# Characterising the bidirectional biomolecular interactions between the intestinal microbiota and host drug metabolism

Thesis submitted by

*Marine Patricia Magali Letertre*

For the degree of

**Doctor of Philosophy (PhD) of Imperial College London**

**and the Diploma of Imperial College London (DIC)**

Supervised by

Dr Jonathan R. Swann

Prof Ian D. Wilson

Division of Nutrition Research
Department of Metabolism, Digestion and Reproduction
Faculty of Medicine
Imperial College London

# Characterising the bidirectional biomolecular interactions between the intestinal microbiota and host drug metabolism

Thesis submitted by

*Marine Patricia Magali Letertre*

For the degree of

**Doctor of Philosophy (PhD) of Imperial College London**

**and the Diploma of Imperial College London (DIC)**

Supervised by

Dr Jonathan R. Swann

Prof Ian D. Wilson

Division of Nutrition Research
Department of Metabolism, Digestion and Reproduction
Faculty of Medicine
Imperial College London

## 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 to the standard referencing practices in the discipline.

## Copyright declaration

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

# Acknowledgments

*First and foremost, I would like to thank my supervisors, Prof Ian Wilson and Dr Jonathan Swann. I found myself particularly lucky to have had such an enthusiastic and charismatic team of supervisors. Their guidance, availability and support, especially at the beginning of the PhD when everything can be confusing, helped me get through it all. Ian, your dedication to your students, your passion for LC-MS, your scientific thinking, your advice on life, allowed so many students to successfully obtain their Doctorate. I am glad I am one of them and I will keep in mind what you have taught me for the rest of my career… What could possibly go wrong? Jon, your passion for science and the hundreds of ideas you have every minute made my PhD such an intense and exciting experience. I have learned a lot from both of you and I wish I will be as accomplished as you one day.*

*I am also thankful to the STRATiGRAD PhD programme and especially Servier Technologies, for their funding which allowed me to perform my work. Matthieu Gaudin, who supported my application and believed in my motivation. Perrine Masson, who kindly took time to share her knowledge in the field and bring fresh ideas on my projects, as did Dr Pierre Barbier Saint Hilaire, Dr Erwan Wagner and Dr Bernard Walther. I am also particularly grateful to Dr Charles Pineau (Protim Platform), who was the first supporter of my ability to execute a PhD and has always been willing to guide me, whenever I needed it or wherever I was in the world.*

*Obviously, and I realised as I start writing this paragraph that it will be long, the three years of my PhD in CSM would not have been bearable without having met so many nice people. My beloved Kiana, of course, outside your infinite help with 16S and coding, you have always been there to listen to me and support me in the difficulties that I have had. As were Cecilia and Natasa, whose smiles made me continuously look on the bright side. The whole Greek team of course, Antonis, Alex(s), Panos, Eirini, and the CSM members that I had the chance to cross paths with, Simone, Vince, Olôf, Erika, Tsz, Pete, Rabiya, Beckie, Giorgia, Elze, Michael, Eric, Toby, Liz, Virag …, offered me such a magic atmosphere where the combination of shared knowledge and huge solidarity at work, addiction to gym classes and moments of craziness at the pub, made you all unforgettable to me. I would also like to thank Kate, Alex, Antonis, Nyasha and Leanne, as well as the whole NPC team and the Waters engineers, for their patience teaching me how to deal with UPLC-MS set-up and troubleshooting, as well as answering my numerous questions ("the work of the devil!", right Ian?). I would also like to acknowledge Cecilie Alstad and Annie Thaikkatil for their help in my research. Both were remarkable students who fully invested themselves and from whom I have also learned... Team work!*

*Then a very special thank you that I would like to send across the Channel, to my friends and my family waiting for us back in France. To the Krolombanais(es), the Garden team, the Credi team, the Simon's family and friends from Brittany in general, although you like to laugh about me playing with poo in a lab all day, your adorable support texts (often attached with new-born pictures lately!), gave me so much courage during the past three years. To the best parents and my brother, you have always believed in me more than I have myself, and you always supported my choices along the way, as long as I was happy and that I was following my passion, I dedicate this thesis to you. The beautiful moments and countless parties that we all shared with you Malouins, got us through rough times, we have missed you.*

*Finally, the most important thank you is addressed to my love, Erwan, for his support, his determination, his patience and his sense of humour. You have always followed me and believed in me since we first met, never doubting my abilities and choices, trying to understand what the mysterious technique of mass spectrometry really is and constantly cheering me up when things were not always going as planned. Although I know it was difficult at times, you keep taking care of me, particularly these last few months. I would have never been able to do it without you, this thesis is also your achievement, thank you.*

## Abstract

In pharmacological and toxicological studies, the aim is to improve drug efficacy while limiting toxicity. It is now recognised that the gut microbiota, which constitutes trillions of microbial cells and contains 150 time more genes that the human genome itself, can synthesise enzymes and metabolites that can influence the metabolism and toxicity of drugs in the host. Conversely, drugs themselves have also been identified to exert an impact on the community structure of the gut microbiota and thus modify its overall functionality. Such bidirectional interactions can have implications for host health and the outcomes of treatment strategies and therefore, it is essential that these complex events are further understood. In this work a combination of metataxonomic and metabolomic approaches have been used to explore various drug-microbiota-host interactions.

Firstly, a common chemotherapeutic, methotrexate, was studied. When administrated to rats at different doses, this drug was found to have a dose-dependent effect on the faecal microbial profile of the animals. Conversely, colonic bacteria in these animals were found to express an enzyme that was able to detoxify methotrexate by converting it into the non-toxic metabolite, 2,4-diamino-*N*-10-methylpteroic acid (DAMPA). Bacteria belonging to the phylum Firmicutes were found to correlate with this detoxification process suggesting that measuring this bacterial group (or their enzymes) prior to treatment with methotrexate could be used to prevent high dose MTX-induced toxicity.

Being able to systematically investigate the capacity of an individual's microbiome to directly metabolize a specific drug could help to inform treatment strategies to maximise efficacy and minimise toxicity. To achieve this, an *in vitro* model was developed to monitor the concentration of a drug following its incubation with faecal samples from different human donors. Despite technical limitations and the need for further development, this work demonstrated the potential utility of this model for rapidly monitoring the drug degradation capacity of an individual's microbiome.

The potential for the gut microbiome to indirectly modulate the drug metabolism of the host was then explored in mice. Here, a high-tyrosine diet was fed to mice to increase the microbial production of *p*-cresol in the presence and absence of antibiotics. This work furthered our understanding of the influence of gut-derived *p*-cresol on the sulfation capacity of the host liver and its effect on the phase II drug metabolism of paracetamol.

Finally, the benefits of inhibiting a specific bacterial enzyme implicated in the toxicity of irinotecan was explored to improve the outcomes of this drug. The metabolic perturbations induced by the inhibitor were found to be restricted to the gut environment with minimal biochemical consequences for the host. This demonstrates that selective inhibition of gut microbial enzymes could be an attractive target to complement the development of future drugs.

These investigations highlight the importance of taking into consideration the gut microbiota in clinical, pharmaceutical and toxicological studies. The gut microbiota can explain a notable proportion of toxicity and variability in drug responses and evaluating an individual's gut microbiota prior to drug treatment could help to improve drug outcomes. As technology advances and costs reduce, there is increasing potential to measure and use this microbial metric in the clinical setting to facilitate the implementation of stratified and personalised medicine.

# Table of contents


Statement of originality ..... 2
Copyright declaration ..... 2
Acknowledgments ..... 3
Abstract ..... 4
Table of contents ..... 5
List of Figures ..... 9
List of Tables ..... 13
Abbreviations ..... 14

**Chapter 1. Introduction ..... 18**
1. The gut microbiota ..... 18
1.1. What and where is the microbiome? ..... 18
1.2. Gut microbial influence over the host metabolic system ..... 20
1.3. The gut microbiota in health and disease ..... 21
1.4. Factors influencing the gut microbiota ..... 22
1.5. Inter- and intra-individual variability of the gut microbiota ..... 23
2. Drug metabolism ..... 24
2.1. Xenobiotic metabolism by the host ..... 24
2.2. Enterohepatic circulation ..... 26
2.3. Role of the gut microbiota in shaping drug metabolism ..... 28
2.4. Role of drugs in shaping the gut microbiota ..... 30
2.5. Challenge ..... 31
3. Systems Medicine ..... 32
3.1. Omics sciences ..... 32
3.2. DNA-based omics ..... 33
3.3. RNA-based omics ..... 33
3.4. Metabolomics ..... 34
3.5. Integration of the "-omics" sciences ..... 36
4. Thesis objectives ..... 36
Thesis outline ..... 38

**Chapter 2. Analytical background ..... 39**
1. Metabolic profiling (Metabolomics) ..... 40
1.1. $^{1}$H NMR spectroscopy ..... 40
1.2. Untargeted UPLC-MS ..... 47
1.3. Targeted UPLC-MS ..... 62

1.4. Multivariate Statistics ..... 72
1.5. Metabolite Identification ..... 81
1.6. Univariate Statistics ..... 79
1.7. Comparison between $^{1}$H NMR spectroscopy and UPLC-MS ..... 84
2. Amplicon sequencing ..... 85
2.1. Principle ..... 85
2.2. Data pre-processing ..... 87
2.3. Data analysis ..... 87

**Chapter 3. Investigating the two-way interaction between methotrexate and the gut microbiota in Sprague Dawley rats ..... 90**
1. Introduction ..... 91
2. Materials and methods ..... 94
2.1. Chemical and reagents. ..... 94
2.2. Study design. ..... 94
2.3. Faecal DNA extraction for amplicon sequencing. ..... 95
2.4. Faecal water preparation for UPLC-MS analysis. ..... 95
2.5. Urine sample preparation for UPLC-MS. ..... 96
2.6. UPLC-MS. ..... 96
2.7. Data Analysis for the amplicon sequencing. ..... 97
2.8. Targeted UPLC-MS data analysis. ..... 97
2.9. Untargeted UPLC-MS data analysis. ..... 98
2.10. Metabolite identification. ..... 98
2.11. Correlogram. ..... 99
3. Results ..... 99
3.1. Quantification of MTX and its metabolites 7-OH-MTX and DAMPA. ..... 99
3.2. MTX modulated the urinary metabolic profile ..... 101
3.3. MTX effects on the faecal metabolic profile ..... 103
3.4. MTX effect on the faecal microbial profile ..... 106
3.5. Correlation between DAMPA faecal excretion and intestinal microbial profile ..... 109
4. Discussion. ..... 113
5. Conclusions ..... 116

**Chapter 4. Development of an *in vitro* model to investigate direct drug-microbiome interactions 117**
1. Introduction ..... 118
2. General Experimental Section ..... 121
2.1. Chemicals ..... 121

2.2. Ethical considerations ..... 121
2.3. Faecal homogenate preparation ..... 122
2.4. Selection of the drugs and range of concentration to test ..... 122
3. Development of a rapid *in vitro* model ..... 123
3.1. Experimental part ..... 123
3.2. Results ..... 126
3.3. Conclusions ..... 130
4. Optimization experiments ..... 130
4.1. Experimental section ..... 131
4.2. Results ..... 133
4.3. Conclusion ..... 136
5. Refinement of the *in vitro* model ..... 137
5.1. Experimental section ..... 137
5.2. Results ..... 140
5.3. Conclusion ..... 148
6. General discussion ..... 149
6.1. Strengths ..... 149
6.2. Limitations ..... 150
6.3. Perspectives and future work ..... 151
7. Conclusion ..... 151

**Chapter 5. Dietary manipulation of gut microbial metabolism and its influence on host paracetamol metabolism ..... 152**
1. Introduction ..... 153
2. Experimental section ..... 156
2.1. Chemicals and reagents ..... 156
2.2. Animal handling ..... 157
2.3. UPLC-MS targeted analysis ..... 158
2.4. $^{1}$H NMR spectroscopy: ..... 160
3. Results ..... 162
3.1. Effect of high tyrosine diet on urine and plasma tyrosine metabolism ..... 162
3.2. Effect of high tyrosine diet on APAP metabolites detected in the urine, plasma and liver 164
3.3. Effect of APAP intake on tyrosine metabolism in the urine, plasma and liver ..... 167
3.4. Effect of high tyrosine diet and APAP administration on the hepatic metabolic profile 169
3.5. Multi-blocks integration of the targeted UPLC-MS data ..... 170
3.6. $^{1}$H-NMR spectroscopic analysis of liver aqueous extracts ..... 171
4. Discussion ..... 172

4.1. Antibiotic treatment ..... 172
4.2. High-tyrosine diet ..... 173
4.3. APAP treatment ..... 173
5. Conclusion ..... 175

**Chapter 6. Investigating the metabolic perturbations induced by the selective inhibition of bacterial β-glucuronidases, enzymes known to influence drug toxicity ..... 176**
1. Introduction ..... 177
2. Experimental Section ..... 181
2.1. Chemicals ..... 181
2.3. $^{1}$H NMR spectroscopy sample preparation ..... 182
2.4. UPLC-MS sample preparation ..... 182
2.6. UPLC-MS analysis ..... 184
3. Results ..... 187
3.1. Plasma samples ..... 187
3.2. Liver tissue ..... 187
3.3. Gall bladder tissue ..... 187
3.8. Network analysis ..... 198
4. Discussion ..... 203
5. Conclusion ..... 205

**Chapter 7. General Discussion and Perspectives ..... 206**
1. General discussion ..... 206
2. Limitations ..... 209
3. Perspectives and further work ..... 209
4. General conclusion ..... 211

Appendices ..... 228
Publications ..... 259

# List of Figures

## Chapter 1

Figure 1-1. Exposed view of the gastrointestinal tract, along with its physiological properties and most abundant phyla. ..... 20
Figure 1-2. Xenobiotic transport (represented on the left side) and metabolism (represented on the right side) across an animal cell. ..... 24
Figure 1-3. Schematic representation of a) enterohepatic circulation and the different molecular reactions occurring on xenobiotics in the liver, the kidney and in b) the microbiota located in the gastrointestinal (GI) tract.. ..... 27
Figure 1-4. Representation of what Integrative Systems Medicine research takes into consideration to study human health. ..... 32
Figure 1-5. Overview of the multifaceted approach adopted to deepen the understanding of the complex tripartite drug-microbiome-host health interaction. It involves direct interactions (methotrexate, MTX) or indirect interaction. The latter is exemplified by competition between acetaminophen and the microbial metabolite p-cresol for sulfation, and the potential benefit of targeting the microbiome to reduce irinotecan induced-toxicity. ..... 37

## Chapter 2

Figure 2-1. A schematic for the superconducting magnet of an NMR spectrometer. ..... 40
Figure 2-2. $^{1}H$ nuclear magnetic moment μ spinning around the axis the magnetic field $B_0$ at a momentum angle (P) in a β and α energy levels, respectively presenting a magnetic quantum number of - 1/2 and +1/2 and separated by an energy difference of ΔE. ..... 42
Figure 2-3. Radiofrequency 90° pulse applied on an aligned spin to $B_0$ (initial position $M_0$). ..... 43
Figure 2-4. $^{1}H$ NMR signal multiplicity principle according Pascal's triangle and exemplified with ethanol. ..... 44
Figure 2-5. Different 1D and 2D NMR spectroscopy methods. ..... 46
Figure 2-6. Liquid chromatography principle, showing how the analytes contained in a sample can be separated after passage through the column followed by elution and detection. ..... 48
Figure 2-7. Electrospray ionization (ESI) principle. ..... 51
Figure 2-8. Ion optics of the Xevo G2 Q-Tof mass analyser used in the Chapter 6. ..... 53
Figure 2-9. Different chromatographic and MS traces. ..... 56
Figure 2-10. Tyrosine to *p*-cresol pathway. ..... 64
Figure 2-11. UHPLC-MS chromatogram of the twelve standards and five deuterated internal standards. ..... 65
Figure 2-12. Principle of principal component analysis. ..... 74
Figure 2-13. Effect of mean-centring and log-transforming on PCA scores plot. ..... 75
Figure 2-14. An example of the influence of the centring and scaling on the standard deviation of an UPLC-MS dataset according to the intensity of the feature. ..... 77

Figure 2-15. Supervised analysis, cross validation and permutation test. PLS-DA score plot where each point is a sample and where the class affiliation ("case" and "control" samples which are in red and black respectively) is taking into consideration to build the model.. ..... 78
Figure 2-16. Representation of the 16S rRNA gene, composed of conserved region (pink) and variable regions (blue). ..... 85
Figure 2-17. Amplicon sequencing workflow. It starts by DNA extraction, followed by the amplification of the 16S rRNA sequence and sequencing, prior OTUs clustering based on sequences similarity and identification by matching to databases..... 86
Figure 2-18. Investigation of the influence of distance metric and ordination method on β diversity. ..... 89

## Chapter 3

Figure 3-1. Summary of MTX metabolism.. ..... 92
Figure 3-2. Quantification of MTX, 7-OH-MTX and DAMPA using a semi-targeted UPLC-MS approach. ..... 100
Figure 3-3. PCA scores plots of the urine metabolic profile, plotted according the dose of MTX administrated..... 101
Figure 3-4. CA-PLS-DA models performed on the metabolic profile of the urine samples. ..... 102
Figure 3-5. PCA scores plots of the faecal metabolic profiles plotted according the dose of MTX administered. ..... 104
Figure 3-6. CA-PLS-DA models performed on the metabolic profile of the faecal samples. ..... 105
Figure 3-7. Investigation of the β and α diversities of the faecal samples. ..... 107
Figure 3-8. Taxonomic profiles of the faecal samples of the rodents representing the relative abundance of the Family in each of the MTX dose-related group pre- and post- MTX treatment (0 and 48 h respectively). ..... 108
Figure 3-9. Taxonomic profiles of the faecal samples of the rodents representing the relative abundance of the phyla in each of the MTX dose-related group pre- and post- MTX treatment (0 and 48 h respectively). ..... 109
Figure 3-10. Correlogram integrating the genera present pre-MTX-dose in the faecal samples with MTX concentration measured in urine at 12h and in faeces at 48h, as well as the quantity of DAMPA excreted in faeces samples at 48 h. ..... 110

## Chapter 4

Figure 4-1. *In vitro* models allowing the investigation of biomolecular mechanisms between bacterial species and intestines epithelium..... 120
Figure 4-2. Location of the samples across the 96-well plate used in the development of the *in vitro* model. ..... 124
Figure 4-3. *In vitro* model workflow..... 124
Figure 4-4. APAP concentrations measured in the first attempt to develop an *in vitro* model. ..... 127
Figure 4-5. APAP controls concentrations measured in the first attempt to develop an *in vitro* model. ..... 128
Figure 4-6. SSZ concentrations measured in the first attempt to develop an *in vitro* model. ..... 129

Figure 4-7. SSZ control concentrations measured in the first attempt to develop an *in vitro* model.. ..... 130
Figure 4-8. APAP concentrations in the first optimization experiment testing different diluents ($H_2O$, PBS, basal medium and gut microbiota medium) at two temperature (room temperature and 37°C). ..... 133
Figure 4-9. APAP concentrations in the second optimization experiment testing plastic *vs* glass containers. ..... 134
Figure 4-10. APAP concentrations in the third optimization experiment testing incubation under aerobic or anaerobic conditions. ..... 135
Figure 4-11. APAP concentrations in the fourth optimization experiment testing incubation under aerobic or anaerobic conditions, both with normal and fixed (4 % PFA) faecal homogenate. ..... 136
Figure 4-12. Detailed pre-incubation steps added to the refined *in vitro* model. ..... 138
Figure 4-13. Refined *in vitro* model workflow. ..... 139
Figure 4-14. APAP concentrations measured in the refined an *in vitro* model. ..... 141
Figure 4-15. SSZ concentrations measured in the refined an *in vitro* model.. ..... 143
Figure 4-16. MTX concentrations measured in the refined an *in vitro* model. ..... 145
Figure 4-17. APAP-G concentrations measured in the refined an *in vitro* model. ..... 147

## Chapter 5

Figure 5-1. molecular structure of *p*-cresol and paracetamol (APAP). ..... 154
Figure 5-2. Schematic representation of the competition between *p*-cresol, obtained from the microbial degradation of tyrosine from the diet, and the common used orally administered analgesic APAP. ..... 155
Figure 5-3. High tyrosine diet effect on tyrosine metabolites in a) urine and b) plasma. ..... 163
Figure 5-4. Boxplots of the APAP metabolites measured in the liver samples at a) 1h following APAP administration and b) 6h following APAP administration. ..... 164
Figure 5-5. Boxplots of the APAP metabolites measured in the plasma samples at a) 1 h following APAP administration and b) 6 h following APAP administration. ..... 165
Figure 5-6. Boxplots of the APAP metabolites measured in the urine samples at a) 1 h following APAP administration and b) 6 h following APAP administration. ..... 166
Figure 5-7. Boxplots of the tyrosine metabolites measured in the a) liver samples b) plasma samples and c) urine samples, at 1 h and 6 h following APAP administration. ..... 168
Figure 5-8. Investigation of pCS concentration impact on Sulfate/Glucruonide ratio. ..... 169
Figure 5-9. Integration of the urine and plasma tyrosine metabolic profile to perform a a) sparse PLS-DA model and b) a circos plot, highlighting correlations between tyrosine metabolites.. ..... 170
Figure 5-10. Scores plots from pairwise PCA models comparing the hepatic metabolic profiles of the different treatment groups 1 h and 6 h post APAP treatment. ..... 171

## Chapter 6

Figure 6-1. The metabolism of irinotecan (CPT-11).. ..... 177
Figure 6-2. Results of the high-throughput screening of a chemical library to underline compounds able to inhibit bacterial β-glucuronidase enzyme. ..... 178

Figure 6-3. Decreased toxicity of irinotecan (CPT-11) in the presence of the first inhibitor of β-glucuronidase (Inhibitor 1)..... 179
Figure 6-4. Structure of the inhibitor 9 (Inh-9), which inhibits the bacterial β-glucuronidase..... 180
Figure 6-5. CA-PLS-DA score plots and the kernel density estimation (KDE) obtained for the plasma, liver tissue and gall bladder samples. ..... 188
Figure 6-6. CA-PLS-DA score plots with the kernel density estimation (KDE) obtained for the colon tissue extracts. ..... 189
Figure 6-7. Heatmap of the metabolites annotated as impacted by the Inh-9 treatment by UPLC-MS ESI RP + in the colon tissues. ..... 190
Figure 6-8. CA-PLS-DA score plots and KDE for the models performed on the faecal and colon content extracts..... 191
Figure 6-9. Heatmap of the metabolites identified as impacted by the Inh-9 treatment by UPLC-MS ESI RP + in the colon contents. ..... 192
Figure 6-10. Unsupervised analysis performed on the UPLC-MS and $^{1}$H NMR spectroscopic urine sample dataset. ..... 194
Figure 6-11. Supervised analysis performed on the UPLC-MS and $^{1}$H NMR spectroscopic urine sample dataset. ..... 195
Figure 6-12. Heatmap of the putatively annotated metabolites found significantly different between the Inh-9-treated animals and the control animals, in the urine samples by $^{1}$H NMR spectroscopy, UPLC-MS ESI RP + and ESI RP –.. ..... 197
Figure 6-13. Molecular network of the metabolites annotated in the urine samples analysed by UPLC-MS. ..... 199
Figure 6-14. Molecular network of based on the metabolites annotated in the colon contents. ..... 201
Figure 6-15. Molecular network of the colon tissues. ..... 202

# List of Tables

## Chapter 2

Table 2-1. Summary of the analytical techniques use in each Chapter and on which samples it has been applied. ..... 39
Table 2-2. Common nuclei analysed by NMR and their respective spin quantum number (*I*) which varies according their number of protons and neutrons. ..... 41
Table 2-3. Chromatographic and Mass Spectrometric Details for the Determination of Tyrosine and Targeted Catabolites in Mouse Urine and Plasma..... 66
Table 2-4.QCs monitoring in urine. ..... 68
Table 2-5. QCs monitoring in plasma. ..... 68
Table 2-6. Concentration of the low and high QCs for each compound measured initially, and following periods of 28h and 390h in the autosampler at 4˚C. ..... 70
Table 2-7. Percentage bias and CV of the low and high QCs for each compound following a period of 28h and 390h in the autosampler at 4˚C compared to the initial concentration measured..... 70

## Chapter 4

Table 4-1. Approximate concentrations of APAP, SSZ and MTX in the gut. ..... 123

# Abbreviations

$^{1}$H NMR: proton nuclear magnetic resonance
5-HIAA: 5-hydroxyindoleacetic acid
7-OH-MTX: 7-hydroxy-methotrexate
AA: acetic acid
AB: antibiotic
ACN: acetonitrile
ACPI: atmospheric pressure chemical ionization
ADME: absorption, distribution, metabolism and elimination
ADRs: adverse drug reactions
AO: aldehyde oxidase
APAP: paracetamol
APAP-C-: paracetamol-cysteine
APAP-G: paracetamol glucuronide
APAP-GSH: paracetamol-glutathione
APAP-NAC: paracetamol-*N*-acetylcysteine
APAP-S: paracetamol sulfate
ASVs: amplicon sequence variants
ATP: adenosine triphosphate
$B_0$: magnetic field
BAs: bile acids
BEH: ethylene bridge hybrid
BM: basal medium
BMI: body mass index
BPI: base peak ion
*C. diff*: *Clostridium difficile*
CA-PLS-R/DA: covariate-adjusted projection to latent structure-regression or discriminant analysis
CC: calibration curve
CCS: collision-cross section
CE: capillary electrophoresis
CFU: colony-forming unit
CI: chemical ionization
CID: collision-induced dissociation
COSY: correlation spectroscopy
CPDG2: carboxypeptidase glutamate 2
CPMG: carr-purcell-meiboom-gill
CPT-11: irinotecan
CV: coefficient de variance
CYP450: cytochrome P450
$D_2O$: deuterium oxide
DAMPA: 2,4-diamino-*N*-10-methylpteroic acid
DDA: data dependant acquisition
DESI: desorption electrospray ionization
DHAP: dihydroxyacetone phosphate
DHFR: dihydrofolate reductase
DMSO: dimethyl sulfoxide
DNA: deoxyribonucleic acid
EI: electron impact ionization

EIC: extracted ion chromatograms
ER: endoplasmic reticulum
ESI +/-: electrospray positive or negative ionization
EtOH: ethanol
FA: formic acid
FAD: flavin adenine dinucleotide
FAs: fatty acids
FDA: Food and Drug Administration
FDR: false discovery rate
FGCP: folylpolyglutamate carboxypeptidase
FID: free induction decay
FISH: fluorescence *in situ* hybridisation
FPGS: folypolylglutamate synthetase
FT: Fourier transformation
FWER: family-wise error rate
FWHM: full width at half maximum
G3P: glycerol-3-phosphate
GC: gas chromatography
GI: gastrointestinal
GMM: gut microbiota medium
GSTs: glutathione *S*-transferase
HAs: heterocyclic aromatic amines
HCl: hydrochloric acid
HD: high dose
HILIC: hydrophobic interaction liquid chromatography
HMBC: heteronuclear multiple bond correlation
HMDB: Human metabolome database
HPLC: high-performance liquid chromatography
HSQC: heteronuclear single quantum coherence
HSS: high strength silica
HT: high tyrosine
i.v.: intravenous
*I*: spin quantum number
IACUC: Institutional Animal Care and Use Committee
ICR: ion cyclotron resonance
IM: ion mobility
Inh-9: inhibitor-9
IPA: isopropanol
IS: internal standard
J: coupling constant
J-Res: Coupling constant-resolved spectroscopy
KDE: kernel density estimate
L-DOPA: L-3,4-dihydroxyphenylalanine
LLOQ: lower limit of quantification
LOESS: local non-linear regression
*m/z*: mass-to-charge ratio
MALDI: matrix-assisted laser desorption ionization
MCCV: Monte-Carlo cross validation
MeOH: methanol

MoNA: MassBank of North America
MRM: multiple ion monitoring
mRNA: messenger RNA
MS: mass spectrometry
MTX: methotrexate
MTX: methotrexate
MW: molecular weight
$NaN_3$: sodium azide
NAPQi: N-acetyl p-benzoquinone imine
NMDS: non-metric multidimensional scaling
NOESY: nuclear overhauser effect spectroscopy
NPC: National Phenome Centre
NSAIDs: non-steroidal anti-inflammatory drugs
O-PLS-R/DA: orthogonal-projection to latent structure-regression or discriminant analysis
OTUs: operational taxonomic units
p.o.: per oral
PA: phosphatidic acid
PAG: phenylacetylglycine
PAPS: 3'-Phosphoadenosine-5'-phosphosulfate
PBS: phosphate-buffered saline
PCA: principal component analysis
*p*CG: *p*-cresol glucuronide
PCoA: principal coordinate analysis
PCR: polymerase chain reaction
*p*CS: *p*-cresol sulfate
PD: pharmacodynamic
PERMANOVA: permutational multivariate analysis
PFA: paraformaldehyde
Phase I: reactions of functionalization
Phase II: reactions of conjugation
*p*HB: *p*-hydroxybenzoic acid
*p*HLA: *p*-hydroxyphenyl lactic acid
*p*HPA: *p*-hydroxyphenyl acetic acid
*p*HPP: *p*-hydroxyphenyl propionic acid
PK: pharmacokinetic
PLS-R/DA: partial least squares regression or projection to latent structure-regression or discriminant analysis
ppm: parts per million
$Q^2$: define the predictive power of a multivariate statistical model
QC: quality control
Q-Tof: quadrupole- time of flight
*R*: resolving power
$R^2$: define the goodness of a multivariate statistical model
RA: rheumatoid arthritis
RCV: robustness of the cross validation ($R^2/Q^2$)
RDP: ribosomal database project
REIMS: rapid evaporative ionization
RF: radiofrequency
RMSE: root mean squared error
RMSEP: root mean squared error of the prediction

RP: reversed phase
rRNA: ribosomal ribonucleic acid
Rs: chromatographic resolution
S/G: sulfate to glucuronide ratio
S/N: Signal to noise ratio
SCFA: short-chain fatty acids
SD: standard deviation
SIL: stable isotope labelled
SIM: single ion monitoring
SN-38: 7-ethyl-10-hydroxy-camptothecin
SSZ: sulfasalazine
STOCSY: statistical total correlation spectroscopy
$T_0$: dead volume
THF: tetrahydrofolate
TIC: total ion current
TMA: trimethylamine
TMAO: trimethylamine-oxide
TOCSY: total correlation spectroscopy
TOF: time of flight
TQ: triple quadrupole
Tr: retention time
TSP: (3-(trimethylsilyl)-[2,2,3,3-$d_4$] propionic acid
U(H)PLC: ultra-high-performance liquid chromatography
UDP-GA: uridine 5'-diphospho-glucuronic acid
UDP-GT: uridine 5'-diphospho-glucuronosyltransferase
ULOQ: upper limit of quantification
γ: gyromagnetic ratio
μ: magnetic moment

# Chapter 1. Introduction

## 1. The gut microbiota

### 1.1. What and where is the microbiome?

#### *1.1.1. Definition*

The gastrointestinal (GI) tract hosts 38 trillion bacterial cells and various other microorganisms (Sender et al., 2016). This population of microorganisms are collectively referred to as the gut microbiota. It comprises archaea, bacteria, eukaryotes and viruses which cohabit within the mammalian organism. From the Eukaryotes, the fungi account for around 200 species belonging to four phyla (Ascomycota, Basidiomycota, Micosporidia and Zygomycota - Hamad et al., 2016), although this makes up only 0.03 % of the total faecal microbiota (Ott et al., 2008). Due to limited availability of analytical tools, Archaea and viruses are understudied, but are still known to play a role in methane production and immune responses, respectively. However, these organisms only constitute a small proportion of the microbiota, comparable to that of the fungi, making the bacterial community the most abundant in the GI tract. Research on the impact of the gut microbiota and especially bacteria on human health has considerably increased over the past twenty years as the gut microbiome, which is the collective term for all the genetic material of the microorganisms inhabiting the gut, encodes a vast range of metabolic potential.

#### *1.1.2. Localisation and composition of the gut microbiota*

The gastrointestinal tract encompasses the mouth, the oesophagus, the stomach, the small intestine (composed successively of the duodenum, jejunum and ileum), then the caecum, the colon (divided in the ascending, the transverse and the descending colon) and finishes at the rectum. The stomach contains a relatively low abundance of microorganisms ($10^1$ to $10^3$ Colony-Forming Unit (CFU)/mL) because of the acidic pH, the presence of the antimicrobial nitric oxide from gastric food processing, the bactericidal properties of the bile acids, coming from the bile duct upstream of the duodenum, and the high concentration of oxygen (Nardone and Compare, 2015; Donaldson et al., 2016). These conditions favour colonisation by facultative anaerobes. *Lactobacillus*, *Streptococcus*, *Staphylococcus*, *Prevotella*, *Veillonella*, *Rothia* and *Haemophilus* are the genera the most abundant in the stomach (Clarke et al., 2019; Nardone and Compare, 2015), along with some pathogens such as *Helicobacter pylori* which can easily colonise the stomach and which can lead to ulcers or gastric cancer (Benktander

et al., 2018). The microbial density and diversity increases moving along the GI tract from the small intestine to the colon (Donaldson et al., 2016). The small intestinal microbiota is enriched, particularly in the jejunum ($10^4$ to $10^7$ CFU/mL; Figure 1-1), with Firmicutes, particularly *Clostridium* and *Streptococcus* species, as well as Proteobacteria, especially *Escherichia coli* (Zoetendal et al., 2012). The density of these bacteria is directly related to the nutrient concentrations, as nutrients and small carbohydrates are easily accessible by the microbiota in these proximal regions of the small intestine. Such dietary components support rapid growth of microbes (Zoetendal et al., 2012). However, once in the caecum and the colon, simple nutrients become less abundant and are replaced by non-digestible polysaccharides and more complex carbohydrates. The digestion is then slower, promoting a higher proportion of strict anaerobes such as Bacteroidetes (Donaldson et al., 2016). Interestingly, Firmicutes species are more present in the deep part of the intestinal villi, formed by folds of the epithelium, and mucosal layer of the intestines in which mucus, secreted by enterocytes and perceived as a source of nutrients, is denser. In the large intestine this mucus layer is separated into two distinct zones (Figure 1-1). The inner zone is attached to the epithelium and is tightly packed to restrict the presence of bacteria, usually a small number of Firmicutes and Proteobacteria (e.g. *Actinobacter*) can be found here. The outer zone, which is in contact with the gut lumen, is more dispersed and is richly colonised by a variety of micro-organisms (Johansson et al., 2008). This includes Bacteroidetes, such as *Bacteroides acidifaciens* and *Bacteroides fragilis*, as this phyla is capable of polysaccharide degradation (Donaldson et al., 2016).

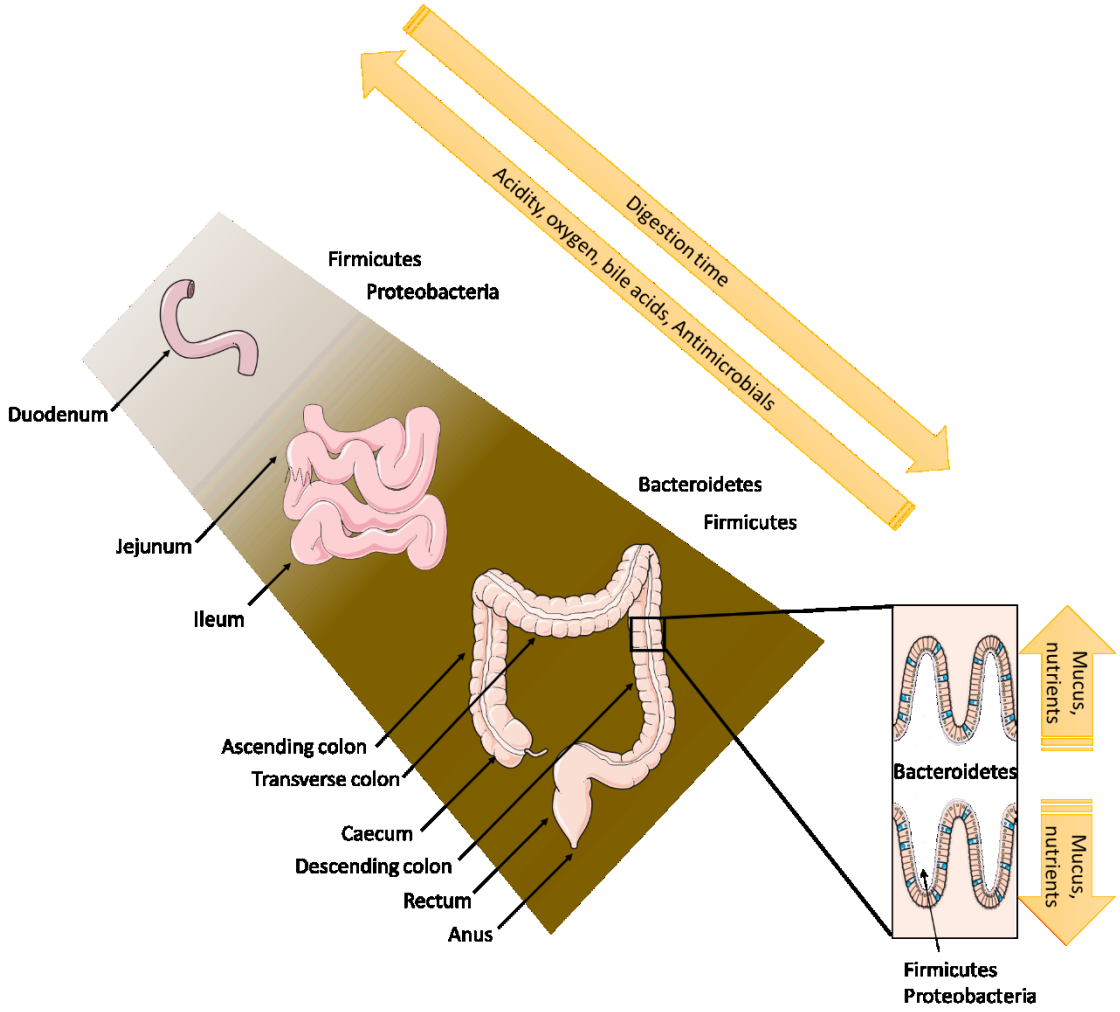



**Figure 1-1. Exposed view of the gastrointestinal tract, along with its physiological properties and most abundant phyla.** Adapted from **(Donaldson et al., 2016)**.

## 1.2. Gut microbial influence over the host metabolic system

The gut microbiome encodes enzymes with a vast range of metabolic potential which can influence the degradation and absorption of nutrients and vitamins (Kau et al., 2011). One of the most important metabolite classes produced by the digestion of fibres and complex carbohydrates, are short-chain fatty acids (SCFAs), such as acetate, propionate or butyrate. Firmicutes are the main microbes responsible for SCFAs production making them available as an energy source for other microbes and most importantly for colonic epithelial cells to ensure their growth. SCFAs impact the intestinal transit, lowering the pH of the colon and inhibit pathogen growth while promoting water and sodium absorption, as well as being involved in metabolic pathways such as cholesterol synthesis (Nicholson et al., 2012).

The production of metabolites by the gut microbiome naturally increases the diversity of the dietary compounds available to the host metabolic system. Choline metabolism by bacterial enzymes allows for instance the production of various choline-derivatives, such as methylamine, dimethylamine, trimethylamine (TMA) or TMA-*N*-oxide (TMAO), through the action of *Faecalibacterium* or *Bifidobacterium* genera. These compounds influence lipid and glucose metabolism. Phenolic derivatives also give rise to a broad range of metabolites, such as cinnamoylglycine, phenylacetylglycine (PAG), or catabolites of tyrosine, such as *p*-hydroxyphenyl acetic acid, *p*-cresol and its host-derived conjugates *p*-cresol glucuronide and *p*-cresol sulfate. Firmicutes have been demonstrated to be involved in the production of phenolic compounds. Similarly, indole derivatives (indole, indole acetic acid, serotonin) are produced by Firmicutes and Proteobacteria and are involved in gut-brain exchange, development of the epithelial cell barrier and regulation of the immune system. Polyamines, such as putrescine, spermidine, are produced by the same phyla as indoles and also influence the inflammatory response of the host. The list of small molecules produced by the gut microbiota extends to a wide array of metabolites including succinate, urea, lysine, hippurate and many others (Nicholson et al., 2012).

Bile acids such as cholate and other cholesterol derivatives facilitate the metabolism and the absorption of dietary fat, lipids and vitamins as well as serving as signalling molecules that can regulate host energy, lipid and glucose metabolism. The gut microbiome can also synthesise vitamins, such as vitamin B12 (cobalamin), vitamin B5 (pantothenic acid), niacin (vitamin B3), riboflavin (vitamin B2), biotin, vitamin K, as well as the active form of vitamin B6, pyridoxal phosphate and tetrahydrofolate, cofactors involved in amino acid reactions.

## 1.3. The gut microbiota in health and disease

This metabolic contribution of the gut microbiota has been implicated in disorders such as insulin resistance (Dumas et al., 2006), diabetes (Blandino et al., 2016) and obesity (Gérard, 2016), cancer, as well as neurodegenerative and neurodevelopmental diseases, such as Parkinson's disease (Mulak and Bonaz, 2015), Alzheimer's (Jiang et al., 2017) and autism (Ding et al., 2017). However, the specific role that the intestinal microbiota play across these pathologies is often complicated to determine, firstly due to the complex structure of the gut microbiota and its interference with the host metabolic and cellular profile. It is thus often difficult to identify if an altered microbiota is a cause or a consequence of a disease. However, perturbation of the gut microbiota has certainly been observed in a large number of diseases (Knight et al., 2017), and should, as a consequence, be explored further.

Although others would argue, animals, and especially rodents, have been a useful tool to explore the role of the gut microbiota in health and diseases. Antibiotic treated mice are useful when it comes to suppressing microbial-metabolic pathways and observing the downstream consequences on the metabolic profile of the host or on the metabolic fate of a xenobiotic (example in Chapter 5). The use of germ-free and gnotobiotic animals, which respectively, harbour no bacterial strains or only known bacterial strains, are common tools used in microbiome research. These models enable the inoculation of the rodent GI tract with human faeces, directly after collection, or following culture, to manipulate its genetic material to create a so-called "humanized mice", which can be especially valuable in personalised gut culture (Goodman et al., 2011) or used to explore the complex interaction between diet and gut microbiota (Turnbaugh et al., 2009b). Finally, the transplant of organ or tissue from human to rodent, called a xenograft model can also be used. It is particularly promising in cancer research, as demonstrate a study where tumour organoids were implanted into mice (Pauli et al., 2017). This model enables drug screening to be directly validated *in vivo* and could lead to personalised cancer treatment.

## 1.4. Factors influencing the gut microbiota

The microbiome is sensitive to environmental effects or even psychological factors, often referred to as the exposome, which includes factors such as the host diet, physical activity, exposure to pollutants and therapeutic treatments (Falony et al., 2016). The influence of these factors is particularly great during the development of the infant intestinal microbiota, as the establishment of adult microbiome occurs predominantly during the first three years of life and these initial colonisation trajectories have potential long-term ramifications. The mother's microbiome is the first of those factors as during birth, the mother's vaginal and faecal microbes are passed to the infant, dominated by *Lactobacillus*, *Prevotella* and *Sneathia* species. These communities are replaced by skin-related bacteria, especially *Staphylococcus, Corynebacterium* and *Propionibacterium* species for caesarean-born infants (Dominguez-Bello et al., 2010). Other factors can influence the development of the infant microbiome such as perinatal antibiotic courses, length of gestation (*e.g.* premature birth) and weaning age, as well as if the new-born is breastfed or formula fed (German et al., 2008). Indeed, bacterial diversity and composition of the infant gut microbiota is correlated with breastmilk intake (Pannaraj et al., 2017) and increased *Bifidobacteria* compared to formula fed babies, presenting more facultative anaerobes (Fanaro et al., 2003). In all cases, early-life microbiome development appears to influence the long-term functionality of the infant-microbiome and demonstrates a role in metabolism and immunological protection (Mueller et al., 2015).

Modern western lifestyle has played a key role in the development and evolution of the human microbiome. Comparison of the microbiome of a population leaving in an isolated Amerindian village (presumed to have had no contact with civilized population for 11,000 years) with the microbiome of a US cohort, showed that the microbiome diversity and functionality of the former was the richest ever reported in human metataxonomics studies, with a higher level of *Prevotella* and lower level of *Bacteroides* (Clemente et al., 2015). In addition, 28 gene markers of antibiotic resistance were found in their microbial profile (such as the *cblA* gene encoding for extended-spectrum β-lactamase enzymes), although no antibiotic treatments or contact with antibiotic-treated persons were known. This suggests that antibiotic overuse affects human microbial evolution by uncontrolled environmental contamination (Martinez, 2009; Danner et al., 2019). The role of industrialization in the long-term modification of the gut microbiota has been reported and associated with lower amounts carbohydrate-active enzymes and a higher abundance of gene markers of bacteria for mucus degradation such as *Akkermansia muciniphila* (Sonnenburg and Sonnenburg, 2019). Biogeographical differences in infant intestinal microbiota have also been observed and assumed to be linked to breastfeeding and lifestyle habits between two cohorts, one more industrialized compared to a more rural one (Lawley et al., 2019). Consequently, the evolutionary context of the microbiome, through industrialization and diet habits for instance, is also important to take into consideration (Davenport et al., 2017).

## 1.5. Inter- and intra-individual variability of the gut microbiota

As the factors impacting the gut microbiota composition and diversity include diet, life-style, as well as pollution, natural intra-individual and biogeographical gut microbial variations occur (Kau et al., 2011; Falony et al., 2016). Eckburg et al., 2005, found that the vast majority of the microbiota in the caecum, colon and faeces was dominated by bacteria from the Bacteroidetes and Firmicutes phyla. Although diversity and variability across individuals was reported, this study was only performed on three individuals (Eckburg et al., 2005). In contrast, a study performed on 124 Europeans showed that 40 % of the genes were in common with 50 % of the rest of the cohort, and that 57 bacterial species were shared by more than 90 % the entire cohort (Qin et al., 2010). Interestingly, the authors reported that about 1,150 bacterial species were detected in their experiment and that deeper sequencing would not increase considerably this number due to the redundancy of the bacteria found in the different individuals. Yet, it remains unknown what a “healthy microbiota” looks like. In one study, only 14 genera out of 664 detected were found to be present in 95% of 4,000 individuals measured. Interestingly, the extensive metadata combining 69 covariates (related to diet, phenotype,

medication, birth, etc.) explained only 15% of the overall variance in the microbial profiles (Falony et al., 2016). This suggests that the number of covariates commonly investigated, and assumed to affect the intestinal microbial composition, is largely underestimated. The factor which was found explaining the higher percentage of microbial variation (around 10 %), was therapeutic treatment (Falony et al., 2016).

# 2. Drug metabolism

## 2.1. Xenobiotic metabolism by the host

The term xenobiotic describes all molecules alien to a living organism, such as drugs or environmental pollutants and toxicants, which may influence the endogenous biomolecular reactions with either no net effect, potential beneficial, or toxicological effects on the organism. Once in the host, xenobiotics can be subject to two different types of metabolism, simplistically described as "phases" (Figure 1-2).

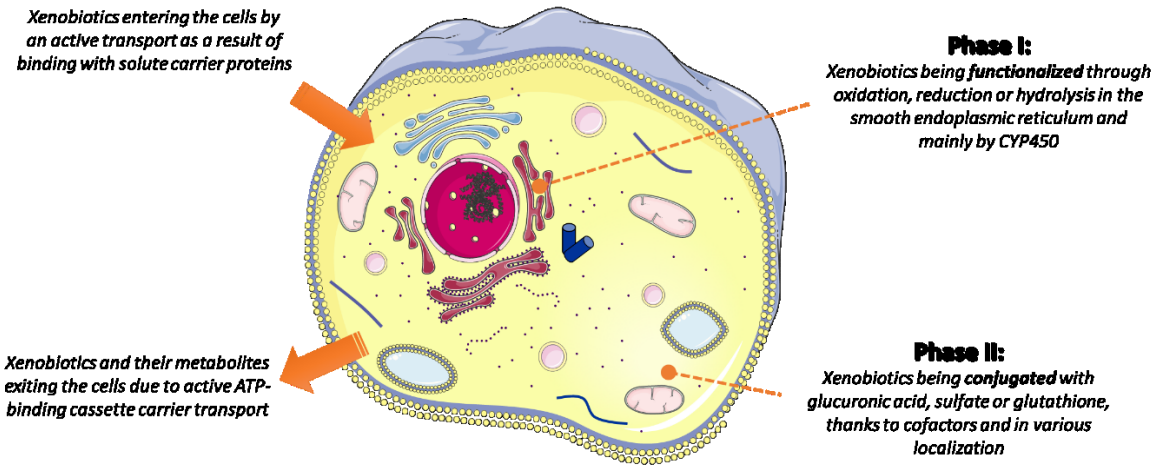



**Figure 1-2. Xenobiotic transport (represented on the left side) and metabolism (represented on the right side) across an animal cell.**

The two most common phases of xenobiotic metabolism are functionalization and conjugation. Functionalization reactions (phase I) encompass oxidations, reductions and hydrolytic biotransformations. Many of these occur in the smooth endoplasmic reticulum (ER) where cytochrome P450 (CYP450) enzymes are found. The CYP450 system presents a broad range of enzymes mainly performing oxidations (monooxygenase, dehydrogenase) and some reductions. Other enzymes which are not part of the CYP450 family are also able to produce functionalization reactions (*e.g.* esterase, oxidase). Conjugation reactions (phase II) can then be performed on the functional group that has been introduced, or exposed after an oxidation or hydrolytic biotransformation reaction, or if a xenobiotic already has a suitable functional group that can be conjugated. The main conjugation reactions are with glucuronic acid, sulfate or glutathione. These conjugation reactions require specific enzymes, such as uridine 5'-diphospho-glucuronosyltransferase (UDP-GT), sulfotransferases and glutathione *S*-transferase respectively, as well as cofactors, UDP-glucuronic acid, 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) and glutathione, respectively. UDP-GT is localized in the ER, with its catalytic domain exposed to the luminal side of the membrane structure, whilst the sulfotransferases are cytosolic. Acetylation glutathione *S*-transferases (GSTs) are found in many subcellular locations including the cytosol, mitochondria, the ER, nucleus and plasma membranes whilst amino acid conjugation occurs in the mitochondria. The movement of xenobiotics into the cells occurs by the action of active transporters called solute carrier proteins. The movement of xenobiotics and their metabolites following Phase I and Phase II reactions, out of the cells, is active via adenosine triphosphate (ATP)-binding cassette transporter proteins, before being mainly excreted in the bile or urine (Dobson et al., 2009; Kell, 2016, 2015).

## 2.2. Enterohepatic circulation

The vast majority of drugs and other xenobiotics are orally ingested, although dermal, respiratory, vaginal and rectal routes can also be used. They first cross the mouth, the pharynx and the oesophagus to reach the stomach. Once in the stomach, acidic pH promotes the dissolution of xenobiotics so that they can be transported to the lowest part of the GI tract, starting with the small intestine. In the small intestine xenobiotics can be metabolised in enterocytes but also come into contact with the gut microbiota. According to their permeability and solubility, xenobiotics and their derivatives can make it through the intestinal epithelium of the small intestine (Amidon et al., 1995). However, even the most commonly used drugs do not combine the optimal permeability and solubility properties to be easily or entirely absorbed (Dahan et al., 2009). As a result, many are not completely absorbed and a portion of the dose enters the large intestine, composed of the caecum, the colon and the rectum, before being re-absorbed or excreted through the rectum. Those compounds which have been absorbed through the small intestine, travel to the liver via the hepatic portal vein. It is in the liver that most xenobiotic metabolism occurs, when the compounds enter hepatocytes (although the gut wall, kidney and lung may also be important in determining the metabolic fate of certain drugs). The metabolic reactions that xenobiotics undergo up until that stage, either in the small intestine, gut wall or the liver, are called “first-pass” metabolism. This first-pass metabolism decreases the proportion of non-metabolized xenobiotics that eventually reaches the systemic circulation. Once metabolised by the hepatic system, the drug and its metabolites are available for either urinary excretion via the kidneys, with or without further additional metabolism in the podocytes, or enter the enterohepatic circulation to be distributed to the lymphoid organs, which includes the gall bladder. When biliary excretion occurs, drugs and/or their metabolites re-enter in the intestines and can either be directly excreted in the faeces or be re-absorbed and brought back into the liver and the enterohepatic circulation. During its second passage through the gastrointestinal tract, enterocytes and the gut microbiota can affect the drug metabolism for the second time (Figure 1-3a).

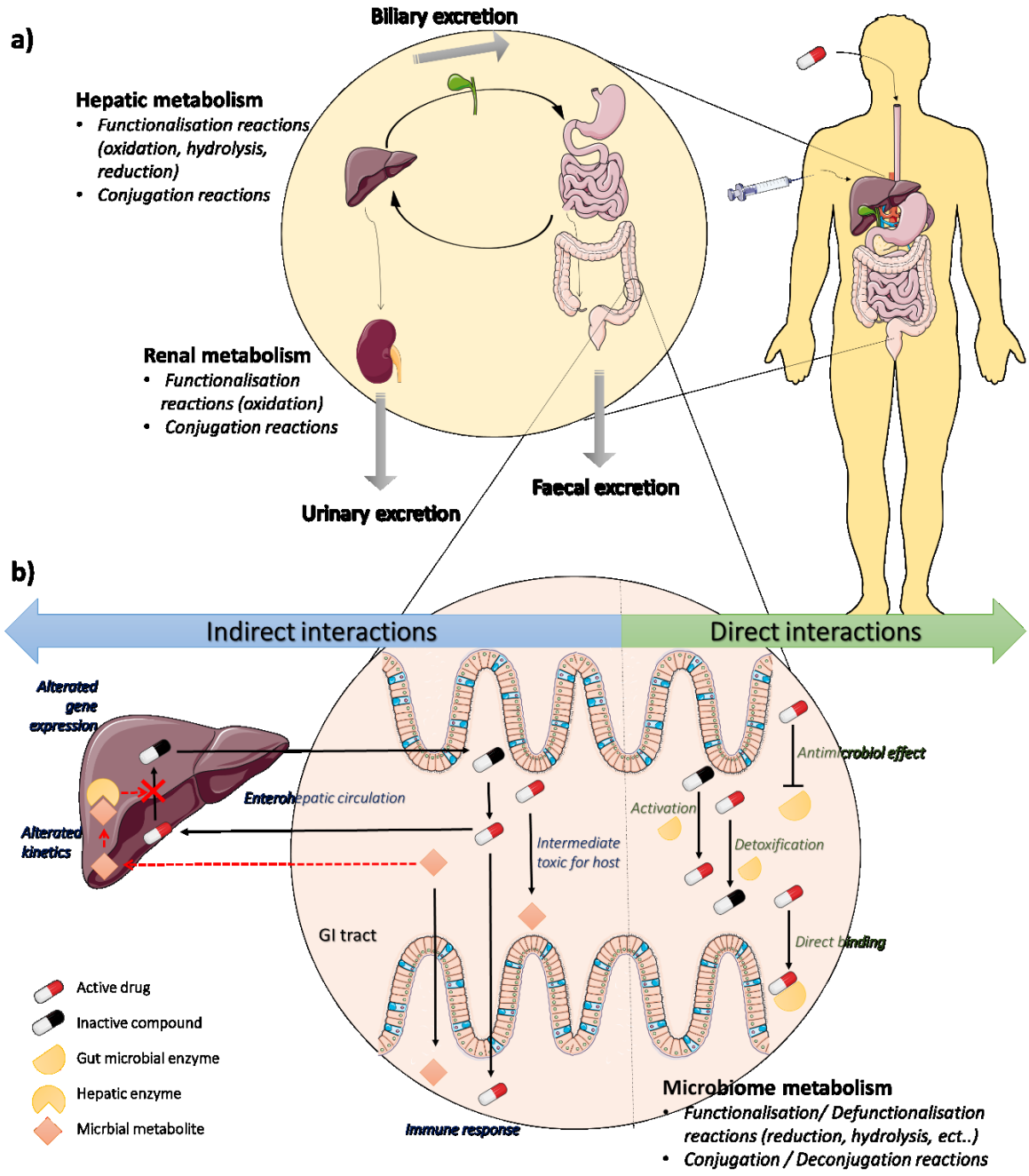



**Figure 1-3. Schematic representation of a) enterohepatic circulation and the different molecular reactions occurring on xenobiotics in the liver, the kidney and in b) the microbiota located in the gastrointestinal (GI) tract.** Adapted from (Haiser and Turnbaugh, 2012; Carmody and Turnbaugh, 2014; Maier et al., 2018).

Studying drug absorption, distribution, metabolism and elimination (ADME) is called drug disposition. It allows for a deeper understanding of how a drug is affected by the host organism and distributed in the organism. The distribution and elimination of the drug can be monitored by studying how drug concentrations (*e.g.* in blood) vary through time after administration (this is known as pharmacokinetics). It is used to understand the relationship between its concentration in different organs and its pharmaco- and toxico-dynamic effects on them. In pharmacodynamics, drugs can target either human-cells or bacterial-cells (Haiser and Turnbaugh, 2012). In pharmacokinetics, a broad range of factors can affect a drug and its metabolism, this includes host genetics (pharmacogenomics), polymorphisms associated with host genetic material (pharmacogenetics), and the host microbiome (pharmacomicrobiomics; Lam et al., 2019). Even though extensive research has been done to explore the biotransformations that drugs undergo when in contact with human-cells, the potential for microbial-cells to influence drug metabolism has only been gathering attention in recent years.

## 2.3. Role of the gut microbiota in shaping drug metabolism

Containing 150 times more genes than the human genome (Qin et al., 2010), the microbiome is able to produce enzymes and metabolites that can affect drug metabolism both directly or indirectly with the potential to modulate their toxicities and/or their efficacies (Figure 1-3b).

### *2.3.1. Direct interactions*

A broad range of reactions resulting from the interaction of the drug with the microbial population of the GI tract has been observed. An important reaction is hydrolysis, which often results in a more toxic product, but also acetylation/deacetylation, decarboxylation, dihydroxylation and demethylation (reviewed in Wilson and Nicholson, 2016). Reductions are also an important metabolic capability of the gut microbiota from a pharmacological perspective as these reactions can activate prodrugs. Thus, bacterially mediated reductions can release the active compounds in the organism to exert their pharmacological effect. Prodrugs are usually used to improve the bioavailability of the active ingredient, to avoid its absorption in the small intestine for instance and force a release in the colon. A good example of this is the prodrug sulfasalazine which releases the active compound 5-aminosalicylic acid following azo reduction by the intestinal bacteria (Peppercorn and Goldman, 1972). Inactivation or detoxification of the active compound can also be performed following a direct interaction with a bacterial enzyme as has been reported for digoxin, degraded by *Eggerthella lenta*, an Actinobacteirum, into the non-active metabolite dihydrodigoxin (Saha et al., 1983). Direct binding

between a drug and a microbe has also been observed for the dopamine precursor L-3,4-dihydroxyphenylalanine (L-DOPA- Niehues and Hensel, 2009). In this case, L-DOPA must cross the blood-brain barrier before it can be activated to dopamine in Parkinson's patients. However, it has been shown *in vitro* that L-DOPA can bind to the outer membrane of *Helicobacter pylori*, which could partly explain the difference in bioavailability across patients which received oral dose of L-DOPA (Niehues and Hensel, 2009). Finally, direct influence of the drug itself on the gut microbiota composition can occur, as a quarter of ~ 1,000 orally administrated drugs were recently reported to present off-target antibacterial activity (Maier et al., 2018; discussed further below).

### *2.3.2. Indirect interactions*

Within the indirect interactions occurring between xenobiotics and the gut microbiota, there is the creation of microbial intermediates which make the parent compound toxic. Although this is more often encountered with dietary metabolites, it has been demonstrated through a health scandal in China in 2008 that it could also be applied to xenobiotics. Melamine-tainted products precipitated and thus generated kidney stones and other complications which led to several deaths in the population. Research on the origin of those stones proved the presence of cyanuric acid, a structural analogue of melamine, produced from melamine, by the single species *Klebsiella terrigena* (Zheng et al., 2013). However, the dual presence of cyanuric acid and melamine causes their mutual precipitation in renal tubules and irrevocable downstream consequences (Kobayashi et al., 2010). Alterated kinetics due to microbial metabolites have also been reported, especially for the well-known anti-pyretic and analgesic drug paracetamol, showing higher level of the toxic metabolite NAPQi in the presence of high concentrations of the microbial metabolite *p*-cresol (see Chapter 5). Modified gene expression can also be induced or controlled by a drug, even central drug-metabolism enzyme such as increased expression of CYP450 in the liver following the activation of specific enzymes or transporters involved in the detoxification of xenobiotics (Toda et al., 2009). Finally, intestinal microbiota can also impact on the enterohepatic recycling of xenobiotics and their metabolites. For instance, the active form of the chemotherapeutic agent, irinotecan, a hydroxylated metabolite, is conjugated to glucuronic acid before biliary excretion. However, once the conjugated drug reaches the GI tract, bacterial β-glucuronidase cleaves back the glucuronic acid and releases the active compound in the GI tract, where it is toxic for the gut, causing dose limiting diarrhoea, and enterohepatic recycling (see Chapter 6).

Haiser and Turnbaugh, 2013, reported that the impact of the gut microbiota on drug metabolism has been investigated for only 40 drugs. However, *in vitro* tools have also been developed to study the microbial metabolism of 271 oral drugs (Zimmermann et al., 2019a). This study was able to show that more that 65 % of these oral drugs could be degraded by at least one bacterial isolate out of 76 tested, although this does not mean that such effects would be observed *in vivo*. All together, these results show the importance of considering the gut microbiota as a factor in inter-individual variability in drug response and hence in personalised medicine

## 2.4. Role of drugs in shaping the gut microbiota

As mentioned above, many factors such as the diet, or exposure to pollutants or drug therapy, can impact on the diversity and the functionality of the gut microbiota, which can have potential ramifications for host health. Medication has even been reported to explain the largest part of microbial variance in large cohorts (Falony et al., 2016). Drugs are in general divided into two categories based on their main molecular target group: (1) human-targeted drugs, which are intended to exert their pharmacological effects on human cells, and (2) antibiotics, aimed at targeting microbial cells (Haiser and Turnbaugh, 2012).

### *2.4.1.Antibiotics*

Maurice et al. (2013) investigated the effect of both antibiotics and human-targeted drugs on the human microbiota by combining flow cytometry coupled with fluorescence-activated cell counting, amplicon sequencing and metatranscriptomics. By looking at bacterial cell physiology, 75 % of antibiotic treatments increased bacterial cell damage. Moreover, by looking at the effect of the antibiotics on the microbial profile, it was observed that most of the bacteria which were affected (although in different ways) were from the phyla Firmicutes, which include the most prevalent and active bacteria in healthy human gut microbiota, especially *Clostridiales* (Maurice et al., 2013). Overall, antibiotic exposure increased the gene expression of bacteria involved in a broad range of pathways, such as transporter RNA synthesis and then translation, vitamin synthesis as well as phosphate transport. Interestingly, host-targeted drugs did not show any impact on microbial cell physiology but were still enhancing the expression of genes involved in membrane transport, the pentose phosphate pathway and xenobiotic metabolism (Maurice et al., 2013).

### *2.4.2. Drugs with antimicrobial off-target effects*

Subsequentially, other drugs have been reported to affect the composition of the gut microbiota, among others, methotrexate (Nayak et al., 2019), metformin (Forslund et al., 2015) and non-steroidal anti-inflammatories (Rogers and Aronoff, 2016). A third category of drugs has even been distinguished from the host-targeted drugs and the antibiotics (Maier and Typas, 2017). They include anti-infective drugs, which consist of antiprotozoal, antiparasitic, antiviral and antifungal compounds. The same authors developed a systematic method to explore the role of 1,197 oral drugs from a chemical library on intestinal microbiota (Maier et al., 2018). Their results proved that, using 40 common gut bacterial isolates, 24% of the human-targeted drugs and 58% of the anti-infectives could inhibit the growth of at least one bacterial strain. While the anti-infective drugs affected the growth of a restricted number of strains, the host-targeted drugs affected the most prevalent species commonly found in healthy human gut microbiota (e.g. *Roseburia intestinalis*, *Eubacterium rectale* and *Bacteroides vulgatus*). Moreover, these species are involved in important host-metabolism regulatory pathways such as short-chain fatty-acids. Such findings prove the necessity of considering the off-target effects that might result from host-targeted drugs on the microbiota and their long-term consequences on human health (i.e. microbial diversity reduction following long- or repeated exposure to drugs (Clemente et al., 2015)) and increased risk of antibiotic resistance induced by host-targeted drugs presenting a similar mode of action to that of antibiotics (Maier et al., 2018).

## 2.5. Challenge

In summary, the role that the gut microbiota plays in a drug metabolism is important to study to (1) understand how interindividual microbiome variability may dictate how the microbiota can impact the drug metabolism, its efficacy and toxicity, (2) establish how a drug, even a human-targeted, can shape the gut microbiota diversity and functionality and thus its ramification on the host health and (3) help to understand and use individual microbiota to personalize drug strategies, improve outcomes and limit adverse drug reactions. In order to study the complex microbiome-drug-host metabolic profile system, systems medicine was used.

# 3. Systems Medicine

## 3.1. Omics sciences

Systems medicine is an interdisciplinary area of research which considers human health as part of a global and complex system with different compartments (Figure 1-4). Since its application in the clinical setting it has been adapted from the term systems biology (Kirschner, 2016). It aims to correlate human phenotypes to their exposure to environmental factors (exposome), and to understand the biomolecular mechanisms occurring at different levels. One means of studying biomolecular mechanisms at different levels is based on exploring biological samples using different analytical techniques, gathered under the name of "-omics" sciences. The "-omics" sciences are top-down approaches that include global metabolic profiling (metabolomics), protein sequencing (proteomics), RNA-sequencing approaches and DNA-sequencing approaches. As opposed to bottom-up techniques, mainly based on experimental data, biotransformation rates and kinetics, top-down techniques aim to holistically capture information from across the different compartments of a biological system to understand how the system as a whole is functioning and responding to different stimuli (Bruggeman and Westerhoff, 2007). The results obtained following "-omics" experiments generate large datasets composed of numerous variables. The interpretation of these datasets involves the use chemometrics, based on multivariate statistical methods, in order to associate specific variables (*e.g.* a metabolite, a protein, a gene or bacterium) with a specific health condition, nutrition state, therapeutic treatment or exposure to environmental pollutants.

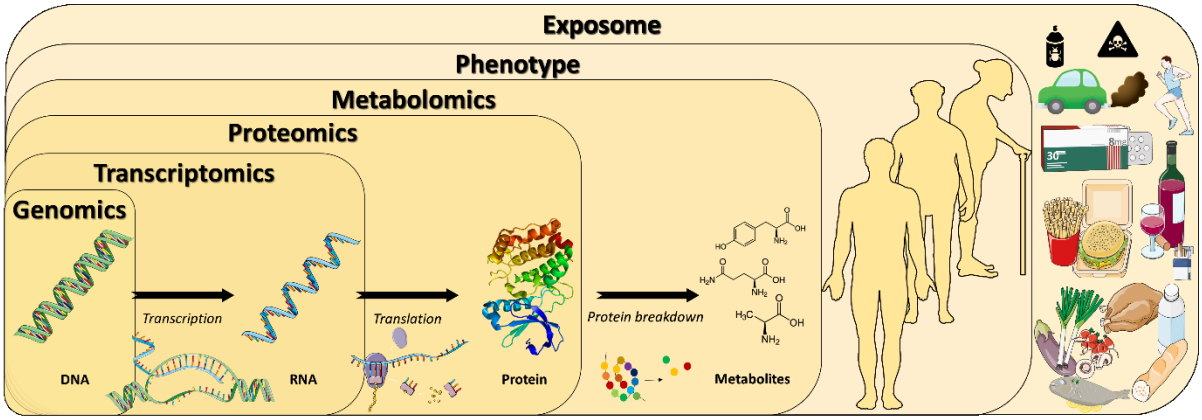



**Figure 1-4. Representation of what Integrative Systems Medicine research takes into consideration to study human health.** Chemometrics allow underlying differences, similarities or correlations between the "-Omics" sciences, phenotype and exposome to be found.

## 3.2. DNA-based omics

As a general rule, genomics is based on DNA sequencing. It combines two different techniques which are often confused. Although the term of metagenomics is often used for both techniques, herein the distinctions between amplicon sequencing, based on the sequencing of only one DNA sequence, that for the 16S rRNA gene, and metagenomics, based on the sequencing of all the genes encoded by the DNA, will be used here.

### *3.2.1.Amplicon sequencing*

Amplicon sequencing, also called 16S ribosomal RNA (rRNA) gene sequencing or metataxonomics, gives information about "*Which bacteria are present?*". The principle of amplicon sequencing is based on DNA extraction followed by a polymerase chain reaction (PCR) to amplify the 16S rRNA genes (Schmidt et al., 1991) then sequencing of the amplicon library. The 16S rRNA gene is present in all prokaryotes, and in consequence offers valuable insights into phylogenetic relationships. It is probably the most accessible technique for exploring microbial diversity and composition because of its modest cost. This technique has been applied in Chapter 3 and is further described in Chapter 2 (Analytical background).

### *3.2.2.Metagenomics*

Metagenomics, also called direct shotgun DNA sequencing, answers the question "*Which bacteria are present and what are their potential functions?*". This sequencing technique allows the sequencing of all microbial genes at once (Weber and Myers, 1997). It has proved to be more exhaustive than amplicon sequencing as it does not target only one gene, but looks at all the microbial DNA, the metagenome, in an untargeted way, which provides a much more complete picture of the microbiome of interest (Holst-Jensen et al., 2016).

## 3.3. RNA-based omics

Metatranscriptomics is based on messenger RNA (mRNA) sequencing. It is by far the most cost-extensive technique to explore the gut microbiota, or indeed any other biological sample. This technique provides an answer to the questions "*Which bacteria are present, what are their potential functions, and which ones are currently active?*". The output obtained with this technique is, in consequence, much more informative than amplicon sequencing or metagenomic sequencing, as it

gives insight into bacterial gene expression changes as a result of diet changes, xenobiotics exposure or diseases, or another stimulus. Metatranscriptomics has become quite popular and has been applied particularly to the nutrition-host-microbiome interaction in humanized mice (Turnbaugh et al., 2009b; McNulty et al., 2011) or humans (Hoyles et al., 2018). However most of the time, it is still used to complement amplicon or shotgun sequencing or, cell based techniques (Maurice et al., 2013), as well as metabolic profiling.

## 3.4. Metabolomics

### *3.4.1. Principles*

Metabolomics was defined as the determination of all the metabolites which are small molecules generally weighing under 1000 Da present in a cell (Fiehn, 2002). However, this definition has expanded to encompass all types of biological samples. It is based on high information content analytical methods, which allow a "snapshot" of all these metabolites to be acquired. Currently the two main methods are nuclear magnetic resonance (NMR) spectroscopy (usually proton [$^1$H]) and mass spectrometry (MS) which is coupled with separation techniques such as ultra-performance liquid chromatography (UPLC), gas chromatography (GC), capillary electrophoresis (CE) or ion mobility (IM), and allows a two- or three-dimensional separation of the analytes. Both NMR spectroscopy and MS techniques can be used in an untargeted or a targeted approach. The targeted approach, most commonly applied in MS-based techniques, is by definition more sensitive than untargeted approaches, as experimental parameters can be adjusted in order to focus only on few metabolite(s) of interest, which can be from the same class of metabolites or a particular biochemical pathway or a specific drug and its metabolites. The untargeted approach offers a broader, and less biased, view of all the metabolites present in a biological sample, up to the limit of the metabolites which can be lost during sample handling or preparation and limited by the sensitivity of the analytical method used. $^1$H NMR spectroscopy is known to be less sensitive than MS-based techniques, limiting the discovery of significant biomolecular changes and biomarkers. However, $^1$H NMR spectroscopy has the strong advantage of being a robust and repeatable technique, both through time and between laboratories (Ward et al., 2010). MS-based techniques, especially UPLC-MS, are somewhat less robust and although targeted assays are generally comparable between laboratories (Orellana et al., 2014; Nye et al., 2019) untargeted methods require careful quality control (QC, biological pool sample) procedures to be used to demonstrate robustness and repeatability over time (Gika et al., 2008; Zelena et al., 2009). In addition, as is well-known, mass spectrometers and chromatography equipment require extensive and meticulous maintenance to ensure reproducible results.

The term metabolomics is often used interchangeably with global metabolic profiling and metabonomics. Although, in practice, there is little or no recognized difference between the two first terms, metabolomics and metabonomics do differ in definition. Metabolomics is the study of the metabolic profile in a biological sample at a given time point to explore a specific condition (Fiehn, 2002; Goodacre et al., 2004), while metabonomics is the study of the changes in metabolic profile in an organism following a stimulus, such as disease, drug intake or toxins exposure (Nicholson et al., 1999; Lindon et al., 2003). However, most of the research exploring the global metabolic profile of an organism which underwent a nutritional, physiological, therapeutic or environmental change, used the term metabolomics. As such and to avoid any confusion the term used by the authors of the literature discussed will be used and metabolomics while discussing the results reported throughout this thesis.

### *3.4.2. Pharmacometabonomics*

As for metabolomics/metabonomics, pharmacometabolomics and pharmacometabonomics are often used interchangeably as terms which describe the use of metabolic phenotyping to predict the efficacy/toxicity of drug treatment on animals or patients (Clayton et al 2006, 2009: Kaddurah-Daouk and Weinshilboum, 2014). By comparing metabolic profiles before drug administration, the aim is to find features in the profile that accurately predict the response of a patient to treatment and, in consequence, promote the development of personalized medicine (Nicholson et al., 2010; Wilson, 2009). It has highlighted for instance that high urinary concentrations of *p*-cresol sulfate in humans were actually inversely correlated with lower concentration of acetaminophen sulfate, suggesting a competition mechanism between this microbial metabolite and the widely used drug (Clayton et al., 2009, see also Chapter 5). The metabolic profile obtained before and after a stimuli is increasingly used in parallel with pharmacogenomics in order to map metabolic and genomic variables together (Neavin et al., 2016), which has proved to be a useful combination in an early phase clinical trial (Oh et al., 2018).

### 3.5. Integration of the "-omics" sciences

As exemplified through the combined use of transcriptomics and genomics analysis, or the yet new but promising researches where pharmacometabolomics aids pharmacogenomics (Katsila et al., 2016; Manzoni et al., 2018; Neavin et al., 2016; Oh et al., 2018), the integration of different "-omics" datasets is a powerful method. It can be by combining two metabolomics datasets or methods together, such as targeted and untargeted UPLC-MS, which was found useful in "metabolome-microbiome wide association studies" (Melnik et al., 2017). It can also consists in the integration of untargeted UPLC-MS-based metabolomics data with DNA sequencing techniques of the microbiome (Zimmermann et al., 2019a). Taking into consideration all the different but complementary compartments of the "omics" science (Figure 1-4), genomics, transcriptomics, proteomics and metabolomics, as well as epigenomics or pharmacogenomics, promotes a better understanding of the biological mechanistic of a specific health condition. For Autism Spectrum Disorder for instance, a new patient-stratification has been described thanks to multi-omic researches (Higdon et al., 2015), which is a step closer towards personalized medicine. However, prior to their full implementation in biomedical sciences computational challenges will need to be addressed with respect to the integration of such large datasets and interpretation of the outcome with any certainty (Manzoni et al., 2018).

## 4. Thesis objectives

The general aim of the research described in this thesis was to gain a deeper understanding of the two-way biomolecular interactions which transpire between the intestinal microbiota and the metabolism of drugs, as well as the complexity of the biological interactions shaping the tripartite system drug-microbiota-host. As previously explained, a broad range of interactions can occur, from direct to indirect mechanisms of the microbiota on drug pharmacological and toxicological profile, to effect of drugs on the microbial communities' composition and functions through time. As a result, a multifaceted approach was adopted to explore these different types of interactions. In the end, the aim was to clarify the biomolecular mechanisms behind these interactions and the potential ramifications for host health. Such studies will aid in the assessment of the need to routinely implement taking the gut microbiota into consideration in clinical and pharmaceutical research aimed at promoting the development of personalized medicine.

The overall aims of this research summarized in Figure 1-5, are the following:

- To understand if the bidirectional interaction that the drug methotrexate presents with the gut microbiota, by directly interacting with bacterial enzymes through a detoxification process, with consequences affecting the structure of the microbiota, correlates with methotrexate toxicity.
- To develop an *in vitro* model able to monitor drug degradation by the human gut microbiota and explore the possibility of also determining the impact of the drug on the microbiome structure within the same model.
- To design a *in vivo* model enabling a clear mechanistic understanding of the impact of a high-tyrosine diet and induced high concentration of *p*-cresol on sulfation capacity and therein its indirect competition with acetaminophen to be obtained
- To determine the possibility of targeting the microbiota by inhibiting bacterial β-glucuronidase and reduce or supress the adverse drug reactions observed during irinotecan chemotherapy.

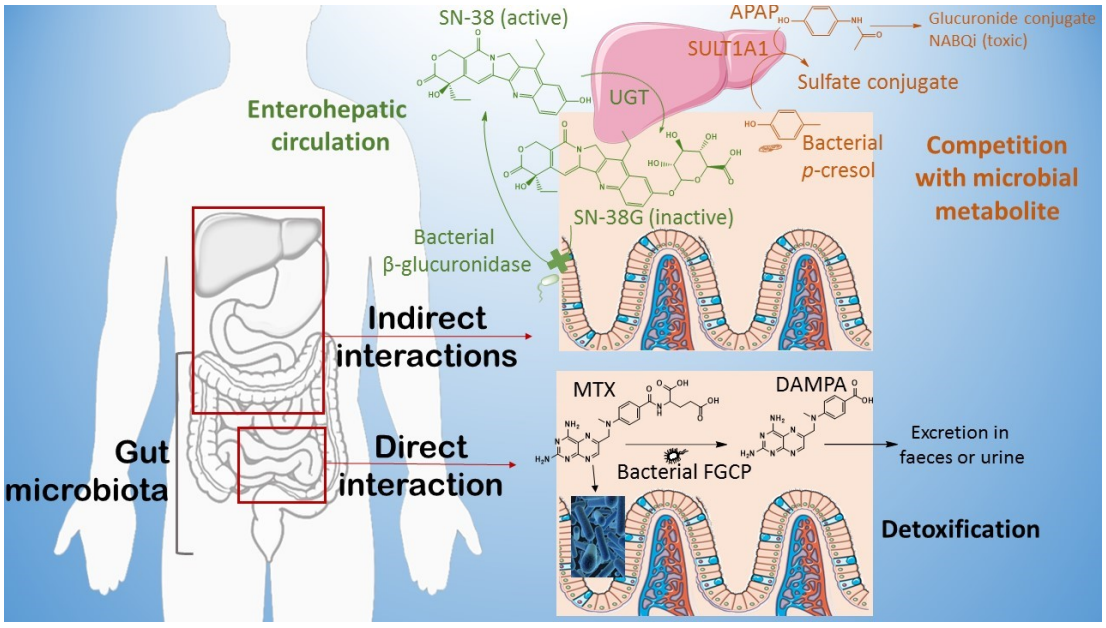
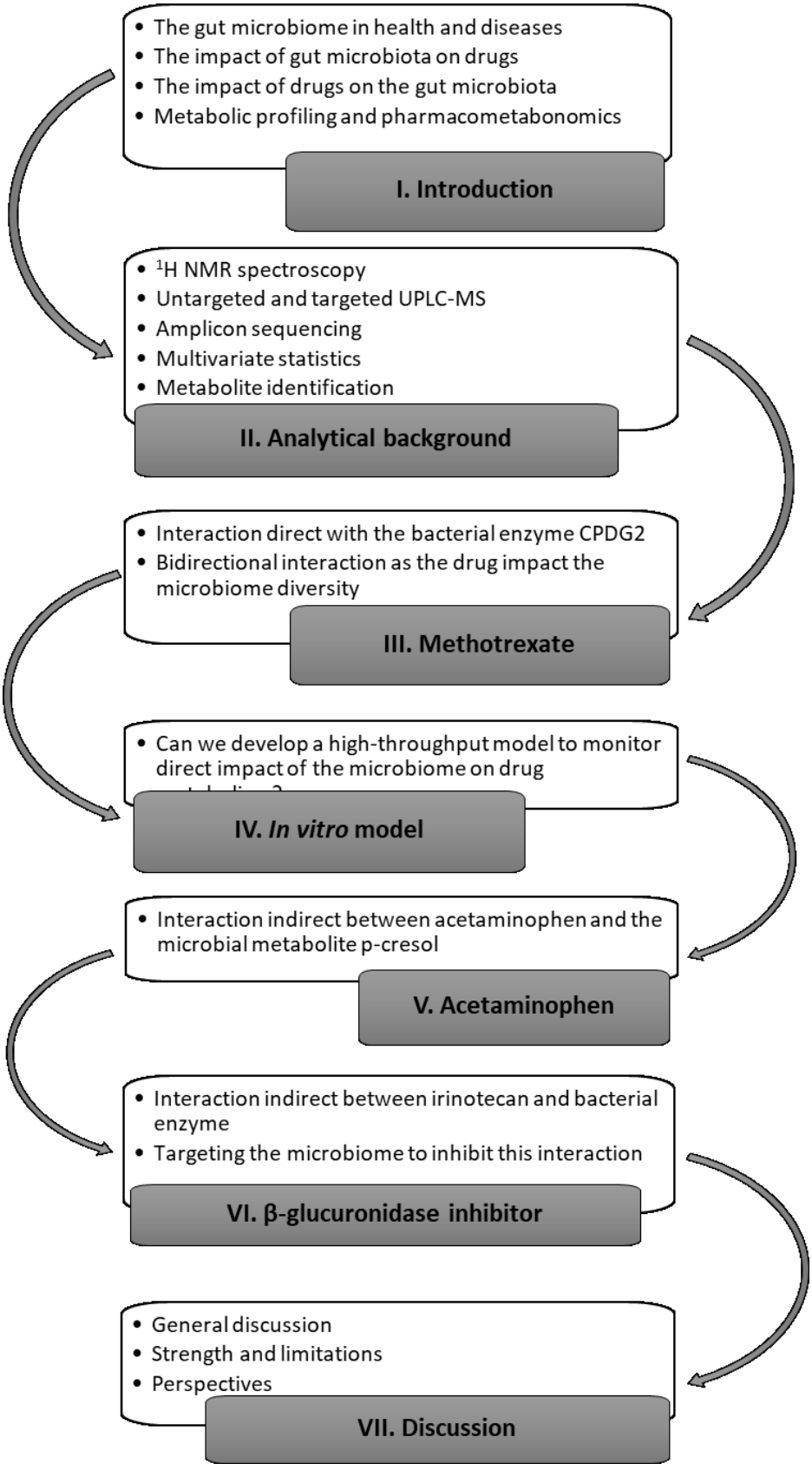



**Figure 1-5. Overview of the multifaceted approach adopted to deepen the understanding of the complex tripartite drug-microbiome-host health interaction. It involves direct interactions (methotrexate, MTX) or indirect interaction. The latter is exemplified by competition between acetaminophen and the microbial metabolite *p*-cresol for sulfation, and the potential benefit of targeting the microbiome to reduce irinotecan induced-toxicity.**

# Thesis outline



# Chapter 2. Analytical background

To conduct the research described in this thesis, a variety of omics techniques were applied. The techniques used within each project are summarised in Table 2-1. Metabolic profiling was performed using both ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) and $^{1}$H nuclear magnetic resonance (NMR) spectroscopy, and thus the theory behind each one of them are described in this chapter. Targeted UPLC-MS was also used to quantify drugs of interest (specific to each chapter) and to target a specific class of metabolites (tyrosine and a specific set of catabolites). A method to target those specific compounds was thus developed for a "fit-for-purpose" validation applied in Chapter 5. This validation was used as an example in this chapter to describe targeted UPLC-MS requirements. Finally, amplicon sequencing was also used in the Chapter 3, and in consequence its use is also presented in this chapter.

**Table 2-1. Summary of the analytical techniques use in each Chapter and on which samples it has been applied.**

| | ***Chapter 3*** | ***Chapter 4*** | ***Chapter 5*** | ***Chapter 6*** |
|---|---|---|---|---|
| ***$^{1}$H NMR spectroscopy*** | X | X | Mouse Liver<br>Cell supernatants<br>Mouse faeces | Mouse urine, plasma, liver, colon tissues, gall bladder |
| ***Untargeted UPLC-MS*** | Rat urine<br>Rat faeces | X | X | Mouse urine, faeces, liver, colon tissues, colon contents, gall bladder |
| ***Targeted UPLC-MS*** | Rat urine<br>Rat faeces | Human faecal culture | Mouse urine, liver and plasma | Mouse plasma, faeces |
| ***Amplicon sequencing*** | Rat faeces | X | X | X |

# 1. Metabolic profiling (Metabolomics)

## 1.1. $^{1}$H NMR spectroscopy

### *1.1.1. NMR spectrometer*

An NMR instrument is composed of three parts: (1) a superconducting magnet, into which the sample is placed, (2) a spectrometer, where free induction decay (FID) signals are detected and measured, and (3) a computer, from which the instrument is controlled and FID signals transformed into NMR spectra. The most complex part is the superconducting magnet, which provides a strong and uniform magnetic field at a specific resonance frequency. The $^{1}$H NMR experiments reported herein were performed on a 600 MHz spectrometer. The samples, in 5, 3- or 1.7-millimetre tubes, were inserted via the top centre of the superconducting magnet and lowered into the probe (Figure 2-1). The probe allows the sample to receive and emit radiofrequency pulses. Superconducting-wire coils envelop the probe to produce the magnetic field. However, to ensure the production of this field, it also requires a low temperature which is achieved by the use of liquid helium, which presents an absolute temperature of 4 K. To achieve this, an inner liquid helium tank encases the probe and the magnet, and an outer liquid nitrogen tank surrounds the liquid helium tank. The whole set up is enclosed in a vacuum chamber.

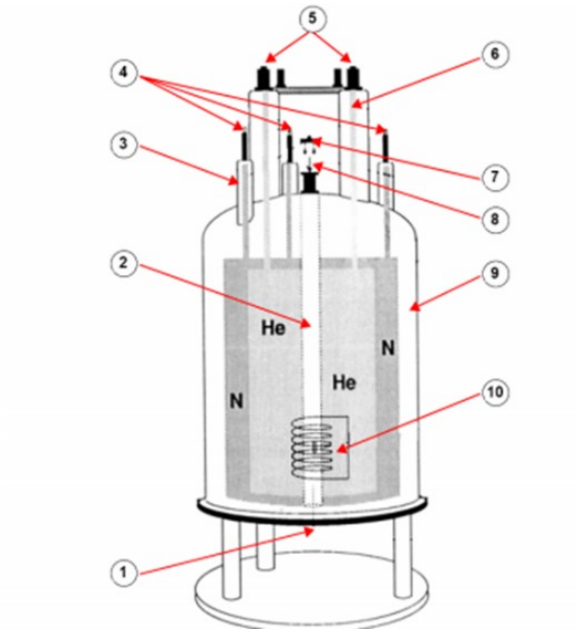



**Figure 2-1. A schematic for the superconducting magnet of an NMR spectrometer.** The different elements of the superconducting magnet are the following: (1) probe (2) bore, (3) nitrogen towers, (4) nitrogen ports, (5) helium ports, (6) helium towers, (7) metal plug, (8) sample insert, (9) vacuum chamber, (10) magnet. Source: AVANCE Beginners BRUKER Guide (version 006).

### 1.1.2. *Theory*

The principle of NMR spectroscopy is based on the physical and mechanical response of an atom nucleus to radiofrequency (RF) pulses which are applied in the presence of a magnetic field ($B_0$). Each atom nucleus is composed of protons and neutrons, which together define the spin quantum number ($I$) of an atom. Atoms, such as $^{1}H$ and $^{13}C$ that have a nucleus with an odd number of protons or neutrons, have a $I = ½$ and are capable to emit an electric current, called the nuclear magnetic moment ($\mu$), in the presence of a magnetic field. Common nuclei analysed by NMR are reported in Table 2-2.

**Table 2-2. Common nuclei analysed by NMR and their respective spin quantum number (*I*) which varies according their number of protons and neutrons.**

| Number of protons | Number of neutrons | Spin quantum number (*I*) | Common nuclei analysed |
|---|---|---|---|
| Odd | Even | 1/2 | $^{1}H$, $^{19}F$, $^{31}P$ |
| | | 3/2 | $^{11}B$, $^{35}Cl$, $^{79}Br$ |
| Even | Odd | 1/2 | $^{13}C$ |
| | | 5/2 | $^{17}O$ |
| Odd | Odd | 1 | $^{2}H$, $^{14}N$ |

In the absence of magnetic field, atom nuclei are present in a neutral and disorganised state (Figure 2-2). Once a magnetic field is applied, a nucleus aligns itself with the field. The number of different alignments that are possible for an atom is obtained by $2I + 1$. Each alignment is at a specific energy level which is described by the magnetic quantum number ($m_I$), described by the following rule:

$$m_I = I, I-1, I-2, \ldots . -I$$

Most commonly, metabolic profiling is based on the measurement of the $^{1}H$ nuclear μ. This is because hydrogens are prevalent in biological molecules with $^{1}H$ having a relative abundance of 99.98 %. As indicated above, the $I$ of $^{1}H$ is ½. In consequence, it can present two different magnetic quantum numbers in the presence of a magnetic field ($m_I = -½$ and $m_I = +½$). As shown is Figure 2-2, the negative magnetic quantum is explained by the nuclear μ spinning in an anti-parallel direction to the magnetic field (β state) and a positive magnetic quantum by a parallel spin (α state). The nuclear $\mu$ is proportionally associated with the momentum angle ($P$) of the spin against the $B_0$, according the gyromagnetic ratio (γ), specific to each nucleus:

$$\mu = \gamma P$$

The α state requires slightly less energy than the β state. The difference between these two energy levels is described as ΔE (Figure 2-2). It represents the energy that requires a nuclear $\mu$ to pass from one state to another.

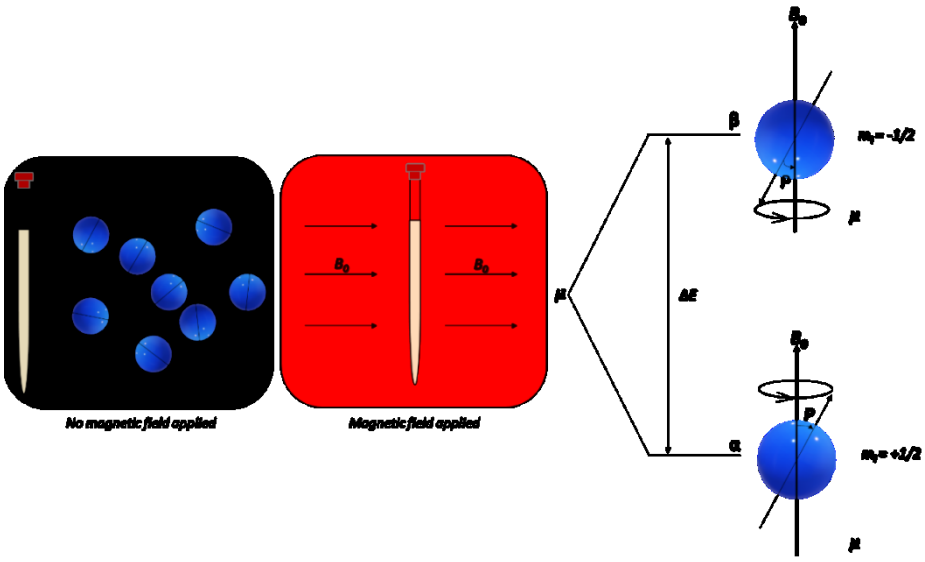



**Figure 2-2. $^{1}$H nuclear magnetic moment $\mu$ spinning around the axis the magnetic field $B_0$ at a momentum angle (*P*) in a β and α energy levels, respectively presenting a magnetic quantum number of - 1/2 and +1/2 and separated by an energy difference of ΔE.**

As an atom nucleus contains several protons, a certain number will be orientated in the α state ($N_\alpha$) and others in β state ($N_\beta$). There is a slight excess of nuclei in the α state and it is this excess that provides the signal to be measured. The distribution between each state can be calculated by the Boltzmann equation (with k being the Boltzmann constant and T the temperature in Kelvin):

$$\frac{N_\alpha}{N_\beta} = e^{-\Delta E/kT}$$

The ability for the nuclei to oscillate between the α and β states depends on the application of a specific $B_0$, as the energy difference (ΔE) between α and β is linked to the magnetic field by the gyromagnetic ratio (γ), which is specific to each nucleus, and the universal constant of Planck (h ~ *6.626 x 10$^{-34}$ J*s*). On the other side, the energy difference (ΔE) is directly proportional to the Larmor frequency at which a nucleus will oscillate. The $^{1}$H Larmor frequency being of 600 MHz and the $^{1}$H gyromagnetic ratio of 42.58 MHz/Tesla, the magnetic field that needs to be applied on samples to measure $^{1}$H NMR spectroscopy is equal to 14.1 Tesla.

$$\Delta E = h\gamma B_0 = h\nu$$

Once a magnetic field has been applied and the nuclei aligned, a short radiofrequency (RF) pulse is sent to the nuclei, which will orientate the spin at a 90 or 180˚ angle, according to the RF, which usually last for 2 to 10 µs. Once this pulse is applied, nuclei resonate with the RF energy and oscillate according its ΔE described above. Once the pulse is stopped the nuclei relax back to align with $B_0$ (Figure 2-3). The time taken for the spin to return to its original alignment to $B_0$ is called the relaxation time. During the relaxation time, the oscillations produced are registered as a structure-based signal, called a free induction decay (FID). The application of the Fourier transformation to the FID allows the instrument to convert the time-scaled FID signal into a frequency-scaled NMR spectrum.

