Metabolomics Approaches for Gastrointestinal Disease Research and Application to Clinical Trials Investigating Inflammatory Bowel Disease and Coeliac Disease

A thesis presented by

Patricia Kelly

Supervised by Dr Nicholas JW Rattray

In fulfilment of the requirement for the degree of

Doctor of Philosophy

2025

Strathclyde Institute of Pharmacy and Biomedical Science

University of Strathclyde

Table of Contents

Thesis Abstract

Chronic inflammatory diseases of the gut, including inflammatory bowel disease (IBD) and coeliac disease (CoD), present major diagnostic and therapeutic challenges due to their heterogeneous clinical presentation, reliance on invasive biomarkers, and limited specificity of existing tests. Liquid chromatography-based mass spectrometry (LC-MS)-based metabolomics provides a powerful means of characterising small molecule signatures of disease and treatment response. This thesis advances the field by establishing optimised high-throughput LC-MS workflows for gastrointestinal metabolomics and applying them to large-scale clinical cohorts to identify novel disease-relevant metabolic alterations.

A monophasic faecal extraction protocol was developed and systemically optimised, ensuring broad metabolite coverage and reproducibility for both untargeted and targeted analyses. Applied to paediatric CoD cohorts, this method revealed three major groups of candidate biomarkers. Firstly, we identified a panel of 12 CoD-specific, non-treatment responsive metabolites spanning bile acids and amino acid derivatives that remain persistently altered despite adherence to a gluten-free diet (GFD). Secondly, we note a group of treatment-responsive metabolites, including amino acid dipeptides and indole and purine related metabolites, which normalised following dietary treatment. Finally, we identified treatment dependent, non-disease-specific metabolites driven by dietary change rather than CoD itself, such as indole-derived compounds and acylcarnitines.

In parallel, a rapid LC-MS workflow for urine was developed and systematically optimised to ensure robust application in large-scale clinical studies. Eight individual parameters were sequentially evaluated, spanning across sample preparation, LC and MS elements of the workflow. This iterative workflow produced a high-throughput protocol with a 6.5-minute data collection time, while maintaining peak resolution, reproducibility, and broad metabolite coverage. This optimised protocol was used to analyse 1094 urine samples from IBD patients and healthy controls, representing the largest urinary metabolomics study of IBD performed to date.

Finally, a critical review was conducted on the impact of food additives on gut inflammation. This synthesis underscored the dual potential of additives as either inflammatory or

therapeutic modulators, reinforcing diet as both a confounder and a therapeutic axis in gastrointestinal disease.

Collectively, this thesis provides methodological advancement that strengthens standardisation in LC-MS-based gastrointestinal metabolomics and delivers biological insights into the pathophysiology of disease. This work contributes to the development of non-invasive biomarkers for future clinical translation of metabolomics in gut health.

Acknowledgements

The completion of this thesis requires a heartfelt thanks to many individuals who have supported me along the way and allowed my love of science to flourish. I am grateful to be surrounded by those who continuously inspire me to dream big, work hard, and embrace every opportunity along the way.

First and foremost, I would like to express my deepest gratitude to my supervisor, Nicholas Rattray, for his unwavering guidance, expertise, and support throughout my PhD journey. Your encouragement, critical insights, and knowledge have been invaluable in helping me grow both as a researcher and as an individual. Your dedication to research and education have inspired me to approach challenges with curiosity and resilience, and to always strive for excellence in everything I do. I am truly fortunate to have had the opportunity to work under your mentorship and for helping me set the foundations for my future career.

I am deeply grateful to Konstantinos Gerasimidis for his mentorship as a secondary supervisor, for the advice and thoughtful feedback that have helped shaped the direction of my research. I would also like to thank the members of my research groups, for the supportive environment, challenging questions, and constructive input, all of which strengthened this project. I sincerely acknowledge the clinical teams involved in facilitating sample collection and ethical approvals and the patients who made this research possible through their generosity and trust.

A special thank you goes to Gillian, our mass spectrometry technician, for her constant help, patience, and good humour in the lab. From keeping the instruments running smoothly to sharing practical advice and encouragement, you made even the challenging days easier. I am truly grateful for your support and kindness throughout the project.

I would like to acknowledge Shimadzu for funding my PhD and thank my industry supervisors Jonathan McGeehan, Chris Titman, and Stephen Brookes, for always supporting my work and providing a constant source of laughter throughout my studies and conference trips.

To my parents, both my best friends and my biggest supporters throughout my life, who have always encouraged me to pursue my love for science and growth in academia. To my mum, you have shaped who I am as an individual and have sacrificed yourself in ways more than I could ever truly capture, always reminding me to never lose sight of what truly matters. My dad, thank you for being my driving force and showing me the value of hard work and resilience, and always ensuring I enjoy the achievements as well as the plans. I am forever grateful for your unconditional support, encouragement, and love, and for being the best parents I could have hoped for. Your belief in me has been a constant source of strength and this achievement is yours as much as it is mine.

A special thank you to my brother Iain, for your constant wisdom, sense of reason, and true kindness. You've always known how to bring balance and keep the mood light when things get challenging, and your ability to make everyone feel at ease is something I truly admire. And to my sister, Roisín, for being the hilarious whirlwind of chaos that you are. You've kept me on my toes every step of the way, ensuring that I never take myself too seriously. Being your sister is an honour that I carry proudly with me.

A heartfelt thank you to my boyfriend, Ross, for being my source of motivation and my greatest cheerleader throughout the final stages of my PhD. You have an incredible way of making me smile, instilling a sense of positivity for the future, and reminding me that I am capable of achieving the things I set my mind to. Whether it's offering a listening ear, sharing our joys, or simply reminding me to take things one step at a time, you've always known exactly what I need. I'm grateful beyond words to have you by my side.

