry. PLoS ONE 2009;4(5): e5440.

312. Scalbert A, Brennan L, Fiehn O, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 2009;5(4): 435-58-58.

313. Wishart DS. Metabolomics: applications to food science and nutrition research. Trends in Food Science & Technology 2008;19(9): 482-93.

314. Stamler J, Elliott P, Dennis B, et al. INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). J Hum Hypertens 0000;17(9): 591-608.

315. Crockford DJ, Holmes E, Lindon JC, et al. Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets:â□ ‰ Application in Metabonomic Toxicology Studies. Analytical Chemistry 2005;78(2): 363-71.