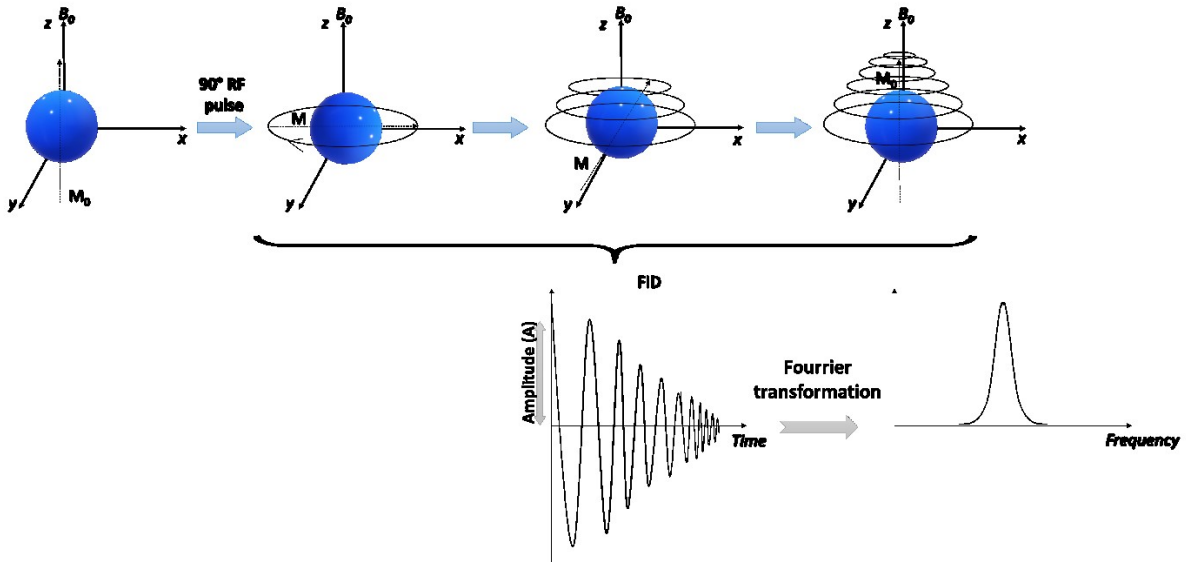



**Figure 2-3. Radiofrequency 90° pulse applied on an aligned spin to $B_0$ (initial position $M_0$).**

In a complex biological sample, each nuclear µ and thus nucleus FID will depend of the surroundings. Indeed, during its relaxation time reaching back the initial alignment with the z axis ($T_1$), or $B_0$, a nucleus releases its energy to the sample, which is called the spin-lattice relaxation. Nuclei from the other metabolites will also release their energy to the sample during their own spin lattice relaxation. The energy released by all the nuclei will create a local magnetic field (B'). This local magnetic field will influence the excitation state of the nuclei present in a sample, which is known as the shielding effect. The effective magnetic field perceived by the nucleus, called $B_{eff}$, is thus linked to the $B_0$ applied by the following equation (where σ is the shielding constant):

$$B_{eff} = B_0(1 - \sigma)$$

A second parameter depending on the environment of the nuclei will also influence their relaxation. Metabolites contain several atom nuclei that can interact with one another according to their electronegativity, as they are on the same scaffold. These interactions can reach up to three chemical bonds. In this scenario, the magnetic field created by a nucleus is sent to nearby nuclei and can interfere with the $B_0$, delaying the relaxation time in the x-y axis plan, also called spin-spin relaxation time ($T_2$). It has the consequence of splitting of the NMR signal obtained, which is defined by the multiplicity of the signal.

The spin-spin coupling and the multiplicity of the signal complies with the Pascal triangle. In that, it means that if a proton is less than three bonds away from n other protons, it will present a n + 1 signal. For instance, with ethanol (Figure 2-4), the $-CH_3$ directly bound to a $-CH_2$ will have a (2 + 1) = 3 signal, called a triplet. In return, the group $-CH_2$ group bound to the $-CH_3$ will have a (3 + 1) = 4 signal, called a quadruplet. The frequency distance separating each peak within a signal is called the coupling constant (*J*) and is the same between each peak within a signal/multiplet. The $^1H$ of the hydroxyl group does not present any coupling because oxygen being electronegative, it will attract the electrons towards itself, not allowing the small magnetic field B' to travel across three consecutives bonds. This results in a more shielded signal.

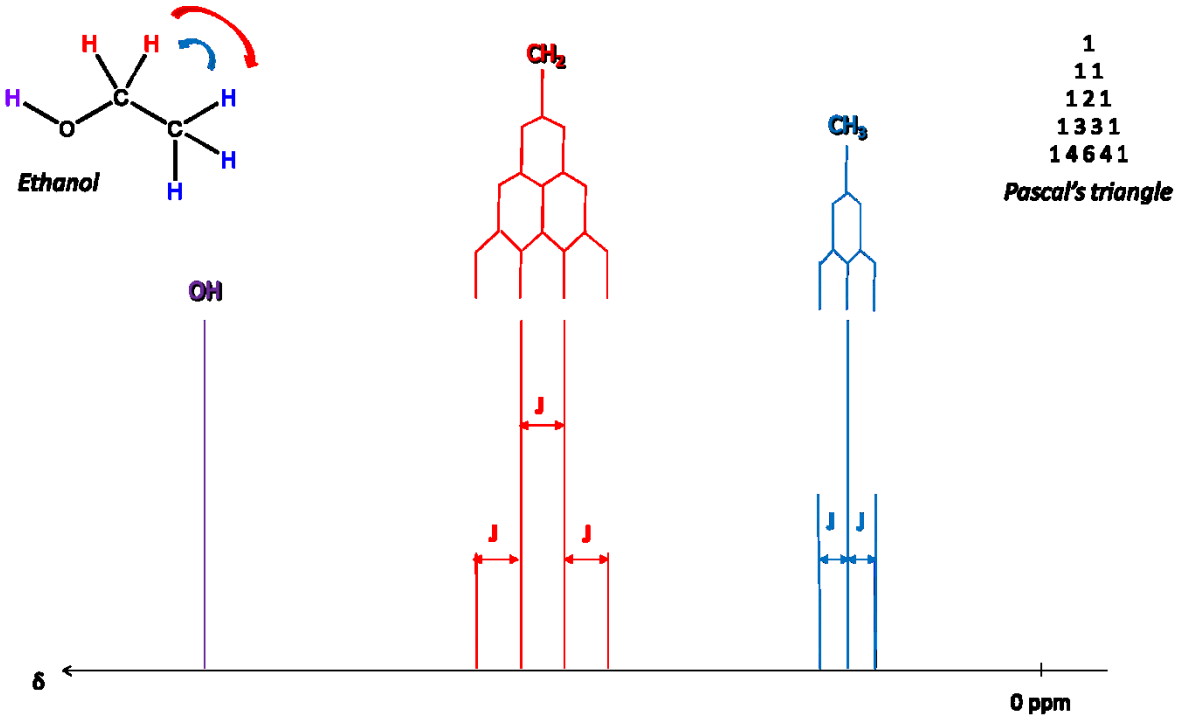



**Figure 2-4. $^1H$ NMR signal multiplicity principle according Pascal's triangle and exemplified with ethanol.**

Such observations are a valuable tool to determine the close atomic neighbourhood and elucidate molecular structure. This principle is routinely used in 1D NMR spectroscopy, where the signal frequency is measured on the x axis and its intensity as the area under the curve on the y axis. Nuclear Overhauser Effect Spectroscopy (NOESY), is the routine technique to acquire a $^{1}H$ NMR profile. The use of Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence is sometimes preferred when samples contain proteins and/or lipids, as these large macromolecules produce broad spectral peaks that often overlap and mask those arising from small molecular weight metabolites. The CPMG pulse sequence is able to selectively attenuate these peaks from the macromolecules enabling more metabolic information to be measured. This is particularly useful in plasma or liver samples for instance where lipoproteins and lipids are present. In complex biological mixtures, the large number of signals can easily become difficult to interpret and analyse. Indeed, signal overlapping between different atom nuclei and from different metabolites can easily occur. Identification of metabolites present can therefore be difficult. In this instance 2D NMR techniques can be used to assist metabolite identification. Here, the signal is registered across two frequency axes, x and y, and its intensity is measured as the volume under the 3D signal on the z axis. Typical 2D NMR techniques gathered two types of experiments, the homo- and the heteronuclear ones (Figure 2-5). The homonuclear NMR experiments explore the coupling that exists between the same type of atomic nuclei, such as two $^{1}H$. The two main homonuclear experiments that exist are correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY). COSY allow signals to be acquired that can go up to three bonds, and TOCSY to more than three (Figure 2-5). Heteronuclear experiments measure the energy exchange between different types of atomic nuclei, such as $^{1}H$ and $^{13}C$. Heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond correlation (HMBC), are the two main heteronuclear techniques. HMBC allow signals to be acquired for two atom nuclei directly bound and HSQC acquires a signal between atom nuclei separated by a larger number of bonds (Figure 2-5).

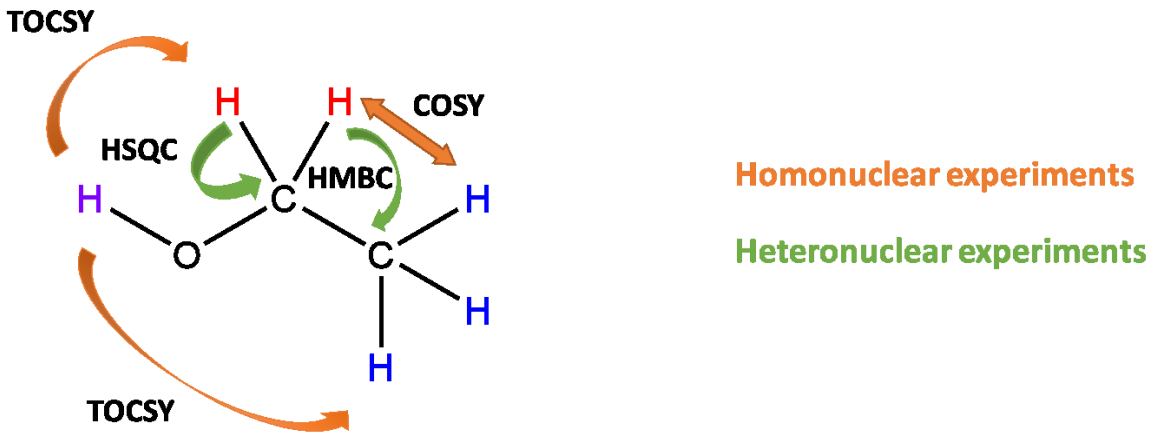



**Figure 2-5. Different 1D and 2D NMR spectroscopy methods.** COSY: correlation spectroscopy; TOCSY: total correlation spectroscopy; HSQC: heteronuclear single quantum coherence; HMBC: heteronuclear multiple bond correlation.

### *1.1.3. NMR spectrometer calibration*

Each NMR spectrometer has instrumental parameters that need to be adjusted before an analysis is performed. These include phasing, baseline correction and chemical shift calibration. To smooth the FID signals and increase the NMR spectrometer resolution, adopisation is first performed during the Fourier transformation. However, by doing so, it increases the signal/noise ratio. A correct threshold between resolution and signal/noise ratio needs to be determined prior to an experiment. Phasing is also done in order to obtain symmetrical signals, as well as baseline correction. Finally, chemical shift calibration needs to be done, particularly if the aim is to obtain quantitative data. The chemical shift (δ) measured by an NMR experiment is expressed in parts per million (ppm) and can be calculated by the following equation:

$$\delta = \frac{\left[\nu_{analyte}\,(Hz) - \nu_{ref}(Hz)\right] 10^6}{Spectrometer\ frequency\ (Hz)}$$

The chemical shift ($\delta$), is in consequence, always the same from one instrument to another, which confers to NMR spectroscopy its robustness and allows easy inter-laboratory method transfers and comparison. However, for that, as the frequency of the analyte is defined according the frequency of a reference, a universally recognised reference compound needs to be added to each sample. In aqueous samples, the preferred reference is trimethylsilylpropionic acid (TSP).

#### *1.1.4. NMR Data pre-processing*

Once the data have been generated, the raw data were imported into Matlab (version R2018a, The Mathworks Inc.), where an in-house script was used to pre-process and perform the multivariate statistical analysis. Different pre-processing steps need to be carefully performed. The first step is the removal of non-biologically relevant signals, such as those arising from residual water, the TSP peak, and sometimes solvent signals such as those for DMSO. Then, spectral alignment can be performed, either automatically or manually to adjust for subtle alterations in peak position due to variation in pH across samples. Normalisation can then be applied (Craig et al., 2006). Normalisation aims to remove any unwanted and systematic (technical) variation between the samples, which is unrelated to biological variation. This unwanted noise can be due to the nature of the sample (urine dilution), to the sample preparation or the analytical experiment. For urine, the objective of the normalisation is typically to remove the variation induced by differences in dilution across individuals. When it comes to cells or liver extracts, the starting weight of the material is consistent across all samples and so no dilution factor is expected, and normalisation should be unnecessary. However, experimental errors can be observed, such as errors linked to the sample preparation (*e.g.* pipette error, protein extraction, *etc*…). The most common normalisation, and the one used throughout this thesis, is median fold change, initially called probabilistic quotient normalisation. Its principle is based on dividing each variable by the median difference between the sample set and a reference spectra for each variable (Dieterle et al., 2006).

## 1.2. Untargeted UPLC-MS

#### *1.2.1. Liquid Chromatography*

*Principle.* The sample of interest is injected through a high-pressure liquid mobile phase which flows into a chromatographic column containing a stationary phase. A separation of the metabolites present in the sample occurs then, based on their physico-chemical properties. The separation depends on the variable affinities that the metabolites have with the stationary phase, where the analytes are more or less retained causing them to elute at different retention times (Tr).

*Choice of the chromatography*. Many kinds of chromatographic separation exist, but the main mode of separation employed in this thesis was reversed phase (RP) chromatography. RP-LC is widely used in metabolomics and lipidomics (a variant of metabolomics focusing only on lipids). Reversed-phase (RP), or partition, chromatography employs a non-polar stationary phase and a polar aqueous-organic mobile phase. In this mode of separation, polar compounds have more affinity for the mobile phase than the stationary phase and will in consequence be eluted first (green compound, Figure 2-6). Less or non-polar compounds will share more affinity with the stationary phase, will be retained on the column for longer. Untargeted metabolic profiling aims to cover as many metabolites as possible and RPLC is limited to medium polar to non-polar compounds. When the highest possible coverage is required, the combined use of RP with a separation mode such as hydrophobic interaction liquid chromatography (HILIC) will allow the range to be extended to polar compounds.

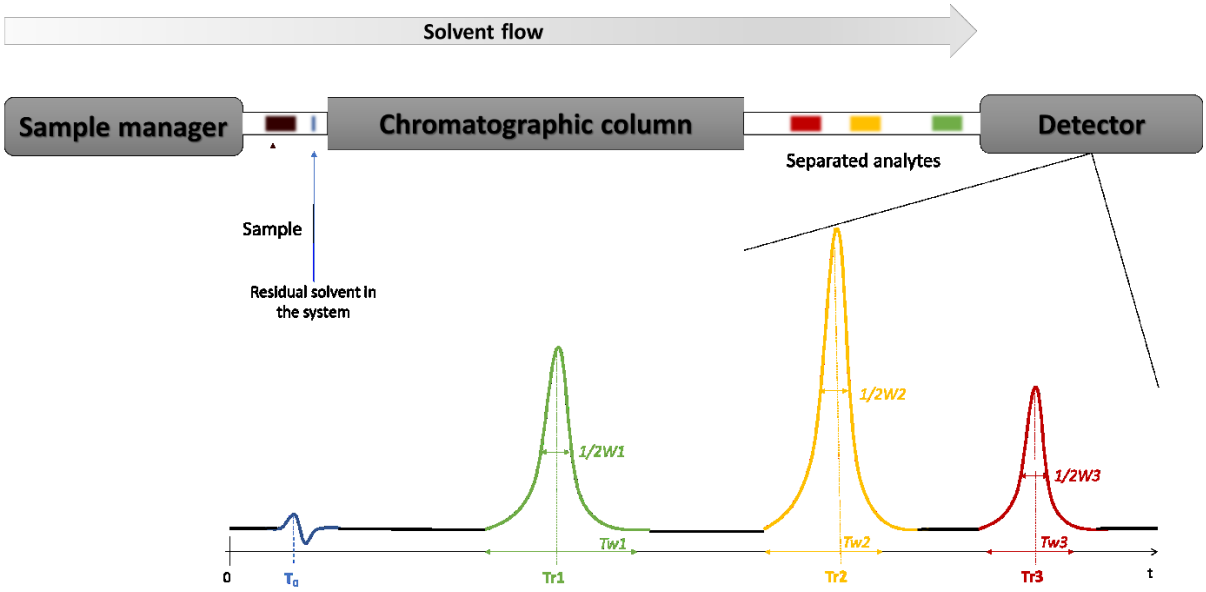



**Figure 2-6. Liquid chromatography principle, showing how the analytes contained in a sample can be separated after passage through the column followed by elution and detection.** Unretained compounds, eluting in the void volume, elute at a time $T_0$.

*Choice of the column.* The first parameter to select for the column is the size of the particles constituting the stationary phase. Ultra (high) Performance Liquid Chromatography (UPLC) uses sub 2-micron particles as opposed to High-Performance Liquid Chromatography (HPLC) which generally uses 3-micron, or larger, particles. UHPLC provides greater separation efficiency, although at the cost of much higher back pressures, and is in consequence preferred in metabolomics. Indeed, the use of UHPLC results in sharper chromatographic peaks and thus a better separation of the metabolites

thereby reducing ion suppression. The type of the stationary phase also matters. The main columns that are used for reversed-phase liquid chromatography (RPLC) are based on C18-bonded silica or, as mainly used in this work C18-bonded ethylene bridge hybrid (BEH) silica particles. The C18 phases provide good retention for "medium" polar and non-polar analytes. Alternative phases such as those based on the shorter C8 carbon have also been popular for the analysis of lipids, whilst the BEH phenyl, which retains aromatic compounds, has also found application.

However, these columns do not provide good retention of polar compounds. Poorly retained metabolites are likely to elute in the solvent front ($t_0$, Figure 2-6) where high levels of ion suppression are often observed, which can complicate or prevent the integration, quantification or identification of metabolites. The ideal retention is described by a retention time factor (k) between 2 and 10, which can be calculated according the following equation:

$$k = \frac{t_r - t_0}{t_0}$$

High strength silica (HSS) T3 columns are useful to retain more polar compounds and increase their chromatographic resolution (Rs) as they use highly bonded trifunctional C18 alkyl phase. A well resolved peak is usually defined by a resolution (Rs) superior to 1.5. It enables accurate chromatographic peak integration and thus quantification, as well as aiding metabolite identification. It can be calculated by the following equation:

$$Rs = \frac{2\,(Tr2 - Tr1)}{(Tw2 - Tw1)}$$

*Choice of the mobile phases and the gradient.* The mobile phases used for gradient RPLC are usually based on water as the weak eluent (solvent A) and an organic solvent (solvent B- *e.g.* acetonitrile, methanol, isopropanol) used to obtain elution. Gradient chromatography elution is achieved by changing the proportion of A and B during the time of the analysis (in RPLC going from a low to high percentage of B). Changing the proportion of A and B during the analysis changes the affinity of analytes initially retained by the stationary phase, which will result in partitioning into the mobile phase, exiting the column and entering the detector. A longer gradient with a lower slope can provide a better separation of the analyte. However, if too long, a gradient can result in broad peak shapes and reduces peak intensity. A compromise is often needed to obtain the required Rs with an

acceptable sensitivity. Modifiers are often added in the mobile phase to influence its pH, the solubilisation of some compounds and enhance ionization of the ion source of the mass spectrometer (see next paragraph). The most commonly used modifiers are formic acid (FA) and acetic acid (AA), usually added at a concentration of 0.1 % in each of the solvents constituting the mobile phase. Other parameters to select for the gradient are the temperature of the column and the flow rate of the mobile phase through the column. An increased temperature, by lowering viscosity and reducing back pressure, allows higher flow rates and enable a better peak resolution (Plumb et al., 2006). However, temperatures that are too high risk analytes being degraded (Fang et al., 2015). In consequence, a compromise needs to be found to ensure good peak shape, symmetry and resolution, while not losing any information.

#### *1.2.2. Mass spectrometry*

Once eluted from the chromatographic column, the analytes are introduced into the mass spectrometer. The separation of the metabolites is here based on the mass-to-charge (*m/z*) ratio of the analytes which can be ionized. The mass spectrometer is composed of three components. The first is the ion source, which ionises the analytes coming from the column. The second is the analyser, which separates the ions according to their mass-to-charge ratio. The third is the detector, which detects and counts all the ions produced by the source and separated by the analyser. Those three elements are contained in a single compartment maintained at high vacuum.

*Ion sources.* Numerous ion sources are used in modern mass spectrometry. These include electrospray ionization (ESI), desorption electrospray ionization (DESI), matrix assisted laser desorption ionization (MALDI), chemical ionization (CI), atmospheric pressure chemical ionization (ACPI) and rapid evaporative ionization (REIMS). These are generally considered to be soft ionization techniques (in contrary to hard ionization techniques, such as electron ionization (EI)) as the energy used to achieve ionization is not too high. This has the advantage that it avoids excessive fragmentation of the analytes of interest and thereby enables the observation of both the parent metabolites and the fragments, also called the fragment, or product, ions. The most used ionization method in LC-MS-based metabolomics, and the one used here, is ESI (Want et al., 2005), which can be applied in both positive or negative ionization modes. In global profiling, and to ensure a good coverage of the metabolites, samples are usually analysed using both positive and negative ESI ionization, as some metabolites ionize in only one mode, some in both and some simply do not ionize. Dual ionization mode can also be applied although it can cause a loss in sensitivity due to the time that it takes to switch the polarity back and forth in global profiling, but this can be less of a problem in targeted assays (see paragraph 1.3.).

Ionization in ESI occurs at the junction between the chromatography, where solvent and analytes are liquid and under atmospheric pressure, and the mass spectrometer, in which the solvent and analytes are transformed into gas phase and under vacuum conditions. It starts by the liquid eluting out from the column and reaching the ion source by exiting a capillary. At the surface of this capillary a high voltage is applied which results in the formation of a highly charged solvent cone (called the Taylor cone) which then forms charged droplets (Figure 2-7). Due to the presence of nitrogen, high temperature, and an electric field created between the capillary tip and the extractor cone (also called the ground plate or mass spectrometer inlet), the solvent evaporates from the charged droplets. This creates an increase in charge density at the surface of the droplets and induces repulsion charges inside them, causing dissociation into smaller droplets. This desorption then creates ions which enter the mass inlet. As the *m/z* ratio is proportional to the molecular weight of the analytes mass spectrometry in metabolomics is usually based on the study of a singly charged species, often the molecular ion (dimers or trimers can also be seen although these are less common). However, due to the presence of solvent or buffer ions (such as formate, acetate, ammonium, sodium or potassium ions), or ions originating from the glassware used to prepare or store solvent, the features can often be detected as adducts.

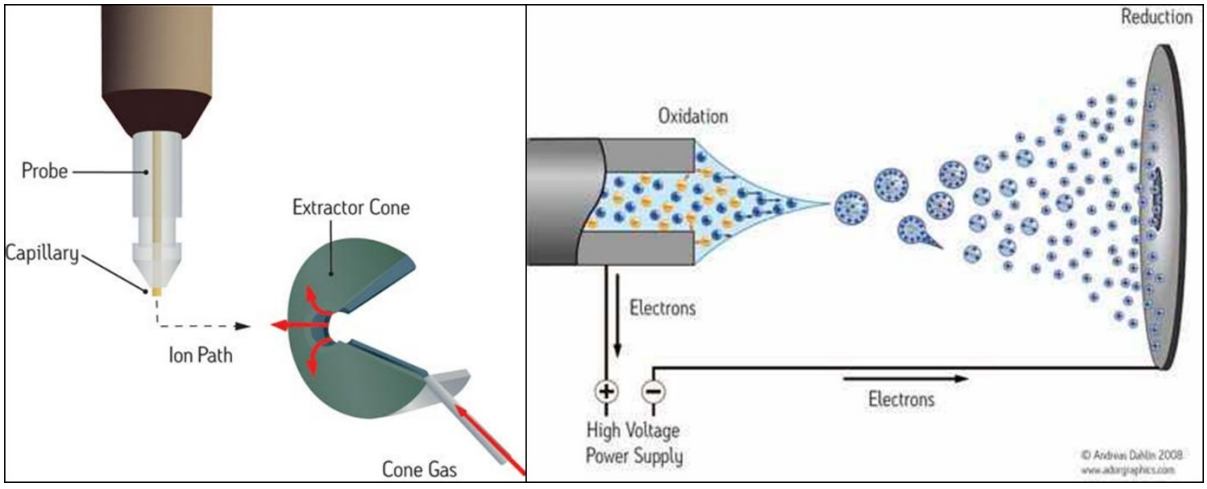



**Figure 2-7. Electrospray ionization (ESI) principle.** The mixture of solvent and sample reaches the ion source through the probe and the capillary, where a high voltage is applied to force the formation of droplets and their dispersion into ions, attracted by an electrical field to the extractor cone. Source: Common ionization, Waters Corporation website.

*Mass Analysers.* The second part of the mass spectrometer allows the separation of the ions according their *m/z* by applying an electrical or magnetic field in the mass analyser. According the type of analysers used, the *m/z* will be separated according their trajectory stability (quadrupole analyser), their resonance frequency (ion trap, Fourier transformation-ion cyclotron resonance - FT-ICR - or Orbitrap) or their velocity (time of flight of TOF). The different mass analysers are defined by several parameters that will describe how well the ions will be separated. These parameters are the mass range, speed of acquisition, sensitivity, resolution, and accuracy.

- The mass range describes the scope of the *m/z* ratios that can be detected in a sample during an acquisition (in metabolomics the usual range is from *m/z* 50 to 1200).
- The speed of acquisition is the time that a mass analyser takes to cover the mass range of interest.
- The sensitivity is determined by taking three parameters, which are the efficiency of the detector, the transmission and the duty cycle. The efficiency of the detector simply represents how efficient the detector can be to detect all the ions present. The transmission is the ratio between all the ions which reaches the detector and all the ions which entered the mass analyser. In contrast, the duty cycle is measured by the ratio between one single ion formed in the ion source and the proportion of the same specific ion being efficiently detected. In consequence, if only one ion is monitored, the duty cycle will be 100 %. If several ions are measured, the duty cycle will be decreased according to the time spent monitoring each ion (called the dwell time).
- Accuracy is defined by how different the measured *m/z* compares to the theoretical *m/z*. The accuracy is tightly linked to the resolution, as a mass spectrometer with a low resolution will have a low mass accuracy. The mass accuracy can be expressed in part per million (ppm) and calculated according the following equation:

$$ppm = \left( \frac{(\frac{m}{z})_{measured} - (\frac{m}{z})_{theoretical}}{(\frac{m}{z})_{theoretical}} \right) x\ 10^6$$

- Certainly, the most important parameter defining a mass analyser is its resolution, also called its resolving power (*R*), which is the ability to separate two ions of similar *m/z*. It can be calculated by measuring the mass of a peak divided by the mass difference to that of the *m/z* of the similar ion. However, it can also be calculated on a single peak by replacing the mass difference by the peak width at its half height (full width at half maximum or FWHM).

$$\mathrm{R} = \frac{\mathrm{m}}{\Delta \mathrm{m}} = \frac{\mathrm{m}}{\mathrm{FWHM}}$$

The instruments with which the resolution and the accuracy are the lowest are instruments using a quadrupole or a triple-quadrupole (TQ) mass analysers and which are generally used to perform targeted assay, i.e. to monitor known ions and their fragments (further explained in paragraph 1.3). However, quadrupoles are still used in untargeted global profiling, as they can be combined with another mass analyser, providing both resolution and sensitivity. This is notably the case for Q-Tof instruments which combine a quadrupole with a time of flight analyser. Instruments providing high resolution and mass accuracy include ion traps, Tof and Q-Tof's. Tof and Q-Tof mass spectrometers enable the acquisition of the maximum number of *m/z* measurements while measuring them with high accuracy. Such measurements facilitate the identification based on the precise calculation of the empirical chemical formula of metabolites and their fragments. These instruments also permit the use of fast scan speeds with a good dynamic range. The advantage of the Q-Tof compared to the Tof is that it offers the possibility of doing MS/MS (further discussed in the paragraph 1.5), which enables the acquisition of fragmentation patterns for specific ions and makes the identification of metabolites easier. Finally, ultra-high resolution can be achieved using orbitrap and FT-ICR mass analysers, which currently provide the highest mass resolution and mass accuracy, although these instruments are costly compared to QTofs and often more complicated to use, with slower duty cycles. The mass analysers used here for global profiling were Q-Tof analysers and an example of the ion optics of the Waters Xevo G2 Q-Tof is presented in Figure 2-8.

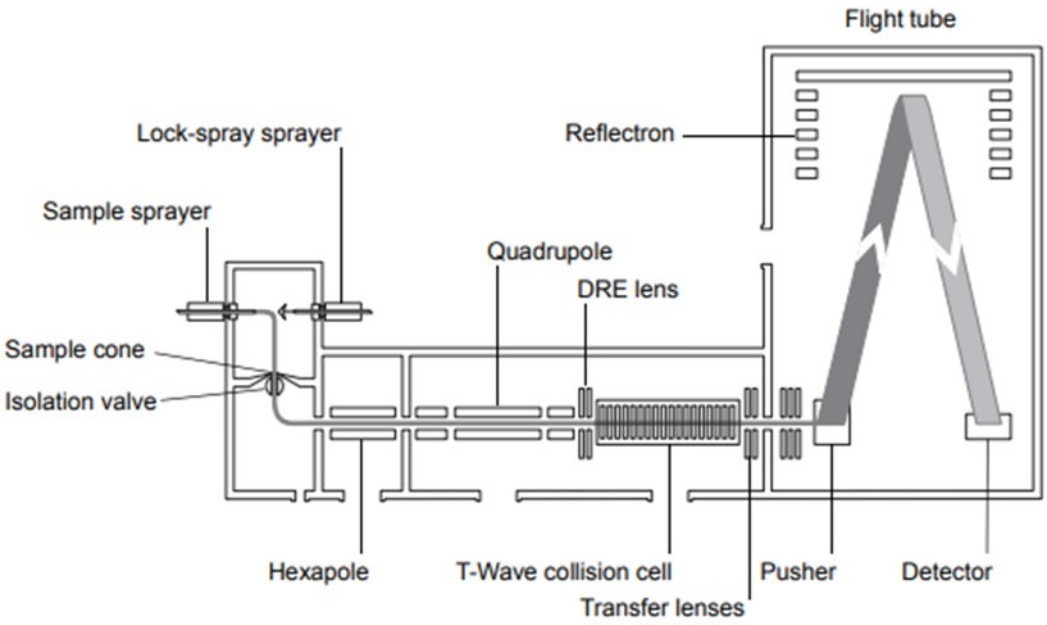



**Figure 2-8. Ion optics of the Xevo G2 Q-Tof mass analyser used in the Chapter 6.** Source: Waters Xevo G2 QTof Operator's Overview and Maintenance Guide.

Q-Tof analysers are, as explained above, constructed using two analysers, a quadrupole and a Tof, separated by a collision cell (Figure 2-8). The quadrupole is made of four hyperbolic rods and an electric field is applied on each of the parallel rod pairs. According to the electric field applied, a *m/z* filtration is made. Indeed, the trajectory of an ion will be influenced by the RF applied through the electric field and if an ion hits one of the rods its charge will be neutralised, and the neutral molecule formed will not be detected. However, if a specific RF is applied to maintain a stable trajectory for the targeted ion, only this ion will be capable to reaching the detector. While this can be a useful tool in a targeted assay, the aim of untargeted global profiling is to measure as many ions as possible. To achieve this, the electrical current applied between the two-opposite cross-pairs of rods oscillates, which will maintain the trajectory of ions over the mass range of the mass analyser. Once the ions are separated according their trajectory in the quadrupole, the ions will pass into in the collision cell (Figure 2-8) where collision-induced dissociation (CID) will occur due to the presence of a neutral gas, such as helium, argon or nitrogen. The collision between the ions and the neutral gas atoms will transform the kinetic energy of the ions into an internal energy which will cause the parent ion to fragment into a number of smaller ions. Finally, these fragments, as well as the non-fragmented parents, will reach the second analyser, the Tof (Figure 2-8). This analyser is placed at the exit of the collision cell, where the ions interact with an ion pusher before to be accelerated by an electric field towards a field-free flight tube. In this tube the ions are not separated by their energy differences anymore but by differences in their velocity, which defines the time that the ions take to accelerate and fly to the detector. Heavier ions will naturally reach the detector with a longer time of flight than lighter ions. This separation, in combination with the quadrupole separation and the fragmentation of the collision cell, allows the detector to measure ions and fragments with a very rapid duty cycle (5 to 30 times per second) with high resolution, which makes Q-Tof mass analysers valuable tools for metabolomics. Moreover, a reflectron, made of two lenses oppositely charged and forming an electrostatic reflective current (simply called an ion mirror) can be added in the middle of the Tof tube. Once the reflectron is reached by the ions, they are sent in the opposite direction following a “V” shape (Figure 2-8), or sometime even a “W” shape. The use of a reflectron compensates for induced kinetic differences at the entrance of the Tof, which can cause peaks to broaden. In addition, this approach has the advantage increasing the length of the Tof tube and, in consequence, increases the separation of the ions, thereby improving their resolution. However, each improvement to increase the resolution of an instrument often comes with the drawback of sensitivity loss and a compromised need to be taken.

Two different Q-Tof instruments have been used in the experiments reported here: a Xevo G2 (Figure 2-8) and a Synapt Q-Tof. The latter is a hybrid version of the first, which has three compartments instead of a collision cell. Those compartments can be used as collision cells. However, the middle one can also serve as an ion mobility drift cell and provide an additional dimension of separation. The separation is due to the presence of a neutral gas at low pressure, usually helium, in the presence of an electric field. The separation is essentially "gas phase electrophoresis" and ion move through the neutral gas at a rate which depends upon their size, shape and charge. As well as providing a separation the shape of the molecule, in terms of its "averaged collision-cross section (CCS) value" can be measured providing another characteristic physicochemical property of a molecule that can be used as further means of improving metabolite identification in metabolomics (Nye et al., 2019) and lipidomics (Paglia et al., 2015).

*Ion detectors.* The ions which successfully reach the mass detector are then acquired and can be displayed in the form of mass chromatograms with a time scale, which corresponds to the elution order. The default chromatogram is called total ion current (TIC), which is the sum of all the ions coming at each time point, or scan, that the detector has performed (Figure 2-9.a). In other words, a mass spectrum is orthogonally present behind each time point. As there are always some residuals ions, such as the ones coming from the solvents, the baseline obtained is never at zero, which can bring complexity to the mass spectrum and metabolite identification. Base peak ion (BPI) traces are thus often preferred as these shows only the most intense ions at each point of the mass chromatogram (Figure 2-9.a). Extracted ion chromatograms (EIC) can also be observed by selecting only one *m/z*, allowing the retention time of a specific ion to be determined. For example, the ions presenting an *m/z* of 455.18 in Figure 2-9.a is for instance eluting 1.30 and 3.69 minutes.

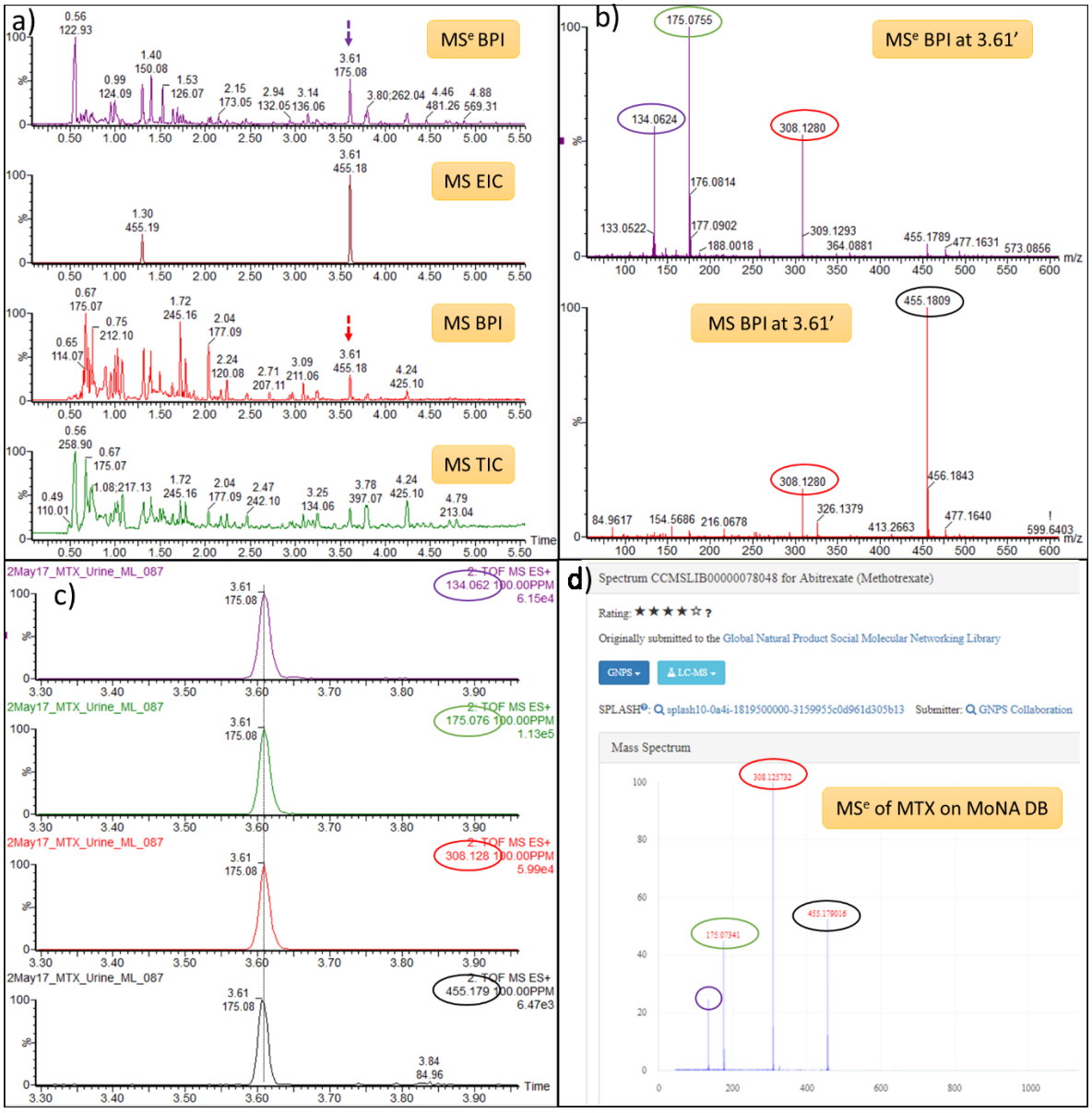



**Figure 2-9. Different chromatographic and MS traces.** a) Rat urine chromatogram seen according the total ion current (TIC) trace which shows all the ions eluting during the chromatographic gradient (green), and the base peak ion (BPI) trace which shows only the most intense ions (red), the extracted ion chromatogram (EIC), which shows only the elution time of the *m/z* 455.18 ions (brown), and the BPI chromatogram of the $MS^E$ acquisition. b) Mass spectrum of the MS BPI at 3.61 minutes (red) and mass spectrum of the $MS^E$ BPI at 3.61 minutes, illustrating decrease in intensity of the parent ion with an *m/z* of 455.18 and an increase in intensities of the daughter ions. c) EIC traces of the *m/z* of the four putative fragment ions to see if their retention time was exactly the same as that of the *m/z* 455.18/3.61 min ion, and therefore in likely from the same molecule. d) Mass spectrum of the drug methotrexate reported in the Massbank of North America (MoNA) database, acquired under $MS^E$ conditions, and showing the same fragmentation pattern (fragment *m/z* and intensities) as the compound observed in b) and eluting at 3.61 minutes in a).

### *1.2.3. Ensuring and monitoring performance monitoring*

It is clear from the practical application of LC-MS to metabonomic analysis that several parameters can induce variations during the analysis of large sample sets. On the UPLC side, the properties of the column can change due to contamination by components of the samples, which can cause retention time drift. On the MS side, contamination of the ion source can influence ionization efficiency, and contamination in the ion guide affects *m/z* signal variation. Regular maintenance of the instruments, such as good practice to prepare the solvents or store the chromatographic columns or regular cleaning of the ion source and ion guide etc., can help. However, different elements need to be carefully prepared and monitored before and along the experiment to ensure stable measurements, and minimize retention time drift or decreases in Rs for the UPLC, or unacceptable changes in accuracy, sensitivity or resolution for the MS.

Prior to the experiment, the chromatographic column needs to be equilibrated, particularly if it is new. In RPLC, it needs to be flushed with high percentage of B until the back pressure of the instrument stabilizes. The proportion of solvent B can then be slowly reduced, step by step, allowing reaching a stable back pressure to be obtained between each one, until the initial conditions of the gradient are reached. Once the column is equilibrated, it needs to be conditioned in order to ensure stability of the stationary phase over time. Indeed, during the time that a sample set is analysed the stationary phase will change, probably by adsorbing some analytes that will potentially elute out from the column, resulting in effects on retention. In practice it has been found that these effects can be minimised by "conditioning" the column with multiple injections of a pooled biological sample, known in untargeted assays as a QC. Sufficient conditioning has been performed when, after several QC injection, "perfect" overlap of the mass chromatograms is observed (which commonly takes about twenty injections, but this is arbitrary and depends on the matrix analysed, the age of the column and its previous use). Gradients without injection as well as injection of blanks (usually corresponding to the proportion of mobile phases A and B in the initial conditions of the gradient) are also important to analyse to assess the "background" noise and the presence, if any, of contaminants in the solvents, in the system or in the lab equipment which serves to prepare the samples (e.g. glass vials or plastic well plates). During the sample set, analysis of QCs, usually prepared by mixing aliquots of the study samples, are injected once every 5 or 10 samples (Dunn et al., 2015, 2011). This allow the stability of the chromatography over time to be monitored and determine if any bias has been introduced into the experiment. It is usually checked by overlapping them or by principal component analysis (PCA, see multivariate statistics paragraphs 1.4.) where all the QCs analysed should cluster together, with no obvious run order effects, in the the PCA score plot. A dilution series of the QCs can also be used to check that the intensity of a possible feature of interest decreases while the dilution factor increases.

In the case of the MS, prior the analysis the instrument detector needs to be tuned and adjusted if needed. A calibration needs to be performed prior each analysis, usually with a solution of sodium formate ($HCO_2Na$, 0.5 mM) which provides clusters that can be detected up to *m/z* 1200 and which allow to optimise the acquisition parameters to obtain the best resolution. To monitor the mass accuracy and stability over the run, a lock mass solution is used, composed of leucine enkephalin ($C_{28}H_{37}N_5O_7$, 200 pg/µL in ACN:$H_2O$ 1:1), which has an a *m/z* of 556.2771 for its $[M+H]^+$ adduct in positive electrospray ionization mode (ESI+) and a *m/z* of 554.2615 for its $[M-H]^-$ adduct in negative ESI (ESI-). It is continuously infused through an orthogonal sprayer into the ion source, constituted of a baffle which rotates, alternatively stopping the sample or the lock mass spray and allowing it to reach the mass analyser. Any *m/z* drift during the analysis will cause a difference between the measured masses of the analytes compared to their theoretical masses, increasing the ppm error and reducing the mass accuracy of the instrument. Monitoring the *m/z* of the lock mass with a scan every 60 seconds or so, allows changes to be detected and any *m/z* drift to be corrected in real time.

#### *1.2.4. Data pre-processing in LC-MS*

As for $^1H$ NMR spectroscopy, data pre-processing is a crucial step as data normalisation in LC-MS, filtering, centring, scaling and transformation can significantly change the raw dataset, directly affecting the multivariate statistics, as well as the potential biomarkers found. For metabolomics there are open source software or commercial software from different instrument manufacturers that can be used to pre-process UPLC-MS data. The use of open source tools is mainly dominated by two softwares (Weber et al., 2017), XCMS first, available as a R-based package (Smith et al., 2006) or online (Tautenhahn et al., 2012), and MZmine or MZmine2 (Pluskal et al., 2010). Both XCMS online and MZmine are workflow packages which include the possibility of performing data treatment post pre-processing. For Waters UPLC-MS instruments, commercially available software are Progenesis QI or Unifi. At the beginning of this project, Progenesis and XCMS were explored and compared. Progenesis was found to be more user-friendly. However, several drawbacks need to be pointed out. To start with, this is not a free-access software. The algorithms used behind the different steps such as the peak alignment or the normalisation are not accessible. This makes it a particular problem to adjust pre-processing parameters to the specific dataset, as it is well-established that in UPLC-MS large variation can occur from one dataset to another depending for instance on the instrument used and its resolution, but also the type of chromatography used or the kind of samples analysed. It has also a problem regarding the output obtained from Progenesis, which reports a lot of features that are just baseline in the raw data, so called “phantom peaks” (data not shown). For all these reasons, the R-package based XCMS was preferred at Progenesis to (pre)-process all of the datasets reported herein.

The first step of the pre-processing is to convert the raw data files into a readable format for XCMS, such as netCDF, mzXML or mzML format files. In the studies reported here mzML formatted files were used with the conversion performed using MZConvert utility of Proteowizard 3.0 (Chambers et al., 2012). Then the workflow is divided into two main parts: (1) the conversion of the data files into a matrix, which is done through successively peak peaking, peak alignment and peak grouping, followed by (2) data cleaning, which includes data normalisation, filtering and scaling.

*Conversion of the data file into a matrix.* The first step in converting a data file into a matrix is peak detection. The algorithm that can be used to perform this step depends on the nature of the dataset. The algorithm "centWave" is the most adapted for high resolution datasets generated in centroided mode (Tautenhahn et al., 2008). Centroided data is the calculation of the intensity-weighted average of the area under the curve of an *m/z* signal to represent it as a single straight line, in contrast to a continuous acquisition of the mass spectrum generating a trace all along the mass range measured. This allows a reduction of the size of the dataset generated, which is quite useful when large sample sets are analysed. Several parameters can be adjusted, such as the ppm error (as explained in paragraph 1.2.2.), the *m/z* error, the minimal and maximal chromatographic peak width, the retention time error, the signal/noise ratio at which signals will be considered (an in consequence all the signals below this S/N will be discard), the minimum number of scans needed to describe a good peak shape (usually defined as 7), etc. As these parameters are likely to vary from one dataset to another, according the chromatography, the mass spectrometer and the kind of samples analysed, it is advisable to carefully determine these parameters prior to performing peak detection, usually by examining the QCs acquired at the beginning, middle and the end of the analysis.

The next step of the pre-processing is peak alignment. It is composed of retention time drift correction (if any are observed) and peak grouping between the different samples of the dataset. Three main methods exist for retention time drift correction. The first one is based on a linear smoothing, which assumes that the observed drift is linear across the time course of the analysis. However, most of the time this is not the case, with some chromatographic peaks eluting with exactly the same retention time throughout the run, whilst others can vary. A local non-linear regression is then preferred, through the use of the LOESS algorithm (Smith et al., 2006). This polynomial smoothing fixes the most intense or consistent peaks found in all the samples first and then warps the peaks which are not that consistent in between. However, if no peaks are found to be consistent from a sample to another, warping the chromatograms of all the samples compared to a reference sample, carefully selected by the analyst, is possible by using the OBI-Warp algorithm (Prince and Marcotte, 2006). A lot of confusion is found in the literature for this step as it is sometimes assumed to be a normalisation method (Srivastava, 2017; Veselkov et al., 2011). However, retention time drift is not always observed if the chromatography can be kept stable along the analysis. In the studies reported here, the retention time error was negligible and neither linear smoothing, LOESS nor OBI-warp helped to improve the data alignment. Peak grouping in contrast, was performed which allows the retention time and the mass spectrum behind each retention time to be matched between all the samples. It can be done using the "nearest" algorithm or the "density" algorithm. The density overlaps the peaks by binning whilst the nearest method creates a peak list and matches close signals in the peak list from each sample. The nearest method is in consequence longer but can avoid two sharp peaks in the same bin being considered as one. The nearest method was applied in the untargeted studies reported here.

Finally, because of the S/N ratio which is defined during peak detection, features with smaller intensities will present null values. However, the baseline itself never being at zero, null values are experimentally not possible and could induce bias in the multivariate statistics. Peak filling is thus performed to overcome this problem, which consists of going back to the raw data and replacing the null values by the S/N values or the small intensities detected.

The output obtained from this part of the pre-processing is an extensive table, which is called a matrix. The rows list the samples and the columns list the variables. Each variable is reported as a *m/z* and retention time pair. The matrix also contains the average, the minimum and the maximum values of the *m/z* and the retention time as well as in how many samples it has been detected. Each [*m/z* / retention time] pair is called a feature, and the intensity of each feature in each sample is reported.

*Data cleaning.* As for $^{1}$H NMR spectroscopy, dilution in urine samples or experimental errors, as well as fluctuations due to the stability of the chromatography over the experiment or the MS signal (temperature changes, ionization variation…) need to be corrected for by using normalisation. Nevertheless, normalisation algorithms are not based on biological criteria and in consequence can be seen as an extra manipulation of dataset introducing significant bias (Mertens, 2017). Careful consideration of which normalisation and the need for using it must be undertaken. Normalisation is usually made by the ratio of each variable of the dataset by a constant. This constant can be the intensity of an internal standards spiked at a known concentration (not always practical due to the risk of overlapping with analytes and thus inducing ion suppression), an external standard such as creatinine concentration in urine (which requires a knowledge of the total volume of urine excreted and which assumes that the urinary concentration of creatinine is the same in all the subjects), or the median fold change of the variable in a sample compared to a reference. This technique has been discussed above for $^{1}$H NMR-based metabolomics, but it is also widely used in UPLC-MS metabolomics as even if its efficacy in normalisation is similar to other techniques, it is less skewed by most intense signals (Veselkov et al., 2011).

Filtering is also often performed. Usually, the coefficient of variation of a variable is calculated by the ratio of its standard deviation (relative to a QC) over its mean. Then, only those variables which have a CV % below an arbitrary value, usually 20 or 30% in metabolomics, are retained for future analysis. Other kinds of filtering exist, such as the minimal fraction filter, which remove features that are present in less than a fixed cut-off fraction of a sample group. This type of filtering was explored during the experiments reported herein but, as its use was not found to improve the dataset and was judged to remove too much useful information, CV filtration was preferred.

## 1.3. Targeted UPLC-MS

### *1.3.1. Instrumentation*

Although the resolution and the accuracy of triple quadrupole (TQ) mass analysers are inferior to those of the Q-Tof instruments, TQ provide higher sensitivity and specificity for the targeted analysis of specific compounds (*e.g.* single metabolites or xenobiotics), classes of compounds (*e.g.* bile acids or amino acids) or even a specific metabolic pathways (*e.g.*, tryptophan or tyrosine biosynthesis or catabolism). Targeted analysis can also be used on samples that have already been analysed by untargeted UPLC-MS and which have highlighted the presence of certain compounds as biomarkers. The quantification of biomarkers or metabolites based on the use of standard calibration curves and internal/external standards is then possible. The specific and sensitive detection of a targeted metabolite(s) is made by the measurement of specific *m/z* transitions resulting from the fragmentation of the parent compound to give characteristic fragment daughter ions. The degree of fragmentation, and the number of fragments, depends on the collision energy applied to the parent compound.

As was the case for untargeted profiling, different types of ionization, such as EI, APCI and APPI can be used, in positive and negative modes, and switching back and forth from positive to negative (multiple times) is also possible during the same assay, allowing the detection of compounds with different ionization requirements (see example below). The mode of acquisition is usually MS (single ion monitoring or SIM) where fragment ions are not monitored, or MS/MS (which uses multiple reaction monitoring or MRM). To determine the best transition(s) to monitor (and the best parameters by which to measure them) in MRM-based assays, direct infusion of the standard is usually done while 50 % of mobile phase A and 50 % of mobile phase B is being infused to the mass spectrometer. In such assays it is preferable to monitor several transitions per compound.

During the work performed for Chapter 5, a fit-for-purpose validated method was developed to measure tyrosine and its catabolites. In order to explain the targeted approach used in UPLC-MS, the fit-for purpose validation is described below.

### *1.3.2. Fit-for-purpose validation of a targeted assay for tyrosine and selected catabolites*

The influence of tyrosine on health and disease is extensive. Indeed, tyrosine, a non-essential (or semi-essential) amino acid obtained from the diet and protein degradation, plays a key role in hormonal pathways, vitamin C metabolism, stress management, and is a precursor of dopamine and other neurotransmitters. It is also a precursor of the bacterial metabolite *p*-cresol, which has been growing interest for its impact on host-microbiome reactions. Notably *p*-cresol is suspected of involvement in several diseases such as autism spectrum disorder or obesity. However, *p*-cresol has also been reported to influence drug metabolism and toxicity by depleting the sulfation capacity of the host. Indeed, high concentrations of the bacterial metabolite *p*-cresol have the potential to saturate the enzyme SULT1A1, which is responsible for the sulfation of a broad range of substrates including xenobiotics but also endogenous compounds which might have a role in diseases involving hormone and neurotransmitter metabolism, particularly important in early human development. However, the bacterial impact on the transformation going from tyrosine to *p*-cresol and *p*-cresol-sulfate is still unclear and a fit-for-purpose UHPLC-ESI-MS method has been developed to quantify tyrosine and downstream metabolites such as *p*-cresol, tyramine, *p*-hydroxyphenylacetic acid (a *p*-cresol precursor), and *p*-cresol glucuronide to better understand microbial metabolism from tyrosine to *p*-cresol sulfate (Figure 2-10).

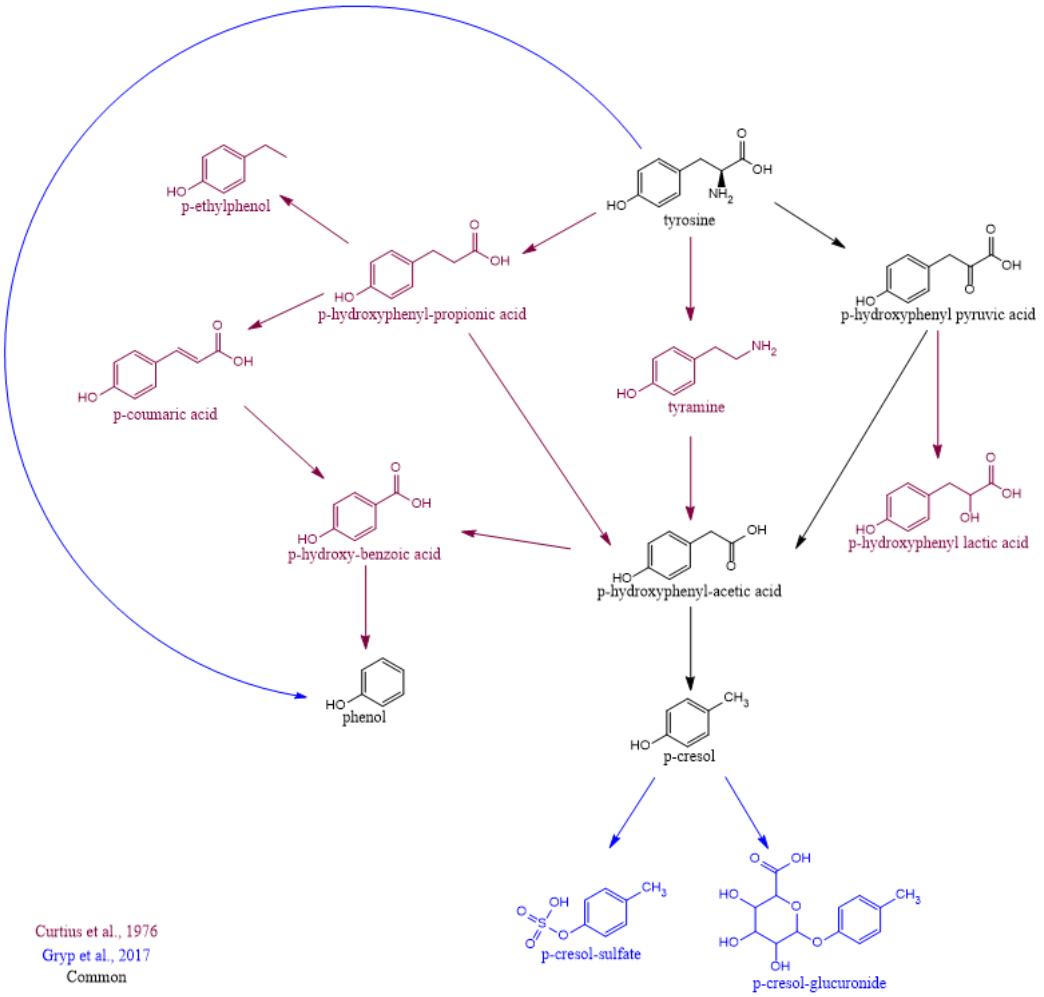



**Figure 2-10. Tyrosine to *p*-cresol pathway.** The 12 analytes detected by UHPLC-MS in the targeted assay are indicated (p-hydroxyphenylpyruvic acid is indicated in the pathway although it could not be successfully included in the targeted compounds). The colour represents the literature which used to derive the metabolic pathway, in either purple **(Curtius et al., 1976)**, blue **(Gryp et al., 2017)**, or black for the compounds common to both studies.

**Sample Analysis.** Samples were analysed by targeting the compounds reported in Figure 2-10 (at the exception of *p*-hydroxyphenylpyruvic acid which was not detected). The chromatographic separation is reported in Figure 2-11. Chemicals used, sample preparation and UPLC-MS parameters are reported in the supplementary information. Calibration curves were run before and after each set of samples with QC samples run at regular intervals throughout the analysis. The assay described below was designed to enable the rapid and reproducible quantification of tyrosine and 11 catabolites in urine and plasma samples (applied to samples in Chapter 5). Derivatization with dansyl chloride was required for the analysis of the phenolic tyrosine catabolites in order to obtain suitable MS-detection properties using positive ESI, whilst the phenolic sulfate and glucuronide conjugates of *p*-cresol were analysed using negative ESI. However, attempts to use dansylation were unsuccessful for the *p*-hydroxyphenylpyruvic acid and, as a consequence, the analysis of this metabolite was not pursued.

The resulting separation (Figure 2-11) showed good stability, providing reproducible retention times for all analytes based on the $t_r$ CVs of the various standards of < 0.81 % across the six replicate injections of the high QCs over the course of the analysis.

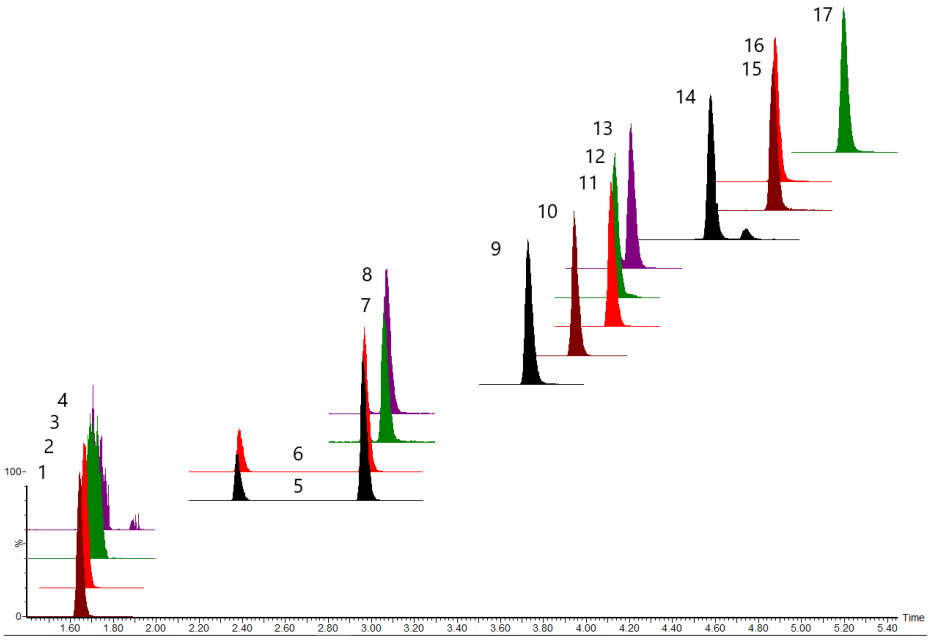



**Figure 2-11. UHPLC-MS chromatogram of the twelve standards and five deuterated internal standards.** (1) *p*-cresol sulfate-$d_7$, (2) *p*-cresol sulfate, (3) *p*-cresol-glucuronide-$d_7$, (4) *p*-cresol glucuronide, (5) tyrosine-$d_7$, (6) tyrosine, (7) tyramine-$d_4$, (8) tyramine, (9) *p*-hydroxyphenyllactic acid, (10) p-hydroxyphenylacetic acid, (11) *p*-hydroxybenzoic acid, (12) *p*-hydroxypropionic acid, (13) *p*-coumaric acid, (14) phenol (15) *p*-cresol-$d_4$, (16) *p*-cresol, (17) *p*-ethyl phenol.

In the case of dansylated phenol, two well separated peaks were observed in the mass chromatogram (Figure 2-11) however, the later eluting peak was also present in the blank and represented a contaminant. With tyrosine, two chromatographic peaks were also observed in the analytes MRM transition window (Figure 2-11), with the same phenomenon also seen for the SIL IS (tyrosine-$d_7$). The likely explanation for this is that tyrosine has two derivatizable groups (both the phenolic -OH and the amino -$NH_2$) and integrating and summing both chromatographic peaks of the calibration curve gave linear responses that could be used for quantification.

For detection and confirmation, selective MRM transitions were developed for each metabolite and SIL IS standard (Table 2-3). In the case of the dansylated analytes the loss of the dansyl moiety from the molecular ion was used and, apart from *p*-hydroxybenzoic acid and *p*-hydroxypropoinic acid, co-elution of these derivatized analytes was not observed. Whilst *p*-cresol sulfate and *p*-cresol glucuronide eluted near each other early in the separation, quantification was possible based on different molecular ions.

**Table 2-3. Chromatographic and Mass Spectrometric Details for the Determination of Tyrosine and Targeted Catabolites in Mouse Urine and Plasma.** q: quantitative; c: confirmatory.

| *Compounds* | *Tr (mins)* | *Parent m/z* | *Daughter m/z* | *Cone voltage (V)* | *Collision energy (eV)* | *MS polarity* |
|---|---|---|---|---|---|---|
| *tyrosine 1 q** | 2.97 | 415.07 | 156.13 | 50.0 | 48.0 | + |
| *tyrosine 1 c** | 2.97 | 415.07 | 170.50 | 50.0 | 28.0 | + |
| *tyrosine 2 q** | 2.39 | 415.07 | 170.50 | 50.0 | 28.0 | + |
| *tyrosine 2 c** | 2.39 | 415.07 | 156.13 | 50.0 | 48.0 | + |
| *tyramine q** | 3.07 | 371.00 | 156.26 | 2.0 | 36.0 | + |
| *tyramine c** | 3.07 | 371.00 | 170.31 | 2.0 | 24.0 | + |
| *p-ethylphenol q** | 5.20 | 356.00 | 156.27 | 32.0 | 36.0 | + |
| *p-ethylphenol c** | 5.20 | 356.00 | 171.23 | 32.0 | 24.0 | + |
| *p-coumaric acid q** | 4.21 | 398.10 | 156.26 | 64.0 | 40.0 | + |
| *p-coumaric acid c** | 4.21 | 398.10 | 171.28 | 64.0 | 22.0 | + |
| *p-hydroxybenzoic acid q** | 4.12 | 372.03 | 171.22 | 10.0 | 20.0 | + |
| *p-hydroxybenzoic acid c** | 4.12 | 372.03 | 156.26 | 10.0 | 36.0 | + |
| *p-hydroxyphenyl acetic acid q** | 3.95 | 386.10 | 156.26 | 62.0 | 36.0 | + |
| *p-hydroxyphenyl acetic acid c** | 3.95 | 386.10 | 171.23 | 62.0 | 24.0 | + |
| *p-hydroxyphenyl lactic acid q** | 3.73 | 416.07 | 171.21 | 8.0 | 26.0 | + |
| *p-hydroxyphenyl lactic acid c** | 3.73 | 416.07 | 156.24 | 8.0 | 40.0 | + |
| *p-cresol q** | 4.88 | 342.07 | 156.26 | 50.0 | 32.0 | + |
| *p-cresol c** | 4.88 | 342.07 | 171.23 | 50.0 | 22.0 | + |
| *phenol q** | 4.58 | 328.03 | 171.23 | 10.0 | 20.0 | + |
| *phenol c** | 4.58 | 328.03 | 156.26 | 10.0 | 30.0 | + |
| *p-cresol sulfate* | 1.70 | 186.71 | 107.20 | 2.0 | 16.0 | - |
| *p-cresol glucuronide* | 1.72 | 283.31 | 107.14 | 18.0 | 32.0 | - |
| *p-hydroxyphenyl propionic acid q** | 4.13 | 400.70 | 156.26 | 10.0 | 40.0 | + |
| *p-hydroxyphenyl propionic acid c** | 4.13 | 400.70 | 171.16 | 10.0 | 24.0 | + |
| *tyrosine $d_7$ 1 q** | 2.96 | 422.13 | 156.13 | 50.0 | 48.0 | + |
| *tyrosine $d_7$ 1 c** | 2.96 | 422.13 | 170.50 | 50.0 | 28.0 | + |
| *tyrosine $d_7$ 2 q** | 2.38 | 422.13 | 170.50 | 50.0 | 28.0 | + |
| *tyrosine $d_7$ 2 c** | 2.38 | 422.13 | 156.13 | 50.0 | 48.0 | + |
| *tyramine $d_4$ q** | 3.06 | 375.00 | 156.26 | 2.0 | 36.0 | + |
| *tyramine $d_4$ c** | 3.06 | 375.00 | 170.31 | 2.0 | 24.0 | + |
| *p-cresol $d_4$ q** | 4.87 | 347.07 | 171.23 | 50.0 | 22.0 | + |
| *p-cresol $d_4$ c** | 4.87 | 347.07 | 156.26 | 50.0 | 32.0 | + |
| *p-cresol sulfate $d_7$ q* | 1.68 | 193.71 | 114.20 | 2.0 | 16.0 | - |
| *p-cresol glucuronide $d_7$* | 1.71 | 290.31 | 114.14 | 18.0 | 32.0 | - |

*(* = Compounds detected following dansyl chloride derivatization).*

The methodology devised for the analysis of tyrosine and its catabolites was, as far as possible, consistent with a "fit for purpose" method, developed based on the guidance of the US Food and drug administration (FDA) (Research, 2019; Tiwari and Tiwari, 2010).

*Analytical range.* A linear fit was used for the calibration curves, without weighting or axis transformation, and applied for peak integration over the range 0.1 to 1000 nM for all analytes (Table 2-S1). For tyrosine, tyramine, *p*-cresol, *p*-cresol sulfate and *p*-cresol glucuronide, where deuterated stable isotope labelled (SIL) internal standards were available, peak area response ratios were calculated and plotted against the nominal concentration.

*Lower limit of quantification (LLOQ).* The LLOQ's for each of the analytes were set as the lowest concentration of the calibration curve with an acceptable level of analyte carry over (from 10 to 250 nM depending upon the analyte (Table 2-S2), see below). The assays were established as being linear up to 1000 nM (used as the ULOQ). Apart from *p*-ethylphenol with an *R2* of 0.972 and *p*-cresol-glucuronide with an *R2* of 0.967, the *R2* values obtained were all better than 0.980, (Table 2-S3). The final concentration ranges used for each of the analytes, as well as the QC concentrations also used to determine accuracy and precision are reported in Table 2-S2.

*Accuracy and Precision.* To monitor the intra-day accuracy and precision during the sample analysis, the low, medium and high QC's (Table 2-S2) were injected every 11 samples. A pragmatic assessment of accuracy (how close the measured concentrations were to the nominal values) and precision (the variability of the result) was made. The sample analysis was judged acceptable if the lower QC was within 20 % and the middle and higher QCs were within 15 % of their nominal concentrations. For acceptance of the analytical data obtained, 2/3 of the QCs (at least 50 % at each concentration) had to meet these expectations.

By these criteria tyrosine and all the catabolites monitored in this assay gave acceptable results for all the metabolites in urine, apart from the middle QC of *p*-cresol-glucuronide in urine (Table 2-4), and for the metabolites present in plasma (Table 2-5). Therefore, particular care needs to be taken in interpreting the concentration of *p*-cresol-glucuronide in the biological samples.

**Table 2-4.QCs monitoring in urine.** q: quantitative; c: confirmatory.

| *Compounds* | *QC low (n=2)* | | | *QC mid (n=7)* | | | *QC high (n=6)* | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** |
| *tyrosine 1 q* | 37.5 | 34.03 | 9.15 | 400 | 393.72 | 1.80 | 800 | 804.78 | 1.61 |
| *tyrosine 1 c* | 37.5 | 30.84 | 0.72 | 400 | 386.27 | 2.12 | 800 | 810.25 | 1.61 |
| *tyrosine 2 q* | 37.5 | 32.88 | 2.67 | 400 | 387.17 | 2.22 | 800 | 803.33 | 1.29 |
| *tyrosine 2 c* | 37.5 | 36.68 | 4.86 | 400 | 398.58 | 2.42 | 800 | 807.19 | 2.10 |
| *tyramine q* | 15.0 | 19.34 | 17.44 | 400 | 386.07 | 2.19 | 800 | 790.96 | 1.93 |
| *tyramine c* | 15.0 | 28.11 | 13.86 | 400 | 433.85 | 8.04 | 800 | 816.88 | 4.10 |
| *p-ethylphenol q* | 150.0 | 127.41 | 0.25 | 400 | 235.68 | 2.30 | 800 | 798.66 | 4.20 |
| *p-ethylphenol c* | 150.0 | 126.56 | 0.01 | 400 | 237.66 | 1.92 | 800 | 800.39 | 4.09 |
| *p-coumaric acid q* | 75.0 | 86.31 | 3.31 | 400 | 254.26 | 1.96 | 800 | 819.10 | 3.48 |
| *p-coumaric acid c* | 75.0 | 86.58 | 1.79 | 400 | 253.01 | 2.32 | 800 | 815.10 | 3.45 |
| *p-hydroxybenzoic acid q* | 125.0 | 139.15 | 0.23 | 400 | 363.86 | 2.29 | 800 | 821.10 | 3.47 |
| *p-hydroxybenzoic acid c* | 125.0 | 141.17 | 1.43 | 400 | 363.94 | 2.32 | 800 | 817.49 | 3.68 |
| *p-hydroxyphenyl acetic acid q* | 75.0 | 81.78 | 1.16 | 400 | 337.15 | 1.91 | 800 | 810.38 | 3.81 |
| *p-hydroxyphenyl acetic acid c* | 75.0 | 80.52 | 0.81 | 400 | 339.45 | 2.18 | 800 | 810.19 | 3.67 |
| *p-hydroxyphenyl lactic acid q* | 125.0 | 147.82 | 2.59 | 400 | 340.56 | 2.43 | 800 | 821.67 | 3.28 |
| *p-hydroxyphenyl lactic acid c* | 125.0 | 147.68 | 2.30 | 400 | 340.92 | 2.52 | 800 | 821.57 | 3.34 |
| *p-cresol q* | 150.0 | 144.19 | 0.67 | 400 | 388.60 | 1.25 | 800 | 799.60 | 0.85 |
| *p-cresol c* | 150.0 | 145.70 | 0.16 | 400 | 388.98 | 1.19 | 800 | 799.25 | 1.57 |
| *phenol q* | 200.0 | 169.31 | 2.61 | 400 | 243.20 | 2.29 | 800 | 821.70 | 3.66 |
| *phenol c* | 200.0 | 170.81 | 0.20 | 400 | 245.06 | 1.91 | 800 | 819.59 | 3.67 |
| *p-cresol sulfate q* | 150.0 | 150.76 | 3.36 | 400 | 356.14 | 9.54 | 800 | 718.28 | 11.13 |
| *p-cresol glucuronide* | 250.0 | 273.54 | 1.66 | 400 | 400.84 | 16.66 | 800 | 763.37 | 14.01 |
| *p-hydroxyphenyl propionic acid q* | 75.0 | 86.97 | 5.32 | 400 | 271.19 | 6.17 | 800 | 775.29 | 4.68 |
| *p-hydroxyphenyl propionic acid c* | 75.0 | 88.40 | 2.86 | 400 | 273.54 | 5.59 | 800 | 791.75 | 3.63 |

**Table 2-5. QCs monitoring in plasma.** q: quantitative; c: confirmatory.

| *Compounds* | *QC low (n = 3)* | | | *QC mid (n = 8)* | | | *QC high (n = 7)* | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** | **Nominal conc (nM)** | **Mean conc (nM)** | **CV %** |
| *tyrosine 1 q* | 37.5 | 33.26 | 4.49 | 400 | 381.69 | 8.53 | 800 | 809.84 | 1.30 |
| *tyrosine 1 c* | 37.5 | 32.94 | 3.25 | 400 | 379.94 | 7.14 | 800 | 805.60 | 1.18 |
| *tyrosine 2 q* | 37.5 | 30.18 | 10.75 | 400 | 379.44 | 5.89 | 800 | 803.20 | 1.89 |
| *tyrosine 2 c* | 37.5 | 31.46 | 2.50 | 400 | 376.87 | 8.45 | 800 | 803.77 | 1.13 |
| *p-hydroxybenzoic acid q* | 125.0 | 142.28 | 2.14 | 400 | 363.10 | 3.11 | 800 | 815.65 | 3.96 |
| *p-hydroxybenzoic acid c* | 125.0 | 143.18 | 3.63 | 400 | 364.11 | 5.31 | 800 | 818.86 | 5.87 |
| *p-hydroxyphenyl lactic acid q* | 125.0 | 150.42 | 0.68 | 400 | 337.38 | 2.48 | 800 | 815.55 | 3.08 |
| *p-hydroxyphenyl lactic acid c* | 125.0 | 150.05 | 2.90 | 400 | 339.64 | 3.50 | 800 | 814.79 | 4.35 |
| *p-cresol q* | 150.0 | 149.51 | 0.28 | 400 | 379.04 | 7.68 | 800 | 801.10 | 0.46 |
| *p-cresol c* | 150.0 | 152.36 | 2.37 | 400 | 376.63 | 8.76 | 800 | 798.43 | 0.74 |
| *p-cresol sulfate q* | 150.0 | 147.25 | 8.13 | 400 | 356.87 | 14.97 | 800 | 739.44 | 9.29 |

*Carryover.* Analyte/SIL carryover was assessed by injecting a double blank (pure diluent which had been through the derivatization process where appropriate) immediately after the top concentration of the calibration curve (1000 nM). Carry over for analytes was considered to be acceptable if the amount detected in the double blanks was ≤ 20% of the LLOQ for the standards and ≤ 5% of an internal standard blank (pure diluent containing only the internal standard mixture that had been through the derivatization process). Both mid and high QCs passed these acceptance criteria. However, the initial LLOQs assessment did not provide acceptable levels of carryover using the 30 % criterion, which is known to be a recurrent issue with the autosampler used during the development and application of this assay. The lowest concentration solutions of the calibration curves were therefore re-purposed to answer this issue. All of the analytes showed then acceptable levels of carryover, with the signals detected being below than 20 % of their respective LLOQ's (Table 2-S4) observed in a double blank sample (containing no analytes, but with derivatization reagents present and taken through the sample preparation process). Similarly, the carryover of the internal standards was acceptable at less than 5 % (Table 2-S4).

*Stability.* To evaluate short term analyte stability at 4 °C in the autosampler the results for the low and high QC samples analysed at both the beginning (0 h) and end (28h) of sample analysis were compared. In addition, the stability of both derivatized analytes and *p*-cresol sulfate and glucuronide, were assessed following 16 days (390h) storage in the autosampler at 4°C. From the resulting data the percentage bias and CV were calculated, and stability was considered to be acceptable if the bias was less than 15 %.

Concentrations of the QCs are reported in Table 2-6. The CV % of the concentrations of the QC at 28h and 390h compared to the initial concentration were calculated and reported in Table 2-7. Compounds were judged stable if presenting a percentage bias below 15 %. Apart from *p*-cresol-glucuronide, all analytes were stable following a period of 28h at 4˚C. However, a longer period of 390h showed significant loss of signal for six compounds (*p*-cresol-sulfate, phenol, *p*-hydroxyphenyl propionic acid, *p*-coumaric acid and *p*-ethylphenol). Tyrosine, tyramine, *p*-hydroxyphenyl lactic acid, *p*-hydroxyphenyl acetic acid, *p*-hydroxybenzoic acid and *p*-cresol demonstrated stability over the period of 390h at 4˚C.

**Table 2-6. Concentration of the low and high QCs for each compound measured initially, and following periods of 28h and 390h in the autosampler at 4˚C.**

| *Concentrations* (nM) | *QC low* | | | | *QC high* | | | |
|---|---|---|---|---|---|---|---|---|
| *Compounds* | Nominal | Initial | 28 hours | 390 hours | Nominal | Initial | 28 hours | 390 hours |
| *tyrosine 1* | 37.5 | 31.2 | 31.8 | 33.5 | 800 | 785.7 | 815 | 814.6 |
| *tyrosine 2* | 37.5 | 34.9 | 32.7 | 31.9 | 800 | 801.5 | 814.6 | 794.8 |
| *tyramine* | 15 | 16.9 | 15 | 17.5 | 800 | 811.5 | 790.8 | 822.1 |
| *p-ethyl phenol* | 150 | 129.2 | 129.7 | 121.4 | 800 | 841.1 | 834.7 | 529.8 |
| *p-coumaric acid* | 75 | 94.1 | 94.4 | 79.9 | 800 | 873.3 | 859.6 | 685.3 |
| *p-hydroxybenzoic acid* | 125 | 145 | 145.9 | 123.1 | 800 | 857.8 | 850.5 | 702.3 |
| *p-hydroxyphenylacetic acid* | 75 | 86.5 | 88.3 | 76.8 | 800 | 843.1 | 841.4 | 700.6 |
| *p-hydroxyphenyllactic acid* | 125 | 154.1 | 155.2 | 132.6 | 800 | 855.2 | 854 | 711.6 |
| *p-cresol* | 150 | 148 | 147.7 | 145.3 | 800 | 788.4 | 819.6 | 799.2 |
| *phenol* | 200 | 180.2 | 173.4 | 146.3 | 800 | 873.4 | 848.8 | 568.2 |
| *p-cresol sulfate* | 150 | 126 | 133.8 | 192.7 | 800 | 700 | 752.2 | 909.4 |
| *p-cresol glucuronide* | 250 | 357.4 | 215.1 | 78.1 | 800 | 1172.6 | 1167.1 | 95.9 |
| *p-hydroxyphenylpropionic acid* | 75 | 86.7 | 95.5 | 75.1 | 800 | 813.1 | 907.5 | 653.1 |

**Table 2-7. Percentage bias and CV of the low and high QCs for each compound following a period of 28h and 390h in the autosampler at 4˚C compared to the initial concentration measured.**

| *Compounds* | *CV % compared to initial concentration* | | | |
|---|---|---|---|---|
| | QC low | | QC high | |
| | 28h | 390h | 28h | 390h |
| *tyrosine 1* | 1.3 | 5.0 | 2.6 | 2.6 |
| *tyrosine 2* | 4.6 | 6.4 | 1.1 | 0.6 |
| *tyramine* | 8.4 | 2.5 | 1.8 | 0.9 |
| *p-ethyl phenol* | 0.3 | 4.4 | 0.5 | 32.1 |
| *p-coumaric acid* | 0.2 | 11.5 | 1.1 | 17.1 |
| *p-hydroxybenzoic acid* | 0.4 | 11.6 | 0.6 | 14.1 |
| *p-hydroxyphenyl acetic acid* | 1.5 | 8.4 | 0.1 | 13.1 |
| *p-hydroxyphenyl lactic acid* | 0.5 | 10.6 | 0.1 | 13.0 |
| *p-cresol* | 0.1 | 1.3 | 2.7 | 1.0 |
| *phenol* | 2.7 | 14.7 | 2.0 | 29.9 |
| *p-cresol sulfate* | 4.2 | 29.6 | 5.1 | 18.4 |
| *p-cresol glucuronide* | 35.2 | 90.7 | 0.3 | 120.0 |
| *p-hydroxyphenyl propionic acid* | 6.8 | 10.1 | 7.8 | 15.4 |

*Recovery*. In the absence of analyte-free matrix, analytical recovery was assessed for plasma using the pooled QC sample, into which the SIL ISs were added before and after the protein precipitation step (Table 2-S5). Both *p*-cresol-sulfate-$d_7$ and *p*-cresol-glucuronide-$d_7$ show essentially quantitative recoveries (higher than 95%). For the second signal of the tyrosine-$d_7$, the recovery was also quite high with a percentage of 94.9% but the first signal shows a recovery percentage of only 88.4%, which have an average recovery for the two tyrosine signals of 91.7%. Tyramine-$d_4$ and *p*-cresol-$d_4$ provided recoveries of 84.0% and 75.2% respectively.

*Analyte/Blank Interferences.* Blank interferences with the analytes targeted in the assay were assessed by integrating the area under the signal obtained for the analytes of interest in the double blank (diluent not containing standards or internal standards, but which had been taken through the derivatization process, if any). Signals found for the double blank that were < 15% of the signals obtained in the LLOQ were not considered to be interferences.

In general, there were no interfering signals in the double blank compared to the LLOQ sample (Table 2-S6). All the compounds passed the acceptance criteria, with the exception of the first-eluting signal of the two ones detected for tyrosine.

*Data Processing.* The data were acquired using MassLynx software (4.1, Waters Corporation) and processed using the TargetLynx package. Statistical analyses were performed using the GraphPad package in Prism Software -version 8.1.2 (California, U.S.A).

## 1.4. Multivariate Statistics

### *1.4.1. Principle*

The third component of metabolomics, after NMR spectroscopy and mass spectrometry-based techniques, is chemometrics which involves the use of mathematical and statistical techniques, including multivariate statistical analysis, to improve the understanding of the datasets generated by the analysis. As explained above, the datasets obtained, from both $^{1}$H NMR spectroscopy or UPLC-MS, are in the form of a matrix containing *N* samples (or observations) in rows, and *K* variables (or features) in columns. The number of variables obtained by doing metabolic profiling can reach thousands. Performing univariate statistics, such as a student's *t* test, relies on taking into consideration each variable, one at a time. The correlation between the variables are ignored and likely to induce error in biomarker discovery such as false positives. By projecting the samples into a multidimensional space, their distribution between each other gives an insight regarding the similarities or differences that the samples share in terms of their metabolic profile. Two main kinds of multivariate statistics exist, which depends on whether the class the sample belongs to is taken into account during the analysis. If the sample class is not taken into consideration in the algorithm, the resultant model is considered to be unsupervised. The main method used to perform unsupervised analysis is principal component analysis (PCA) (Pearson K., 1901). When the analyst specifies prior to the statistical calculations which sample class (or other response variable) each sample belongs to, the model is supervised. The two main supervised methods are partial least squares or projection to latent structures (PLS; Wold et al., 2001) and orthogonal projection to latent structures (OPLS; Trygg and Wold, 2002). However, with the increasing popularity of -omics sciences and thus the use of chemometrics, various derivative methods have emerged, such as the covariate-adjusted projection to latent structures (CA-PLS; Posma et al., 2018). Both unsupervised and supervised methods are discussed below.

### *1.4.2. Unsupervised analysis*

PCA works by identifying the largest sources of variance in the multidimensional dataset (Eriksson et al., 2013). It projects the samples (or observations) into independent and orthogonal principal components, depending on the contribution of the different variables (metabolites) in the samples. The principal components (PC) are axes that follow the variation of the samples in space. The first principal component represents the largest variance between the samples ($PC_1$- Figure 2-12), the second component, which is orthogonal to the first one, represents the second largest variance ($PC_2$- Figure 2-12), and so on.

The position of the samples along these principal components provides a score (*t*) for the sample for each component. Scores represent the distance separating the sample from the PC origin. The distance of a sample to the origin of $PC_1$ will result in the score ($t_1$) and to the origin of $PC_2$ will result in the score ($t_2$) (Figure 2-12.a). The distribution of the scores on the different principal components can be visualised through a scores plot, which can be seen in 2 or 3D, where each axis is a principal component (Figure 2-12.a). The closer samples cluster in their scores, the more similar the metabolic profiles of the samples are.

To explore what variables (metabolites) contribute to each PC, the loadings plots can be studied. The loading (*P*) value describes the contribution of the variable to the PC. The loading value indicates the direction of influence of the variable to the PC and the weight of the variable on this PC is calculated by the cosine of the angle between the variable and PC (Figure 2-12.b). Thus, on a loadings plot (Figure 2-12.b), each point is a variable and the further a variable is from the centre of the plot the greater influence this variable has on the PC (and movement of samples (scores) along this PC).

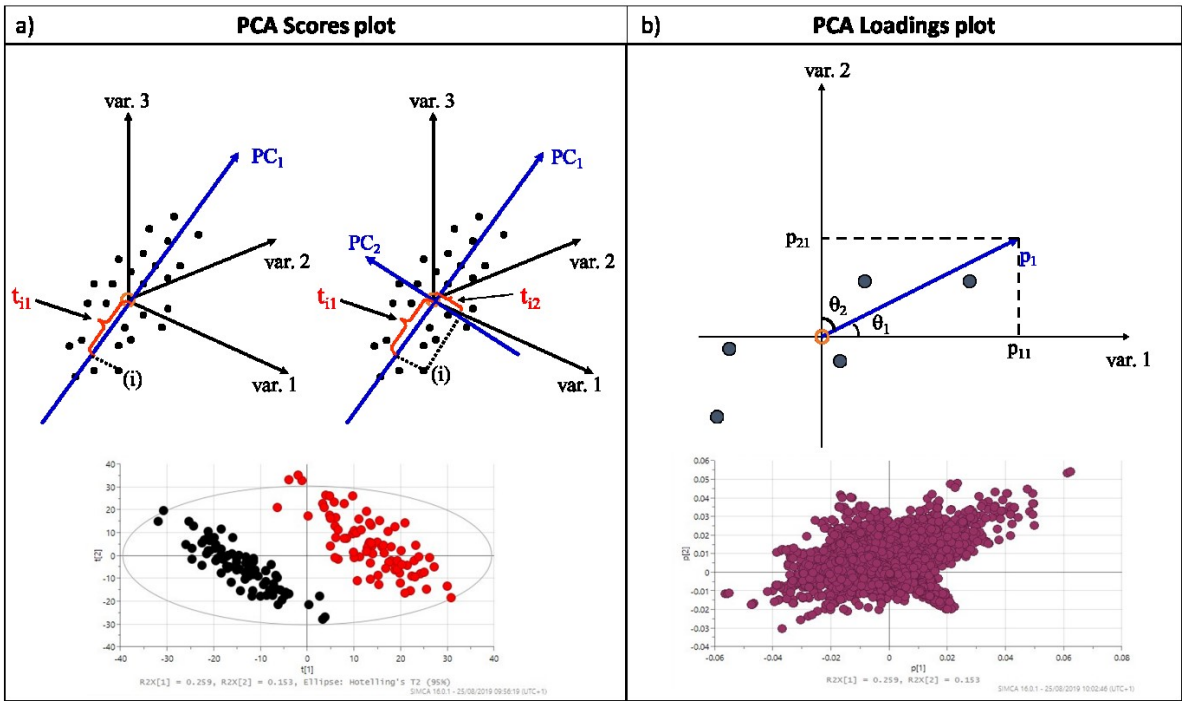



**Figure 2-12. Principle of principal component analysis.** a) Schematic explanation of how principal component (PC) are built to explain the largest variance possible between the samples, which are projected into the plane space described by the PCs according the scores (t). This results in a PCA scores plot with the Hotelling's $T^2$ ellipse (95 % confidence), where each point is a sample and where a clear separation can be observed between the "case" and the "control" samples (red and black respectively). b) Representation of the weights that two variables have on PC1, which are calculated by the cosine of the angle between the PC and each of the variable. An example of the resulting loadings plot, where each point is a variable and show that variable further away from the centre are the variables which influence the most the separation between the two sample classes observed in the PCA scores plot. Credit: Henrik Antti / Umea University.

The number of principal components in a model needs to be high enough to explain as much of the variance ($R^2$) as possible, but not too high to avoid overfitting the model, which will reduce the predictability of the model, measured by $Q^2$ (Eriksson et al., 2013). The higher the $R^2$ and $Q^2$ values the more robust and trustworthy the model is. PCA is useful for observing the general distribution trends of the model and highlighting the presence of outliers, using the Hotelling's $T^2$ ellipse with a default α of 0.05 (Hotelling, 1933).

Due to the extensive intensity variation of the features (dynamic range), two features of different intensities will not exert the same influence on the multivariate statistical model. Mean-centering, scaling and transforming can thus be used to regulate the influence of the features with different intensities. Mean-centering subtracts the mean of each variable across the samples from the variable of each sample. This allows variables to be centred around the mean so that samples with variables higher than the mean will have a positive value and samples with variables lower than the mean will have a negative value (Figure 2-13).

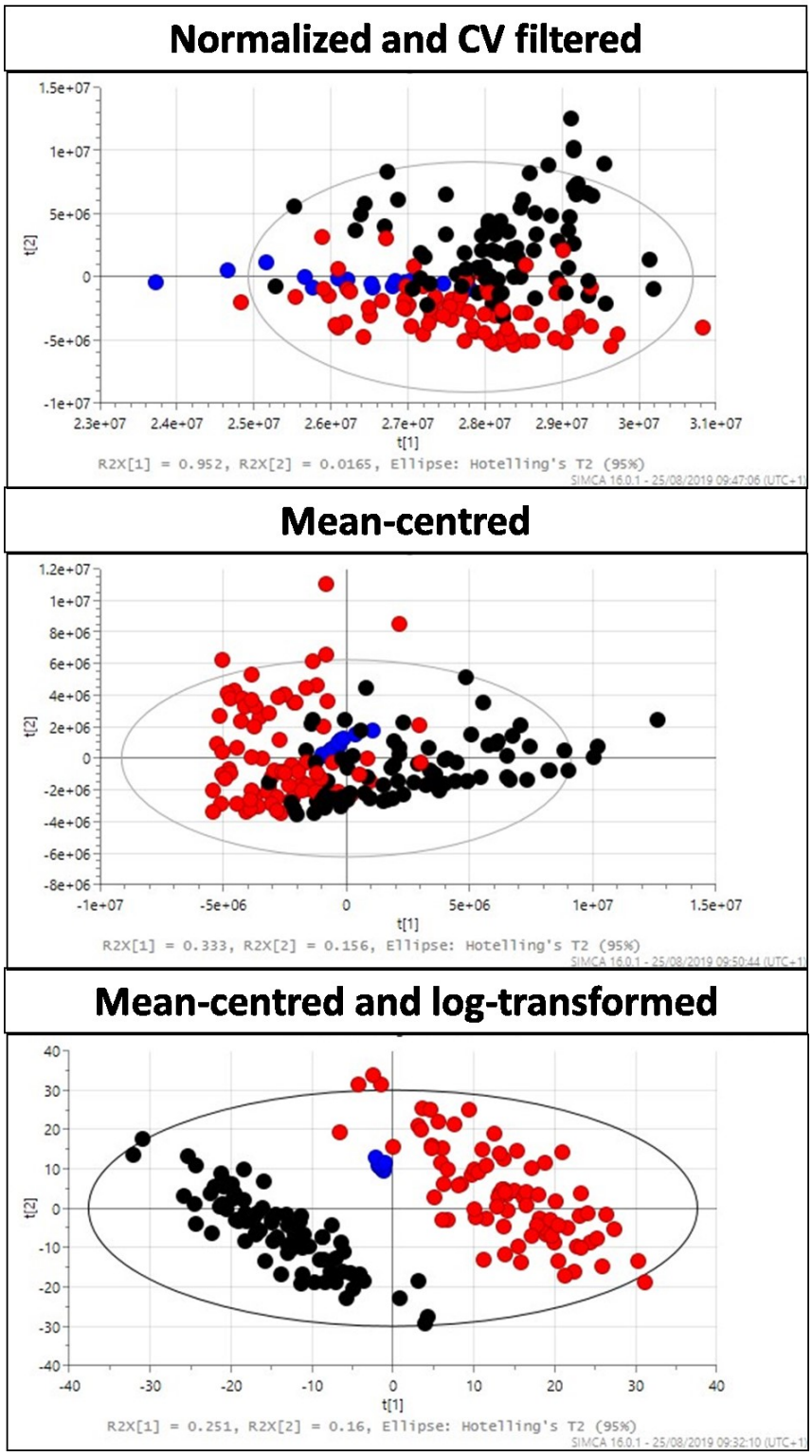



**Figure 2-13. Effect of mean-centring and $log_{10}$-transforming on PCA scores plot.** PCA score plots of - the normalized and CV-filtered datasets, where QCs are not clustered and sample separation can be observed only thanks to the colouring (black for controls and red for cases); - the mean-centred dataset, where QCs are clustered but where separation is still unclear; - the mean-centred and $log_{10}$-transformed dataset where the influence of the variables across the all dynamic range have been removed and thus a clear separation can be observed. The scores t[1] and t[2] represent the distance between a sample to the origin of $PC_1$ and $PC_2$, respectively.