To my friends who have stood by me throughout everything, thank you. Especially to Emma, for being my sidekick every step of the way, my scientific illustration motivator, and someone who I can count on no matter what. And to Zoe, Erin, Maria, and Callum for making my PhD experience so incredible.

And finally, to my Nana. I am grateful for the years I got to spend with you and for the memories that will stay with me forever. I will always remember your kind smile and beautiful soul and you waving us goodbye from your spot at the front window. I will carry the love you gave with me always. I wish you were here to celebrate this achievement with us, but I know you would be proud. Thank you for being such a special part of my life.

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Copyright Statement

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.5. Due to acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: 

Date: 27.02.25

List of Figures

CHAPTER 1

This immunological shift gives rise to an inflammatory state in the gastrointestinal tract. **(C)** Circular model of the chronic nature of IBD. The cycle of inflammation occurring in IBD persists due to the chronic progression and amplification of disease.

Figure 5. Pathophysiology of Coeliac Disease. Coeliac disease is characterised by an inappropriate immune response to dietary gluten, resulting in inflammation and damage to the small intestinal mucosa. In the small intestinal lumen, gluten proteins undergo partial digestion by proteases, producing gliadin peptides. In susceptible individuals, increased intestinal permeability allows translocation of gliadin peptides into the lamina propria. Here, gliadin is deamidated by tissue transglutaminase 2 (TG2). Antigen-presenting cells (APCs) present deamidated gliadin peptides to $CD4^+$ T cells, which become activated and secrete pro-inflammatory cytokines such as IFN-γ, TNF, and IL-2. These T cells also stimulate B cells to produce antibodies against gliadin and TG2. The combined effects of epithelial barrier disruption, cytokine-mediated inflammation, and autoantibody production lead to villous atrophy, crypt hyperplasia, and chronic inflammation.

Figure 6. The Central Dogma of Molecular Biology. The flow of information from DNA to metabolites through transcription, translation, and metabolism, together forming the backbone of -omics fields.

Figure 7. Metabolomics Workflow. Protocols for metabolomic analysis generally follow a method consisting of sample collection, extraction, data acquisition, data analysis, and biological interpretation.

Figure 8. Metabolomics Data Acquisition. Untargeted metabolomics approaches generate hypotheses by identifying as many metabolites as possible in a sample, providing an indication of those which may be involved in disease. Absolute quantification of a pre-defined chemically characterised set of metabolites can be performed by targeted metabolomics to test such hypotheses. MRM, multiple reaction monitoring; QQQ, triple quadrupole.

Figure 9. Schematic representation of the internal architecture of an Orbitrap LC-MS system. The diagram illustrates the integration of liquid chromatography with a mass

spectrometer. In the LC module, analytes are separated based on their physicochemical properties as they pass through a chromatographic column. The eluent enters the ESI source, where analytes are converted into gas-phase ions. These ions are guided through the various ion optics and mass filtering components. AQT, quadrupole mass filter with Advanced Quadrupole Technology; HPLC, high performance liquid chromatography.

CHAPTER 2

Graphical Abstract. Overview of experimental design for metabolomics method optimisation. LC-MS method development was carried out on samples from patients with gastrointestinal disease to maximise metabolite coverage.

Figure 1. The effect of sample weight on features of metabolomic analysis. 1 μL of 20 mg and 50 mg sample was injected onto a C18 column (n = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of sample weight. PCA score plots demonstrating extracted faecal metabolites between different sample weights. Discrimination between 20 mg (blue) and 50 mg (orange) samples was characterised by a variability of 53.1%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of m/z features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each sample weight was assessed. (**f**) A metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in 20 mg and 50 mg samples. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using an unpaired t-test. * $p < 0.05$, ***$p < 0.001$.

Figure 2. Untargeted metabolite class analysis of sample weight. **(A)** Comparison of the total number of metabolites identified by chemical class in 20 mg and 50 mg samples (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in 20 mg and 50 mg samples. Data were expressed as mean ± SEM and statistical significance was assessed using unpaired t-test.

Figure 3. The effect of extraction solvents, MeOH, MeOH/H_2O, and $CHCl_3$/MeOH, on features of metabolomic analysis. 1 μL of each extraction sample was injected onto a C18

column (*n* = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of extraction solvent. PCA score plots demonstrating extracted faecal metabolites between different extraction solvents. Discrimination between extraction solvents MeOH (light blue), MeOH/H_2O (orange), and $CHCl_3$/MeOH (dark blue) was characterised by a variability of 40.2%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m*/*z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each extraction solvent was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using one-way ANOVA. *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 4. Untargeted metabolite class analysis of extraction solvent. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted with MeOH, MeOH/H_2O, and $CHCl_3$/ MeOH (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted with MeOH/ H_2O, MeOH, and $CHCl_3$/ MeOH. Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *$p < 0.05$, **** $p < 0.0001$.

Figure 5. The effect of cellular disruption methods, bead beating, sonication, and freeze-thaw cycles, on features of metabolomic analysis. 1 μL of each extraction sample was injected onto a C18 column (*n* = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of disruption method. PCA score plots demonstrating extracted faecal metabolites between bead beating, sonication, and freeze-thaw cycles. Discrimination between extraction solvents A, bead beating (dark blue); B, sonication (orange) and C, freeze-thaw cycles (light blue) was characterised by a variability of 33.5%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m*/*z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each disruption method was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *$p < 0.05$, ** $p < 0.01$.

Figure 6. Untargeted metabolite class analysis of cellular disruption method. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted using bead beating, sonication, and freeze-thaw cycles (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted using bead beating, sonication, and freeze-thaw cycles. Data are expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA, **** $p < 0.0001$.

Figure 7. The effect of sample-solvent ratio on features of metabolomic analysis. 1 μL of each extraction sample was injected onto a C18 column ($n = 3$), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of sample-to-solvent ratio. PCA score plots demonstrating extracted faecal metabolites between different ratios. Discrimination between extraction solvents 1:5 (dark blue), 1:10 (orange) and 1:20 (light blue) was characterised by a variability of 33.3%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m/z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each sample-to-solvent-ratio was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *$p < 0.05$ ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure 8. Untargeted metabolite class analysis of sample-to-solvent ratio. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted using sample-to-solvent ratios of 1:5, 1:10, 1:20 (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted using 1:5, 1:10, 1:20. Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA., ** $p < 0.01$, **** $p < 0.0001$.

Figure 9. Comparison of individual optimization experiments. Total number of putatively identified metabolites given by optimal parameters of each experiment. Experiment 1, Analysis of Extraction Weight; Experiment 2, Analysis of Extraction Solvent; Experiment 3; Analysis of Cellular Disruption Method; Experiment 4, Analysis of Sample-to-Solvent Ratio.

Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. **p < 0.01, *** p < 0.001, **** p < 0.0001.

Figure 10. PCA of metabolomic profiles based on untargeted analysis of gastrointestinal disease. PCA score plots demonstrating extracted faecal metabolites between patient groups. Principle Component 1 directionality describes the variance between CD (dark blue), CoD (orange) and HC (light blue) and explains 17.7% of the total variance of the data. QCs are shown in green. The samples were performed in triplicate and are shown as individual datapoints to represent the variance in the dataset.

Figure 11. PCA of the metabolomic profiles based on targeted analysis of gastrointestinal disease. PCA score plots demonstrating extracted faecal metabolites between CD (dark blue), CoD (orange) and HC (light blue). The discrimination between (**a**) CD vs. HC, (**b**) CoD vs. HC, and (**c**) CD vs. Co was characterised by variabilities of 34.5%, 31.3%, and 10.5%, respectively. The samples were performed in triplicate and are shown as individual datapoints to represent the variance in the dataset.

Figure 12. Central network analysis of developed metabolite extraction method. Circles shown in green represent metabolites successfully extracted using the developed method and circles shown in red represent metabolites not found using the developed method. C1P, Ceramide-1-phosphate; AA, Arachidonic acid; EPA, Eicosapentanenoic acid; DGLA, Dihomo-gamma linolenic acid.

CHAPTER 2 SUPPLEMENTARY

Figure S1. Untargeted differential analysis of sample weight showing volcano plot of altered metabolites, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased in 50 mg samples compared to 20 mg samples are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

Figure S2. Untargeted differential analysis of extraction solvent showing volcano plot of altered metabolites between **(A)** MeOH vs. MeOH/ H_2O, **(B)** $CHCl_3$/ MeOH vs. MeOH/ H_2O, and **(C)** $CHCl_3$/ MeOH vs MeOH, plotted as log2 fold change vs -log10P. Metabolites that are

significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

Figure S3. Untargeted differential analysis of extraction solvent showing the volcano plot of altered metabolites between **(A)** sonication vs. bead beating, **(B)** freeze-thaw vs. bead beating and **(C)** freeze-thaw vs. sonication, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite levels were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

Figure S4. Untargeted differential analysis of sample-solvent ratio showing volcano plot of altered metabolites between **(A)** 1:10 vs. 1:5, **(B)** 1:20 vs. 1:5 and **(C)** 1:20: vs. 1:10, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased are highlighted red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

Figure S5. Untargeted differential analysis of cell lysis techniques. Volcano plot of **(A)** HC vs. CD; **(B)** CoD vs. CD **(C)** HC vs. CoD, for all patients. Log2 fold change vs. -log10P. Metabolites. Metabolites that are significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

Figure S6. Summary of the developed methodology pipeline. Multi-parameter analysis showed that 50 mg samples give the strongest MS output, and from the extraction solvents analysed, MeOH is the most effective. Additionally, cellular metabolite release is optimal using bead beating as the cell lysis method. Combining optimised parameters provides an experimental protocol for faecal metabolite extraction that can be used for metabolomic analysis.

CHAPTER 3

Figure 1. Overall experimental workflow following stages of sample collection, metabolite extraction using optimised method, data acquisition and analysis, and biological interpretation.

Figure 2. Metabolome profile for the cross-sectional (A) and prospective (B) cohorts. Principal Coordinates–Canonical Variate Analysis (PC-CVA) of the faecal metabolome across study groups (TCD, UCD, and HC) and throughout treatment timepoints (before GFD, 6 months on GFD, and 12 months on GFD) show group centroids and 95% confidence ellipses based on canonical variates. Boxplots show comparison of the CV1 values between groups.

Figure 3. Principal Component Analysis (PCA) of targeted amino acid profiles across study groups. Each point represents an individual sample, and ellipses indicate the 95% confidence interval for each group.

Figure 4. Metabolome profile of children with newly diagnosed coeliac disease compared to healthy controls (n = 82). (A) Scores plot of the orthogonal partial least square discriminant analysis (OPLS-DA) model with R^2Y= 0.555, Q^2= 0.267. (B) Volcano plot of significantly differential faecal metabolites comparing children with newly diagnosed coeliac disease compared to healthy controls, $p < 0.05$, fold change = 2. (C) Box and whisker plots of the top significantly differential faecal metabolites from the untargeted analysis and (D) the targeted analysis.