Scaling, such as univariate or Pareto, can also be used. Univariate scaling divides all the features by their standard deviation across all the samples. This gives the same weight to all the features. However, in UPLC-MS, the dynamic range of a dataset is quite heterogeneous, which thus increase the influence of baseline signal and decrease the influence of high intensity signals. The use of univariate scaling was in consequence not assessed. Pareto scaling dividing each variable by the square root of its standard deviation. This reduces the influence of features with large intensities but does not amplify the importance of noise. Another scaling, or more precisely transforming, approach is $\log_{10}$ transforming the variables. As some variables are only seen at very low intensity, applying $\log_{10}$-transformation to them can induce higher variability between them than in the non-transformed data. The usual way to prevent this possible drawback is to add a constant to all the $\log_{10}$ transformed values, usually 20 for aqueous extracts (Masson et al., 2011). Figure 2-13 shows how $\log_{10}$-transformation eliminate the influence of different *m/z* values across the broad dynamic range, allowing a clear separation to be observed, in accordance with the biological classes.

The effect of both Pareto and $\log_{10}$-transformation were assessed during the studies reported herein and an example generated on how to assess this effect on a UPLC-MS dataset is shown in Figure 2-14. Ideally, the standard deviation of each mean intensity needs to be similar across all the dynamic range. It was commonly found that $\log_{10}$ transformation slightly reduced the standard deviation of the most intense features compared to the standard deviation without scaling, but Pareto scaling actually increases the standard deviation of the most intense features, which can be explained by the fact that Pareto reduced the importance given to the baseline (Figure 2-14). Mean-centring was not surprisingly the pre-processing step allowing to obtain the lowest standard deviation across the dynamic range. These figures were generated for every study in this report and in all cases, mean-centred and $\log_{10}$ transformation were both performed before the multivariate statistical analysis.

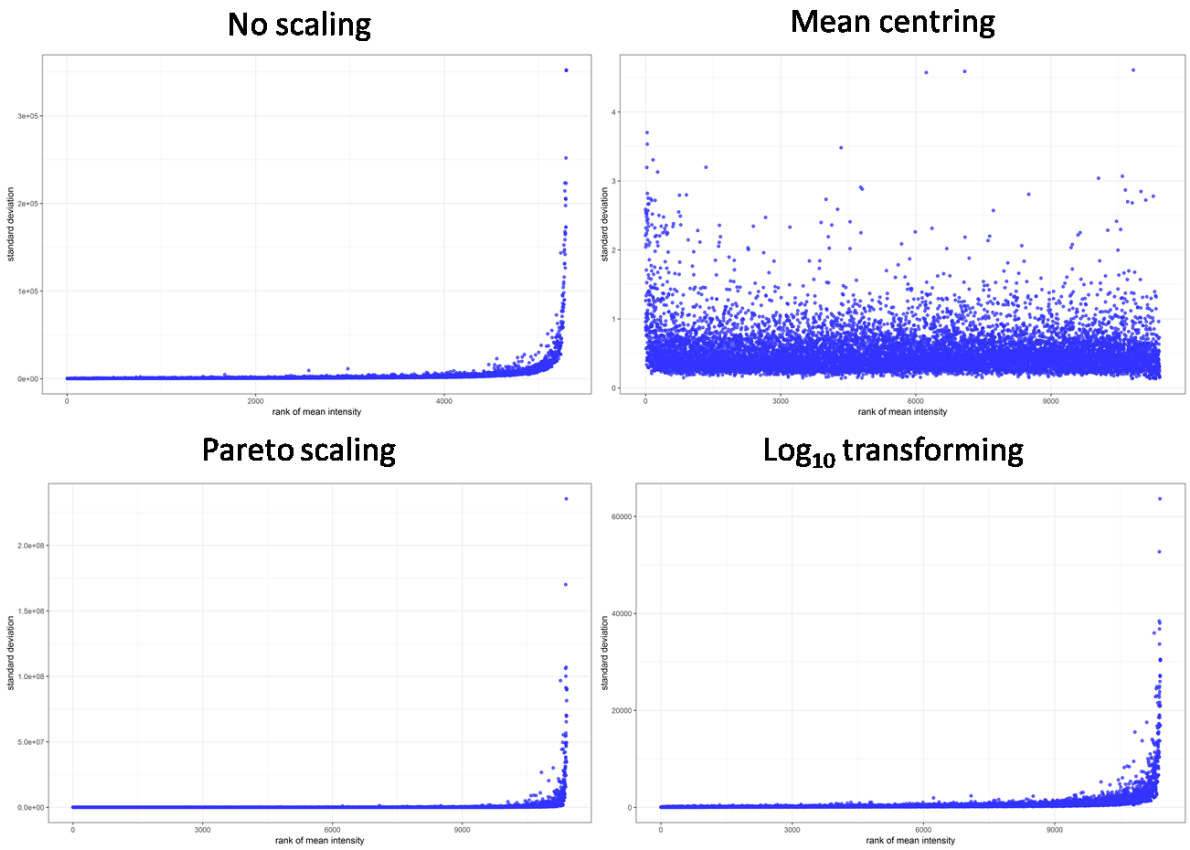



**Figure 2-14. An example of the influence of the centring and scaling on the standard deviation of an UPLC-MS dataset according to the intensity of the feature.** The more the intensity increases the more the standard deviation also increases. This observation is decreased by $log_{10}$-transforming the dataset or regulated using mean-centring. Pareto scaling did not decrease the standard deviation of high intensity features.

### *1.4.3.Supervised analysis*

A clear separation within the PCA score plot between the different classes is not always easy to obtain, particularly in complex study design or in large human phenotyping studies, where a large number of factors can influence the metabolic profile of the samples and the intervention/class is not the largest source of variance. PLS models can then be useful, as the class of each sample (*e.g.* case and control samples), usually described in a Y matrix, is known before the model gets generated. It is thus naturally seen as the regression derivatives of PCA models. If more than two classes are described in the Y matrix, the model generated is a PLS-regression (PLS-R) and if only two classes are described in the Y matrix, the model generated is a PLS-discriminant analysis (PLS-DA). PLS models minimise the variables which do not vary between the classes and maximise the variables that differ between them (Figure 2-15). The model is in consequence optimised, which allows a better interpretation of the metabolic variation between classes or associated with a response (Y) variable.

As for PCA, a PLS model can be over-fitted and thus decrease the predictive power of the model ($Q^2$). To ensure that the model has a robust predictive power, cross-validation can be performed. It is performed by dividing the samples into a training and a test set. The test set represent 1/7$^{th}$ of the dataset selected randomly. The model is then build as described above on the remaining 6/7$^{th}$ of the data (training set). The predictive power of the new model is tested by assigning the class identity of 20% of the excluded data (the test set) and assessing the error in prediction. This is repeated several times and the global result allows $Q^2Y$ to be calculated. It identifies if the initial model can accurately predict the class of the samples and if the metabolic difference between the classes is genuine. The significance of the predictive power ($Q^2Y$) of the model obtained following cross-validation can then be assessed by performing a permutation test. The permutation test randomly assign the samples to classes, generating a new "junk" dataset on which a new model is built. This new model results in a $R^2Y$ and a $Q^2Y$. This iteration is done several times (usually 1,000 times to obtain a satisfactory level of validation). The true model is found to be robust when its $R^2Y$ and the $Q^2Y$ are superior to all the ones calculated by each permutation test and is found to be significant when is $Q^2Y$ is significantly different ($p < 0.05$) from the permutated $Q^2Y$ values (Figure 2-15).

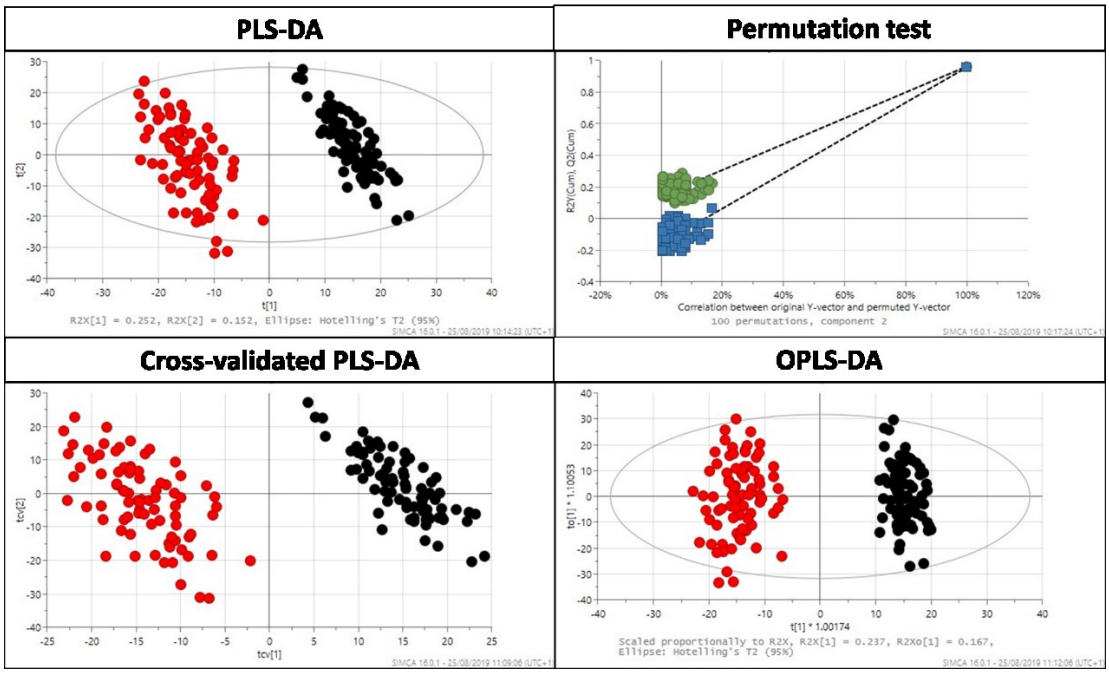



**Figure 2-15. Supervised analysis, cross validation and permutation test.** PLS-DA score plot where each point is a sample and where the class affiliation ("case" and "control" samples which are in red and black respectively) is taking into consideration to build the model. Permutation test using 100 iteration cycles and proving that the $R^2Y$ (green) and the $Q^2Y$ (blue) of the model is superior to the $R^2Y$ and $Q^2Y$ of the iteration cycles. Cross-validated PLS-DA shows the robustness of the PLS-DA, as results are similar to the not cross validated model. OPLS-DA which summarize the PLS-DA components explaining the co-variance between the samples into one orthogonal component.

Finally, when a large number of variables exhibit a common behaviour between classes (*i.e.* metabolic variation unrelated to Y), another algorithm is used to remove them from the model and generate a model that is easier to interpret, as it focusses only on those variables that are different between the classes. This model is described as an orthogonal signal correction of the PLS model (O-PLS, Figure 2-15). The predictive power of this model remains exactly the same as the PLS model. The first component is the predictive component, and the second the orthogonal one. Useful features come with these models, such as the "S-plot". An S-plot is similar to the loading plot in the sense that it represents the variables contributing to the predictive component.

CA-PLS is an extension of the PLS method and has been used in the work reported in this thesis. This technique adjusts for cofounders which can influence the model and cause bias in the distribution of the samples. This is a valuable tool in metabonomic studies where a lot of parameters can interact with the general metabolic profile of the host (Posma et al., 2018). In the case of animal studies for instance, the covariates which were adjusted for were the age and sex of the animals, animal housing, and biological or experimental replicates. A kernel density estimate (KDE) plot can be generated from the CA-PLS model which shows the distribution of the scores from each class and if they overlap. The most important parameters to look at are $R^2Y$ and $Q^2Y$, like in the previously explained models, as well as the robustness of the cross-validation (RCV), which is the ratio between the $Q^2Y$ and the $R^2Y$. Closer the RCV is to 1 the better it is. Logically, when a model as a poor RCV, this means that we can badly predict the separation between each group, meaning there is no separation between the two different groups. P-values were also used in addition of CA-PLS-DA in order to appreciate the significance of the separation between the two classes studied (further explained in paragraph 1.6).

## 1.5. Univariate Statistics

The drawback of using univariate statistics in metabolomics has been extensively discussed (Broadhurst and Kell, 2006; Saccenti et al., 2014), due to the fact that of correlation between the variables are not taken into consideration and thus the increased risk of assignment error. However, there are two cases where the use of univariate statistics in metabolomics is still seen as valuable (Broadhurst and Kell, 2006). The first one is when multivariate statistics allows the presence of metabolites which can potentially be biomarkers to be verified. Validating these findings using univariate analysis and by comparing the relative quantification of the potential biomarkers between two groups can provide more confidence in the finding, although the results between multivariate and univariate statistics can differ. The second scenario in which univariate analysis can be used to

advantage is when using targeted methods. Absolute quantification can easily be represented in box plots or column graphs and the complementary used of univariate methods allow to determine the significance of metabolite distribution between classes.

The type of univariate method to apply to the metabolite concentrations depends on their distribution across the samples. A normal distribution will allow the use of simple tests such as the student t-test or ANOVA. However, in metabolomics, variability in analytical measurement, especially in mass spectrometry force the use of non-parametric test due to the acquisition of data not normally distributed. Additionally, restricted sample group size in animal studies, as in this report, rarely allows a normal distribution, which leads to the use of non-parametric tests, such as the Wilcoxon rank sum test to perform a comparison between two classes and Kruskal-Wallis test to perform a comparison between more than two classes. The Wilcoxon rank sum test often gives similar p-values, as it is based on rank, exploring the different possibilities to gather the two classes of interest. A limited number of possibilities thus exist, and p-values will not be continuous but correspond to a limited number of values.

Output in multiple group comparisons are described as true positive, false positive or Type I error, described as a metabolite wrongly assigned as significantly different between two classes, false negative or Type II error, which is described as metabolites incorrectly assigned as being non-significant between two classes, and finally true negative. To take into consideration the high risk of errors already discussed, as it is especially troubling when a false positive leads to the wrong assignment of metabolite as a disease biomarker, the use of multiple testing comparison has been introduced (Broadhurst and Kell, 2006). Two main kinds of approach exist, either by controlling the family-wise error rate (FWER) or by controlling the false discovery rate (FDR). The FWER predicts the chances of finding at least one false positive, while the FDR calculates the expect proportion of false discovery. In other terms, FWER will adjust the p-value in order to ensure that at least 95 % of the assignments are not false positives, and FDR will adjust the p-value to ensure that less than 5 % of the assignments will be false positives (Truntzer and Ducoroy, 2017), which makes FWER more stringent. The most commonly used FWER method is the Bonferroni correction and the most common one for an FDR approach is Benjamini-Hochberg which was used in the calculations reported herein.

## 1.6. Metabolite Identification

Once significant variables or features are highlighted by the multivariate statistics, identification can be attempted in order to biologically interpret the metabolic variation induced into the two classes. Metabolite identification is described as one of the major bottlenecks in metabolic profiling studies, especially in MS-based techniques. Indeed, thanks to the robustness of $^{1}$H NMR spectroscopic analysis databases are more easily produced and more trustworthy than for LC-MS-based techniques. The workflow to identify compounds within each of the analytical techniques is described below. Nevertheless, as a general rule, feature signals are compared and matched with databases. If a compound is found to be statistically different but cannot be identified, it is an unknown compound annotated at a level 4 of confidence, the lowest according to the criteria used by the Metabolomics Standards Initiative (Sumner et al., 2007). If the chemical class of the compound can be determined, the annotation level is 3. If a compound can be determined by comparison to database, it is a level 2 of annotation. Emergence of sub-confidence groups in the level 2 annotation emerges, as annotation to a specific metabolite can be done only on one orthogonal parameter (e.g. *m/z* values matching to database) or two orthogonal parameters (e.g. *m/z* values and retention time) without spiking the corresponding authentic standard. The sub-confidence group are used herein and reported at 2b and 2a respectively. Finally, if the compound matches with at least two orthogonal parameters (e.g. the *m/z* and the retention time) compared to an authentic standard which has been spiked into a sample, then the level of confidence is 1 and the term identification can be used (although care still needs to be taken in the case of isomeric compounds e.g., 1 and 3 methyl histidine etc.) in contrast to the term annotation for the level 2, 3 and 4 (Sumner et al., 2007). It is important always to report the level of confidence at which a metabolite is annotated/identified, in order to facilitate comparison between studies and avoid any misleading biological interpretation.

### *1.6.1. Metabolite identification based on a $^{1}$H NMR spectrum*

As already discussed, the robustness of $^{1}$H NMR spectroscopy, and transferability of interlaboratory/database spectra allowed the chemical shifts of an extensive list of common biological compounds from in-house or on-line databases (e.g. the human metabolome database (HMDB Wishart et al., 2018), or Metlin (Smith et al., 2005) or reported in the literature) to be used for identification purposes. Software where the NMR spectrum can be aligned and peaks fitted to a compound reference spectrum, such as a the Chenomx NMR suite, are also available whilst they are

commercial and often expensive. Usually, and particularly for well-studied matrix such as urine, a major part of the metabolite can be assigned by database matching. However as overlapping between compounds are often observed in $^{1}$H NMR spectroscopy, poor resolution of the signals often precludes metabolite annotation with absolute certainty. Statistical analysis such as statistical total correlation spectroscopy (STOCSY; Cloarec et al., 2005), allows the correlated chemical shifts coming from the same metabolite to be discovered through the use of a colour code thereby facilitating interpretation. J-resolved spectra (J-Res) are routinely acquired and these give insight into the coupling constants (J) of the signals, which allows the multiplicity of each signal to be more clearly assessed. Two-dimensional NMR analysis, as explained in the NMR theoretical section above, can also be used to help annotation. Only $^{1}$H 1D NMR spectroscopy and J-Res datasets were used in the experiments reported here.

### *1.6.2. Metabolite identification based on UPLC-MS data*

Mass spectrometry-based metabolite identification is obtained by matching the parent and the fragment's *m/z*'s to known compounds gathered in in-house or in on-line databases, such as those also used for NMR spectroscopy, or LC-MS specific ones, such as Lipidmaps or MoNA. CEU Mass Mediator was used here as a tool to explore different databases simultaneously (Gil-de-la-Fuente et al., 2019). The advantage of having the daughter ions *m/z* is that it allows the reconstitution of the fragmentation pattern of the molecule (which depends on the collision energy applied). To obtain fragmentation information, several modes of MS acquisition are available. Untargeted profiling on Waters mass spectrometers is usually performed in MS or $MS^{E}$ acquisition mode. $MS^{E}$ has the advantage of sequentially applying high and low collision energies to peaks entering the mass spectrometer, which allows the collection of both parent fragment ions, (Figure 2-9.b). This is useful for metabolite identification, although as several parents can be eluting at the same retention time, attribution of the daughter-parent couples is often challenging. However, it is easily applicable to any data and the only drawback is the increased size of the raw data generated. Data-dependant acquisition (DDA) can also easily be applied to a pool samples during an analysis and enables the selection (in the quadrupole) of only those ions with *m/z* with an intensity higher than a specified value, which are then fragmented (in the collision cell). However, in consequence, a significant loss of sensitivity is observed and that is why it is performed only on one pooled sample rather than the whole set of samples. It also assumes than the features of interest are contained in the high intensity features. However, even if this it is not the case DDA can always give useful fragmentation insight on

the most abundant metabolites in a sample. Another mode of acquisition, also based on the isolation of compounds in the quadrupole before fragmenting it, is tandem MS, also called MS/MS. This is performed only once the samples have already been analysed by MS or $MS^E$ mode and that a list of features of importance (e.g. significantly changed between two subject classes) has been identified by multivariate statistical analyses. This list of features is then specified by the analyst while setting up the MS method and it will provide confirmation (1) that the features are present and (2) show they are fragmented, which will allow matching with databases at a level 2 of annotation, and with more certainty than with $MS^E$ which can fragment several parents at the same time.

Several strategies were used to match features of interest for which DDA, $MS^E$ or MS/MS data were obtained. Firstly, the use of an in-house database (NPC database supported by the MRC and NIHR) built on the same RP UPLC-MS method applied to the experiment reported here, which allowed the putative annotation of features based on both their *m/z* and their retention time. However, the annotation made on a feature (*m/z* and retention time pair) list by an algorithm, particularly around the dead volume, are often found to be either incorrect or there are several different annotations for the same feature. It is crucial therefore that each annotation is verified by the analyst by exploring the raw data and comparing it with experimental data generated on the standard. Statistical methods such as correlation across the features is also a valuable tool as it highlights the presence of features which belongs to the same metabolite. Indeed, as a single metabolite can produce several product ions, such as dimers, adducts, isotopes or fragments, a list of important features that in fact all come from the ions generated by the same metabolite can be reported by multivariate statistics as significant. Matching each one of them through accurate mass searches of databases would be time-consuming and could lead to false assignments. Taking into consideration the correlation between all of these features will gather together all the ions derived from a single metabolite and simplify data analysis. Although this is a useful tool, it can sometimes result in incorrect correlations, also miss-correlations, being reported. As always, checking the raw data needs to be done to ensure the reliability of the correlation, or even identifying new ones. To do so aligning extracted ion chromatograms of *m/z* signals observed in a same mass spectrum and validate that they elute at exactly the same retention time increases the chances that these ions are coming from the same molecule (Figure 2-9.c). If no match was found based on the in-house database or through accurate mass search on on-line databases, whilst a correlation is still shown between different features, *in silico* searches based on predicting CID (collision-induced dissociation) and matching not only the parent *m/z* but also the daughter's *m/z* can be useful. However, this kind of tool rarely gives successful hits and should also be considered as providing a lower confidence level of annotation. In the worst-case scenario, i.e. when no hit has been found through databases or *in silico* search, exploring the

neutral losses between different *m/z* might give insight in the fragmentation pattern and the compound class, such as the loss of 176 Da between two co-eluting *m/z* would naturally be explained by the loss of a glucuronic acid. Finally, once as many as metabolites possible have been assigned, factors such as the relationship between their observed retention time and their polarity (log P) can be explored, and most importantly the biological meaning of the metabolite annotation must also be validated.

## 1.7. Comparison between $^{1}$H NMR spectroscopy and UPLC-MS

$^{1}$H NMR spectroscopy is a non-invasive technique, as the sample can be recovered and used in a following experiment (providing that the sample preparation needed for NMR analysis such as $D_2O$ addition, does not interfere with the following experiment). In contrast, ESI UPLC-MS analysis, due to the nature of the technique, is destructive but, as only relatively small volume is required for the UPLC-MS analysis (as low as few µL of an often very diluted sample) this does not necessarily cause problems. NMR spectroscopy also present the advantage of being quantitative if correctly calibrated to the reference signal prior the analysis (e.g. TSP signal), and methodology is both reproducible and easily transferred between laboratories. However as previously discussed, the main drawback of NMR spectroscopy is its relative lack of sensitivity, together with the potential for extensive signal overlap in the $^{1}$H NMR spectrum of a biological samples and in consequence making the metabolite identification more difficult. UPLC-MS has potentially greater sensitivity, but this is structure dependent and viable even within classes, and the potential for signal overlap due to analyte co-elution is also present. Also, multiple ionization modes (positive and negative) and chromatographic systems (RP and HILIC) are needed to increase metabolite coverage. In addition, there is still room for improvement in terms of reproducibility and reproducibility. Moreover, the complete coverage of the metabolites contained in a biological sample is impossible to obtain by using only one analytical method. This is due to multiple factors (sensitivity limitations, the loss of metabolites during sample collection, handling, preparation and analysis (e.g. non-retention/non-elution)). In consequence, the use of both NMR spectroscopy and MS-based techniques through an integrated platform seem to be a reasonable option, to maximize metabolome coverage, facilitate metabolite identification and built more robust models through the use of multiple dataset integrations.

# 2. Amplicon sequencing

## 2.1. Principle

Amplicon sequencing, or 16S rRNA gene sequencing, aims to study microbial communities within a biological sample. The 16S rRNA gene is conserved in all the prokaryotes and does not significantly evolve, which makes it a useful phylogenetic marker. It is composed of nine variable regions and those regions are separated by conserved regions (Figure 2-16). The analysis is based on DNA extraction and quantification, amplification by polymerase chain reaction (PCR) and sequencing of targeted hypervariable regions on the 16S rRNA gene. The conserved regions around the targeted variable regions serve as primers for the sequencing. The targeted regions are often V1/V2 or V3/V4, which have a restricted number of base pairs (bp) compared to the total base pairs of the gene (~1,500 bp).

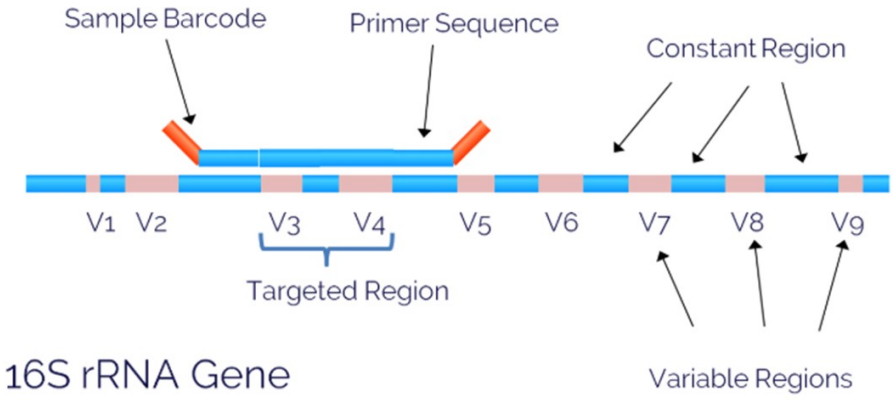



**Figure 2-16. Representation of the 16S rRNA gene, composed of conserved region (blue) and variable regions (pink).** Source: 16S rRNA gene sequencing webpage from LC Sciences.

Different platforms exist to perform the sequencing but a common one, and the one used herein, is Illumina's MiSeq Platform. Once the sequences have been generated, bioinformatics is used to pre-process and analyse the data. Pre-processing is usually based on the operational taxonomic units (OTUs) clustering method based on similarity of the targeted hypervariable regions (Figure 2-17). Once the OTUs are clustered, they are compared to databases such as ribosomal database project (RDP) or Silva database. Due to the limited number of regions targeted and the fact that some bacteria have variable regions which share a high degree of similarity between each other, the identification of bacteria to species level is not possible. Identification to the genus level is usually made by using an arbitrary value of 97 % of similarity between the variable regions targeted of the organisms present in the samples and the ones found in databases.

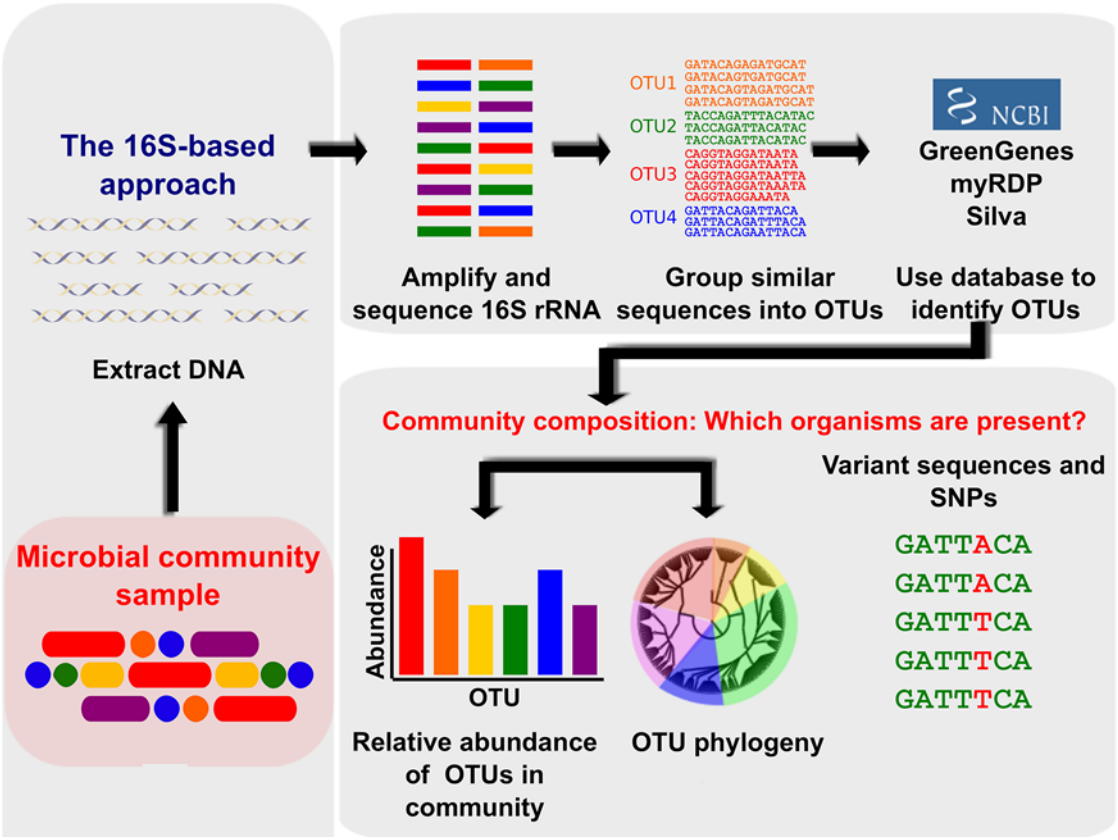



**Figure 2-17. Amplicon sequencing workflow.** It starts by DNA extraction, followed by the amplification of the 16S rRNA sequence and sequencing, prior OTUs clustering based on sequences similarity and identification by matching to databases. Finally, the microbial community of the sample is investigated. Adapted from **(Morgan and Huttenhower, 2012)**.

Whilst it is commonly applied, consensus around a common pipeline has been difficult to achieve. The use of another method based on amplicon sequence variants (ASVs) instead of OTUs is also being discussed and seen by some as a more definitive, reproducible and comprehensive technique (Callahan et al., 2017). ASV-based methods being quite recent at the time of the amplicon sequencing experiment performed here, is in need of further evaluation by the microbiome-focused community, the pre-processing was performed using the more traditional method based on OTU clustering.

## 2.2. Data pre-processing

The pre-processing can be performed using different bioinformatics pipelines, such as QIIME, UPARSE or mothur for OTUs-clustering method or DADA2 for ASVs methods. The one used for the pre-processing of the data from the amplicon sequencing experiment reported in Chapter 3 was mothur (Liu et al., 2012). It consists of three main steps. The first is removing the errors that could have been induced through the PCR or the sequencing. To do so, the two fastq files generated following the sequencing, which correspond to the pair-end reads (one forward and one reverse), are paired up. Any pairs with an ambiguous base-pair (*i.e.* reads which do not overlap) are then removed.

The next step is to process the improved sequences obtained in the previous step. It starts by removing the duplicated sequence to save time during the alignment. The alignment is done to the reference bacterial database SILVA. Trimming poorly aligned sequences, which is defined by sequence starting before the 25%-tile or finishing after the 75%-tile of all the sequences, is then performed, but this can re-introduce duplicates, which are once again removed. A first pseudo-clustering step is performed to check for the presence of chimeras induced by PCR and these are removed. The sequences are then organised according to their abundance, before being classified according to the RDP using the Wang approach (Wang et al., 2007). Any sequences classified as not bacterial are removed. Finally, an output file is created where the cluster of the OTUs is reported and their specific sequences are binned into phylotypes to obtain OTU taxonomy. Further analysis of the OTU-clusters obtained can be performed within the Mothur pipeline, but here it was exported as a non-normalised matrix into Phyloseq.

## 2.3. Data analysis

Phyloseq is an R-based package which offers a broad range of valuable tools. This was found to be more flexible than other available packages (McMurdie and Holmes, 2013). As for metabolomics, amplicon sequencing generates extensive matrices with N number of samples and K variables, which here are OTUs. For each sample, the number of reads (also called the sequencing depth) will vary extensively, which can induce bias in the abundance and diversity analysis. This is frequently discussed in the literature and, in consequence brings discussion to the matter of normalisation.

Several normalisation methods are available, but the two most common are proportion, also called total sum scaling, which divides the number of counts per OTU by the total number of counts in the sample, and rarefying. Rarefying is by far the most criticized (McMurdie and Holmes, 2014). Its principle is based on cutting the sequencing depth of all the samples to the sample presenting the smallest sequencing depth. However, this can be quite variable from one sample to another, which can result in a large proportion of the data generated being discarded. However, rarefying is a compulsory step before exploring the α diversity, which measures the diversity within samples before comparing it to other samples. Indeed, if a sample presents a high sequencing depth, and in consequence coverage, it is very likely that it will represent a more diverse bacterial community compared to a sample with a lower sequence coverage. Several indices can be used to study the α diversity, such as Chao1 which focuses on the richness, which is a measure of the total number of different species present in a sample, or Shannon and Simpson, which focus on the richness and the evenness, which compares the proportion of each species in a sample. The different indices were investigated in the experiment reported here and the results were similar between indices.

Comparing the degree of similarity between samples (β diversity), as well as exploring the taxonomic profiles in a sample class are two important points to study following amplicon sequencing. But to avoid ignoring a large part of the data, the use of other normalisation methods other than rarefying was investigated. Different packages are available to explore sequence read counts, such as edgeT, edgeR, metagenomeSeq, baySeq, DESeq2, ALDEx2, and they contain their own normalisation methods. Comparison of these packages has already been performed (Weiss et al., 2017; Thorsen et al., 2016) and although some variations were observed between them regarding false positive rate and how they act in the presence of high sparsity, the normalisation effect was negligible compared to the distance metric used to explore β diversity. Therefore, proportion normalisation was used before exploring different distance metrics, such as unweighted UniFrac, weighted Unifrac, or Bray-Curtis, and coordination methods, such as principal coordinate analysis (PCoA) and non-metric multidimensional scaling (NMDS) (Figure 2-18). The outcome was similar for each possibility. However, dissimilarity was chosen to be explored through the Bray-Curtis metric, although it does not consider phylogeny compared to Eucledian Unifrac metrics, as it is not affected by null values. PCoA and NMDS were preferred over PCA as it allows the use of similarity distance such as Bray-Curtis. No major difference was seen between PCoA and NMDS ordination (Figure 2-18). Finally, p-values can be calculated to appreciate the significance of the β diversity distribution by using statistical tests such as permutational multivariate analysis of variance (PERMANOVA).

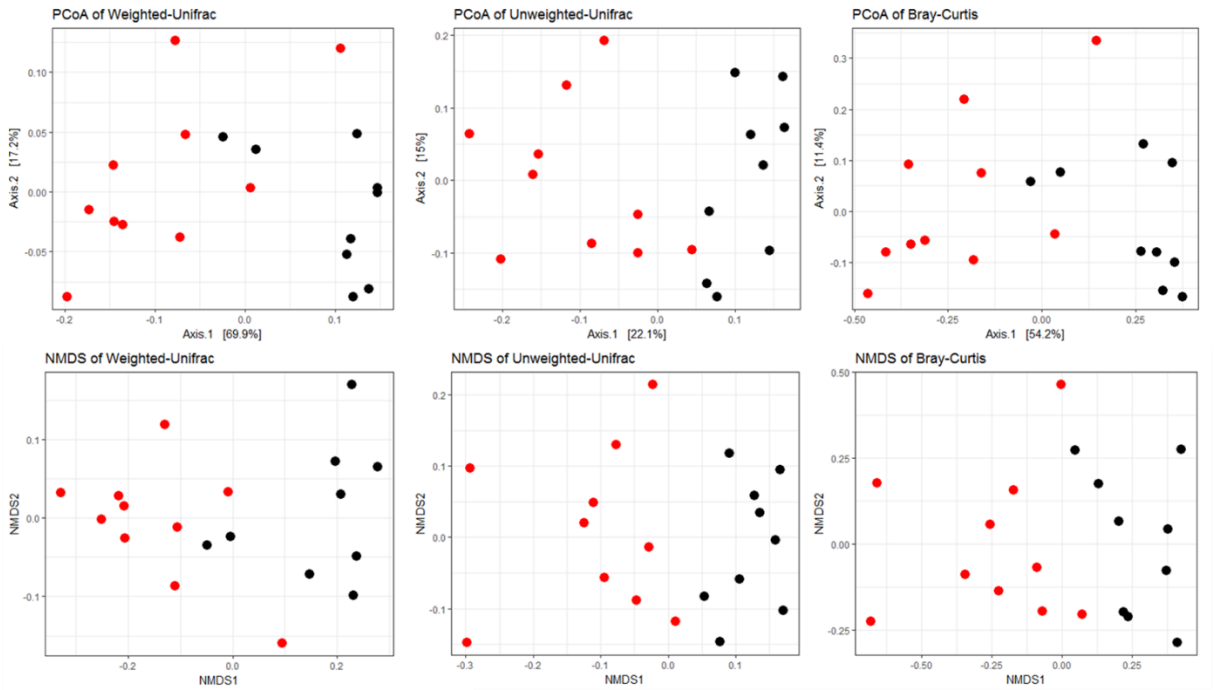



**Figure 2-18. Investigation of the influence of distance metric and ordination method on β diversity.** The distance metrics used were weighted-Unifrac, unweighted-Unifrac and Bray-Curtis and the ordination methods were NMDS or PCoA. No real difference was observed between the repartition of the case group samples (red) compared to the control samples group (black). The percentage displayed for each axis represents the percentage of variation.

# Chapter 3. Investigating the two-way interaction between methotrexate and the gut microbiota in Sprague Dawley rats

**Aim:**

- To explore the impact of methotrexate on the metabolic and microbial profile as well as the impact of gut microbial metabolism on methotrexate detoxification and how the two actions interact.

**Objectives:**

- Apply untargeted and targeted UPLC-MS approaches to explore the impact of methotrexate on the global urinary and faecal metabolic profiles of rats and to quantify a microbial-derived methotrexate metabolite and determine its excretion.

- Use amplicon sequencing to characterise the dose-dependent impact of methotrexate treatment on the faecal microbial profile of rats.

# 1. Introduction

Methotrexate (MTX, amethopterin, 4-amino-4-deoxy-*N*-10-methylpteroylglutamic acid) is a folate analogue used to treat a range of different pathologies. At doses < 15 mg/kg MTX is used to treat autoimmune diseases such as rheumatoid arthritis (RA) or psoriasis. Higher doses, usually between 15 and 1000 mg/kg (Nathan et al., 2006), are used in the treatment of neoplastic diseases, mainly for acute lymphocytic leukaemia, lymphoma and osteosarcoma (Rattu et al., 2013). At low doses, MTX demonstrates anti-inflammatory properties, but with prolonged use, or with high doses, it can induce nephrotoxicity, life-threatening myelosuppression, gastrointestinal toxicity and hepatotoxicity (Buchen et al., 2005). MTX is metabolized in the liver by aldehyde oxidase (AO) into 7-hydroxy-MTX (7-OH-MTX) and both the parent compound and the metabolite are transported into cells by carrier proteins (Genestier et al., 2000; Goldman and Matherly, 1985). These metabolites can undergo polyglutamation and the products of this reaction inhibit dihydrofolate reductase (DHFR) and influence folate metabolism (Allegra et al., 1985). This inhibition blocks tetrahydrofolate (THF) production from folic acid, inhibiting purine and pyrimidine synthesis and downstream DNA synthesis (Allegra et al., 1987).

The intestinal microbiota is increasingly recognised to participate in the processing of various xenobiotics and influences host responses to various compounds (Carmody and Turnbaugh, 2014; Clarke et al., 2019). However, this understanding remains limited. The gut microbiota can directly participate in drug metabolism, as occurs with MTX. Following absorption and metabolism of MTX and 7-OH-MTX, these molecules are excreted into the bile and subsequently into the gastrointestinal tract (Figure 3-1). Here, they undergo further processing by the gut microbiota (Zaharko et al., 1969, Valerino et al., 1972). A bacterial enzyme, carboxypeptidase glutamate 2 (CPDG2; also referred to as folylpolyglutamate carboxypeptidase [FGCP]), can cleave the terminal glutamate residues of MTX and 7-OH-MTX (Valerino et al., 1972; Schmiegelow, 2009) resulting in metabolites such as 2,4-diamino-*N*-10-methylpteroic acid (DAMPA) and 7-hydroxy-DAMPA, respectively (Widemann et al., 2000). As these metabolites do not have activity towards DHFR inhibition, glutamate hydrolysis by CPDG2 is seen as a detoxication process (Donehower et al., 1979; Widemann et al., 1999). As such, CPDG2 is approved as a rescue agent in cancer patients presenting delayed MTX clearance and acute nephrotoxicity (Rattu et al., 2013). Several clinical studies have since reported positive results of the use of CPDG2 as a rescue agent (Buchen et al., 2005; Widemann et al., 2012; Svahn et al., 2017), although MTX and its relatively insoluble metabolites are still presume to precipitate in kidneys, causing nephrotoxicity (Buchen et al., 2005), as well as causing severe gastrointestinal toxicity such as vomiting, diarrhoea, oral mucositis and, in some cases, resulted in the death of patients (Buchen et al., 2005; Shea et al., 2013).

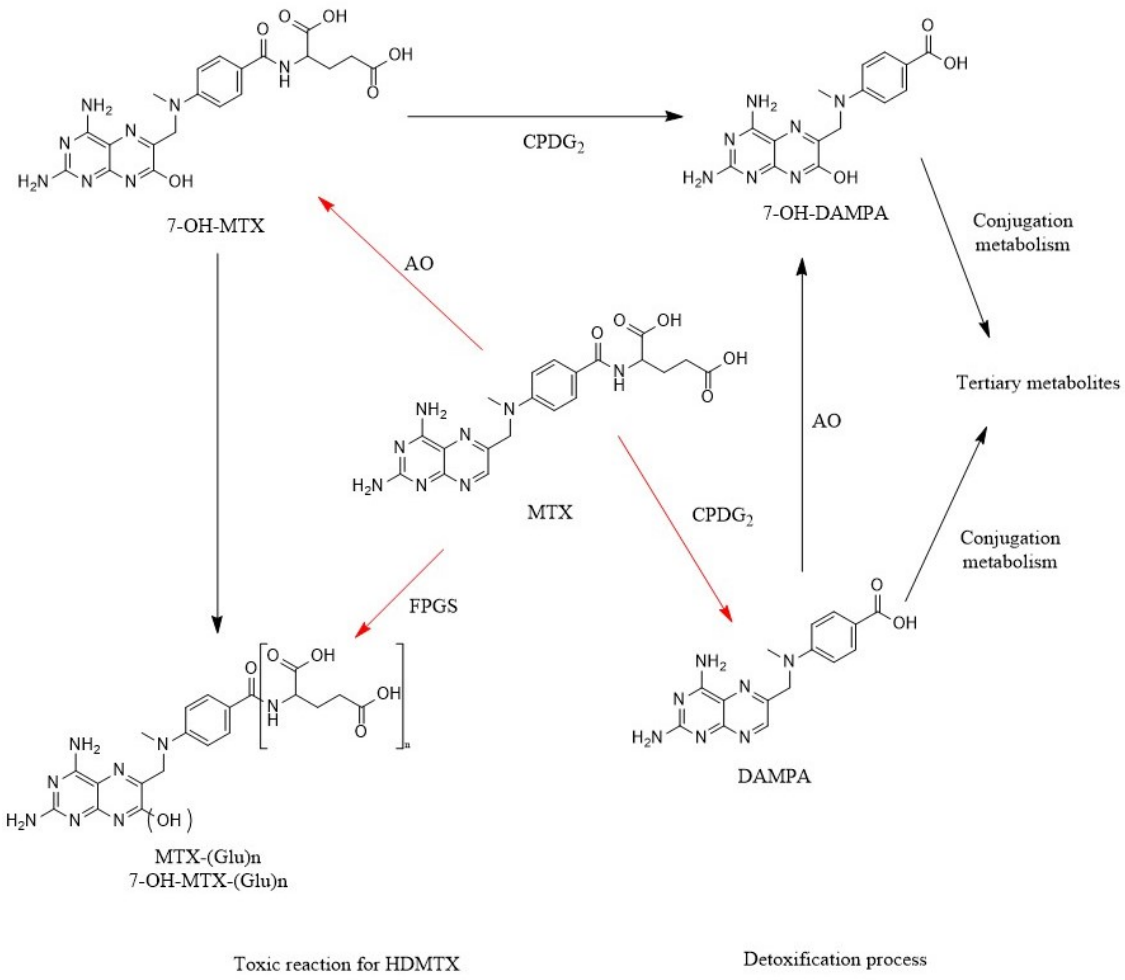



**Figure 3-1. Summary of MTX metabolism.** Red arrows represent primary metabolism (AO: Aldehyde oxidase, CPDG2: Carboxypeptidase glutamate 2, FPGS: folypolylglutamate synthetase) from MTX towards the active metabolite 7-OH-MTX and their polyglutamated form, as well as towards the inactive metabolite DAMPA. Secondary metabolism can occur to give 7-OH-DAMPA. Tertiary metabolism is suspected, but not detected, on the DAMPA forms, towards sulfated, or glucuronidated conjugates, as reported inhuman **(Widemann et al., 2000)**.

Therapeutic treatment can also impact the community structure of the gut microbiota modulating its diversity and functionality. MTX has recently been reported to inhibit the growth of 30 % of 40 representative gut bacterial strains (Maier et al., 2018) and of 84 % of 43 bacterial isolates which had a combined relative abundance covering 43 % of the human gut microbiota (Nayak et al., 2019). Studies exploring the impact of MTX on the gut microbiota have previously focused on the anti-inflammatory role of MTX in RA. At low doses (*i.e.* below 15 mg/kg), MTX in RA patients increased the species richness and diversity of the microbiota and restored the perturbation of the microbiota associated with RA (Zhang et al., 2015; Chen et al., 2016). In a recent study (Nayak et al., 2019), MTX treatment (10 or 50 mg/kg) of humanized gnotobiotic mice decreased the relative abundance of Bacteroidetes and increased that of the Firmicutes. Such alterations were further demonstrated as potential interaction with off-target bacterial enzymes, such as DHFR in *Escherichia coli* or *Lactobacillus casei*, as DHFR enzyme is conserved in all domains of life (Bolin et al., 1982). However, higher doses of MTX (90 mg/kg) administrated to rats with conventional microbiota caused mucositis and a reduction, by 705-fold, of bacterial diversity, compared to untreated control rats (Fijlstra et al., 2014). Using a cell enumeration method (flow-FISH), the majority of the alterations were seen to have occurred via a relative decrease in anaerobes belonging to the Firmicutes, accompanied by a relative increase in *Bacteroides*. This was in contrast to the study performed in humanized gnotobiotic mice, which showed a decrease in abundance of the phylum Bacteroidetes (Nayak et al., 2019). In a rat model, abundance of *Bacteroides* increased by 49 %, whilst a decrease (by 58 %) was seen in bacteria including *Clostridium*, *Ruminococcus*, *Eubacterium* and *Bifidobacterium* (Fijlstra et al., 2014). By perturbing the gut microbiota, high doses of MTX may modify the abundance, or functionality, of bacterial enzymes responsible for MTX hydrolysis, such as CPDG2, which could result in downstream consequences for the inactivation/detoxification of MTX into DAMPA. Initial CPDG2 concentrations in such a bidirectional interaction could perhaps explain differences in patient therapeutic responses to MTX, and also toxicities. As such, a dose-dependent two-way interaction is proposed between MTX and the gut microbiota.

To further explore the bidirectional interaction between MTX and the gut microbiota, the impact of drug-free dose vehicle and three different doses of MTX (low [10 mg/kg], medium [40 mg/kg] and high [100 mg/kg]) were compared in male Sprague Dawley rats. The faecal microbiota was characterised using 16S rRNA gene amplicon sequencing, while the urinary and faecal metabolic phenotypes were measured using a UPLC-MS-based metabolomic approach. In addition, the metabolism of MTX was investigated using a semi-quantitative UPLC-MS assay.

# 2. Materials and methods

## 2.1. Chemical and reagents.

For sample preparation and UPLC-MS, LC-MS Chromasolv acetonitrile (ACN) and methanol (MeOH) were purchased from Honeywell (Seelze, Germany), Optima LC/MS Grade water with 0.1 % formic acid (FA) and analytical reagent grade FA were from Fisher Scientific Ltd (Loughborough, UK). For UPLC-MS mass accuracy and calibration, leucine enkephalin acetate salt (hydrate) and sodium formate, as well as sodium azide for faecal sample preparation were obtained from Sigma-Aldrich (Gillingham, U.K.). For chemical standards, MTX, DAMPA, 7-OH-MTX were obtained from Toronto Chemicals (Toronto, Canada).

## 2.2. Study design.

As reported in detail elsewhere (Hardwick et al., 2014) the animal study was performed at the University of Arizona and was approved by the Institutional Animal Care and Use Committee (IACUC) and in accordance with the NIH guideline. Male Sprague Dawley rats (*n* = 22), 7-8 weeks old, weighing between 200 and 250 g (Harlan Laboratories, Indianapolis, IN) were fed *ad libitum* with a control diet (choline sufficient and amino acid-defined) (Dyets, Bethlehem, PA) and separated into four MTX dose-related sub-groups (0, 10, 40 and 100 mg/kg, *i.p.*, dissolved in 0.3 M sodium bicarbonate). The animals were randomly placed 12 h prior to MTX administration into metabolism cages. Urine samples were collected between -6 and 0 h (pre-dose, 0 h), 6 and 12 h (12 h), 18 and 24 h (24 h) and between 36 and 48 h (48 h); a total of 5 samples per time point were collected. Faecal samples were collected between -12 h and -6 h and between -6 h and 0 h and combined for analysis (pre-dose, 0h), and between 6-12 h (12 h) and 36-48 h (48 h) post dose. Due to dose-dependent diarrhoea in some treated rats, it was not possible to collect all of the post dose faecal samples (see the results section for details).

## 2.3. Faecal DNA extraction for amplicon sequencing.

DNA extraction was carried out on faecal samples (240 mg [+/- 27 mg]) using the Qiagen QIAamp PowerFaecal DNA kit (Mo Bio, Carlsbad, CA, USA) according to the manufacturer's protocol, with the exception that the homogenisation of the samples was achieved using a Bullet Blender Strom (speed 8, 3 min) (Chembio Ltd, St Albans, UK). The Qubit dsDNA BR assay kit (Life Technologies Ltd, Paisley, UK) was used to quantify the extracted DNA. Dilution of each sample was performed to obtain a final concentration of 5 ng/mL. Amplification of the V1 and V2 regions of the 16S rRNA gene were performed using the primers 341F (forward) and 805R (reverse), as previously reported (Mullish et al., 2018). Library pooling and quantification was carried out with NEBNext Library Quant Kit for Illumina (New England Biolabs, Hitchin, UK) and the pooled libraries were denatured before sequencing, which was performed on an Illumina MiSeq platform (Illumina Inc., Saffron Walden, UK) by using the MiSEq Reagent Kit v3 (Illumina) and paired-end 300 bp chemistry.

## 2.4. Faecal water preparation for UPLC-MS analysis.

Faecal samples were defrosted and 50 mg (+/- 1.4 mg) were weighed for the first biological replicate and 51 mg (+/- 1.3 mg) for the second. Samples were mixed with 1 mm diameter zirconium beads (Stratech Scientific Ltd, Ely, UK) and 1.1 mL of $H_2O$ (3.3 mM $NaN_3$) before being vortexed and homogenised with a Precellys 24 bead beater (40 s per cycle, 6,500 Hz speed, 2 cycles). After 20 min centrifugation (17,000 *g*) the supernatants were collected and 1.1 mL of $H_2O$ (3.3 mM $NaN_3$) was added to the residue, followed by an additional step of bead beating and centrifugation to optimise metabolite extraction. The two supernatants were combined and 250 µL of the combined faecal extract were transferred to a 96-well plate. For quality control (QC) purposes, QC samples were prepared by mixing 150 µL of each sample from the animals which had not received any MTX to provide a pooled sample (Gika et al., 2007; Broadhurst et al., 2018) from which 250 µL was also placed into the 96-well plate. The samples were concentrated under a nitrogen flow for 3 h. Then 120 µL of MeOH:$H_2O$ (1:1) was added to each sample and the plates were centrifuged (5 min, 700 *g*). For each centrifuged sample 100 µL was diluted with 100 µL of acidified $H_2O$ (0.1 % FA) and transferred, in a randomised order, into 96-well plates and placed into the LC-MS autosampler at 4 ˚C.

## 2.5. Urine sample preparation for UPLC-MS.

Urine samples were prepared as described previously (Gray et al., 2016). Briefly, 20 µL of urine samples were mixed with 60 µL of MeOH and stored at -20˚C overnight for protein removal. Samples were centrifuged (5 min, 15,000 *g*, 4˚C) and 25 µL of supernatant was transferred into a 96-well plate in a randomised order. For a QC sample a 30 µL aliquot each control sample was pooled. A volume of 225 µL of water was added to each sample and the plates were centrifuged (5 min, 700 *g*) before transfer to the autosampler at 4 ˚C.

## 2.6. UPLC-MS.

Analysis was performed as previously described (Lewis et al., 2016) on a Waters Acquity I-class UPLC system (Waters Corp., Milford, MA) with separation on a HSS T3 1.8 µM column (2.1 mm i.d. x 150 mm). The temperature of the column was set at 45 °C with the autosampler temperature at 4 °C. The volume of sample injected was 5 µL. The mobile phase used was water with 0.1 % FA (v/v) (solvent A) and ACN with 0.1 % FA (v/v) (solvent B). The initial solvent composition was composed of 99 % of A with a flow rate of 0.6 mL/min. For gradient chromatography, the proportion of solvent B was set to increase from 1 % at 0.10 s to reach 55 % at 10 min. The percentage of solvent B was then increased over 0.65 min to 100 % while the flow rate was increased to 0.8 mL/min. This solvent composition was maintained for 1 min as a wash step followed by a re-equilibration step of 1 min with 99 % of solvent A (total run time 12.65 min). Prior to the commencement of the sample analysis the column was conditioned with 20 QCs at the beginning of the run (Gika et al., 2007; Broadhurst et al., 2018) and a QC was injected every 11 samples throughout the analysis to monitor the repeatability of the analysis. Mass spectrometry was performed using a Synapt G2-S mass spectrometer (Waters Corp., Wilmslow, U.K) conducted with electrospray ionization in positive mode (ESI+). The capillary voltage was 1.5 kV and the source temperature was set at 120 °C. The cone gas flow was 50 L/h and the gas used was nitrogen. The desolvation gas temperature was 450 °C at a flow rate of was 900 L/h. The nebulizer gas flow was 6 bars. The acquisition was carried out over the *m/z* range 50-1200. Centroid mode with $MS^E$ acquisition was applied to the experiment with a collision energy ramp of 15 to 45 eV. Leucine encephalin (MW = 556.27 Da, 200 pg/µL in ACN:$H_2O$ 1:1) was used to monitor mass accuracy with a scan collected every 60 s and a cone voltage fixed at 30 V. The data were collected using MassLynx V 4.1 (Waters Corp., Manchester, U.K.). In order to perform semi-targeted analysis for MTX and its metabolites, calibration curves for MTX, DAMPA and 7-OH-MTX were prepared by addition of the analytes to the faecal water (0.05, 0.5, 2, 8, 10, 16, 20 µg/mL) and urine matrices (0.1, 2, 8, 10, 16, 20, 30 µg/mL). Run order and batch effects were assessed using the QC data.

## 2.7. Data Analysis for the amplicon sequencing.

The sequence reads obtained from the faecal DNA extracts were analysed using the Mothur package (v.1.39.5.- Liu et al., 2012) based on the MiSeq SOP pipeline (Kozich et al., 2013). The Silva bacterial database (v.128) was used to process the sequence alignment, and the RDP (v.2014) was used to classify sequences based on the Wang approach (Wang et al., 2007). The entire dataset (number of reads generated per samples reported in Table 3-S1) was exported from Mothur into R (3.4.0 through RStudio v. 1.1.463) where the package phyloseq (v.1.26.1- McMurdie and Holmes, 2012) was used to statistically explore the data. To assess β diversity and taxonomic profile, normalisation was based on the proportion method, which scales each count in a sample by dividing by the counts sum in this sample (McMurdie and Holmes, 2014). Rarefying to the minimal depth of reads observed in the samples (*N* = 4406 reads) was performed to explore the α diversity. Illustration of the taxonomic profile agglomerated at family level was generated for each of the MTX dose groups and by comparing pre- and post-MTX treatment. β diversity was assessed using PCoA of Bray-Curtis dissimilarity matrix and α diversity was assessed using the inverse Simpson index. Drug-related effects such as diarrhoea allowed only a small quantity of faeces to be collected post-dose, often below the amount required for DNA extraction for amplicon sequencing. As a consequence, a low number of samples were available for some groups precluding statistical analysis.

## 2.8. Targeted UPLC-MS data analysis.

The quantification of MTX, DAMPA and 7-OH-MTX by UPLC-MS was achieved using a linear calibration curve and a $\log_{10}$ transformation of the axis, using TargetLynx V 4.2 (Waters). The $R^2$ for the calibration curve of each of the standards was greater than 0.98 (Figure 3-S1). As the faecal samples were prepared in duplicate, concentrations of each of the analytes were averaged. As the total volume of urine samples collected was recorded, it was possible to calculate the total amount of MTX excreted in the urine. However, the total weight of the faecal samples was not recorded and so the quantification of MTX and its metabolites excreted in faeces are reported as the mass of the analytes (μg) per 50 mg of faecal sample.

## 2.9. Untargeted UPLC-MS data analysis.

Pre-processing of the untargeted UPLC-MS analysis data was done using the XCMS software, an R-based package (Tautenhahn et al., 2008; Benton et al., 2010; Smith et al., 2006), after having imported and converted the raw files into the .mZML format using the MSConvert utility of the Proteowizard package 3.0 (Kessner et al., 2008). Peak picking was performed based on the "centWave" method followed by peak grouping using the "nearest" method. The datasets were normalised according to the median fold change method (Veselkov et al., 2011) and a filter was applied to remove any features presenting a coefficient of variation (CV = SD / mean) higher than 20% of the QCs. The output file was imported into SIMCA 14.1 (Umetrics, Umea, Sweden) and both urine and faecal water datasets were mean-centred and $\log_{10}$-transformed with an offset of 20, typically used for aqueous extracts (Rocke and Durbin, 2003). PCA models were built to identify outliers and observe general trends in the data. The datasets were also imported into Matlab R2018a and covariate-adjusted projection to latent structures regression (CA-PLS-R, Posma et al., 2018) models were constructed adjusting for the animals, the cohort and the replicate cofounders. The goodness and the predictable power of each PLS-R models was assessed using the $R^2Y$ and $Q^2Y$ values (respectively), with cut-offs set at 0.25 and 0.1, respectively. Features having a Q-value (*p* value corrected by a multiple testing based on the Benjamini-Hochberg method (Benjamini andhochberg, 1995) with a false discovery rate of 5%) after a Monte-Carlo Cross-Validation (MCCV, using 100 rounds and a partitioning of 3) below 0.05 were selected for metabolite identification.

## 2.10. Metabolite identification.

Features selected as significant according the CA-PLS-R models were putatively annotated using several different approaches. Initially, the *m/z* and the retention times were searched in the "in-house" National Phenome Centre Database (NPC, see acknowledgments). In addition, features were matched with online databases using CEU Mass Mediator (Gil de la Fuente et al., 2018), by comparing the $[M+H]^+$, $[M+H-H_2O]^+$, $[M+Na]^+$ or $[M+K]^+$ adducts, as well as the main fragments (up to four, Tables S1 and S2). Data-dependent acquisition (DDA) and UPLC-MS/MS experiments (using a Xevo G2 Q-ToF [Waters Ltd., Elstree, UK] and the same chromatographic conditions as the original profiling) were performed on the features putatively annotated to support putative identifications. The capillary voltage was 3 kV and the source temperature were set at 120 °C. The cone gas flow was 150 L/h and the gas used was nitrogen. The desolvation gas temperature was 400 °C, the desolvation gas flow was 1000 L/h. Several injections were run with various collision energies, one at 10 eV, another at 30 eV and finally a ramp collision energy starting from 10 eV and going to 40 eV. Leucine encephalin was used for mass accuracy (200 pg/µL in ACN:$H_2O$ 1:1, with scan collection every 60 sec) . The -$\log_{10}$(q-

value) multiplied by the sign of the Manhattan score of the CA-PLS-R models were used to build aheatmap of the metabolites annotated in this way.

### 2.11. Correlogram.

To assess correlation between the faecal excretion of DAMPA, as well as the urinary and faecal excretion of MTX and the microbial taxonomic profile, a correlation matrix was generated. Bacterial family being present in more than one pre-dose faecal sample from three animals treated with 100 mg/kg MTX were selected and correlated with the quantity of MTX excreted in urine samples at 12 h post-dose and with the quantity of MTX and DAMPA excreted in faecal samples at 48 h post-dose (according to the peak excretion of these compounds in each matrix). The correlation method used was Spearman, and the clustering method was based on single linkage clustering. The significance level of the correlation was calculated and added to the correlogram using the *corr.mtest* function in R.

## 3. Results

### 3.1. Quantification of MTX and its metabolites 7-OH-MTX and DAMPA

The concentrations of MTX and its two main metabolites, 7-OH-MTX and DAMPA, were measured in urine and faeces using semi-quantitative targeted UPLC-MS (Figure 3-2). In urine, MTX was the only analyte detected and, of the time points available for analysis (0-6, 6-12, 18-24 and 36-48 h), was predominantly found in the 6-12 h samples. In the faeces, animals receiving 10 mg/kg of MTX excreted the majority of the drug by 12 h post dose and those receiving 40 or 100 mg/kg doses excreted the highest concentrations of the drug in the 36-48 h samples. DAMPA was detected in the faecal extracts over both time-periods (6-12 and 36-48 h) for the animals treated with 10 mg/kg and mainly in the 36-48 h faecal collection for the animals treated with 40 or 100 mg/kg. The faecal concentration of 7-OH-MTX was negligible and was only detected in 6-12 h post-dose samples from animals receiving 10 mg/kg of MTX and 36-48 h collections for the animals treated with 40 mg/kg. The faecal excretion of 7-OH-MTX was particularly low in the animals treated with 100 mg/kg of MTX. Similarly, the faecal concentrations of DAMPA were lower in the animals treated with 100 mg/kg compared to the group treated with 40 mg/kg for the 36-48 h post dose samples.

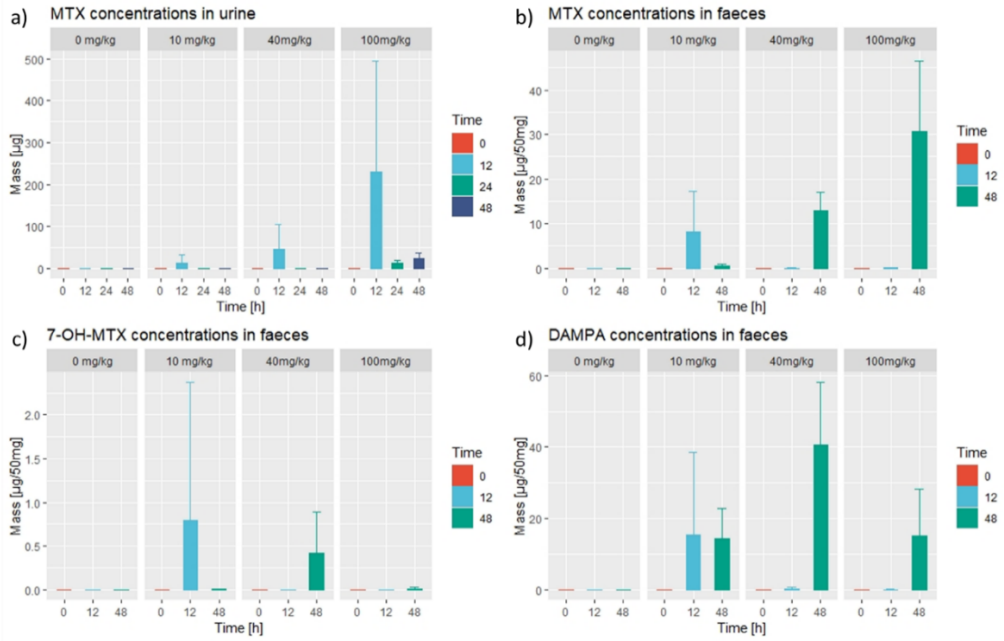



**Figure 3-2. Quantification of MTX, 7-OH-MTX and DAMPA using a semi-targeted UPLC-MS approach.** a) Excretion profile of MTX in urine, reported in total µg. The faecal excretion of b) MTX, c) 7-OH-MTX and d) DAMPA, reported in µg/50 mg of faeces. *N* = 5 for each time point within each dose-related group in urine. For the faecal samples of the untreated rats, *N* = 6 at 0 h, N = 3 at 12 h and *N* = 5 at 48 h; for the rats treated with 10 mg/kg, *N* = 10 at 0 h, *N* = 5 at 12 h and *N* = 4 at 48 h; for the rats treated with 40 mg/kg, *N* = 9 at 0 h, *N* = 3 at 12 h and *N* = 3 at 48 h; and for the rats treated with 100 mg/kg of MTX, *N* = 12 at 0 h, *N* = 3 at 12 h and *N* = 3 at 48 h.

The trace levels at which 7-OH-MTX were measured over the experiment and particularly for those animals treated with 100 mg/kg of MTX, where no metabolite could be detected, could be explained by its possible degradation into 7-OH-DAMPA (Figure 3-2), although no evidence to support this was found in the untargeted MS data. Nevertheless, other metabolites were tentatively annotated in urine, such as 7-OH-MTX-sulfate, 7-OH-DAMPA-glucuronide, DAMPA-sulfate and MTX-glucuronide (Figure 3-2), but because these compounds were found only in trace amounts in the initial $MS^E$ experiment and because MS/MS experiments did not reveal the expected fragments, these annotations were excluded.

## 3.2. MTX modulated the urinary metabolic profile

Principal components analysis (PCA) models were built using the urinary metabolic profiles from each MTX dose to identify sources of variation (Figure 3-3). A modest time-related effect was observed in the control group, particularly at 48 h into the study.

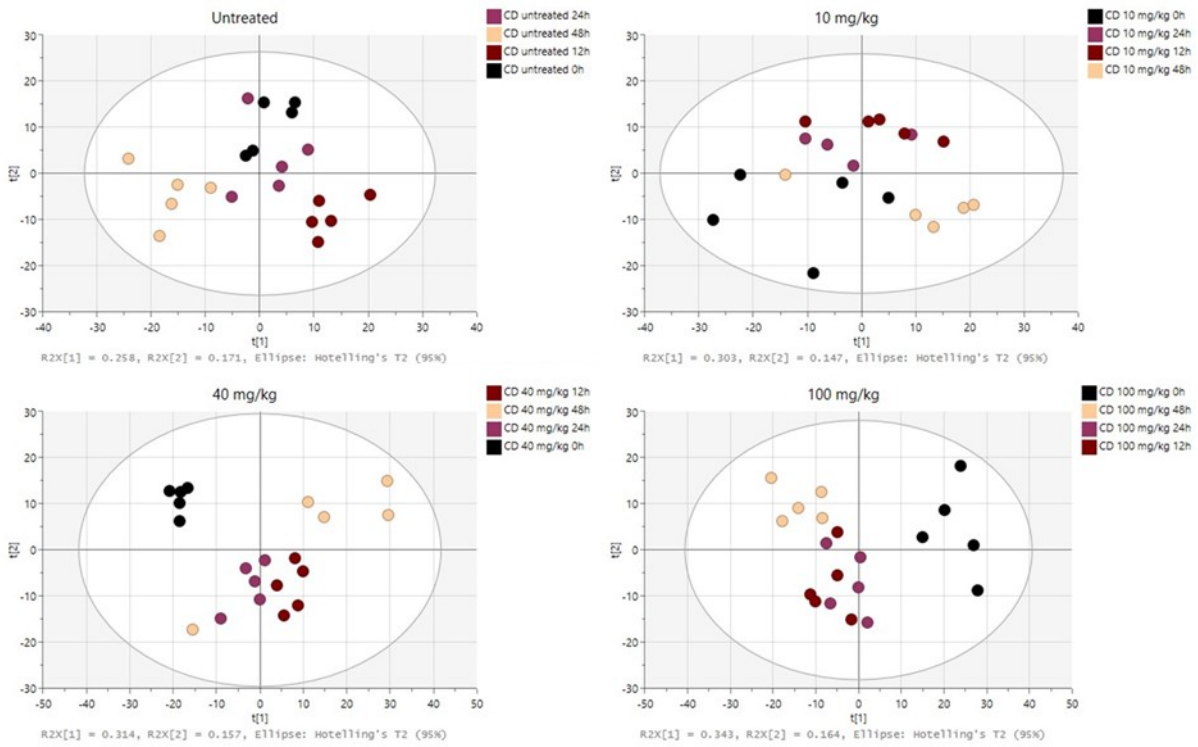



**Figure 3-3. PCA scores plots of the urine metabolic profile, plotted according the dose of MTX administrated.** One PCA scores plot per MTX-dose is displayed with different colours indicating the four time points 0, 12, 24 and 48 h (black, dark red, plum and tan, respectively). *N* =5 per time point per dose-related group. The scores t[1] and t[2] represent the distance between a sample to the origin of $PC_1$ and $PC_2$, respectively.

Supervised multivariate statistical analysis using covariate-adjusted projection to latent structures regression (CA-PLS-R, Posma et al., 2018) analysis, was performed to illuminate the metabolic response following each MTX dose (Figure 3-S2). The skyline significance representations (Figure 3-4a) show that as the dose of MTX increased, the number of urinary features significantly varying with time increased. As seen for the PCA of the untreated (control) animals (Figure 3-3), time-dependent metabolic shifts in the urinary metabolic profiles of the untreated animals were observed by CA-PLS-R with changes in the signal intensity of features over time (Figure 3-4a). The features related to time and found to significantly change in the control group were subtracted from the animal groups treated with MTX, to identify only the features significantly affected by MTX treatment.

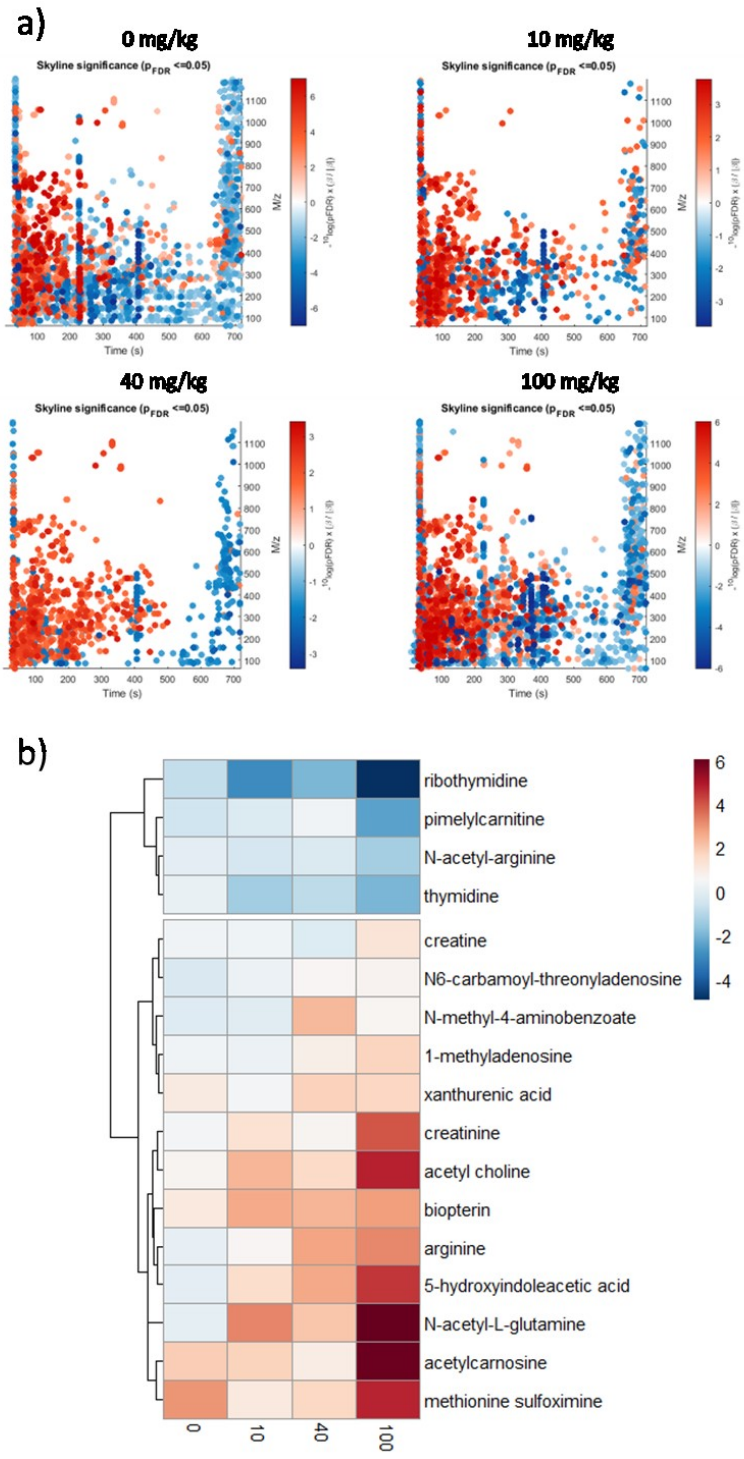



**Figure 3-4. CA-PLS-DA models performed on the metabolic profile of the urine samples.** a) Skyline significance of the CA-PLS-Regression models, performed for each of the MTX dose groups (0, 10, 40 and 100 mg/kg), representing the number of features having a q-value less than 0.05 and which significantly increased (red colour) or decreased (blue colour) over time (0, 12, 24 and 48h). b) Heatmap of the metabolites putatively annotated and significantly influenced by MTX treatment. Colour represents the $-\log_{10}$(q-value) multiplied by the Manhattan sign of the CA-PLS-R models. Red represents an increase of the feature over time and blue indicates a decrease overtime. *N* = 5 per time point per dose-related group.

A large number of features were significantly altered following MTX treatment (1,651 for the urine dataset and 644 for the faecal dataset). Features of interest were matched to an in-house database (based on retention time and *m/z*) and to online databases (based on *m/z*) (Table S2). For a metabolite annotation to be reported it had to meet the criteria of having two *m/z* values match with the values from the database by $MS^E$ and be confirmed by MS/MS as well as biological plausibility. The putative annotations are reported in Table S2 and Table S3, for urine and faecal samples respectively. A heatmap based on the $-\log_{10}$(q-values) of these annotations was constructed from these data (Figure 3-4b). From the clustering analysis of the metabolites two main clusters emerge. One cluster included the metabolites that increased over time following MTX treatment and the other clustered those that decreased. Metabolites that decreased on dosing included ribothymidine, pimelylcarnitine, *N*-acetyl-arginine and thymidine. Among the metabolites that increased, two clusters can be distinguished, one includes metabolites that showed a modest increase across all MTX dosed animals (creatine, N6-carbomoyl-threonyladenosine, *N*-methyl-4-aminobenzoate, 1-methyladenosine and xanthurenic acid), and the other included metabolites whose excretion increased in a dose-dependent manner. These metabolites included acetylcarnosine, which closely clustered with methionine sulfoximine, and *N*-acetyl-L-glutamine, biopterin, arginine, 5-hydroxyindoleacetic acid, creatinine and acetylcholine.

## 3.3. MTX effects on the faecal metabolic profile

From the PCA models built on the faecal metabolic profiles (Figure 3-5), time-related variation was also observed. This occurred in the untreated (0 mg/kg) profiles along the principal component one. A large amount of variation was seen in the 12 h samples and the 48h samples from animals treated with MTX. This effect was dose-dependent being more pronounced at the higher dose.

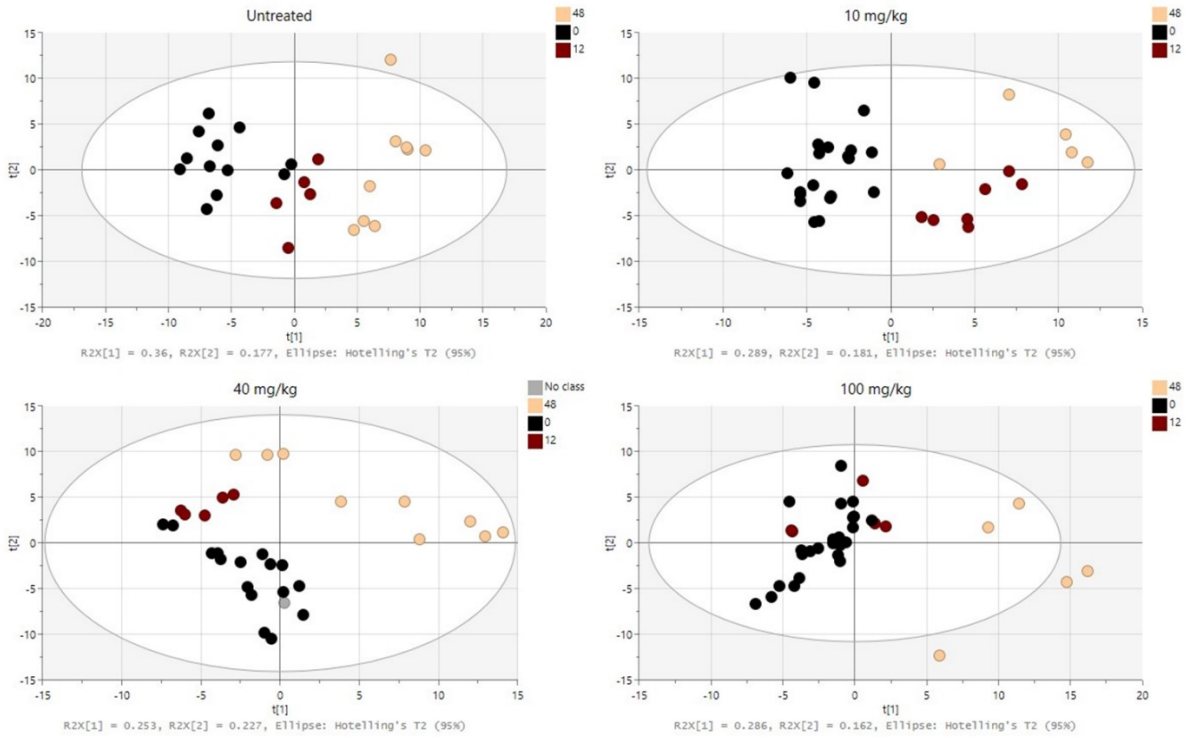



**Figure 3-5. PCA scores plots of the faecal metabolic profiles plotted according the dose of MTX administered.** One PCA scores plot per MTX-dose is displayed with different colours indicating the four time points 0, 12, 24 and 48 h (black, dark red, plum and tan, respectively). For the rats untreated, *N* = 6 at 0h, *N* = 3 at 12h and *N* = 5 at 48h; for the rats treated with 10 mg/kg, *N* = 10 at 0h, *N* = 5 at 12h and *N* = 4 at 48h; for the rats treated with 40 mg/kg, *N* = 9 at 0h, *N* = 3 at 12h and *N* = 3 at 48h; and for the rats treated with 100 mg/kg of MTX, *N* = 12 at 0h, *N* = 3 at 12h and *N* = 3 at 48h. When available, biological replicates were analysed. The scores t[1] and t[2] represent the distance between a sample to the origin of $PC_1$ and $PC_2$, respectively.

CA-PLS-R models were also built to identify metabolic variation occurring over time following the MTX dose (Figure 3-S4). A number of metabolic features significantly changed (after correction for multiple testing using the false discovery rate) over time for each of the MTX dose-related groups (Figure 3-6).

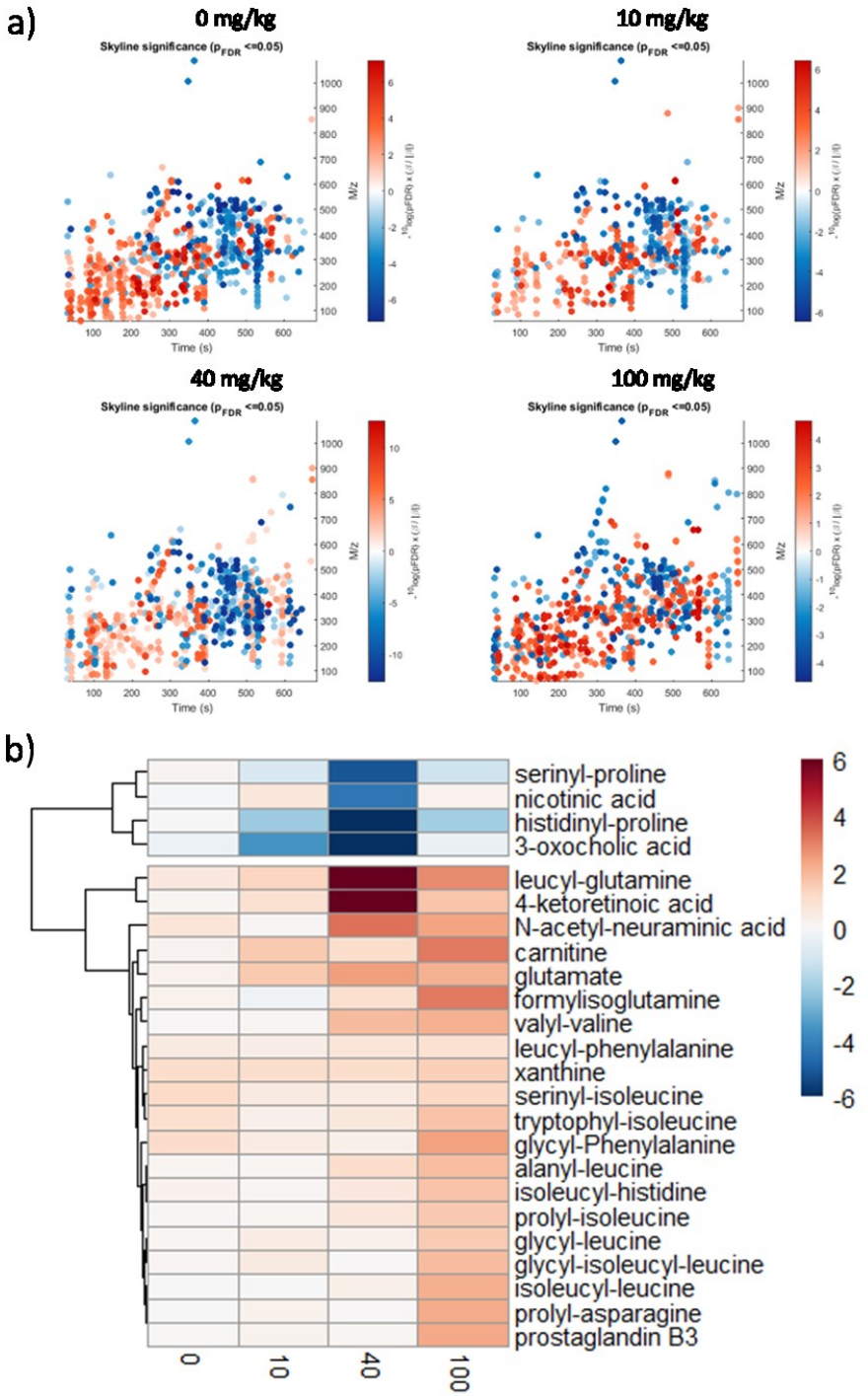



**Figure 3-6. CA-PLS-DA models performed on the metabolic profile of the faecal samples**. a) Skyline significance of the CA-PLS-Regression models, performed for each of the MTX dose groups (0, 10, 40 and 100 mg/kg), representing the number of features having a q-value less than 0.05 and which significantly increased (red colour) or decreased (blue colour) over time (0, 12, 24 and 48 h). b) Heatmap of the metabolites putatively annotated and significantly influenced by MTX treatment. Colour represents the $-\log_{10}$(q-value) multiplied by the Manhattan sign of the CA-PLS-R models. Red represents an increase of the feature over time and blue indicates a decrease overtime. *N* = 6 at 0h, *N* = 3 at 12h and *N* = 5 at 48h; for the rats treated with 10 mg/kg, *N* = 10 at 0h, *N* = 5 at 12h and *N* = 4 at 48h; for the rats treated with 40 mg/kg, *N* = 9 at 0h, *N* = 3 at 12h and *N* = 3 at 48h; and for the rats treated with 100 mg/kg of MTX, *N* = 12 at 0h, *N* = 3 at 12h and *N* = 3 at 48h. When available, biological replicates were analysed.

No major difference was observed between the skyline significance plots of the animals treated with 10 and 40 mg/kg of MTX. However, an increase in biochemical perturbations and their intensity is observed in the skyline significance of the animal group treated with 100 mg/kg of MTX and the lower dosed groups. Indeed, many features seem to show a stronger increase over time following administration of 100 mg/kg of MTX compared to lower doses. The features significantly changing over time in the untreated animal group were subtracted from those of the MTX dose groups to focus the metabolite identification only on the features affected by MTX treatment.

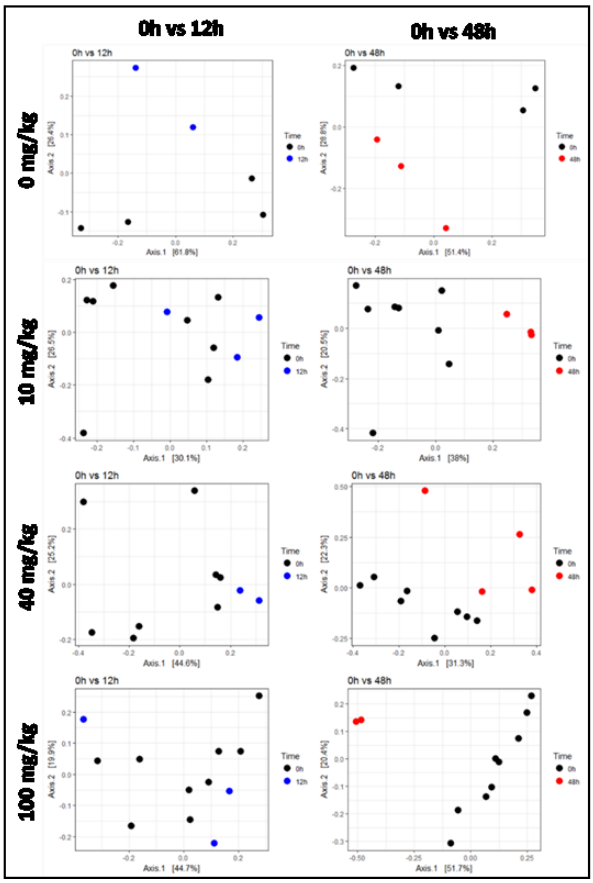
In total, 26 compounds significantly altered by MTX treatment were putatively annotated (Table S3). Of these, fourteen were dipeptides and one a tripeptide. Several of the other metabolites altered were acids, namely nicotinic acid, 3-oxocholic acid, 4-ketoretinoic acid and *N*-acetyl-neuraminic acid (Figure 3-6b). From the cluster analysis of the altered metabolites two main clusters of compounds were observed. One is composed of four compounds which decreased through time in each of the MTX dose groups, and another group composed of the features which increased. Within the cluster of compounds increasing, some increased only in rats treated with 100 mg/kg of MTX (especially dipeptides containing leucine or isoleucine, as well as glycyl-phelylalanine, prolyl-asparagine and prostaglandin B3) whilst others were significantly increased in animals receiving 40 mg/kg. In the case of animals treated with 40 mg/kg of MTX, leucyl-glutamine, 4-ketoretinoic acid, *N*-acetyl-neuraminic acid, octadecatriynoic acid and glutamate increased over time even more than in the case of the animals treated with 100 mg/kg (especially leucyl-glutamine and 4-ketoretinoic acid). A similar observation can be drawn in the cluster of compounds that decreased over time following treatment with MTX, where four of the compounds, serinyl-proline, nicotinic acid, histidinyl-proline and 3-oxocholic acid, decreased even more dramatically in the faecal samples from animals treated with 40 mg/kg.

## 3.4. MTX effect on the faecal microbial profile

Bray-Curtis PCoA was used to assess β-diversity (variation between samples) within the dataset (Figure 3-7a). The PCoA plots comparing the faecal samples of the 12 h post-dose time point with the pre-dose time point indicate no or low dissimilarity levels between the two time points for each of the MTX dose-related animal groups. Greater variance was seen 48 h post-dose compared to the pre-dose time point across all groups however, this was also observed in the control group. This dissimilarity was particularly marked for the faecal samples of the animals treated with 100 mg/kg. α diversity (diversity within samples) was also assessed using the Inverse Simpson index. The averaged α diversity is reported in Figure 3-7b. The α diversity of the untreated animals (0 mg/kg) was stable over time.

For the animals treated with 10 mg/kg, the α diversity of their faecal samples increased 12 h post MTX-dose before decreasing by 48h post MTX-dose. The α diversity of the faecal samples of the animals treated with 40 mg/kg MTX decreased at 12 h post dose and stabilised at 48 h post-dose, which was not the case for the animals treated with 100 mg/kg of MTX, where α diversity progressively decreased for 48h following MTX administration.

a)




b)

| | | Time point | | |
|---|---|---|---|---|
| | MTX dose (mg/kg) | 0h | 12h | 48h |
| Alpha diversity values | 0 | 14.94 | 14.47 | 15.39 |
| | 10 | 15.63 | 19.36 | 10.56 |
| | 40 | 12.30 | 9.66 | 11.18 |
| | 100 | 11.08 | 7.53 | 4.94 |
| Number of samples | 0 | 4 | 2 | 2 |
| | 10 | 5 | 3 | 2 |
| | 40 | 7 | 1 | 3 |
| | 100 | 6 | 2 | 2 |

**Figure 3-7. Investigation of the β and α diversities of the faecal samples.** a) B diversity represented using PCoA of Bray-Curtis distance, displaying the similarity between the samples. For the untreated animals, *N* = 4 at 0 h, *N* = 2 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 10 mg/kg, *N* = 8 at 0 h, *N* = 3 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 40 mg/kg, *N* = 8 at 0 h, *N* = 2 at 12 h and *N* = 4 at 48 h; for the animals which received a dose of 100 mg/kg, *N* = 9 at 0 h, *N* = 3 at 12 h and *N* = 2 at 48 h. b) A diversity values representing the microbial richness of each of the faecal samples of the animals according the MTX dose groups over time. The richness is assessed using the Inverse Simpson index. Low numbers of samples due to diarrhoea and other limitations due to study design do not allow statistical tests between the α diversity of the different groups to be performed.

As both β and α diversity display a stronger shift at the 48h post MTX-dose (Figure 3-7), the taxonomic profile of each of the MTX dose-related groups were aggregated and plotted for the pre- and 48h post-dose periods only (Figure 3-8). The faecal microbiota of the animal group dosed with 100 mg/kg of MTX provided the most notable microbial shift with a striking increase in the relative abundance of *Peptostreptococcaceae* and *Porphyromonadaceae* and a decrease in the relative abundance of *Ruminococcaceae.* This alteration appears to have been dose dependent as it was not seen in the animals treated with 0 or 10 mg/kg of MTX but was visible in animals receiving 40 mg/kg of the drug. A microbial shift was also observed over time for the vehicle-only treated group (0 mg/kg), with an increase in the relative abundance of unclassified Firmicutes and a decrease in the relative abundance of *Erysipelotrichaceae*. This change was also noted in animals that received 10 and 40 mg/kg of MTX, but not those treated with 100 mg/kg of MTX, where the reverse occurred. Also, a decrease in the relative abundance of *Bacteroidaceae*, unclassified *Bacteroidales, Bacteroidetes* and *Mollicutes* was observed for the MTX-treated rats, particularly the 100 mg/kg-dosed animals, compared to the control group (0 mg/kg).

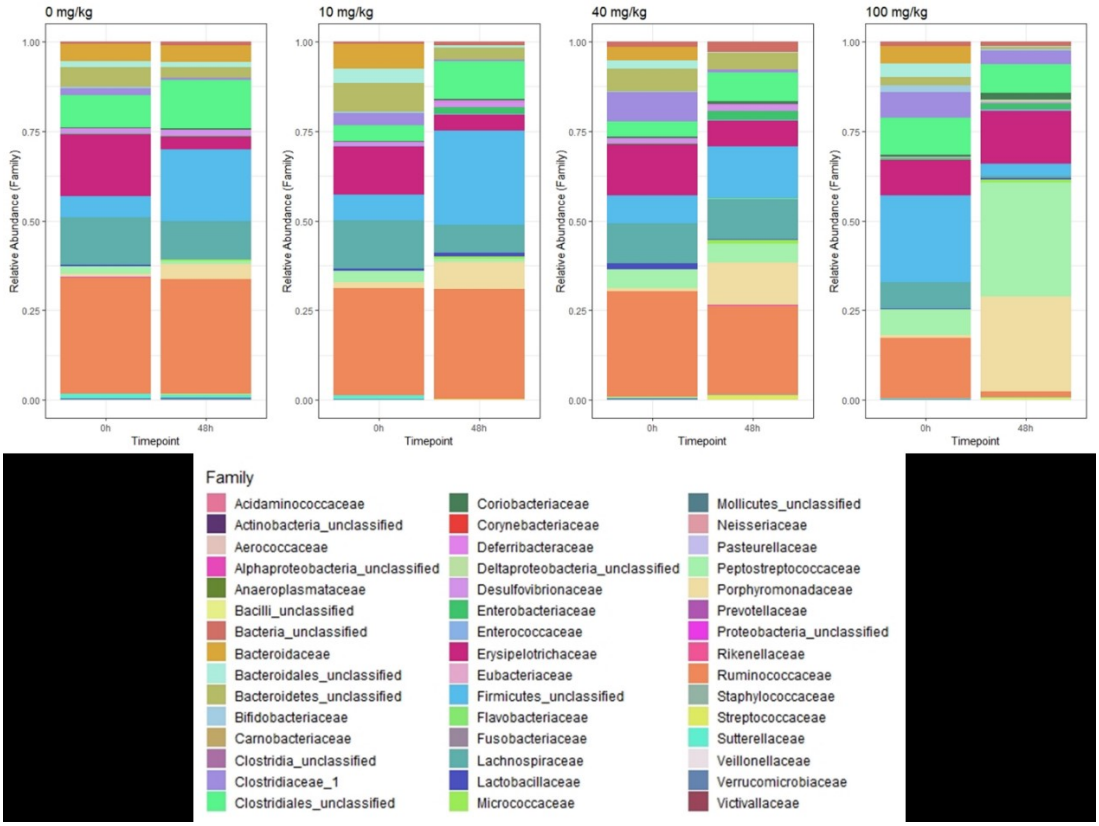



**Figure 3-8. Taxonomic profiles of the faecal samples of the rodents representing the relative abundance of the Family in each of the MTX dose-related group pre- and post- MTX treatment (0 and 48 h respectively).** For the untreated animals, *N* = 4 at 0 h, *N* = 2 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 10 mg/kg, *N* = 8 at 0 h, *N* = 3 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 40 mg/kg, *N* = 8 at 0 h, *N* = 2 at 12 h and *N* = 4 at 48 h; for the animals which received a dose of 100 mg/kg, *N* = 9 at 0 h, *N* = 3 at 12 h and *N* = 2 at 48 h.

Taxonomic profiles of the animals over time following MTX were also assessed at phyla level (Figure 3-9). A different trend was observed in response to the lowest and highest doses of MTX in changes at the phylum level. At 10 mg/kg, a decrease in the relative abundance of Bacteroidetes and an increase in the relative abundance of Firmicutes was observed, while at 40 mg/kg, the trend was reversed with an increase in the relative abundance of Bacteroidetes and a decrease in the relative abundance of Firmicutes, which was exaggerated in the animals treated with 100 mg/kg of MTX.

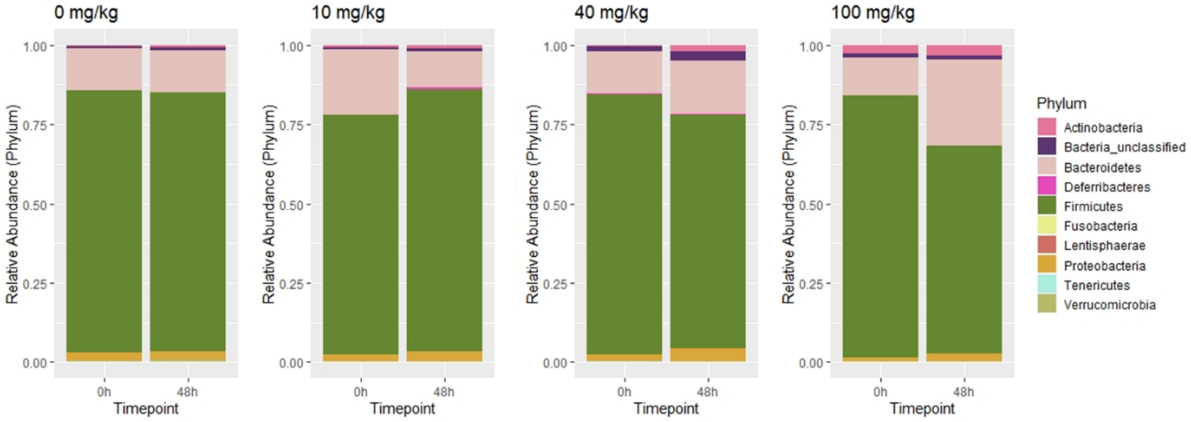



**Figure 3-9. Taxonomic profiles of the faecal samples of the rodents representing the relative abundance of the phyla in each of the MTX dose-related group pre- and post- MTX treatment (0 and 48 h respectively).** For the untreated animals, *N* = 4 at 0 h, *N* = 2 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 10 mg/kg, *N* = 8 at 0 h, *N* = 3 at 12 h and *N* = 3 at 48 h; for the animals which received a dose of 40 mg/kg, *N* = 8 at 0 h, *N* = 2 at 12 h and *N* = 4 at 48 h; for the animals which received a dose of 100 mg/kg, *N* = 9 at 0 h, *N* = 3 at 12 h and *N* = 2 at 48 h.

## 3.5. Correlation between DAMPA faecal excretion and intestinal microbial profile

Correlations analysis between the bacterial family present in the faecal samples of these high dose animals prior to MTX treatment, with the quantity of MTX measured in urine at 12 h and in faeces at 36-48 h, as well as the quantity of DAMPA excreted in faeces samples between 36 and 48 h, were undertaken (Figure 3-10). The excretion of DAMPA in faeces was positively correlated with the pre-dose relative abundance of *Prevotellaceae*, *Anaeroplasmataceae*, *Ruminococcaceae* and *Lactobacillaceae,* and a negative correlation with the pre-dose relative abundance of *Deferribacteraceae* and *Coriobacteriaceae*. Other significant bacterial-metabolite linkages included a negative correlation between *Deferribacteraceae* and the bacterial families *Veillonellaceae*, *Prevotellaceae*, *Pasteurellaceae*, *Acidaminococcaceae* and *Anaeroplasmataceae* (five families showing positive correlations with one another) *and Ruminococcaceae* (Figure 3-10). Similarly, *Coriobacteriaceae* was negatively correlated to *Prevotellaceae*, *Anaeroplasmataceae* and *Ruminococcaceae* (Figure 3-10).

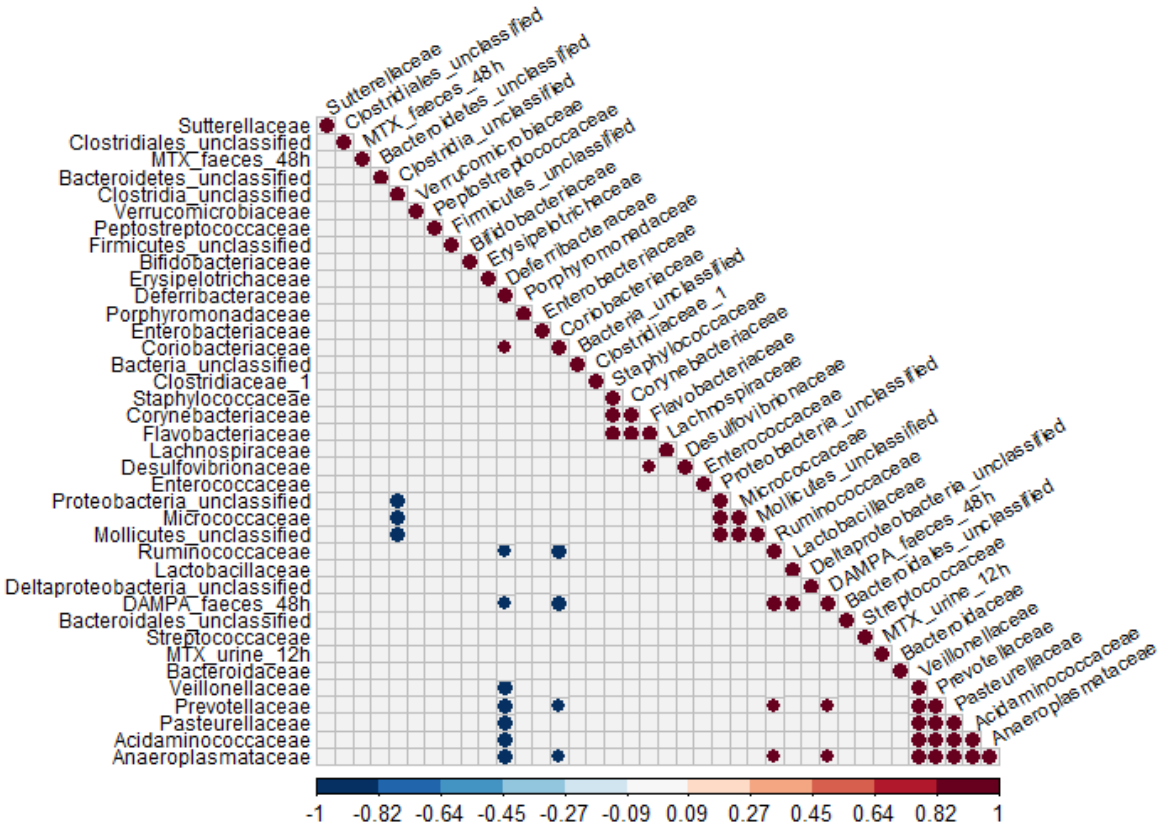



**Figure 3-10. Correlogram integrating the genera present pre-MTX-dose in the faecal samples with MTX concentration measured in urine at 12h and in faeces at 48h, as well as the quantity of DAMPA excreted in faeces samples at 48 h.** Only significant correlations are represented ($p < 0.05$). Represents the significant correlations only ($p < 0.05$) showing positive correlations for DAMPA excreted in faeces with *Prevotellaceae*, *Anaeroplasmataceae*, *Ruminococcaceae* and *Lactobacillaceae* and a negative correlation with *Deferribacteraceae* and *Coriobacteriaceae*. *N* = 3 animals.

Correlation analyses between the bacterial families present in the fecal samples of these high dose animals prior to MTX treatment, with the relative intensities of the putatively annotated metabolites measured in urine and in feces at 36-48 h were also performed (Figure 3-11). The relative abundances of *Prevotellaceae* and *Anaeroplasmataceae*, previously found to be positively correlated with the excretion of DAMPA in feces at 48h (Figure 3-10), were found positively correlated with glutamate excreted in feces and 5-hydroxyindole acetic acid in urine at 48h (Figure 3-11). Similarly, the relative abundances of *Defferibacteraceae* and *Coriobacteriaceae*, previously found negatively correlated with the excretion of DAMPA in feces at 48h (Figure 3-10), were found positively correlated with glutamate excreted in feces and 5-hydroxyindole acetic acid in urine at 48h (Figure 3-11). The relative abundances of *Prevotellaceae* and *Anaeroplasmataceae* were negatively correlated with the relative abundances of *Defferibacteraceae* and *Coriobacteriaceae* in both analyses (Figures 3-10 and 3-11). Finally, the relative abundances of *Ruminococcaceae* and *Lactibacillaceae*, previously found positively correlated with the excretion of DAMPA in feces at 48h (Figure 3-10), were also positively correlated

with N-methyl-4-aminobenzoate excreted in urine at 48h. In addition, *Ruminococcaceae* was also positively correlated with methionine sulfoximine excreted in urine at 48h (Figure 3-11).

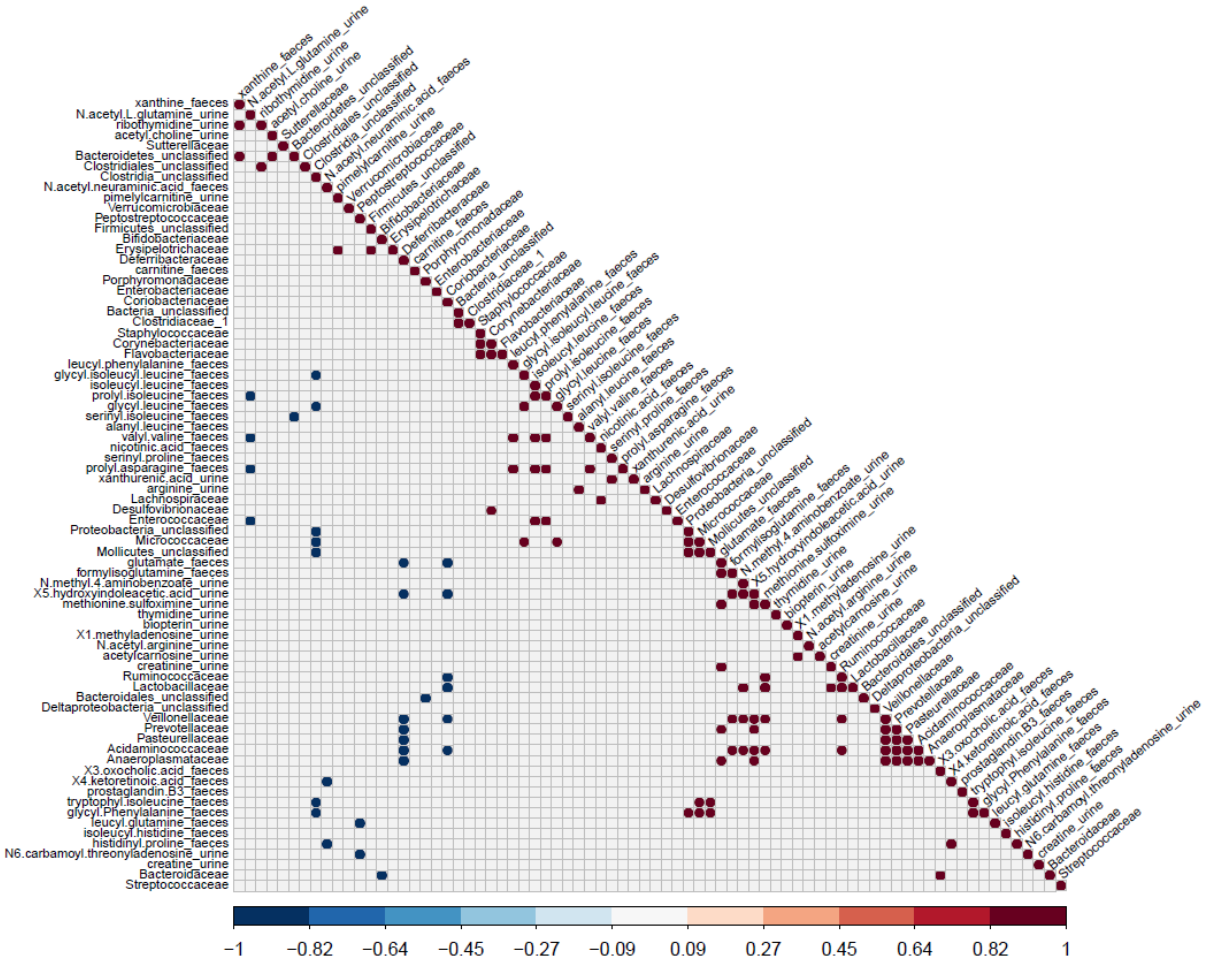



**Figure 3-11. Correlogram integrating the bacteria present pre-MTX-dose in the faecal sampls at family level with putatively annotated compounds excreted in urine and faeces samples at 36-48 h. Represents the significant correlations only ($p < 0.05$) showing positive correlations between the relative abundance of *Prevotellaceae*, *Anaeroplasmataceae* and glutamate in faeces and 5-hydroindoleacetic acid in urine at 48h, and negative correlation of the same metabolites with the relative abundance of *Defferibacteraceae* and *Coriobacteriaceae*. *N* = 3 animals.**

Correlation analyses between the quantity of MTX measured in urine at 12 h and in feces at 36-48 h, as well as the quantity of DAMPA excreted in feces samples between 36 and 48 h, with the relative intensities of the putatively annotated endogenous metabolites present in the urine and fecal samples of these high dose animals prior to MTX treatment, were also made (Figure 3-12). The excretion of MTX in feces at 48h was found to negatively correlate with prolyl-asparagine in feces and 5-hydroxyindoleacetic acid in the urine samples of the predose animals. At 48h, 5-hydroxyindoleacetic acid was found to be positively correlated with the relative abundances of *Prevotellaceae* and *Anaeroplasmataceae* pre-treatment and negatively correlated with the relative abundances of *Deferribacteraceae* and *Coriobacteriaceae* pre-treatment (Figure 3-11), themselves showing the same correlative patterns with the excretion of DAMPA in feces at 48h (Figure 3-10). The excretion of MTX in urine at 12h was found to be positively correlated with the excretion of 4-ketoretinoic acid and prolyl-asparagine in the feces samples of the predose animals (Figure 3-12).

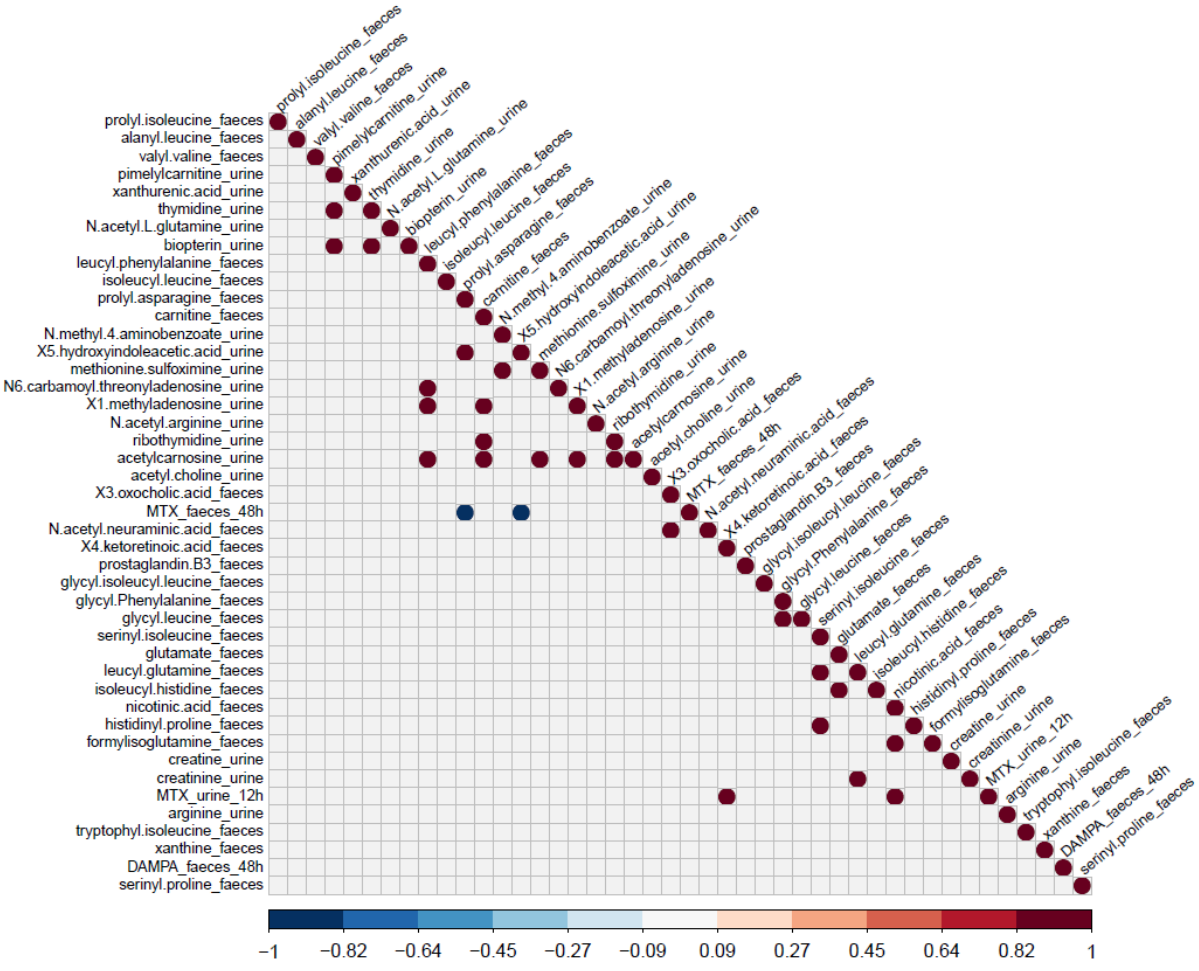



**Figure 3-12. Correlogram integrating the putatively annotated endogenous compounds excreted in urine and feces samples of the animals pre-MTX treatment with MTX concentration measured in urine at 6-12 h and in feces at 36-48 h, as well as the quantity of DAMPA excreted in feces samples at 36-48 h. Represents the significant correlations only ($p < 0.05$) showing negative correlations for 5-hydroxyindole acetic acid in urine samples of the animals pre-MTX treatment with MTX excretion in feces at 48h. *N* = 3 animals.**

# 4. Discussion

A combination of targeted and untargeted UPLC-MS based metabolic profiling approaches were used to assess the biotransformation of MTX and determine its wider biochemical impact on the urinary and faecal metabolome of rats. In parallel, the faecal microbial profiles were defined to characterise the bidirectional interplay between MTX and the microbiota.

The untargeted UPLC-MS results highlighted a number of biochemical pathways affected by MTX, particularly at the 100 mg/kg dose. For example, it is known that MTX impacts DNA synthesis by inhibiting dihydrofolate reductase (DHFR) and by consequence the production of tetrahydrofolate, which is required for purine and pyrimidine synthesis, as well as transmethylation of phospholipids and proteins (Beh et al., 2017). MTX also inhibits thymidine synthesis in host tissues when the plasma concentration of MTX exceeds 10 nM (Chabner and Young, 1973). This is in agreement with the reduced excretion of the pyrimidine nucleosides, namely ribothymidine and thymidine, after MTX administration. In contrast, 100 mg/kg MTX increased the excretion of purine nucleoside-related compounds, such as N6-carbamoyl-threonyladenosine and 1-methyl-adenosine, which is likely to be a consequence of the increase adenosine release previously reported as anti-inflammatory property of MTX (Cronstein et al., 1993). Consistently, xanthine, a purine base produced following the degradation of adenosine monophosphate to uric acid, increased in faeces following MTX exposure.

The urinary excretion of creatinine and its precursor creatine increased following MTX treatment (Figure 3-4b). Interestingly, arginine, a precursor of creatine and a related metabolite *N*-acetylglutamine, were also reduced in the urine following MTX intake. In contrast, urinary *N*-acetylarginine was reduced, possibly due to an increased degradation to produce arginine, or reduced consumption of arginine. When the metabolic profile of the biofluids from this study was studied by $^{1}$H NMR spectroscopy, creatine was also found to be increased in the urine samples (Kyriakides et al., 2014). Other urinary metabolites to be changed using $^{1}$H NMR spectroscopic analysis following MTX treatment were methylamine, dimethylamine, TMAO, formiminoglutamic acid, alanine, phenylacetylglycine, succinate, citrate, 3-indoxyl sulfate, hippurate and formate. MTX-induced elevations in the urinary excretion of creatine could also reflect disruptions to energy metabolism, as creatine is required for the production of phosphocreatine and thus ATP. The potential for MTX treatment to impact on energy metabolism has been previously reported (Kyriakides et al., 2014) and hypothesised to be related to reduced nutrient intake explained by MTX-induced gut epithelial loss and inflammation, morphological damage and malabsorption, previously reported in rats (Naruhashi et al., 2000; Hardwick et al., 2014). Naruhashi *et al.* (Naruhashi et al., 2000) previously reported a disruption to intestinal peptide transporters such as the PEPT1 following MTX exposure, which may

explain alterations in the faecal excretion of dipeptides following the 40 and 100 mg/kg MTX doses, reflecting an impairment in protein digestion or absorption. Similarly, 5-hydroxyindoleacetic acid, a gastroenteritis marker (Ilkhanizadeh et al., 2001), was increased in urine following MTX treatment along with, in the faecal samples, 4-ketoretinoic acid, involved in the maintenance of the epithelial tissue (Heise et al., 2006) and *N*-acetyl-neuraminic acid, found in the gut mucus membrane (among others), which were also found to be increased following MTX treatment. Other metabolites related to gut function were also perturbed by MTX intake. Faecal carnitine and urinary acetylcholine, both implicated into bile acid secretion, were found to increase, while the bile acid 3-oxocholic acid was found to decrease in faeces.

Faecal glutamate excretion increased over time following MTX treatment, which may have arisen from the degradation of MTX into DAMPA by the bacterial enzyme CPDG2 releasing glutamate. The semi-quantification of MTX, 7-OH-MTX and DAMPA in faecal water samples identified two distinct effects occurring depending of the dose of MTX administered to the animals. At the lowest MTX dose (10 mg/kg) both the drug and its metabolite 7-OH-MTX were excreted in the faeces at 12 h while the microbial metabolite DAMPA was excreted from 12 to 48 h. At the higher dose levels (40 or 100 mg/kg) MTX and DAMPA excretion was delayed until 48 h. Surprisingly, DAMPA excretion following the 100 mg/kg MTX dose was lower than that measured during the same time period for the animals administered with 40 mg/kg MTX. The MTX excretion profile was found to be in agreement with previous report of the same study using an alternative MS platform ( Hardwick et al., 2014). Saturation of CPDG2, the enzyme that cleaves glutamate from MTX to produce DAMPA may explain this observation. Alternatively, a shift in the microbial community structure following the highest dose of MTX may reduce CPDG2 activity. Indeed, the faecal microbial profile was modulated by MTX with notable changes occurring 48 h post-dose being most pronounced at the 100 mg/kg dose. Interestingly, at the phylum level, the microbial shift induced by 10 mg/kg MTX included a decrease in Bacteroidetes and an increase in the Firmicutes. This was consistent with previous findings (Nayak et al., 2019). Such changes were then reversed at the 100 mg/kg dose level. This could explained than the antimicrobial effect of MTX, only recently reported (Nayak et al., 2019) and confirmed herein, is specific to the dose of MTX administrated.

The gene encoding for the bacterial enzyme CPDG2, extracted from the intestinal brush border of pigs and humans, was already studied for its H475Y polymorphism (Devlin et al., 2000). For the folates, the consequence of this particular type of polymorphism is a reduction of their absorption through the intestinal epithelium. This could potentially be applied to MTX absorption by the same process, which will increase MTX exposure in the intestinal tract and thus MTX's intestinal toxicity (histological changes, diarrhoea or weight loss). Alternatively, the impact of MTX on the community structure and functionality of the intestinal microbiota may explain the intestinal toxicity associated with this xenobiotic. However even if the gene encoding for CPDG2 has already been extracted from human jejunal brush border (Devlin et al., 2000), the bacterial strains producing this enzyme is, as far as we are aware, unknown.

Correlative studies show that DAMPA in faeces was positively associated with *Prevotellaceae*, *Anaeroplasmataceae*, *Ruminococcaceae* and *Lactobacillaceae* and was negatively associated with the relative abundance of *Defferibacteraceae* and *Coriobacteriaceae* (Figure 3-10). Interestingly, *Prevotellaceae* and *Anaeroplasmataceae* as well as *Defferibacteraceae* and *Coriobacteriaceae* show the same correlative pattern with the excretion of glutamate in faeces and the excretion of 5-hydroxyindoleacetic acid in urine at 48h (Figure 3-11). This could mean that the bacterial enzyme CPDG2 which cleaves MTX into DAMPA and glutamate might be produced by bacteria from the families *Prevotellacea* and *Anaeroplasmataceae*. However, the latest metabolite, above explained to be a gastroenteritis marker, was found to be negatively correlated with the excretion of MTX in feces at 48h (Figure 3-12). The excretion of the metabolite 4-ketoretinoic acid, above explained to maintain gut epithelial layer, was found to be positively correlated with the excretion of MTX in urine at 12h (Figure 3-12).

*Prevotellaceae* and *Anaeroplasmataceae* were strongly correlated with *Ruminococcaceae* and *Lactobacillaceae*, all of them positively correlating with DAMPA excretion in fecal samples, but also with the relative abundances of *Veillonellaceae*, *Pasteurellaceae* and *Acidaminococcaceae*, which might express a symbiotic effect between those bacterial families. Out of those seven bacterial families, one is part of the phylum Bacteroidetes (*Prevotellaceae*), one is from the Tenericutes (*Anaeroplasmataceae*), one is from the Proteobacteria (*Pasteurellaceae*) and the other four are from the Firmicutes (*Ruminococcaceae*, *Lactobacillaceae*, *Veillonellaceae* and *Acidaminococcaceae*). However, as shown in Figure 3-4, the relative abundance of the phylum Firmicutes decreased following a dose of 100 mg/kg of MTX. This demonstrate that higher abundance of certain bacteria mainly from the Firmicutes phylum pre-MTX treatment are positively associated with higher concentrations of DAMPA in faecal samples. These bacteria could be producing the bacterial enzyme CPDG2, responsible for DAMPA production through MTX metabolism. Although further studies are

required to confirm this, understanding the bacterial strains that can produce CPDG2 and determining the expression/activity of this enzyme in the gut microbiome of a patient prior to MTX administration could enable a patient's risk of MTX toxicity/reduced efficacy to be estimated prior to drug therapy.

# 5. Conclusions

The use of semi-quantitative targeted drug/metabolite profiling combined with metataxonomic analysis further characterised the gut microbial involvement in the hydrolysis of MTX to DAMPA and glutamate, two non-toxic metabolites. The complementary use of untargeted UPLC-MS-based endogenous metabolic and metataxonomic profiling helped to illuminate the impact that, MTX can also have on gut microbial diversity, with corresponding reciprocal effects on its metabolism. Specifically, the excretion of DAMPA in the faeces at 48h was associated with several bacterial families present in the pre-treatment faecal samples. These bacteria, which could produce CPDG2 to cleave the glutamate residue of MTX or other folate analogues, were mainly from the phylum Firmicutes, with their abundance reduced by high doses of MTX, leading to a decrease in clearance of MTX and an increase in gastrointestinal toxicity. This could have relevance for chronically administered xenobiotics that modify the functional capacity of the microbiome over time resulting in consequences for drug metabolism and toxicity. Unfortunately, the origin of CPDG2 production by these bacteria by exploring database was unclear and could not be confirmed. Further *in vitro* culture experiments could prove that these bacteria can metabolise MTX. Application to personalised medicine, by assessing the level of these bacteria in patients prior to prescribing them high doses of MTX, could potentially avoid exposing cancer patients to delayed clearance of MTX and the resultant adverse drug effects such as nephrotoxicity as well as a weakened immune system, thereby improving patient outcomes.

# Chapter 4. Development of an *in vitro* model to investigate direct drug-microbiome interactions

**Aim:**

- Develop a rapid and cost-effective anaerobic temperature-controlled culture system that can be inoculated with human faeces to assess the capacity for an individual's microbiome to metabolize a drug.

**Objectives:**

- Test the model using 96-well plate in which human faecal homogenates from five donors to assess their capacity to metabolize sulfasalazine, paracetamol, methotrexate and paracetamol-glucuronide over time.

- Use a targeted UPLC-MS based approach to measure drug concentrations and their metabolites in the culture

- Optimise the methodology in term of drug solubility and bacterial faecal community.

# 1. Introduction

As presented in the introduction chapter, the intestinal microbiota has been shown to impact the fate of xenobiotics through both direct and indirect mechanisms (Carmody and Turnbaugh, 2014), which can contribute to inter-individual variability in drug efficacy and toxicity. Gut microbes possess enzymes that can directly act on drugs or their metabolites, which can generate active ingredients, inactivate a drug, or produce metabolites that are toxic for the host. Understanding this metabolic potential and how it varies across individuals will assist in advancing drug treatment strategies either to harness or minimise this microbial contribution. In addition, this highlights the necessity for a clinically useful tool that can be used to assess the potential for an individual's microbiome to modify a drug prior to treatment. This information may be relevant for predicting adverse drug reactions or alterations in drug efficacy and can be used to inform therapeutic strategies.

Recently, striking progress has been made in the development of models to explore the impact of the gut microbiota on drug metabolism (Park et al., 2017). Indeed, Zimmermann *et al.* created a computational pharmacokinetic model that could distinguish whether a drug metabolite was of either host or bacterial origin (Zimmermann et al., 2019b). This model takes into consideration absorption, exposure and bioavailability of the drug as well as intestinal transit time and enzyme activity. It was successfully applied to three drugs, namely brivudine, sorivudine and clonazepam. However, prior to use, the model parameters require optimisation through conventional and germ-free mouse experiments (Zimmermann et al., 2019b). The same research team published another study, reporting an *in vitro* system (Zimmermann et al., 2019a) in which 76 bacterial isolates were anaerobically incubated with 271 oral drugs during a 12 h time period. The results found that 176 drugs out of the 271 tested were degraded by at least one bacterial strain and this allowed phyla to be correctly grouped according their drug-metabolizing capacity. This demonstrates that metabolic functions are specific to phylogenetic classification. To obtain these results, the authors developed a technique for exploring the DNA functionality enhancement of specific strains which reveals the bacterial genes responsible for drug metabolism. These genes were clustered with drug metabolites measured by untargeted LC-MS (Zimmermann et al., 2019a). However, the results were inconsistent with the outcomes obtained when the drugs were incubated with human faecal cultures, proving that the metabolic fate of a drug in contact with the microbiome requires consideration of the bacterial species as a community rather than a single entity. Another recent *in vitro* tool has been used to explore the microbiome-drug metabolism interaction, but this time in the other direction, *i.e.* through the investigation of the impact of drugs on the gut microbiota. In this model drugs were incubated anaerobically with a mixture of 40 human bacterial isolates, which cover 60% of the median relative

abundance estimate of bacterial species from human faeces (Maier et al., 2018). However, it was acknowledged that it was difficult to grow some bacterial species and, most importantly, that the bacterial strains selected for the study were a restricted part of the gut microbiota (Maier et al., 2018). Moreover, inter-individual variability in the microbiome cannot be explored with a bacteria community pooled from several donors.

To study the biomolecular interactive mechanisms occurring between bacterial species and intestinal epithelium, a broad range of *in vitro* tools already exist. For instance, *in vitro* 2D static culture models, such as centrifuge tubes containing Caco2 cells, also called "Human oxygen-bacteria anaerobic" or HoxBan system (Figure 4-1; Sadaghian Sadabad et al., 2015). Alternatively, the trans-well model, based on a well-plate containing nanopores, thus allowing communication between the basolateral membrane, located in the bottom of the well-plate, and the apical membrane located on top, is valued for its simplicity and robustness (Figure 4-1; Park et al., 2017). However, *in vitro* 3D culture models provide a better appreciation of the spatial structure of the tissues, the Host-Microbiome Interaction (HMI$^{TM}$) model for instance, which contains bacterial species in a luminal compartment, and enterocytes in a tissue compartment (Marzorati et al., 2014). The advantage of this tool is that it can be connected to vessels which model intestinal organs (*e.g.* small intestine), and that fluid flow through the membrane of the two compartments which mimics the blood circulation (Figure 4-1). However, this model lacks intestinal and immune cells as well as epithelium. Organoids have also been applied to host-microbiome interaction studies following injection of the bacterium *Helicobacter pylori* into a stomach organoid (Figure 4-1; McCracken et al., 2014). As patient cells, or microbiome cells from patients can easily be implanted into the organoids, this model shows potential for personalised medicine investigations (Park et al., 2017). However, both *in vitro* 2D and 3D tissue models, which are by convention static cultures, are limited in time. Indeed, due to the static nature of these models, bacteria growth is difficult to control, which will rapidly induce bacterial overgrowth and will reduce the viability of the models. Micro-physiological systems are one of the most advanced systems to explore the gastrointestinal tract, such as the "Human-Microbiome Crosstalk" (HuMIX) model, which consists of assemblies of micro-engineered chambers allowing the control of various parameters seen in 2D or 3D models but most importantly controlled aerobic/anaerobic conditions (Figure 4-1; Shah et al., 2016). An even more advanced tool, the "gut-on-a-chip", offers the enormous advantage of modelling peristalsis. This system has been successfully used in experiments with various human cells and microbiome cultures, in the presence of probiotics and antibiotics which further impacted the epithelium morphology of the gut (Figure 4-1; Kim et al., 2016). However, such models are time-consuming, expensive, and their use cannot be envisaged in a high-throughput clinical setting.

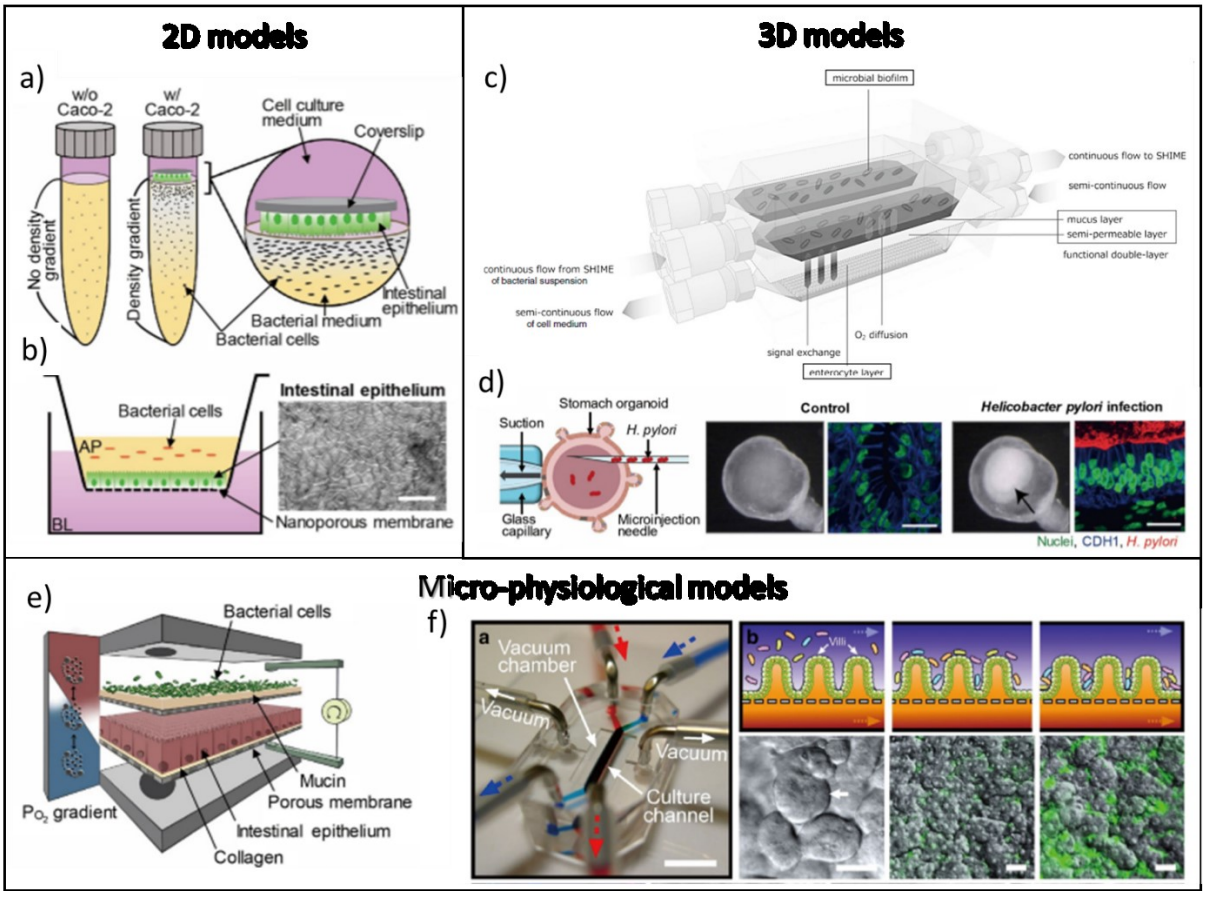



**Figure 4-1.** ***In vitro*** **models allowing the investigation of biomolecular mechanisms between bacterial species and intestines epithelium.** The 2D models are represented by a) the Caco-2 cell cultures in a centrifuge tube which allows the growth of a bacterial species b) the Trans-well culture model, where the use of Caco-2 cells applied on the apical side of a nanoporous membrane allow the growth of an epithelium layer prior to explore its short-term interactions with bacterial species added on the apical side of the model. The 3D models represented are c) the HMI™ model which is composed of two compartments, a lower one containing intestinal cells and an upper one which contains a microbial biofilm, separated by a semi-permeable mucus layer allowing metabolic and microbial exchange and d) the 3D organoid model which consists in the injection of a bacterial pathogen into an gastric organoid, allowing to observed on which side of the epithelium the colonization occurs using immunofluorescence. Micro-physiological model examples are e) the HuMIX model which combine microbial and intestinal epithelium Caco-2 cells separated by a mucus layer and cultured in anaerobic conditions and f) the organ-on-a-chip microdevice which consists in a micro engineered device with culture medium flowing through and possibility to induce peristatic movement allowing the formation of intestinal villi (represented by schematic drawings) with or without the presence of bacterial cells (shown by fluorescence images). Figures from **(Park et al., 2017)**, except for Figure 4-1.c. which were from **(Marzorati et al., 2014)**.

In summary, no method is currently available that allows the rapid investigation of direct interactions between the complex microbiota community and therapeutic drugs or to establish the potential for an individual's microbiota to degrade a specific drug. Given the large amount of inter-individual variability in this capacity, a model would be highly valuable to inform drug treatment strategies. Most of the existing *in vitro* models available to study the impact of the gut microbiota on health are not optimised to study microbiota-drug interactions, or if so, require time-consuming–experiments, complex laboratory setups, or are limited to study the interaction of a single bacterial strain with a specific drug. Therefore, in this chapter a tool was developed to study the potential of an individual's faecal microbial community to metabolise a specific drug. By studying an individual's microbial community, inter-individual variation in the capacity of their microbiome to biotransform drugs can be elucidated.

## 2. General Experimental Section

### 2.1. Chemicals

Acetaminophen (paracetamol, APAP), acetaminophen-(methyl-$d_3$), methotrexate (MTX), *p*-acetamidophenyl $\beta$-D-glucuronide, 1.0 M hydrochloric acid solutions (HCl) and dansyl chloride were purchased from Sigma-Aldrich (Gillingham, UK). Anaerobic sachets, optima LC-MS grade 0.1% formic acid (FA) in water, optima LC-MS Water and analytical reagent grade FA were purchased from Fisher Scientific UK Ltd. (Loughborough, UK). Dimethyl sulfoxide (DMSO) was purchased from VWR Chemicals (Lutterworth, UK). Acetonitrile (ACN), methanol (MeOH) and 2-isopropanol (IPA) were obtained from Honeywell (Seezle, Germany). All chemicals were of analytical grade and solvents were LC-MS grade quality. Ostro plates were purchased from Waters (Wilmslow, UK).

### 2.2. Ethical considerations

This *in vitro* study was approved by the University of Roehampton Research Ethics Committee (ethics ref. LSC 18/236). All of the faecal donors had the experimental procedure explained and were given the opportunity to ask questions. All donors provided verbal consent for the use of their faeces in the study.

## 2.3. Faecal homogenate preparation

Faecal samples were obtained from healthy human volunteers (body mass index (BMI) ~20) who had no known metabolic or gastrointestinal diseases. Faecal samples were collected from donors the same morning as the experiment was carried out and put on dry ice immediately. For each donor, faeces were diluted 1:10 (w/v) in phosphate-buffer saline (PBS) and homogenised for 2 min (400 beats/min) (0.1 g/mL) with 10 mm glass beads added before the solution was vortexed for 5 minutes.

## 2.4. Selection of the drugs and range of concentration to test

As the objective of the study was to develop an *in vitro* model allowing to monitor the degradation of specific drugs according their interactions with the bacteria present in human faeces, different drugs were selected to cover different types of interactions with the gut microbiota. Paracetamol (APAP), was selected as a negative control, as it has not been previously reported to be degraded by bacteria present in the human GI tract, although it has been reported to be *N*-deacetylated to *p*-aminophenol while being incubated with rat cecum content (Smith and Griffiths, 1974). The second xenobiotic selected was sulfasalazine (SSZ). This served as a positive control as it is a pro-drug known to be metabolized by bacterial azo-reductases to release the active ingredient 5-aminosalicylic acid and sulfapyridine (Peppercorn and Goldman, 1972). As explained in the previous chapter, MTX is hydrolysed by the bacterial enzyme CPDG2 to give the inactive metabolite DAMPA. The activity of CPDG2 is assumed to vary in humans, which could explain inter-individual variation between patients in clearing MTX and therefore, differences in MTX-induced toxicities. As such, MTX was selected to validate this hypothesis by assessing variation in its bacterially mediated degradation across individuals.

The concentrations studied during the different experiments were calculated as an approximate concentration present in the gut, assuming that the bulk of the initial oral dose reaches the colon (indicated in Table 1). An example calculation for MTX is provided below:

*e.g.* If MTX is administered at 100 mg/kg x 80 % not absorbed = 80 mg/kg.

Assuming a human weighing around 70 kg so 70 x 80 = 5,600 mg.

Considering that ~ 9 L of liquid passes through the GI tract per day: 5,6 g / 9 L = 0.6 mg/mL.

For the remaining drugs, for which the percentage absorption was not found, it was likely to be overestimated and the value presented in the table represents a maximum.

**Table 4-1. Approximate concentrations of APAP, SSZ and MTX in the gut.**

| *Drugs* | *Molar mass (g/mol)* | *Dose* | *Absorption* | *Bioavailability* | *Time* | *Conc in GI tract (mg/mL)* | *Conc in GI tract (mM)* |
|---|---|---|---|---|---|---|---|
| Methotrexate (MTX) | 454.44 | LDMTX < 15 mg/kg<br>HDMTX >20 mg/kg (up to 1,500 mg/kg with a rescue agent, here 100 mg/kg picked) | 90%<br>20% | 60%<br>< 60 % | 1 to 2h | 0.6 | 1.4 |
| Paracetamol (APAP) | 151.163 | 1 x 4 g /day | + | 70-90 % | 15 to 60 min | 0.13 | 0.88 |
| Sulfasalazine (SSZ) | 398.39 | Acute CD: 3 to 6 g/day | 10% | < 15% | | 0.6 | 1.5 |

# 3. Development of a rapid *in vitro* model

The objective of this experiment was to determine if drug degradation could be monitored when incubating it with faeces from human donors in a miniaturised and rapid *in vitro* model using a 96 well-plate format. The drugs tested for this first development attempt were APAP and SSZ.

## 3.1. Experimental part

### *3.1.1. Medium*

The basal medium used for the preliminary study was composed of peptone water (2 g/L); yeast extract (2 g/L), NaCl (0.1 g/L), $K_2HPO_4.3H_2O$ (0.04 g/L), $KH_2PO_4$ (0.04 g/L), $NaHCO_3$ (2 g/L), $MgSO_4.7H_2O$ (0.01 g/L), $CaCl_2.6H_2O$ (0.01 g/L), Tween 80 (2 mL/L), haemin (50 mg/L), vitamin K1 (10 µL/l), L-cysteine (0.5 g/l), bile salts (0.5 g/L), resazurin (1 mg/l) and distilled water. The medium were autoclaved before to be used in the model.

### *3.1.2. Incubation experiment*

To develop a rapid *in vitro* model a 96 well plate was used. The experimental layout of the plate is presented in Figure 4-2. Solutions of APAP and SSZ, diluted in either warm $H_2O$ or $NH_4OH$ 2.94 M respectively, were prepared at four different concentrations (0, 1, 10, 100 mM) to be further diluted 1 in 10 to obtain a final concentration during the incubation of 0, 0.1, 1, or 10 mM, either in PBS (row A), basal medium (row B), or in a mixture of basal medium and faecal homogenates from five healthy donors (row C to G). Row H contained the drug at the middle concentration (1 mM), mixed with faecal homogenate from all of the donors without any basal medium. Five plates were prepared for each drug and terminated as specific time points (0, 1, 2, 6 and 24 h).

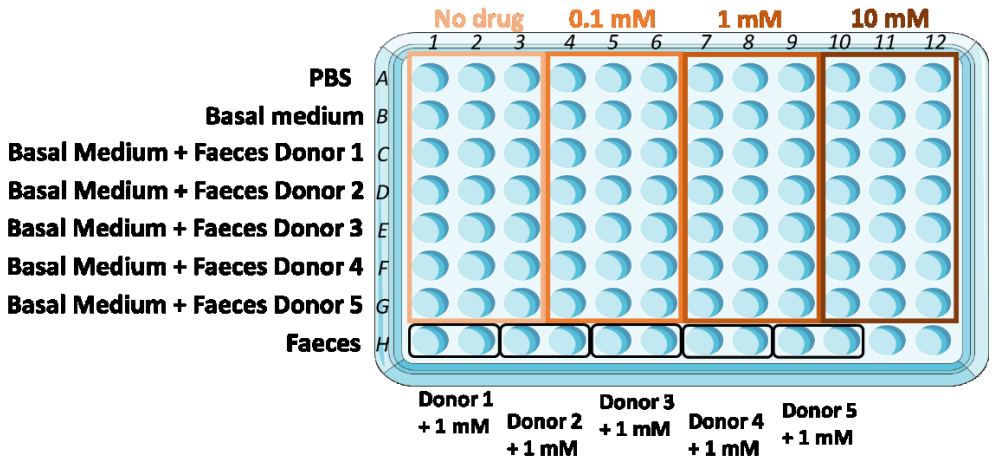



**Figure 4-2. Location of the samples across the 96-well plate used in the development of the *in vitro* model.**

Once the plates were prepared (in a specific, descending order to control the starting time-points of incubation (24, 6, 2, 1 and 0 h)), they were placed in anaerobic jars with anaerobic sachets and fixed on shakers placed in cabinets at 37˚C (Figure 4-3). Once the incubation of each plate was finished, it was taken out of the anaerobic jar, centrifuged at 700 *g* at 4˚C for 20 min, and put directly on ice. The resulting supernatant was collected into a new plate. All plates were then placed in storage at -80˚C.

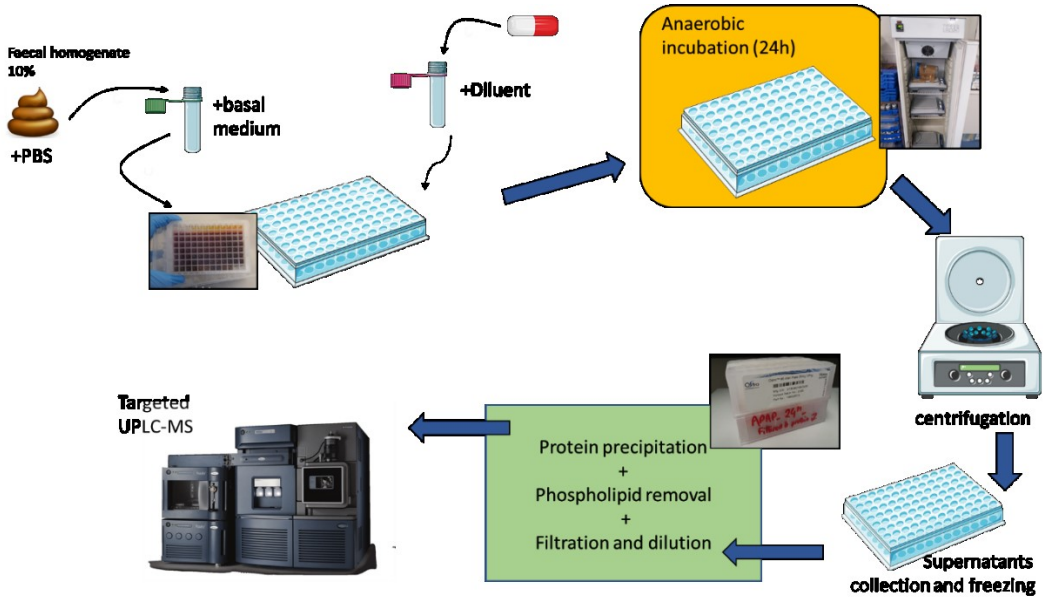



**Figure 4-3. *In vitro* model workflow.** From preparation of the 96-well plates, to the incubation, supernatant collection, preparation and dilution for targeted UPLC-MS.

### *3.1.3. UPLC-MS*

*APAP sample preparation for UPLC-MS analysis.* Plates were defrosted and centrifuged at 700 *g* and 4˚C for 10 minutes. A volume of 150 µL of the supernatant was added to Ostro plates placed on top of a collection plate and mixed with 450 µL ACN. The plate duplex was centrifuged at 700 *g* and 4˚C for 2 minutes to filter and collect the filtrate into the collection plates (concentrations at that stage were 0 M, 25 µM, 250 µM, 2,500 µM). The solution obtained from the collection plates were diluted by mixing with water to give APAP concentrations of 500 nM in all the wells. All UPLC-MS plates were stored at 4˚C until UPLC-MS analysis (Figure 4-3).

*SSZ sample preparation for UPLC-MS analysis.* The same protocol than the one applied for APAP was applied for the plates incubated with SSZ, at the exception that the final concentration used for UPLC-MS analysis was 100 nM in order to meet the linear range of the drug.

*UPLC-MS parameters.* The analysis was performed using an Acquity I-class UPLC system interfaced with a Xevo TQ-S (Waters Corp., Wilmslow, UK). The column used was an Acquity UPLC HSS T3 (2.1 x 50 mm, 1.8 µm). Mobile phase A consisted of 0.1% FA in water and mobile phase B consisted of 0.1% FA in ACN. The gradients used for each of the drugs are reported in Table 4-S1. $H_2O$:ACN (5:95 v:v) was used as a strong wash solvent (1000 µL) and $H_2O$:ACN (95:5 v:v) as a weak wash (1000 µL). The temperature of the column was maintained at 40 ˚C and 5 µL of sample were injected in full loop injection mode. Positive electrospray ionization (ESI) was used to monitor the transitions reported in Table 4-S2. The capillary voltage used was 2 kV, the source temperature was 150 °C and the desolvation temperature 600°C, apart from the detection of ESC, which was performed using a capillary voltage of 1 kV and a desolvation temperature of 500 °C. The cone gas flow was 150 L/h, the collision gas flow 0.15 mL/min, the desolvation gas flow 1000 L/h and the nebuliser gas flow 7 Bar.

*Data acquisition and treatment.* The data were acquired using Masslynx (Waters, MA). TargetLynx (Waters, MA) was used to process the data acquired from samples and quantify the compounds of interest. Within-run QCs were used to assess the accuracy and precision of the run, which were judged acceptable if the CV compared to the nominal concentration at each QC level was below 20 % for the low QC and below 15 % for the mid and high QCs (data not shown). A calibration curve (CC) was run before and after each plate (the concentration of the calibration curves used are reported in Table 4-S3) and the calibration curves were integrated according a linear function and a weighting of 1/X (where X is the concentration).

## 3.2. Results

As shown in Figure 4-3, a first attempt to develop a rapid *in vitro* model was made. To do so, 96 well-plates of 1 mL well volume were used. The basal medium chosen did not contained any glucose, in order to deplete any source of carbon and force any reactions that could potentially happen between the drug of interest and the faecal microbiota. The drug and the basal medium were both directly incubated with the five faecal homogenates, in five different plates that were stopped at different time following the beginning of the incubation (0, 1, 2, 6 or 24 h), providing five different kinetic time points. Once the incubation had been terminated, the plates were centrifuged, and the supernatants collected and prepared for UPLC-MS analysis. The drugs used for this first attempt toward a miniaturised model were once again APAP as a negative control and SSZ as a positive control. The concentrations tested were 0 mM, 0.1 mM, 1 mM and 10 mM.

### *3.2.1. Paracetamol (APAP)*

The initial concentrations of APAP determined by UPLC-MS were different from those expected (Figure 4-4). The initial dose of APAP provided in the basal medium was 100 µM however, the concentration detected was 34.96 +/- 15.49 µM for the 0 h samples. This concentration did increase directly from 1 h (63.85 +/- 29.67 µM) falling to a maximum of 77.80 +/- 16.55 µM at 6 h. At the 1,000 µM dose level, 208.50 +/- 111.54 µM was detected in the 0 h samples, 201.30 +/- 125.84 µM in the 1 h samples and 178.99 +/- 70.94 µM in the 2 h samples before reaching the expected concentration (1,064.84 +/- 138.49 µM) at the 6 h collection time point. Similarly, at 10,000 µM, the concentration of APAP in the media was observed to slowly increase from a value of 1,477.71 +/- 824.05 µM at 0 h to 7,515.56 +/- 1,558.12 µM at the 6 h collection time point. From 6 h, once the maximal concentrations were reached, no major degradation was observed, except when APAP was incubated at 100 µM with the faecal sample of donor 3 (blue line, Figure 4-4).

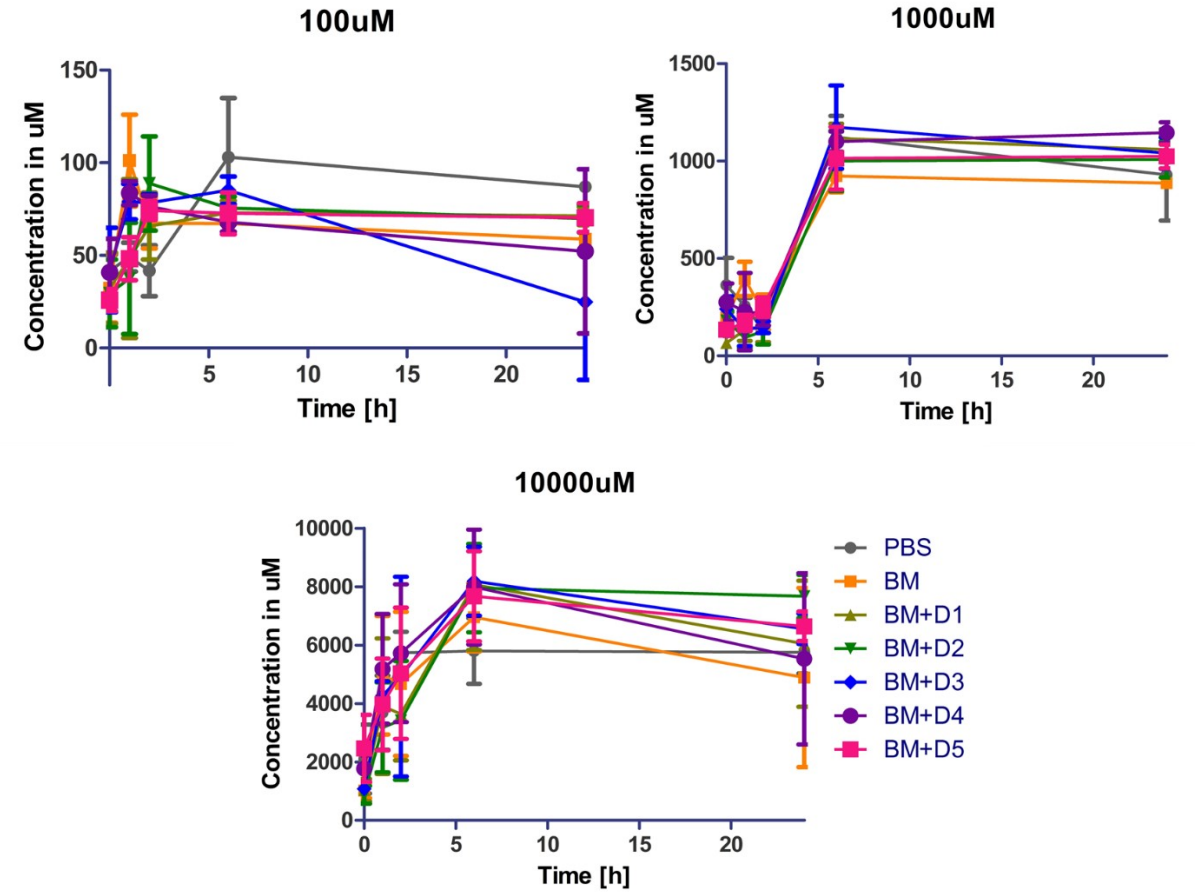



**Figure 4-4. APAP concentrations measured in the first attempt to develop an *in vitro* model.** APAP was incubated in 96 well-plates with faecal samples from five healthy human donors (D1 to D5) and basal medium (BM), or in basal medium alone without any faecal sample as a control (orange line), or in PBS as another control (grey line). Collection time points were 0 h (immediately after addition of the drug), 1, 2, 6 and 24 h. Concentration spiked in the wells are indicated in the top of each graph (100, 1,000 and 10,000 µM). Each point is the mean of the triplicates and the error bars are the standard deviations.

As indicated in Figure 4-2, row H of each plate was used for another type of control, where only faecal samples and the drug were incubated together, without the basal medium. This control was made at the middle concentration (1,000 µM) of the drug, and in duplicate for each of the faecal donors, on each of the five plates incubated at different time points. Interestingly, concentration of the drug in the different faecal samples (measured at 482.04 +/- 100.55, Figure 4-5) at 0 h was higher than when the drug was incubated with both faecal sample and basal medium (208.50 +/- 111.54 µM, Figure 4-4). This could suggest a solubility issue of the drug under these conditions for the first few hours of incubation before an equilibrium was reached (around 6 h). No degradation of APAP was observed between the 6 and the 24 h collection time points, although significant variability was observed, particularly for donor 1 at the 6 h time point.

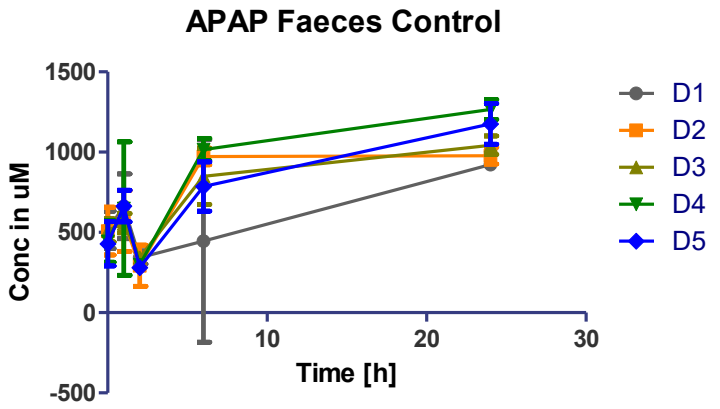



**Figure 4-5. APAP controls concentrations measured in the first attempt to develop an *in vitro* model.** Control samples from the miniaturized *in vitro* model where APAP was incubated with faecal samples alone from the five healthy human donors (D1 to D5) without basal medium (BM). Collection time points were 0 h (immediately after the addition of the drug), 1, 2, 6 and 24 h. Concentration spiked in the wells was 1,000 µM. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *3.2.2.Sulfasalazine*

High amount of variability was observed for the first three collection time points (0, 1 and 2 h). As a result, these data were not considered reliable. The concentration of the drug even exceeded the amount initially spiked in the model, as it appeared that when 100 µM was incubated, the measured concentration reached 122.72 +/- 42.03 µM in the 1 h samples, and when 1,000 µM was incubated, the measured concentration reached 1,496.80 +/- 372.94 µM in the 6 h samples. For the highest concentration (10,000 µM), the initial concentration observed was 5,808.60 +/- 2287.50 µM. Nevertheless, the concentration of SSZ was observed to decrease over time, particularly between the 6 h and the 24 h collection time points. However, this may indicate a stability issue with the drug rather that microbial metabolism as degradation occurred at a similar rate in the PBS control (except at the lowest concentration 100 µM).

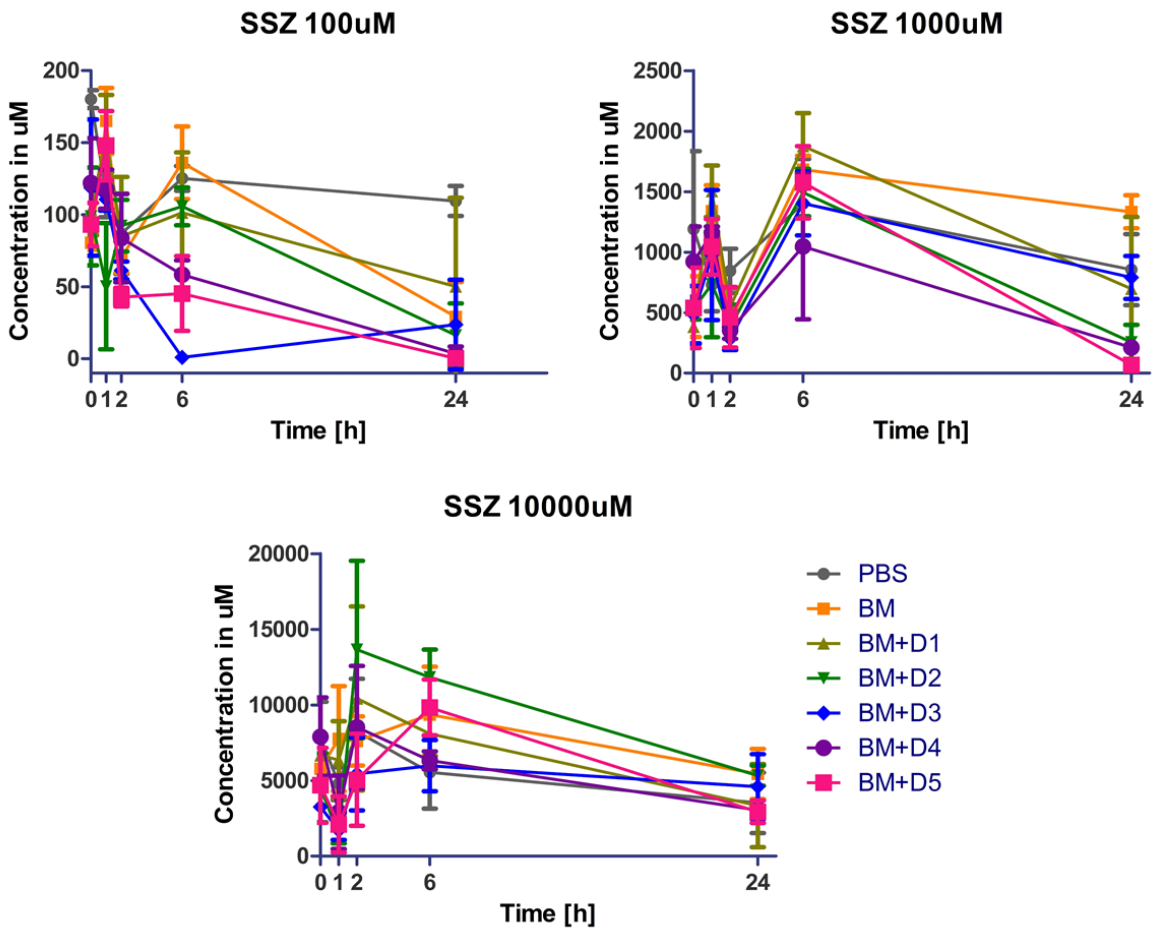



**Figure 4-6. SSZ concentrations measured in the first attempt to develop an *in vitro* model.** Results obtained for the miniaturized *in vitro* model where SSZ was incubated in 96 well-plates, with either one faecal sample from the five healthy human donors (D1 to D5) and basal medium (BM), or just in glucose-free basal medium without any faecal sample as a control (orange line), or in PBS as another control (grey line). Collection time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

No pattern difference was noted when SSZ was incubated in the faecal samples without any basal medium (Figure 4-7), compared to the incubation of SSZ in faecal samples and basal medium (Figure 4-6). Although extensive variability was observed for the first three time points.

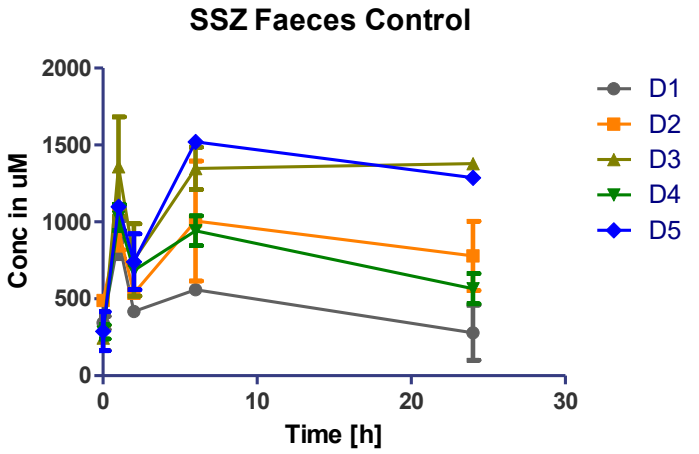



**Figure 4-7. SSZ control concentrations measured in the first attempt to develop an *in vitro* model.** Controls during the first attempt toward a miniaturized *in vitro* model where SSZ was incubated just with the faecal sample from the five healthy human donors (D1 to D5) but without basal medium (BM). Collection time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

### 3.3. Conclusions

In this study, significant variability in drug concentrations was observed. These results are best explained by drug solubility issues, particularly during the early time points of the experiment, which limits the interpretation of the resulting data. As such, further optimisation of the *in vitro* model was required. One optimisation step could involve a pre-incubation step so the drug can reach equilibrium before starting the incubation or finding an "ideal" solvent to ensure solubility, is required.

## 4. Optimization experiments

Optimization experiments were performed to address the problems encountered in the initial experiments, notably on the solubility issues and the fact a 6 h incubation seem to be needed before to reach an equilibrium. The optimization experiments consisted in:

- Test temperature (room temperature *vs* 37°C) and diluents ($H_2O$, PBS, basal medium, gut model medium (GMM)) influence on APAP concentration over time.
- Test glass and plastic containers effect on APAP concentration over time.
- Determinate the effect of anaerobic and aerobic pre-incubation on APAP concentration over time.
- Explore APAP interactions with faecal fixed and non-fixed cells.

## 4.1. Experimental section

### *4.1.1. Medium*

In comparison to the basal medium previously used and described, a gut microbiota medium (GMM) was also prepared as previously reported (Goodman et al., 2011) with peptone water (5 g/L), tryptone (5 g/L), yeast extract (4.5 g/L), NaCl (4.5 g/5L), KCl (4.5 g/L), casein (3 g/L), pectin (2 g/L), xylan (2 g/L), arabinogalactan (2 g/L), $NaHCO_3$ (1.5 g/L), $MgSO_4$ (1.25 g/L), guar gum (1 g/L), inulin (1 g/L), cysteine.HCl (0.8 g/L), $KH_2PO_4$ (0.5 g/L) and bile salts (0.4 g/L) which were dissolved in 4 L of $H_2O$ before adding starch dissolved in $H_2O$ (5 g/L), mucin (paste in $H_2O$, 4 g/L), $CaCl_2.6H_2O$ (0.15 g/L), $FeSO_4.7H_2O$ (5 mg/L), hemin in a few drops of 1M NaOH (0.05 g/L), tween 80 (1 mL/L) and vitamin K (10 µL/L). This medium was autoclaved before use.

### *4.1.2. Incubation experiment*

*Sample preparation.* A fresh stock solution of APAP was prepared in MeOH at a concentration of 500 mM. Once prepared, samples were stored at 4˚C until analysis.

*Optimisation experiment 1 – testing temperature and diluents.* APAP stock solution (500 mM in MeOH) was diluted in five different diluents: $H_2O$, DMSO, PBS, basal medium and GMM to obtain a final concentration of 50 mM. The five solutions were incubated at either room temperature (RT) or 37 ˚C for 24 h. For each of the incubations (5 diluents x 2 temperatures = 10 APAP solutions), 12 µL were sampled at three time-points (0, 6 and 24 h), diluted in $H_2O$ (500 µM) and stored at 4˚C until analysis.

*Optimisation experiment 2 – testing glass vs. plastic container.* APAP stock solution (500 mM in MeOH) was diluted in GMM (50 mM). This solution was transferred in triplicate to a 96-well plate, identical to the one used for the *in vitro* model. In parallel, the solution was added in triplicate to glass vials. All were incubated at 37 ˚C for 24 h. For each of the incubations (2 containers x triplicate = 6 APAP solutions), 12 µL were sampled at three time-points (0, 6 and 24 h), diluted in $H_2O$ (500 µM) and stored at 4˚C until analysis.

*Optimisation experiment 3 – testing anaerobic vs. aerobic pre-incubation of drug.* APAP stock solution (500 mM in MeOH) was diluted in GMM (50 mM). This solution was transferred into nine glass vials. Three of the glass vials were incubated aerobically and 12 µL were collected at three time-points (0, 6 and 24 h), diluted in 1,188 µL of $H_2O$ (final concentration = 500 µM) and stored at 4˚C until analysis. Due to the need to establish and maintain anaerobic conditions, it was not possible to sequentially sample the anaerobic incubations, so the six other glass vials were placed into an anaerobic jar with

an anaerobic sachet. Three were incubated for 6 h and three for 24 h. The anaerobic sachet placed into the jar was changed when the three vials of the 6 h were taking out of the incubation. From each vial, 12 µL were collected, and diluted as described above in $H_2O$ (to 500 µM) and stored at 4 ˚C until analysis.

*Optimisation experiment 4 – testing drug interactions with fixed/non-fixed cells.* APAP stock solution (500 mM in MeOH) was diluted in GMM (100 mM). The solution was separated into two batches of 5 mL to perform a pre-incubation step of 24 h under two different conditions, one conducted anaerobically and the other aerobically, both at 37˚C for 24 h. Faecal homogenates (10 wt. %) were prepared in phosphate-buffered saline (PBS, 0.1 M phosphate buffer solution, pH 7.4) with 10 mm glass beads added before being vortexed for 5 minutes. The faecal homogenate was also pre-incubated anaerobically for 24 h at 37 ˚C. Paraformaldehyde (PFA; 4 %) solution in PBS was prepared by mixing 2 g of PFA, 30 mL of LC-MS grade water and 100 µL of 1 M NaOH and gently warmed to aid dissolution. Then 16.6 mL of PBS were added to the solution before it was pH adjusted with 100 µL 1 M HCl (pH 7.2). Finally, the volume was increased to 50 mL with LC-MS grade water. The use of PFA in the incubation allows the fixation of the bacterial cells stopping their metabolic activity.
Post pre-incubation, 300 µL PFA, 400 µL of the pre-incubated APAP in GMM and 100 µL of the pre-incubated faecal homogenate were mixed into 24 vials (50 mM) with four different conditions; three aerobically pre-incubated PFA(+), three aerobically pre-incubated PFA(-), three anaerobically pre-incubated PFA(+), three anaerobically pre-incubated PFA(-). All were then incubated under anaerobic conditions at 37 ˚C. After 6h 50% were terminated and the remainder after 24h. For all incubations, samples of 12 µL were collected, diluted in 1,188 µL $H_2O$ (500 µM) and stored at 4 ˚C until analysis.

#### *4.1.3. UPLC-MS*

*Calibration curve (CC), Internal Standard (IS) and Quality Control (QC) samples.* APAP was dissolved in MeOH and diluted in $H_2O$ to have 14 solutions for the calibration curve as indicated in Table 4-S3. Calibration curves were injected between each optimisation experiment. Internal standard APAP-d3 was included and spiked in the calibration curve solutions at a concentration of 350 nM. APAP QCs were provided at three concentrations: low, medium and high (5 nM, 150 nM and 500 nM, respectively).

*UPLC-MS parameters.* The same instrument and parameters to monitor APAP transitions were used as those employed during the first attempt made toward a miniaturised *in vitro model*. The MS parameters for detection of APAP-d3 are reported in Table 4-S2.

## 4.2. Results

### *4.2.1. Optimization experiment 1*

To ensure that the drug was completely solubilised before being used in the *in vitro* model, the APAP solution in MeOH was diluted into four different diluents, $H_2O$, PBS, basal medium and GMM, prior to an incubation either at room temperature (RT) or 37 °C. No real difference was observed between either the different diluents or between the two temperatures tested (Figure 4-8). However, the concentration of APAP in GMM (44.91 +/- 2.96 mM) was the closest to expected (50 mM), particularly when incubated at 37 °C (45.65 +/- 4.21 mM), and as both this medium and this temperature seemed more suitable to reproduce GI tract conditions, the GMM and the incubation at 37 °C were used for further development.

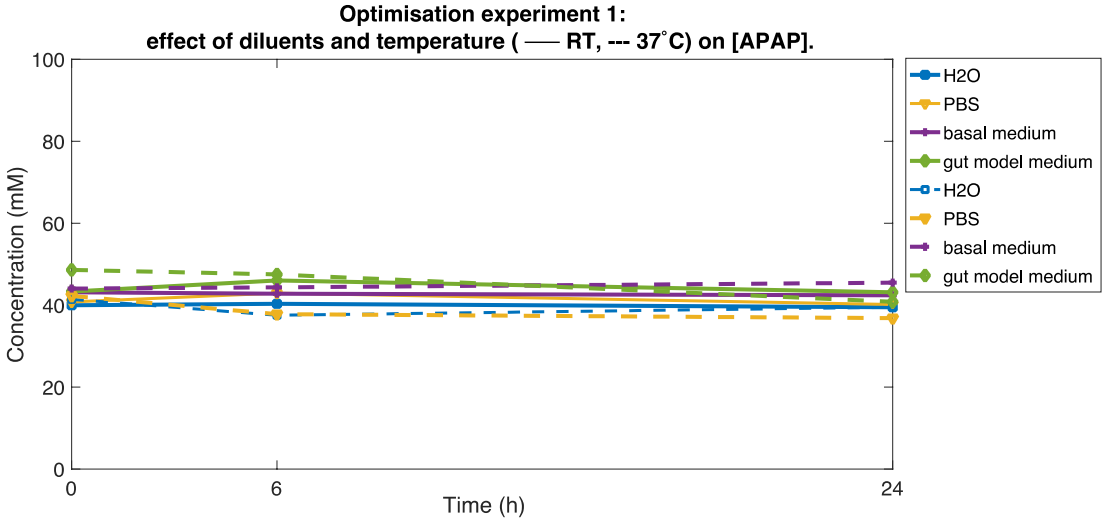



**Figure 4-8. APAP concentrations in the first optimization experiment testing different diluents ($H_2O$, PBS, basal medium and gut microbiota medium) at two temperature (room temperature and 37°C).** Drug concentrations were stable but slightly below (42.12 +/- 3.01 mM) the concentration expected (50 mM). Incubation of APAP in GMM at 37°C gave the closest concentration (45.65 +/- 4.21 mM). Each point is the mean of the triplicates and the error bars are the standard deviations.

### *4.2.2. Optimisation experiment 2*

The possibility of APAP binding to the walls of the model (plastic well plate) during the incubation was investigated by using plastic wells versus glass vials for incubation of the drug. No significant difference was observed at 0 or 24 h ($p = 0.448$ and $p = 0.08$, respectively, *t*-test) between the two conditions (Figure 4-9), while a significant difference was observed at the 6 h ($p = 0.046$, t-test), indicating that the use of plastic well-plates for the incubation had not resulted in any binding.

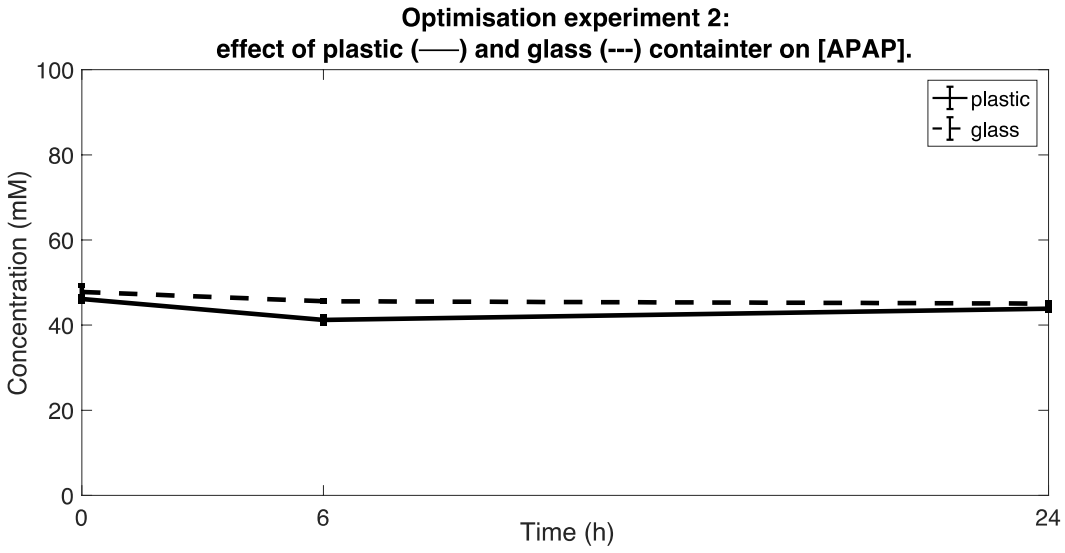



**Figure 4-9. APAP concentrations in the second optimization experiment testing plastic *vs* glass containers.** Significant difference was observed in drug concentrations between the two culture materials following 6 h of incubation (p = 0.046, t-test) but not initially (0 h) and not after 24 h of incubation. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *4.2.3. Optimisation experiment 3*

As faecal bacteria growth is predominantly oxygen sensitive and requires anaerobic incubation conditions, the effect of aerobic and anaerobic conditions on drug concentrations during the incubation was investigated. APAP in GMM (50 mM) was incubated under both conditions with drug concentration measured by UPLC-MS. The results suggest that at 6 h, neither anaerobic (46.09 +/- 1.87 mM) nor aerobic (47.69 + 1.11 mM) conditions had significant effect on drug concentrations (p = 0.2845, *t*-test), nor at 24 h (50.83 +/- 5.88 mM vs 45.34 +/- 1.01 mM, respectively, p value = 0.2452, *t*-test- Figure 4-10).

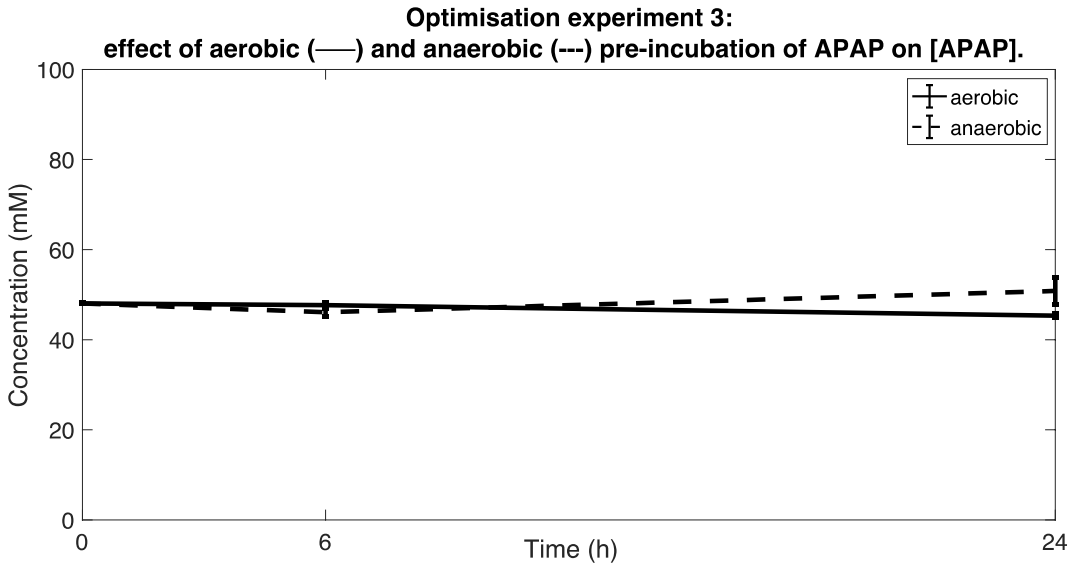



**Figure 4-10. APAP concentrations in the third optimization experiment testing incubation under aerobic or anaerobic conditions.** Drug concentrations were stable and close to the 50 mM expected (aerobic: 46.52 +/- 1.60 mM; anaerobic: 48.33 +/- 3.71 mM) and no statistical differences were observed between the samples incubated under both conditions at any of the time points. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *4.2.4. Optimisation experiment 4*

To ensure that no other non-bacterial components in faeces interacted with the drug once mixed into the model, pre-incubated faecal homogenate was spiked with 4 % PFA to fix the microbial cells and halt their metabolism prior to being incubated with APAP in GMM (50 mM). APAP concentrations were measured with UPLC-MS and showed a significant difference across 0 h samples ($p = 7.6x10^{-7}$ - ANOVA) due to a high concentration of APAP measured when incubated without PFA in anaerobic condition (59.94 +/- 0.65 mM) compared to without PFA in aerobic condition (50.71 +/- 0.08 mM), with PFA in anaerobic condition (50.97 +/- 0.30 mM) and with PFA in aerobic condition (48.18 +/- 0.30 mM). However, after 6 and 24 h of incubation, no more significant difference was observed ($p = 0.971$ and $p = 0.621$, respectively, ANOVA - Figure 4-11).

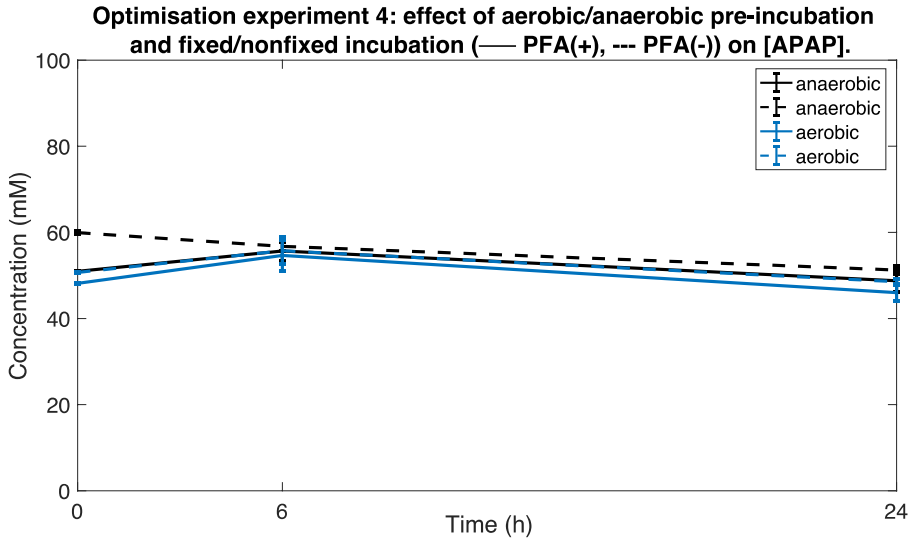



**Figure 4-11. APAP concentrations in the fourth optimization experiment testing incubation under aerobic or anaerobic conditions, both with normal and fixed (4 % PFA) faecal homogenate.** Significant difference was observed between the conditions at 0 h but not at 6 and 24 h ($p = 7.60 \times 10^{-7}$, $p = 0.971$ and $p = 0.621$, respectively, ANOVA). Each point is the mean of the triplicates and the error bars are the standard deviations.

## 4.3. Conclusion

Various model design conditions were tested using APAP in order to explore the solubility problem identified in the first study. The different diluents, room temperature or 37 °C, the containers tested (glass or plastic) and the aerobic/anaerobic conditions had no effect on drug concentrations. These experiments cannot explain the variability in the initial drug concentrations seen in the model during the first experiments. However, these findings contribute to the development of this *in vitro* model as these parameters were found not to influence the results. This validates the use of the intestinal-inspired composition of the GMM and incubation at body temperature (37 °C) mimicking real host conditions. It was also shown that neither aerobic nor anaerobic conditions had any effect on APAP concentrations. However, a significant difference was observed in the concentration of APAP when incubated in plastic *vs* glass container at 6 h, but no significant was observed at 24 h. Due to the lower cost and their convenience, it was decided to adopt the plastic well-plates as a container for the incubation, and, in consequence, a pre-incubation step of the drug in GMM at 37˚C in aerobic conditions during 24 h was added to the protocol. Lastly, drug-faeces interactions were tested by fixing microbial cells in the faecal homogenate with 4 % PFA prior the addition of APAP. Significant difference was observed only initially (0 h) but not at 6 or 24 h following the beginning of the incubation, showing that no interactions, independent of the bacteria, were identified to impact on drug concentrations and confound the model after at least 6 h. A pre-incubation step for the faecal samples at 37 ˚C of 24 h was thus also implemented, under anaerobic conditions to ensure the survival of the bacteria.

# 5. Refinement of the *in vitro* model

A pre-incubation step with the drug in GMM and separately, the faecal homogenate in GMM, was implemented. The drugs tested in these experiments were APAP and SSZ, as well as methotrexate (MTX) and the glucuronidated form of paracetamol (APAP-G), which arises from host phase II drug metabolism. APAP-G enables bacterial β-glucuronidase activity to be studied in the faecal samples, which is explored in Chapter 6.

## 5.1. Experimental section

### *5.1.1. Medium*

The medium used for these experiments was the gut microbiota medium tested in the preceding optimisation experiments and was prepared as described previously.

### *5.1.2. Incubation experiment*

*Pre-incubation of the faecal homogenates.* Once prepared, faecal homogenates were collected and diluted in GMM [2.572 mg/mL] before pre-incubation under anaerobic conditions on a shaker at 37 ˚C for 24 h.

*Preparation and pre-incubation of drugs in GMM.* Stock solutions (24 mM) were prepared for all the test drugs and then diluted in four intermediate solutions (0.48 mM, 4.8 mM, 24 mM and 3 mM) using MeOH for the three drugs APAP, SSZ and APAP-G, and DMSO for MTX. To produce solutions in GMM for plate preparation, solutions with concentration of 0.48 mM, 4.8 mM and 24 mM were diluted in GMM (18 µM, 180 µM, 900 µM, respectively). Solutions with a concentration of 3 mM were diluted in $H_2O$ (180 µM) for control. GMM alone serve as the 0 M drug concentration. Final drug concentrations of 0, 10, 100 and 500 µM were used for selected wells during the incubation. Plates were then incubated aerobically on a shaker at 37 ˚C for 24 h.

*Plate preparation, incubation and termination.* After completing a 24 h pre-incubation step of plates with the drug in GMM and a separate 24 h pre-incubation step of faecal homogenate in GMM, the two components were mixed in a ratio 5:4, respectively, to obtain the final drug concentrations (0 M, 10 µM, 100 µM and 500µM, Figure 4-12). The rest of the workflow was the same as that applied during the first attempt towards a miniaturized *in vitro* model.

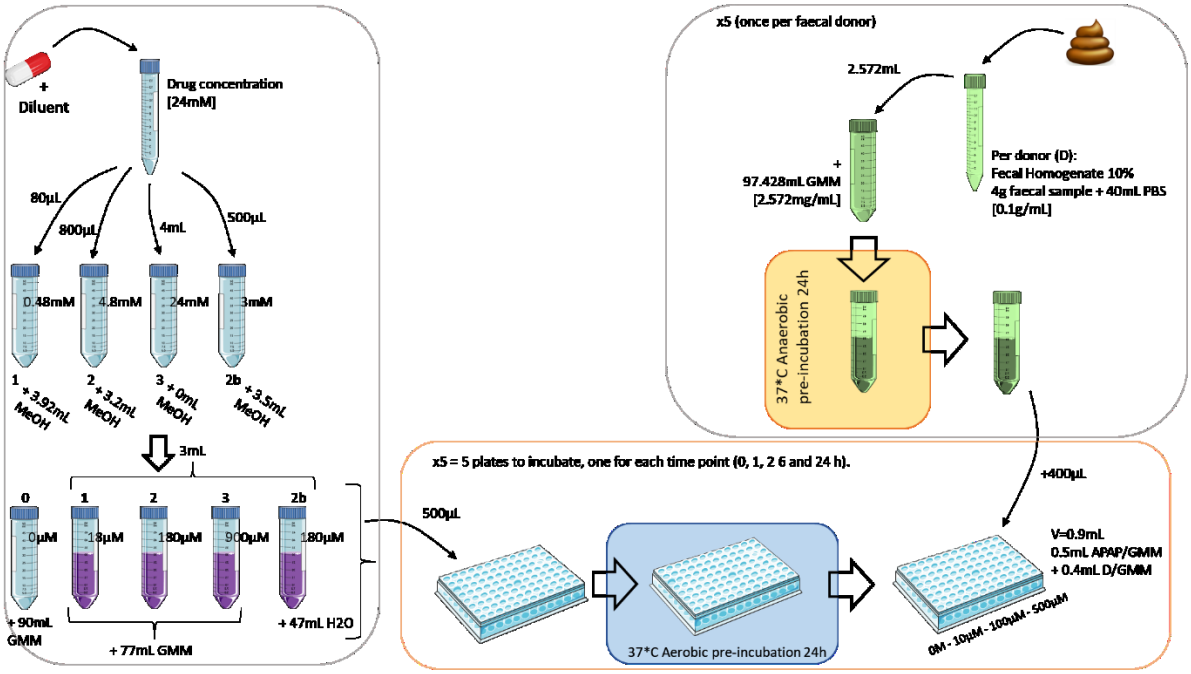



**Figure 4-12. Detailed pre-incubation steps added to the refined *in vitro* model.** The drug was dissolved and diluted at different concentration in a suitable solvent, in order keep the same volume of solvent when dilution into GMM is made (here 3 mL). Following dilution in GMM, pre-incubation directly in the plate in aerobic conditions at 37 ˚C for 24 h. The faecal homogenate was diluted in GMM and pre-incubated anaerobically at 37 ˚C for 24 h. Next, the faecal homogenate was added to the 96-well plate containing the pre-incubated drug and incubated anaerobically at 37 ˚C before to initiate the anaerobic incubation of the drug and the faecal homogenate together.

*5.1.3. UPLC-MS*

*Sample preparation for APAP and APAP-G UPLC-MS analysis.* Compared to the first attempt, plate preparation for the refined *in vitro* model experiment attempt was similar, apart from an addition of 30 µL of APAP-d3 (IS) in the Ostro plate when the compounds studied were APAP and APAP-G, and a reduction in the volume of cold ACN to 420 µL. Centrifugation was used to transfer the solution to collection plates (concentrations at that stage were 0 M, 2.5 µM, 25 µM and 250 µM). Further dilutions in water were made to obtain a final concentration of 500 nM in all the wells. All UPLC-MS plates were stored at 4 ˚C until UPLC-MS analysis (Figure 4-13).

*Sample preparation for SSZ UPLC-MS analysis.* The same protocol was applied for the plates incubated with SSZ, with the exception that no internal standard was available here so the volume of ACN added to the wells was 450 µL. Two consecutive dilutions were needed to obtain a final concentration for UPLC-MS of 90 nM.

*Sample preparation for MTX UPLC-MS analysis.* MTX plates were prepared in a similar way as for SSZ but the final concentration for UPLC-MS analysis was 16.2 nM.