Figure 5. Treatment-responsive faecal metabolites in patients with CoD. Boxplots showing the relative intensities of metabolites significantly altered between untreated coeliac disease (UCD, red) and treated coeliac disease (TCD, blue). Statistical significance is indicated by *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Figure 6. Treatment responsive metabolites in all participants. Boxplots showing the relative intensities of metabolites significantly altered in both untreated coeliac disease (UCD, red) and HCs (green) compared with treated coeliac disease (TCD, blue). These metabolites

reflect changes driven by a GFD rather than CoD status. Statistical significance is indicated by *p < 0.05, **p < 0.01, ***p < 0.001.

Figure 7. Most influential metabolites. (A) Boxplots of CoD-specific metabolites. (B) Receiver operating characteristic (ROC) curve using the top 10 metabolites from the VIP plot. (C) Combined ROC.

Figure 8. Statistically significant differences (log2 fold change) in metabolite levels between coeliac disease diagnosis and follow-up time points on a gluten-free diet.

Figure 9. Metabolomics analysis comparing treated coeliac disease (TCD) and healthy controls (HC) with unaffected siblings. (A) OPLS-DA scores plot for TCD vs siblings (ellipses = 95% confidence intervals). (B) Volcano plot of differential metabolite features for TCD vs siblings (p < 0.05, fold change > 1.5, log2FC = TCD/S). (C) OPLS-DA scores plot for HC vs siblings (ellipses = 95% confidence intervals). (D) Volcano plot of differential faecal metabolite features for HC vs siblings (p < 0.05, fold change > 1.5, log2FC = HC/S).

CHAPTER 3 SUPPLEMENTARY

Figure S1. Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing UCD vs. HC.

Figure S2. Orthogonal partial least square discriminant analysis (OPLS-DA) and variable importance for treated vs untreated coeliac disease (TCD vs UCD). (A) OPLS-DA scores plot comparing TCD vs UCD with points representing individual samples and ellipses showing 95% confidence intervals, R2Y = 0.767, Q2 = 0.0963. (B) Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing TCD vs. UCD.

Figure S3. Orthogonal partial least square discriminant analysis (OPLS-DA) and variable importance for treated coeliac disease (TCD) vs HCs. (A) OPLS-DA scores plot comparing TCD

vs HC with points representing individual samples and ellipses showing 95% confidence intervals, R2Y = 0.843, Q2 = 0.548. (B) Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing TCD vs. HC.

Figure S4. tTG and PedsQL-GS levels in coeliac disease patients throughout 6 and 12 months on a GFD.

CHAPTER 4

Graphical Abstract. Overview of Experimental Design for Untargeted Urinary Metabolomics Optimisation. Eight parameters were optimised across the protocol, including sample preparation, LC and MS analytical conditions of the experimental pipeline. Outcomes were measured by peak quality attributes, analysis time, and metabolite detection. AGC, automatic gain control; MSMS, tandem mass spectrometry.

Figure 1. Peak Quality Factor (PQF) metric description. The **(A)** zig-zag quality factor, **(B)** FWHM2Base, **(C)** jaggedness, and **(D)** modality quality factors are described, using example peaks from acquired data.

Figure 2. The effect of the extraction solvent on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** the peak performance of creatinine, as measured by the quantification of peak quality factors (PQF)s, Zigzag, FWHM2Base, Jaggedness, and Modality indices, **(C)** average peak rating of all metabolites, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated creatinine peak rating. Creatinine was not detected when IPA/H_2O or MeOH were used as the extraction solvents.

Figure 3. The effect of the dilution factor on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was

further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 4. The effect of chromatography gradient elution time. The effects of the chromatography analysis time on untargeted urinary metabolomics were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** FWHM2Base, **(F)** Jaggedness, and **(G)** Modality indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 5. The effect of injection volume on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by **(F)** the FWHM2Base. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zigzag, jaggedness, and modality indices are not shown as all parameters tested gave a zero value.

Figure 6. The impact of flow rate on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** Zigzag, **(G)** FWHM2Base, **(H)** Jaggedness, and **(I)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method.

Figure 7. The impact of gradient curve on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected

creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 8. Comparison of AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected per full scan AGC parameter, **(B)** the associated peak performance of creatinine, as measured by the quantification of peak quality factors (PQF)s, Zigzag, FWHM2Base, Jaggedness, and Modality indices, **(C)** the number of metabolites detected per MSMS scan AGC parameter, and **(D)** peak performance characteristics for the MSMS scan settings. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 9. Comparison of full scan AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** FWHM2Base, **(F)** Jaggedness, and **(G)** Modality indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 10. Comparison of MSMS scan AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** Jaggedness, **(F)** Modality, and **(G)** FWHM2Base indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

Figure 11. Method Optimisation Overview. Comparison of **(A)** the original 15-minute method and **(B)** shortened 10-minute chromatography gradient.

Figure 12. Final optimised workflow for the analysis of urine samples using untargeted UHPLC-MS and application of the optimised method to a clinical trial for urine metabolomics analysis. Schematic representation of the optimised method, which includes the selection of specific sample preparation and LC-MS parameters to ensure the comprehensive and reproducible profiling of urinary metabolites. Illustration of the clinical trial workflow, consisting of three independent studies which aimed to measure the global urine metabolic profile. The total analysis time of the urine samples for all three studies was 182.3 hours and over 1500 metabolites were putatively identified.

Figure 13. Number of metabolites with coefficient of variation (CV) ≤ 20% across three normalisation methods. The bar plot shows the total number of metabolites meeting a CV threshold of ≤ 20% for non-normalised data, creatinine normalised data, and probabilistic quotient normalisation (PQN) normalised data.

CHAPTER 4 SUPPLEMENTARY

Figure S1. Comparison of chromatograms obtained from untargeted LC-MS analysis using different solvent systems for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – Acetonitrile (1.90E9), red - Acetonitrile/ H_2O (2.5E9), blue – H_2O (2.3E9), orange – IPA (1.81E9), pink – IPA/ ACN (2.22E9), green – IPA/ H_2O (2.07E9), brown – IPA/MeOH (2.05E9), light blue – MeOH (2.39E9), grey – MeOH/ ACN (2.44E9), purple – MeOH/H_2O (2.22E9).

Figure S2. The effect of the extraction solvent on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** Zigzag, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. *Creatinine was not detected when IPA/H_2O or MeOH were used as the extraction solvents.

Figure S3. Comparison of chromatograms obtained from untargeted LC-MS analysis using different extraction solvent dilution factors for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – dilution factor 1 (1.7E9), red – dilution factor 2 (2.21E9), green – dilution facctor 5 (1.7E9), blue – dilution factor 10 (1.34E9).

Figure S4. Comparison of chromatograms obtained from untargeted LC-MS analysis using different injection volumes for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – 0.5 (5.09E8), red – 1 (8.3E8), green – 2 (1.31E9), blue – 5 (2.07E9).

Figure S5. Comparison of chromatograms obtained from untargeted LC-MS analysis using different flow rates for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – 0.25 mL/min (1.86E9), red – 0.3 mL/min (2.42E9), green – 0.4 mL/min (2.31E9), blue – 0.5 mL/min (2.35E9), pink – 0.6 mL/min (2.30E9).

Figure S6. Comparison of chromatograms obtained from untargeted LC-MS analysis using different gradient curve parameter values for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – gradient curve of 3 (2.17E9), red – 5 (1.96E9), green – 7 (1.84E9).

Figure S7. Principal component analysis (PCA) plots comparing three normalisation strategies, evaluating two study groups. (A) Non-normalised, (B) Creatinine normalised, and (C) PQN normalised urine data.

CHAPTER 5

Graphical Abstract. Mechanisms of food additives in the prevention or promotion of gastrointestinal inflammation. Food additives have been shown to influence gut microbial

composition and their released metabolites, reactive oxygen species (ROS) and antioxidant balance, immune function, and epithelial barrier integrity.

Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram of literature database search. Selection criteria including paper identification, screening and eligibility, excluding duplicate papers, unavailable papers, and those unspecific to GI inflammation, and final selection of articles included in the review, according to food additive classification.

Figure 2. Food additives implicated in gastrointestinal inflammation. Food additives, grouped according to their functional class, (colours, preservatives, antioxidants, sweeteners, and emulsifiers) shown to either enhance or mitigate gastrointestinal inflammation. P80; polysorbate 80, CMC; carboxymethylcellulose, SSL; sodium stearoyl lactylate, SMS; sorbitan monostearate, KGM; konjac glucomannan.

Figure 3. Food colourants impacting gastrointestinal inflammation. The inner circle represents food colours that demonstrate inflammatory properties; outer circle represents food colours that demonstrate therapeutic properties, in the context of GI health. Food additives are shown with their maximum absorbance wavelengths: Riboflavin: 440 nm [24], Curcumin: 425 nm [25], Lutein: 445 nm [26], Fast green FCF: 620 nm [27], Brilliant blue FCF: 630 nm [28], Anthocyanins: 520 nm [29], Beta-carotene: 470 nm [30], Tartrazine: 426 nm [31], Sunset yellow FCF: 480 nm [32], Allura red AC: 504 nm [33].

Figure 4. Overview of mechanistic effects of food additives on the GI system. (**A**) Mechanisms of food additives promoting intestinal inflammation. Food additives which exert a pro-inflammatory effect disrupt epithelial barrier integrity, for example through increased intestinal permeability, epithelial cell loss, and decreased mucin production, promoting translocation of bacteria into the intestinal lumen, where an adaptive immune response is elicited. (**B**) Mechanisms of food additives promoting intestinal healing. Food additives which exert an anti-inflammatory effect restore intestinal homeostasis via epithelial barrier and microbial restructuring, which is associated with a downregulated inflammatory immune response, e.g., decreased production of pro-inflammatory cytokines and ROS. Muc2: mucin2; MLCK: mysosin light-chain kinase; 5-HT: 5-hydroxytrypamine; STAT3: signal transducer and

activator of transcription; CAT: catalase; SOD: superoxide dismutase; GSH-Px: glutathione peroxidase; ZO-1: zonula occludens 1; SCFA: short chain fatty acid; COX-2: cyclooxygenase-2; MAPK: mitogen-activated protein kinases; GST: glutathione S-transferase; GSH: glutathione; iNOS: inducible nitric oxide synthase; myeloperoxidase; MDA: malondialdehyde.

Figure 5. Potential opportunities for utilising mechanistic knowledge of food additives, upon further human research and controlled trials. Future applications may include altered intake of food additives, pre/ probiotics, novel targeted therapies, personalised nutrition, exclusive enteral nutrition, and unravelling mechanisms of dietary-associated disease.

Figure 6. Food additives present in EEN formulas, adapted from Logan M et al., 2020 [140]. (**A**) Food additives in EEN formulas with inflammatory potential; (**B**) Food additives in EEN formulas with therapeutic potential.

Figure 7. Roadmap to an evidence-driven future of improved food additive utilisation within the food and pharmaceutical industries.

List of Tables

Table S2. Targeted metabolomics experiment elution gradient. Mobile phase A, 99.9% H_2O + 0.1% formic acid; Mobile phase B, 99.9% Acetonitrile + 0.1% formic acid.

Table S3. Overview of Untargeted Metabolite Identification Levels.

Table S4. List of metabolites included in targeted metabolomics method.

Table S5. Parameters of Compound Discoverer workflow.

CHAPTER 3

Table 1. Patient demographics of cross-sectional and prospective cohorts.

CHAPTER 3 SUPPLEMENTARY

Table S1. Quantification of amino acids in coeliac disease. Values given as mean + SEM.