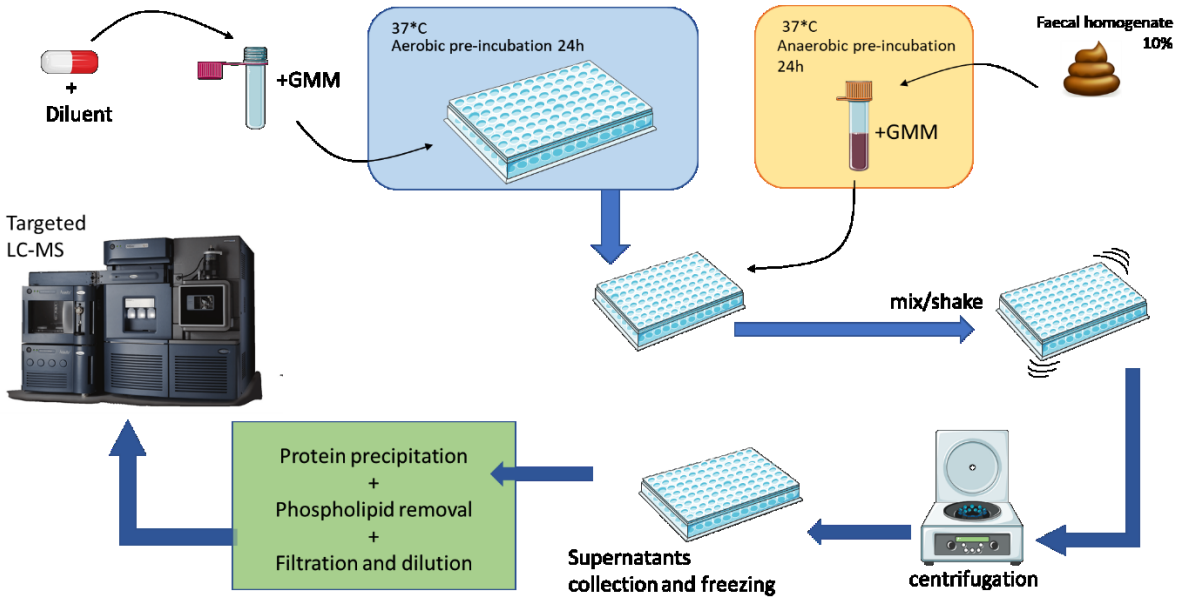



**Figure 4-13. Refined *in vitro* model workflow.** The drug was diluted in GMM and directly pre-incubated in the plate in aerobic conditions at 37 ˚C for 24 h. The faecal homogenate was diluted in GMM and pre-incubated anaerobically at 37 ˚C for 24 h. Next, the faecal homogenate was added to the 96-well plate containing the pre-incubated drug and incubated anaerobically at 37 ˚C. Supernatants were collected at 0, 1, 2, 6 and 24 h after initiation of incubation with faecal homogenate and then prepared for UPLC-MS targeted analysis.

*UPLC-MS parameters.* The UPLC-MS analysis was the same as previously reported for the first attempt toward an *in vitro* model. As MTX and APAP-G were also analysed with the refined *in vitro* model, the gradient, the MS parameters and the linear range are reported in supplementary information (Tables 4-S1, 4-S2 and 4-S3, respectively).

## 5.2. Results

APAP and SSZ were selected as negative and positive controls, respectively. MTX was selected as donor-dependent variation in efficacy has been observed previously across patients (see Chapter 3). Finally, APAP-glucuronide (APAP-G) was investigated to assess bacterial β-glucuronidase activity in the model (explored in Chapter 6). Concentrations tested through this second attempt to develop an *in vitro* model were 0, 10, 100 and 500 µM.

### *5.2.1.Acetaminophen*

For the three concentrations tested, initial drug concentrations at 0, 1 and 2 h across the four APAP concentrations were affected by a confounding factor causing lower concentrations than expected to be measured at the 1 and 2 h time points before increases in the concentration were observed at the 6 h time point (Figure 4-14). Drug concentrations stabilised after 6 h, but concentrations remained either lower or higher than expected when the concentration was supposed to be 10 µM (7.16 +/- 1.93 µM) or at 500 µM (744.08 +/- 59.03 µM) but was correct for 100 µM (109.15 +/- 8.69 µM). Surprisingly, APAP was detected at low concentrations in a control well with no drug spiked in. However, no degradation of APAP in any of the donors or controls, *i.e.* PBS, GMM and $H_2O$ was measured, which is in accordance with the hypothesis that APAP is suitable as a negative control.

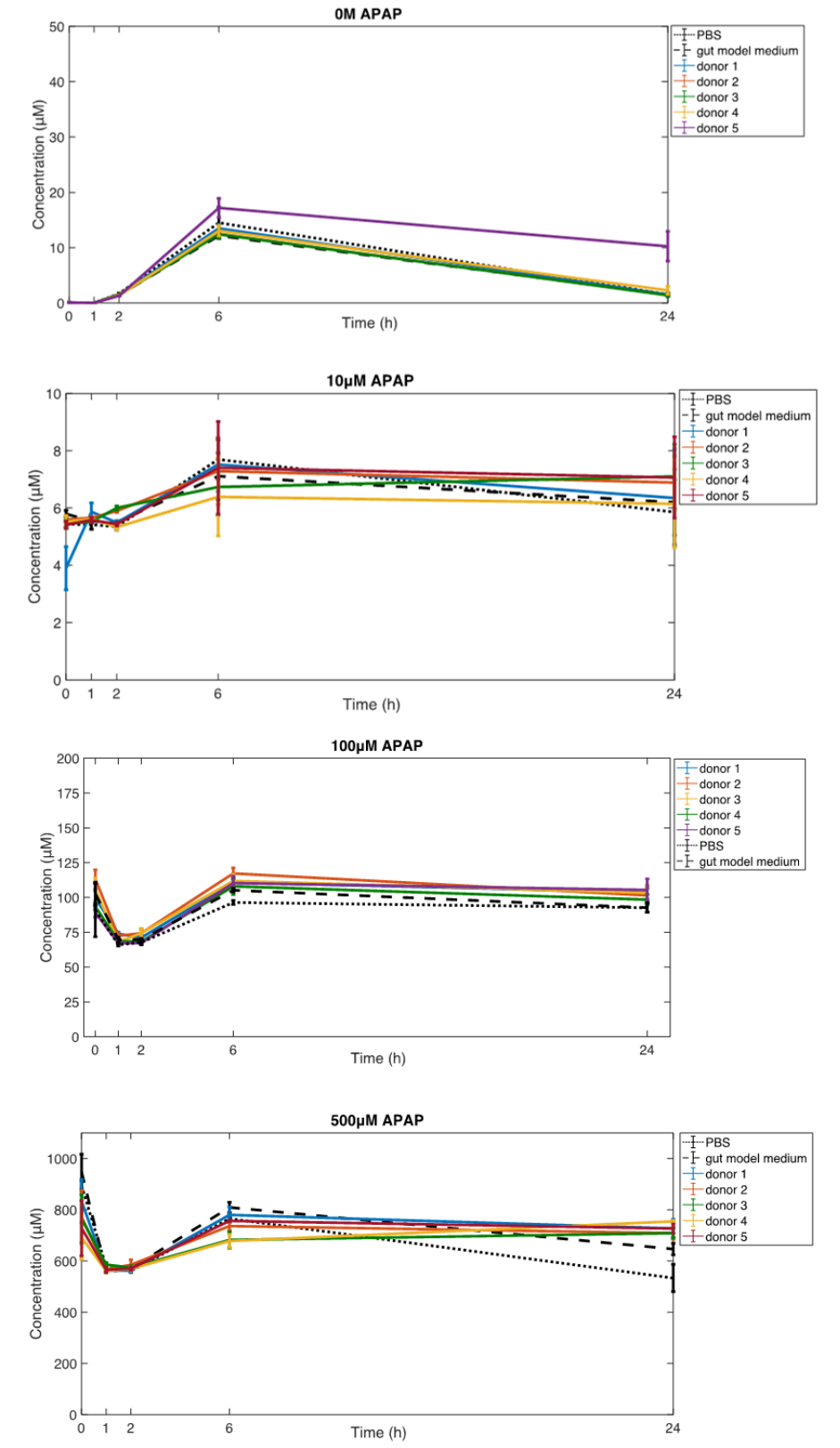



**Figure 4-14. APAP concentrations measured in the refined an *in vitro* model.** APAP was incubated in 96 well-plates, at four concentrations (0, 10, 100, 500 µM, indicated on top of the graphs) with either a faecal sample from one of the five healthy human donors and gut microbial medium (GMM), or as control, in GMM, PBS of faecal homogenate alone. Sampling time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *5.2.2.Sulfasalazine*

Initial SSZ concentrations were in accordance (11.90 +/- 8.65, 100.67 +/- 18.51, 584.04 +/- 238.81 µM) with those expected (10, 100 or 500 µM), although a small increase above the highest expected concentration was observed (Figure 4-15). Moreover, degradation of the drug in all donors was noted, and even though the same observation was made in the controls (PBS or GMM), the degradation occurred quicker in the presence of faecal samples. Interestingly, variation was noted in the velocity of degradation across the donors and this variation was conserved from one concentration to another. Overall, faecal samples from donors 1 and 2 showed total degradation, at a faster rate, than the faecal samples of donors 3, 4 and 5. This indicates donor-specific degradation patterns.

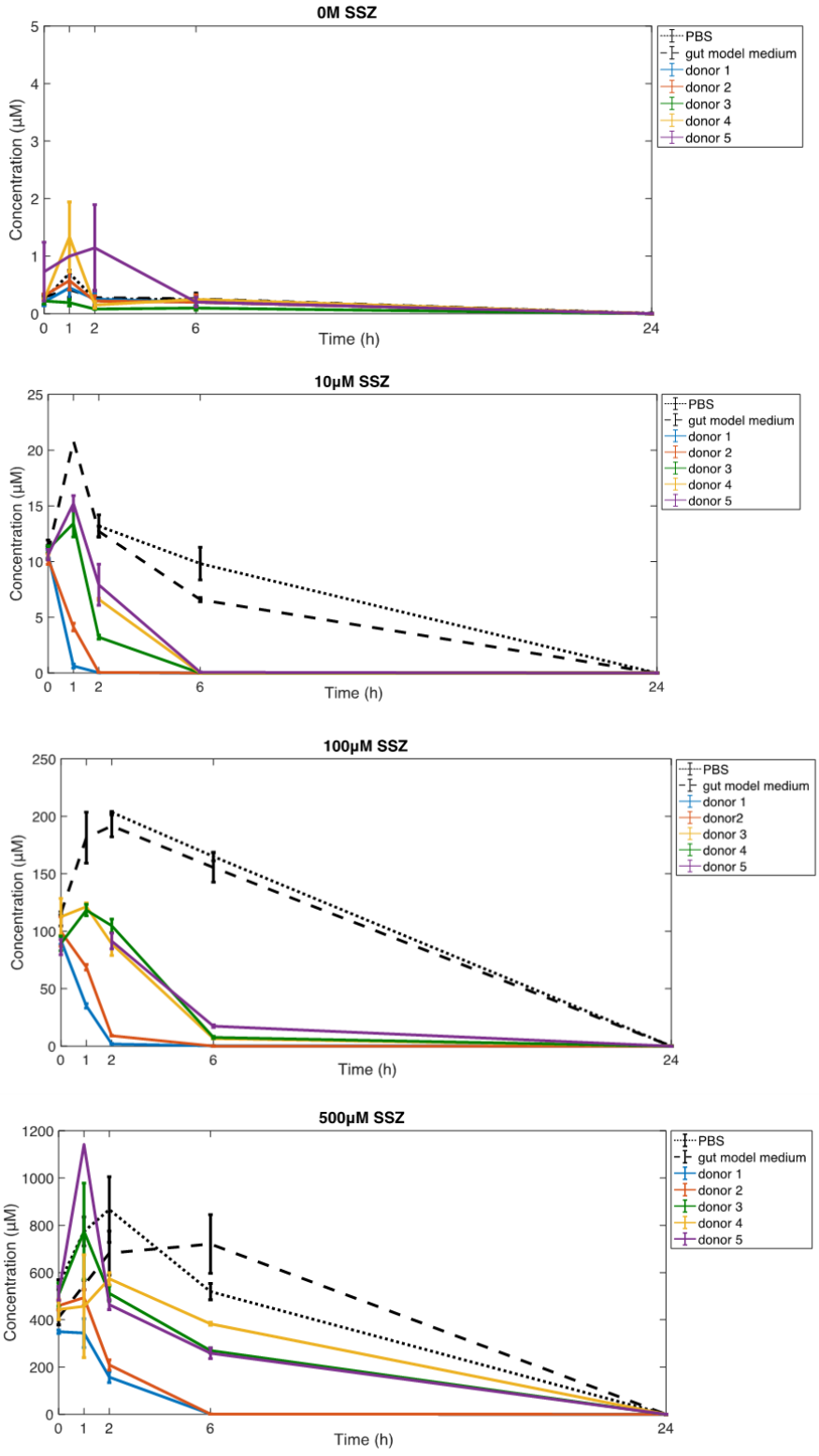



**Figure 4-15. SSZ concentrations measured in the refined an *in vitro* model.** SSZ was incubated in 96 well-plates, at four concentrations (0, 10, 100, 500 µM, indicated on top of the graphs) with either a faecal sample from one of the five healthy human donors and gut microbial medium (GMM), or as control, in GMM, PBS of faecal homogenate alone. Sampling time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *5.2.3.Methotrexate*

At the initial time point, the measured concentration of MTX (9.98 +/- 2.49 µM) was only consistent with the amount added for the 10 µM incubation (Figure 4-16). For the two highest doses the concentrations measured were higher than expected (242.71 +/- 110.56 µM instead of 100 µM and 1162.92 +/- 435.63 µM instead of 500 µM) and unexpected patterns were again observed in the first three time points for the highest concentrations. Also, 6 µM of the drug where detected in the wells in which no drug had been spiked in for all the five time-point plates. No degradation was really observed over time. However, variability was observed in the quantities of MTX between the donors without that any donor-specific pattern was underlined.

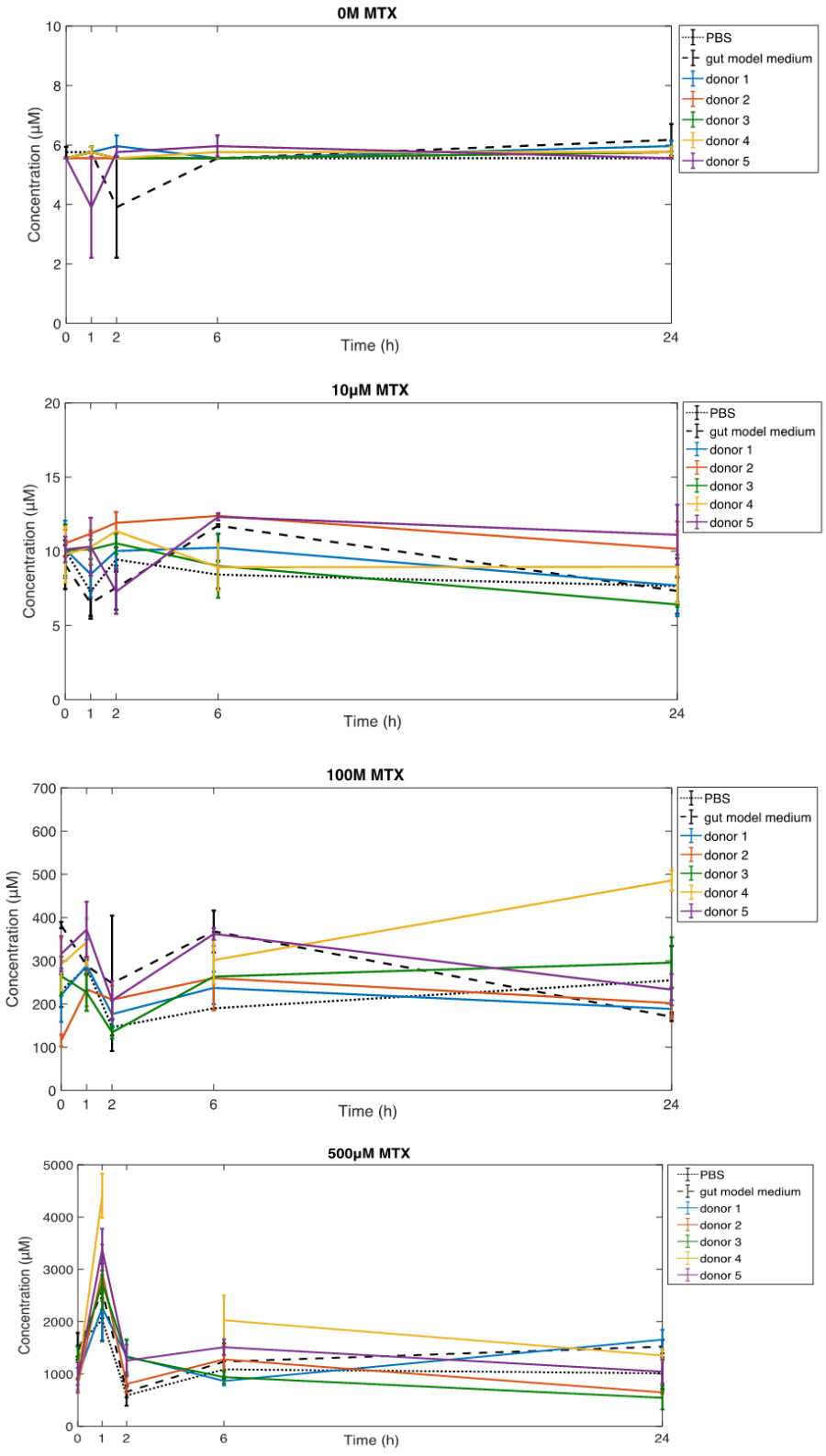



**Figure 4-16. MTX concentrations measured in the refined an *in vitro* model.** MTX was incubated in 96 well-plates, at four concentrations (0, 10, 100, 500 µM, indicated on top of the graphs) with either a faecal sample from one of the five healthy human donors and gut microbial medium (GMM), or as control, in GMM, PBS of faecal homogenate alone. Sampling time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

### *5.2.4.Acetaminophen-glucuronide*

As for MTX, no degrading pattern in the concentration of APAP-G were measured by UPLC-MS (Figure 4-17). The measured concentrations of APAP-G were at higher concentrations than expected (16.05 +/-1.58 instead of 10 µM, 309.46 +/- 99.46 µM instead of 100 µM and 2,048.15 +/- 1,129.76 µM instead of 500 µM). It did not stabilise as for the other compounds as it kept increasing over the 24h of incubation. However, degradation appeared to have occurred for both PBS and GMM controls in the higher concentrations (3.77 +/- 0.95 µM in PBS and 10.50 +/- 2.65 µM in GMM at 24 h when incubated with 500 µM of APAP-G).

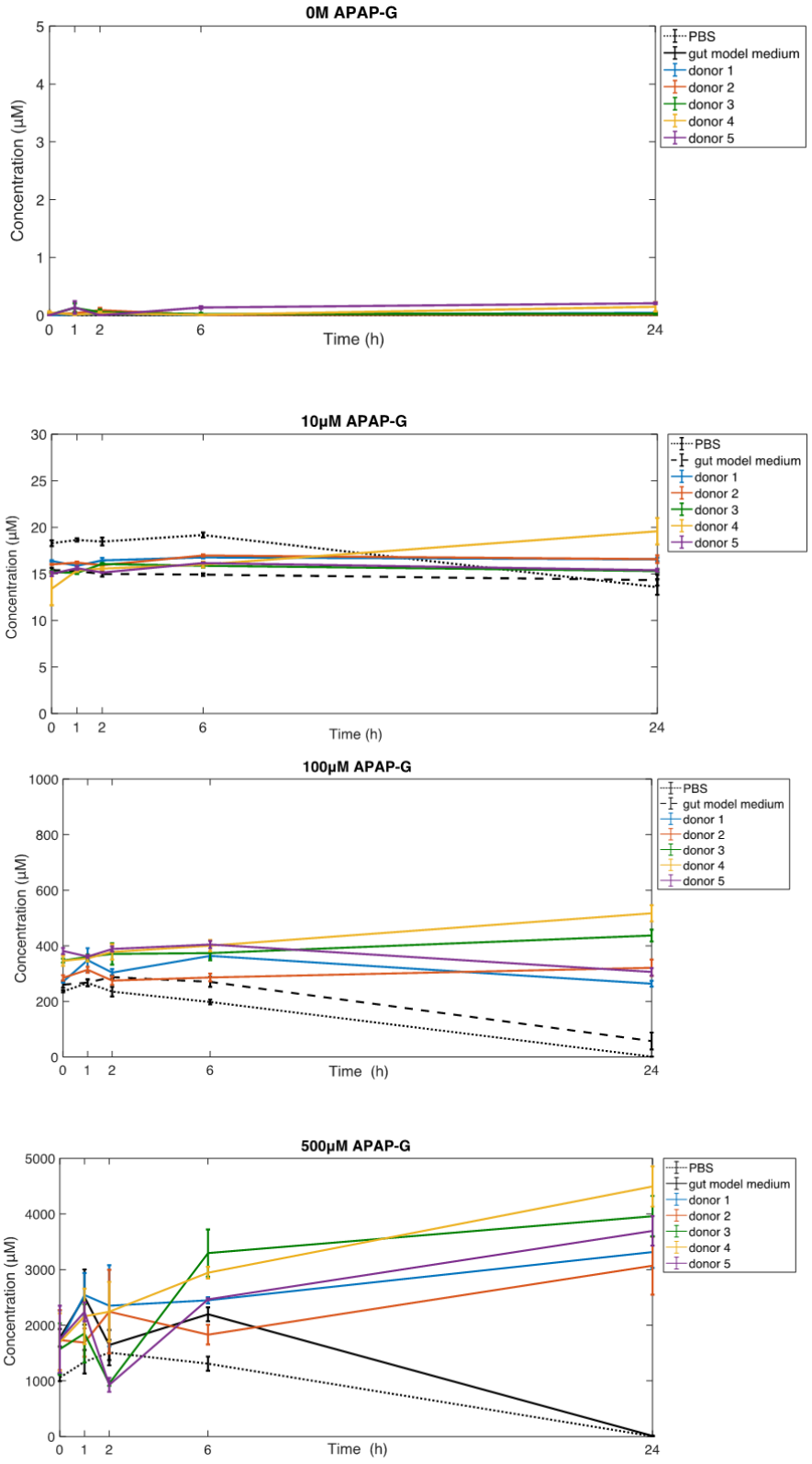



**Figure 4-17. APAP-G concentrations measured in the refined an *in vitro* model.** APAP-G was incubated in 96 well-plates, at four concentrations (0, 10, 100, 500 µM, indicated on top of the graphs) with either a faecal sample from one of the five healthy human donors and gut microbial medium (GMM), or as control, in GMM, PBS of faecal homogenate alone. Sampling time points were 0 h (just after spiking in the drug), 1, 2, 6 and 24 h. Each point is the mean of the triplicates and the error bars are the standard deviations.

## 5.3. Conclusion

The results obtained from APAP and SSZ obtained during the refinement experiment (Figures 4-14 and 15, respectively) compared to the results obtained during the first attempt to develop the *in vitro* model (Figures 4-4 and 4-6, respectively) were slightly improved, as less variability was observed at the first three time points and the degradation patterns seem consistent with our expectations (APAP being a negative control and SSZ a positive control). However, when the model was applied to the compounds MTX and APAP-G (which were run at the same time and using the same faecal samples) the results indicate methodological issues. The expected metabolites for each drug (DAMPA for MTX and APAP for APAP-G) were also measured by UPLC-MS but neither of them was detected. An important detail was noted between the *in vitro* models applied to APAP and SSZ compared to MTX and APAP-G whereby, after the pre-incubation and during the incubation itself, the faecal samples used during the *in vitro* models applied to APAP and SSZ, emitted a characteristic odour, indicative of a viable culture. However, this odour was absent when the model was applied to MTX and APAP-G, which could suggest an issue with the bacterial culture during this incubation, possibly explaining the lack of degradation of these test compounds. Finally, detection of the drugs by UPLC-MS in the wells where no drugs had actually been added, especially for APAP and MTX, suggests an eventual evaporation of the solutions which could condensate on the cap-mat and lead to the contamination of other wells. This would also explain the high variability between wells containing the same concentrations and is consistent with variable volumes observed in the wells of a same plate following the incubation.

# 6. General discussion

To study the potential for an individual's gut microbiota to modulate the bioavailability of an orally ingested drug, attempts were made to develop a simple, cost-effective, 96 well plate-based, *in vitro* model that could rapidly assess this capacity based on an individual's faecal sample.

## 6.1. Strengths

This protocol was designed to be applied to a range of drugs, using faeces from human donors, as a preliminary screen to inform any future *in vivo* studies, offering extensive access to the bacterial community and microbial metabolites of the colon. Most importantly, it provides the opportunity to explore inter-individual variability in microbially-mediated drug metabolism. Indeed, such variability was observed with microbial SSZ degradation in the *in vitro* model (Figure 4-15). The inter-individual variability could potentially be explained by a subset of faecal donors able to reduce SSZ more rapidly than other donors, thus releasing the active ingredient in the gut sooner modifying its pharmacological properties. Careful consideration was given to developing a cost-effective and efficient model, utilising a 96-well plate. This number of wells would enable a high-throughput model able to evaluate a range of drugs (with replicates) in a single patient with the addition of control samples as well. The nature of this model also enables a variety of different concentrations to be studied simultaneously.

UPLC-MS sample preparation included filtration with Ostro plates to remove unwanted contaminants such as, *e.g.* protein and phospholipids if present, to clean-up samples prior analysis and avoid ion suppression. For each analysis, a single or a selection of specific compounds were of interest. Consequently, targeted UPLC-MS was preferred to deliver accurate metabolite quantification based on the use of calibration curves. Accordingly, MRMs transitions and suitable MS parameters for each of the test drugs were chosen based on published data (Ambavaram et al., 2013; Gu et al., 2011; Guo et al., 2007) and short UPLC-MS gradients were developed for optimal separation and detection (Tables S1, S2 and S3). Calibration curves and within-run QCs were used in every analysis to assess linearity, concentration and accuracy during the run, and an IS was used where feasible. Untargeted UPLC-MS could also be implemented to explore the dynamic metabolic profile of the faecal homogenates following their different incubation times with a specific drug.

## 6.2. Limitations

Several limitations exist for the model we attempted to develop. Firstly, the use of MeOH or DMSO to dissolve the drugs probably affect the culture of the bacteria, even if these solvents represent only 1% of the total volume incubated. This could be explored by comparing the metabolic or the microbial profile of faecal cultures with or without solvent. Also, the model being *in vitro* and not *in vivo*, does not simulate any of the host metabolic activities, such as drug absorption, hepatic, or extrahepatic, metabolism or enterohepatic recirculation. The absence of enterocytes and intestinal epithelium of 3D morphology does not replicate the gastrointestinal cavities in which drug encounters bacteria *in vivo*. It simply allows an investigator to determine if a drug is degraded if placed in contact with a human faecal sample. Another limitation of this model is thus the assumption that the faecal microbial populations and their activity *in vitro* reflect those found in the colon. Although previous research reported that human faecal transplants in germ free mice were able to reproduce normal bacterial variation along the gut (Turnbaugh et al., 2009b), and that three-stage fermentation system with temperature, pH and nutrients control has proved to accurately model human colon (Macfarlane et al., 1998), miniaturized cultures such as the one reported herein have not yet been demonstrated to accurately model the diversity and functionality of the gut microbiota. Analysis of the bacterial profile of the samples just after collection and following culture, with and without drugs, could help determining how representative the bacterial community obtained after the incubation is compared to the real bacterial community. Moreover, it appears that the incubation step of the faecal samples is sensitive to bacterial loss of viability. As was discussed for the MTX and APAP-G experiments, the characteristic odour of a viable culture was not present for these incubations. This may suggest that subtle variation in the plate preparation may have killed some members of the bacterial community, especially strictly anaerobic ones. Preparation in an anaerobic cabinet or culture of the faecal samples for several days under anaerobic conditions, as has already been reported (Goodman et al., 2011), could be an option to maintain the metabolic activity of the bacterial community. Finally, careful consideration needs to be taken regarding the choice of the solvent. Discrepancies in the initial concentrations of the drugs and changes throughout the experiment suggest that there were solubility issues. Optimisation of the solvents for delivering the drugs may overcome these issues but care must be taken as solvents such as DMSO can impact the microbiome community and their growth, potentially inducing false negative results.

### 6.3. Perspectives and future work

Although the results obtained in the experiments reported above were not entirely satisfactory, the *in vitro* model has the strong potential to rapidly determine the capacity of an individual's microbiome to degrade a specific drug, and thus help toward personalised therapeutic treatments. Such information would provide supportive evidence that the drug PK profile might partly be shaped by interaction with the gut microbiota. Such interactions could, in certain circumstances, explain inter-individual variability in drug responses and toxicity. It could also be applied in clinical setting, medical research and drug development. For instance, for hospitalised patients, it could help identify optimal drug treatments or dose levels prior to administration. It could also be useful in pharmaceutical industry for determining the interaction of new drug candidates with faecal bacterial contents. If further optimisation of the model was successful, measuring gut microbial diversity could also be pursued, either using fluorescence *in situ* hybridisation (FISH) analysis, or amplicon sequencing (as described in Chapter 2). This could enable not only the measurement of the impact of the gut microbiota on drug metabolism but also the effect of a given drug on the microbial diversity of the faecal samples.

## 7. Conclusion

Using the compounds APAP, SSZ, MTX and APAP-G, a promising start has been made in developing a rapid *in vitro* model that could be used to investigate the effect of the complex gut microbiota on the metabolism of specific drugs. This model has potential to determine the direct interactions occurring between the microbiome and a specific drug and promote the development of personalised therapeutic treatment. However, this methodology requires further development to answer the current issues and properly assess the feasibility of incorporating such a tool in routine clinical or of pharmacological investigations. Consequently, future studies should aim to develop standardised methodologies, based on this model, and assessed against a much wider range of drugs. A valuable complimentary technique to this model would be to explore, in addition to just the disappearance of the drug, both drug metabolite concentrations and the impact of the drug on the gut microbiota diversity and composition. This would help to make it a completer and more informative model on the bidirectional interaction between the gut microbiota and the drug(s) under investigation.

# Chapter 5. Dietary manipulation of gut microbial metabolism and its influence on host paracetamol metabolism

**Aim:**

- To investigate the impact of a high tyrosine diet on the gut microbial production of *p*-cresol and related metabolites in mice.

- Determine the impact of diet-induced changes in microbial tyrosine metabolites on the host metabolism of paracetamol.

**Objectives:**

- To establish a mouse model feeding a high tyrosine diet to enhance the microbial production of *p*-cresol and subsequently increase the hepatic conjugation of these microbial metabolites in the host.

- To determine the contribution of the gut microbiota to the dietary-induced modulation of conjugative drug metabolism by feeding the high tyrosine diet in combination with antibiotics.

- To characterise tyrosine and paracetamol metabolism in the mouse using two targeted UPLC-MS-based methods to measure metabolites in urine, plasma and liver samples.

- To investigate the wider impact of a high-tyrosine diet, antibiotic treatment and paracetamol treatment on the global hepatic metabolic profile using $^{1}$H NMR spectroscopy.

# 1. Introduction

*N*-Acetyl-*p*-aminophenol (APAP), commonly known as paracetamol or acetaminophen is the most commonly used antipyretic and analgesic drug (Sharma and Mehta, 2013) and is categorised under the "Essential Medicines" list by the World Health Organization. The safe dosage limit for adults is recommended to be 1-2 g/day to treat mild pain or fever symptoms (Graham et al., 2013), with a maximum dosage limit of 4 g/day for adults (Mazaleuskaya et al., 2015). Despite its wide therapeutic window, APAP overdose is one of the most commonly reported drug overdoses (Mazaleuskaya et al., 2015). Because of this, it is the main cause of acute liver failure in the US and other major developed countries (Hodgman and Garrard, 2012).

In humans, orally administrated APAP gets rapidly absorbed in the duodenum and transported to the liver. It then undergoes hepatic conjugation reactions, mainly involving sulfation (20 to 30 %) and glucuronidation (55 to 60%; Hodgman and Garrard, 2012), leading to the detoxification of the drug and rapid excretion *via* urine or bile (Mast et al., 2015). APAP also follows two other oxidation reactions, one leading to a catechol derivative (< 5 %) and another one during which it gets oxidized by cytochrome P450 (CYP450) enzymes (*i.e*. CYP2E1 and CYP3A4) leading to the formation of a toxic metabolite called *N*-acetyl *p*-benzoquinone imine (NAPQi; 5 to 15 %; Hodgman and Garrard, 2012)). NAPQi can either be detoxified by glutathione (GSH) conjugation, leading to the formation of APAP-cysteine (APAP-C), which can get further acetylated to APAP-*N*-acetylcysteine (APAP-NAC), or it can bind to thiol functions of cell proteins causing toxicity.

One emerging factor that has been shown to influence the efficacy and toxicity of APAP in human is the microbial metabolite *p*-cresol (Clayton et al., 2009). *p*-Cresol is produced from the bacterial degradation of tyrosine (Smith and Macfarlane, 1997). Tyrosine undergoes degradation to tyramine and *p*-hydroxyphenylacetate (*p*HPA) before being converted to *p*-cresol as a result of putrefaction by certain bacterial strains (*e.g. Clostridium difficle [C. difficile]*) found in the gastrointestinal tract (Sivsammye and Sims, 1990). *p*-Cresol has bacteriostatic properties that can confer a growth advantage to *C. difficile* over other bacterial strains (Passmore et al., 2018). Once absorbed from the gut, *p*-cresol is conjugated to produce *p*-cresyl glucuronide (*p*CG) and *p*-cresyl sulfate (*p*CS). However, *p*-cresol and APAP are structurally similar (Figure 5-1).

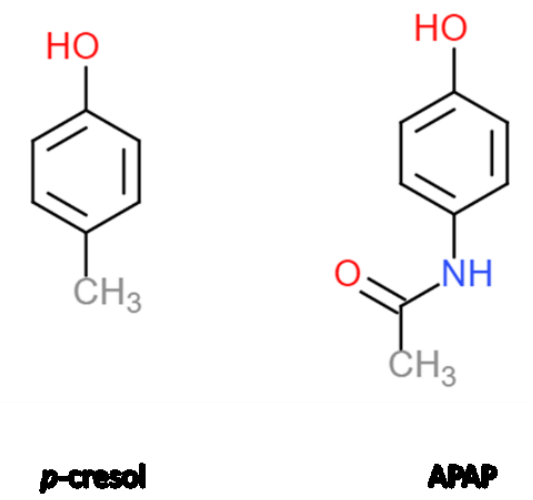



**Figure 5-1. Molecular structure of *p*-cresol and paracetamol (APAP).**

Because of their structural similarity, both can serve as substrate to SULT1A1 enzymes to undergo sulfation in the liver and colonic mucosa (Teubner et al., 2007). As such, both compounds compete for the binding site of the enzyme as well as the co-factor (3'-phosphoadenosine 5'-phosphosulfate) PAPS. Studies have shown that in humans, *p*-cresol is primarily converted to *p*CS indicating the high efficiency of *p*-cresol as SULT1A1 ligands (de Loor et al., 2005). High exposure of the liver to *p*-cresol prior to APAP treatment has been hypothesized to deplete the hepatic sulfate pool of the host and as a result, skew the metabolism of APAP towards other pathways (Clayton et al., 2009). This can lead to an increase in the generation of the hepatotoxic NAPQi, but also to a saturation of enzymes like glutathione-*S*-transferase (GST) which plays a major role in NAPQi detoxification (Ketterer Brian et al., 1983). This competition between APAP and *p*-cresol for host phase II (conjugative) drug metabolizing systems is represented in Figure 5-2.

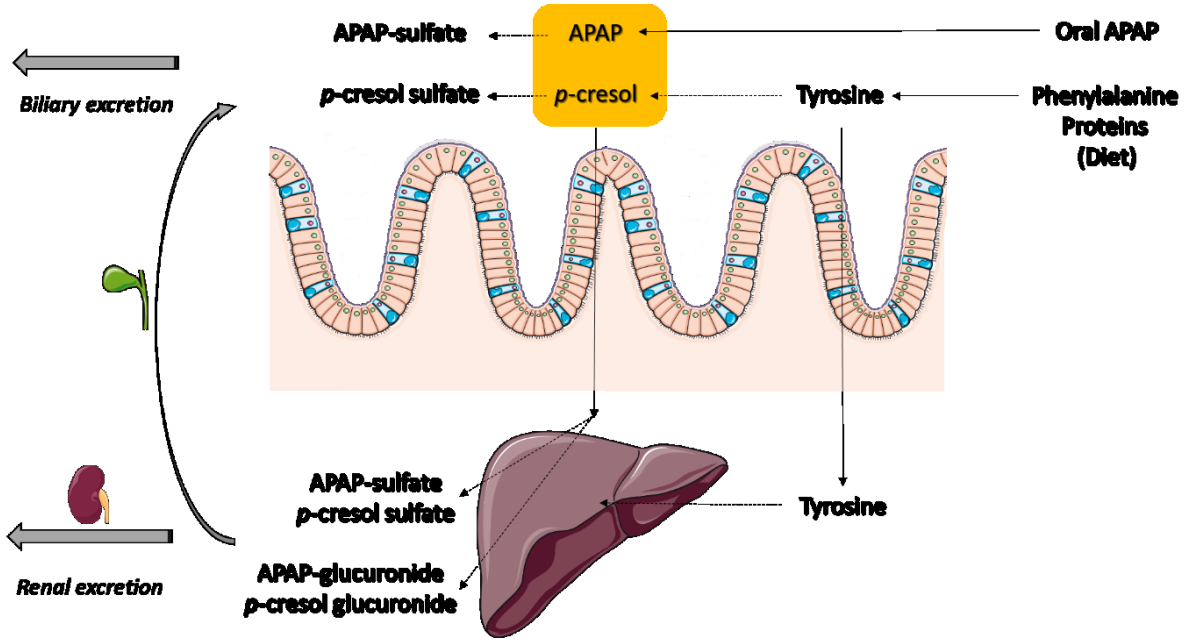



**Figure 5-2. Schematic representation of the competition between *p*-cresol, obtained from the microbial degradation of tyrosine from the diet, and the common used orally administered analgesic APAP.** Both are conjugated with sulfate and glucuronide by bacterial and hepatic enzymes, before urinary or biliary excretion occurs.

Nevertheless, the exact indirect mechanism occurring between the microbial metabolite *p*-cresol and APAP metabolism is unclear. Moreover, sulfation through SULT1A1 is a major pathway implied in the elimination of other xenobiotics and endogenous compounds, such as hormone or neurotransmitter crucial for early human development (James and Ambadapadi, 2013). It is then crucial to understand how high concentrations of *p*-cresol obtained from a dietary route can modify sulfation capacity to influence host xenobiotic and endogenous metabolism.

As *p*-cresol is produced by the bacterial degradation of tyrosine and that tyrosine is an amino acid abundant in proteins, it is then assumed than individual consuming a high-protein diet will have a higher level of *p*-cresol and its conjugated derivatives. This observation has also been made in people consuming either a low-fibre (Khosroshahi et al., 2019) or a high-fat diets (Wan et al., 2019). Perturbation in the gut microbiota was also observed in the high fat diet study, with decreased relative abundance of the phylum Firmicutes and increased relative abundance of Bacteroidetes (Wan et al., 2019). Consequently, it is also key to understand how the microbiome can process a high-tyrosine diet and influence the production of *p*-cresol.

In this study, a high tyrosine diet (diet based on 5 % by weight of tyrosine) was administrated to 16 female BALB/mice for a period of 2 weeks to increase the bacterial production and host exposure to *p*-cresol and its conjugated derivatives. A subtoxic dose of APAP (300 mg/kg, p.o.) was administered to the mice after 2 weeks on the high tyrosine diet to observe how the high tyrosine diet would alter APAP metabolism. The hepatic metabolic profiles of the mice were explored using $^{1}$H-NMR spectroscopy and two targeted UPLC-MS based assays were used to selectively quantify tyrosine catabolites and APAP metabolites in urine, plasma and liver. To demonstrate/understand the contribution of the microbiota in this diet-based model an additional study group was included where the microbial metabolism was suppressed with antibiotics and the high tyrosine diet was supplied. The results were compared and validated with control mice group.

# 2. Experimental section

## 2.1. Chemicals and reagents

**Tyrosine assay:** The chemicals used to perform the tyrosine targeted assay are the same as those reported in the analytical background chapter (supplementary information for the paragraph II.1.3.).
**APAP assay:** Paracetamol (APAP), APAP-glucuronide (APAP-G; sodium salt), deuterated APAP internal standard (APAP-$d_3$) and dansyl chloride were purchased from Sigma Aldrich (Gillingham, U.K.). APAP-sulfate (APAP-S; potassium salt), cysteinyl APAP (APAP-C; trifluoracetic acid salt), *N*-acetylcysteine APAP (APAP-NAC; disodium salt) and deuterated internal standards APAP-S-$d_3$ (potassium salt), APAP-G-$d_3$ (sodium salt), APAP-C-$d_5$ (trifluoracetic acid salt), APAP-NAC-$d_5$ (sodium salt) were purchased from Toronto research chemicals (Toronto, Canada).
**Mass spectrometry:** For sample preparation and UPLC-MS, LC-MS Chromasolv acetonitrile (ACN), isopropanol (IPA) and methanol (MeOH) were purchased from Honeywell (Seelze, Germany). Optima LC/MS Grade $H_2O$ with 0.1% formic acid (FA) and analytical reagent grade formic acid were purchased from Fisher Scientific Ltd (Loughborough, U.K.). Dansyl chloride and ammonium hydroxide (28% w/v in water), as well as leucine, enkephalin, acetate salt (hydrate) and sodium formate for UPLC-MS mass accuracy and calibration, leucine enkephalin acetate salt (hydrate) and sodium formate, were obtained from Sigma-Aldrich (Gillingham, U.K.).
**$^{1}$H NMR spectroscopy:** Potassium phosphate salts ($KH_2PO_4$), TSP (3-(trimethylsilyl)-[2,2,3,3-$d_4$] propionic acid sodium salt, sodium azide ($NaN_3$), methanol (MeOH) and $D_2O$ were all purchased from Sigma-Aldrich (Gillingham, U.K.).

## 2.2. Animal handling

The animal study was conducted by Janvier Labs (Genest-Saint-Isle, France) in accordance with the 2010/63/UE guideline and its transposition in the French law. The protocol was validated by the ethical committee of the Janvier LABS animal experimentation (n°78). Forty BALB/c female mice were randomly assigned to three diet groups - high tyrosine (HT) (*n* = 16), high tyrosine group plus antibiotic treatment (AB) (*n* = 16) and a control group (C) (*n* = 8). The control diet consisted of a standard AIN-93G diet and the high-tyrosine diet was fortified with 5% tyrosine (compared to 0.98% for the control diet, Table 5-S1) based on a previous study (Shiba et al., 2014). Mice were eight weeks old at the beginning of the dietary study and were maintained on their respective diets for 14 days. Antibiotic treatment (4 mg/mL of streptomycin and 2 mg/mL of penicillin G in the animal's drinking water) was provided on days 11-14 for the AB group. Urine was collected from all animal groups at three different time points – pre-administration of the assigned diet (PD), and after 7 and 14 days from the beginning of the study. Blood collection by the mandibular vein was made prior mixing with lithium heparin as an anticoagulant and centrifugation at 2,690 *g* for 15 minutes to collect plasma. Plasma collection was undertaken at two time points – pre-administration of the assigned diet (PD) and after 14 days from the beginning of the study. On day 15 of the administration of the assigned diet, the animal groups received a subtoxic dose of APAP (300mg/kg, p.o.). Half the number of animals *(n* = 8 for the HT and AB groups, *n* = 4 for the CT group) were euthanized with carbon dioxide one-hour post APAP treatment (1 h) and the other half were euthanized six hours post-APAP treatment (6 h). Prior to euthanasia, urine samples were collected from the urinary meatus. Once euthanized, liver samples were collected as well as blood via cardiac puncture. Plasma were extracted as described above. Samples were stored at -80°C until mailing and then at -40°C until analysis.

## 2.3. UPLC-MS targeted analysis

### *2.3.1. UPLC-MS targeted assay for tyrosine and its catabolites*

The method of preparation of the calibration curve solutions and of the QC samples along with the UPLC-MS method and parameters used, are described in the paragraph II.1.3 Targeted UPLC-MS of the Analytical Background (Chapter 2). Sample preparation is described below.

**Urine and plasma sample preparation:** Urine and plasma were prepared by mixing an aliquot of 10 or 20 µL (depending on the initial volume of sample available) with ACN: IPA (1:1, v/v) and IS mixture at a final ratio of 1:3:1 v/v/v. The IS standard mix used for urine samples was at 3.2 µM whilst that for plasma samples was 20 µM. The QC biological pool sample was prepared by collecting 20 µL from each sample and mixing them together. A blank QC pool sample (QC-B) was prepared by mixing 20 µL of $H_2O$, 20 µL of QC biological sample and 60 µL of ACN: IPA (1:1, v/v). A QC biological sample with internal standard (QC-IS) was prepared by adding 20 µL of IS mix to 20 µL of QC sample and 60µL of ACN: IPA (1:1; v/v). A QC biological pool sample with internal standards and standard (QC-IS-S) was prepared by mixing 30 µL of IS mix, standard mix (intermediate solution 1 at 10 µM) and 60µL of ACN: IPA (1:1; v/v). The rest of the protocol was the same as that used for the samples. The samples were vortexed and left overnight at -20°C for protein precipitation, followed by centrifugation at 11,250 *g* for 10 minutes at 4°C. A volume of 30 µL of the supernatants were collected and placed in 96 well-plates. Derivatization of the analytes was performed by adding 24 µL of 7% $NH_4OH$ and 6 µL of a 12.5 mg/mL solution of dansyl chloride, in order to derivatise the phenolic hydroxyl group of the targeted compounds. The plates were foil heat sealed and incubated at 65$^{o}$C for 30 minutes. Due to the design of the animal study, concentrations of the analytes were uncertain prior to sample analysis (especially for those mice given the HT diet). Therefore, to ensure that, following dilution, the samples lay within the linear ranges of the calibration curves analysis of various dilutions of the pooled QC samples were performed prior to committing the bulk of the samples to the analysis. Based on these a 1:50 dilution with water was used with the mouse urine samples in order to analyse tyrosine, tyramine, *p*-ethylphenol, *p*-cresol, phenol and *p*-hydroxyphenylpropioinic acid (giving final IS concentrations of 6.4 nM). A further dilution of 1:6 (final dilution 1:300) using water was necessary to analyze *p*-coumaric acid, *p*-hydroxybenzoic acid, *p*-hydroxyphenyl acetic acid, *p*-hydroxyphenyl lactic acid and *p*-cresol sulfate resulting final in IS concentrations of 1.067 nM). A dilution of 1:50 using water was done to analyse tyrosine and all its catabolites in the plasma samples, giving a final IS concentration of 40 nM.

**Liver sample preparation:** Liver samples (*n* = 40) were initially weighed out between 50-60 mg in bead beating tubes filled with 1 mm zirconium beads. A volume of 1,380 µL of cold $H_2O$:ACN:IPA (2:1:1, v/v/v) were added to all the samples along with 120 µL of tyrosine IS mix at 20 µM. The samples were homogenized using a Precellys 24 bead beater (40 sec per cycle, 6,500 Hz speed, 2 cycles). The samples were then centrifuged at 11,250 *g* for 10 minutes at 4°C. The supernatants from each sample was pipetted and stored for the APAP analysis as well (described below). A volume of 100 µL of each of the supernatants was collected and placed into 96 well plates in a randomized order and derivatized by adding 80 µL of $NH_4OH$ (7% aq) and 20 µL of dansyl chloride (12.5 mg/mL in acetone). Sample plates were heat-foiled and incubated at 65°C for 30 minutes. Samples were firstly analyzed without dilution (IS concentration = 800 nM) but was later diluted in water (1:50) in order to analyze tyrosine within the linear range (IS concentration = 16 nM).

### *2.3.2. UPLC-MS targeted assay for APAP and its metabolites*

The method of preparation of the calibration curve solutions and of the QC samples along with the UPLC-MS method and parameters used, was developed and validated by R. Dargue (R. Dargue, 2019; not yet published). A brief description of them is given in supplementary information. Sample preparation is described below.

**Urine and Plasma sample preparation.** Urine (*n* = 40) and Plasma (*n* = 40) samples were thawed and 5, 10 or 20 µL (according the volume of sample available) were mixed with MeOH in a 1:3 ratio. Protein precipitation was performed at -20°C for 40 minutes. Samples were centrifuged at 11,250 *g* at 4 °C for 15 minutes. A volume of 5 µL of each supernatant was collected and added with 10 µL of SIL IS mix (described in supplementary information) and 235 µL of $H_2O$ for a 1:50 dilution. Urine samples required a further dilution of 1:300 to assay the following compounds- APAP, APAP-NAC, APAP-GSH and a dilution of 1:7500 with water to assay the following compounds- APAP-S, APAP-C and APAP-G. The final concentrations of internal standards after their respective dilutions contained 666.7 ng/mL APAP-$d_3$, 37.34 ng/mL APAP-NAC-$d_5$, 0.03 ng/mL paracetamol-glutathione-$d_3$ (APAP-GSH-$d_3$),142.1 ng/mL APAP-S, 23.3 ng/mL APAP-G, 34.6 ng/mL APAP-C. Plasma samples required a dilution of 1:50 with water and the final IS standard concentrations were 800 ng/mL APAP-$d_3$, 137.6 ng/mL APAP-S-$d_3$, 27.4 ng/mL APAP-G-$d_3$, 33.52 ng/mL APAP-C-$d_5$, 44.8 ng/mL APAP-NAC-$d_5$ and 0.04 ng/mLAPAP-GSH-$d_3$.

**Liver sample preparation:** Supernatants collected from the liver aqueous extraction performed for the tyrosine pathway targeted analysis were used for this section of the study. A volume of 5 µL of each of the sample supernatants were mixed with 5 µL of SIL IS mix and 240 µL of $H_2O$ (1:50 dilution). All APAP metabolites were within the respective linear ranges with this dilution. The final IS concentrations were 400 ng/mL APAP-$d_3$, 68.8 ng/mL APAP-S-$d_3$, 13.7 ng/mL APAP-G-$d_3$, 16.76 ng/mL APAP-C-$d_5$, 22.4 ng/mL APAP-NAC-$d_5$ and 0.02 ng/mL APAP-SG-$d_3$.

#### *2.3.3. Data analysis*

The raw LC-MS/MS data obtained from the tyrosine targeted assays were treated as described in the Analytical background (Chapter 2). Those obtained for the APAP targeted assay were processed using the TargetLynx package of the MassLynx software (version 4.1., Waters Corporation). The raw data pre-processing and peak integration for all compounds were done individually using linear calibration curve response with a log axis transformation. For all the compounds except for APAP-GSH, internal standards were spiked in the samples and calibration curves in order to monitor their peak area response and correct the nominal concentration of the corresponding standards. The results acquired from both targeted assays were then imported into R and box plots were made in R using the ggpubr package (version 0.2.3). Univariate statistics within each diet group was done using Wilcoxon test. Multivariate statistical analyses, data integration and modelling for all the acquired MS data was done using MixOmics R package (version 6.1.1).

## 2.4. $^1$H NMR spectroscopy

#### *2.4.1. Liver-sample extraction and preparation for NMR analysis*

Frozen liver samples were thawed, and 50-60 mg of tissue were placed in bead beating tubes filled with zirconium beads (1 mm diameter, Stratech Scientific Ltd, Ely, UK). The tubes were then filled with 300 µL of pre-chilled $CHCl_3$:MeOH (2:1, v:v) and $H_2O$ (300 µL). Samples were homogenised with a Precellys 24 bead beater (40 sec per cycle, 6,500 Hz speed, 2 cycles). The samples were later centrifuged at 11,250 *g* for 10 minutes at 4 $^oC$. Aqueous and organic layers were carefully separated into separate tubes after discarding the fat layer in-between. Pellets were resuspended with 300 µL of pre-chilled $CHCl_3$: MeOH (2:1, v:v) and $H_2O$ (300 µL). The samples were vortexed and spun at 11,250 *g* for 10 minutes at 4$^oC$. The aqueous and organic layers were again separated and added to the previous aliquots, thereby achieving a double extraction. The aqueous extracts were dried overnight in a savant vacuum concentrator set at 30 $^oC$ in vacuum aqueous mode and the organic extracts were

left in a fume hood to dry. The samples were then stored at -40°C until the day of $^{1}$H NMR spectroscopic analysis. The dried aqueous extracts were then thawed and resuspended using 540 µL $D_2O$ and 60 µL urine buffer (9:1; v:v). $^{1}$H NMR spectroscopy urine buffer was prepared by dissolving 10.2 *g* of $KH_2PO_4$ in 40 mL $D_2O$ and 50mg of TSP (3-(trimethylsilyl)-[2,2,3,3-$d_4$] propionic acid sodium salt, used for spectral calibration); 6.5 mg of $NaN_3$ (sodium azide, in order to prevent bacterial contamination) in 3-5 mL of $D_2O$ and buffer pH was adjusted to 7.4 using KOH 1M solution. The final volume was made up to 50 mL using $D_2O$. Each liver sample was vortexed followed by centrifugation at 11,250 *g* for 10 minutes at 4°C. A volume of 580 µL of supernatant of each sample was pipetted into 5mm (outer diameter) NMR tubes.

### *2.4.2.$^{1}$H-NMR NOESY spectral acquisition of aqueous liver extracts*

The samples were randomised and one dimensional (1D) $^{1}$H NMR spectra was obtained using a Bruker 600 MHz spectrometer (Bruker, Rheinstetten, Germany) at an operating frequency of 600.13 MHz and a constant temperature of 300 K. The samples were referenced to TSP at $\partial$ 0 ppm. The NMR pulse sequence used a recycle delay (RD)-90°-$t_1$-90$^0$-$t_m$-90°-acquisiton, with $t_1$ = 3 µs, $t_m$ = 100 ms and 90° pulse ~10 µs. Water peak suppression was performed by using selective irradiation during RD of 4 s. The number of dummy scans used were 4 and the actual number of scans were set at 32 with a spectral resolution of 20 ppm.

### *2.4.3.$^{1}$H-NMR spectral pre-processing and multivariate statistical analysis*

Spectral phase correction and baseline correction were automatically performed by Topspin software (version 3.5; Bruker BioSpin GmbH, USA). The acquired raw spectra was imported into MATLAB (*R2*108b) using an in-house script to carry out multivariate statistical analyses. Peaks related to the water ($\partial$ 4.7-5.1 ppm) and TSP ($\partial$ 0 ppm) were removed. The raw spectra were subjected to pre-processing steps- manual alignment (recursive peak alignment method) and normalisation (probabilistic quotient method). Both, PCA and OPLS-DA models were generated with Pareto scaling used for the PCA models.

# 3. Results

## 3.1. Effect of high tyrosine diet on urine and plasma tyrosine metabolism

The tyrosine pathway targeted by UPLC-MS and the fit-for purpose method validation performed on the urine and plasma samples is reported in the analytical background (Chapter 2). Urine samples were collected from the three treatment groups (Control [CT], High tyrosine group [HT] and antibiotic treated high tyrosine fed group [AB]) at three different sampling points (Pre-diet [PD], 1-week post-high tyrosine diet [1w] and 2- week post-high tyrosine diet [2w]). In both HT and AB animals a significant increase was noted in urinary tyrosine, tyramine and *p*-hydroxyphenylacetate (*p*HPA) concentrations compared to the control group at 1 week and 2 weeks on the diet (Figure 5-3.a). Although a significant increase in urinary tyrosine and tyramine was also noted at 1 w and 2 w in the control compared to the PD sampling point, the increase was more significant for the HT and AB groups. Following AB treatment, from days 11 to 14 of the diet, the urinary excretion of *p*-cresol sulfate (*p*CS) and *p*-cresol glucuronide (*p*CG) was found to be lower compared to the HT group. Again, a significant increase was also observed in urinary *p*CS and *p*CG in the 1w and 2w samples of the control group compared to the PD samples of the control group, but this increase was lower compared to the increase seen in the HT group. However, no significant difference was observed in urinary *p*-cresol sulfate or *p*-cresol glucuronide between the control and the HT group at 1 w or 2 w.

Plasma samples were collected at the pre-diet and 2-week time points (Figure 5-3.b). Tyrosine, *p*-hydroxyphenyllactic acid (*p*HPL) and *p*-hydroxybenzoic acid (*p*HB) were significantly increased in the plasma from both the HT and AB groups at 2 w compared to the PD plasma samples of the respective groups. At 2 weeks, significant differences were observed between the CT and both HT and AB animals for plasma tyrosine and *p*HPL. However, no significant differences were observed in the plasma tyrosine, *p*HPL and *p*HB concentrations between the HT and the AB groups.

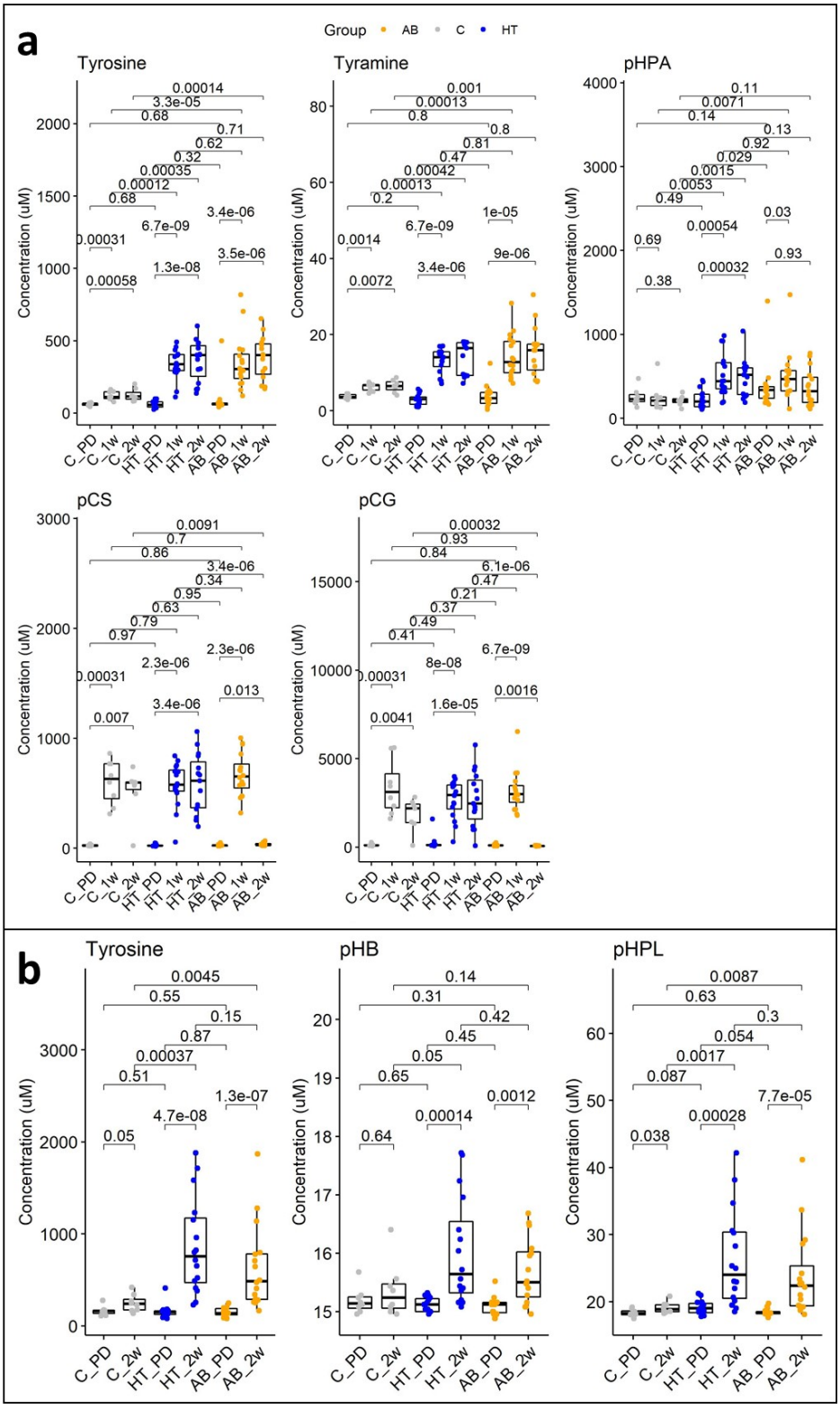



**Figure 5-3. High tyrosine diet effect on tyrosine metabolites in a) urine and b) plasma.** Whiskers in the box plots represent the full range quantiles and boxplots the 25, 50 and 75 % quantiles. Individual animals are indicated by a dot in the plot. Statistical tests within each group were done using Wilcoxon test. Only metabolites with significant p-values are reported on the graph. pHPA: p-hydroxyphenylacetic acid, pCS: p-cresol sulfate, pCG: p-cresol glucuronide; pHB: p-hydroxybenzoic acid; pHPL: p-hydroxyphenyllactic acid; C: control in grey; HT: high tyrosine diet in blue; AB: high tyrosine plus antibiotic treatment in orange; PD: pre-diet; 1w: 1week; 2w: 2 week.

## 3.2. Effect of high tyrosine diet on APAP metabolites detected in the urine, plasma and liver

A sub-toxic oral dose of 300 mg/kg of APAP was administered to all treatment groups and half of the groups were euthanized 1 hr post APAP treatment (Control, *n* = 4; HT, *n* = 8; AB, *n* = 8) and the other half 6 hr post APAP treatment. All APAP metabolites: APAP, APAP-S, APAP-G, APAP-GSH, APAP-C and APAP-NAC were quantified. Linearity across APAP measurement reported a $r^2$ superior to 0.990 for all the compounds and QCs monitored for each urine, plasma and liver were within the acceptance criteria range (Tables V.S3, V.S4 and V.S5 respectively).

The APAP hepatic metabolic profile is shown at 1 h (Figure 5-4.a) and 6 h following APAP treatment (Figure 5-4.b). The most abundant APAP metabolites in liver were APAP-GSH and APAP-C. Pair-wise comparison between the groups at 6 h show significant differences between control and AB animals in hepatic APAP and APAP-GSH concentrations. No significant differences were observed in APAP metabolism between HT and control groups at any of the time points. In the HT group, one animal which was euthanized in the 1 h group consistently showed higher concentration of APAP and APAP metabolites compared to the other animals.

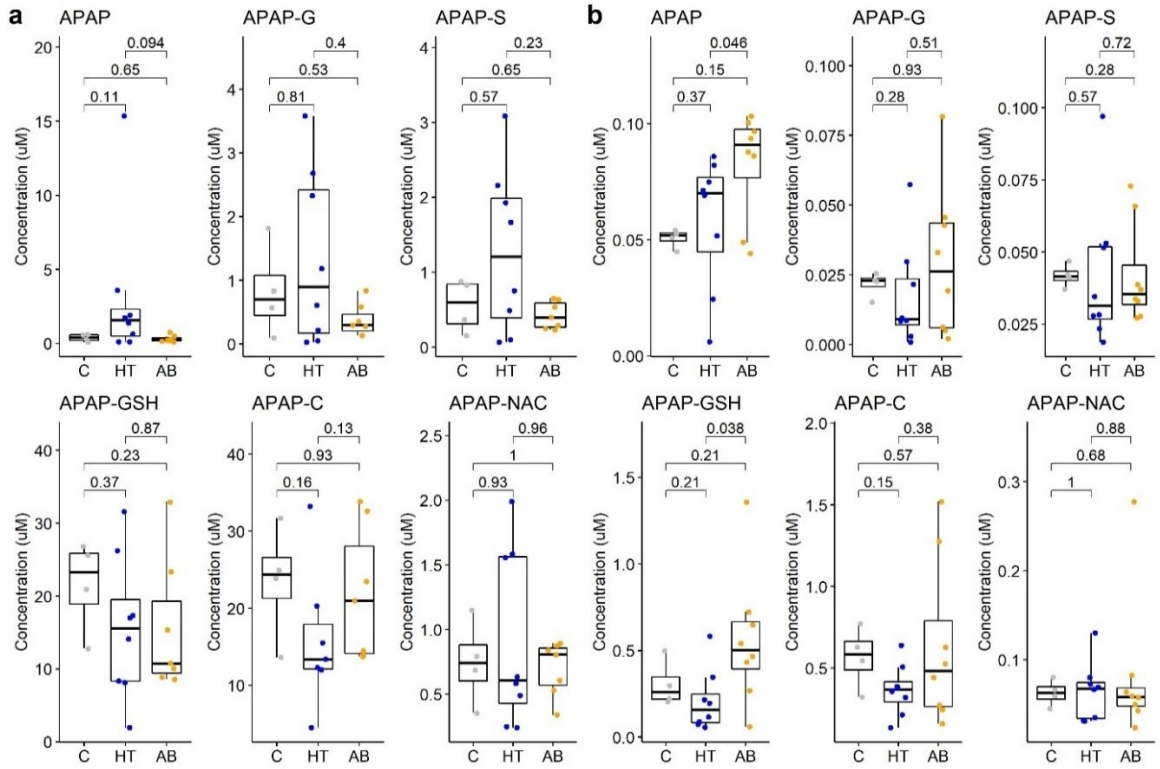



**Figure 5-4. Boxplots of the APAP metabolites measured in the liver samples at a) 1h following APAP administration and b) 6h following APAP administration.** Diet-related groups are indicated on the x axis. Whiskers in the box plots represent the full range quantiles and boxplots the 25, 50 and 75 % quantiles. Individual animals are indicated by a dot in the plot. Statistical tests within each group were done using Wilcoxon test. C: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

APAP plasma metabolic profiles are illustrated at 1 h (Figure 5-5.a) and 6 h following APAP treatment (Figure 5-5.b). The parent drug, as well as APAP-G were the most concentrated APAP-related compounds in plasma. Low concentrations of APAP-GSH were noted. A significant difference was observed between the control and the HT group for APAP-C concentrations at 6 h, being higher in the control group. No significant differences were observed in the pair-wise comparisons between the groups. Once again, one animal from the HT group euthanized at 1 h following APAP administration showed higher concentrations across all the metabolites.

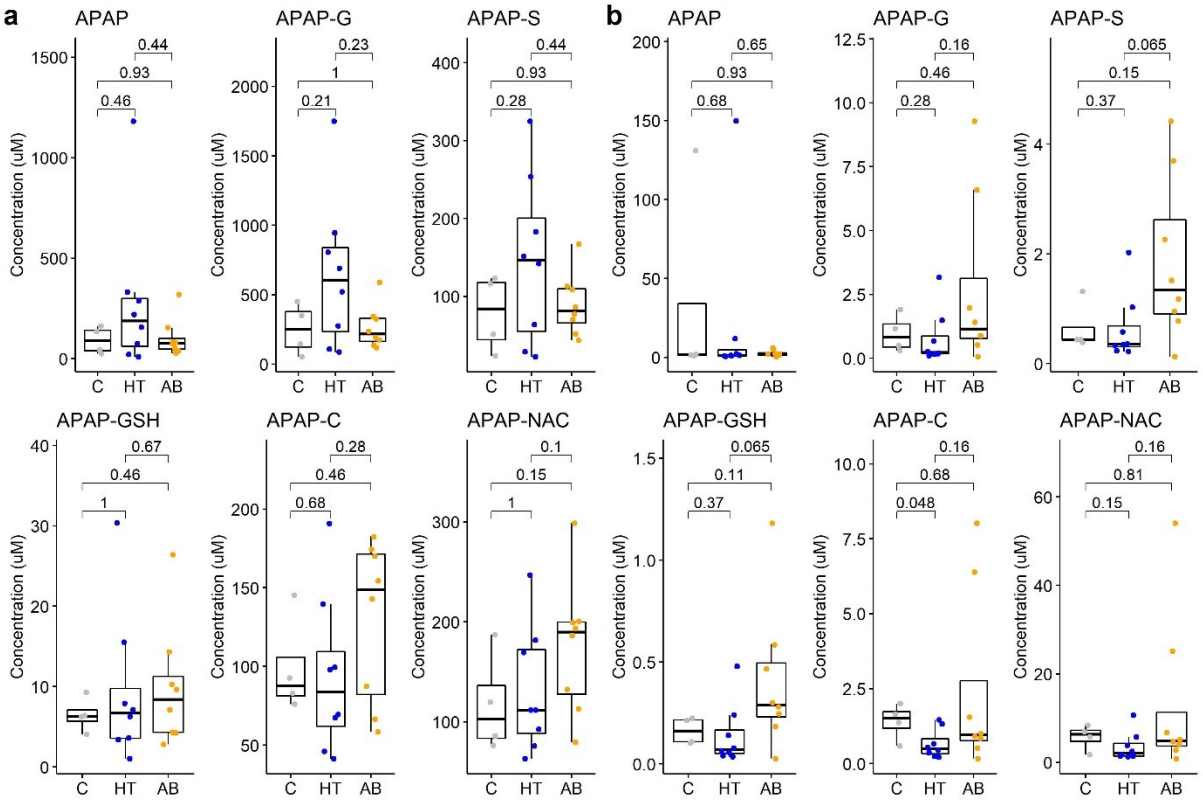



**Figure 5-5. Boxplots of the APAP metabolites measured in the plasma samples at a) 1 h following APAP administration and b) 6 h following APAP administration.** Diet-related groups are indicated on the x axis. Whiskers in the box plots represent the full range quantiles and boxplots the 25, 50 and 75 % quantiles. Individual animals are indicated by a dot in the plot. Statistical tests within each group were done using Wilcoxon test.: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

The concentration ranges of APAP metabolites show that the majority of the drug was removed from the liver and plasma at 6 h compared to the 1 h samples (Figure 5-4 and Figure 5-5). However, similar urinary concentrations of APAP metabolites were observed between the 1 (Figure 5-6.a) and 6 h time points (Figure 5-6.b). The concentration range was also much higher for the APAP metabolites in urine compared to both liver and plasma, with the exception of APAP-GSH, which was present at higher concentrations in the liver and plasma compared to the urine. APAP-G was the most concentrated of the APAP metabolites in urine. Finally, the same animal from the HT group euthanized at 1 h which showed higher concentrations of APAP metabolites in the liver and plasma samples also demonstrated higher concentrations in urine. Another HT animal, but this time euthanized at 6 h, also showed a higher concentration of APAP metabolites in urine samples.

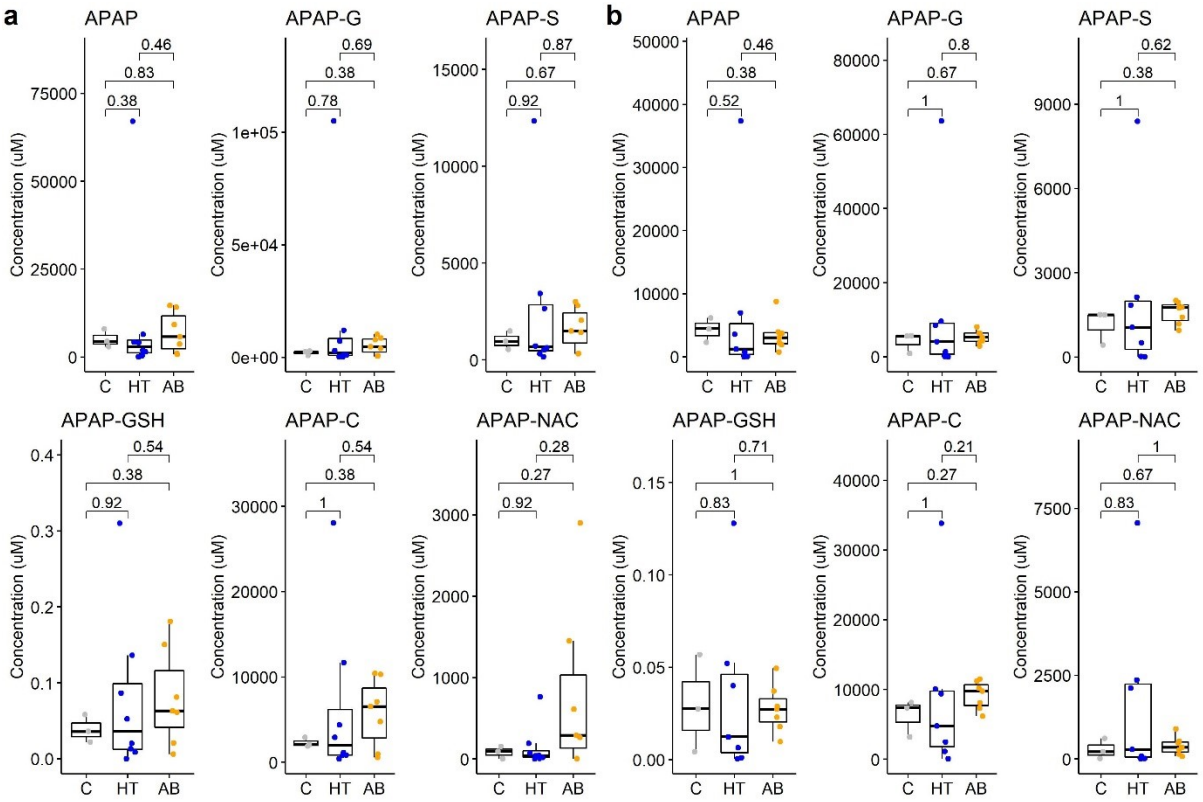



**Figure 5-6. Boxplots of the APAP metabolites measured in the urine samples at a) 1 h following APAP administration and b) 6 h following APAP administration.** Diet-related groups are shown on the x axis. Whiskers in the box plots represent the full range quantiles and boxplots the 25, 50 and 75 % quantiles. Individual animals are indicated by a dot in the plot. Statistical tests within each group were done using Wilcoxon test. C: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

### 3.3. Effect of APAP intake on tyrosine metabolism in the urine, plasma and liver

To investigate how tyrosine metabolism was modified by APAP intake, the tyrosine-related metabolites were studied at 1 h and 6 h post APAP treatment in the liver (Figure 5-7.a), plasma (Figure 5-7.b) and urine samples (Figure 5-7.c). The fit-for-purpose validated tyrosine targeted method was thus applied to liver extracts. The CV % of the QC over the run is reported in Table 5-S6. In the liver, only tyrosine was above the limit of detection and a significant decrease was seen in the AB group at 6 h compared to the 1 h time point. In plasma, tyrosine increased in the HT and AB groups compared to the control group at both 1 and 6 h post APAP dosing. For *p*HB and *p*HPL, significantly greater amounts were observed in the plasma of HT diet fed mice compared to those on the control diet at 6 h. However, this significant increase was observed at 1 h when comparing the control and the AB group. Indeed, the three metabolites detected in plasma (tyrosine, *p*HB and *p*HPL) were higher in the HT group at 6 h compared to 1 h. Similar to APAP metabolites, urine samples had a higher concentration of tyrosine metabolites compared to plasma and liver samples. As for plasma, significant increase of most of the metabolites (tyrosine, tyramine, *p*HPA and *p*HPL) was observed at the 6 h time point when comparing the control group with both HT and AB groups. However, no significant difference was observed between the HT and the AB groups for those metabolites. As observed prior to APAP treatment, urinary *p*CS and *p*CG were supressed in the antibiotic treated group, and this observation remained at 1 and 6 h following APAP treatment.

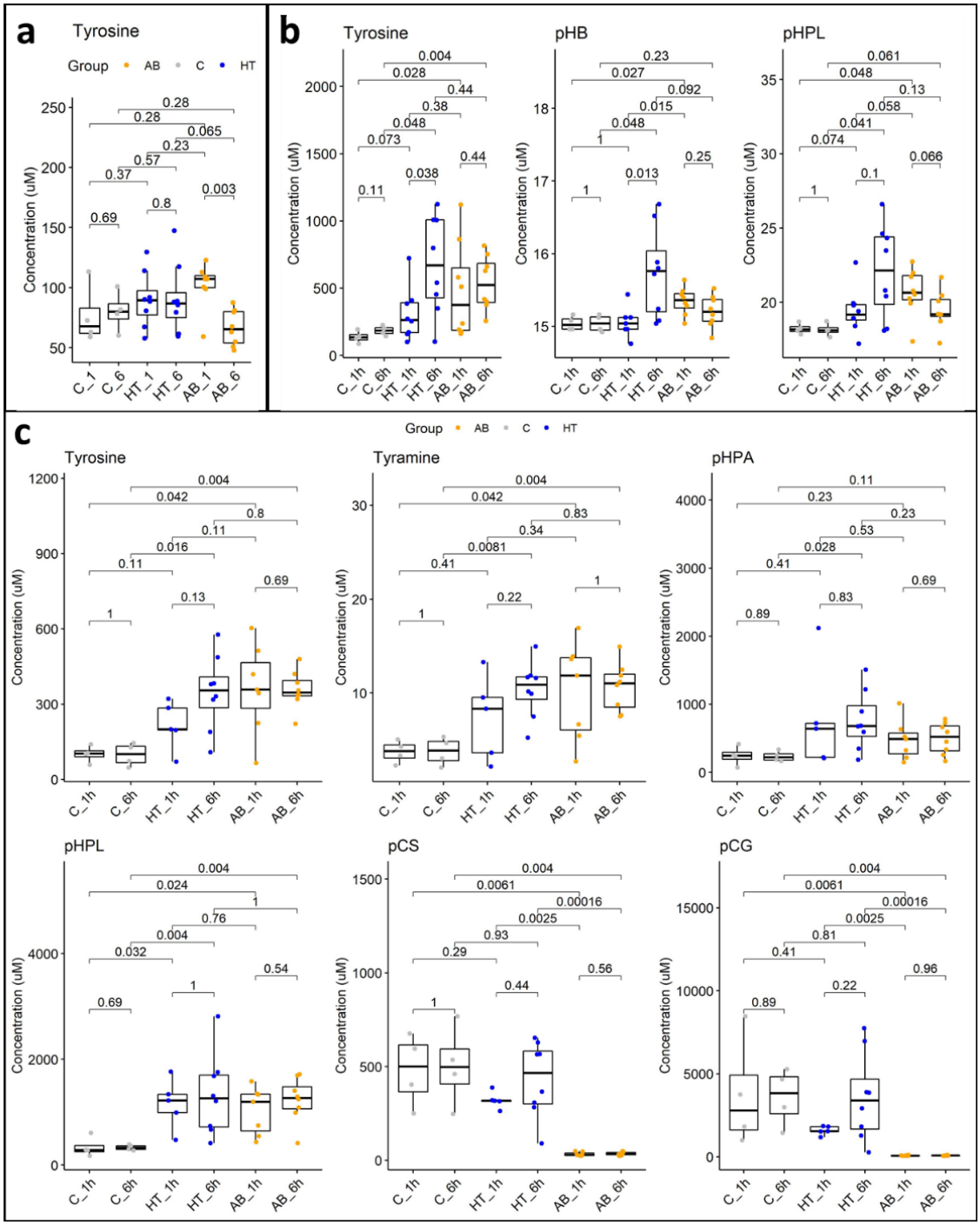



**Figure 5-7. Boxplots of the tyrosine metabolites measured in the a) liver samples b) plasma samples and c) urine samples, at 1 h and 6 h following APAP administration.** Whiskers in the box plots represent the full range quantiles and boxplots the 25, 50 and 75 % quantiles. Individual animals are indicated by a dot in the plot. Statistical tests within each group were done using Wilcoxon test. C: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

## 3.4. Effect of high tyrosine diet and APAP administration on the hepatic metabolic profile

The sulfate to glucuronide ratio was investigated to explore the difference in conjugation reactions between the different treatment groups for both APAP and *p*-cresol (Figure 5-8.a). S/G ratios were below 1 for both compounds, showing that a higher proportion of APAP and *p*-cresol was glucuronidated than sulfated. As inter-animal variability was observed in the concentration of both tyrosine and APAP metabolites, the APAP-S/APAP-G ratio was plotted per animal and categorised per group, as was the concentration of *p*CS prior to APAP-treatment (Figure 5-8.b). No correlation was found between the APAP-S/APAP-G ratio and the *p*CS pre-APAP treatment concentration for the HT and the AB group. However, an increase in *p*CS pre-APAP concentration is inversely associated to APAP-S/APAP-G ratio in the control group. AB treatment in Figure 5-8.b. shows a reduced concentration of *p*CS compared to the control and HT groups. However, a significant increase was observed in the *p*CS/*p*CG ratio when comparing the AB group to both the control and HT group, at both 1 and 6 h time points. This indicates an increase in the proportion of *p*-cresol sulfation compared to glucuronidation in the AB group, probably due to lower amount of *p*-cresol requesting sulfation and thus causing no depletion of sulfation capacity.

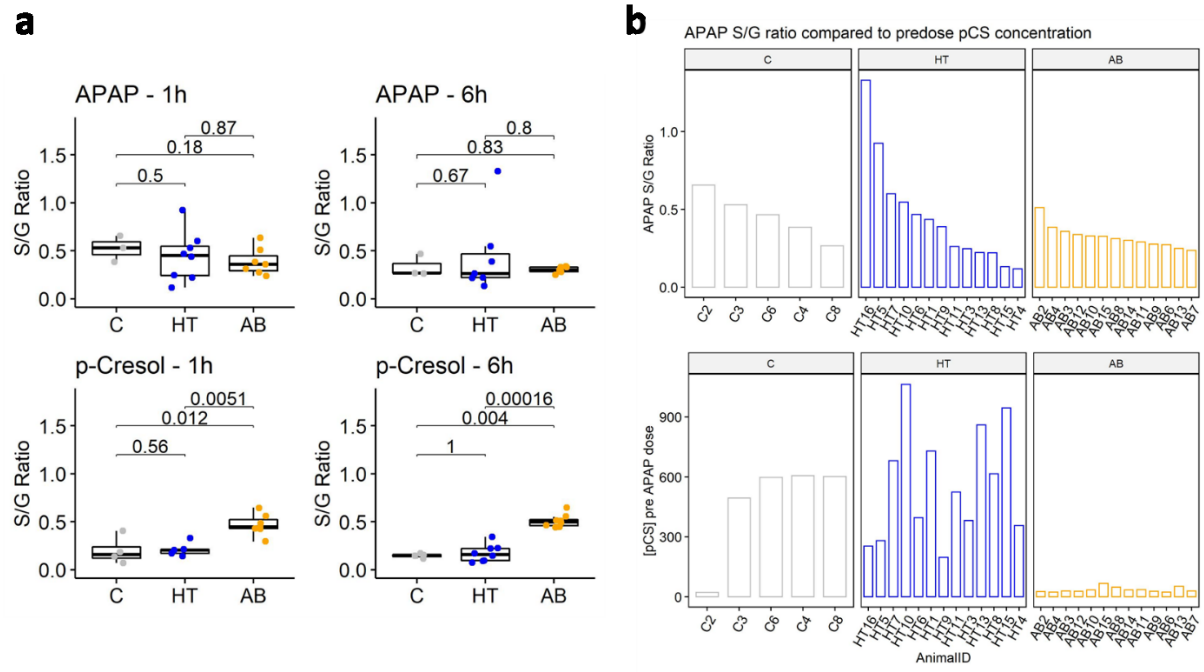



**Figure 5-8. Investigation of pCS concentration impact on Sulfate/Glucruonide ratio. a) APAP and p-cresol sulfate to glucuronide ratio in urine samples at 1 and 6 h.** Wilcoxon test has been performed to calculate p-value. b) Inter-animal variability of APAP-S/APAP-G ratio (top panel) compared to the urinary concentration of pCS in the pre-APAP samples (bottom panel). C: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

## 3.5. Multi-blocks integration of the targeted UPLC-MS data

Multi-block integration was performed on the tyrosine and APAP metabolites to determine if the combination of both the diet and treatment were strong enough to separate the samples by their treatment groups during multivariate analysis. Although no separation was observed on the Sparse PLS-DA (sPLS-DA) due to APAP treatment (data not shown), a specific sample distribution due to the diet was observed when combining the tyrosine-related metabolic profiles from the urine and plasma samples at 2 weeks, for the urine samples (Figure 5-9.a). Correlation between the metabolites measured in the urine and in the plasma samples were made through a circos plot (Figure 5-9.b). Interestingly, *p*CS and *p*CG were only positively correlated with *p*HB in plasma. These two correlations were found to be associated only with the second component of the sPLS-DA, along which AB urine samples separated from the HT and control urine samples.

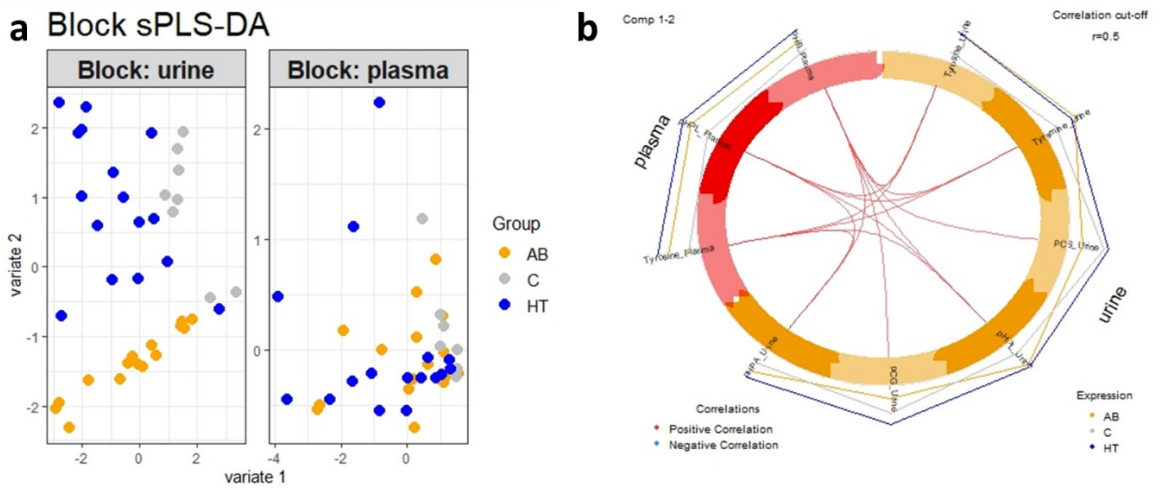



**Figure 5-9. Integration of the urine and plasma tyrosine metabolic profile to perform a a) sparse PLS-DA model and b) a circos plot, highlighting correlations between tyrosine metabolites.** The metabolites detected in the plasma are shown in red and tyrosine metabolites measured in the urine are shown in yellow. The correlation cutoff was set at 0.5 and the plot shows the co-relation between each metabolite within and among all sample groups (red for positive correlation and blue for negative correlation). The outside of the circus plot represents the mean relative abundance of each metabolites according the groups. C: control in grey; HT: High tyrosine in blue; AB: High tyrosine plus antibiotic treatment in orange.

## 3.6. $^{1}$H-NMR spectroscopic analysis of liver aqueous extracts

To investigate the wider hepatic biochemical changes induced by tyrosine and APAP intake, untargeted $^{1}$H NMR spectroscopy-based metabolic profiling was performed on liver tissue collected from groups 1 h and 6 h post APAP-treatment. Pair-wise principal component analysis (PCA) models were built to identify metabolic changes in the hepatic profiles at 1 h (Figure 5-10.a, b and c) and 6 h (Figure 5-10.d, e and f). From the scores plots obtained from these PCA models, no clear metabolic variation was observed between any of the groups. The absence of metabolic variation between the different groups was confirmed with the construction of supervised orthogonal projection to latent structures-discriminant analysis (OPLSDA) loading plots (Figure 5-S7).

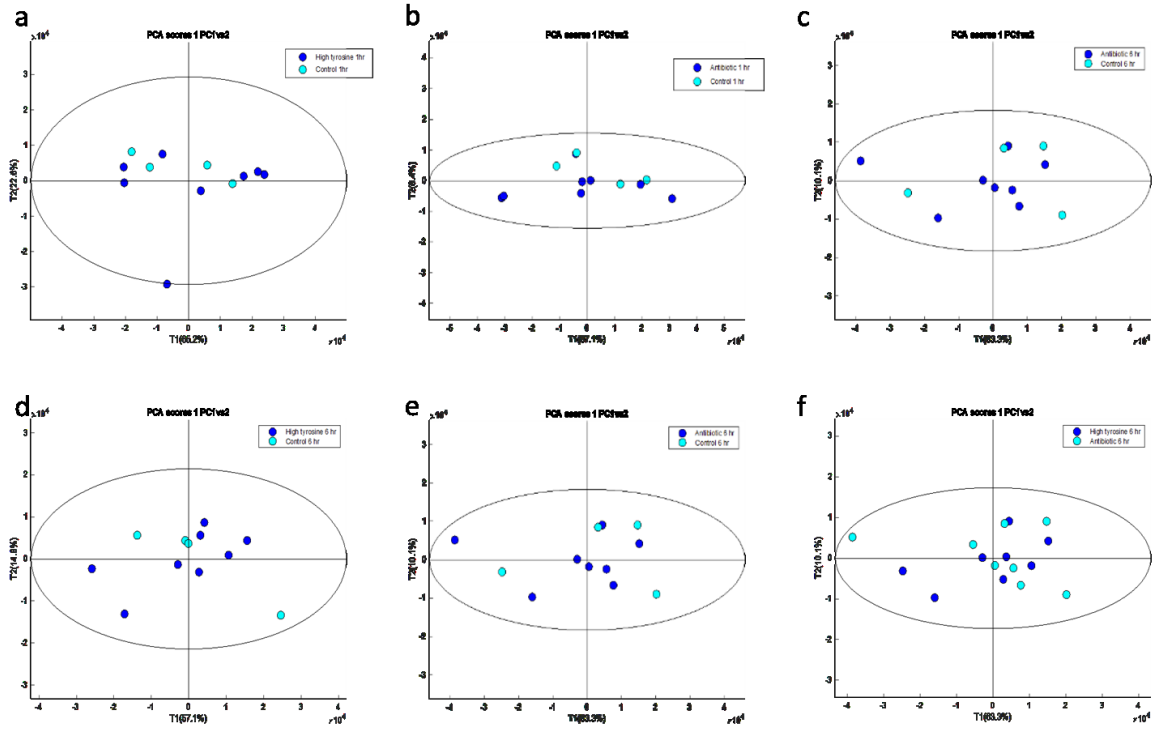



**Figure 5-10. Scores plots from pairwise PCA models comparing the hepatic metabolic profiles of the different treatment groups 1 h and 6 h post APAP treatment.** a. control 1 h vs HT 1 h b. control 1 h vs AB 1 h c. HT 1 h vs AB 1 h d. control 6 h vs HT 6 h e. control 6 h vs AB 6 h f. HT 6 h vs AB 6 h.

# 4. Discussion

Even though APAP metabolism has been extensively studied for many years, the influence of gut microbial metabolites or host-microbial co-metabolites on APAP metabolism remains an underexplored area of interest. It was demonstrated in 1974 that incubating APAP with caecal contents from rats resulted in the direct metabolism of APAP by the microbiota to *p*-aminophenol through *N*-deacetylation (Smith and Griffiths, 1974). However, it has only recently been shown that the gut microbiota can influence the conjugative metabolism of APAP through the production of *p*-cresol (Clayton et al., 2009). Here, the potential to manipulate the microbial production of *p*-cresol in mice using a high tyrosine diet was investigated. To do so, a female BALB/c mouse model was chosen due to previous reports demonstrating that its intestinal microbiota can overproduce *p*-cresol when provided with a high tyrosine diet (Shiba et al., 2014). An antibiotic treated group was also included to attenuate the metabolic activity of the microbiota and evaluate the importance of microbial metabolism in producing *p*-cresol following a high tyrosine diet. Finally, whether this manipulation could modulate host sulfation capacity and APAP metabolism was determined by using a sub-toxic dose (300 mg/kg) of APAP. A sub-toxic dose was selected to avoid hepatotoxic formation involving centrilobular necrosis (Kyriakides et al., 2016).

## 4.1. Antibiotic treatment

To assess if tyrosine metabolites were either produced by host or microbiome-mediated enzymes, antibiotic treatment was used in conjunction with the high tyrosine diet. A reduction in *p*CS and *p*CG was observed after antibiotic treatment compared to the HT group and was still observed 1 h and 6 h following APAP treatment. This confirms the involvement of the microbiota in the production of these microbial-host co-metabolites. Moreover, AB treated mice excreted different tyrosine-related metabolites in the urine compared to the control and HT group. AB treatment appeared thus to have the greatest impact on tyrosine metabolism in the mice, compared to the other groups. However, the urinary excretion of tyramine and its metabolite *p*HPA, a *p*-cresol precursor, was observed to be increased in both the HT and AB groups compared to the control group. This indicates that the metabolism of tyrosine to *p*HPA is performed by the host. Similarly, *p*HPL and *p*HB showed the same plasma concentrations in the AB group as the HT group indicating that both *p*HPL and *p*HB are also produced by host-mediated metabolic reactions. Interestingly, the reduction of *p*CS and *p*CG in the AB group result in a significantly higher *p*CS/*p*CG ratio (Figure 5-7.a). This suggests that the *p*-cresol is preferentially sulfated and as lower amounts of *p*-cresol is produced in the AB treated mice the sulfation capacity is not depleted, compared to the HT and the control groups, and thus an increase in *p*CS production is seen compared to the production of *p*CG, leading to higher *p*CS/*p*CG ratio.

## 4.2. High-tyrosine diet

Naturally, higher urinary excretion of tyrosine in the HT and AB groups were observed compared to the control animals, as a result of the high tyrosine diet. This indicates that not all the ingested tyrosine was metabolized by the host or the microbiota. A similar increase in tyrosine concentration was noticed in plasma samples of the HT group indicating that tyrosine was present in the circulation before being excreted in the urine or bile. In addition, this may suggest an increase in gut membrane permeability, as it has already been found to be a consequence of the higher presence of phenolic compounds due to lipid saturation (Heipieper et al., 1991; Keweloh et al., 1991).

Concentrations of *p*CS and *p*CG in the HT group were similar to those of the control group. This suggests that the bacterial enzymes responsible for producing *p*-cresol may have been saturated by the HT diet and higher amounts of *p*-cresol production is not immediately possible. This is in accordance with the increased urinary excretion of tyrosine. This indicates that a HT diet may need to be provided for a period superior to 2 weeks to promote the growth of bacteria that are able to utilise tyrosine and produce *p*-cresol. This may indicate that acute dietary changes in humans do not lead to immediate differences in *p*-cresol exposure although further work is necessary to validate this. In plasma, high concentrations of *p*HPL and *p*HB were observed in the HT and AB groups compared to the control group, showing that the metabolism of tyrosine also followed other degradation routes (Figure 2-10 in Chapter 2) rather than solely being degraded sequentially to tyramine, *p*HPA, *p*-cresol.