Table S2. Top differentially abundant metabolites identified in the prospective cohort, with corresponding VIP scores in the cross-sectional cohort.

CHAPTER 4

Table S1. Overall Summary of LC-MS protocol parameter optimisation.

CHAPTER 4 SUPPLEMENTARY

Table S1. Compound Discoverer metabolite filter settings.

Table S2. Compound Discoverer workflow settings.

Table S3. Peak Quality Factor Descriptions.

Table S4. Untargeted metabolomics experiment elution gradient. Mobile phase A, 99.9% water + 0.1% formic acid, mobile phase B, 99.9% ACN + 0.1% formic acid.

Table S5. Coefficient of Variation (CV) values for metabolites detected in three normalisation methods. Metabolites were detected at MSI level 2.

CHAPTER 5

Table 1. Overview of food additive clinical trials in GI health. Clinical trial status information was last verified in September 2025.

Table 2. Overview of enteral nutrition clinical trials in GI health, CD-TREAT; Crohn's Disease Treatment-with-EATing, PEN; partial enteral nutrition, SEN; standard enteral nutrition, EEN; exclusive enteral nutrition. Clinical trial status information was last verified in September 2025.

CHAPTER 5 SUPPLEMENTARY

Table S1. Effects of colours suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory food colours, green E numbers represent anti-inflammatory food colours. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

Table S2. Effects of preservatives suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory preservatives, green E numbers represent anti-inflammatory preservatives. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

Table S3. Effects of sweeteners suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory sweeteners, green E numbers represent anti-

inflammatory sweeteners. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

Table S4. Effects of emulsifiers, thickeners, and stabilisers exerting inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory additives, green E numbers represent anti-inflammatory additives. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

Table of Abbreviations

Abbreviation	Full Form
AA	Arachidonic acid
Ace-K	Acesulfame potassium
ACN	Acetonitrile
ACR	Albumin-to-creatinine ratio
AD	Alzheimer's disease
ADI	Accepted daily intake
AGC	Automatic gain control
ALA	Alpha-linolenic acid
APC	Antigen presenting cell
AQT	Advanced quadrupole technology
ATG16L1	Autophagy-related 16 like 1
AUC	Area under curve
BA	Bile acid
C1P	Ceramide-1-phosphate
CARD	Caspase recruitment domain
CAT	Catalase
CD	Crohn's disease
CD-TREAT	Crohn's disease treatment-with-EATing
$CD4^+CD25^+FOXP3^+$	Forkhead box p3-expressing Treg subset
CDAI	Crohn's disease activity Index
CDEIS	Crohn's disease endoscopic index of severity
CDQ	Coeliac disease specific questionnaire
$CHCl_3$	Chloroform
CKD	Chronic kidney disease
CMC	Carboxymethylcellulose
CoD	Coeliac disease
COX-2	Cyclooxygenase-2
CRP	C-reactive protein
CTLA4	Cytotoxic T-lymphocyte antigen

CVA	Canonical variate analysis
DAI	Disease activity index
DC	Dendritic cell
DDA	Data dependent acquisition
DGLA	Dihomo-gamma linolenic acid
DHA	Docosahexaenoic acid
DSS	Dextran sulfate sodium
EAA	Essential amino acids
EEN	Exclusive enteral nutrition
EFA	Essential fatty acids
EFSA	European food standards agency
ELISA	Enzyme-linked Immunosorbent Assay
EMA	Anti-endomysial antibodies
ENIP	Exclusive or partial enteral nutrition in healthy individuals
EPA	Eicosapentaenoic acid
ERK1	Extracellular signal-regulated kinase 1
ESI	Electrospray ionisation
ESPGHAN	European society for paediatric gastroenterology, hepatology, and nutrition
FA	Formic acid
FC	Faecal calprotectin
FFQ	Food frequency questionnaire
FN1	Fibronectin 1 (FN1)
FWHM2Base	Full width at half maximum to base width
GALT	Gut associated lymphoid tissue
GC-MS	Gas chromatography-mass spectrometry
GFD	Gluten free diet
GIP	Gluten immunogenic peptide
GIT	Gastrointestinal tract
GLP-2	Glucagon-like peptide 2

GLP-R	Glucagon-like peptide- receptor
GRAS	Generally regarded as safe
GSH-Px	Glutathione peroxidase
H&E	Haematoxylin and eosin
H_2O	Water
HC	Healthy control
HLA	Human leukocyte antigen
HMDB	Human metabolome database
IBD	Inflammatory bowel disease
IEC	Intestinal epithelial cell
IFN	Interferon
IGF-1	Insulin-like growth factor 1
iNOS	Inducible nitric oxide synthase
iPENS	Intensive post exclusive enteral nutrition study
JAK	Janus kinase
JAM	Junctional adhesion molecule
JNK	June N-terminal kinase
KGM	Konjac glucomannan
LCFA	Long chain fatty acid
LLE	Liquid-liquid extraction
LPA	Lysophosphatidic acid
LPC	Lysophosphatidylcholine
LRR	Leucine-rich repeat
m/z	Mass-to-charge ratio
MAPK	Mitogen-activated protein kinase
MDA	Malondialdehyde
MDP	Muramyl dipeptide
MeOH	Methanol
MLCK	Myosin light chain kinase
MMP	Matrix metalloproteinase
MMP-2	Matrix metalloproteinase-2
MPO	Myeloperoxidase

MS	Mass spectrometry
MSMS	Tandem mass spectrometry
mTOR	Mammalian target of rapamycin
Muc2	Mucin2
NF-κB	Nuclear factor kappa B
NHDC	Neohesperidin dihydrochalcone
NLRC4	NLR family CARD domain-containing protein 4
NMR	Nuclear magnetic resonance
NMR	Nuclear magnetic resonance
NOD	nucleotide oligomerisation domain containing protein
Notch-1	Neurogenic locus notch homolog protein 1
Nrf2	Nuclear factor erythroid 2–related factor 2
OPLS-DA	Orthogonal partial least squares discriminant analysis
P80	Polysorbate 80
Panx1	Pannexin 1
PC	Phosphatidylcholine
PCA	Principal component analysis
PCA3	Prostate cancer antigen 3
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PedsQL-GS	pediatric quality of life inventory gastrointestinal symptoms
PEN	Partial enteral nutrition
PFPP	Pentafluorophenylpropyl
PG	Prostaglandin
PI3K	Phosphoinositide 3-kinase
PQF	Peak quality factor
PRISMA	preferred reporting items for systematic reviews and meta-analyses
PRR	Pattern recognition receptor
pTregs	Peripherally derived Foxp3+ regulatory T cells
PUFA	Polyunsaturated fatty acid

QC	Quality control
qPCR	quantitative polymerase chain reaction
ROS	Reactive oxygen species
rRNA	Ribosomal ribonucleic acid
S1P	Sphingosine-1-phosphate
SAA	Serum amyloid A
SCFA	Short chain fatty acid
SEN	Standard enteral nutrition
SM	Sphingomyelin
SMD	Small molecule drug
SMS	Sorbitan monostearate
SOD	Superoxide dismutase
SPE	Solid phase extraction
SSL	Sodium stearoyl lactylate
STAT	Signal transducer and activator of transcription
T6SS	Type VI secretion systems
TCD	Treated coeliac disease
TFF	Trefoil factor
TJ	Tight junction
TLR	Toll-like receptor
TNBS	2,4,6-Trinitrobenzene sulfonic acid
TNFAIP3	Tumor necrosis factor alpha-induced protein 3
TNFS12	TNF superfamily member 1
T_{RM}	Tissue-resident memory T cells
tTegs	Thymically derived Foxp3+ regulatory T cells
tTG	Tissue transglutaminase
UC	Ulcerative colitis
UCD	Untreated coeliac disease patients
UCDAI	Ulcerative colitis disease activity index
UHPLC-MS	Ultra-high-performance liquid chromatography-mass spectrometry
V:C	Villous-to-crypt ratio

VEGF	Vascular endothelial growth factor
VIP	Variance in projection
ZO-1	Zonula occludens-1

Research Publications and Posters Presented

Papers Published

Kelly PE, Ng HJ, Farrell G, McKirdy S, Russell RK, Hansen R, Rattray Z, Gerasimidis K, Rattray NJW. *An Optimised Monophasic Faecal Extraction Method for LC-MS Analysis and Its Application in Gastrointestinal Disease.* Metabolites. 2022 Nov 14;12(11):1110.

van den Driest L, **Kelly PE**, Marshall A, Johnson CH, Lasky-Su L, Lannigan A, Rattray Z , Rattray NJW. A gap analysis of UK biobank publications reveals SNPs associated with intrinsic subtypes of breast cancer (2024).

Cunningham MR, Rattray NJW, McFadden Y, Berardi D, Daramy K, **Kelly PE**, Galbraith A, Lochiel I, Mills L, Scott Y, Chalmers S, Lanningan A, Rattray Z. *A roadmap for patient-public involvement and engagement (PPIE): Recounting the untold stories of breast cancer patient experiences* (2024).

Kelly PE, Farrel G, Russell RK, Hansen R, McGrogan P, Edwards C, Broadhurst D, Gerasimidis K, Rattray NJW. Metabolic Dysregulation Driven by Coeliac Disease is Ameliorated by a Gluten Free Diet (in draft).

Kelly PE, Farrell G, Gkikas K, Kerbiriou C, White B, Gerasimidis K, Rattray NJW. Optimisation of a Dilute-and-Shoot UHPLC-MS Method for High-Throughput Urinary Metabolomics (in draft).

Kelly PE, Al Noumas L, van den Driest L, Rattray Z, Gerasimidis K, Rattray NJW. Understanding Food Additives in Inflammatory Bowel Disease: Challenging Perceptions to Improve Gastrointestinal Health (in draft).

Oral Presentations

Developing Metabolomics Approaches in Gut Health. Scottish Metabolomics Network Symposium 2023.

Alterations in the Intestinal Metabolome of Children with Coeliac Disease. European Society for Paediatric Gastroenterology, Hepatology, and Nutrition Conference, Milan 2024.

Metabolomics Applications in Clinical Trials of Gut Disease. University of Strathclyde Cellular Basis of Disease Seminary, February 2025.

Poster Presentations

Targeted Metabolomics Profiling of Inflammatory Bowel and Coeliac Diseases. **Kelly PE**, Ng HJ, Farrell. G, McKirdy S, Gerasimidis K, Rattray NJW (Poster presentation, BMSS, Manchester 2022).

Alterations in the Intestinal Metabolome of Children with Coeliac Disease. **Kelly PE**, Ng HJ, Farrell G, McKirdy S, Gerasimidis K, Rattray NJW (Poster presentation, BMSS, Manchester 2023).

An Optimised Monophasic Faecal Extraction Method for LC-MS Analysis and its Application in Gastrointestinal Disease. **Kelly PE**, Ng HJ, Farrell G, McKirdy S, Russell RK, Hansen R, Rattray Z, Gerasimidis K, Rattray NJW (Poster presentation, Metabolomics and Human Health Gordon Research Conference, Tuscany 2023).

Utilisation of Metabolomics Approaches to Investigate Gastrointestinal Health and Disease (Poster presentation, Metabololomics conference, Osaka 2024).

Molecular Mechanisms of Coeliac Disease and the Effect of a Gluten Free Diet. **Kelly PE**, Ng HJ, Farrell G, McKirdy S, Gerasimidis K, Rattray NJW (Poster presentation, ESPGHAN conference, Milan 2024).