## 4.3. APAP treatment

The potential for changes in bacterial *p*-cresol production to alter the metabolism of APAP was then explored in these mice. Interestingly, in urine there was a significant increase in the concentration of all APAP metabolites when compared to plasma and liver, which is our primary site of interest as major APAP conjugates are excreted through urine. Metabolism of APAP across the different matrices was consistent with common drug metabolism routes, with APAP-GSH the most abundant APAP metabolite in the liver, APAP and APAP-G in plasma and APAP-G in urine.

In plasma, significantly less APAP-GSH in the plasma of the AB animals could be explained by the absence of *p*CS and *p*CG in these animals. By suppressing the microbial enzymes that produce *p*-cresol, the sulfate and glucuronic acid pool will not be depleted. If these pools are not depleted, APAP will be conjugated to APAP-G and APAP-S at a higher rate and this will result in a lower proportion of APAP undergoing glutathione (GSH) conjugation. This could indicate that modulating the gut microbiota through AB treatment and thus decreasing production of *p*-cresol could serve to prevent sulfation depletion and drug toxicity.

In the HT group, urinary APAP-G excretion was higher compared to APAP-S, suggesting a higher rate of glucuronidation compared to sulfation. This was confirmed by plotting the S/G ratio of both *p*-cresol and APAP (Figure 5-7.a), which both demonstrated ratio values inferior to one (indicating a higher proportion of glucuronide-conjugate compared to sulfate-conjugate). Interestingly, inverse relationship was observed between the APAP-S/APAP-G ratio and the pre-treatment *p*CS concentration in the control group, which is in accordance with a previous study in humans (Clayton et al., 2009). However, this inverse association was not observed in the HT group, likely due to the saturation of conjugation enzymes in the HT animals, as described above. Similarly, no significant difference was observed in the *p*CS/*p*CG ratio between the HT and the control groups. As previously discussed, *p*CS/*p*CG ratio was found to significantly increase in the AB group compared to the HT and control groups, due to lower level of p-cresol produced, which allow the sulfation capacity to be preserved compared to the HT and control group and this increase *p*CS/*p*CG ratio. The clinical significance of this finding may be that phase II drug metabolism could be different in individuals consuming large and stable amounts of tyrosine over a long period of time (*e.g.* from a high meat diet) compared to individuals consuming lower amounts of tyrosine (*e.g.* from a plant based or vegan diet). However, this observation was not confirmed by the results reported herein. Similar *p*CS concentrations were observed in the HT group compared to the control group, and high inter-animal variability was observed in the HT group. Saturation of the enzymes responsible for *p*-cresol production could explain this observation and thus longer exposure to high-tyrosine diet might be needed to entrain the microbiome onto this substrate and see marked increases in *p*-cresol production.

The fact that this was not observed in the HT diet is related to inter-animal variation, which was noticed in the concentrations of most of the metabolites. Indeed, *p*CS concentration was fairly uniform within the control group as well as within the AB group but presented huge inter-variability in the HT animals, potentially due to variability in the microbiota of the animals and its adaptation to the HT diet. The results were consistent throughout as the same animals showed consistently high metabolite concentrations compared to other animals. These higher concentration values were not considered as outliers in this study as this could have a deeper biological implication, most likely involving the difference in microbiota composition among animals and their ability to adapt to the HT diet over a short period of 2 weeks.

# 5. Conclusion

The HT diet did not produce a model with increased *p*CS and *p*CG excretion as expected and previously reported (Shiba et al., 2014). This could be explained by the 2014 study providing the control animals with a tyrosine-free diet (0 % against 0.98 % in the study reported herein). However, as total removal of tyrosine from the diet would not represent a realistic dietary condition and could induce a metabolic deficiency, potentially adding a bias in the experiment, this option was not considered. Administration of the HT diet through a longer period of time than two weeks might allow the microbiota to adapt further and allow a higher production of *p*CS and/or *p*CG. This study still provides an insight into the competition between APAP conjugation metabolism and *p*CS when looking at the control diet group. This allows the previous hypothesis by (Clayton et al., 2009) to be confirmed suggesting a higher concentration of *p*-cresol would promote APAP glucuronidation over APAP sulfation, meaning that *p*-cresol depletes the host sulfate pool. Our results suggest that when the concentration of *p*-cresol is low or "normal", *p*CS can be used as a sensitive marker of APAP metabolism and toxicity. However, this marker is lost when the system producing *p*-cresol and its conjugates is saturated by high amounts of dietary tyrosine. This study has a major clinical significance pertaining to the threat posed by incomplete APAP conjugative metabolism leading to the production of toxic APAP metabolites in liver eventually causing hepatotoxicity. This study could be further extended to study gut tissues and gut contents, as well as gall bladders, kidneys, muscles and brains, as APAP and *p*-cresol metabolism also involves a complex inter-organ communication as the metabolites are drained into circulation or bile for further distribution or elimination. Amplicon sequencing of the faecal contents could also be performed to explore (1) the impact that APAP has on the gut microbiota and (2) perform correlative study between microbial profiles and *p*CS production, which could potentially explain the extensive inter-animal variability by an increase presence of specific bacteria.

# Chapter 6. Investigating the metabolic perturbations induced by the selective inhibition of bacterial β-glucuronidases, enzymes known to influence drug toxicity

**Aim:**

- Explore the impact of selectively inhibiting bacterial β-glucuronidases in the gut on the host metabolic phenotype.

**Objectives:**

- $^{1}$H NMR spectroscopy and UPLC-MS will be used to measure the metabolic profiles of the liver, gall bladder, and colon, as well as colon contents, faeces, plasma, and urine of mice treated with a targeted bacterial β-glucuronidase inhibitor and compared to those of control animals.

- Assess the possibility of using this inhibitor to selectively manipulate the activity of the gut microbiome to limit the toxicity induced by bacterial β-glucuronidases during irinotecan chemotherapy.

# 1. Introduction

Irinotecan, also called CPT-11, is a common chemotherapeutic treatment for colorectal cancer, and it is also part of the first-line regimen for pancreas cancer. It is a prodrug which is activated by the host carboxylesterase in the systemic circulation to its active form, SN-38. This active form targets topoisomerase I which regulates DNA synthesis in tumour cells. When the active form of irinotecan reaches the liver, it is conjugated with a glucuronide group by a UDP-glucuronosyltransferase which inactivates the drug and facilitates its elimination through the bile (Gupta et al., 1994). Once transported into the GI tract the glucuronide group on the inactivated drug can be hydrolysed by bacterial β-glucuronidase enzymes liberating the active form of the drug (SN-38) in the gut (Figure 6-1). The release of SN-38 into the GI tract causes severe intestinal toxicity (Tobin et al., 2003; Stein et al., 2010), such as diarrhoea and often results in the discontinuation of the treatment (Mathijssen et al., 2001; Ma and McLeod, 2003).

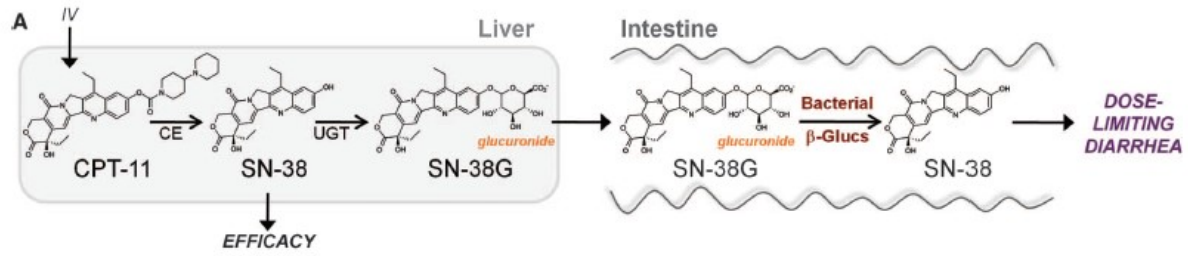



**Figure 6-1. The metabolism of irinotecan (CPT-11).** It is first activated into the active metabolite SN-38 followed by its conjugation with UGT producing SN38-glucuronic acid (SN-38G). SN-38G reaches the intestines where bacterial β-glucuronidase cleaves the glucuronic acid, releasing the aglycone SN-38 into the GI tract and causing severe adverse drug reactions **(Wallace et al., 2010)**.

Cleavage of the glucuronide group provides the bacteria with a source of carbon and this is a clear example of how microbial enzymatic activity can impact on host drug metabolism and dramatically increase the toxicity of irinotecan. Given the key role of the gut microbiota in potentiating this toxicity, with implications for treatment strategies, manipulation of this microbial metabolism may enable this toxicity to be attenuated allowing higher doses to be administered and the efficacy to be improved. Inhibition of bacterial β-glucuronidases has been explored using antibiotic treatment, which blocks bacterial β-glucuronidase activity but also non-selectively transforms the overall structure and diversity of the gut microbiota, which may have broad implications for the immune system, barrier against pathogens (including nosocomial infections) and loss of beneficial metabolites arising from the microbiome. The potential effects on the immune system of the patient are an area of concern for patients with colorectal cancer. This lack of selectivity led Wallace et al., 2010, to screen a chemical library to find an inhibitor specific to this bacterial enzyme. Four candidate compounds were found to potentially inhibit the bacterial enzyme and Inhibitor 1 has been characterised as the most potent and selective inhibitor after molecular modelling (Figure 6-2.C).

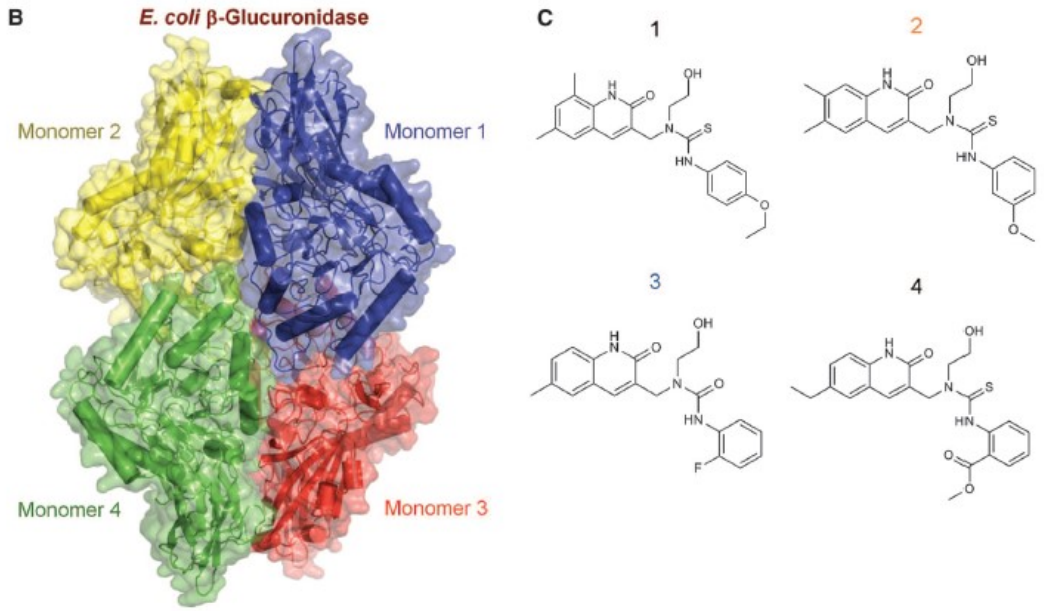



**Figure 6-2. Results of the high-throughput screening of a chemical library to underline compounds able to inhibit bacterial β-glucuronidase enzyme.** (B) Crystal structure of the β-glucuronidase encoded by *Escherichia coli*. (C) Inhibitors 1 to 4 which were found to successfully inhibit bacterial β-glucuronidase follwing high-throughput screening **(Wallace et al., 2010)**.

In mice this inhibitor was found to significantly limit adverse drug reactions induced by irinotecan (Figure 6-3) and also to prevent changes in the abundance of *Enterobacteriacea* caused by this compound (Bhatt et al., 2020). Also, no variation has been observed in the PK profile of irinotecan and its metabolites when co-administrated with the β-glucuronidase inhibitor (Wallace et al., 2015). This inhibitor was later found to be more efficient for Firmicutes and Proteobacteria β-glucuronidases and ineffective for β-glucuronidases from *Bacteroidetes fragilis* (Wallace et al., 2015). This highlights diversity in the β-glucuronidases across different species. Bacterial β-glucuronidase enzymes also impact other xenobiotics, such as non-steroidal anti-inflammatory drugs (NSAIDs- LoGuidice et al., 2012) and heterocyclic aromatic amines (HAs- Humblot et al., 2007). Targeted inhibition of bacterial β-glucuronidases is an attractive approach to limit adverse drug reactions induced by these enzymes during the course of several therapeutic treatments whilst preserving the overall structure of the gut microbiome.

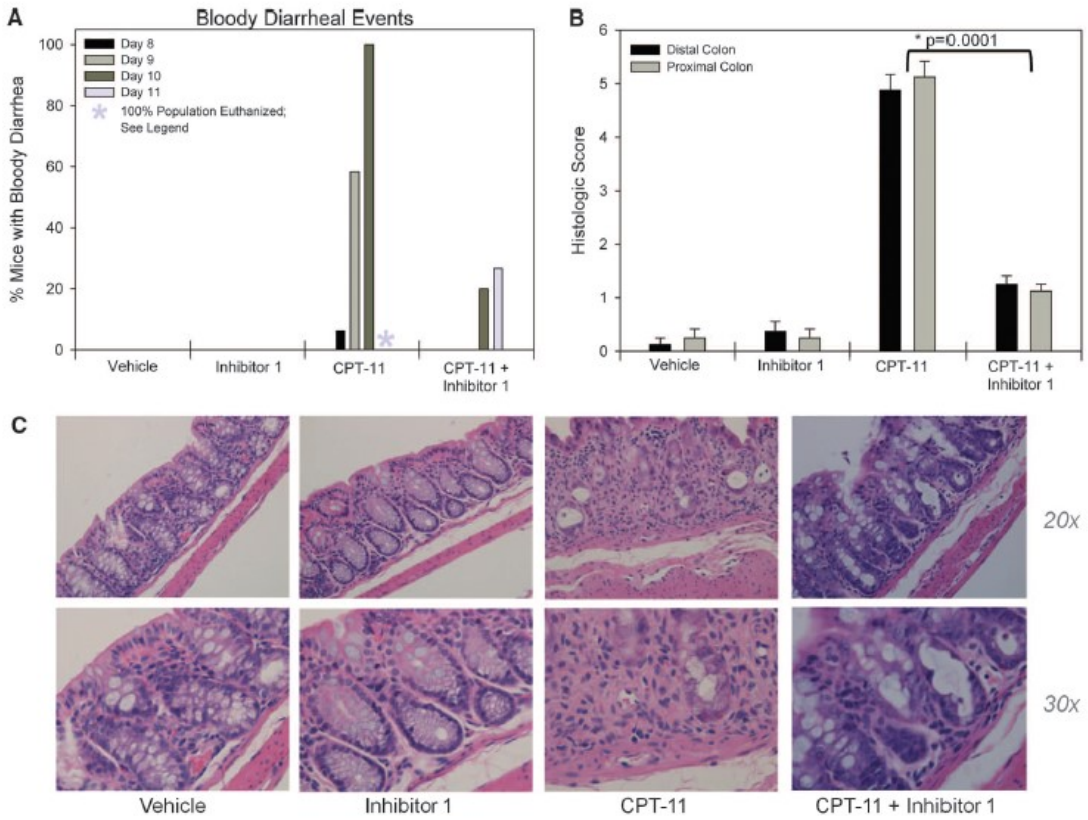



**Figure 6-3. Decreased toxicity of irinotecan (CPT-11) in the presence of the first inhibitor of β-glucuronidase (Inhibitor 1).** (A) Bloody diarrhoea percentage in vehicle mice and inhibitor 1 mice (0 %), in CPT-11 treated mice (starting from ~ 5 % at day 8 following the treatment and reaching 100% at day 10) and in CPT-11 + inhibitor 1 treated mice (< 25 %). (B) Histologic score of distal and proximal colon, showing a significant difference between the mice treated only with CTP-11 and the one treated with CPT-11 + inhibitor 1. (C) Colon histology showing that the co-administration of inhibitor 1 with CPT-11 protect the damage that can be induced by administration of CPT-11 on its own, as the structure of the colon tissue in the co-administrated mice is more similar to the one observed in the vehicle or inhibitor 1 treated mice **(Wallace et al., 2010)**.

While inhibitors of β-glucuronidase have been demonstrated in mice to prevent ADRs from CPT-11 (Wallace et al., 2010), there is currently limited understanding of the influence of these bacterial enzymes on wider host metabolism. This work aims to understand the downstream impact of inhibiting bacterial β-glucuronidases on host metabolic pathways. Following previous research from Redinbo and co-workers, the scaffold of the inhibitor of β-glucuronidase has been improved and the optimised inhibitor is termed Inhibitor 9 (Inh-9, Figure 6-4; Pellock et al., 2018). To determine the effect of the Inh-9 on the metabolic profile of the murine host it has been administered without irinotecan, which can induce further metabolic changes and make the understanding more complex. Three oral doses of 20 μg of Inh-9 were administered to mice at 0, 24, 48 h (*n* = 8/group) and compared to a group of seven control mice that did not receive the inhibitor. Urine and faecal samples were collected from each animal pre-dose and at 24 and 48 h. After a wash-out period of 48 h, the mice were euthanized 96 hours after the start of the study. At necropsy, plasma, colon, colon contents, liver and gall bladder were collected. On these samples, $^{1}$H NMR spectroscopy and/or UPLC-MS analysis were performed.

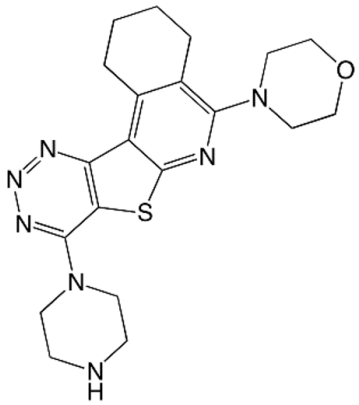



**Figure 6-4. Structure of the inhibitor 9 (Inh-9), which inhibits the bacterial β-glucuronidase.**

# 2. Experimental Section

## 2.1. Chemicals

Dimethyl sulfoxide (DMSO) was purchased from Hybri-Max (US). Analytical reagent grade formic acid and optima LC-MS grade water with and without 0.1% formic acid were obtained from Fisher Chemical (Loughborough, UK). Acetonitrile (ACN), methanol (MeOH) and isopropanol (IPA) were purchased from Honeywell (Seelze, Germany). For UPLC-MS mass accuracy and calibration, leucine enkephalin acetate salt hydrate and sodium formate, as well as sodium azide for faecal sample preparation were obtained from Sigma-Aldrich (Gillingham, U.K.). The β-glucuronidase inhibitor Inh-9 was provided by Professor Matthew Redinbo from the University of Carolina at Chapel Hill. The molar mass of the compound is 447.98 g/mol.

## **2.2. Study design**

Fifteen C57/BL6 mice were weighed, and pre-treatment urine and faecal samples were collected before eight of the mice received their first dose of 20 µg of Inh-9 per oral gavage (dissolved in DMSO at a concentration of 2 mg/mL prior to being diluted with sterile 0.9 % saline and pre-warmed). Control mice received a vehicle solution (sterile 0.9 % saline solution). The distribution of the animals in the cage was as follows: a first cohort was composed of two cages (two females in the first one, 10 weeks old, treated with the vehicle solution; two females in the second cage, 10 weeks old, treated with Inh-9); a second cohort was composed of 6 cages (one with two females, 10 weeks old, treated with the vehicle solution; one with two females, 10 weeks old, treated with Inh-9 solution; one with one male, 8.5 weeks old, treated with the vehicle solution; one with two males, 8.5 weeks old, treated with Inh-9; the next with two males, 10 weeks old, treated with the vehicle solution; and finally one with two males of 10 weeks old, treated with Inh-9). When available, urine and faecal samples were then collected at 24 h and 48 h, just prior to the second (at 24 h) and third (at 48 h) administration of the inhibitor or saline solution. After a wash-out period of 48h after the last treatment, urine and faeces as well as plasma, liver, colon tissues, colon contents and gall bladders were collected at 96 h. Collection of the samples was made in the same order as the animals were treated.

## 2.3. $^{1}$H NMR spectroscopy sample preparation

Urine samples were thawed, vortexed and centrifuged at 10,621 *g* for 10 minutes at 4˚C, before being diluted (1:1) with 30 µL urine buffer (1.5 M $KH_2PO_4$, 2 mM $NaN_3$, 1 % TSP solution, pH = 7.4) and mixed (Dona et al., 2014). The samples were centrifuged at 20,783 *g* for 10 minutes at room temperature and 50 µL were transferred into a capillary NMR tube (1.7 mm). Plasma samples were thawed, vortexed and mixed in a 1:1 ratio with 100 µL of plasma buffer (0.075 M $NaH_2PO_4$, 2 mM $NaN_3$, 1 % TSP solution, pH = 7.4) (Dona et al., 2014). The samples were centrifuged for 5 minutes at 15,294 *g* at 4 °C and 180 µL was transferred into 3mm NMR tubes. Liver and colon tissues were thawed on ice. Approximately 20-30 mg of each sample was weighed and placed into special bead beating tubes. A volume of 300 µL of $CHCl_3$:MeOH (2:1) was added, as well as ~100 µL of 1 mm Zirconium beads before vortexing and two cycles of bead beating (40 s at 6,500 Hz speed- Percellys bead beater). A volume of 300 µL of $H_2O$ was added to the homogenate before centrifugation at 20,783 *g* for 10 minutes at room temperature. The organic and aqueous layers were carefully separated by pipette before the addition of 300 µL of $CHCl_3$:MeOH (2:1) and 300 µL of $H_2O$. Each sample was vortexed and centrifuged again, and each of the layers separated and combined with the corresponding previous one. Organic extracts were left under the fume hood overnight and aqueous extracts were dried using the Savant vacuum concentrator (180 min, 45°C, V-AQ mode). Dried extracts were stored at -40°C until the day of the analysis, during which the aqueous extracts were thawed, resuspended in 700 µL of $D_2O$:$H_2O$ (9:1, 1 mM TSP), vortexed and centrifuged at 6,797 *g* for 10 minutes at 4°C. A volume of 600 µL was then transferred into 5 mm NMR tube.

## 2.4. UPLC-MS sample preparation

The urine samples were prepared according a published method (Gray et al., 2016). Briefly, the samples were thawed and vortexed, before being mixed with cold MeOH (1:3), vortexed and stored overnight at -20°C. A 10 minute centrifugation at 4°C, 14,000 *g* was then performed, before a 1:10 dilution of the supernatant with $H_2O$ in 96 well-plates, which was centrifuged just before the UPLC-MS analysis. Sample were analysed in in a randomized order. If available, a biological duplicate was prepared.

Liver samples were prepared as indicated in a previous method (Want et al., 2013). Briefly, samples were thawed and a first replicate of 50.04 mg (+/-0.30 mg) was collected followed by a second replicate of 50.10 mg (+/-0.30 mg). Zirconium beads (~100 µL) and 1.2 mL of cold MeOH:$H_2O$ (1:1) were then added to the samples before homogenisation with two cycles of Percellys bead beater (40 seconds, 6,500Hz), centrifugation at 10,000 *g* (10 min, 4°C) and the supernatant (250 µL) was collected. A pooled sample was also prepared for quality control purposes. All the samples were concentrated overnight under nitrogen flow at room temperature and the dried extracts were kept at -40 °C until the day of the UPLC-MS analysis. The same protocol was applied for the colon tissue preparation, with the exception of a volume of 1.5 mL of cold MeOH: $H_2O$ (1:1) instead of 1.2 mL to 48.93+/-3.18 mg. Also, the limited quantity of the samples available did not allow the preparation of biological replicates, but technical replicates were analysed. The gall bladders were collected and frozen intact. The gall bladder metabolite extraction was then performed on the entire organ, which presented very small and variable weights (1.93 +/- 2.26 mg), by keeping the ratio 0.8 mg: 200 µL of cold MeOH: $H_2O$ (1:1). The rest of the protocol was the same as for the liver or colon tissues metabolite extraction. The liver and colon tissue extracts were resuspended in 120 µL of MeOH: H2O (1:1), while the gall bladders were resuspended into 43.5 µL, before to being vortexed, centrifuged (5 min, 20,783 *g*, room temperature), randomised and analysed.

Faeces and colon contents were prepared in a similar way as the liver tissues. Although the solvent used for the metabolite extraction was $H_20$ with 3.3 mM $NaN_3$ to prevent bacterial overgrowth, and the weight of the samples was variable with some samples providing less than 10 mg of material. The volume of the solvent used was by consequence adapted to the weight of each sample maintaining a ratio of 50 mg sample: 1.1 mL of solvent. The homogenisation was made with two cycles of Percellys bead beater (40 seconds, 6,500Hz), followed by a centrifugation at 13,000 *g* (20 min, 4°C). Supernatants were collected before to repeat the bead beating steps with another volume of $H_20$ (3.3mM $NaN_3$) to optimize metabolite extraction. After centrifugation, both supernatants were combined, and the resultant aqueous extracts were kept at -40 °C until the day of the analysis. Concentration of the samples under nitrogen flow overnight was done, before re-suspending in MeOH: $H_2O$ (1:1), in proportion to the ratio 50 mg of initial sample: 528.9 µL, vortexed, centrifuged (5 min, 20,783 *g*, room temperature) and analysed.

### 2.5. $^{1}$H NMR spectroscopic analysis

The $^{1}$H NMR analysis was performed on a Bruker 600 MHz spectrometer (Bruker Biospin, Karlsruhe, Germany) operating at 300K. The parameters of the acquisition were as previously reported for urine, serum and plasma (Dona et al., 2014). Standard 1D Noesy experiments were performed for all samples except plasma incorporating a water pre-saturation step. For the plasma samples, a Carr-Purcell-Meiboom-Gill (CPMG) experiment was performed with water pre-saturation. For the tissue samples, 64 scans were acquired after 4 dummy scans for each sample, and 4 dummy scans were followed by 32 scans for the urine and plasma samples. The acquisition was imported into Matlab (version *R2*014a, The Mathworks Inc.). The residual water region was removed (δ 4.82 ± 0.15). For the urine dataset, the urea (δ 5.92 ± 0.24) and DMSO signals (δ 3.18 ± 0.02 and δ 2.73 ± 0.02) were removed. The urine and the liver datasets were normalised using the probabilistic quotient method before manual alignment of the spectral data. The colon tissues and plasma datasets were not normalised. Principal Components Analysis (PCA) was performed for each dataset to identify any outliers. One plasma spectrum from a control mouse was identified as an outlier due to its $^{1}$H NMR spectra being off and the difficulty to align it properly to the $^{1}$H NMR spectra of the other samples and was removed from future analyses. For the urine dataset, mean-centring and Pareto scaling were applied. For the plasma, liver and colon tissues, the datasets were mean centred only. Metabolite annotation was performed by database comparison (in-house libraries and Chenomx Suite) as described in the analytical background chapter. The $-\log_{10}$(q value) of the metabolites successfully annotated by UPLC-MS were used to generate heatmaps, as explained in the analytical background chapter as well, in the aim to visually assess their variation in the mice group treated with the inhibitor compared to the control mice group.

### 2.6. UPLC-MS analysis

Analysis was performed as previously reported (Lewis et al., 2016) on a Waters Acquity I-class UPLC system (Waters Corp., Milford, MA) with separation on an HSS T3 1.8 µM column (2.1 mm i.d. x 150 mm). The temperature of the column was set at 45 °C with the autosampler temperature at 4 °C. The volume of sample injected was 1 µL. The mobile phase used was

water with 0.1 % of formic acid (solvent A) and acetonitrile with 0.1 % of formic acid (solvent B). The initial mobile phase composition was composed of 99 % of A with a flow rate of 0.6 mL/min. The proportion of solvent B was set to increase from 1 % at 0.10 min to reach 55 % at 10 min. The percentage of solvent B was then gradually increased in 0.15 min intervals of 10% before 100 % was reached at 10.70 min. This solvent composition was maintained as for 0.95 min as a wash step followed by a re-equilibration step of 2 min with 99 % of solvent A (total run time 12.65 min). Volumes of 1000 µL of weak wash solvent ($H_2O$: ACN 3:1) and of 1000 µL of strong wash solvent (IPA) were used. Fifteen QC pool samples were used at the beginning of each matrix analysis to condition the column (Gika et al., 2007; Broadhurst et al., 2018) and a QC was injected every 5 samples throughout the analysis to monitor its repeatability.

The Acquity UPLC-MS system was interfaced with the Waters Xevo G2 Q-ToF mass spectrometer (Waters Corp., Wilmslow, U.K). In positive electrospray ionization (ESI +), the capillary voltage was 1.5 kV, the source temperature was set at 120°C, the cone gas flow was 50 L/h, the desolvation gas temperature was 600 °C and the desolvation gas flow was 1000 L/h. The same parameters were applied to negative electrospray ionization (ESI -), apart from the capillary voltage which was 1 kV, and the cone gas flow which was 100 L/h. The acquisition was carried out over the *m/z* range 50-1200. Centroid mode with $MS^E$ acquisition was applied to the experiment with a RAMP collision energy going from 15 to 45 eV. Leucine encephalin (MW = 556.27 Da, 200 pg/µL in ACN:$H_2O$ 1:1) was infused for mass accuracy at a flow rate of 10 µL/mL with scan acquisition every 60 seconds. The data were collected using MassLynx V 4.1 (Waters Corp., Manchester, U.K.).

The raw datafile were firstly imported on R and converted into mzML format using ProteoWizard 3.0. The pre-processing of the data was performed with an XCMS script (Smith et al., 2006). The peak picking was performed using centWave method (parameters adapted according raw data exploration). Peak grouping was performed using the nearest method. Normalisation of the datasets was performed using median fold change normalisation and filtered based on the coefficient of variation of each variable, using a threshold of 20 % compared to the QC pool sample. The percentage of null values compared to the total number of features detected was of 0.009 % of the dataset. Investigation of the influence of the features across the dynamic range of each model was carried out following different transformation/scaling in order to know which one was the most adapted to homogenise the

influence of all the features. Mean-centring and $log_{10}$-transformation, with an offset of 20, were found to fulfil the most this requirement. Unsupervised analyses were performed using SIMCA 14.1 (Umetrics, Umea, Sweden) in order to assess the presence of outliers. Supervised analyses were performed in Matlab *R2*018a to produce covariate-adjusted projection to latent structures discriminant analysis (CA-PLS-DA) of all the samples (control *vs* Inh-9) excluding the pre-dose for urine and faeces and at 96 h for the rest of the matrices (Posma et al., 2018). Supervised models were in consequence adjusted for any bias which could have been introduced due to cohort, sex, age, cage or replicate cofounders. The goodness of fit and the predictable power of each PLS-DA model was assessed by studying the $R^2$ and $Q^2$ values (respectively), which were both set to be above 0.05 and 0.01, respectively. Wilcoxon rank sum test was used to calculate a p-value and to estimate the significance of the separation along the unique component of the models. If the p value was found to be significant ($\alpha < 0.05$) and the skyline significance plot reported features having a Q-value (P-value corrected by a multiple testing based on the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) with a false discovery rate of 5%) after a Monte-Carlo Cross-Validation (MCCV, using 100 rounds and a partitioning of 5) below 0.05, then these features were selected for metabolite identification. Features found to significantly influence the model were putatively annotated using different approaches. Initially, a correlation script was used to highlight features derived from the same metabolite (isotopes, adducts, dimers, *etc*..). Features were also matched with the *m/z* and the retention times of the compounds in the National Phenome Centre (NPC) in-house database acquired with the same method applied here (Lewis et al., 2016) to help the putative annotation of the observed features. In addition features were also matched with online databases (HMDB, Metlin and LipidMaps) using CEU Mass Mediator (Gil de la Fuente et al., 2018), by comparing the $[M+H]^+$, $[M+H-H_2O]^+$, $[M+Na]^+$ or $[M+K]^+$ adducts for ESI+ datasets and $[M-H]^-$ and $[M+Cl]^-$ and $[M-H-H2O]^-$ for ESI- datasets, as well as the main fragments. The $-log_{10}$(q value) of the metabolites successfully annotated by UPLC-MS were used to generate heatmaps, as explained in the analytical background chapter, in the aim to visually assess their variation in the mice group treated with the inhibitor compared to the control mice group. Molecular networks were also generated in MetExplore (online version 2.22.24, (Cottret et al., 2010)), based on the BioSource *Mus Musculus* iSS1393 (MetExplore ID 5677) in which some of the compounds identified had to be implemented based on their CHeBI or KEGG identifier.

## 3. Results

$^{1}$H NMR spectroscopy was performed on plasma and urine samples and tissue extracts from the colon and the liver. All the collected matrices except plasma (urine, faeces, colon contents, colon tissues, liver and gall bladders), were analysed by reversed-phase untargeted UPLC-MS in both positive and negative ESI modes (Table 6-S1).

### 3.1. Plasma samples

Unsupervised analyses performed on the $^{1}$H NMR spectroscopy dataset of the plasma samples did not show any sign of clustering by treatment group (Figure 6-S1). Supervised CA-PLS-DA analysis and Wilcoxon rank sum test used to analyse the significance of the treatment group separation along the predictive component of the models showed no significant difference ($\alpha < 0.05$) between the groups for the plasma samples measured by $^{1}$H NMR spectroscopy (Figure 6-5).

### 3.2. Liver tissue

Unsupervised analyses performed on the $^{1}$H NMR spectroscopic and UPLC-MS datasets of the liver extracts did not reveal any metabolic variation between the animals treated with the β-glucuronidase inhibitor and the animals treated with the saline solution (Figure 6-S1). Supervised CA-PLS-DA analysis and Wilcoxon rank sum test used to analyse the significance of the separation on the predictive component of the models showed no significant difference ($\alpha < 0.05$) for either the $^{1}$H NMR spectroscopy or the UPLC-MS ESI- datasets of the liver extracts (Figure 6-5). A significant difference was observed in the hepatic metabolic profiles measured by UPLC-MS ESI+ ($p = 0.002$), although following Monte-Carlo cross-validation and false positive discovery rate correction, no significant features were positively associated with this separation (Figure 6-5).

### 3.3. Gall bladder tissue

Unsupervised analyses performed on the UPLC-MS datasets of the gall bladders did not show any sign of clustering by treatment group (Figure 6-S1). Supervised CA-PLS-DA analysis and Wilcoxon rank sum test used to analyse the significance of the treatment group separation along the predictive component of the models also showed no significant metabolic variation in the gall bladder profiles ($\alpha < 0.05$) between the treatment groups measured by UPLC-MS ESI+ and UPLC-MS ESI- (Figure 6-5).

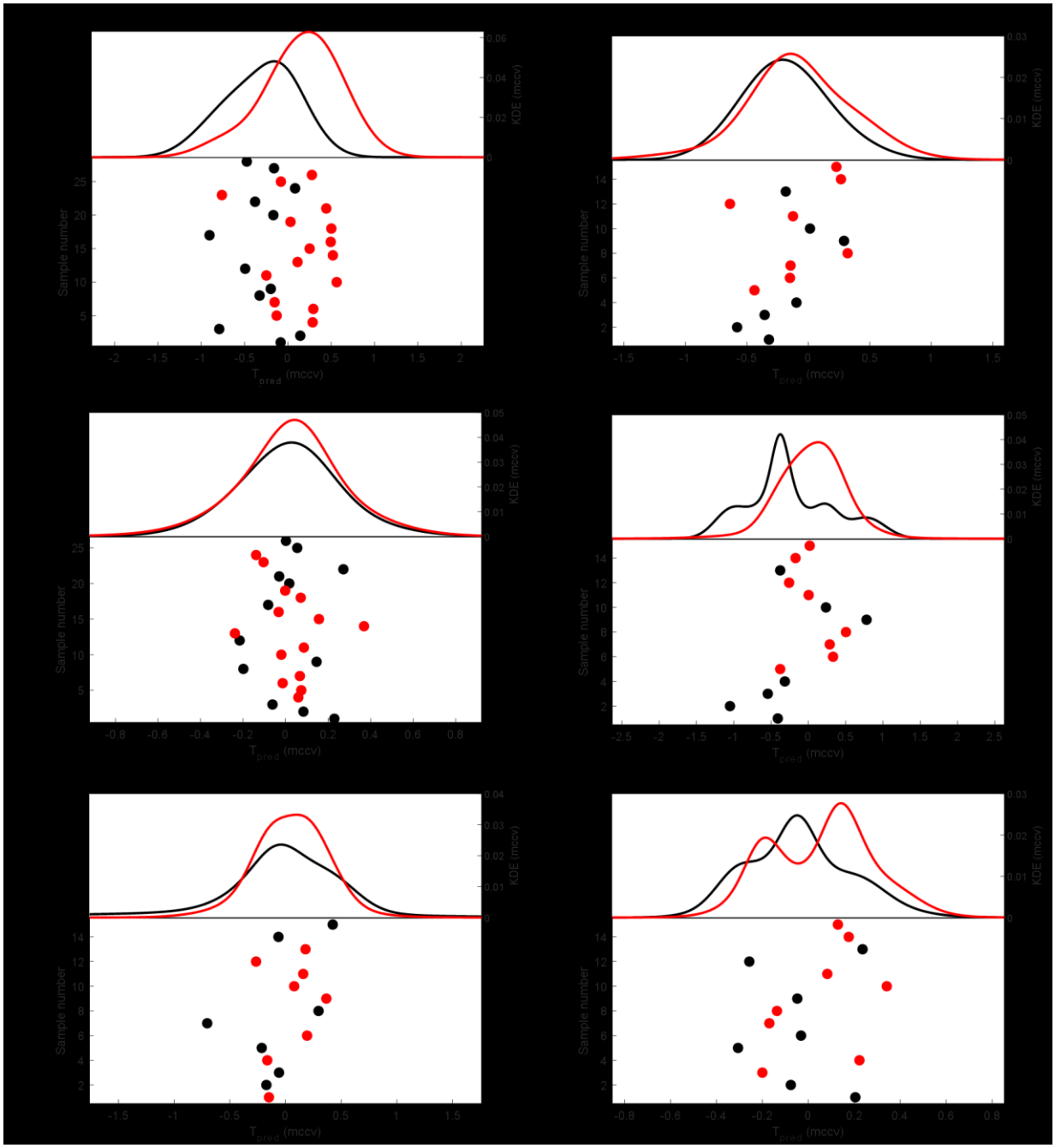



**Figure 6-5. CA-PLS-DA score plots and the kernel density estimation (KDE) obtained for the plasma, liver tissue and gall bladder samples.** The analysis performed on the UPLC-MS ESI+ (p = 0.001), the UPLC-MS ESI- (p = 0.938) and the $^{1}$H NMR spectroscopic (p = 0.613) datasets of the liver extracts are represented on the left panel. The analysis performed on the UPLC-MS ESI+ (p = 0.694) and the UPLC-MS ESI- (p = 0.463) datasets of the gall bladders and on the $^{1}$H NMR spectroscopic (p = 0.535) dataset of the plasma samples are reported on the right panel. Each dot represents a sample and the black colour represents samples from the control group while the red colour represents samples from the animals treated with the inhibitor.

### 3.4. Colon tissue

For the colon tissue extracts, no separation was observed between the treatment groups based on the unsupervised analyses performed on the $^{1}$H NMR spectroscopic and UPLC-MS datasets (Figure 6-S2). Supervised CA-PLS-DA analysis in combination with the Wilcoxon rank sum test showed that there were no significant differences ($\alpha < 0.05$) between the colon profiles from the two treatment groups in either the $^{1}$H NMR spectroscopic or the UPLC-MS ESI- datasets (Figure 6-6). A significant difference was observed for the UPLC-MS ESI+ dataset of the colon tissue extracts ($p = 2.75\text{x}10^{-4}$ - Figure 6-6).

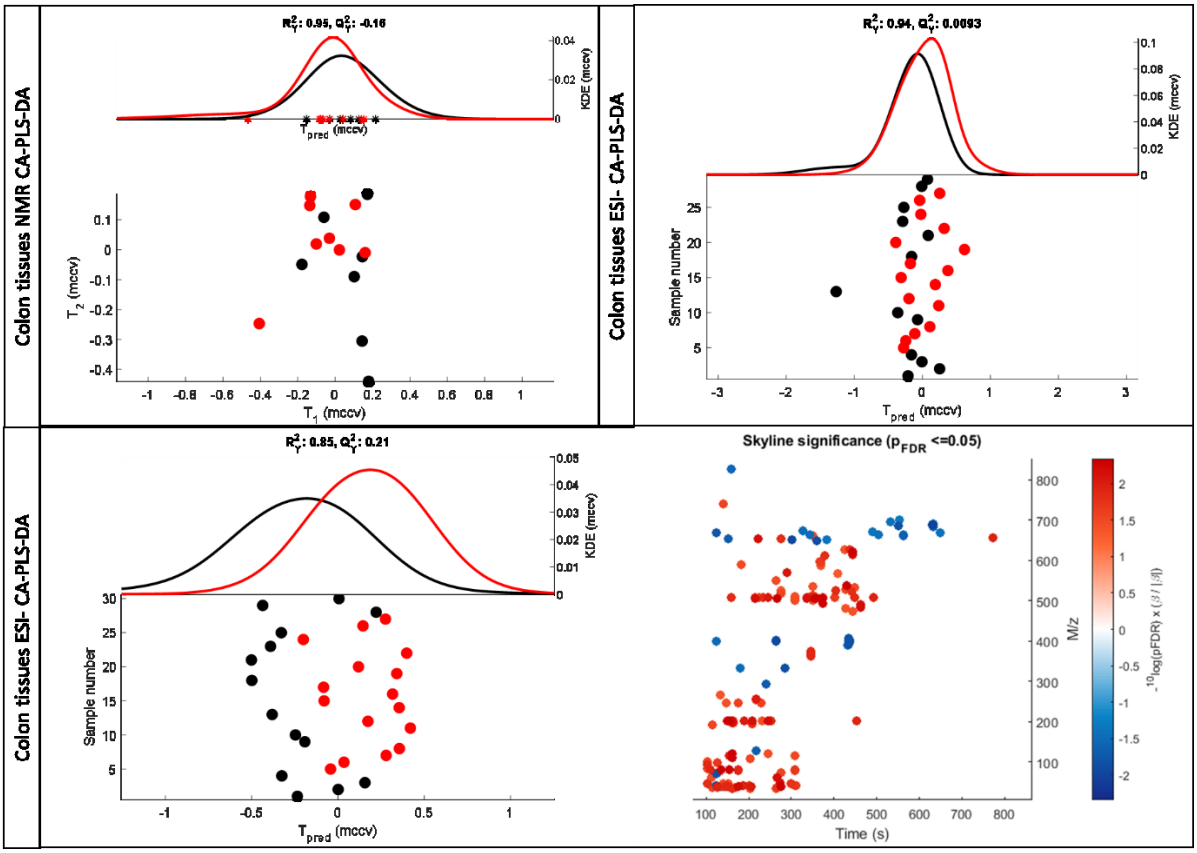



**Figure 6-6. CA-PLS-DA score plots with the kernel density estimation (KDE) obtained for the colon tissue extracts.** The analysis performed by $^{1}$H NMR spectroscopy (p = 0.152) are reported on the top left panel and by UPLC-MS ESI- (p = 0.245) dataset reported on the top right panel. The bottom panel shows the CA-PLS-DA score plot of the UPLC-MS ESI+ (p = $2.75\text{x}10^{-4}$) colon tissue extracts dataset with the KDE on the left side. For the score plots, each dot represents a sample and the black colour represents samples from the control group while the red colour represents samples from the animals treated with the inhibitor. The Skyline significance on the right side of the bottom panel, displaying features found to be significantly differ between the metabolic profiles of the mice treated with the β-glucuronidase and the control animals. Features coloured in red if they increased in the treated animals and in blue if they decreased in the treated animals compared to the vehicle-dosed animals ($\alpha < 0.05$).

For the UPLC-MS datasets, putative metabolite annotation following successful deconvolution and/or in-house database and on-line database matching was performed on features found significantly changed following bacterial β-glucuronidase treatment for colon tissues in ESI+ mode (Table 6-S2). The relative changes in the putatively annotated metabolites (increased or decreased) following the treatment is represented in Figure 6-7. An increase in the fatty-acid tetracosahexaenoic acid was observed in colon tissues. Changes were observed in bile acid species in the colonic tissue with nutriacholic acid increased with the inhibitor as well as a mild increase in glycocholic acid compared to the control animals. Overall, an increase was found for all the molecules annotated in the colon tissues of the treated animals, with the most significant increase in 4-(2-aminophenyl)-2,4-dioxobutanoate, nutriacholic acid, N-acetyl leucine, glycocholic acid, tetracosahexanoic acid, NG,NG-dimethyl-arginine and a slight increase in tyrosine (Figure 6-7).

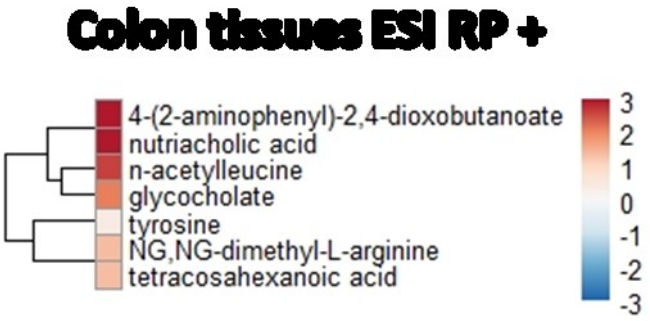



**Figure 6-7. Heatmap of the metabolites annotated as impacted by the Inh-9 treatment by UPLC-MS ESI RP + in the colon tissues.** The colours represent the -$\log_{10}$(q values) x the Manhattan sign obtained in the CA-PLS-DA models, which allow to assessment the increase (red) or the decrease (blue) of a metabolites in the animals which received the inhibitor treatment compared to the control animals.

### 3.5. Colon contents

No significant variation was observed in the UPLC-MS ESI- metabolic profiles of the colon contents between the treatment groups (Figure 6-S2). However, a significant difference was observed for the CA-PLS-DA model built on the UPLC-MS ESI+ dataset of the colon ($p$ = 5.76x$10^{-4}$, Figure 6-8). Treated animals had a lower abundance of features with a *m/z* between 50 to 600 mainly, eluting up to 8 minutes into the analysis ($n$ = 255), and an increase in features having a *m/z* ~ 650 Da eluting after 8 minutes ($n$ = 57).

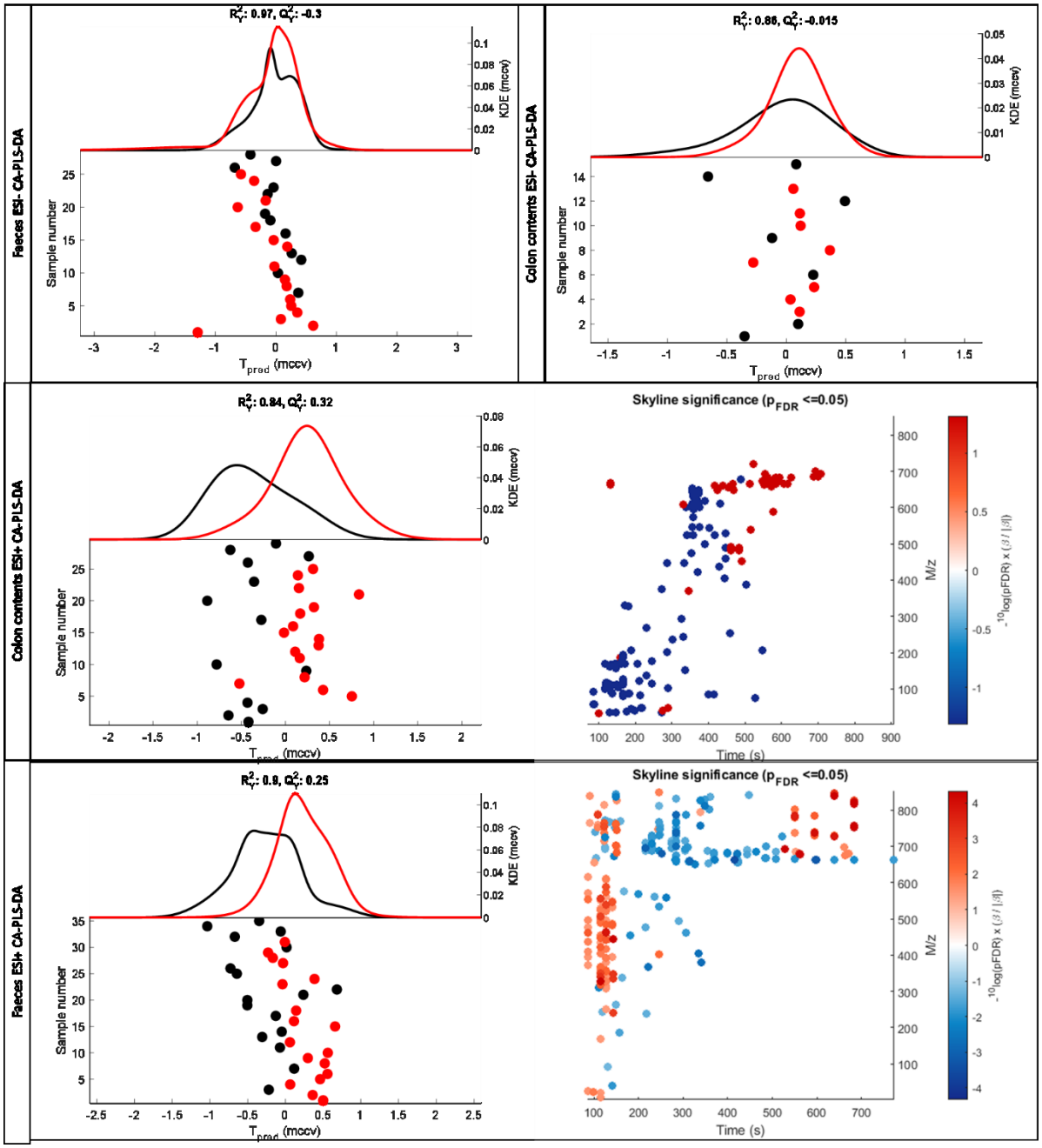



**Figure 6-8. CA-PLS-DA score plots and KDE for the models performed on the faecal and colon content extracts.** UPLC-MS ESI- datasets of the faecal extracts (p = 0.981) is reported on the top left corner and of the colon content extracts (p = 0.463) on the top right corner. Middle panel shows the CA-PLS-DA score plots and the Skyline significance obtained from the UPLC-MS ESI+ dataset obtained on the colon content extracts (p = $5.76 \times 10^{-4}$) and bottom panel shows the UPLC-MS ESI- dataset obtained on the faecal extracts (p = $7.17 \times 10^{-4}$). On the score plots, each dot represents a sample and the black colour represents samples from the control group while the red colour represents samples from the animals treated with the inhibitor. The Skyline significance plot displays features found to significantly differ between the metabolic profiles of the mice treated with the β-glucuronidase and the control animals. Features coloured in red if they increased in the treated animals and in blue if they decreased compared to the vehicle animals ($\alpha < 0.05$).

Putative metabolite annotation following successful deconvolution and/or in-house database and on-line database matching was performed on features found to be significantly changed following bacterial β-glucuronidase treatment for colon contents in ESI+ mode (Table 6-S3). The relative changes in the putatively annotated metabolites (increased or decreased) following treatment is represented in Figure 6-9.

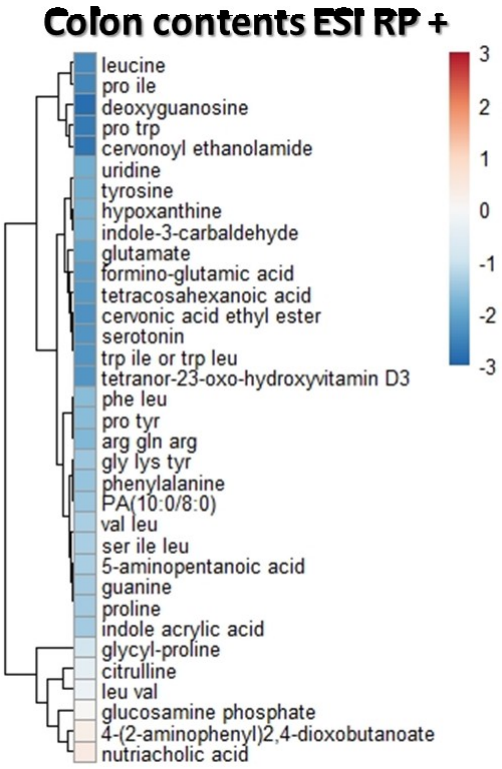



**Figure 6-9. Heatmap of the metabolites identified as impacted by the Inh-9 treatment by UPLC-MS ESI RP + in the colon contents.** The colours represent the $-\log_{10}$(q values) x the Manhattan sign obtained in the CA-PLS-DA models, which allow to assess the increase (red) or the decrease (blue) of a metabolites in the animals which received the inhibitor treatment compared to the control animals.

Similar to colon tissues, changes in lipid-like molecules were also observed, such as the decrease in the fatty-acid tetracosahexaenoic acid in colon contents (which increases in colon tissues), as well as the decrease of the vitamin D derivative tetranor-23-oxohydroxyvitamin D3. The phosphatidic acid PA(10:0/8:0) was also annotated and found to be reduced in the treated mice compared to the control animals. Overall, only nutriacholic acid, 4-(2-aminophenyl)-2,4-dioxobutanoate and glucosamine phosphate was also found to be significantly increased in the treated group compared to the control group (Figure 6-9). Several dipeptides and tripeptides were also annotated (glycyl-proline, prolyl-tyrosine, leucyl-valine, prolyl-isoleucine, glycyl-lysyl-tyrosine, seryl-isoleucyl-leucine, phenylalanyl-leucine, prolyl-tryptophan, arginyl-glutamyl-arginine, tryptohan-leucine) and were found to all decrease in the animals treated with the β-glucuronidase inhibitors compared to the controls. The same observation can be drawn for the indoles, such as indole-3-carbaldehyde and indole acrylic acid which were both found to decrease in the colon contents of the treated animals. Various compounds were also reduced in the treated group compared to the control group, such as leucine, deoxyguanosine, uridine, tyrosine, hypoxanthine, glutamate, formino-glutamic acid, or serotonin.

### 3.6. Faecal samples

No difference was also observed in the PCA scores plot comparing the faecal profiles measured by UPLC-MS ESI-. However, modest clustering was observed between the faecal samples of the treated animals compared to the control animals in the UPLC-MS ESI+ dataset (Figure 6-S2). This observation was confirmed by the supervised CA-PLS-DA analysis and Wilcoxon rank sum test ($\alpha < 0.05$). A significant difference was seen in the CA-PLS-DA model built on the UPLC-MS ESI+ dataset from the faecal samples ($p = 7.17 \times 10^{-4}$, Figure 6-8). Surprisingly, metabolite annotations were not possible as all the features found to have been significantly affected by the treatment appeared to be lipids, in particularly sterol lipids (data not shown). Unfortunately, the metabolite extraction, as well as the UPLC-MS RP method were both optimised for small molecules observation rather than lipids. Therefore, the metabolite annotation was difficult and the confidence in the annotation was too low to be reported.

### 3.7. Urine samples

For the urine metabolic profiles measured by $^{1}$H NMR spectroscopy and UPLC-MS, a separation was observable in the PCA score plots, but it was subsequently attributed to the sex of the animals, rather than inhibitor treatment (Figure 6-10). Indeed, metabolite identification by $^{1}$H NMR spectroscopy showed that this separation was due to trimethylamine (TMA) and trimethylamine-*N*-oxide (TMAO) (loading plot, Figure 6-10).

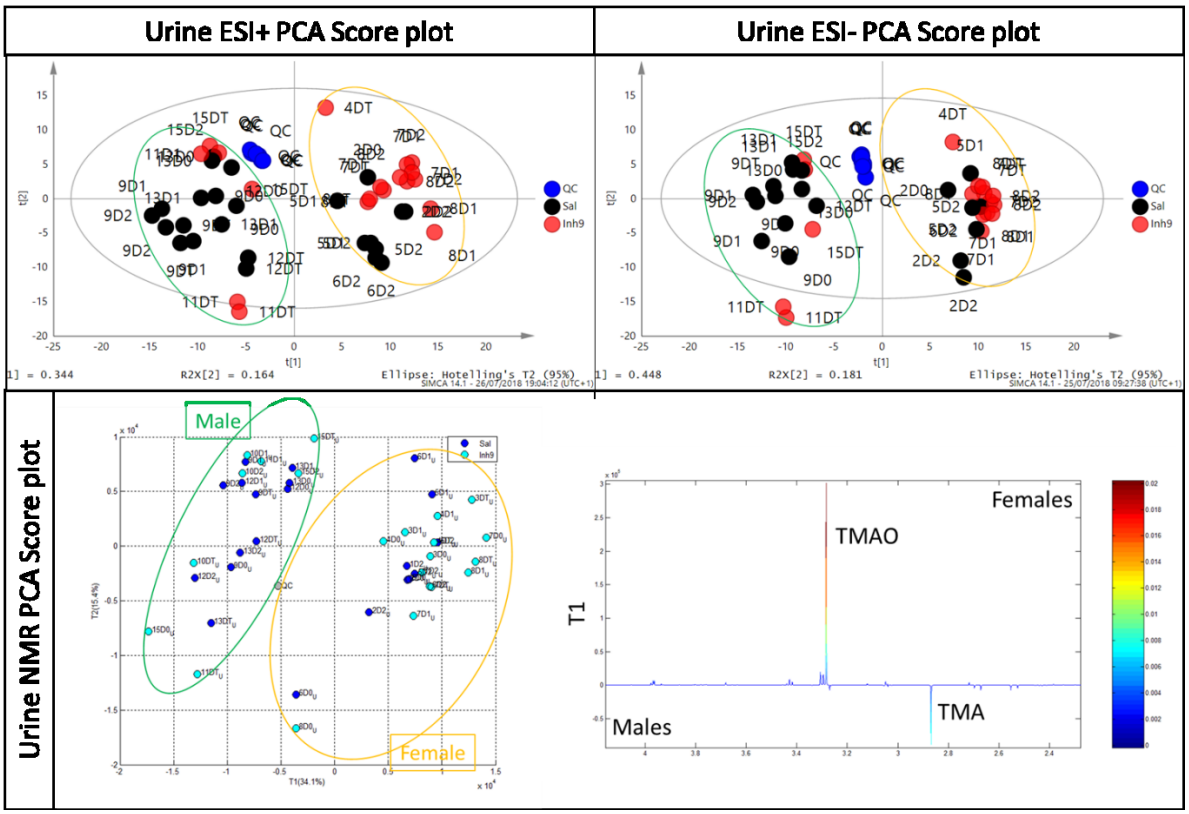



**Figure 6-10. Unsupervised analysis performed on the UPLC-MS and $^{1}$H NMR spectroscopic urine sample dataset.** PCA score plots obtained by both UPLC-MS ESI+ and UPLC-MS ESI- of the urine samples as well as the PCA score plot obtained by $^{1}$H NMR spectroscopy for the urine sample and the loading plot and showing the metabolites driving the separation along principal component 1. The loading plots linked to the NMR PCA showed that the separation between male and female mice was due to higher concentrations of TMAO in females and of TMA in males.

The CA-PLS-DA models adjusted for cofounders that could bias the model, such as sex, indicated significant metabolic variation related to inhibitor treatment, for the $^{1}$H NMR spectral profiles ($p$ = 0.005), the UPLC-MS ESI+ ($p = 2.67 \times 10^{-6}$) and the UPLC-MS ESI- ($p = 4.55 \times 10^{-6}$) datasets of the urine samples (Figure 6-11).

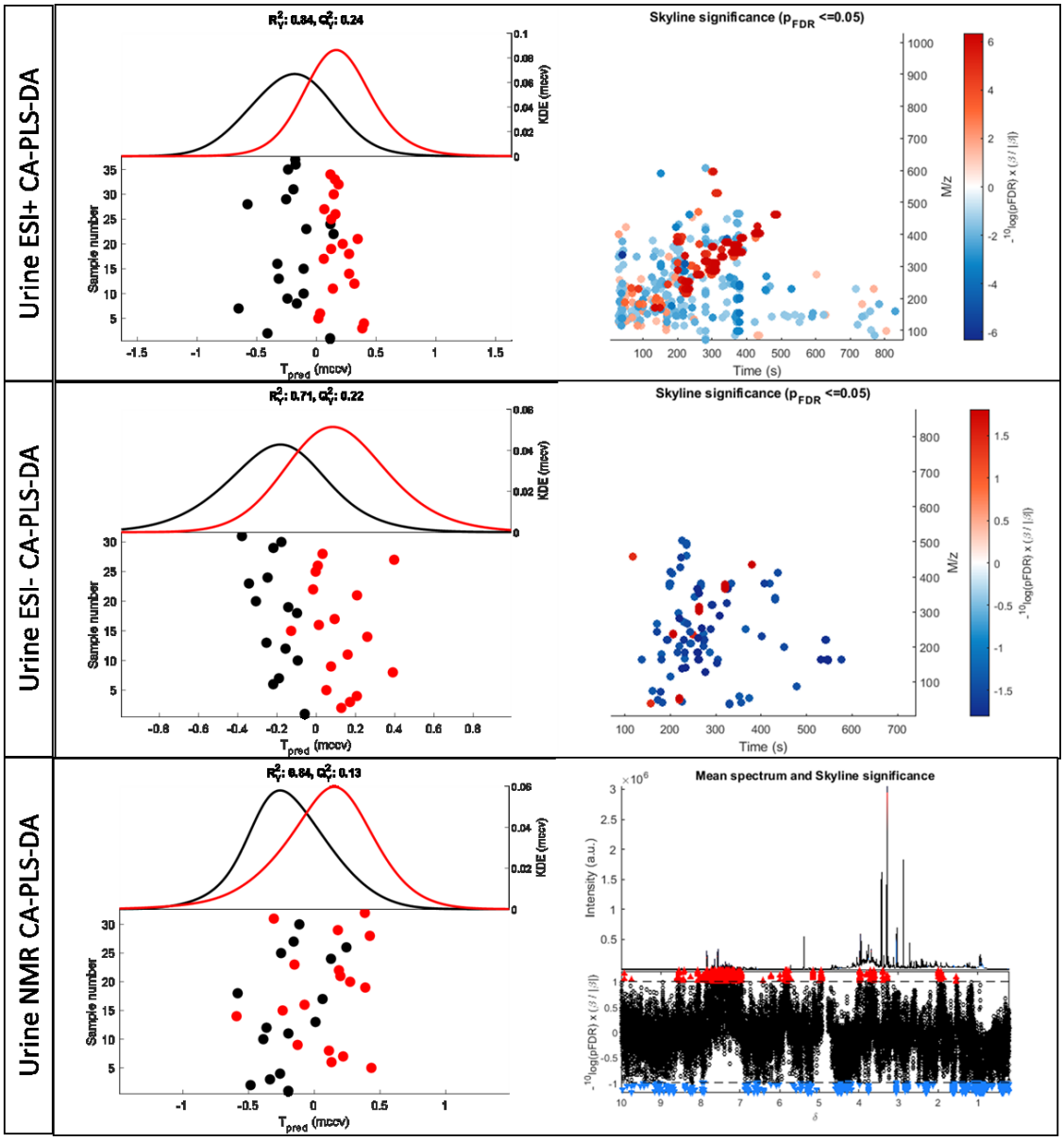



**Figure 6-11. Supervised analysis performed on the UPLC-MS and $^{1}$H NMR spectroscopic urine sample dataset.** CA-PLS-DA models obtained for both UPLC-MS ESI+ ($p = 2.67x10^{-6}$) and UPLC-MS ESI- ($p = 4.55x10^{-6}$) datasets are reported in the top and middle panel respectively and the CA-PLS-DA model obtained for the $^{1}$H NMR spectroscopic ($p = 0.005$) dataset is reported on the bottom panel. The left side shows the score plots and the right panel the skyline significance. On score plots, each dot represents a sample and the black colour represents samples from the control group while the red colour represents samples from the animals treated with the inhibitor. The Skyline significance plot displays features found to significantly differ between the metabolic profiles of the mice treated with the β-glucuronidase inhibitor and the control animals is reported on the right side of each panel. Features are coloured in red if they increased in blue if they decreased in the treated animals compared to the vehicle-dosed animals ($\alpha < 0.05$ for UPLC-MS and $\alpha < 0.1$ for $^{1}$H NMR spectroscopy).

Putative metabolite annotation of features found significant changes following bacterial β-glucuronidase treatment for urine samples analysed by both $^{1}$H NMR spectroscopy, and UPLC-MS in ESI+ mode (Table 6-S3) and ESI- mode (Table 6-S4). Urinary variation (increase or decrease) of the

metabolites successfully annotated in the treated mice compared to the control is reported in Figure 6-12. $^{1}$H NMR spectroscopy results show increases in TMAO, 3-indoxyl sulfate (3-IS), phenylacetylglycine (PAG), orotate, hippurate, *p*-cresol sulfate (*p*CS) and *p*-cresol glucuronide (*p*CG) in the animals that received the β-glucuronidase inhibitor, as well as a decrease in the excretion of creatinine, TMA and trans-aconitate. Similarly, to $^{1}$H NMR spectroscopy, UPLC-MS ESI– results showed an increase in PAG, *p*CS and *p*CG in the urine of the treated mice, along with phenyl glucuronide and allantoin. In the urine ESI+ dataset, hydroxyprolyl-methionine, *N*-acetyl glutamine and *N*-acetyl lysine were the annotated compounds found to be significantly higher in the treated mice compared to the controls. Similarly, to the colon tissue and the colon contents, 4-(2-aminophenyl)-2,4-dioxobutanoate was found to be increase in both ionization modes. However, the dipeptide glycyl-proline was found to increase in the urine of the treated group were found to increase in urine while it was found to decrease in the colon contents. A larger number of compounds were found to be decreased in the treated animals based on the urine ESI+ dataset, such as uridine and leucine (similar to colon contents), NG,NG-dimethyl-arginine (found to increase in to colon tissues), tyrosine (found to decrease in colon contents but increase in colon tissues), as well as trigonelline, biotin, hexanoylglycine and octenoylglycine, glycyl-aspartate, or 3-hydroxyanthranilic acid found to be decreased.

Several additional glucuronide-compounds were found to be affected also in the urine ESI+ dataset, such as dihydroxyquinoline-β-glucuronide, or dihydrocaffeic acid glucuronide, which were found slightly reduced in the mice that received the inhibitor compared to the control mice. Moreover, three other compounds from which the MS spectra of the glucuronide form were not available on database, were annotated and supposedly in the form of a glucuronide because of a neutral loss of 176 Da observed from the $[M+H]^{+}$ ion. These compounds include a 2-indole-carboxylic acid glucuronide, which was found to slightly increase in the treated group, phenylalanine glucuronide and xanthurenic acid glucuronide which were found to decrease in the urine samples of the treated mice compared to the control mice. In the urine sample analysed in the negative ionization mode, several additional glucuronide-compounds were significantly affected by the treatment. Consistent with the positive ionization mode data, phenylalanine glucuronide, dihydroxyquinoline-β-glucuronide, dihydrocaffeic acid glucuronide and xanthurenic acid glucuronide were excreted in lower amounts by the mice receiving the inhibitor compared to those receiving the saline. However as reported above, phenyl-glucuronide and *p*-cresol glucuronide were found to increase in the treated mice. Additional conjugated metabolites changed with inhibitor intake. As its glucuronide homologue, *p*-cresol sulfate significantly increased following the inhibitor while *p*-coumaric acid sulfate and ferulic acid 4-*O*-sulfate, were both significantly decreased. Another metabolite from tyrosine catabolism, *p*-

hydroxyphenyllactic acid, was significantly reduced in the urine following inhibitor treatment. Other compounds which were found lower in the urine samples of the treated animals were sucrose, glycyl-aspartate, pimelic acid and porphobilinogen, as well as hexanoylglycine and octenoylglycine (similarly to the ESI+ dataset). Others compounds significantly increased following inhibitor treatment compared to the controls were 4-pyridoxate, glycerophosphocholine and glycerol-phosphate, cinnamoyglycine, N-acetyl glutamic acid and N-acetyl histidine, arabinose, hippuric acid and succinic acid. The fatty acyl tetranor-5-$NO_2$-CLA was also detected in urine in both ionization modes showing a slight increase in the urine samples of the treated mice.

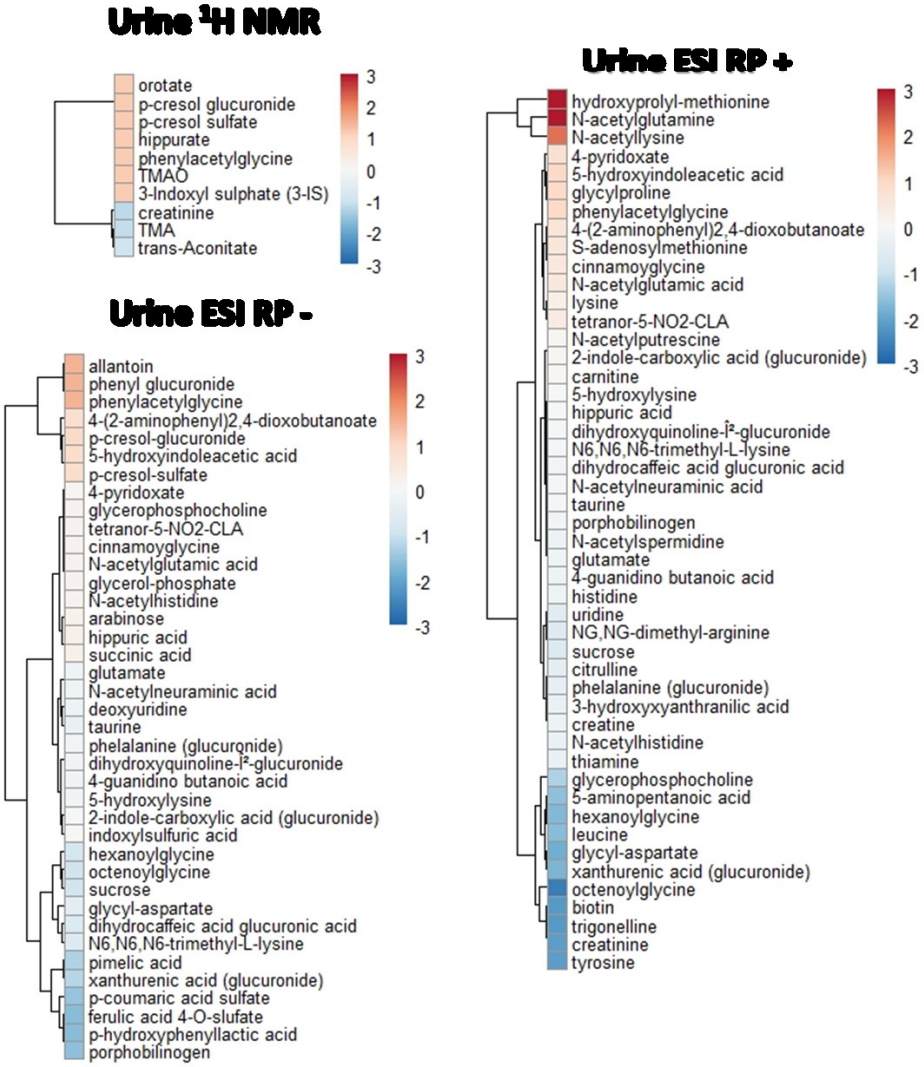



**Figure 6-12. Heatmap of the putatively annotated metabolites found significantly different between the Inh-9-treated animals and the control animals, in the urine samples by $^1$H NMR spectroscopy, UPLC-MS ESI RP + and ESI RP –.** The colours represent the -$\log_{10}$(q values) x the Manhattan sign obtained in the CA-PLS-DA models, which allow the increase (red) or the decrease (blue) of a metabolites in the animals which received the inhibitor treatment compared to the control animals to be assessed.

## 3.8. Network analysis

A molecular network was generated by importing 53 of the metabolites annotated by UPLC-MS in the urine samples (Figure 6-13). In total, the urine metabolites were involved in 54 pathways (listed in Figure 6-13), with 8 pathways related to transport. The pathways most related to the urinary metabolites included the biosynthesis of tyrosine, phenylalanine and tryptophan, biotin metabolism, dipeptide degradation, taurine metabolism, biomass, lysine metabolism, phenylalanine metabolism urea cycle, methionine metabolism, histidine metabolism and tyrosine metabolism (coverage ≥ 10). From the metabolites annotated, many were involved in reactions that occur in the cytoplasm or in the mitochondrion. Two reactions were involved in the transport of metabolites between the mitochondrion and the cytoplasm. The first one is part of the glycolysis/gluconeogenesis pathway and involved a reaction between glycerol-3-phosphate (G3P) from the cytoplasm and flavin adenine dinucleotide oxidized (FAD ox) from the mitochondrion to produce dihydroxyacetone phosphate (DHAP) and the reduced FAD (FAD red). The second is part of the urea cycle/amino group metabolism and involved in the two ways passive diffusion of creatine. Transport of taurine from cytosol to peroxisome is also underlined as well as the transport of 3,4-dihydroxy-L-phenylalanine from the tyrosine pathway, across the apical surface of the lumen to the cytoplasm (Figure 6-13).

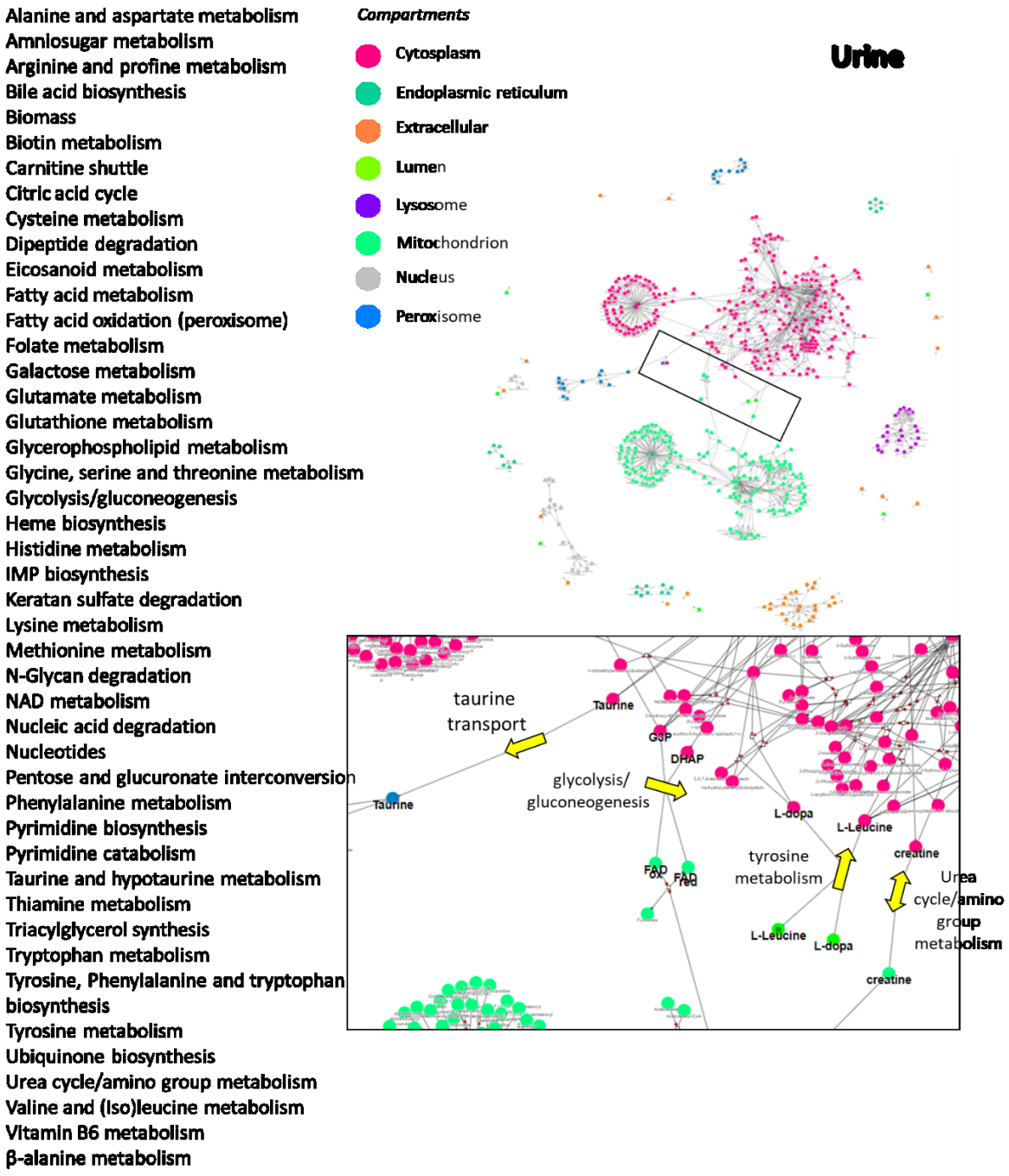



**Figure 6-13. Molecular network of the metabolites annotated in the urine samples analysed by UPLC-MS.** The list of pathways in which those metabolites are involved is listed on the right side (except the transport pathways). The colours of the metabolites represent the compartments, listed on the middle top of the picture. A zoom is made on the reactions doing the junction between the cytoplasm network (pink) and the mitochondrion network (blue green). G3P: glycerol-3-phosphate (G3P), FAD ox: flavin adenine dinucleotide oxidized (FAD ox), FAD red: flavin adenine dinucleotide reduced, DHAP: dihydroxyacetone phosphate.

Molecular network analysis of the colon content extracts was generated by importing 30 of the metabolites annotated by UPLC-MS. A total of 38 pathways were affected, mainly in the cytoplasm (Figure 6-14). The ones presenting the most coverage were tyrosine, phenylalanine and tryptophan biosynthesis, the salvage pathway, dipeptide degradation, purine catabolism, phenylalanine metabolism, biomass and nucleic acid degradation (coverage ≥ 10). In the mitochondrion, arginine and proline metabolism, nucleotides, folate metabolism, valine and (iso)leucine metabolism, tyrosine metabolism, lysine metabolism, glutamate metabolism, phenylalanine metabolism and cysteine metabolism were the most represented pathways. Folate metabolism was also reported to take place in the lysosome and extracellularly (Figure 6-14). No transport connections from one compartment to another were highlighted in the colon content molecular network.

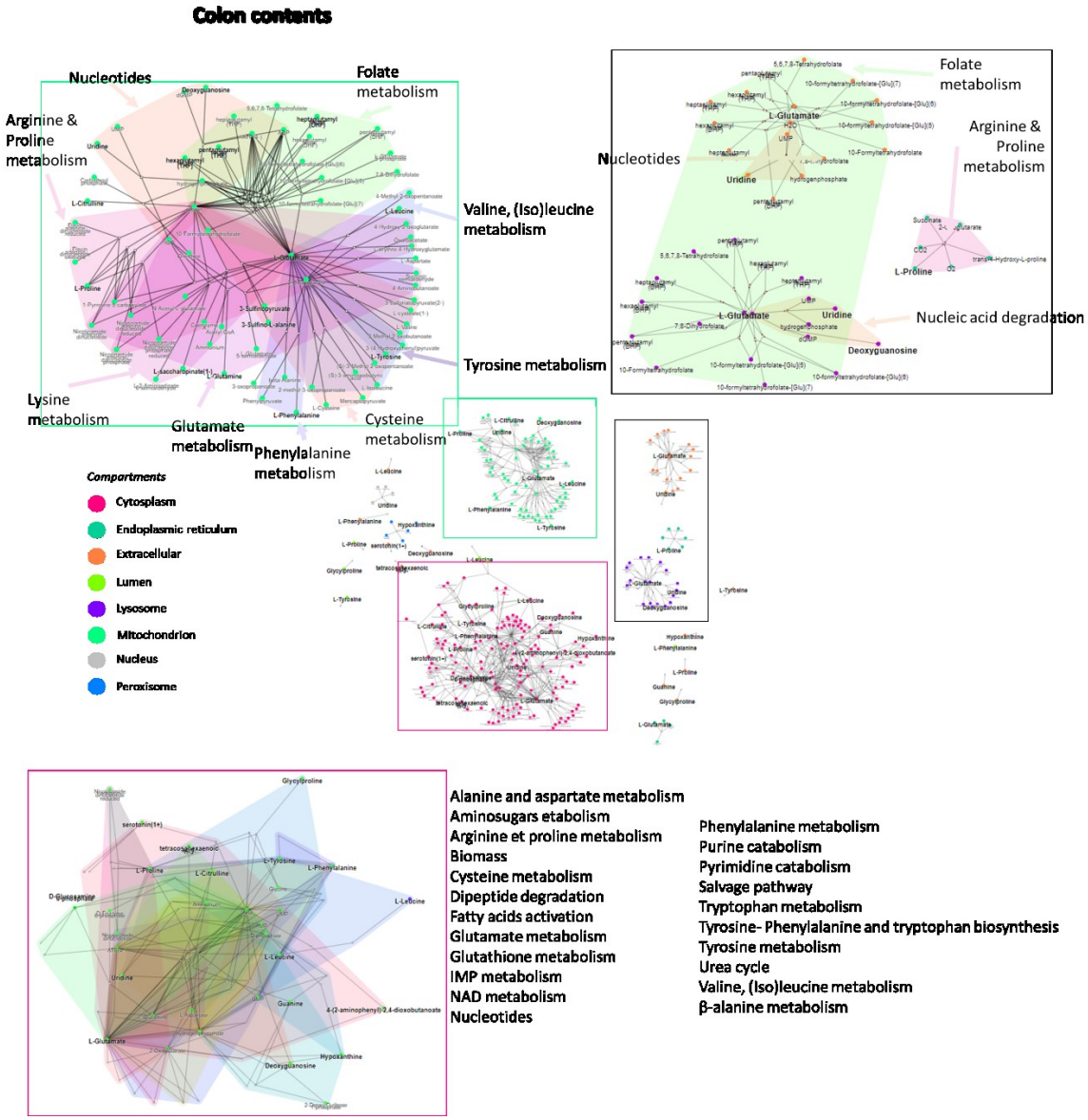



**Figure 6-14. Molecular network of based on the metabolites annotated in the colon contents.** A global view of the molecular network is shown in the middle, with the colour of the metabolites corresponding to the compartments where the reactions are taking place. A zoom on the reactions occurring in the cytoplasm is made on the bottom, with the metabolites which have been imported displayed, and with the list of pathways affected on the left side. A zoom on the reactions occurring in the mitochondrion was also made (top left corner) with the annotated metabolites which were imported in bold and the pathways affected directly around. The same was performed to zoom on the reactions occurring on the lysosome, the ER and outside the cell (top right corner).

The molecular network of the colon tissues was generated by importing 7 of the metabolites annotated by UPLC-MS (Figure 6-15). Once again, tyrosine, phenylalanine and tryptophan biosynthesis were the pathway the most represented by the metabolites, followed by food processing-related pathways, tyrosine metabolism, biomass, tryptophan metabolism, bile acid biosynthesis and fatty acid activation (coverage ≥ 1). As for the colon contents, no metabolite was identified as part of a transport reaction doing the junction from one compartment to another.

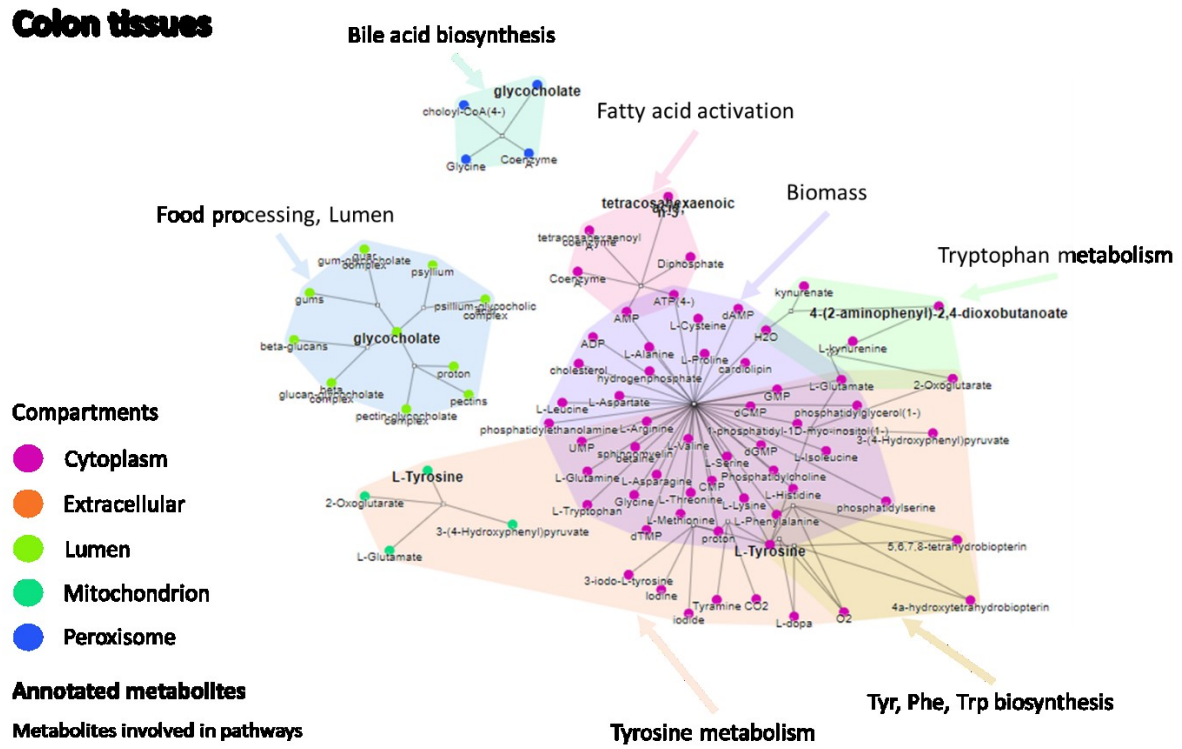



**Figure 6-15. Molecular network of the colon tissues.** Annotates metabolites which have been imported are highlighted in bold and with the pathways affected indicated around. The colour of the metabolite indicates the compartment were the reactions are taking place.

# 4. Discussion

With the central role that gut microbial metabolism has in the potentiation of irinotecan gut toxicity, the feasibility of selectively attenuating this microbial metabolism to modulate drug toxicity outcomes was explored. While collaborators previously demonstrated that blocking bacterial β-glucuronidases could reduce irinotecan toxicity (Figure 6-1 (Wallace et al., 2015, 2010)), it was unknown if there were wider ramifications for microbial metabolism, microbe-host biochemical interactions and also host metabolism. As such, this chapter studied the metabolic disruptions to the host caused by the inhibition of bacterial β-glucuronidases in the mouse. This was studied locally in the gut and at the gut wall and globally throughout the biochemical system of the mice.

Bile acids, such as glycocholic acid and nutriacholic acid, were increased in the colon wall and contents following bacterial β-glucuronidase inhibition. Bile acids are secreted into the gut as glycine and taurine conjugates (Goto et al., 2007). The small proportion (1-5 %) of conjugated bile acids not re-absorbed in the small intestine reach the colon where they become modified by the colonic microbiota. Firstly, bile salt hydrolase enzymes cleave the glycine or taurine residue from the bile acids before further modifications occur to the steroid nucleus resulting in the production of secondary bile acids. Most of the bacteria present in the large intestine are able to deconjugate bile acids with specificity towards different bile acids. It has been shown *in vitro* that *Clostridium perfringens* and *Bacteroides fragilis* are particularly influential in deconjugating the primary bile acid, glycocholic acid (Masuda, 1981). It is unclear why inhibiting bacterial β-glucuronidase activity increases the abundance of this bile acid in colon tissue, but glucuronidated bile acids are from interest as it has been hypothesized that they could potentially be a non-neglectable energy source to bacteria able to process them (Pellock and Redinbo, 2017). However, glycocholic acid has been reported to be an activator of β-glucuronidases in the human liver suggesting that these changes may promote the up-regulation of the host glucuronidases to compensate for the loss of this bacterial activity (Ho, 1985).

In the urine metabolic profile, obtained by both $^{1}$H NMR spectroscopy and UPLC-MS, several metabolites arising from microbial metabolism were altered by the enzyme inhibitor. This included an increase in TMAO, hippurate, or 5-hydroxyindoleacetic acid (5-HIAA), *N*-acetylputrescine, succinate and a decrease in creatinine, creatine, TMA, biotin. Interestingly, 5-HIAA is formed by the degradation of serotonin, which was found to decrease in the colon contents of the treated mice. This suggests that the inhibition of bacterial β-glucuronidases has a wider impact on gut microbial metabolism including the metabolism of bile acids (glycocholic acid), choline (production of TMAO from TMA), vitamins (biotin), polyamines (putrescine derivatives), indoles (5-HIAA or serotonin), and lipids

(Nicholson et al., 2012). Bacterial tyrosine metabolism was also altered with the inhibitor. PAG was found to be increased in the treated mice whilst tyrosine and its microbial degradation products, *p*-hydroxyphenyllactic acid and *p*-coumaric acid sulfate, were decreased compared to the control animals.

Inhibition of the bacterial β-glucuronidases resulted in an increase in the urinary excretion of phenol glucuronide, alongside other glucuronides such as *p*-cresol glucuronide. While this is expected, there was a lower excretion of other glucuronides following treatment, such as xanthurenic acid glucuronide, phenylalanine glucuronide, dihydrocaffeic acid glucuronide, dihydoxyquinoline-β-glucuronide and 2-indole-carboxylic acid glucuronide. A potential explanation is that Inh-9 has a specificity towards the enzymes from the Firmicutes and Proteobacteria and not towards those from Bacteroidetes (Wallace et al., 2015). Indeed, there is a well-described set of six distinct active site geometries for gut microbial β-glucuronidases, and these enzymes have differential activities with distinct substrates (Pollet et al. and Redinbo, Structure, 2018). This could only affect the bacterial enzymes in charge of the hydrolysis of the glucuronic acid placed on non-substituted phenols, or only para-substituted phenols, which is the case of *p*-cresol glucuronide and the active ingredient of SN-38. Secondly, the lack of glucuronidase activity decreases the cleavage of some glucuronide-conjugated compounds, which is assumed to decrease the bacterial quantity of glucuronic acid available to perform further glucuronidation.

Nevertheless, the biochemical perturbations associated with the inhibition of the bacterial β-glucuronidases appeared to be restricted to the colonic contents, colon wall, faeces, and urine metabolic profiles. No effect was seen in other parts of the global metabolic system of the mice, including liver, gall bladder, and plasma. This demonstrates the feasibility of selectively inhibiting bacterial β-glucuronidases using Inh-9 to limit irinotecan ADRs without impacting on the host metabolic system. The fact that Inh-9 treatment was found to affect principally gut microbial co-metabolites, in the colon tissues, colon contents, faeces and urine could indicate the potential impact of Inh-9 on the microbial community and its functionality. However, collaborators have previously demonstrated that the β-glucuronidase inhibitor does not modulate the community structure of the intestinal microbiota (Bhatt et al., 2020; Wallace et al., 2015). Moreover, irinotecan has been found to increase the abundance of Proteobacteria (68 %) in mice but such changes were reduced (14 %) when irinotecan was co-administered with the inhibitor. Similarly, bacterial α diversity was reduced in mice following irinotecan treatment, but not when the drug was co-administered with the inhibitor (A. P. Bhatt et al., submitted). Thus, Inh-9 successively targets the microbiome without impacting host metabolism demonstrating that specific manipulation of this bacterial activity can be used to optimise drug treatment strategies.

# 5. Conclusion

The potential of inhibiting bacterial β-glucuronidases as a strategy to limit ADRs induced by irinotecan treatment by blocking the release of the active ingredient SN-38 in the gut was assessed in mice. Inhibition of bacterial β-glucuronidases altered the abundance of several microbial co-metabolite classes, such as bile acids, lipids, dipeptides, indoles, as well as choline-, polyamine-, vitamin-derivatives and tyrosine-related metabolites. However, bacterial β-glucuronidase inhibition does not induce significant changes in metabolic profile of the global host system. This demonstrates that specific inhibition of the bacterial β-glucuronidase is a usable strategy to limit the toxicity of irinotecan chemotherapy potentially allowing the dose range to be expanded and efficacy to be improved. With our increasing appreciation for the influence of the gut microbiome on host drug metabolism and toxicity there is an increasing demand to develop novel strategies to fine tune this influence for favourable drug outcomes.

# Chapter 7. General Discussion and Perspectives

## 1. General discussion

This thesis aimed to assess the potential for the gut microbiota to directly and indirectly influence host drug metabolism and toxicity and for drugs to impact on the microbes themselves. In addition, it aimed to develop a tool to determine this gut microbial potential for drug metabolism prior to drug treatment and to explore methods to modulate gut microbial metabolism to reduce adverse drug reactions without negatively impacting the host. This was achieved through several studies discussed in the previous chapters.

First, direct interactions between the chemotherapeutic drug, methotrexate (MTX) and the bacterial enzyme CPDG2 was studied in rats. DAMPA, a non-toxic metabolite of MTX produced by the bacterial enzyme CPDG2 as a detoxification reaction (Zaharko et al., 1969; Valerino et al., 1972), was quantified in faecal samples. Correlation with amplicon sequencing data indicated that the faecal excretion of DAMPA was positively associated with bacteria mainly from the phylum Firmicutes, which suggests that this bacterial phylum is able to produce the enzyme CPDG2 to degrade MTX. While further analysis is required to validate this hypothesis, determining the abundance of these bacteria in patients prior to dosing may have potential to predict MTX treatment outcomes. Such information could help to inform this chemotherapeutic treatment contributing to the personalisation of this strategy. Patients identified to have a low relative abundance of Firmicutes could be proactively supplemented with CPDG2 to mitigate against MTX toxicity rather than used as a rescue agent as is currently the case (Rattu et al., 2013). Conversely, high doses of MTX were shown to perturb the gut microbial community, increasing the relative abundance of Bacteroidetes and decreasing the relative abundance of Firmicutes, presumably as a result of gut toxicity. Chronic exposure to MTX could modify the composition and functional capacity of the microbiome with a subsequent impact on its ability to detoxify the drug as well as other xenobiotics. Such perturbations could also have implications for the immune system of cancer patients, which is known to be altered by chemotherapeutic treatment. These results demonstrate the direct two-way interactions that occur between MTX and the gut microbiota, a relationship that could be altered through chronic dosing. A clearer understanding of this relationship in humans has great potential to enhance the effectiveness of MTX and minimise toxic outcomes.

Attempts were made to develop an *in vitro* model that could assess the potential for an individual's microbiome to have a direct impact on a specific drug. This model was successfully applied to paracetamol (APAP) and sulfasalazine (SSZ), which were used as negative and positive controls, respectively. This model highlighted the variability in the functional capacity of the microbiome across different donors to degrade the prodrug SSZ. This reinforces the contribution of the intestinal microbiota to variation in drug responses and the value in determining this metabolic capability prior to drug administration. However, this *in vitro* model requires further development as a number of technical issues need to be resolved. Specifically, there were issues maintaining the viability of the faecal cultures and problems with the solubility of the drugs in the model. As a result of these issues the *in vitro* model could not be used to study MTX and paracetamol-glucuronide (APAP-G). If this model can be developed successfully, it would allow the metabolic capacity of the entire faecal microbial community to be studied simultaneously, rather than a single bacterial isolate or a restricted set of bacterial isolates (Haiser and Turnbaugh, 2012; Maier et al., 2018; Zimmermann et al., 2019a). If correlated with 16S rRNA gene sequencing and applied to a large study population, this simple model could be used to identify bacterial groups that are associated with the drug degradation profile. This *in vitro* model could also be used to rapidly and simultaneously profile several therapeutic treatments against an individual's gut microbiota prior to drug selection. This could enable the most efficient and/or the least toxic treatment options to be selected without harm to the patient.

Attempts were also made to explore how the diet could interact with the microbiota to liberate microbial metabolites that influence the host phase II (conjugative) metabolism of APAP. Here, a high tyrosine diet was fed to female BALB/c mice with the aim of increasing the microbial production of *p*-cresol in the gut, as demonstrated previously (Shiba et al., 2014). Increased host exposure to *p*-cresol was hypothesised to increase its hepatic sulfation generating *p*-cresol sulfate (*p*CS) while depleting the sulfation capacity of the liver, a key drug metabolism reaction involved in APAP metabolism. Unfortunately, the high tyrosine diet was not observed to increase the production of *p*CS as it had been anticipated, presumably due to a saturation of the bacterial enzymes responsible for producing *p*-cresol. This suggests that a pronounced diet-induced change in the production of microbial metabolites may require longer term exposures to promote the growth of certain bacterial groups, such is the case with lifestyle choices for example, vegetarianism. Nevertheless, it was still possible to investigate the competitive interaction between *p*-cresol and APAP. Interestingly, in the control mice the excretion of *p*CS was found to be inversely associated with the sulfation of APAP, as evidenced by the APAP sulfate/glucuronide ratio. Thus, it appears that increased host exposure to *p*-cresol indeed depletes the sulfation capacity of the liver, which leads to a greater amount of APAP being glucuronidated rather sulfated. Furthermore, the loss of this microbial *p*-cresol production

following antibiotic treatment resulted in an increased APAP-S/APAP-G ratio due to lower hepatic *p*-cresol exposure. These observations provide further confirmation of a previous hypothesis drawn in a human study performed by (Clayton et al., 2009) and could have clinical significance by limiting sulfation capacity of the host. Regarding APAP metabolism, this can lead to the production of hepatotoxic metabolites (*i.e*. NAPQi). Increased risk of elevated *p*-cresol concentrations in the host has been observed in populations following high-fat (Wan et al., 2019) and low-fibre diets (Khosroshahi et al., 2019). Taking into consideration the Western-type diet adopted in most industrialised countries, many individuals are likely to be exposed to high amounts of *p*-cresol, with potential downstream consequences for their capacity to sulfate and clear endogenous and xenobiotic compounds. This highlights that indirect interactions between microbial metabolites and the host drug metabolizing systems should also be factored in when considering the microbiome in the individualisation of therapeutic strategies.

Finally, bacterial β-glucuronidase enzymes are known to potentiate the toxicity of irinotecan. Targeted inhibition of this bacterial activity has been demonstrated to be an effective strategy to blunt this toxic outcome (Wallace et al., 2010). This demonstrates the potential utility of selectively modulating the activity of the microbiome to promote favourable drug outcomes. However, with the fine balance between the microbiota and the host, as well as the multiparametric biochemical exchange occurring between these two complex systems, it was not clear if selective inhibition of the microbiota would have wider downstream effects on host metabolism and ultimately host health. Therefore, work was undertaken to characterise the wider biochemical impact, if any, of that inhibition on the host and microbial global metabolic profile. Importantly, inhibition of bacterial β-glucuronidases was shown not to impact the wider metabolic profile of the mouse host and the metabolic changes were limited to alterations in microbial metabolites within the gut (colon tissues and colon contents), as well as in the faecal and urine samples. Inhibition of bacterial β-glucuronidases was found to affect microbial metabolites including bile acids, lipids, vitamins, choline and polyamine derivatives, as well as indoles and tyrosine metabolites such as *p*CG and phenol-glucuronide. The increased amounts of these latest metabolites is not anticipated to have downstream consequences for host health as they are predominantly excreted in the urine and no differences were observed in the liver tissue or plasma. Inhibition of the β-glucuronidases appeared to be specific to those of bacteria from the phyla, Firmicutes and Proteobacteria but not the Bacteroidetes (Wallace et al., 2015). This suggests that there are different levels of specificity that can be targeted for inhibition. These results demonstrate the potential for manipulating an individual's microbiota to improve drug outcomes. In this example, the gastrointestinal toxicity induced by irinotecan can be limited allowing dose levels to be increased to potentially enhance its effectiveness. A similar strategy could be applied to other classes of drugs,

such as non-steroidal anti-inflammatory drugs (NSAIDs), which can also be de-glucuronidated in the gut (LoGuidice et al., 2012). In combination with *in vitro* models, an individual's microbial β-glucuronidase capacity could be evaluated prior to dosing and the inhibitor could be delivered accordingly.

Collectively, the multifaceted approach adopted here enables a better comprehension of the biomolecular mechanisms behind the tripartite drug-microbiota-host interaction and its potential ramifications for host health. Together, these results show the importance to assess the biomolecular exchange happening between a drug and an individual microbiota, the need to further explore this variable in relation pharmacological and toxicological studies, and how it could be used in pharmaceutical and clinical context as a strategy to manipulate the microbiome to improve drug outcomes and thus facilitate the development of personalised therapeutic treatment.

## 2. Limitations

Exploring drug-microbiome-host tripartite interactions would benefit from more studies being performed in humans rather than rodents. Indeed, the gut microbiome of human is assumed to be different from a mouse microbiota. In consequence, conclusions drawn from rodent models require extra care when translating their importance to humans. Also, even more extensive inter-individual variability is expected in humans compared to mice, which can lead to extensive variation in drug response and toxicity. Intravenous treatment and its impact on the host microbiota as well as how drugs administrated by this route are metabolised by the gut microbiota once reaching the intestine could also be explored. Finally, the work reported in this thesis is restricted to the study of three drugs and their respective interactions with the microbiota and thus does not cover all the possible interactions which can occur between a drug and gut microbiota. However, investigating the interaction of a single drug with the gut microbiota requires time-consuming studies and often the use of animals which, as discussed above, limit the translation of the results. The development of a model to perform high-throughput screening with human gut microbiota is thus needed in the field. Unfortunately, difficulties in *in vitro* model development to investigate the impact of the gut microbiota on a larger number of drugs were encountered and suggest the need for further optimisation.

## 3. Perspectives and further work

Further development of the promising *in vitro* model to explore an individual microbiota capacity to metabolise a drug will be undertaken in a wider project, within which additional precautions such as performing the protocol directly in an anaerobic cabinet to substitute the use of anaerobic jar is

seriously considered. Moreover, microbial community investigation would be performed by FISH techniques or next-generation sequencing. Regarding the degradation of the drug of interest, the detection and quantification of drug metabolites produced following an interaction with the gut microbiota, if known, could be targeted by UPLC-MS/MS. The correlation between drug degradation, drug-metabolite appearance and microbial community would not only allow an individual microbiome's capacity to metabolise a drug to be determined but it would also provide biomolecular mechanistic insights into how it is metabolised and what microbial perturbation it induces. Such a tool would enable pharmaceutical companies to systematically assess the potential interaction that a new drug candidate has with the gut microbiota. The increasing number of drugs found to have their metabolism influenced by the gut microbiota (Zimmermann et al., 2019a) or which have been found to influence the microbiota structure (Maier et al., 2018) prove the necessity of including the gut microbiota in toxicological assessment for future drugs.

Future perspectives are also envisioned considering the manipulation of the gut microbiota as a therapeutic strategy to limit adverse drug reactions. The selective inhibition of β-glucuronidase to limit irinotecan has demonstrated that well-designed microbiome-targeted treatments could be adapted and applied to other drugs to improve their outcomes. For example, bacterial *p*-cresol formation has been shown to influence the hepatic sulfation pool. Manipulating this production is an attractive target and one that is being pursued further. Indeed, additional research performed with collaborators has identified that *Clostridium difficile* produces *p*-cresol to provide the pathogen with a growth advantage over other bacteria facilitating its persistence in vulnerable individuals (Passmore et al., 2018). As such, selective inhibition of the enzyme producing *p*-cresol could not only prevent sulfation depletion but also prevent *C. diff* overgrowth, which could have life-changing application in the treatment of *C. difficile* infections. Manipulation of the gut microbiota structure could also be performed by administrating pre- and/or probiotics. Prebiotics are nutrients to promote the growth of beneficial bacteria in the gut and probiotics are beneficial bacteria directly administrated to the patient. A combined use of pre- and probiotics is called synbiotics. Their use has already been of growing interest to reduce disease symptoms such as inflammatory bowel disease (Bernstein, 2014). It has also been investigated to prevent cancer and in chemotherapy (Scott et al., 2018). It could eventually promote the growth of beneficial bacteria over bacteria that can be a threat to induce drug toxicity. However, the variability in individual microbiome is such that researchers face difficulties to clearly conclude the long-term beneficial effects of pre- and probiotics and this area will require further investigation before it can be routinely applied in clinical settings.

# 4. General conclusion

The influence of the gut microbiota on therapeutic treatments is not negligible, neither is the influence of therapeutic treatments on microbiome diversity and functionality. Having a better understanding of this bidirectional interaction is essential to advance to personalised medicine and improve human health. Not only will this enhance the effectiveness of drug therapy, it will also limit drug toxicity. The combined use of omics techniques has been proven to be a useful tool to explore this tripartite interaction and constant progress in expanding chemometrics should open up promising perspectives to deepen our understanding of the role of the intestinal microbiota in therapeutic treatment.

# References


Allegra, C.J., Chabner, B.A., Drake, J.C., Lutz, R., Rodbard, D., Jolivet, J., 1985. Enhanced inhibition of thymidylate synthase by methotrexate polyglutamates. J. Biol. Chem. 260, 9720–9726.

Allegra, C.J., Hoang, K., Yeh, G.C., Drake, J.C., Baram, J., 1987. Evidence for direct inhibition of de novo purine synthesis in human MCF-7 breast cells as a principal mode of metabolic inhibition by methotrexate. J. Biol. Chem. 262, 13520–13526.

Ambavaram, V.B.R., Nandigam, V., Vemula, M., Kalluru, G.R., Gajulapalle, M., 2013. Liquid chromatography–tandem mass spectrometry method for simultaneous quantification of urapidil and aripiprazole in human plasma and its application to human pharmacokinetic study. Biomed. Chromatogr. 27, 916–923. https://doi.org/10.1002/bmc.2882

Amidon, G.L., Lennernäs, H., Shah, V.P., Crison, J.R., 1995. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm. Res. 12, 413–420.

Beh, S.C., Kildebeck, E., Narayan, R., Desena, A., Schell, D., Rowe, E.S., Rowe, V., Burns, D., Whitworth, L., Frohman, T.C., Greenberg, B., Frohman, E.M., 2017. High-dose methotrexate with leucovorin rescue: For monumentally severe CNS inflammatory syndromes. J. Neurol. Sci. 372, 187–195. https://doi.org/10.1016/j.jns.2016.11.012

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benktander, J., Barone, A., Johansson, M.M., Teneberg, S., 2018. Helicobacter pylori SabA binding gangliosides of human stomach. Virulence 9, 738–751. https://doi.org/10.1080/21505594.2018.1440171

Benton, H.P., Want, E.J., Ebbels, T.M.D., 2010. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. Bioinforma. Oxf. Engl. 26, 2488–2489. https://doi.org/10.1093/bioinformatics/btq441

Bernstein, C.N., 2014. Antibiotics, probiotics and prebiotics in IBD. Nestle Nutr. Inst. Workshop Ser. 79, 83–100. https://doi.org/10.1159/000360713

Bhatt, A.P., Pellock, S.J., Biernat, K.A., Walton, W.G., Wallace, B.D., Creekmore, B.C., Letertre, M.M., Swann, J.R., Wilson, I.D., Roques, J.R., Darr, D.B., Bailey, S.T., Montgomery, S.A., Roach, J.M., Azcarate-Peril, M.A., Sartor, R.B., Gharaibeh, R.Z., Bultman, S.J., Redinbo, M.R., 2020. Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1918095117

Blandino, G., Inturri, R., Lazzara, F., Di Rosa, M., Malaguarnera, L., 2016. Impact of gut microbiota on diabetes mellitus. Diabetes Metab. 42, 303–315. https://doi.org/10.1016/j.diabet.2016.04.004

Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., Kraut, J., 1982. Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate. J. Biol. Chem. 257, 13650–13662.

Broadhurst, D., Goodacre, R., Reinke, S.N., Kuligowski, J., Wilson, I.D., Lewis, M.R., Dunn, W.B., 2018. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 14, 72. https://doi.org/10.1007/s11306-018-1367-3

Broadhurst, D.I., Kell, D.B., 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2, 171–196. https://doi.org/10.1007/s11306-006-0037-z

Bruggeman, F.J., Westerhoff, H.V., 2007. The nature of systems biology. Trends Microbiol. 15, 45–50. https://doi.org/10.1016/j.tim.2006.11.003

Buchen, S., Ngampolo, D., Melton, R.G., Hasan, C., Zoubek, A., Henze, G., Bode, U., Fleischhack, G., 2005. Carboxypeptidase G2 rescue in patients with methotrexate intoxication and renal failure. Br. J. Cancer 92, 480–487. https://doi.org/10.1038/sj.bjc.6602337

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119

Carmody, R.N., Turnbaugh, P.J., 2014. Host-microbial interactions in the metabolism of therapeutic and diet-derived xenobiotics. J. Clin. Invest. 124, 4173–4181. https://doi.org/10.1172/JCI72335

Chabner, B.A., Young, R.C., 1973. Threshold Methotrexate Concentration for In Vivo Inhibition of DNA Synthesis in Normal and Tumorous Target Tissues. J. Clin. Invest. 52, 1804–1811.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920. https://doi.org/10.1038/nbt.2377

Chen, J., Wright, K., Davis, J.M., Jeraldo, P., Marietta, E.V., Murray, J., Nelson, H., Matteson, E.L., Taneja, V., 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome Med. 8. https://doi.org/10.1186/s13073-016-0299-7

Clarke, G., Sandhu, K.V., Griffin, B.T., Dinan, T.G., Cryan, J.F., Hyland, N.P., 2019. Gut Reactions: Breaking Down Xenobiotic-Microbiome Interactions. Pharmacol. Rev. 71, 198–224. https://doi.org/10.1124/pr.118.015768

Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R., Nicholson, J.K., 2009. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc. Natl. Acad. Sci. U. S. A. 106, 14728–14733. https://doi.org/10.1073/pnas.0904489106

Clayton, T.A., Lindon, J.C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., Provost, J.-P., Le Net, J.-L., Baker, D., Walley, R.J., Everett, J.R., Nicholson, J.K., 2006. Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 440, 1073–1077. https://doi.org/10.1038/nature04648

Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcón, Ó., Lander, O., McDonald, J., Cox, M., Walter, J., Oh, P.L., Ruiz, J.F., Rodriguez, S., Shen, N., Song, S.J., Metcalf, J., Knight, R., Dantas, G., Dominguez-Bello, M.G., 2015. The microbiome of uncontacted Amerindians. Sci. Adv. 1. https://doi.org/10.1126/sciadv.1500183

Cloarec, O., Dumas, M.-E., Craig, A., Barton, R.H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J.C., Holmes, E., Nicholson, J., 2005. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. Anal. Chem. 77, 1282–1289. https://doi.org/10.1021/ac048630x

Cottret, L., Wildridge, D., Vinson, F., Barrett, M.P., Charles, H., Sagot, M.-F., Jourdan, F., 2010. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. Nucleic Acids Res. 38, W132-137. https://doi.org/10.1093/nar/gkq312

Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K., Lindon, J.C., 2006. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Anal. Chem. 78, 2262–2267. https://doi.org/10.1021/ac0519312

Cronstein, B.N., Naime, D., Ostad, E., 1993. The antiinflammatory mechanism of methotrexate. Increased adenosine release at inflamed sites diminishes leukocyte accumulation in an in vivo model of inflammation. J. Clin. Invest. 92, 2675–2682. https://doi.org/10.1172/JCI116884

Curtius, H.C., Mettler, M., Ettlinger, L., 1976. Study of the intestinal tyrosine metabolism using stable isotopes and gas chromatography-mass spectrometry. J. Chromatogr. 126, 569–580.

Dahan, A., Miller, J.M., Amidon, G.L., 2009. Prediction of solubility and permeability class membership: provisional BCS classification of the world's top oral drugs. AAPS J. 11, 740–746. https://doi.org/10.1208/s12248-009-9144-x

Danner, M.-C., Robertson, A., Behrends, V., Reiss, J., 2019. Antibiotic pollution in surface fresh waters: Occurrence and effects. Sci. Total Environ. 664, 793–804. https://doi.org/10.1016/j.scitotenv.2019.01.406

Davenport, E.R., Sanders, J.G., Song, S.J., Amato, K.R., Clark, A.G., Knight, R., 2017. The human microbiome in evolution. BMC Biol. 15, 127. https://doi.org/10.1186/s12915-017-0454-7

de Loor, H., Bammens, B., Evenepoel, P., De Preter, V., Verbeke, K., 2005. Gas chromatographic-mass spectrometric analysis for measurement of p-cresol and its conjugated metabolites in uremic and normal serum. Clin. Chem. 51, 1535–1538. https://doi.org/10.1373/clinchem.2005.050781

Devlin, A.M., Ling, E., Peerson, J.M., Fernando, S., Clarke, R., Smith, A.D., Halsted, C.H., 2000. Glutamate carboxypeptidase II: a polymorphism associated with lower levels of serum folate and hyperhomocysteinemia. Hum. Mol. Genet. 9, 2837–2844. https://doi.org/10.1093/hmg/9.19.2837

Dieterle, F., Ross, A., Schlotterbeck, G., Senn, H., 2006. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. Anal. Chem. 78, 4281–4290. https://doi.org/10.1021/ac051632c

Ding, H.T., Taur, Y., Walkup, J.T., 2017. Gut Microbiota and Autism: Key Concepts and Findings. J. Autism Dev. Disord. 47, 480–489. https://doi.org/10.1007/s10803-016-2960-9

Dobson, P.D., Lanthaler, K., Oliver, S.G., Kell, D.B., 2009. Implications of the dominant role of transporters in drug uptake by cells. Curr. Top. Med. Chem. 9, 163–181. https://doi.org/10.2174/156802609787521616

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., Knight, R., 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. U. S. A. 107, 11971–11975. https://doi.org/10.1073/pnas.1002601107

Dona, A.C., Jiménez, B., Schäfer, H., Humpfer, E., Spraul, M., Lewis, M.R., Pearce, J.T.M., Holmes, E., Lindon, J.C., Nicholson, J.K., 2014. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal. Chem. 86, 9887–9894. https://doi.org/10.1021/ac5025039

Donaldson, G.P., Lee, S.M., Mazmanian, S.K., 2016. Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 14, 20–32. https://doi.org/10.1038/nrmicro3552

Donehower, R.C., Hande, K.R., Drake, J.C., Chabner, B.A., 1979. Presence of 2,4-diamino-N10-methylpteroic acid after high-dose methotrexate. Clin. Pharmacol. Ther. 26, 63–72. https://doi.org/10.1002/cpt197926163

Dumas, M.-E., Barton, R.H., Toye, A., Cloarec, O., Blancher, C., Rothwell, A., Fearnside, J., Tatoud, R., Blanc, V., Lindon, J.C., Mitchell, S.C., Holmes, E., McCarthy, M.I., Scott, J., Gauguier, D., Nicholson, J.K., 2006. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proc. Natl. Acad. Sci. U. S. A. 103, 12511–12516. https://doi.org/10.1073/pnas.0601056103

Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., Nicholls, A.W., Wilson, I.D., Kell, D.B., Goodacre, R., Consortium, T.H.S.M. (HUSERMET), 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat. Protoc. 6, 1060–1083. https://doi.org/10.1038/nprot.2011.335

Dunn, W.B., Lin, W., Broadhurst, D., Begley, P., Brown, M., Zelena, E., Vaughan, A.A., Halsall, A., Harding, N., Knowles, J.D., Francis-McIntyre, S., Tseng, A., Ellis, D.I., O'Hagan, S., Aarons, G.,

Benjamin, B., Chew-Graham, S., Moseley, C., Potter, P., Winder, C.L., Potts, C., Thornton, P., McWhirter, C., Zubair, M., Pan, M., Burns, A., Cruickshank, J.K., Jayson, G.C., Purandare, N., Wu, F.C.W., Finn, J.D., Haselden, J.N., Nicholls, A.W., Wilson, I.D., Goodacre, R., Kell, D.B., 2015. Molecular phenotyping of a UK population: defining the human serum metabolome. Metabolomics Off. J. Metabolomic Soc. 11, 9–26. https://doi.org/10.1007/s11306-014-0707-1

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A., 2005. Diversity of the Human Intestinal Microbial Flora. Science 308, 1635–1638. https://doi.org/10.1126/science.1110591

Eriksson, L., Byrne, T., Johansson, E., Trygg, J., Vikström, C., 2013. Multi- and Megavariate Data Analysis Basic Principles and Applications. Umetrics Academy.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., Tito, R.Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., Garcia, R., Tigchelaar, E.F., Eeckhaudt, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., Raes, J., 2016. Population-level analysis of gut microbiome variation. Science 352, 560–564. https://doi.org/10.1126/science.aad3503

Fanaro, S., Chierici, R., Guerrini, P., Vigi, V., 2003. Intestinal microflora in early infancy: composition and development. Acta Paediatr. Oslo Nor. 1992 Suppl. 91, 48–55.

Fang, M., Ivanisevic, J., Benton, H.P., Johnson, C.H., Patti, G.J., Hoang, L.T., Uritboonthai, W., Kurczy, M.E., Siuzdak, G., 2015. Thermal Degradation of Small Molecules: A Global Metabolomic Investigation. Anal. Chem. 87, 10935–10941. https://doi.org/10.1021/acs.analchem.5b03003

Fiehn, O., 2002. Metabolomics--the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

Fijlstra, M., Ferdous, M., Koning, A.M., Rings, E.H.H.M., Harmsen, H.J.M., Tissing, W.J.E., 2014. Substantial decreases in the number and diversity of microbiota during chemotherapy-induced gastrointestinal mucositis in a rat model. Support. Care Cancer 23, 1513–1522. https://doi.org/10.1007/s00520-014-2487-6

Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdottir, V., Krogh Pedersen, H., Arumugam, M., Kristiansen, K., Yvonne Voigt, A., Vestergaard, H., Hercog, R., Igor Costea, P., Roat Kultima, J., Li, J., Jørgensen, T., Levenez, F., Dore, J., MetaHIT Consortium, Bjørn Nielsen, H., Brunak, S., Raes, J., Hansen, T., Wang, J., Dusko Ehrlich, S., Bork, P., Pedersen, O., 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528, 262–266. https://doi.org/10.1038/nature15766

Genestier, L., Paillot, R., Quemeneur, L., Izeradjene, K., Revillard, J.-P., 2000. Mechanisms of action of methotrexate. Immunopharmacology 47, 247–257. https://doi.org/10.1016/S0162-3109(00)00189-2

Gérard, P., 2016. Gut microbiota and obesity. Cell. Mol. Life Sci. CMLS 73, 147–162. https://doi.org/10.1007/s00018-015-2061-5

German, J.B., Freeman, S.L., Lebrilla, C.B., Mills, D.A., 2008. Human Milk Oligosaccharides: Evolution, Structures and Bioselectivity as Substrates for Intestinal Bacteria. Nestle Nutr. Workshop Ser. Paediatr. Programme 62, 205–222. https://doi.org/10.1159/000146322

Gika, H.G., Macpherson, E., Theodoridis, G.A., Wilson, I.D., 2008. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 871, 299–305. https://doi.org/10.1016/j.jchromb.2008.05.048

Gika, H.G., Theodoridis, G.A., Wingate, J.E., Wilson, I.D., 2007. Within-Day Reproducibility of an HPLC−MS-Based Method for Metabonomic Analysis: Application to Human Urine. J. Proteome Res. 6, 3291–3303. https://doi.org/10.1021/pr070183p

Gil de la Fuente, A., Godzien, J., Fernández López, M., Rupérez, F.J., Barbas, C., Otero, A., 2018. Knowledge-based metabolite annotation tool: CEU Mass Mediator. J. Pharm. Biomed. Anal. 154, 138–149. https://doi.org/10.1016/j.jpba.2018.02.046

Gil-de-la-Fuente, A., Godzien, J., Saugar, S., Garcia-Carmona, R., Badran, H., Wishart, D.S., Barbas, C., Otero, A., 2019. CEU Mass Mediator 3.0: A Metabolite Annotation Tool. J. Proteome Res. 18, 797–802. https://doi.org/10.1021/acs.jproteome.8b00720

Goldman, I.D., Matherly, L.H., 1985. The cellular pharmacology of methotrexate. Pharmacol. Ther. 28, 77–102. https://doi.org/10.1016/0163-7258(85)90083-X

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B., 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 22, 245–252. https://doi.org/10.1016/j.tibtech.2004.03.007

Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., Gordon, J.I., 2011. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc. Natl. Acad. Sci. U. S. A. 108, 6252–6257. https://doi.org/10.1073/pnas.1102938108

Goto, T., Myint, K.T., Sato, K., Wada, O., Kakiyama, G., Iida, T., Hishinuma, T., Mano, N., Goto, J., 2007. LC/ESI-tandem mass spectrometric determination of bile acid 3-sulfates in human urine 3beta-Sulfooxy-12alpha-hydroxy-5beta-cholanoic acid is an abundant nonamidated sulfate. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 846, 69–77. https://doi.org/10.1016/j.jchromb.2006.08.013

Graham, G.G., Davies, M.J., Day, R.O., Mohamudally, A., Scott, K.F., 2013. The modern pharmacology of paracetamol: therapeutic actions, mechanism of action, metabolism, toxicity and recent pharmacological findings. Inflammopharmacology 21, 201–232. https://doi.org/10.1007/s10787-013-0172-x

Gray, N., Adesina-Georgiadis, K., Chekmeneva, E., Plumb, R.S., Wilson, I.D., Nicholson, J.K., 2016. Development of a Rapid Microbore Metabolic Profiling Ultraperformance Liquid Chromatography–Mass Spectrometry Approach for High-Throughput Phenotyping Studies. Anal. Chem. 88, 5742–5751. https://doi.org/10.1021/acs.analchem.6b00038

Gryp, T., Vanholder, R., Vaneechoutte, M., Glorieux, G., 2017. p-Cresyl Sulfate. Toxins 9. https://doi.org/10.3390/toxins9020052

Gu, G.-Z., Xia, H.-M., Pang, Z.-Q., Liu, Z.-Y., Jiang, X.-G., Chen, J., 2011. Determination of sulphasalazine and its main metabolite sulphapyridine and 5-aminosalicylic acid in human plasma by liquid chromatography/tandem mass spectrometry and its application to a pharmacokinetic study. J. Chromatogr. B 879, 449–456. https://doi.org/10.1016/j.jchromb.2010.12.034

Guo, P., Wang, X., Liu, L., Belinsky, M.G., Kruh, G.D., Gallo, J.M., 2007. Determination of methotrexate and its major metabolite 7-hydroxymethotrexate in mouse plasma and brain tissue by liquid chromatography-tandem mass spectrometry. J. Pharm. Biomed. Anal. 43, 1789–1795. https://doi.org/10.1016/j.jpba.2006.12.034

Gupta, E., Lestingi, T.M., Mick, R., Ramirez, J., Vokes, E.E., Ratain, M.J., 1994. Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. Cancer Res. 54, 3723–3725.

Haiser, H.J., Turnbaugh, P.J., 2013. Developing a metagenomic view of xenobiotic metabolism. Pharmacol. Res. 69, 21–31. https://doi.org/10.1016/j.phrs.2012.07.009

Haiser, H.J., Turnbaugh, P.J., 2012. Is It Time for a Metagenomic Basis of Therapeutics? Science 336, 1253–1255. https://doi.org/10.1126/science.1224396

Hamad, I., Raoult, D., Bittar, F., 2016. Repertory of eukaryotes (eukaryome) in the human gastrointestinal tract: taxonomy and detection methods. Parasite Immunol. 38, 12–36. https://doi.org/10.1111/pim.12284

Hardwick, R.N., Clarke, J.D., Lake, A.D., Canet, M.J., Anumol, T., Street, S.M., Merrell, M.D., Goedken, M.J., Snyder, S.A., Cherrington, N.J., 2014. Increased Susceptibility to Methotrexate-Induced

Toxicity in Nonalcoholic Steatohepatitis. Toxicol. Sci. 142, 45–55. https://doi.org/10.1093/toxsci/kfu156

Heipieper, H.J., Keweloh, H., Rehm, H.J., 1991. Influence of phenols on growth and membrane permeability of free and immobilized Escherichia coli. Appl. Environ. Microbiol. 57, 1213–1217.

Heise, R., Mey, J., Neis, M.M., Marquardt, Y., Joussen, S., Ott, H., Wiederholt, T., Kurschat, P., Megahed, M., Bickers, D.R., Merk, H.F., Baron, J.M., 2006. Skin retinoid concentrations are modulated by CYP26AI expression restricted to basal keratinocytes in normal human skin and differentiated 3D skin models. J. Invest. Dermatol. 126, 2473–2480. https://doi.org/10.1038/sj.jid.5700432

Higdon, R., Earl, R.K., Stanberry, L., Hudac, C.M., Montague, E., Stewart, E., Janko, I., Choiniere, J., Broomall, W., Kolker, N., Bernier, R.A., Kolker, E., 2015. The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. Omics J. Integr. Biol. 19, 197–208. https://doi.org/10.1089/omi.2015.0020

Ho, K.-J., 1985. Human β-glucuronidase. Studies on the effects of pH and bile acids in regard to its role in the pathogenesis of cholelithiasis. Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol. 827, 197–206. https://doi.org/10.1016/0167-4838(85)90203-1

Hodgman, M.J., Garrard, A.R., 2012. A review of acetaminophen poisoning. Crit. Care Clin. 28, 499–516. https://doi.org/10.1016/j.ccc.2012.07.006

Holst-Jensen, A., Spilsberg, B., Arulandhu, A.J., Kok, E., Shi, J., Zel, J., 2016. Application of whole genome shotgun sequencing for detection and characterization of genetically modified organisms and derived products. Anal. Bioanal. Chem. 408, 4595–4614. https://doi.org/10.1007/s00216-016-9549-1

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441. https://doi.org/10.1037/h0071325

Hoyles, L., Fernández-Real, J.-M., Federici, M., Serino, M., Abbott, J., Charpentier, J., Heymes, C., Luque, J.L., Anthony, E., Barton, R.H., Chilloux, J., Myridakis, A., Martinez-Gili, L., Moreno-Navarrete, J.M., Benhamed, F., Azalbert, V., Blasco-Baque, V., Puig, J., Xifra, G., Ricart, W., Tomlinson, C., Woodbridge, M., Cardellini, M., Davato, F., Cardolini, I., Porzio, O., Gentileschi, P., Lopez, F., Foufelle, F., Butcher, S.A., Holmes, E., Nicholson, J.K., Postic, C., Burcelin, R., Dumas, M.-E., 2018. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat. Med. 24, 1070–1080. https://doi.org/10.1038/s41591-018-0061-3

Humblot, C., Murkovic, M., Rigottier-Gois, L., Bensaada, M., Bouclet, A., Andrieux, C., Anba, J., Rabot, S., 2007. Beta-glucuronidase in human intestinal microbiota is necessary for the colonic genotoxicity of the food-borne carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline in rats. Carcinogenesis 28, 2419–2425. https://doi.org/10.1093/carcin/bgm170

Ilkhanizadeh, B., Owji, A.A., Tavangar, S.M., Vasei, M., Tabei, S.M., 2001. Spot urine 5-hydroxy indole acetic acid and acute appendicitis. Hepatogastroenterology. 48, 609–613.

James, M.O., Ambadapadi, S., 2013. Interactions of cytosolic sulfotransferases with xenobiotics. Drug Metab. Rev. 45, 401–414. https://doi.org/10.3109/03602532.2013.835613

Jiang, C., Li, G., Huang, P., Liu, Z., Zhao, B., 2017. The Gut Microbiota and Alzheimer's Disease. J. Alzheimers Dis. JAD 58, 1–15. https://doi.org/10.3233/JAD-161141

Johansson, M.E.V., Phillipson, M., Petersson, J., Velcich, A., Holm, L., Hansson, G.C., 2008. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc. Natl. Acad. Sci. U. S. A. 105, 15064–15069. https://doi.org/10.1073/pnas.0803124105

Kaddurah-Daouk, R., Weinshilboum, R.M., 2014. Pharmacometabolomics: Implications for Clinical Pharmacology and Systems Pharmacology. Clin. Pharmacol. Ther. 95, 154–167. https://doi.org/10.1038/clpt.2013.217

Katsila, T., Konstantinou, E., Lavda, I., Malakis, H., Papantoni, I., Skondra, L., Patrinos, G.P., 2016. Pharmacometabolomics-aided Pharmacogenomics in Autoimmune Disease. EBioMedicine 5, 40–45. https://doi.org/10.1016/j.ebiom.2016.02.001

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., Gordon, J.I., 2011. Human nutrition, the gut microbiome, and immune system: envisioning the future. Nature 474, 327–336. https://doi.org/10.1038/nature10213

Kell, D.B., 2016. Implications of endogenous roles of transporters for drug discovery: hitchhiking and metabolite-likeness. Nat. Rev. Drug Discov. 15, 143–143. https://doi.org/10.1038/nrd.2015.44

Kell, D.B., 2015. The transporter-mediated cellular uptake of pharmaceutical drugs is based on their metabolite-likeness and not on their bulk biophysical properties: Towards a systems pharmacology. Perspect. Sci., Proceedings of the Beilstein Bozen Symposium 2014 – Chemistry and Time 6, 66–83. https://doi.org/10.1016/j.pisc.2015.06.004

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinforma. Oxf. Engl. 24, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Ketterer Brian, Coles Brian, Meyer David J., 1983. The role of glutathione in detoxication. Environ. Health Perspect. 49, 59–69. https://doi.org/10.1289/ehp.834959

Keweloh, H., Diefenbach, R., Rehm, H.-J., 1991. Increase of phenol tolerance of Escherichia coli by alterations of the fatty acid composition of the membrane lipids. Arch. Microbiol. 157, 49–53. https://doi.org/10.1007/BF00245334

Khosroshahi, H.T., Abedi, B., Ghojazadeh, M., Samadi, A., Jouyban, A., 2019. Effects of fermentable high fiber diet supplementation on gut derived and conventional nitrogenous product in patients on maintenance hemodialysis: a randomized controlled trial. Nutr. Metab. 16. https://doi.org/10.1186/s12986-019-0343-x

Kim, H.J., Li, H., Collins, J.J., Ingber, D.E., 2016. Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. Proc. Natl. Acad. Sci. U. S. A. 113, E7-15. https://doi.org/10.1073/pnas.1522193112

Kirschner, M., 2016. Systems Medicine: Sketching the Landscape. Methods Mol. Biol. Clifton NJ 1386, 3–15. https://doi.org/10.1007/978-1-4939-3283-2_1

Knight, R., Callewaert, C., Marotz, C., Hyde, E.R., Debelius, J.W., McDonald, D., Sogin, M.L., 2017. The Microbiome and Human Biology. Annu. Rev. Genomics Hum. Genet. 18, 65–86. https://doi.org/10.1146/annurev-genom-083115-022438

Kobayashi, T., Okada, A., Fujii, Y., Niimi, K., Hamamoto, S., Yasui, T., Tozawa, K., Kohri, K., 2010. The mechanism of renal stone formation and renal failure induced by administration of melamine and cyanuric acid. Urol. Res. 38, 117–125. https://doi.org/10.1007/s00240-010-0254-9

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kyriakides, M., Hardwick, R.N., Jin, Z., Goedken, M.J., Holmes, E., Cherrington, N.J., Coen, M., 2014. Systems level metabolic phenotype of methotrexate administration in the context of non-alcoholic steatohepatitis in the rat. Toxicol. Sci. Off. J. Soc. Toxicol. 142, 105–116. https://doi.org/10.1093/toxsci/kfu160

Kyriakides, M., Maitre, L., Stamper, B.D., Mohar, I., Kavanagh, T.J., Foster, J., Wilson, I.D., Holmes, E., Nelson, S.D., Coen, M., 2016. Comparative metabonomic analysis of hepatotoxicity induced by acetaminophen and its less toxic meta-isomer. Arch. Toxicol. 90, 3073–3085. https://doi.org/10.1007/s00204-015-1655-x

Lam, K.N., Alexander, M., Turnbaugh, P.J., 2019. Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. Cell Host Microbe 26, 22–34. https://doi.org/10.1016/j.chom.2019.06.011

Lawley, B., Otal, A., Moloney-Geany, K., Diana, A., Houghton, L., Heath, A.-L.M., Taylor, R.W., Tannock, G.W., 2019. Comparison of the development of the fecal microbiota of Indonesian and New Zealand children during the first year of life reveals differences in bifidobacterial taxa and microbiota complexity at 12 months. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.01105-19

Lewis, M.R., Pearce, J.T.M., Spagou, K., Green, M., Dona, A.C., Yuen, A.H.Y., David, M., Berry, D.J., Chappell, K., Horneffer-van der Sluis, V., Shaw, R., Lovestone, S., Elliott, P., Shockcor, J., Lindon, J.C., Cloarec, O., Takats, Z., Holmes, E., Nicholson, J.K., 2016. Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. Anal. Chem. 88, 9004–9013. https://doi.org/10.1021/acs.analchem.6b01481

Lindon, J.C., Holmes, E., Nicholson, J.K., 2003. Peer Reviewed: So What's the Deal with Metabonomics? Anal. Chem. 75, 384 A-391 A. https://doi.org/10.1021/ac031386+

Liu, C.M., Aziz, M., Kachur, S., Hsueh, P.-R., Huang, Y.-T., Keim, P., Price, L.B., 2012. BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. BMC Microbiol. 12, 56. https://doi.org/10.1186/1471-2180-12-56

LoGuidice, A., Wallace, B.D., Bendel, L., Redinbo, M.R., Boelsterli, U.A., 2012. Pharmacologic Targeting of Bacterial β-Glucuronidase Alleviates Nonsteroidal Anti-Inflammatory Drug-Induced Enteropathy in Mice. J. Pharmacol. Exp. Ther. 341, 447–454. https://doi.org/10.1124/jpet.111.191122

Ma, M.K., McLeod, H.L., 2003. Lessons learned from the irinotecan metabolic pathway. Curr. Med. Chem. 10, 41–49.

Macfarlane, G.T., Macfarlane, S., Gibson, G.R., 1998. Validation of a Three-Stage Compound Continuous Culture System for Investigating the Effect of Retention Time on the Ecology and Metabolism of Bacteria in the Human Colon. Microb. Ecol. 35, 180–187. https://doi.org/10.1007/s002489900072

Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., Patil, K.R., Bork, P., Typas, A., 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555, 623–628. https://doi.org/10.1038/nature25979

Maier, L., Typas, A., 2017. Systematically investigating the impact of medication on the gut microbiome. Curr. Opin. Microbiol. 39, 128–135. https://doi.org/10.1016/j.mib.2017.11.001

Manzoni, C., Kia, D.A., Vandrovcova, J., Hardy, J., Wood, N.W., Lewis, P.A., Ferrari, R., 2018. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief. Bioinform. 19, 286–302. https://doi.org/10.1093/bib/bbw114

Martinez, J.L., 2009. Environmental pollution by antibiotics and by antibiotic resistance determinants. Environ. Pollut. Barking Essex 1987 157, 2893–2902. https://doi.org/10.1016/j.envpol.2009.05.051

Marzorati, M., Vanhoecke, B., De Ryck, T., Sadaghian Sadabad, M., Pinheiro, I., Possemiers, S., Van den Abbeele, P., Derycke, L., Bracke, M., Pieters, J., Hennebel, T., Harmsen, H.J., Verstraete, W., Van de Wiele, T., 2014. The HMI™ module: a new tool to study the Host-Microbiota Interaction in the human gastrointestinal tract in vitro. BMC Microbiol. 14, 133. https://doi.org/10.1186/1471-2180-14-133

Masson, P., Spagou, K., Nicholson, J.K., Want, E.J., 2011. Technical and biological variation in UPLC-MS-based untargeted metabolic profiling of liver extracts: application in an experimental toxicity study on galactosamine. Anal. Chem. 83, 1116–1123. https://doi.org/10.1021/ac103011b

Mast, C., Lyan, B., Joly, C., Centeno, D., Giacomoni, F., Martin, J.-F., Mosoni, L., Dardevet, D., Pujos-Guillot, E., Papet, I., 2015. Assessment of protein modifications in liver of rats under chronic treatment with paracetamol (acetaminophen) using two complementary mass spectrometry-based metabolomic approaches. J. Proteomics 120, 194–203. https://doi.org/10.1016/j.jprot.2015.03.014

Masuda, N., 1981. Deconjugation of bile salts by Bacteroids and Clostridium. Microbiol. Immunol. 25, 1–11. https://doi.org/10.1111/j.1348-0421.1981.tb00001.x

Mathijssen, R.H., van Alphen, R.J., Verweij, J., Loos, W.J., Nooter, K., Stoter, G., Sparreboom, A., 2001. Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 7, 2182–2194.

Maurice, C.F., Haiser, H.J., Turnbaugh, P.J., 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell 152, 39–50. https://doi.org/10.1016/j.cell.2012.10.052

Mazaleuskaya, L.L., Sangkuhl, K., Thorn, C.F., FitzGerald, G.A., Altman, R.B., Klein, T.E., 2015. PharmGKB summary: pathways of acetaminophen metabolism at the therapeutic versus toxic doses. Pharmacogenet. Genomics 25, 416–426. https://doi.org/10.1097/FPC.0000000000000150

McCracken, K.W., Catá, E.M., Crawford, C.M., Sinagoga, K.L., Schumacher, M., Rockich, B.E., Tsai, Y.-H., Mayhew, C.N., Spence, J.R., Zavros, Y., Wells, J.M., 2014. Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature 516, 400–404. https://doi.org/10.1038/nature13863

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLOS Comput. Biol. 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P.J., Holmes, S., 2012. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac. Symp. Biocomput. Pac. Symp. Biocomput. 235–246.

McNulty, N.P., Yatsunenko, T., Hsiao, A., Faith, J.J., Muegge, B.D., Goodman, A.L., Henrissat, B., Oozeer, R., Cools-Portier, S., Gobert, G., Chervaux, C., Knights, D., Lozupone, C.A., Knight, R., Duncan, A.E., Bain, J.R., Muehlbauer, M.J., Newgard, C.B., Heath, A.C., Gordon, J.I., 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci. Transl. Med. 3, 106ra106. https://doi.org/10.1126/scitranslmed.3002701

Melnik, A.V., da Silva, R.R., Hyde, E.R., Aksenov, A.A., Vargas, F., Bouslimani, A., Protsyuk, I., Jarmusch, A.K., Tripathi, A., Alexandrov, T., Knight, R., Dorrestein, P.C., 2017. Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. Anal. Chem. 89, 7549–7559. https://doi.org/10.1021/acs.analchem.7b01381

Mertens, B.J.A., 2017. Transformation, Normalization, and Batch Effect in the Analysis of Mass Spectrometry Data for Omics Studies, in: Datta, S., Mertens, B.J.A. (Eds.), Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry, Frontiers in Probability and the Statistical Sciences. Springer International Publishing, Cham, pp. 1–21. https://doi.org/10.1007/978-3-319-45809-0_1

Morgan, X.C., Huttenhower, C., 2012. Chapter 12: Human Microbiome Analysis. PLOS Comput. Biol. 8, e1002808. https://doi.org/10.1371/journal.pcbi.1002808

Mueller, N.T., Bakacs, E., Combellick, J., Grigoryan, Z., Dominguez-Bello, M.G., 2015. The infant microbiome development: mom matters. Trends Mol. Med. 21, 109–117. https://doi.org/10.1016/j.molmed.2014.12.002

Mulak, A., Bonaz, B., 2015. Brain-gut-microbiota axis in Parkinson's disease. World J. Gastroenterol. 21, 10609–10620. https://doi.org/10.3748/wjg.v21.i37.10609

Mullish, B.H., Pechlivanis, A., Barker, G.F., Thursz, M.R., Marchesi, J.R., McDonald, J.A.K., 2018. Functional microbiomics: Evaluation of gut microbiota-bile acid metabolism interactions in

health and disease. Methods, Functional Microbiomics 149, 49–58. https://doi.org/10.1016/j.ymeth.2018.04.028

Nardone, G., Compare, D., 2015. The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? United Eur. Gastroenterol. J. 3, 255–260. https://doi.org/10.1177/2050640614566846

Naruhashi, K., Nadai, M., Nakao, M., Suzuki, N., Nabeshima, T., Hasegawa, T., 2000. Changes In Absorptive Function Of Rat Intestine Injured By Methotrexate. Clin. Exp. Pharmacol. Physiol. 27, 980–986. https://doi.org/10.1046/j.1440-1681.2000.03380.x

Nathan, P.C., Whitcomb, T., Wolters, P.L., Steinberg, S.M., Balis, F.M., Brouwers, P., Hunsberger, S., Feusner, J., Sather, H., Miser, J., Odom, L.F., Poplack, D., Reaman, G., Bleyer, W.A., 2006. Very high-dose methotrexate (33.6 g/m(2)) as central nervous system preventive therapy for childhood acute lymphoblastic leukemia: results of National Cancer Institute/Children's Cancer Group trials CCG-191P, CCG-134P and CCG-144P. Leuk. Lymphoma 47, 2488–2504. https://doi.org/10.1080/10428190600942769

Nayak, R.R., Alexander, M., Stapleton-Grey, K., Ubeda, C., Scher, J.U., Turnbaugh, P.J., 2019. Perturbation of the human gut microbiome by a non-antibiotic drug contributes to the resolution of autoimmune disease. bioRxiv 600155. https://doi.org/10.1101/600155

Neavin, D., Kaddurah-Daouk, R., Weinshilboum, R., 2016. Pharmacometabolomics informs Pharmacogenomics. Metabolomics Off. J. Metabolomic Soc. 12. https://doi.org/10.1007/s11306-016-1066-x

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-Gut Microbiota Metabolic Interactions. Science 336, 1262–1267. https://doi.org/10.1126/science.1223813

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica Fate Foreign Compd. Biol. Syst. 29, 1181–1189. https://doi.org/10.1080/004982599238047

Nicholson, J.K., Wilson, I.D., Lindon, J.C., 2010. Pharmacometabonomics as an effector for personalized medicine. Pharmacogenomics 12, 103–111. https://doi.org/10.2217/pgs.10.157

Niehues, M., Hensel, A., 2009. In-vitro interaction of L-dopa with bacterial adhesins of Helicobacter pylori: an explanation for clinicial differences in bioavailability? J. Pharm. Pharmacol. 61, 1303–1307. https://doi.org/10.1211/jpp/61.10.0005

Nye, L.C., Williams, J.P., Munjoma, N.C., Letertre, M.P.M., Coen, M., Bouwmeester, R., Martens, L., Swann, J.R., Nicholson, J.K., Plumb, R.S., McCullagh, M., Gethings, L.A., Lai, S., Langridge, J.I., Vissers, J.P.C., Wilson, I.D., 2019. A comparison of collision cross section values obtained via travelling wave ion mobility-mass spectrometry and ultra high performance liquid chromatography-ion mobility-mass spectrometry: Application to the characterisation of metabolites in rat urine. J. Chromatogr. A 1602, 386–396. https://doi.org/10.1016/j.chroma.2019.06.056

Oh, J., Yi, S., Gu, N., Shin, D., Yu, K.-S., Yoon, S.H., Cho, J.-Y., Jang, I.-J., 2018. Utility of Integrated Analysis of Pharmacogenomics and Pharmacometabolomics in Early Phase Clinical Trial: A Case Study of a New Molecular Entity. Genomics Inform. 16, 52–58. https://doi.org/10.5808/GI.2018.16.3.52

Orellana, G., Vanden Bussche, J., Van Meulebroek, L., Vandegehuchte, M., Janssen, C., Vanhaecke, L., 2014. Validation of a confirmatory method for lipophilic marine toxins in shellfish using UHPLC-HR-Orbitrap MS. Anal. Bioanal. Chem. 406, 5303–5312. https://doi.org/10.1007/s00216-014-7958-6

Ott, S.J., Kühbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., Drews, O., Weichert, W., Timmis, K.N., Schreiber, S., 2008. Fungi and inflammatory bowel diseases: Alterations of composition and diversity. Scand. J. Gastroenterol. 43, 831–841. https://doi.org/10.1080/00365520801935434

Paglia, G., Angel, P., Williams, J.P., Richardson, K., Olivos, H.J., Thompson, J.W., Menikarachchi, L., Lai, S., Walsh, C., Moseley, A., Plumb, R.S., Grant, D.F., Palsson, B.O., Langridge, J., Geromanos, S., Astarita, G., 2015. Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. Anal. Chem. 87, 1137–1144. https://doi.org/10.1021/ac503715v

Pannaraj, P.S., Li, F., Cerini, C., Bender, J.M., Yang, S., Rollie, A., Adisetiyo, H., Zabih, S., Lincez, P.J., Bittinger, K., Bailey, A., Bushman, F.D., Sleasman, J.W., Aldrovandi, G.M., 2017. Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. JAMA Pediatr. 171, 647–654. https://doi.org/10.1001/jamapediatrics.2017.0378

Park, G.-S., Park, M.H., Shin, W., Zhao, C., Sheikh, S., Oh, S.J., Kim, H.J., 2017. Emulating Host-Microbiome Ecosystem of Human Gastrointestinal Tract in Vitro. Stem Cell Rev. Rep. 13, 321–334. https://doi.org/10.1007/s12015-017-9739-z

Passmore, I.J., Letertre, M.P.M., Preston, M.D., Bianconi, I., Harrison, M.A., Nasher, F., Kaur, H., Hong, H.A., Baines, S.D., Cutting, S.M., Swann, J.R., Wren, B.W., Dawson, L.F., 2018. Para-cresol production by Clostridium difficile affects microbial diversity and membrane integrity of Gram-negative bacteria. PLOS Pathog. 14, e1007191. https://doi.org/10.1371/journal.ppat.1007191

Pauli, C., Hopkins, B.D., Prandi, D., Shaw, R., Fedrizzi, T., Sboner, A., Sailer, V., Augello, M., Puca, L., Rosati, R., McNary, T.J., Churakova, Y., Cheung, C., Triscott, J., Pisapia, D., Rao, R., Mosquera, J.M., Robinson, B., Faltas, B.M., Emerling, B.E., Gadi, V.K., Bernard, B., Elemento, O., Beltran, H., Dimichelis, F., Kemp, C.J., Grandori, C., Cantley, L.C., Rubin, M.A., 2017. Personalized In Vitro and In Vivo Cancer Models to Guide Precision Medicine. Cancer Discov. 7, 462–477. https://doi.org/10.1158/2159-8290.CD-16-1154

Pearson K., 1901. LIII. On lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin Philos. Mag. J. Sci. 2, 559–572. https://doi.org/10.1080/14786440109462720

Pellock, S.J., Creekmore, B.C., Walton, W.G., Mehta, N., Biernat, K.A., Cesmat, A.P., Ariyarathna, Y., Dunn, Z.D., Li, B., Jin, J., James, L.I., Redinbo, M.R., 2018. Gut Microbial β-Glucuronidase Inhibition via Catalytic Cycle Interception. ACS Cent. Sci. 4, 868–879. https://doi.org/10.1021/acscentsci.8b00239

Pellock, S.J., Redinbo, M.R., 2017. Glucuronides in the gut: Sugar-driven symbioses between microbe and host. J. Biol. Chem. 292, 8569–8576. https://doi.org/10.1074/jbc.R116.767434

Peppercorn, M.A., Goldman, P., 1972. The role of intestinal bacteria in the metabolism of salicylazosulfapyridine. J. Pharmacol. Exp. Ther. 181, 555–562.

Plumb, R.S., Rainville, P., Smith, B.W., Johnson, K.A., Castro-Perez, J., Wilson, I.D., Nicholson, J.K., 2006. Generation of ultrahigh peak capacity LC separations via elevated temperatures and high linear mobile-phase velocities. Anal. Chem. 78, 7278–7283. https://doi.org/10.1021/ac060935j

Pluskal, T., Castillo, S., Villar-Briones, A., Oresic, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11, 395. https://doi.org/10.1186/1471-2105-11-395

Posma, J.M., Garcia-Perez, I., Ebbels, T.M.D., Lindon, J.C., Stamler, J., Elliott, P., Holmes, E., Nicholson, J.K., 2018. Optimized Phenotypic Biomarker Discovery and Confounder Elimination via Covariate-Adjusted Projection to Latent Structures from Metabolic Spectroscopy Data. J. Proteome Res. 17, 1586–1595. https://doi.org/10.1021/acs.jproteome.7b00879

Prince, J.T., Marcotte, E.M., 2006. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal. Chem. 78, 6140–6152. https://doi.org/10.1021/ac0605344

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, Shaochuan, Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg,

A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, Shengting, Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, Songgang, Qin, N., Yang, H., Wang, Jian, Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium, Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Le Roux, K., Leclerc, M., Maguin, E., Melo Minardi, R., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., de Vos, W., Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, Jun, 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65. https://doi.org/10.1038/nature08821

Rattu, M.A., Shah, N., Lee, J.M., Pham, A.Q., Marzella, N., 2013. Glucarpidase (Voraxaze), a Carboxypeptidase Enzyme for Methotrexate Toxicity. Pharm. Ther. 38, 732–744.

Research, C. for D.E. and, 2019. Bioanalytical Method Validation Guidance for Industry [WWW Document]. US Food Drug Adm. URL http://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry (accessed 7.15.19).

Rey, F.E., Faith, J.J., Bain, J., Muehlbauer, M.J., Stevens, R.D., Newgard, C.B., Gordon, J.I., 2010. Dissecting the in vivo metabolic potential of two human gut acetogens. J. Biol. Chem. 285, 22082–22090. https://doi.org/10.1074/jbc.M110.117713

Rocke, D.M., Durbin, B., 2003. Approximate variance-stabilizing transformations for gene-expression microarray data. Bioinformatics 19, 966–972. https://doi.org/10.1093/bioinformatics/btg107

Rogers, M. a. M., Aronoff, D.M., 2016. The influence of non-steroidal anti-inflammatory drugs on the gut microbiome. Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. 22, 178.e1-178.e9. https://doi.org/10.1016/j.cmi.2015.10.003

Saccenti, E., Hoefsloot, H.C.J., Smilde, A.K., Westerhuis, J.A., Hendriks, M.M.W.B., 2014. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics 10, 361–374. https://doi.org/10.1007/s11306-013-0598-6

Sadaghian Sadabad, M., von Martels, J.Z.H., Khan, M.T., Blokzijl, T., Paglia, G., Dijkstra, G., Harmsen, H.J.M., Faber, K.N., 2015. A simple coculture system shows mutualism between anaerobic faecalibacteria and epithelial Caco-2 cells. Sci. Rep. 5, 17906. https://doi.org/10.1038/srep17906

Saha, J.R., Butler, V.P., Neu, H.C., Lindenbaum, J., 1983. Digoxin-inactivating bacteria: identification in human gut flora. Science 220, 325–327. https://doi.org/10.1126/science.6836275

Salvatore, S.R., Vitturi, D.A., Baker, P.R.S., Bonacci, G., Koenitzer, J.R., Woodcock, S.R., Freeman, B.A., Schopfer, F.J., 2013. Characterization and quantification of endogenous fatty acid nitroalkene metabolites in human urine. J. Lipid Res. 54, 1998–2009. https://doi.org/10.1194/jlr.M037804

Schmidt, T.M., DeLong, E.F., Pace, N.R., 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173, 4371–4378. https://doi.org/10.1128/jb.173.14.4371-4378.1991

Schmiegelow, K., 2009. Advances in individual prediction of methotrexate toxicity: a review. Br. J. Haematol. 146, 489–503. https://doi.org/10.1111/j.1365-2141.2009.07765.x

Scott, A.J., Merrifield, C.A., Younes, J.A., Pekelharing, E.P., 2018. Pre-, pro- and synbiotics in cancer prevention and treatment—a review of basic and clinical research. ecancermedicalscience 12. https://doi.org/10.3332/ecancer.2018.869

Sender, R., Fuchs, S., Milo, R., 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLOS Biol. 14, e1002533. https://doi.org/10.1371/journal.pbio.1002533

Shah, P., Fritz, J.V., Glaab, E., Desai, M.S., Greenhalgh, K., Frachet, A., Niegowska, M., Estes, M., Jäger, C., Seguin-Devaux, C., Zenhausern, F., Wilmes, P., 2016. A microfluidics-based in vitro model of the gastrointestinal human–microbe interface. Nat. Commun. 7. https://doi.org/10.1038/ncomms11535

Sharma, C.V., Mehta, V., 2013. Paracetamol: mechanisms and updates. Contin. Educ. Anaesth. Crit. Care Pain mkt049. https://doi.org/10.1093/bjaceaccp/mkt049

Shea, B., Swinden, M.V., Tanjong Ghogomu, E., Ortiz, Z., Katchamart, W., Rader, T., Bombardier, C., Wells, G.A., Tugwell, P., 2013. Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis. Cochrane Database Syst. Rev. CD000951. https://doi.org/10.1002/14651858.CD000951.pub2

Shiba, T., Kawakami, K., Sasaki, T., Makino, I., Kato, I., Kobayashi, T., Uchida, K., Kaneko, K., 2014. Effects of intestinal bacteria-derived p-cresyl sulfate on Th1-type immune response in vivo and in vitro. Toxicol. Appl. Pharmacol. 274, 191–199. https://doi.org/10.1016/j.taap.2013.10.016

Sivsammye, G., Sims, H.V., 1990. Presumptive identification of Clostridium difficile by detection of p-cresol in prepared peptone yeast glucose broth supplemented with p-hydroxyphenylacetic acid. J. Clin. Microbiol. 28, 1851–1853.

Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. Ther. Drug Monit. 27, 747–751.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. 78, 779–787. https://doi.org/10.1021/ac051437y

Smith, E.A., Macfarlane, G.T., 1997. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. Microb. Ecol. 33, 180–188. https://doi.org/10.1007/s002489900020

Smith, G.E., Griffiths, L.A., 1974. Metabolism of N-acylated and O-alkylated drugs by the intestinal microflora during anaerobic incubation in vitro. Xenobiotica Fate Foreign Compd. Biol. Syst. 4, 477–487. https://doi.org/10.3109/00498257409052100

Sonnenburg, E.D., Sonnenburg, J.L., 2019. The ancestral and industrialized gut microbiota and implications for human health. Nat. Rev. Microbiol. 17, 383–390. https://doi.org/10.1038/s41579-019-0191-8

Srivastava, A., 2017. Automated Alignment of Mass Spectrometry Data Using Functional Geometry, in: Datta, S., Mertens, B.J.A. (Eds.), Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry, Frontiers in Probability and the Statistical Sciences. Springer International Publishing, Cham, pp. 23–43. https://doi.org/10.1007/978-3-319-45809-0_2

Stein, A., Voigt, W., Jordan, K., 2010. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther. Adv. Med. Oncol. 2, 51–63. https://doi.org/10.1177/1758834009355164

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics Off. J. Metabolomic Soc. 3, 211–221. https://doi.org/10.1007/s11306-007-0082-2

Svahn, T., Mellgren, K., Harila-Saari, A., Åsberg, A., Kanerva, J., Jónsson, Ó., Vaitkeviciene, G., Stamm Mikkelssen, T., Schmiegelow, K., Heldrup, J., 2017. Delayed elimination of high-dose methotrexate and use of carboxypeptidase G2 in pediatric patients during treatment for acute lymphoblastic leukemia. Pediatr. Blood Cancer 64, n/a-n/a. https://doi.org/10.1002/pbc.26395

Tautenhahn, R., Böttcher, C., Neumann, S., 2008. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics 9, 504. https://doi.org/10.1186/1471-2105-9-504

Tautenhahn, R., Patti, G.J., Rinehart, D., Siuzdak, G., 2012. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal. Chem. 84, 5035–5039. https://doi.org/10.1021/ac300698c

Teubner, W., Meinl, W., Florian, S., Kretzschmar, M., Glatt, H., 2007. Identification and localization of soluble sulfotransferases in the human gastrointestinal tract. Biochem. J. 404, 207–215. https://doi.org/10.1042/BJ20061431

Thorsen, J., Brejnrod, A., Mortensen, M., Rasmussen, M.A., Stokholm, J., Al-Soud, W.A., Sørensen, S., Bisgaard, H., Waage, J., 2016. Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome 4, 62. https://doi.org/10.1186/s40168-016-0208-8

Tiwari, G., Tiwari, R., 2010. Bioanalytical method validation: An updated review. Pharm. Methods 1, 25–38. https://doi.org/10.4103/2229-4708.72226

Tobin, P.J., Dodds, H.M., Clarke, S., Schnitzler, M., Rivory, L.P., 2003. The relative contributions of carboxylesterase and beta-glucuronidase in the formation of SN-38 in human colorectal tumours. Oncol. Rep. 10, 1977–1979.

Toda, T., Saito, N., Ikarashi, N., Ito, K., Yamamoto, M., Ishige, A., Watanabe, K., Sugiyama, K., 2009. Intestinal flora induces the expression of Cyp3a in the mouse liver. Xenobiotica Fate Foreign Compd. Biol. Syst. 39, 323–334. https://doi.org/10.1080/00498250802651984

Truntzer, C., Ducoroy, P., 2017. Statistical Approach for Biomarker Discovery Using Label-Free LC-MS Data: An Overview, in: Datta, S., Mertens, B.J.A. (Eds.), Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry, Frontiers in Probability and the Statistical Sciences. Springer International Publishing, Cham, pp. 177–201. https://doi.org/10.1007/978-3-319-45809-0_10

Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). J. Chemom. 16, 119–128. https://doi.org/10.1002/cem.695

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., Gordon, J.I., 2009a. A core gut microbiome in obese and lean twins. Nature 457, 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., Gordon, J.I., 2009b. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci. Transl. Med. 1, 6ra14. https://doi.org/10.1126/scitranslmed.3000322

Valerino, D.M., Johns, D.G., Zaharko, D.S., Oliverio, V.T., 1972. Studies of the metabolism of methotrexate by intestinal flora. I. Identification and study of biological properties of the metabolite 4-amino-4-deoxy-N 10 -methylpteroic acid. Biochem. Pharmacol. 21, 821–831.

Veselkov, K.A., Vingara, L.K., Masson, P., Robinette, S.L., Want, E., Li, J.V., Barton, R.H., Boursier-Neyret, C., Walther, B., Ebbels, T.M., Pelczer, I., Holmes, E., Lindon, J.C., Nicholson, J.K., 2011. Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. Anal. Chem. 83, 5864–5872. https://doi.org/10.1021/ac201065j

Wallace, B.D., Roberts, A.B., Pollet, R.M., Ingle, J.D., Biernat, K.A., Pellock, S.J., Venkatesh, M.K., Guthrie, L., O'Neal, S.K., Robinson, S.J., Dollinger, M., Figueroa, E., McShane, S.R., Cohen, R.D., Jin, J., Frye, S.V., Zamboni, W.C., Pepe-Ranney, C., Mani, S., Kelly, L., Redinbo, M.R., 2015. Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem. Biol. 22, 1238–1249. https://doi.org/10.1016/j.chembiol.2015.08.005

Wallace, B.D., Wang, H., Lane, K.T., Scott, J.E., Orans, J., Koo, J.S., Venkatesh, M., Jobin, C., Yeh, L.-A., Mani, S., Redinbo, M.R., 2010. Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. Science 330, 831–835. https://doi.org/10.1126/science.1191175

Wan, Y., Wang, F., Yuan, J., Li, J., Jiang, D., Zhang, J., Li, H., Wang, R., Tang, J., Huang, T., Zheng, J., Sinclair, A.J., Mann, J., Li, D., 2019. Effects of dietary fat on gut microbiota and faecal

metabolites, and their relationship with cardiometabolic risk factors: a 6-month randomised controlled-feeding trial. Gut 68, 1417–1429. https://doi.org/10.1136/gutjnl-2018-317609

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Want, E.J., Cravatt, B.F., Siuzdak, G., 2005. The expanding role of mass spectrometry in metabolite profiling and characterization. Chembiochem Eur. J. Chem. Biol. 6, 1941–1951. https://doi.org/10.1002/cbic.200500151

Want, E.J., Masson, P., Michopoulos, F., Wilson, I.D., Theodoridis, G., Plumb, R.S., Shockcor, J., Loftus, N., Holmes, E., Nicholson, J.K., 2013. Global metabolic profiling of animal and human tissues via UPLC-MS. Nat. Protoc. 8, 17–32. https://doi.org/10.1038/nprot.2012.135

Ward, J.L., Baker, J.M., Miller, S.J., Deborde, C., Maucourt, M., Biais, B., Rolin, D., Moing, A., Moco, S., Vervoort, J., Lommen, A., Schäfer, H., Humpfer, E., Beale, M.H., 2010. An inter-laboratory comparison demonstrates that [H]-NMR metabolite fingerprinting is a robust technique for collaborative plant metabolomic data collection. Metabolomics Off. J. Metabolomic Soc. 6, 263–273. https://doi.org/10.1007/s11306-010-0200-4

Weber, J.L., Myers, E.W., 1997. Human Whole-Genome Shotgun Sequencing. Genome Res. 7, 401–409. https://doi.org/10.1101/gr.7.5.401

Weber, R.J.M., Lawson, T.N., Salek, R.M., Ebbels, T.M.D., Glen, R.C., Goodacre, R., Griffin, J.L., Haug, K., Koulman, A., Moreno, P., Ralser, M., Steinbeck, C., Dunn, W.B., Viant, M.R., 2017. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. Metabolomics Off. J. Metabolomic Soc. 13, 12. https://doi.org/10.1007/s11306-016-1147-x

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27. https://doi.org/10.1186/s40168-017-0237-y

Weiss, S.J., Xu, Z., Amir, A., Peddada, S., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vazquez-Baeza, Y., Birmingham, A., Knight, R., 2015. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data (No. e1408). PeerJ PrePrints. https://doi.org/10.7287/peerj.preprints.1157v1

Widemann, B.C., Balis, F.M., Adamson, P.C., 1999. Dihydrofolate Reductase Enzyme Inhibition Assay for Plasma Methotrexate Determination Using a 96-Well Microplate Reader. Clin. Chem. 45, 223–228.

Widemann, B.C., Jayaprakash, N., Howard, S.C., Daugherty, C., Chauhan, N., King, T., Rush, J., 2012. Clinical trial and compassionate use experience with glucarpidase for methotrexate toxicity. J. Clin. Oncol. 30, 6530–6530. https://doi.org/10.1200/jco.2012.30.15_suppl.6530

Widemann, B.C., Sung, E., Anderson, L., Salzer, W.L., Balis, F.M., Monitjo, K.S., McCully, C., Hawkins, M., Adamson, P.C., 2000. Pharmacokinetics and Metabolism of the Methotrexate Metabolite 2,4-Diamino-N 10-methylpteroic Acid. J. Pharmacol. Exp. Ther. 294, 894–901.

Wilson, I.D., 2009. Drugs, bugs, and personalized medicine: Pharmacometabonomics enters the ring. Proc. Natl. Acad. Sci. 106, 14187–14188. https://doi.org/10.1073/pnas.0907721106

Wilson, I.D., Nicholson, J.K., 2016. Gut microbiome interactions with drug metabolism, efficacy, and toxicity. Transl. Res. J. Lab. Clin. Med. https://doi.org/10.1016/j.trsl.2016.08.002

Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., Scalbert, A., 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 46, D608–D617. https://doi.org/10.1093/nar/gkx1089

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst., PLS Methods 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Zaharko, D.S., Bruckner, H., Oliverio, V.T., 1969. Antibiotics alter methotrexate metabolism and excretion. Science 166, 887–888.

Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D., Halsall, A., HUSERMET Consortium, Wilson, I.D., Kell, D.B., 2009. Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. Anal. Chem. 81, 1357–1364. https://doi.org/10.1021/ac8019366

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, Junhua, Tang, L., Li, Yin, Lan, Z., Chen, B., Li, Yanli, Zhong, H., Xie, H., Jie, Z., Chen, Weineng, Tang, S., Xu, Xiaoqiang, Wang, X., Cai, X., Liu, S., Xia, Y., Li, Jiyang, Qiao, X., Al-Aama, J.Y., Chen, H., Wang, L., Wu, Q., Zhang, F., Zheng, W., Li, Yongzhe, Zhang, M., Luo, G., Xue, W., Xiao, L., Li, Jun, Chen, Wanting, Xu, Xun, Yin, Y., Yang, H., Wang, Jian, Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Yingrui, Wang, Jun, 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat. Med. 21, 895–905. https://doi.org/10.1038/nm.3914

Zheng, X., Zhao, A., Xie, G., Chi, Y., Zhao, L., Li, H., Wang, C., Bao, Y., Jia, Weiping, Luther, M., Su, M., Nicholson, J.K., Jia, Wei, 2013. Melamine-induced renal toxicity is mediated by the gut microbiota. Sci. Transl. Med. 5, 172ra22. https://doi.org/10.1126/scitranslmed.3005114

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., Goodman, A.L., 2019a. Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature 570, 462–467. https://doi.org/10.1038/s41586-019-1291-3

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., Goodman, A.L., 2019b. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. Science 363. https://doi.org/10.1126/science.aat9931

Zoetendal, E.G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C.C.G.M., Troost, F.J., Bork, P., Wels, M., de Vos, W.M., Kleerebezem, M., 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J. 6, 1415–1426. https://doi.org/10.1038/ismej.2011.212

# Appendices

## **Supplementary data for Chapter 2:** Analytical Background

### Supplementary information for the paragraph 1.3. Targeted UPLC-MS

**Chemicals and Reagents.** L-tyrosine, *p*-ethyl phenol, *p*-coumaric acid, 4-hydroxybenzoic acid, *p*-cresol were obtained from Fisher Scientific Ltd (Loughborough, U.K.), tyramine, 4-hydroxyphenylpropionic acid, 4-hydroxyphenylacetic acid and phenol from Sigma Aldrich (Gillingham, U.K.) whilst 4-hydroxyphenyllactic acid were purchased from Flurochem (Hadfield, U.K.). *P*-cresol glucuronide, and the deuterated internal standards *p*-tyramine-$d_4$-hydrochloride, *p*-cresol sulfate-$d_7$ potassium salt and *p*-cresol glucuronide-$d_7$ were obtained from Toronto Research Chem (Toronto, Canada) whilst p-cresol-2,3,5,6-$d_4$ was sourced from Sigma-Aldrich (Gillingham, U.K.) and L-tyrosine-$d_7$ from Cambridge Isotope Laboratories (Andover, USA). Dansyl chloride, leucine enkephalin acetate salt (hydrate) and sodium formate, were obtained from Sigma-Aldrich (Gillingham, U.K.). LC-MS Chromasolv acetonitrile (ACN), isopropanol (IPA) and methanol (MeOH) were purchased from Honeywell (Seelze, Germany) ammonium hydroxide (28% w/v in water) Optima LC/MS Grade $H_2O$ with 0.1% formic acid (FA) and analytical reagent grade formic acid were purchased from Fisher Scientific Ltd (Loughborough, U.K).

***Stock solution, Calibration curve and IS mixture preparation.*** The twelve compounds targeted in this assay were weighed individually at an initial concentration of 20 mM. The solvent used for dissolution was either 1 M HCl (tyrosine), ethanol (tyramine), acetonitrile (*p*-ethylphenol, *p*-coumaric acid, *p*-hydroxyphenylpropionic acid, *p*-hydroxyphenylbenzoic acid and *p*-hydroxyphenyllactic acid) or water (*p*-hydroxyphenylacetic acid, *p*-cresol, phenol, *p*-cresol sulfate and *p*-cresol glucuronide). An intermediate concentration stock solution was prepared at 10 µM in water and then further diluted using water to 50 nM.

The stable isotope-labelled (SIL) analogues *p*-cresol-$d_4$, L-tyrosine-$d_7$, *p*-tyramine-$d_4$-hydrochloride used as internal standards were individually dissolved in either water, 1 M HCl or ethanol respectively, whilst *p*-cresol sulfate-$d_7$-(potassium salt) and *p*-cresol glucuronide-$d_7$ were dissolved in methanol (at 5 mM except for L-tyrosine at 2.3 mM). Intermediate-IS stock solutions (3.2 µM and 20 µM) were made for L-tyrosine-$d_7$ by diluting in water.

Calibration curve (CC) and quality control (QC) solution preparation is shown in **Table S1**. All these solutions except the initial blank were also mixed with 420 µL of $NH_4OH$ (7 %; v/v; aq.) and 180 µL of dansyl chloride (12.5 mg/mL in acetone) to perform the derivatization.

A blank QC pool sample (QC-B) was prepared by mixing 20 µL of $H_2O$, 20 µL of QC biological sample and 60 µL of ACN: IPA (1:1, v/v). A biological QC sample with internal standard (QC-IS) was prepared by adding 20 µL of IS mix to 20 µL of the QC sample and 60µL of ACN: IPA (1:1; v/v). A pool biological QC sample with internal standards and standards (QC-IS-S) was prepared by mixing 30 µL of IS mix, standard mix (intermediate solution 1 at 10 µM) and 60µL of ACN: IPA (1:1; v/v). The rest of the protocol was the same as that for the samples.

**UHPLC-ESI-MS/MS Analysis.** The chromatographic separation was performed with Acquity UHPLC solvent manager system coupled to a 2777C autosampler (Waters, Wilmslow, U.K.) on an ACQUITY UPLC® BEH C8, 1.7 µm column (1 mm x 100 mm, Waters, Wilmslow, U.K.). Column and autosampler temperatures were 65 and 4°C respectively. The mobile phases were composed of water containing 0.1% FA (v/v) as solvent A and ACN:IPA (1:1) with 0.1% FA (v/v) as solvent B. Multilinear reversed-phase gradient elution, at 0.2 mL/min, was performed starting at 1% B and increasing (from 0.1 min) to 10% B at 0.75 min then to 25% B at 2.00 min, 60% B at 5.50 min, and 99% B at 5.60 min. This solvent composition was maintained until 6.50 min when was returned to the starting composition of 1% B at 6.60 min and held until 7.50 min for column re-equilibration. A sample volume of 1,000 µL was used with a weak wash solvent of $H_2O$: ACN (90:10 v:v) containing 0.2% FA (v/v) and a strong wash solvent of IPA with 0.5% FA (v/v) added. One weak solvent wash cycle, one strong solvent wash cycle and another weak solvent wash cycle were used between each injection to remove sample-derived injection system contamination.

The eluent from the column was directed to a Xevo-TQS mass spectrometer (Waters, Wilmslow, U.K.). Electrospray ionization (ESI) was used in positive ionization mode to detect the dansylated analytes with switching to negative ESI for the *p*-cresol sulfate and glucuronide conjugates. Apart from *p*-cresol-sulfate and glucuronide and their deuterated IS's, two MRM transitions were monitored for each compound, one for quantification (q) and one for confirmation (c). Details of the MRMs used, together with the cone voltages and the collision energies used for each analyte can be found in Table 2-1.

The desolvation gas used was nitrogen and the collision gas used was argon. The source temperature was set to 150°C and the desolvation temperature to 550°C. The capillary voltage was fixed at 2.50 kV for MRMs in positive ionization mode and 1.50 kV for MRMs in negative ionization mode. The cone gas flow was fixed to 150 L/hr, the desolvation gas flow at 1000 L/hr, the collision gas flow at 0.15 mL/min and the nebulizer gas flow to 7 Bar.

**Table 2-S1. Preparation of the calibration curve and QC solutions*.**

| ID | Intermediate standard solution used | V of inter sol (µL) | IS vol of 3.2µM (µL) | V of $H_2O$ (µL) | Derivatization reagent present | IS conc after deriv, (nM) | Conc Standard after Derivatization (nM) |
|---|---|---|---|---|---|---|---|
| Blank | / | / | / | 600 | No | 0 | 0 |
| D-blank | / | / | / | 600 | Yes | 0 | 0 |
| IS-blank | / | / | 150 | 450 | Yes | 400 | 0 |
| CC 1 | 2 | 2.4 | 150 | 447.6 | Yes | 400 | 0.1 |
| CC 2 | 2 | 12 | 150 | 438 | Yes | 400 | 0.5 |
| CC 3 | 2 | 24 | 150 | 426 | Yes | 400 | 1 |
| CC 4 | 2 | 120 | 150 | 330 | Yes | 400 | 5 |
| CC 5 | 2 | 240 | 150 | 210 | Yes | 400 | 10 |
| CC 6 | 2 | 360 | 150 | 90 | Yes | 400 | 15 |
| CC 7 | 1 | 2.4 | 150 | 447.6 | Yes | 400 | 20 |
| CC 8 | 1 | 3 | 150 | 447 | Yes | 400 | 25 |
| CC 9 | 1 | 4.5 | 150 | 445.5 | Yes | 400 | 37.5 |
| CC 10 | 1 | 6 | 150 | 444 | Yes | 400 | 50 |
| CC 11 | 1 | 9 | 150 | 441 | Yes | 400 | 75 |
| CC 12 | 1 | 12 | 150 | 438 | Yes | 400 | 100 |
| CC 13 | 1 | 15 | 150 | 435 | Yes | 400 | 125 |
| CC 14 | 1 | 18 | 150 | 432 | Yes | 400 | 150 |
| CC 15 | 1 | 21 | 150 | 429 | Yes | 400 | 175 |
| CC 16 | 1 | 24 | 150 | 426 | Yes | 400 | 200 |
| CC 17 | 1 | 30 | 150 | 420 | Yes | 400 | 250 |
| CC 18 | 1 | 36 | 150 | 414 | Yes | 400 | 300 |
| MidQC | 1 | 48 | 150 | 402 | Yes | 400 | 400 |
| CC 19 | 1 | 60 | 150 | 390 | Yes | 400 | 500 |
| CC 20 | 1 | 72 | 150 | 378 | Yes | 400 | 600 |
| CC 21 | 1 | 84 | 150 | 366 | Yes | 400 | 700 |
| HighQC | 1 | 96 | 150 | 354 | Yes | 400 | 800 |
| CC 22 | 1 | 108 | 150 | 342 | Yes | 400 | 900 |
| CC 23 | 1 | 120 | 150 | 330 | Yes | 400 | 1000 |

**Middle QC and high QC are indicated but low QC varying from one compound to another it is not reported here but in Table 2-S2.*

**Table 2-S2. Concentrations of the calibration curve (CL = Calibration levels) and of the QCs in nM.**

| Compounds | CL1 LLOQ | CL2 | CL3 | CL4 | CL5 | CL6 | CL7 | CL8 ULOQ | QC low | QC mid | QC high |
|---|---|---|---|---|---|---|---|---|---|---|---|
| tyrosine | 10 | 25 | 50 | 100 | 250 | 600 | 900 | 1000 | 37.5 | 400 | 800 |
| tyramine | 5 | 25 | 50 | 100 | 250 | 600 | 900 | 1000 | 15 | 400 | 800 |
| *p*-ethyl phenol | 50 | 75 | 125 | 200 | 250 | 600 | 900 | 1000 | 150 | 400 | 800 |
| *p*-coumaric acid | 25 | 50 | 100 | 175 | 250 | 600 | 900 | 1000 | 75 | 400 | 800 |
| *p*-hydroxybenzoic acid | 37.5 | 75 | 150 | 200 | 300 | 600 | 900 | 1000 | 125 | 400 | 800 |
| *p*-hydroxyohenyl acetic acid | 25 | 50 | 100 | 175 | 250 | 600 | 900 | 1000 | 75 | 400 | 800 |
| *p*-hydroxyohenyl lactic acid | 37.5 | 75 | 150 | 200 | 300 | 600 | 900 | 1000 | 125 | 400 | 800 |
| *p*-cresol | 50 | 75 | 125 | 200 | 300 | 600 | 900 | 1000 | 150 | 400 | 800 |
| phenol | 100 | 125 | 150 | 175 | 250 | 600 | 900 | 1000 | 200 | 400 | 800 |
| *p*-cresol sulfate | 50 | 75 | 125 | 200 | 300 | 600 | 900 | 1000 | 150 | 400 | 800 |
| *p*-cresol glucuronide | 175 | 200 | 300 | 500 | 600 | 700 | 900 | 1000 | 250 | 400 | 800 |
| *p*-hydroxyphenyl propionic acid | 25 | 50 | 100 | 175 | 250 | 600 | 900 | 1000 | 75 | 400 | 800 |

**Table 2-S3. Linearity of the calibration curve obtained through the analysis of a minimum of three calibration curves per matrix and LLOQ.**

| Compounds | $R^2$ | LLOQ (nM) |
|---|---|---|
| tyrosine 1 | 0.999 | 10 |
| tyrosine 2 | 0.999 | 10 |
| tyramine | 0.998 | 5 |
| *p*-ethyl phenol | 0.972 | 50 |
| *p*-coumaric acid | 0.991 | 25 |
| *p*-hydroxybenzoic acid | 0.996 | 37.5 |
| *p*-hydroxyohenyl acetic acid | 0.997 | 25 |
| *p*-hydroxyphenyl lactic acid | 0.994 | 37.5 |
| *p*-cresol | 0.999 | 50 |
| phenol | 0.981 | 100 |
| *p*-cresol sulfate | 0.999 | 50 |
| *p*-cresol glucuronide | 0.967 | 175 |
| *p*-hydroxyphenyl propionic acid | 0.990 | 25 |

**Table 2-S4. Carry over results for analytes and SIL IS's**

| *Compounds* | *Area DB* | *Compared to (nM)* | *Area* | *Carry over (%)* |
|---|---|---|---|---|
| *tyrosine 1* | 257 | 10 | 2955 | 8.7 |
| *tyrosine 2* | 171 | 10 | 1161 | 14.7 |
| *tyramine* | 0 | 5 | 121 | 0.0 |
| *p-ethyl phenol* | 7019 | 50 | 47903 | 14.7 |
| *p-coumaric acid* | 1365 | 25 | 11703 | 11.7 |
| *p-hydroxybenzoic acid* | 2372 | 37.5 | 19612 | 12.1 |
| *p-hydroxyohenyl acetic acid* | 4777 | 25 | 38013 | 12.6 |
| *p-hydroxyohenyl lactic acid* | 2514 | 37.5 | 16701 | 15.1 |
| *p-cresol* | 5593 | 50 | 29432 | 19.0 |
| *phenol* | 1981 | 100 | 10735 | 18.5 |
| *p-cresol sulfate* | 648 | 50 | 4250 | 15.2 |
| *p-cresol glucuronide* | 0 | 175 | 16 | 0.0 |
| *p-hydroxyphenyl propionic acid* | 166 | 25 | 2133 | 7.8 |
| *tyrosine-$d_7$ 1* | 108 | 400 | 51115 | 0.2 |
| *tyrosine-$d_7$ 2* | 136 | 400 | 17809 | 0.8 |
| *tyramine-$d_4$* | 0 | 400 | 7942 | 0.0 |
| *p-cresol-$d_4$* | 294 | 400 | 22271 | 1.3 |
| *p-cresol sulfate-$d_7$* | 57 | 400 | 37398 | 0.2 |
| *p-cresol glucuronide-$d_7$* | 0 | 400 | 1007 | 0.0 |

**Table 2-S5. Analytical recovery of the internal standards spiked before and after the protein precipitation step for the QC plasma pool sample.**

| *Compounds* | *% Recovery* |
|---|---|
| *tyrosine-$d_7$* | 91.7 |
| *tyramine-$d_4$* | 84.0 |
| *p-cresol-$d_4$* | 75.2 |
| *p-cresol sulfate-$d_7$* | 101.3 |
| *p-cresol glucuronide-$d_7$* | 95.9 |

**Table 2-S6. Analyte to double blank interference.**

| *Compounds* | *Blank Interference (%)* |
|---|---|
| *tyrosine* | 17.9 |
| *tyramine* | 0.0 |
| *p-ethyl phenol* | 3.0 |
| *p-coumaric acid* | 0.0 |
| *p-hydroxybenzoic acid* | 1.1 |
| *p-hydroxyohenyl acetic acid* | 0.0 |
| *p-hydroxyohenyl lactic acid* | 4.1 |
| *p-cresol* | 1.2 |
| *phenol* | 3.5 |
| *p-cresol sulfate* | 2.9 |
| *p-cresol glucuronide* | 0.0 |
| *p-hydroxyphenyl propionic acid* | 0.9 |

**Supplementary data for Chapter 3:** A two-way interaction between methotrexate and the gut microbiota in Sprague Dawley rats

**Table 3-S1. Number of reads per samples after 16S rRNA gene sequencing pre-processing and prior importation into phyloseq for data investigation.**

| SAMPLE ID | READS COUNT |
|---|---|
| 10_C_01 | 13189 |
| 10_C_02 | 13680 |
| 11_C_01 | 13129 |
| 11_C_02 | 6182 |
| 11_C_04 | 15638 |
| 11_C_08 | 12801 |
| 12_C_01 | 16269 |
| 13_D_04 | 15603 |
| 14_D_01 | 17043 |
| 14_D_02 | 17909 |
| 14_D_08 | 14648 |
| 15_D_02 | 16451 |
| 16_D_02 | 15168 |
| 16_D_08 | 11959 |
| 17_E_01 | 12937 |
| 17_E_02 | 19481 |
| 17_E_08 | 10813 |
| 18_E_01 | 15943 |
| 18_E_08 | 14209 |
| 19_E_01 | 15241 |
| 19_E_08 | 16935 |
| 1_A_08 | 15753 |
| 20_E_01 | 19037 |
| 20_E_02 | 14600 |
| 21_F_01 | 16923 |
| 21_F_02 | 18220 |
| 22_F_01 | 18429 |
| 22_F_08 | 17227 |
| 23_F_01 | 18852 |
| 23_F_02 | 18097 |
| 23_F_08 | 18209 |
| 24_F_01 | 17406 |
| 24_F_02 | 4406 |
| 25_G_01 | 15517 |
| 25_G_02 | 16624 |
| 25_G_04 | 17014 |
| 25_G_08 | 6745 |
| 26_G_01 | 18234 |
| 26_G_02 | 18370 |
| 27_G_01 | 11336 |
| 27_G_04 | 13117 |
| 28_G_01 | 17482 |
| 28_G_02 | 15805 |
| 29_H_01 | 12397 |
| 29_H_04 | 14809 |
| 29_H_08 | 11919 |
| 2_A_02 | 18056 |
| 2_A_08 | 15282 |
| 30_U_01 | 16765 |
| 30_U_02 | 21982 |
| 31_H_01 | 15485 |
| 31_H_08 | 12619 |
| 32_H_01 | 13381 |
| 32_H_02 | 12724 |
| 33_I_02 | 14062 |
| 33_I_04 | 13899 |
| 33_I_08 | 18292 |
| 34_I_01 | 18632 |
| 34_I_02 | 14898 |
| 35_I_01 | 11371 |
| 35_I_02 | 12434 |
| 35_I_04 | 17798 |
| 35_I_08 | 16482 |
| 36_I_01 | 13159 |
| 36_I_02 | 14176 |
| 36_I_08 | 17722 |
| 37_I_01 | 10524 |
| 37_I_02 | 13151 |
| 38_I_01 | 14775 |
| 38_I_02 | 13278 |
| 39_J_02 | 10433 |
| 3_A_02 | 11592 |
| 40_J_01 | 17328 |
| 40_J_02 | 15795 |
| 41_J_01 | 12942 |
| 41_J_02 | 14064 |
| 42_J_01 | 17428 |
| 42_J_02 | 18626 |
| 42_J_08 | 19071 |
| 43_J_01 | 12622 |
| 43_J_04 | 11234 |
| 44_J_01 | 24371 |
| 44_J_02 | 14249 |
| 4_A_01 | 17464 |
| 4_A_02 | 13804 |
| 5_B_01 | 12107 |
| 5_B_02 | 15130 |
| 5_B_04 | 14601 |
| 5_B_08 | 15550 |
| 6_B_01 | 18907 |
| 7_B_02 | 15476 |
| 8_B_01 | 17688 |
| 8_B_02 | 12730 |
| 8_B_08 | 21445 |
| 9_C_01 | 11312 |
| 9_C_04 | 10702 |
| 9_C_08 | 15240 |
| BLANK-P1 181 | 53 |
| BLANK-P2 257 | 71 |
| BLANK-P3-V1 248 | 80 |
| BLANK-P3-V2 324 | 23 |
| POSITIVE-CONTROL 30404 | 23749 |

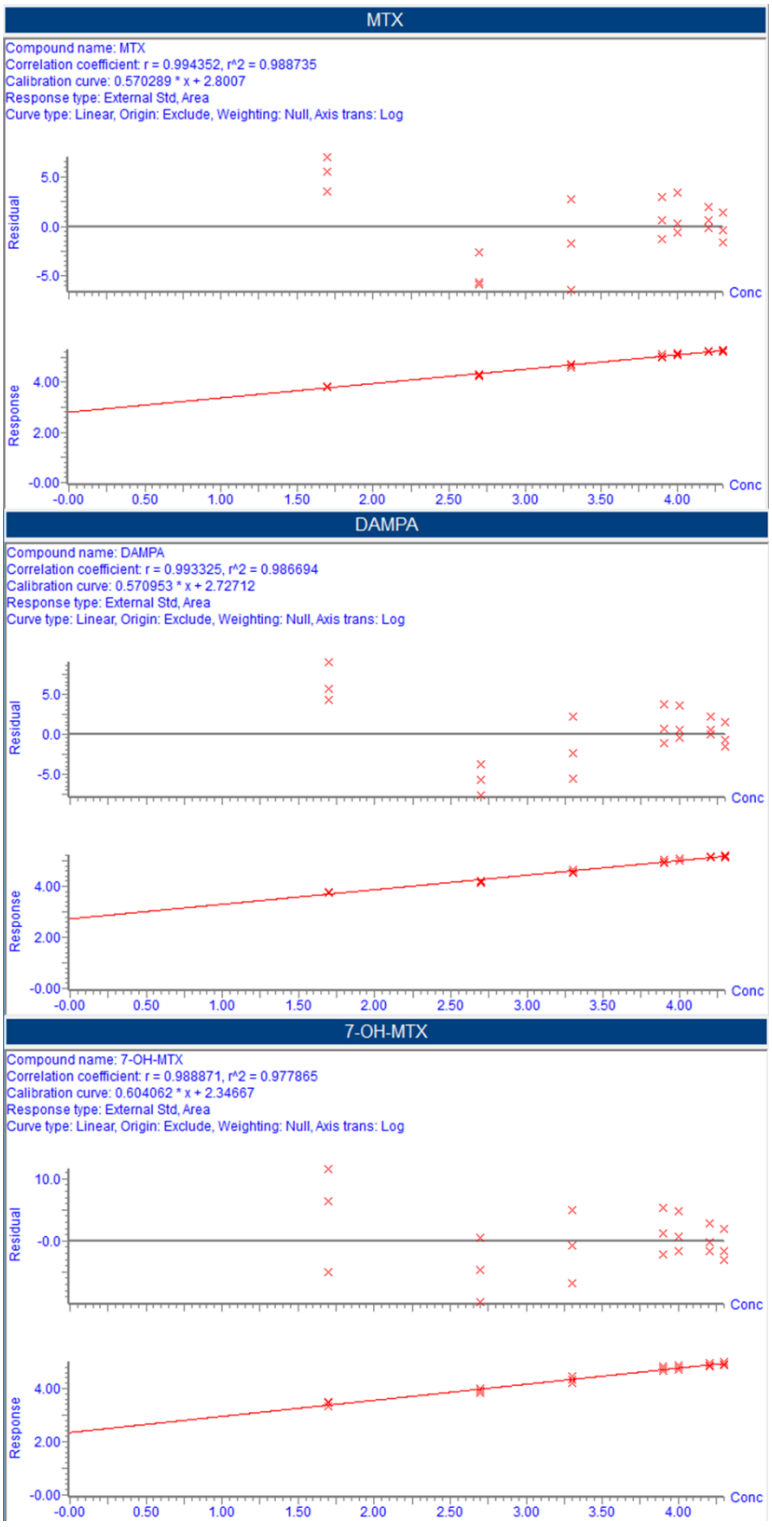



**Figure 3-S 1. Screenshot of TargetLynx V4.1 showing the three calibration curves for MTX, DAMPA and 7-OH-MTX used during the faecal sample analysis.** Linear curve with log axis was used, all the $R^2$ for each of the standards for both the urine and the faecal samples were superior to 0.97.

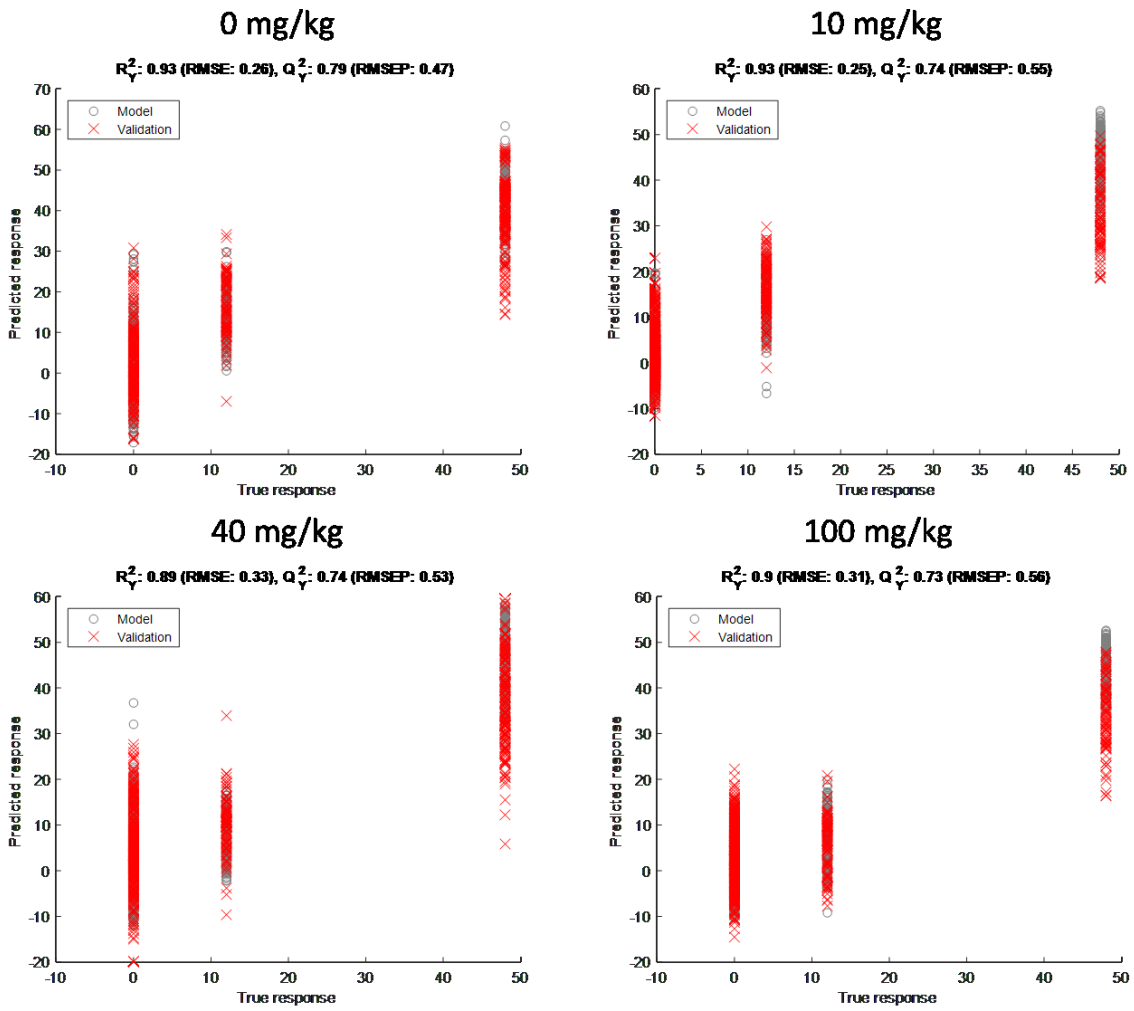



**Figure 3-S2. CA-PLS-R (covariate-adjusted projection to latent structures regression) analysis performed over time (0, 12 and 48h) on the faecal samples of each of the MTX dose groups.** $R^2Y$ indicate how well the model fits the data, the $Q^2Y$ how well the model predicts the data. Root mean squared error (RMSE) and root mean squared error of the prediction (RMSEP) represent the root mean squared error the root mean squared error of the prediction respectively. The model set is represented by grey circles and the predictions by red crosses.

**Table 3-S2. Putative annotations for metabolites detected in the urine samples.** The database reference number of each putative annotation is indicated, as well as if the matching was done with as spectra coming from real measurements or in silico calculation, and finally if validation of the annotation by MS/MS or DDA is reported as well as matches with the NPC database. Confidence of annotation are indicated in ID level **(Sumner et al., 2007)**. Frag = fragment; theo = theoretical; obs = observed

| putative ID | Adduct | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | Database | In silico or exp | MS/MS or DDA | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| arginine | $[M+H]^+$ | $C_6H_{14}N_4O_2$ | 2a | 175.0828 | 175.1187 | 158.093 | 130.098 | 116.071 | 112.087 | 158.057 | 130.053 | / | 112.056 | 0.61 | Metlin 13 | ESI Q-TOF | Yes | Yes |
| creatinine | $[M+Na]^+$ | $C_4H_7N_3O$ | 2a | 136.0504 | 136.0481 | 114.069 | 86.0741 | / | / | 114.068 | 86.0722 | / | / | 0.65 | Metlin 8 | ESI Q-TOF | / | Yes |
| creatine | $[M+H]^+$ | $C_4H_9N_3O_2$ | 2a | 132.0786 | 132.0768 | 114.066 | 90.0555 | 87.0557 | / | 114.068 | 90.0568 | 87.0566 | / | 0.67 | Metlin 7 | ESI Q-TOF | Yes | Yes |
| acetyl choline | $[M+H]^+$ | $C_7H_{15}NO_2$ | 2a | 146.12 | 146.118 | 87.0448 | 64.0165 | 60.0818 | / | 87.0459 | / | 60.0826 | / | 0.69 | Metlin 57 | ESI Q-TOF | / | Yes |
| ribothymidine | $[M+Na]^+$ | $C_{10}H_{14}N_2O_6$ | 2b | 281.0753 | 281.0744 | 259.093 | 133.048 | 127.05 | 110.023 | 259.094 | 133.052 | 127.052 | / | 0.95 | Metlin 85160 | ESI Q-TOF | Yes | / |
| N-acetyl-L-glutamine | $[M+Na]^+$ | $C_7H_{12}N_2O_4$ | 2a | 211.0711 | 211.0689 | 189.087 | 172.06 | 130.05 | 84.0444 | 189.089 | 172.061 | 130.052 | 84.045 | 1.07 | Metlin 58344 | ESI Q-TOF | Yes | Yes |
| acetylcarnosine | $[M+Na]^+$ | $C_{11}H_{16}N_4O_4$ | 2b | 291.1070 | 291.1064 | 269.124 | 251.114 | 209.058 | 156.077 | 269.126 | 251.115 | 209.058 | 156.077 | 1.08 | HMDB0012881 | In silico | Yes | / |
| N-acetyl-arginine | $[M+H]^+$ | $C_8H_{16}N_4O_3$ | 2b | 217.1320 | 217.1298 | 200.103 | 175.117 | 158.083 | / | 200.105 | 175.12 | 158.084 | / | 1.08 | Metlin 58241 | ESI Q-TOF | Yes | / |
| 1-methyladenosine | $[M+H]^+$ | $C_{11}H_{15}N_5O_4$ | 2b | 282.1216 | 282.1200 | 264.109 | 206.067 | 150.077 | / | / | 206.019 | 150.08 | / | 1.40 | Metlin 6888 | ESI Q-TOF | Yes | / |
| N-acetyl-L-glutamic Acid | $[M+Na]^+$ | $C_7H_{12}N_2O_4$ | 2a | 212.054 | 212.053 | 148.061 | 130.05 | 102.055 | 84.045 | 148.065 | 130.052 | 102.06 | 84.048 | 1.49 | Metlin 3325 | ESI Q-TOF | Yes | Yes |
| biopterin | $[M+Na]^+$ | $C_9H_{11}N_5O_3$ | 2b | 260.078 | 260.075 | 238.094 | 220.083 | 194.068 | 178.073 | 238.096 | 220.085 | 194.069 | 178.073 | 1.58 | Metlin 247 | ESI Q-TOF | Yes | / |
| thymidine | $[M+H]^+$ | $C_{10}H_{14}N_2O_5$ | 2a | 265.0830 | 265.0795 | 243.114 | 127.05 | 117.055 | 99.0439 | 243.056 | 127.054 | 117.057 | / | 2.25 | Metlin 3375 | ESI Q-TOF | / | Yes |
| succinoadenosine | $[M+H]^+$ | $C_{14}H_{17}N_5O_8$ | 2b | 384.1179 | 384.1150 | 252.073 | 236.078 | 206.067 | 192.052 | 252.076 | 236.08 | 206.067 | 192.053 | 2.45 | Metlin 5867 | In silico | Yes | / |
| xanthurenic acid | $[M+H]^+$ | $C_{10}H_7NO_4$ | 2b | 206.048 | 206.044 | 188.034 | 178.046 | 160.038 | 132.044 | 188.074 | 178.053 | 160.042 | 132.082 | 2.97 | Metlin 5841 | ESI Q-TOF | Yes | / |
| N6-carbamoyl-L-threonyladenosine | $[M+Na]^+$ | $C_{15}H_{20}N_6O_8$ | 2b | 435.1255 | 435.1235 | 413.142 | 281.099 | 136.062 | 102.016 | 413.143 | 281.101 | 136.064 | 102.055 | 3.14 | Metlin 95993 | In silico | Yes | / |
| pimelylcarnitine | $[M+H]^+$ | $C_{14}H_{25}NO_6$ | 2b | 304.178 | 304.176 | 286.165 | 268.154 | 258.17 | 240.159 | 286.174 | 268.11 | 258.18 | 239.997 | 3.15 | HMDB0013328 | In silico | Yes | / |
| methionine sulfoximine | $[M+H]^+$ | $C_5H_{12}N_2O_3S$ | 2b | 181.0630 | 181.0640 | 163.038 | 135.059 | 118.032 | 102.055 | 163.052 | 135.058 | 118.066 | 102.016 | 3.20 | Metlin 65992 | ESI Q-TOF | Yes | / |
| 5-Hydroxyindoleacetic acid | $[M+H]^+$ | $C_{10}H_9NO_3$ | 2b | 192.0676 | 192.0637 | 146.06 | 110.009 | / | / | 146.063 | 110.011 | / | / | 3.54 | Metlin 2975 | ESI Q-TOF | Yes | / |
| N-Methyl-4-aminobenzoate | $[M+H]^+$ | $C_8H_9NO_2$ | 2b | 152.0720 | 152.0710 | 134.0600 | 108.0810 | 106.0650 | 104.0490 | 134.0616 | 108.0823 | 106.0653 | 104.0100 | 3.66 | Metlin 65994 | In silico | Yes | / |

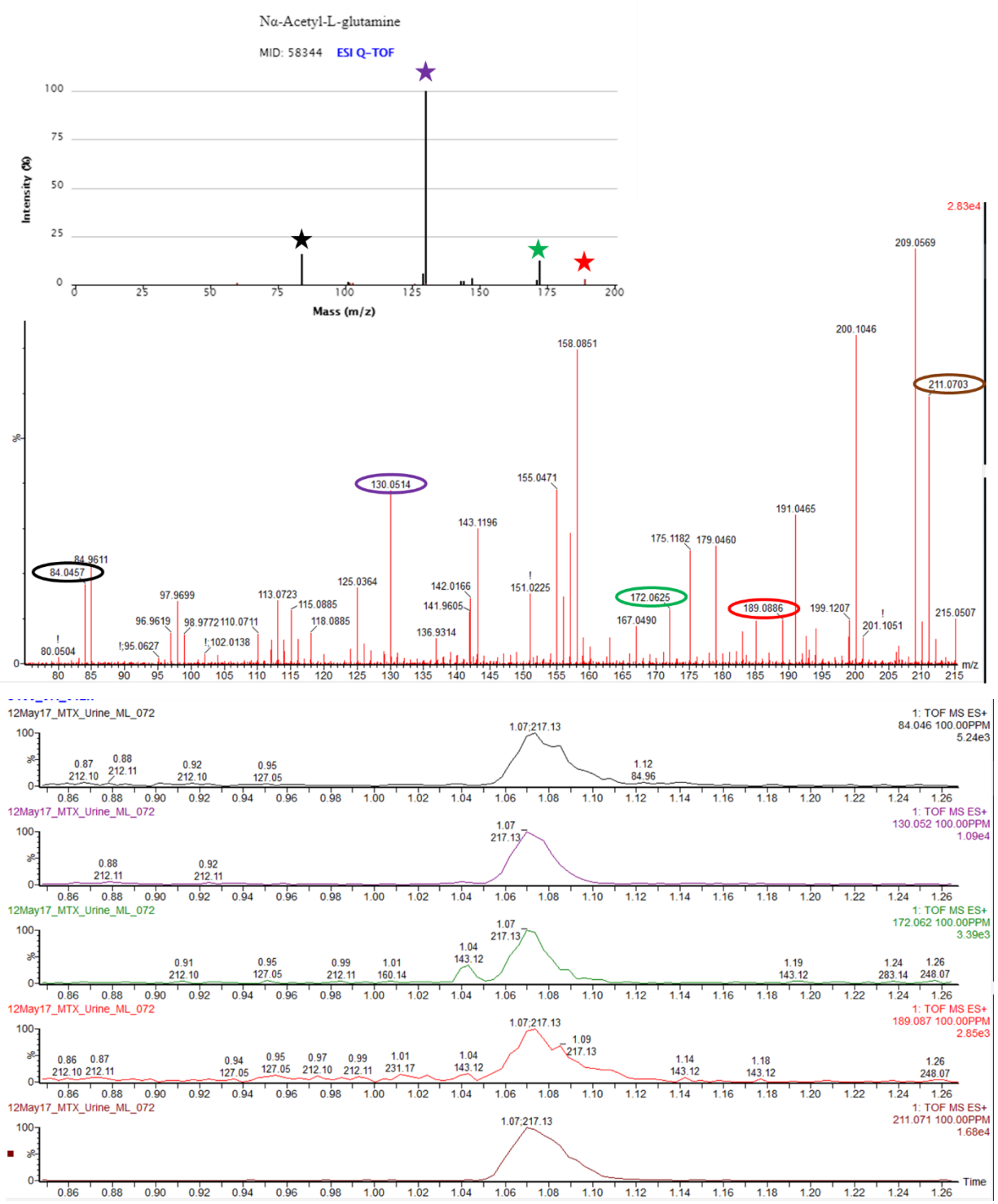



**Figure 3-S3. Example of *N*-acetyl glutamine identification in urine samples of rat treated with 100 mg/kg of MTX and collected in the 6-12h time window. The *m/z* 211.0711, eluting at 1.07 min was found to significantly influenced the samples separation in the supervised model. In house database matching (based on *m/z* and Tr comparison) resulted in a putative annotation of this *m/z* as the $[M+Na]^+$ adduct of *N*-acetyl glutamine. Validation of this putative annotation to a level 2a confidence was done by comparing its fragmentation pattern (*m/z* and intensities of the fragment) to the experimental spectra from Metlin (top left panel) to the raw data. Fragments of the molecule in the raw data show identical retention times.**

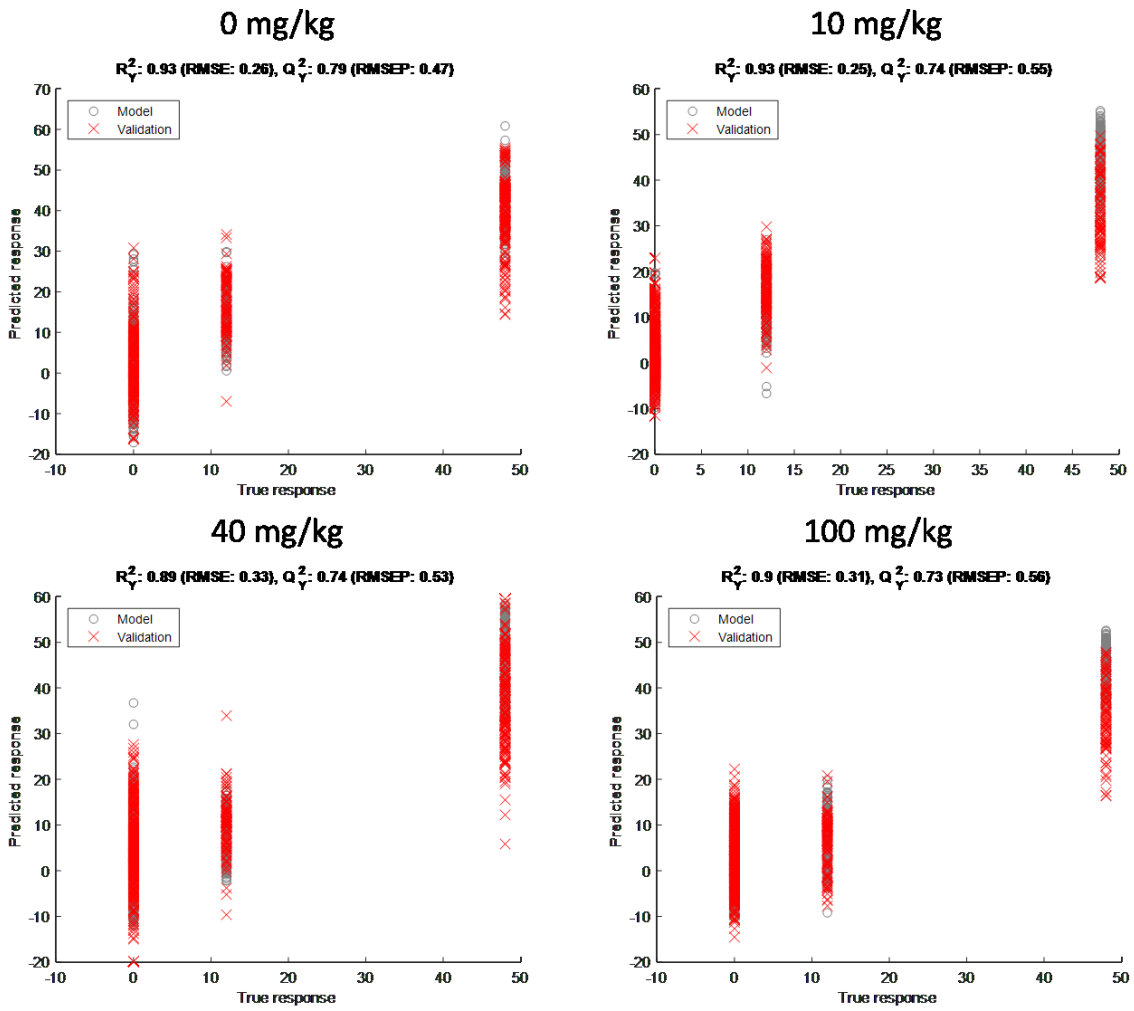



**Figure 3-S4. CA-PLS-R (covariate-adjusted projection to latent structures regression) analysis performed over time (0, 12 and 48h) on the faecal samples of each of the MTX dose groups.** $R^2$Y indicate how well the model fits the data, the $Q^2$Y how well the model predicts the data. RMSE and RMSEP represent the root mean squared error the root mean squared error of the prediction respectively. The model set is represented by grey circles and the predictions by red crosses.

**Table 3-S3. Putative annotations for metabolites detected in the faecal samples.** The database reference number of each putative annotation is indicated, as well as if the matching was done with as spectra coming from real measurements or in silico calculation, and finally if validation of the annotation by MS/MS or DDA is reported as well as matches with the NPC database. Confidence of annotation are indicated in ID level **(Sumner et al., 2007)**. Frag = fragment; theo = theoretical; obs = observed

| putative ID | adduct | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | Database | In silico or exp | MS/MS or DDA | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| carnitine | $[M+H]^+$ | $C_7H_{15}NO_3$ | 2a | 162.113 | 162.113 | 103.04 | 102.092 | 85.0291 | 60.0821 | 103.04 | 102.055 | 85.048 | 60.081 | 0.62 | Metlin 52 | ESI Qtof | Yes | Yes |
| prolyl-asparagine | $[M+H]^+$ | $C_9H_{15}N_3O_4$ | 2b | 230.113 | 230.114 | 213.087 | 195.076 | 133.061 | 98.06 | 213.089 | 195.078 | 133.061 | 98.0602 | 0.65 | Metlin 85910 | In sillico | Yes | / |
| N-acetyl-neuraminic acid | $[M+H]^+$ | $C_{11}H_{19}NO_9$ | 2a | 310.12 | 310.12 | 292.106 | 274.093 | 232.083 | 108.044 | 292.102 | 274.091 | 232.083 | 108.049 | 0.66 | Metlin 24101 | ESI Qtof | Yes | Yes |
| formylisoglutamine | $[M+H]^+$ | $C_6H_{10}N_2O_4$ | 2b | 175.073 | 175.071 | 157.061 | 147.076 | 139.05 | 130.05 | 157.062 | 147.077 | 139.057 | 130.051 | 0.66 | Metlin 63495 | In sillico | Yes | / |
| histidinyl-proline | $[M+H]^+$ | $C_{11}H_{16}N_4O_3$ | 2b | 253.128 | 253.13 | 235.119 | 138.066 | 116.071 | 110.071 | 235.118 | 138.056 | 116.071 | 110.075 | 0.67 | Metlin 85799 | In sillico | Yes | / |
| serinyl-proline | $[M+H]^+$ | $C_8H_{14}N_2O_4$ | 2b | 203.099 | 203.103 | 185.092 | 171.076 | 157.097 | 116.071 | 185.088 | 171.076 | 157.092 | 116.072 | 1.05 | Metlin 85940 | In sillico | Yes | / |
| nicotinic acid | $[M+H]^+$ | $C_6H_5NO_2$ | 2b | 124.04 | 124.039 | 106.029 | 80.0501 | 78.0344 | / | 106.048 | 80.05 | 78.034 | / | 1.11 | Metlin 240 | ESI Qtof | Yes | / |
| xanthine | $[M+H]^+$ | $C_5H_4N_4O_2$ | 2b | 153.041 | 153.04 | 136.013 | 110.035 | 82.0402 | / | 136.015 | 110.035 | 82.041 | / | 1.45 | Metlin 82 | ESI Qtof | Yes | / |
| isoleucyl-histidine | $[M+H]^+$ | $C_{12}H_{20}N_4O_3$ | 2b | 269.159 | 269.161 | 251.15 | 223.155 | 156.077 | 114.091 | 251.146 | 223.157 | 156.076 | / | 1.47 | Metlin 85815 | In sillico | Yes | / |
| leucyl-glutamine | $[M+H]^+$ | $C_{11}H_{21}N_3O_4$ | 2b | 260.16 | 260.16 | 243.134 | 147.077 | 130.05 | 86.0964 | 243.133 | 147.076 | 130.05 | 86.0963 | 1.61 | Metlin 85831 | In sillico | Yes | / |
| glutamate | $[M+H]^+$ | $C_5H_9NO_4$ | 2b | 148.06 | 148.061 | 130.05 | 105.954 | 102.055 | 84.0446 | 130.04 | 105.947 | 102.053 | 84.0445 | 1.77 | Metlin 19 | ESI Qtof | Yes | / |
| valyl-valine | $[M+H]^+$ | $C_{10}H_{20}N_2O_3$ | 2a | 217.155 | 217.155 | 199.144 | 171.149 | 118.086 | 100.076 | 199.144 | 171.149 | 118.086 | / | 2.09 | Metlin 86032 | In sillico | Yes | Yes |
| alanyl-leucine | $[M+H]^+$ | $C_9H_{18}N_2O_3$ | 2a | 203.139 | 203.142 | 185.131 | 157.136 | 132.104 | 86.0979 | 185.128 | 157.134 | 132.102 | 86.0966 | 2.3 | Metlin 442668 | ESI Qtof | Yes | Yes |
| serinyl-isoleucine | $[M+H]^+$ | $C_9H_{18}N_2O_4$ | 2b | 219.134 | 219.134 | 201.123 | 173.128 | 132.102 | 60.0444 | 201.124 | 173.129 | 132.102 | 60.0447 | 2.35 | Metlin 85935 | In sillico | Yes | / |
| glycyl-leucine | $[M+H]^+$ | $C_8H_{16}N_2O_3$ | 2b | 189.125 | 189.123 | 171.113 | 143.118 | 132.102 | 86.0964 | 171.113 | 143.119 | 132.103 | 86.0973 | 2.5 | Metlin 5727 | In sillico | Yes | / |
| prolyl-isoleucine | $[M+H]^+$ | $C_{11}H_{20}N_2O_3$ | 2b | 229.151 | 229.152 | 183.147 | 70.0645 | / | / | 183.143 | / | / | / | 2.78 | Metlin 23879 | ESI Qtof | Yes | / |
| glycyl-Phenylalanine | $[M+H]^+$ | $C_{11}H_{14}N_2O_3$ | 2b | 223.103 | 223.108 | 177.102 | 166.086 | 120.081 | 103.054 | 177.1 | 106.081 | 120.077 | 103.053 | 2.9 | Metlin 85758 | In sillico | Yes | / |
| isoleucyl-leucine | $[M+H]^+$ | $C_{12}H_{24}N_2O_3$ | 2b | 245.186 | 245.186 | 227.175 | 199.18 | 132.102 | 114.091 | 227.176 | 199.18 | 132.102 | 114.097 | 3.45 | Metlin 85817 | In sillico | Yes | / |
| glycyl-isoleucyl-leucine | $[M+H]^+$ | $C_{14}H_{27}N_3O_4$ | 2b | 302.205 | 302.204 | 171.109 | 143.114 | 132.099 | / | 171.107 | 143.113 | 132.095 | / | 3.76 | Metlin 23190 | ESI Qtof | Yes | / |
| leucyl-phenylalanine | $[M+H]^+$ | $C_{15}H_{22}N_2O_3$ | 2b | 279.108 | 279.17 | 233.165 | 132.102 | 120.081 | 103.054 | 233.155 | 132.102 | 120.081 | 103.055 | 4.12 | Metlin 85839 | In sillico | Yes | / |
| tryptophyl-isoleucine | $[M+H]^+$ | $C_{17}H_{23}N_3O_3$ | 2b | 318.182 | 318.181 | 301.155 | 272.176 | 159.092 | 132.102 | 301.156 | 272.182 | 159.093 | 132.103 | 5.02 | Metlin 85979 | In sillico | Yes | / |
| prostaglandin B3 | $[M+H]^+$ | $C_{20}H_{28}O_4$ | 2b | 333.202 | 333.205 | 315.195 | 197.184 | 171.112 | 157.099 | 315.198 | 297.176 | 171.112 | 157.1 | 6.81 | Metlin 36190 | ESI Qtof | Yes | / |
| 4-ketoretinoic acid | $[M+H]^+$ | $C_{20}H_{26}O_3$ | 2b | 315.197 | 315.196 | 297.185 | 269.191 | 251.18 | 149.097 | 297.183 | 269.19 | 251.179 | / | 8.36 | HMDB0006285 | In sillico | Yes | / |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| octadecatriynoic acid | $[M+H]^+$ | $C_{24}H_{38}O_5$ | 2b | 273.186 | 273.185 | 255.174 | 199.148 | 173.132 | 147.117 | 255.211 | 199.149 | 173.134 | 147.118 | 8.6 | Metlin 74210 | In sillico | Yes | / |
| 3-oxocholic acid | $[M+H]^+$ | $C_7H_{15}NO_3$ | 2a | 407.281 | 407.279 | 389.269 | 371.258 | 347.258 | 343.263 | 389.269 | 371.26 | / | 343.264 | 8.62 | Metlin 57908 | In sillico | Yes | / |

# **Supplementary data for Chapter 4:** Development of an *in vitro* model to investigate drug-microbiome interactions

**Table 4-S1. MRMs, cone voltages and collision energies used for the detection of the drug of interests.**

| *Analyte* | *Precursor ion (m/z)* | *Daughter ion (m/z)* | *Cone voltage (V)* | *Collision energy (eV)* |
|---|---|---|---|---|
| ***APAP*** | 152.10 | 92.88 | 30 | 22 |
| ***APAP-d3*** | 155.07 | 111.00 | 30 | 20 |
| ***SSZ*** | 398.98 | 93.95 | 30 | 20 |
| ***MTX*** | 455.10 | 175.10 | 8 | 38 |
| ***APAP-G*** | 328.13 | 152.09 | 30 | 14 |

**Table 4-S2. Concentrations of the calibration curve solutions used for each of the drug of interest.** The concentrations are reported in nM.

| *APAP* | *APAP-G* | *SSZ* | *MTX* |
|---|---|---|---|
| 0 | 0.5 | 0.1 | 1 |
| 1 | 1 | 0.5 | 5 |
| 2 | 5 | 1 | 40 |
| 5 | 10 | 2.5 | 50 |
| 10 | 25 | 5 | 70 |
| 20 | 50 | 10 | 80 |
| 50 | 100 | 15 | 150 |
| 125 | 150 | 20 | |
| 250 | 200 | 125 | |
| 375 | 250 | | |
| 500 | 350 | | |
| 750 | 400 | | |
| 1000 | 500 | | |
| 1500 | 700 | | |

**Table 4-S3. Chromatographic gradients used for the analysis of the different drugs of interest.**

| | *Time (min)* | *Flow rate (mL/min)* | *% B* |
|---|---|---|---|
| ***APAP / APAP-G*** | Initial | 0.6 | 5 |
| | 2 | 0.6 | 16 |
| | 3 | 0.6 | 95 |
| | 4 | 0.6 | 95 |
| | 4.1 | 0.6 | 5 |
| ***SSZ*** | Initial | 0.6 | 10 |
| | 0.5 | 0.6 | 15 |
| | 1.5 | 0.6 | 95 |
| | 2.5 | 0.6 | 95 |
| | 2.6 | 0.6 | 10 |
| ***MTX*** | Initial | 0.6 | 10 |
| | 0.3 | 0.6 | 30 |
| | 0.45 | 0.6 | 35 |
| | 1 | 0.6 | 95 |
| | 1.5 | 0.6 | 95 |
| | 1.6 | 0.6 | 10 |

# Supplementary information for the Chapter 5. Manipulation of Gut Microbial Metabolites and its Influence on Paracetamol Metabolism

**Table 5-S 1. Diet composition for all animal groups in this study.**

| ***Based on AIN-93G diet*** ***(% by weight)*** | *High tyrosine diet (HT)* | *AB High tyrosine diet (AB)* | *Control (AIN-93G diet) (C)* |
|---|---|---|---|
| *Corn Sugar* | 39.75 | 39.75 | 39.75 |
| *Casein Lactic* | 20.00 | 20.00 | 20.00 |
| *Granular Sugar* | 10.00 | 10.00 | 10.00 |
| *Dextrin* | 13.20 | 13.20 | 13.20 |
| *Solka Floc-40* | 5.00 | 5.00 | 5.00 |
| *AIN-93 mineral Mix* | 3.50 | 3.50 | 3.50 |
| *AIN-93 vitamin Mix* | 1.00 | 1.00 | 1.00 |
| *L-Cysteine* | 0.30 | 0.30 | 0.30 |
| *Choline Bitartrate* | 0.25 | 0.25 | 0.25 |
| *Soy Oil* | 7.00 | 7.00 | 7.00 |
| *Tyrosine* | 5.00 | 5.00 | 0.98 |

**Stock solution, Calibration curves and IS mix preparation for APAP assay:**

The analytical protocol used was the same from a previously validated study (R. Dargue,2019, not yet published). Briefly, APAP stock solution was prepared by initially weighing out around 1 mg of all the standards individually (APAP, APAP-G, APAP-S, APAP-C, APAP-NAC, PAP-G) and diluting in MeOH: $H_2O$ (1:1; v/v) to obtain a final concentration of 1 µg/mL stock solution. Final concentration of standard stock solution (Stock A) was 50 µg/ml of APAP, 8.59 µg/mL of APAP-S, 70.3 µg/mL APAP-G, 1.39 µg/mL of APAP-C, 1.75 µg/mL of APAP-NAC and 1.82 µg/mL of APAP-SG. Stock A was serially diluted to produce 8 calibration standard stocks, which was then further diluted in a 1:100 ratio to obtain the required calibration curve range. Stocks were produced with two LLOQ concentrations (LLOQ1 and LLO$Q^2$), in order to avoid any background interference that might occur at low concentration of metabolites. Stable isotope labelled (SIL) IS solutions were prepared by weighing out the compounds and diluting in MeOH: H2O (1:1; v/v). The final concentration of IS in the SIL ISs solution contained 20 µg/mL of APAP-$d_3$, 3.44 µg/mL of APAP-S-$d_3$, 0.685 µg/mL of APAP-G-$d_3$, 0.838 µg/mL of APAP-C-$d_5$, 1.12 µg/mL of APAP-NAC-$d_5$, 0.001 µg/mL of APAP-SG-$d_3$.

**UPLC-MS parameters.** The chromatographic separation of APAP and its metabolites were performed using a 2.1 x 100 mm 1.8 µm 100 Å C18 Acquity HSS T3 column (Waters Corporation, Manchester, UK). A reversed-phase gradient was used with water and 0.1 % (v/v) FA as the mobile phase A and methanol and 0.1 % (v/v) FA as the mobile phase B. Initial chromatographic conditions were 5 % of B until 0.5 minutes, 7 % of B was reached at 1.85 minutes, 8 % of B at 1.9 minute, 10 % of B at 2.5 minutes, 16 % of B at 4.0 minutes, 25 % of B at 5.0 minutes and reaching 95 % of B at 5.1 minutes for one minute wash before to lower down B back to 5 % for 1.4 column minutes re-equilibration. Total length of the gradient was thus 7.5 minutes. The weak wash solvent used was $H_2O$: MeOH (9:1 v:v) and strong wash solvent was 100 % isopropanol. The sample loop was subjected to both weak and strong solvent wash between each injection. The chromatography part was interfaced with a Xevo–TQs mass spectrometer (Waters Corporation, Manchester, UK). Detection of each compound was done using electrospray ionization (ESI) in positive ion mode in a multiple reaction monitoring (MRM) mechanism reported in Table V.S2. The source temperature was fixed at 150 °C and desolvation temperature at 500 °C. The desolvation gas used was nitrogen and the collision gas used was argon. The capillary voltage was fixed at 3 kV for MRMs in positive ion mode. The cone gas flow was fixed at 150 L/hr, the desolvation gas flow at 1000 L/hr, the collision gas flow at 0.13 mL/min and the nebulizer gas flow to 7.0 bar.

**Table 5-S2. MRMs measured for the APAP targeted assay, as well as the cone voltage and collision energy used for each MRM measurement**

| COMPOUND | PARENT ION (*m/z*) | PRODUCT ION (*m/z*) | CONE VOLTAGE (V) | COLLISION VOLTAGE (V) | RT (MIN) |
|---|---|---|---|---|---|
| **APAP** | 152.1 | 110.06 | 30 | 16 | 2.87 |
| **APAP-$D_3$** | 155.1 | 111.0 | 30 | 20 | 2.84 |
| **APAP-C** | 271.0 | 139.99 | 34 | 24 | 2.48 |
| **APAP-C-$D_5$** | 276.2 | 142.81 | 36 | 26 | 2.88 |
| **APAP-S** | 232.1 | 110.09 | 30 | 22 | 2.19 |
| **APAP-S-$D_3$** | 235.0 | 111.02 | 30 | 22 | 2.20 |
| **APAP-G** | 328.1 | 152.09 | 20 | 14 | 1.74 |
| **APAP-G-$D_5$** | 353.2 | 177.12 | 42 | 16 | 1.74 |
| **APAP-NAC** | 335.1 | 152.03 | 6 | 16 | 4.96 |
| **APAP-NAC-$D_5$** | 340.2 | 152.02 | 34 | 18 | 4.94 |
| **APAP-GSH** | 457.2 | 140.0 | 30 | 36 | 3.82 |

**Table 5-S3. APAP QCs monitoring in urine.**

| *Compounds* | *QC low* | *QC mid* | *QC high* |
|---|---|---|---|
| | *CV %* | *CV %* | *CV %* |
| *APAP* | 1.73 | 2.47 | 1.45 |
| *APAP-S* | 6.49 | 5.36 | 1.90 |
| *APAP-C* | 1.91 | 3.13 | 1.13 |
| *APAP-G* | 3.05 | 3.34 | 5.20 |
| *APAP-NAC* | 11.87 | 6.64 | 7.32 |
| *APAP-GSH* | 4.78 | 2.04 | 3.14 |

**Table 5-S4. APAP QCs monitoring in plasma.**

| *Compounds* | *QC low* | *QC mid* | *QC high* |
|---|---|---|---|
| | *CV %* | *CV %* | *CV %* |
| *APAP* | 2.13 | 1.51 | 1.38 |
| *APAP-S* | 8.99 | 5.77 | 1.33 |
| *APAP-C* | 5.10 | 1.37 | 4.32 |
| *APAP-G* | 3.82 | 6.84 | 7.97 |
| *APAP-NAC* | 15.94 | 4.51 | 11.77 |
| *APAP-GSH* | 2.35 | 1.98 | 3.55 |

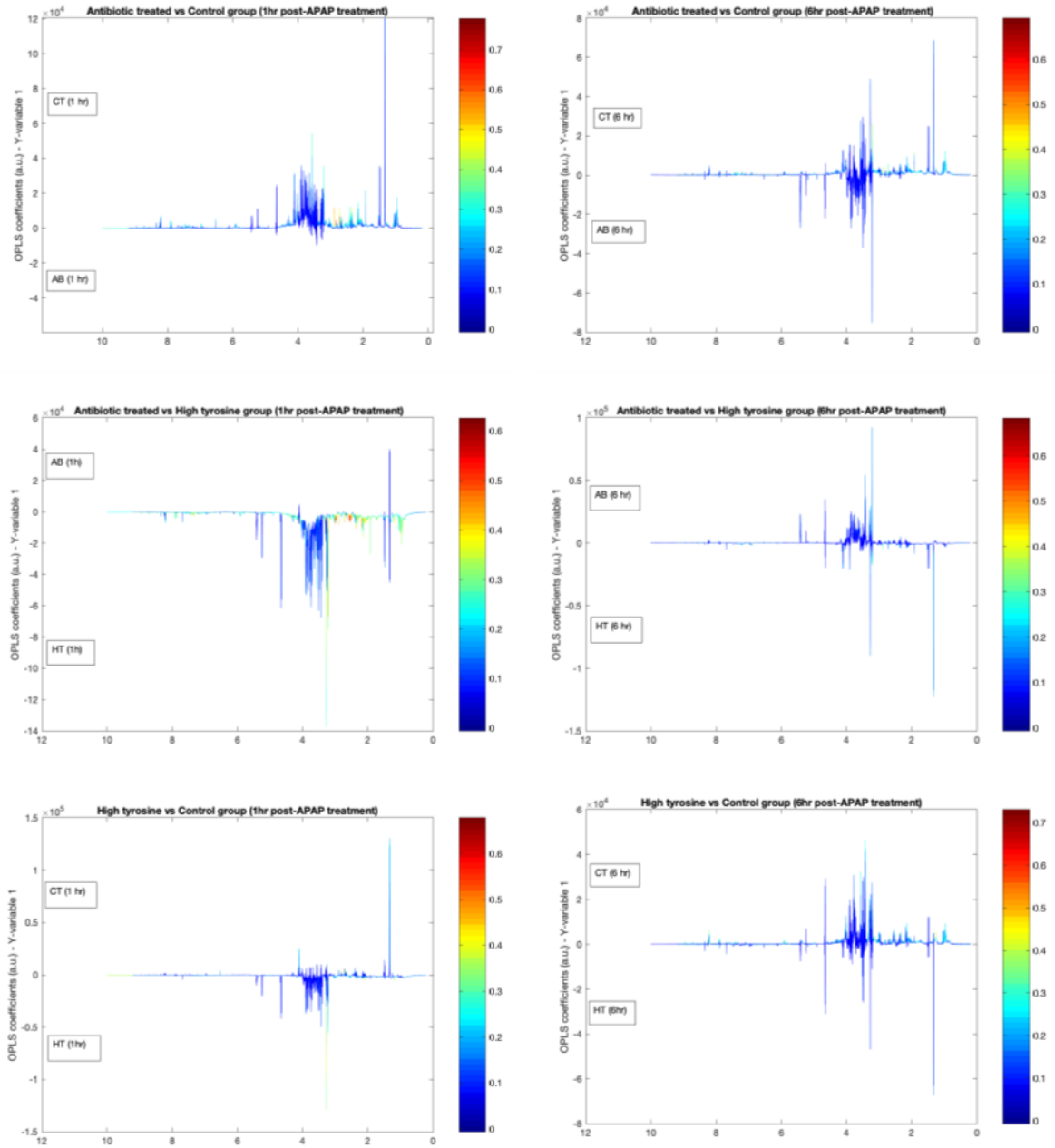
**Table 5-S5. APAP QCs monitoring in liver tissues.**

| *Compounds* | *QC low* | *QC mid* | *QC high* |
|---|---|---|---|
| | *CV %* | *CV %* | *CV %* |
| *APAP* | 0.97 | 1.36 | 0.80 |
| *APAP-S* | 4.82 | 4.67 | 4.74 |
| *APAP-C* | 0.78 | 4.03 | 1.20 |
| *APAP-G* | 6.54 | 2.61 | 6.90 |
| *APAP-NAC* | 27.76 | 9.06 | 2.24 |
| *APAP-GSH* | 3.61 | 0.86 | 0.83 |

**Table 5-S6. Tyrosine QCs monitoring in liver tissues.**

| *Compounds* | *QC low* | *QC mid* | *QC high* |
|---|---|---|---|
| | *CV %* | *CV %* | *CV %* |
| *tyrosine 1 c* | 0.17 | 2.28 | 1.86 |
| *tyrosine 2 c* | 3.05 | 3.60 | 1.59 |
| *tyramine q* | 9.73 | 4.79 | 1.76 |
| ***p**-hydroxybenzoic acid q* | 5.84 | 1.31 | 1.73 |
| ***p**-hydroxy phenyl acetic acid q* | 3.24 | 4.16 | 5.28 |
| ***p**-hydroxy phenyl lactic acid q* | 5.64 | 3.05 | 5.13 |
| ***p**-cresol q* | 2.03 | 1.40 | 0.80 |
| *phenol q* | 0.66 | 2.07 | 4.98 |
| ***p**-cresol sulfate q* | 1.47 | 7.99 | 6.65 |

**Figure 5-S 1. OPLS-DA loading plots proving so significant separation within pair-wise comparison between treatment groups in liver samples.**



Supplementary data for Chapter 6. Investigating the metabolic perturbations induced by the selective inhibition of bacterial β-glucuronidases, enzymes known to influence drug toxicity

**Table 6-S1. Summary of the metabolomics experiments ($^1$H NMR spectroscopy or UPLC-MS ESI+ and ESI-) performed on the samples.**

| MATRICES | $^1$H NMR SPECTROSCOPY | UPLC-MS (ESI +/-) |
|---|---|---|
| **LIVER** | ✓ | ✓ |
| **GALL BLADDER** | - | ✓ |
| **PLASMA** | ✓ | - |
| **COLON TISSUES** | ✓ | ✓ |
| **COLON CONTENTS** | - | ✓ |
| **FAECES** | - | ✓ |
| **URINE** | ✓ | ✓ |

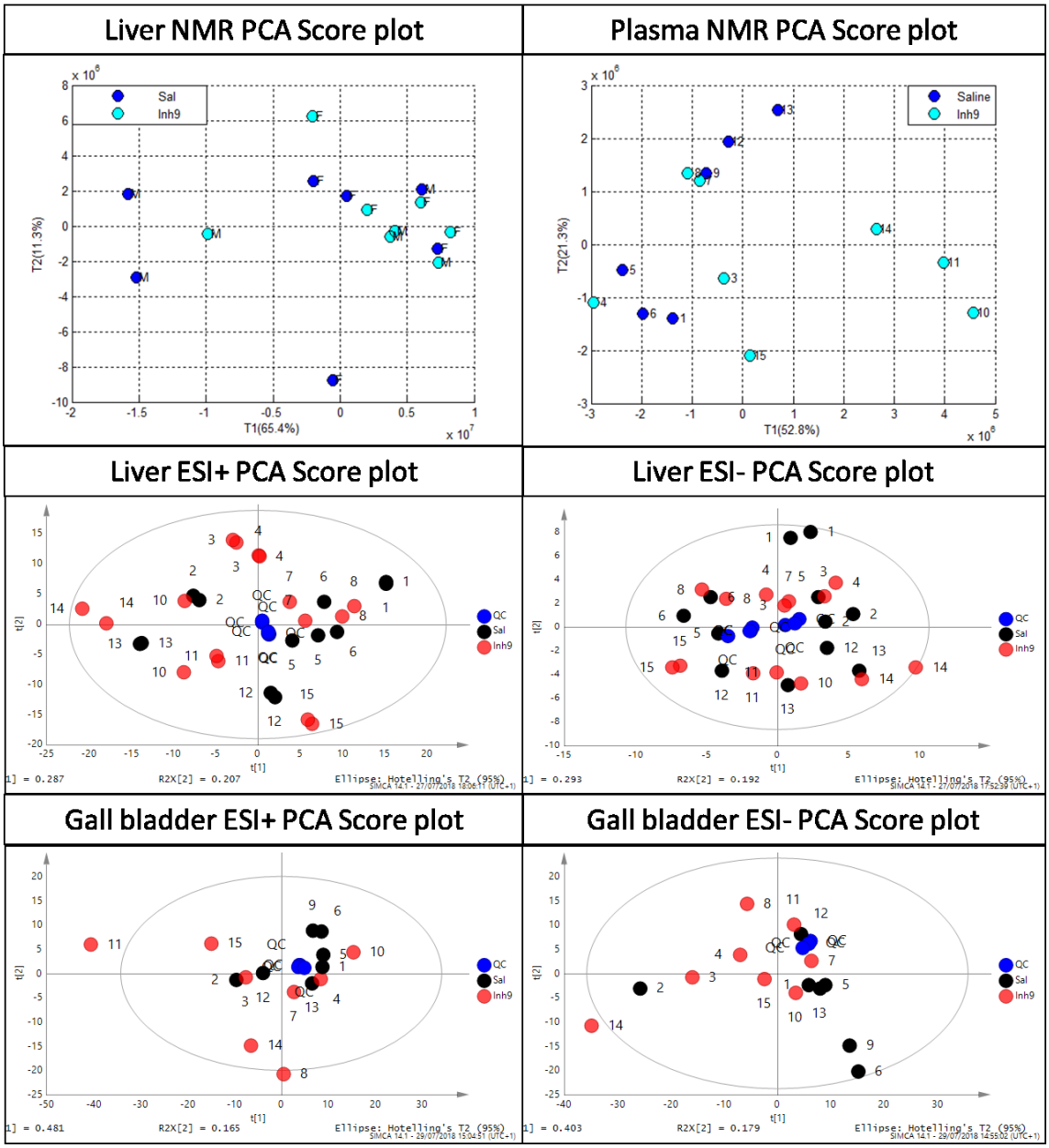



**Figure 6-S1. PCA score plots obtained for the $^{1}$H NMR spectroscopy datasets of the plasma samples and the liver extracts and for the UPLC-MS ESI+ and the UPLC-MS ESI- datasets of the liver and the gall bladders extracts.** The scores t[1] and t[2] represent the distance between a sample to the origin of $PC_1$ and $PC_2$, respectively.

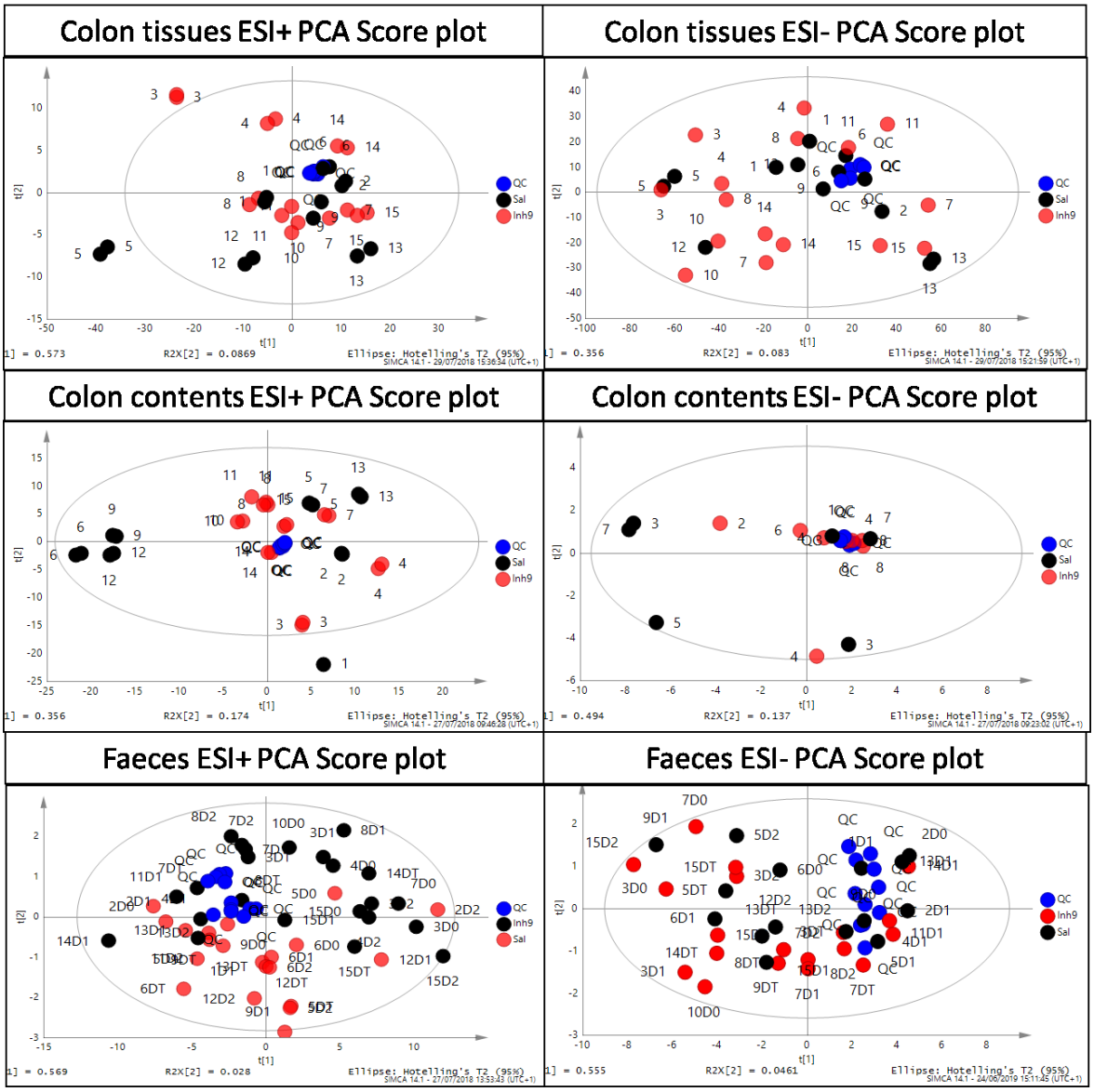



**Figure 6-S2. PCA score plots obtained for the UPLC-MS ESI+ and the UPLC-MS ESI- datasets of the colon tissue, colon content and faecal extracts.** The scores t[1] and t[2] represent the distance between a sample to the origin of $PC_1$ and $PC_2$, respectively.

**Table 6-S2. Metabolite annotations putatively performed on the features found to be significantly affected by the inhibition of the bacterial β-glucuronidase in colon tissues in UPLC-MS ESI+ mode.** All the *m/z* reported correspond to $[M+H]^+$ adducts and the levels of annotation are reported according the levels of confidence described in the Analytical Background Chapter.

| putative ID | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | DB | In silico / exp | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| trimethylsilyl l-alanine | $C_6H_{15}NO_2Si$ | 2b | 162.0938 | 162.0945 | 145.0679 | 144.0839 | 116.089 | 99.0625 | / | 144.0821 | 116.0833 | / | 0.55 | HMDB0094699 | Theo | No |
| NG,NG-dimethyl-L-arginine | $C_8H_{18}N_4O_2$ | 2a | 203.1603 | 203.1497 | 158.1285 | 116.0708 | 88.0875 | 70.0659 | / | 116.0828 | / | 70.0741 | 0.78 | Metlin 6891 | Exp | Yes |
| tyrosine | $C_9H_{11}NO_3$ | 2b | 182.2005 | 182.0864 | 165.0548 | 147.0451 | 136.0753 | 123.0424 | 165.0721 | 147.0603 | 136.0902 | 123.0567 | 1.43 | Metlin 34 | Exp | No |
| 4-(2-aminophenyl)2,4-dioxobutanoate | $C_{10}H_9NO_4$ | 2b | 208.0849 | 208.061 | 190.0504 | 164.0712 | 162.0555 | 120.0449 | 190.0728 | / | 162.0619 | 120.0952 | 3.38 | HMDB0000978 | Theo | No |
| n-acetylleucine | $C_8H_{15}NO_3$ | 2a | 174.0926 | 174.1147 | 156.1035 | 132.1034 | 128.1088 | 86.0987 | 156.098 | 132.1169 | 128.0802 | 86.1016 | 4.23 | Metlin 43866 | Exp | Yes |
| glycocholate | $C_{26}H_{43}NO_6$ | 2a | 466.3685 | 466.3151 | 448.3056 | 430.2936 | 412.2836 | 337.2524 | 448.2846 | 430.2847 | 412.267 | 337.2895 | 8.46 | Metlin 202 | Exp | Yes |
| nutriacholic acid | $C_{24}H_{38}O_4$ | 2b | 391.2386 | 391.2843 | 373.2737 | 355.2632 | 345.2788 | 327.2682 | 373.3144 | 355.2968 | / | 327.258 | 8.93 | HMDB0000467 | Theo | No |
| tetracosahexanoic acid | $C_{24}H_{36}O_2$ | 2b | 357.3146 | 357.2788 | 339.2682 | 321.2577 | 311.2733 | 297.2577 | 339.3076 | 321.2757 | 311.2566 | 297.271 | 9.95 | HMDB0002007 | Theo | No |

**Table 6-S3. Metabolite annotations putatively performed on the features found to be significantly affected by the inhibition of the bacterial β-glucuronidase in colon contents in UPLC-MS ESI+ mode.** All the *m/z* reported correspond to $[M+H]^+$ adducts and the levels of annotation are reported according the levels of confidence described in the Analytical Background Chapter.

| putative ID | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | DB | In silico / exp | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| glutamine | $C_5H_{10}N_2O_3$ | 2a | 147.0765 | 147.0762 | 130.0495 | 102.0547 | 101.071 | 84.0448 | 130.0504 | 102.0561 | 101.0719 | 84.0451 | 0.57 | Metlin 18 | Exp | Yes |
| glutamate | $C_5H_9NO_4$ | 2a | 148.0609 | 148.0607 | 130.0501 | 102.0548 | 84.0446 | / | 130.0503 | 102.0557 | 84.0449 | / | 0.59 | Metlin 19 | Exp | Yes |
| citrulline | $C_6H_{13}N_3O_3$ | 2a | 176.103 | 176.1019 | 159.0757 | 116.0715 | 115.087 | 113.0706 | 159.0766 | 116.0715 | 115.0869 | 113.0706 | 0.61 | Metlin 16 | Exp | Yes |
| glucosamine phosphate | $C_7H_{13}N_3O_4$ | 2b | 260.0458 | 260.053 | 162.0761 | 146.0812 | 144.0655 | 98.9842 | 162.0765 | 146.0819 | 144.0652 | 98.9751 | 0.63 | Metlin 3390 | Theo | No |
| formino-glutamic acid | $C_6H_{10}N_2O_4$ | 2a | 175.0715 | 175.0719 | 157.0613 | 148.061 | 140.0348 | 129.0188 | 157.0604 | 148.0609 | / | / | 0.64 | HMDB0000854 | Theo | Yes |
| 5-aminopentanoic acid | $C_5H_{11}NO_2$ | 2a | 118.0867 | 118.0864 | 101.0602 | 100.076 | 82.0657 | 56.0504 | 101.0608 | 100.0764 | 82.0665 | 56.0505 | 0.67 | Metlin 6902 | Exp | Yes |
| glycyl-proline | $C_7H_{12}N_2O_3$ | 2a | 173.0921 | 173.0915 | 155.081 | 127.0863 | 116.0706 | 70.0655 | 155.0819 | 127.0868 | 116.071 | 70.0662 | 0.68 | Metlin 4185 | Exp | Yes |
| leucine | $C_6H_{13}NO_2$ | 2b | 132.1022 | 132.1016 | 86.0969 | 73.067 | 69.07 | 55.0538 | 86.0973 | / | / | / | 0.96 | Metlin 24 | Exp | No |
| uridine | $C_9H_{12}N_2O_6$ | 2a | 133.0491 | 133.0489 | 115.0372 | 113.0349 | 97.0298 | 85.0289 | 115.0352 | 113.0354 | 97.0285 | / | 1.41 | Metlin 90 | Exp | No |
| tyrosine | $C_9H_{11}NO_3$ | 2b | 182.0821 | 182.0864 | 165.0548 | 147.0451 | 136.0753 | 123.0424 | 165.0553 | 147.0438 | 136.0764 | 123.0444 | 1.43 | Metlin 34 | Exp | No |
| guanine | $C_5H_5N_5O$ | 2b | 152.0559 | 152.0566 | 135.0299 | 110.0355 | 107.0357 | 82.0409 | 135.0297 | 110.0348 | 107.0487 | 82.043 | 1.63 | Metlin 315 | Exp | No |
| hypoxanthine | $C_5H_4N_4O$ | 2b | 137.0458 | 137.0455 | 119.0351 | 110.0348 | 94.0403 | 82.0406 | 119.0389 | 110.0356 | 94.0412 | 82.0404 | 1.64 | Metlin 83 | Exp | No |
| deoxyguanosine | $C_{10}H_{13}N_5O_4$ | 2a | 268.1074 | 268.1031 | 152.0558 | 135.0291 | 117.0535 | 99.0437 | 152.0572 | 135.0323 | 117.0522 | 99.0424 | 1.72 | Metlin 3395 | Exp | Yes |
| proline | $C_5H_9NO_2$ | 2b | 116.0702 | 116.0708 | 70.0659 | 68.0506 | / | / | 70.0662 | 68.0513 | / | / | 1.77 | Metlin 29 | Exp | No |
| phenylalanine | $C_9H_{11}NO_2$ | 2a | 166.0862 | 166.0858 | 149.0591 | 131.0484 | 120.0806 | 103.0537 | 149.0599 | 131.0497 | 120.0811 | 103.0547 | 1.89 | Metlin 28 | Exp | Yes |
| pro tyr | $C_{14}H_{18}N_2O_4$ | 2b | 279.1338 | 279.1329 | 233.1282 | 182.0807 | 70.0651 | / | 233.1484 | 182.0793 | 70.0667 | / | 2.17 - 2.43 | Metlin 23717 | Exp | No |
| leu val | $C_{11}H_{22}N_2O_3$ | 2b | 231.1703 | 231.1671 | 185.1613 | 118.0843 | 86.0955 | 72.0803 | 185.1661 | 118.0859 | 86.0965 | 72.0817 | 2.45 | Metlin 23890 | Exp | No |
| val leu | $C_{11}H_{22}N_2O_3$ | 2b | 213.158 | 213.1666 | 185.1618 | 132.1 | 86.0944 | 72.0796 | 185.165 | 132.1016 | 86.097 | 72.0816 | 2.59 | Metlin 23950 | Exp | No |
| pro ile | $C_{11}H_{20}N_2O_3$ | 2a | 229.1547 | 229.1523 | 211.1388 | 183.1473 | 86.093 | 70.0645 | / | / | 86.0968 | 70.0667 | 2.67 | Metlin 23879 | Exp | Yes |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| indole-3-carbaldehyde | $C_9H_7NO$ | 2b | 146.0612 | 146.0589 | 118.0642 | 91.0528 | 65.0379 | / | 118.0656 | 91.0542 | / | / | 2.69 | HMDB0029737 | Exp | No |
| serotonin | $C_{10}H_{12}N_2O$ | 2b | 177.1031 | 177.102 | 160.0752 | 142.0648 | 132.0806 | 115.0541 | / | 142.0652 | 132.0799 | 155.0546 | 2.78 | Metlin 23 | Exp | No |
| indole acrylic acid | $C_{11}H_9NO_2$ | 2b | 188.0712 | 188.0698 | 170.06 | 146.0602 | 142.0645 | 115.0534 | 170.0628 | 146.06 | 142.0655 | 115.0547 | 2.8 | Metlin 5702 | Exp | No |
| 4-(2-aminophenyl)2,4-dioxobutanoate | $C_{10}H_9NO_4$ | 2b | 208.0592 | 208.061 | 190.0504 | 164.0712 | 162.0555 | 120.0449 | 190.0486 | / | 162.0555 | 120.0443 | 3.38 | HMDB0000978 | Theo | No |
| gly lys tyr | $C_{17}H_{26}N_4O_5$ | 2b | 364.2003 | 367.1946 | 332.1599 | 310.1715 | 186.1209 | 129.1006 | 332.1543 | / | 186.1243 | 129.1013 | 3.53 | Metlin 20072 | Exp | No |
| ser ile leu | $C_{15}H_{29}N_3O_5$ | 2b | 332.2135 | 332.225 | 201.1272 | 173.1321 | 132.1058 | 86.0985 | 201.1245 | 173.1293 | 132.1023 | 86.0974 | 3.57 | Metlin 22825 | Exp | No |
| phe leu | $C_{15}H_{22}N_2O_3$ | 2b | / | 279.1687 | 233.1623 | 132.0994 | 120.0792 | 86.0941 | 233.1669 | 132.0986 | 120.081 | 86.0981 | 3.72 | Metlin 23831 | Exp | No |
| pro trp | $C_{16}H_{19}N_3O_3$ | 2b | 302.1563 | 302.1487 | 205.095 | 188.0686 | 181.1006 | 70.0653 | 205.0879 | 188.0699 | / | 70.0654 | 3.94 | Metlin 24006 | Exp | No |
| arg gln arg | $C_{17}H_{34}N_{10}O_5$ | 2b | 459.2759 | 459.2802 | 442.2558 | 303.1753 | 286.1471 | 215.3198 | 442.2702 | / | 286.144 | / | 4.3 | Metlin 19907 | Exp | No |
| trp ile or trp leu | $C_{17}H_{23}N_3O_3$ | 2b | 318.1697 | 318.1799 | 301.153 | 273.1581 | 159.0906 | 144.0805 | 301.1552 | 273.1696 | / | 144.0855 | 5.26 | Metlin 23956 | Exp | No |
| tetranor-23-oxo-hydroxyvitamin D3 | $C_{23}H_{34}O_3$ | 2b | 359.2561 | 359.2581 | 341.2475 | 323.2369 | 219.1743 | 139.0754 | 341.248 | 323.1792 | 219.1415 | / | 8.73 | HMDB0060114 | Theo | No |
| nutriacholic acid | $C_{24}H_{38}O_4$ | 2b | / | 391.2843 | 373.2737 | 355.2632 | 345.2788 | 327.2682 | 373.2695 | 355.262 | 345.2475 | / | 8.93 | HMDB0000467 | Theo | No |
| cervonoyl ethanolamide | $C_{24}H_{36}O_3$ | 2b | 373.2722 | 373.2737 | 355.2632 | 329.2475 | 311.2369 | 103.039 | 355.2637 | 329.2156 | / | / | 9.11 | HMDB0013627 | Theo | No |
| cervonic acid ethyl ester | $C_{24}H_{36}O_2$ | 2b | 357.279 | 357.2777 | 311.2355 | 215.1773 | 119.0845 | 175.146 | 311.1736 | 215.1841 | 119.0849 | 175.1455 | 9.58 | Metlin 62947 | Exp | No |
| PA(10:0/8:0) | $C_{21}H_{41}O_8P$ | 2b | 453.2591 | 453.2612 | 435.2506 | 373.2949 | 355.2843 | 213.1632 | 435.2499 | 373.2746 | 355.2639 | 213.1849 | 9.71 | HMDB0114778 | Theo | No |
| tetracosahexanoic acid | $C_{24}H_{36}O_2$ | 2b | 357.2791 | 357.2788 | 339.2682 | 321.2577 | 311.2733 | 297.2577 | 339.265 | 321.2613 | / | 293.2396 | 9.95 | HMDB0002007 | Theo | No |

**Table 6-S4. Metabolite annotations putatively performed on the features found to be significantly affected by the inhibition of the bacterial β-glucuronidase in urine samples in UPLC-MS ESI+ mode.** The levels of annotation are reported according the levels of confidence described in the Analytical Background Chapter.

| putative ID | Adduct | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | DB | In silico / exp | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S*-adenosylmethionine | $[M+H]^+$ | $C_{15}H_{22}N_6O_5S$ | 2a | / | 399.1434 | 298.094 | 264.0908 | 250.0918 | 136.0604 | 298.0981 | / | 250.0933 | 136.0638 | 0.55 | Metlin 3289 | Exp | Yes |
| *N*-acetylspermidine | $[M+H]^+$ | $C_9H_{21}N_3O$ | 2b | 188.176 | 188.1755 | 171.1488 | 117.1018 | 100.0758 | 72.0812 | 171.1496 | 117.1032 | 100.0769 | 72.0823 | 0.55 | Metlin 3323 | Exp | No |
| *N6,N6,N6*-trimethyl-L-lysine | $[M+H]^+$ | $C_9H_{20}N_2O_2$ | 2a | 189.1603 | 189.1598 | 172.1332 | 171.1492 | 143.1543 | 114.1277 | / | 171.1496 | / | / | 0.55 | HMDB0001325 | Theo | Yes |
| lysine | $[M+H]^+$ | $C_6H_{14}N_2O_2$ | 2a | 147.113 | 147.1125 | 130.0857 | 106.0021 | 84.081 | 64.0165 | 130.0876 | | 84.0818 | / | 0.55 | Metlin 25 | Exp | Yes |
| 5-hydroxylysine | $[M+H]^+$ | $C_6H_{14}N_2O_3$ | 2a | / | 163.1083 | 145.0977 | 128.0712 | 117.1028 | 99.0922 | / | 128.0723 | 117.1044 | 99.0896 | 0.55 | HMDB0000450 | Exp | Yes |
| histidine | $[M+H]^+$ | $C_6H_9N_3O_2$ | 2b | 156.0779 | 156.0764 | 110.0713 | 95.0606 | 93.0449 | 83.061 | 110.0746 | 95.0602 | 93.0449 | 83.0622 | 0.55 | Metlin 21 | Exp | No |
| taurine | $[M+H]^+$ | $C_2H_7NO_3S$ | 2a | 126.0232 | 126.0214 | 108.0107 | 84.9596 | 78.9847 | 64.9698 | 108.012 | / | 78.9833 | / | 0.56 | Metlin 31 | Exp | Yes |
| glycerophosphocholine | $[M+H]^+$ | $C_8H_{20}NO_6P$ | 2a | 258.1105 | 258.1101 | 184.0732 | 166.0625 | 125.000 | 104.1071 | 184.0755 | / | 125.0012 | 104.1081 | 0.58 | Metlin 370 | Exp | Yes |
| glutamate | $[M+H]^+$ | $C_5H_9NO_4$ | 2a | 148.0621 | 148.0607 | 130.0501 | 102.0548 | 84.0446 | / | 130.0508 | 102.0554 | 84.0448 | / | 0.59 | Metlin 19 | Exp | Yes |
| 5-aminopentanoic acid | $[M+H]^+$ | $C_5H_{11}NO_2$ | 2a | 118.0864 | 118.0864 | 101.0602 | 100.076 | 82.0657 | 56.0504 | / | 100.0768 | 82.0671 | 56.0501 | 0.6 | Metlin 6902 | Exp | Yes |
| carnitine | $[M+H]^+$ | $C_7H_{15}NO_3$ | 2a | 162.1127 | 162.1107 | 103.0385 | 102.0908 | 82.0282 | 60.0812 | 103.0422 | 102.0933 | / | / | 0.6 | Metlin 52 | Exp | Yes |
| thiamine | $[M+H]^+$ | $C_{12}H_{16}N_4OS$ | 2a | 265.1123 | 265.1117 | 144.0476 | 122.0714 | 81.0449 | / | 144.0495 | 122.0719 | 81.0492 | / | 0.62 | Metlin 229 | Exp | Yes |
| creatinine | $[M+H]^+$ | $C_4H_7N_3O$ | 2a | 114.0671 | 114.0687 | 86.0741 | 72.0474 | / | / | 86.0722 | 72.0446 | / | / | 0.63 | Metlin 8 | Exp | Yes |
| sucrose | $[M+Na]^+$ | $C_{12}H_{22}O_{11}$ | 2a | 365.1058 | 365.1051 | 203.0514 | 185.0414 | / | / | 203.0556 | / | / | / | 0.63 | Metlin 137 | Exp | Yes |
| *N*-acetylneuraminic acid | $[M+H]^+$ | $C_{11}H_{19}NO_9$ | 2a | 310.1105 | 310.1123 | 292.101 | 274.0929 | 256.0852 | 250.0927 | 292.1043 | 274.0932 | / | 250.0893 | 0.64 | Metlin 5739 | Exp | Yes |
| trigonelline | $[M+H]^+$ | $C_7H_7NO_2$ | 2a | 138.0557 | 138.0553 | 94.0663 | 92.0506 | 78.0349 | 67.0552 | 94.0666 | 92.0506 | / | 67.0545 | 0.65 | Metlin 273 | Exp | Yes |
| creatine | $[M+H]^+$ | $C_4H_9N_3O_2$ | 2b | 132.0779 | 132.0768 | 115.05 | 114.0661 | 90.0555 | 87.0557 | 115.0516 | 114.0669 | 90.0555 | 87.0555 | 0.65 | Metlin 7 | Exp | No |
| *NG,NG*-dimethyl-arginine | $[M+H]^+$ | $C_8H_{18}N_4O_2$ | 2b | 203.1508 | 203.1497 | 158.1247 | 133.0964 | 116.0704 | 88.0875 | 158.1294 | 133.0965 | 116.0713 | 88.0879 | 0.67 | Metlin 6891 | Exp | No |
| glycyl-aspartate | $[M+H]^+$ | $C_6H_{10}N_2O_5$ | 2a | 191.0682 | 191.0662 | 173.0557 | 155.0451 | 145.0608 | 127.0502 | 173.0563 | 155.0456 | / | 127.0471 | 0.68 | HMDB0028837 | Theo | Yes |
| *N*-acetylhistidine | $[M+H]^+$ | $C_8H_{11}N_3O_3$ | 2b | 198.0873 | 198.0865 | 180.0757 | 156.0759 | 152.081 | 110.0709 | 180.0764 | 156.0771 | 152.0768 | 110.0723 | 0.71 | Metlin 44784 | Exp | No |
| citrulline | $[M+H]^+$ | $C_6H_{13}N_3O_3$ | 2a | 176.1031 | 176.1019 | 159.0757 | 116.0715 | 115.087 | 113.0706 | 159.0758 | 116.0704 | / | 113.0721 | 0.71 | Metlin 16 | Exp | Yes |
| *N*-acetylglutamine | $[M+H]^+$ | $C_7H_{12}N_2O_4$ | 2a | 189.0867 | 189.0869 | 172.0603 | 147.0747 | 130.0496 | 84.0444 | 172.0616 | / | 130.0512 | 84.0455 | 0.82 | Metlin 58344 | Exp | Yes |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4-guanidino butanoic acid | $[M+H]^+$ | $C_5H_{11}N_3O_2$ | 2a | 146.0934 | 146.0917 | 128.0817 | 104.0707 | 87.0447 | 86.0607 | 128.0822 | 104.0714 | 87.0448 | 86.0613 | 0.82 | Metlin 4155 | Exp | Yes |
| glycylproline | $[M+H]^+$ | $C_7H_{12}N_2O_3$ | 2a | 173.0937 | 173.0915 | 155.081 | 127.0863 | 116.0706 | 70.0655 | 155.0821 | 127.0879 | 116.0711 | 70.0663 | 0.82 | Metlin 4185 | Exp | Yes |
| N-acetyllysine | $[M+H]^+$ | $C_8H_{16}N_2O_3$ | 2a | 189.1242 | 189.123 | 172.0961 | 171.1124 | 147.1123 | 129.102 | 172.0972 | 171.1133 | 147.112 | 129.1016 | 0.83 | Metlin 63102 | Exp | Yes |
| N-acetylputrescine | $[M+H]^+$ | $C_6H_{14}N_2O$ | 2a | 131.0901 | 131.184 | 114.0915 | 72.0816 | / | / | 114.0922 | 72.0818 | / | / | 0.83 | HMDB002064 | Exp | Yes |
| N-acetylglutamic acid | $[M+H]^+$ | $C_7H_{11}NO_5$ | 2a | 190.0736 | 190.076 | 172.0611 | 148.0601 | 130.0495 | 102.0559 | 172.063 | 148.06 | 130.0513 | 102.0544 | 1.35 | Metlin 3325 | Exp | Yes |
| uridine | $[M+H]^+$ | $C_9H_{12}N_2O_6$ | 2a | / | 133.0489 | 115.0372 | 113.0349 | 97.0298 | 85.0289 | / | 113.0365 | / | / | 1.38 | Metlin 90 | Exp | Yes |
| tyrosine | $[M+H]^+$ | $C_9H_{11}NO_3$ | 2a | 182.0849 | 182.0864 | 165.0548 | 147.0451 | 136.0753 | 123.0424 | 165.0538 | 147.043 | 136.0755 | 123.0423 | 1.43 | Metlin 34 | Exp | Yes |
| 3-hydroxyxyanthranilic acid | $[M+H]^+$ | $C_7H_7NO_3$ | 2b | 154.0497 | 154.049 | 136.0386 | 108.0441 | 80.0498 | / | 136.0399 | 108.0437 | 80.0501 | / | 1.47 | Metlin 3275 | Exp | No |
| leucine | $[M+H]^+$ | $C_6H_{13}NO_2$ | 2a | 132.1011 | 132.1016 | 86.0969 | 73.067 | 69.07 | / | 86.0973 | / | / | / | 1.56 | Metlin 24 | Exp | Yes |
| 4-pyridoxate | $[M+H]^+$ | $C_8H_9NO_4$ | 2a | 184.0603 | 184.0604 | 166.0495 | 148.0387 | 144.9871 | 138.0558 | 166.0508 | 148.0401 | 144.9589 | 138.0536 | 1.56 - 1.71 | Metlin 239 | Exp | Yes |
| dihydroxyquinoline-β-glucuronide | $[M+H]^+$ | $C_{15}H_{15}NO_8$ | 2b | 338.0869 | 338.087 | 320.0765 | 302.0659 | 292.0816 | 162.055 | 320.0977 | 302.0807 | / | 162.0548 | 1.79 | HMDB0011658 | Theo | No |
| phelalanine | $[M+H]^+$ | $C_9H_{11}NO_2$ | 2a | 166.0849 | 166.0858 | 149.0591 | 131.0484 | 120.0806 | 103.0537 | 149.06 | 131.0492 | 120.0807 | 103.0559 | 2.1 | Metlin 28 | Exp | Yes |
| porphobilinogen | $[M+H]^+$ | $C_{10}H_{14}N_2O_4$ | 2b | 227.1025 | 227.11 | 210.0876 | 182.0926 | 165.0826 | 138.0645 | 210.0745 | 182.0876 | / | / | 2.28 | Metlin 76 | Exp | No |
| dihydrocaffeic acid glucuronic acid | $[M+H]^+$ | $C_{15}H_{18}O_{10}$ | 2b | 359.0874 | 359.0973 | 341.0867 | 323.0761 | 183.0652 | 165.0546 | / | 323.0755 | 183.0666 | 165.055 | 2.38 | HMDB0041720 | Theo | No |
| 2-indole-carboxylic acid | $[M+H]^+$ | $C_9H_7NO_2$ | 2b | 162.0556 | 162.0555 | 144.0449 | 134.0242 | 118.0293 | 116.05 | 144.0438 | / | / | 116.0508 | 2.76 | HMDB0002285 | Theo | No |
| xanthurenic acid (glucuronide) | $[M+H]^+$ | $C_{10}H_7NO_4$ | 2b | 206.0441 | 206.0442 | 188.0336 | 178.0462 | 160.0382 | 132.0435 | 188.0238 | 178.0486 | 160.0389 | 132.0482 | 2.77 | Metlin 5481* | Theo | No |
| 4-(2-aminophenyl)2,4-dioxobutanoate | $[M+H]^+$ | $C_{10}H_9NO_4$ | 2b | 208.0613 | 208.061 | 190.0504 | 164.0712 | 162.0555 | 120.0449 | 190.0499 | 164.0658 | 162.0555 | 120.047 | 3.38 | HMDB0000978 | Theo | No |
| hippuric acid | $[M+H]^+$ | $C_9H_9NO_3$ | 2a | 180.0667 | 180.0672 | 139.0574 | 121.1007 | 105.0338 | 77.0394 | / | 121.1048 | 105.0348 | 77.0402 | 3.58 | Metlin 1301 | Exp | Yes |
| hydroxyprolyl-methionine | $[M+H]^+$ | $C_{10}H_{18}N_2O_4S$ | 2b | 263.1037 | 263.106 | 245.0954 | 227.0849 | 217.1005 | 199.09 | 245.0905 | 227.0887 | 217.0969 | 199.30889 | 3.62 | HMDB0028869 | | No |
| biotin | $[M+H]^+$ | $C_{10}H_{16}N_2O_3S$ | 2a | 245.0923 | 245.0956 | 227.0849 | 184.0819 | 167.0533 | 166.0685 | 227.0839 | / | / | / | 3.7 | Metlin 243 | Exp | Yes |
| phenylacetylglycine | $[M+H]^+$ | $C_{10}H_{11}NO_3$ | 2a | 194.0813 | 194.0808 | 148.0745 | 120.0814 | 91.0546 | 76.0398 | 148.0755 | 120.081 | 91.0549 | 76.0404 | 4.04 | Metlin 4237 | Exp | Yes |
| 5-hydroxyindoleacetic acid | $[M+H]^+$ | $C_{10}H_9NO_3$ | 2b | 192.0663 | 192.0637 | 146.0599 | 110.0086 | 68.9833 | 66.0208 | 146.0607 | 110.0106 | / | / | 4.13 | Metlin 2975 | Exp | No |
| hexanoylglycine | $[M+H]^+$ | $C_8H_{15}NO_3$ | 2b | 174.1123 | 174.1125 | 156.1019 | 128.107 | 99.0804 | 76.0393 | 156.1024 | 128.1125 | 99.082 | 76.0385 | 4.68 | HMDB0000701 | Exp | No |
| cinnamoyglycine | $[M+H]^+$ | $C_{11}H_{11}NO_3$ | 2b | 206.0479 | 206.0464 | 188.0373 | 160.0378 | 131.0488 | 103.0547 | 188.0381 | / | 131.0507 | 103.0556 | 5.34 | Metlin 34534 | Exp | No |
| tetranor-5-NO2-CLA | $[M+Na]^+$ | $C_{14}H_{23}NO_4$ | 2b | 292.1532 | 292.1519 | 270.17 | 252.1594 | / | / | 270.1711 | 252.1606 | 206.1535 | 177.1279 | 5.7 | LMFA01120010 | Theo | No |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| octenoylglycine | $[M+H]^+$ | $C_{10}H_{17}NO_3$ | 2b | 200.13 | 200.1281 | 183.1016 | 182.1176 | 165.091 | 154.1226 | / | 182.1196 | 165.0939 | 154.1247 | 6.2 - 6.31 | HMDB0094801 | Theo | No |

**Table 6-S5. Metabolite annotations putatively performed on the features found to be significantly affected by the inhibition of the bacterial β-glucuronidase in urine samples in UPLC-MS ESI- mode.** The levels of annotation are reported according the levels of confidence described in the Analytical Background Chapter.

| putative ID | Adduct | formula | ID level | *m/z* | Calculated mass | frag 1 theo | frag 2 theo | frag 3 theo | frag 4 theo | frag 1 obs | frag 2 obs | frag 3 obs | frag 4 obs | RT (mins) | DB | In silico / exp | NPC DB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5-hydroxylysine | $[M-H]^-$ | $C_6H_{14}N_2O_3$ | 2b | / | 161.0926 | 144.0661 | 143.0821 | 130.0504 | 117.1028 | 144.0651 | 143.084 | / | / | 0.55 | HMDB0000450 | Theo | No |
| taurine | $[M-H]^-$ | $C_2H_7NO_3S$ | 2a | 124.0069 | 124.0062 | 106.9814 | 94.9808 | 80.9651 | 79.957 | 106.9842 | 94.9783 | 80.9629 | 79.9568 | 0.56 | Metlin 31 | Exp | Yes |
| glycerol-phosphate | $[M+Br]^-$ | $C_3H_9O_6P$ | 2a | 251.0491 | 251.047 | 171.0139 | 114.943 | 86.9067 | 71.045 | 171.0058 | / | / | / | 0.58 | HMDB0000126 | Exp | Yes |
| glutamate | $[M-H]^-$ | $C_5H_9NO_4$ | 2a | 146.0463 | 146.0452 | 128.035 | 102.056 | 85.0293 | / | 128.0353 | / | 85.0279 | / | 0.58 | Metlin 19 | Exp | Yes |
| glycerophosphocholine | $[M-H]^-$ | $C_8H_{20}NO_6P$ | 2b | / | 256.0956 | 242.0813 | 168.0423 | 152.9959 | 78.9586 | 242.0795 | 168.0391 | 152.9913 | 78.9584 | 0.58 | Metlin 370 | Exp | No |
| N6,N6,N6-trimethyl-L-lysine | $[M-H]^-$ | $C_9H_{20}N_2O_2$ | 2a | / | 187.1452 | 130.0874 | 128.0717 | 114.0561 | 74.0248 | / | 128.0698 | 114.0509 | 74.0256 | 0.61 | HMDB0001325 | Theo | Yes |
| allantoin | $[M-H]^-$ | $C_4H_6N_4O_3$ | 2a | 157.0355 | 157.0344 | 140.0056 | 114.0292 | 97.0034 | 71.0251 | 140.0068 | 114.0296 | 97.0027 | / | 0.63 | Metlin 89 | Exp | Yes |
| N-acetylneuraminic acid | $[M-H]^-$ | $C_{11}H_{19}NO_9$ | 2b | 308.0989 | 308.0983 | 290.0875 | 170.0453 | 98.0607 | 87.0091 | 290.0854 | 170.0472 | 98.0619 | 87.0092 | 0.64 | Metlin 5739 | Exp | No |
| glycyl-aspartate | $[M-H]^-$ | $C_6H_{10}N_2O_5$ | 2b | 189.0518 | 189.0517 | 171.0411 | 145.0619 | 132.0302 | 127.0513 | 171.0398 | / | 132.0383 | / | 0.68 | HMDB0028837 | Theo | No |
| arabinose | $[M-H]^-$ | $C_5H_{10}O_5$ | 2a | 149.0452 | 149.0425 | 131.0599 | 113.0226 | 101.0233 | 89.0237 | 131.0348 | 113.0242 | 101.0239 | 89.0246 | 0.68 | Metlin 65497 | Exp | Yes |
| sucrose | $[M-H]^-$ | $C_{12}H_{22}O_{12}$ | 2a | 341.1071 | 341.01085 | 179.0553 | 161.0444 | 143.0342 | 119.0341 | 179.055 | 161.0385 | 143.033 | 119.033 | 0.7 | Metlin 137 | Exp | Yes |
| N-acetylhistidine | $[M-H]^-$ | $C_8H_{11}N_3O_3$ | 2b | 196.0721 | 196.0722 | 178.0617 | 154.0617 | 152.0824 | 134.0718 | / | 154.0614 | 152.0831 | 134.068 | 0.71 | HMDB0032055 | Theo | No |
| 4-guanidino butanoic acid | $[M-H]^-$ | $C_5H_{11}N_3O_2$ | 2b | 144.0767 | 144.0782 | 127.051 | 102.0555 | 84.0448 | 83.0612 | 127.0554 | 102.0576 | / | 83.063 | 0.94 | Metlin 4155 | Exp | No |
| N-acetylglutamic acid | $[M-H]^-$ | $C_7H_{11}NO_5$ | 2b | 188.0561 | 188.0553 | 170.0446 | 144.0654 | 128.0345 | 102.0555 | 170.0468 | 144.0658 | 128.0352 | 102.0566 | 1.38 | Metlin 3325 | Exp | No |
| succinic acid | $[M-H]^-$ | $C_4H_6O_4$ | 2a | 117.0189 | 117.019 | 99.0086 | 73.03 | 55.019 | / | 99.008 | 73.0279 | / | / | 1.41 | Metlin 114 | Exp | Yes |
| tyrosine | $[M-H]^-$ | $C_9H_{11}NO_3$ | 2b | 180.0679 | 180.0654 | 163.0392 | 119.0502 | 93.0346 | 72.0089 | 163.041 | 119.053 | 93.0433 | / | 1.43 | Metlin 34 | Exp | No |
| deoxyuridine | $[M-H]^-$ | $C_9H_{12}N_2O_5$ | 2a | 227.0672 | 227.066 | 184.0593 | 136.0372 | 111.0195 | 94.03 | 184.0613 | 136.042 | 111.009 | 94.0301 | 1.62 | Metlin 91 | Exp | Yes |
| 4-pyridoxate | $[M-H]^-$ | $C_8H_9NO_4$ | 2b | 182.0457 | 182.0448 | 138.0552 | 123.0315 | 120.0442 | 108.0446 | 138.0557 | 123.0329 | 120.0451 | 108.046 | 1.75 | Metlin 239 | Exp | No |
| dihydroxyquinoline-β-glucuronide | $[M-H]^-$ | $C_{15}H_{15}NO_8$ | 2b | 336.0726 | 336.0725 | 318.0619 | 292.087 | 290.067 | 160.0404 | 318.0638 | / | / | 160.0398 | 1.99 | HMDB0011658 | Theo | No |
| phenylalanine (glucuronide) | $[M-H]^-$ | $C_9H_{11}NO_2$ | 2b | 164.0727 | 164.0711 | 147.0445 | 103.0549 | 91.0549 | 72.0094 | / | / | / | / | 2.09 | Metlin 28 | Exp | No |
| porphobilinogen | $[M-H]^-$ | $C_{10}H_{14}N_2O_4$ | 2b | 225.08872 | 225.0872 | 181.097 | 163.0864 | 152.071 | 120.0816 | 181.0954 | 163.0875 | / | / | 2.32 | Metlin 76 | Exp | No |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dihydrocaffeic acid glucuronic acid | $[M-H]^-$ | $C_{15}H_{18}O_{10}$ | 2b | 357.0825 | 357.0827 | 339.0722 | 181.0506 | 163.0401 | 117.0193 | 339.0734 | 181.049 | 163.0456 | 117.0192 | 2.42 | HMDB0041720 | Theo | No |
| p-hydroxyphenyllactic acid | $[M-H]^-$ | $C_9H_{10}O_4$ | 2a | 181.0495 | 181.0496 | 163.0385 | 135.044 | 119.0505 | 72.9928 | 163.0412 | 135.0443 | 119.05 | / | 2.72 | Metlin 34515* | Exp | Yes |
| 2-indole-carboxylic acid (glucuronide) | $[M-H]^-$ | $C_9H_7NO_2$ | 2b | 160.0403 | 160.0399 | 142.0293 | 116.05 | 114.0344 | / | / | 116.053 | / | / | 2.76 | HMDB0002285 | Theo | No |
| xanthurenic acid (glucuronide) | $[M-H]^-$ | $C_{10}H_7NO_4$ | 2b | 204.0296 | 204.029 | 160.0391 | 132.0459 | 116.049 | / | 160.0397 | 132.0432 | 116.0495 | / | 2.85 | Metlin 5481* | Theo | No |
| 5-hydroxyindoleacetic acid | $[M-H]^-$ | $C_{10}H_9NO_3$ | 2a | 190.0504 | 190.0407 | 146.0602 | 144.044 | 131.0382 | / | 146.0606 | 144.0442 | 131.0368 | / | 3.18 - 3.38 | Metlin 2975 | Exp | Yes |
| phenyl glucuronide | $[M-H]^-$ | $C_{12}H_{14}O_7$ | 2b | 269.0659 | 269.0667 | 251.0555 | 175.0243 | 129.0189 | 113.0241 | 251.0556 | 175.024 | 129.0551 | 113.0245 | 3.28 | Metlin 1825 | Exp | No |
| 4-(2-aminophenyl)2,4-dioxobutanoate | $[M-H]^-$ | $C_{10}H_9NO_4$ | 2b | 206.0452 | 206.0453 | 188.0348 | 162.0555 | 160.0399 | 145.029 | 188.0351 | 162.0555 | 160.0398 | 145.0452 | 3.38 | HMDB0000978 | Theo | No |
| p-coumaric acid sulfate | $[M-H]^-$ | $C_9H_8O_6S$ | 2b | 242.9965 | 242.9969 | 224.9863 | 199.0071 | 163.0401 | 145.0295 | / | 199.0062 | 163.0396 | / | 3.49 | HMDB0125166 | Theo | No |
| pimelic acid | $[M-H]^-$ | $C_7H_{12}O_4$ | 2a | 159.0663 | 159.0647 | 141.0545 | 115.0755 | 97.0656 | 95.0502 | / | 115.0766 | 97.0648 | 95.0466 | 3.51 | Metlin 3280 | Exp | Yes |
| indoxylsulfuric acid | $[M-H]^-$ | $C_8H_7NO_4S$ | 2a | 212.0011 | 212.0019 | 132.0448 | 80.9653 | 79.9576 | / | 132.045 | 80.963 | 79.954 | / | 3.53 | Metlin 252 | Exp | Yes |
| hippuric acid | $[M-H]^-$ | $C_9H_9NO_3$ | 2a | 178.053 | 178.0503 | 134.0607 | 132.0449 | 77.0399 | 56.0139 | 134.053 | 132.0449 | 77.0387 | / | 3.61 | Metlin 1301 | Exp | Yes |
| ferulic acid 4-O-slufate | $[M-H]^-$ | $C_{10}H_{10}O_7S$ | 2b | 273.0073 | 273.0074 | 254.9969 | 229.0176 | 193.0506 | 175.0401 | 254.9608 | 229.0183 | 193.0503 | 175.041 | 3.67 | HMDB0029200 | Theo | No |
| phenylacetylglycine | $[M-H]^-$ | $C_{10}H_{11}NO_3$ | 2a | 192.0661 | 192.0631 | 91.054 | 74.0244 | / | / | 91.0561 | 74.0246 | / | / | 4.04 | Metlin 4237 | Exp | Yes |
| p-cresol-glucuronide | $[M-H]^-$ | $C_{13}H_{16}O_7$ | 2b | 283.082 | 283.0818 | 265.0712 | 239.0919 | 175.0243 | 107.0497 | 265.0717 | 239.0977 | 175.0226 | 107.05 | 4.35 | HMDB0011686 | Theo | No |
| p-cresol-sulfate | $[M-H]^-$ | $C_7H_8O_4S$ | 2b | 187.0068 | 187.0065 | 160.9909 | 107.0497 | 105.043 | 80.9646 | / | 107.0499 | 105.0324 | 80.9641 | 4.46 | HMDB0011635 | Theo | No |
| hexanoylglycine | $[M-H]^-$ | $C_8H_{15}NO_3$ | 2b | 172.0968 | 172.0979 | 154.0874 | 128.1081 | 126.0924 | 97.0659 | / | / | / | / | 4.68 | HMDB0000701 | Theo | No |
| cinnamoyglycine | $[M-H]^-$ | $C_{11}H_{11}NO_3$ | 2b | 204.0662 | 204.0656 | 160.0758 | 130.0654 | 117.0693 | 103.0545 | 160.0773 | 130.0654 | 117.0722 | 103.0559 | 5.34 | Metlin 34534 | Exp | No |
| tetranor-5-NO2-CLA | $[M-H]^-$ | $C_{14}H_{23}NO_4$ | 2b | 268.1545 | 268.1554 | 206.1548 | 168.0663 | 154.0507 | 112.0403 | 206.1525 | / | / | / | 5.73 | LMFA01120010 | Exp* | No |
| octenoylglycine | $[M-H]^-$ | $C_{10}H_{17}NO_3$ | 2b | 198.1126 | 198.1136 | 181.087 | 180.103 | 154.1237 | 152.108 | 181.0918 | / | 154.1228 | / | 6.2 - 6.31 | HMDB0094801 | Theo | No |

*(Salvatore et al., 2013)

# Publications

Bhatt, A.P., Pellock, S.J., Biernat, K.A., Walton, W.G., Wallace, B.D., Creekmore, B.C., Letertre, M.M., Swann, J.R., Wilson, I.D., Roques, J.R., Darr, D.B., Bailey, S.T., Montgomery, S.A., Roach, J.M., Azcarate-Peril, M.A., Sartor, R.B., Gharaibeh, R.Z., Bultman, S.J., Redinbo, M.R., 2020. Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1918095117

Nye, Leanne C., Williams, J.P., Munjoma, N.C., Letertre, M.P.M., Coen, M., Bouwmeester, R., Martens, L., Swann, J.R., Nicholson, J.K., Plumb, R.S., McCullagh, M., Gethings, L.A., Lai, S., Langridge, J.I., Vissers, J.P.C., Wilson, I.D., 2019. A comparison of collision cross section values obtained via travelling wave ion mobility-mass spectrometry and ultra high performance liquid chromatography-ion mobility-mass spectrometry: Application to the characterisation of metabolites in rat urine. J. Chromatogr. A 1602, 386–396. https://doi.org/10.1016/j.chroma.2019.06.056

Passmore, I.J., Letertre, M.P.M., Preston, M.D., Bianconi, I., Harrison, M.A., Nasher, F., Kaur, H., Hong, H.A., Baines, S.D., Cutting, S.M., Swann, J.R., Wren, B.W., Dawson, L.F., 2018. Para-cresol production by Clostridium difficile affects microbial diversity and membrane integrity of Gram-negative bacteria. PLOS Pathog. 14, e1007191. https://doi.org/10.1371/journal.ppat.1007191