1.0 CHAPTER 1

INTRODUCTION

The gastrointestinal tract (GI), also known as the gut, is a complex system comprising several key structures and organs including the mouth, oesophagus, stomach, small intestine, and large intestine. This vital organ system is responsible for many key functions including digestion, nutrient absorption, and waste elimination, with important roles extending to immune function [1], hormone regulation [2], and the association with other organs to form functional circuit pathways including the gut-brain axis [3] and the gut-liver axis [4]. Approximately 70% of the body's immune system resides in the gut, which makes the gut a unique space due to the requirement to maintain a balance between protecting the body from harmful triggers and establishing immune tolerance and defence.

The human gut is home to around 100 trillion of symbiotic microorganisms [5] ,collectively known as the host microbiota. The collective composition of the intestinal microbiota, referring to bacteria, viruses, archaea, and protozoa, form a complex and heterogenic network which largely impacts intestinal health. With increasing recognition of the wide range of functions conveyed by the gut microbiota in line with the advancement of sequencing technologies, the focus on the microbiome in gut health has soared in recent years. An individual's microbiome is shaped in early life, where it is suggested that a stable community is reached after three years [6]. The large number of factors contributing to microbial acquisition and development, such as delivery type, method of milk feeding, and antibiotic use, provides a microbial composition unique to the individual. Despite the highly complex and specific nature of the microbiome, there are compositional regularities which exist between healthy individuals: A diverse microbiome rich in *Bacteroidetes* and *Firmicutes,* for example, is found in a healthy and well-tolerated intestine [7]. It can additionally be noted that alterations in the composition occur throughout the lifetime of an individual in line with environmental pressures [8] and with increasing age [9].

Coevolution of the microbiota and host has provided many physiological benefits in maintaining a delicate balance between pathogen invasion and self-tolerance. As such, this close interplay has allowed the development of protective mechanisms to prevent T cell responses being directed towards the resident microbiota. In return for the provision of a

stable anoxic environment and substrates by the host, a myriad of crucial functions is given by the microbiome to maintain this symbiotic relationship. Under normal conditions, a stable environment is provided by the diverse community of microbes which, along with the intestinal epithelium, acts as a physical protective barrier to potential pathogens. Enteric pathogens are first exposed to the microbial layer, to which the structure exhibits colonisation resistance mechanisms. Colonisation resistance is a term describing the specific colonisation and growth inhibition mechanisms provided by microbiota to maintain homeostasis [10]. Microbial structure therefore enhances epithelial integrity, which is evident in the finding that the microbial composition largely impacts the maintenance of tight junctions [11]. One hypothesis describing this phenomenon is the competitive exclusion model, which suggests that the host microbes outcompete pathogens for nutrients, resources, and receptors [12]. While this has proven to be a useful model in explaining direct pathogen inhibition, it is important to highlight additional indirect biomechanical mechanisms that mediate colonisation resistance. These include the pH modification, production of antimicrobial peptides and enzymes, and type VI secretion systems (T6SS) [13]. Immunomodulation is another important feature of the gut microbiome via regulation of innate and adaptive immune homeostasis. Microbial composition influences the subsequent immune response via engagement with pattern recognition receptors (PRRs), e.g., Toll-like receptors (TLRs), to initiate appropriate signalling cascades. This crucial crosstalk between the microbiota and the immune system provides the delicately controlled response which exists in homeostatic states.

An increasing understanding of the extent to which the gut impacts overall wellbeing and the association to an increasing number of diseases has placed gut health at a central area of scientific and medical research with widely extending implications. Chronic diseases of the gut are of increasing prevalence, particularly in the UK and Western countries due to lifestyle and environmental factors and pose significant health challenges. Continued research exploring the onset, development, and management of gut diseases is crucial for improved disease management and treatment. Of particular interest in this project is the investigation of inflammatory bowel disease (IBD) and coeliac disease (CoD).

1.1 Inflammatory Bowel Disease (IBD)

Inflammatory Bowel Disease (IBD) is an autoimmune disorder describing chronic inflammation of the gastrointestinal tract (GIT) and involves disruption of normal intestinal homeostasis. This presents symptomatically as abdominal pain, diarrhoea, anaemia, weight loss, and fatigue, which together reduce patients' overall quality of life. IBD follows a relapsing and remitting disease course, consisting of repeating cycles of active flare-ups followed by periods of quiescent disease. These waves of disease activity are unpredictable and show great variation between patients. Although IBD may present at any age, bimodal age distribution of disease onset shows a large peak in adolescents as well as a smaller peak in older adults [14]. In 2019, 4.9 million IBD cases were reported globally [15], with disease incidence and prevalence expected to continue rising, projected at an average annual 5% increase [16].

Two main forms of IBD exist: Crohn's Disease (CD) and Ulcerative Colitis (UC). Although they each follow a similar inflammatory process with overlapping symptoms, the two related conditions have significant differences, primarily in terms of the location and extent of inflammation (**Figure 1**). CD is characterised by transmural inflammation at any point throughout the gastrointestinal tract, although presentation occurs primarily in the colon and ileum, which can extend throughout the entire intestinal wall. This inflammation presents as non-continuous patches, often with deep fissures in affected areas, giving a cobblestone endoscopic appearance with thickened walls. By contrast, the inflammation caused by UC is localised to the large intestine and is limited to the inner lining of the gut. Furthermore, this inflammation is continuous and spread proximally, commonly marked by pseudopolyps. It is important to note that in approximately 5-15% of cases, patients present with features that are insufficient to definitively classify the disease as either CD or UC, leading to a diagnosis of indeterminate colitis (IC) [17, 18].

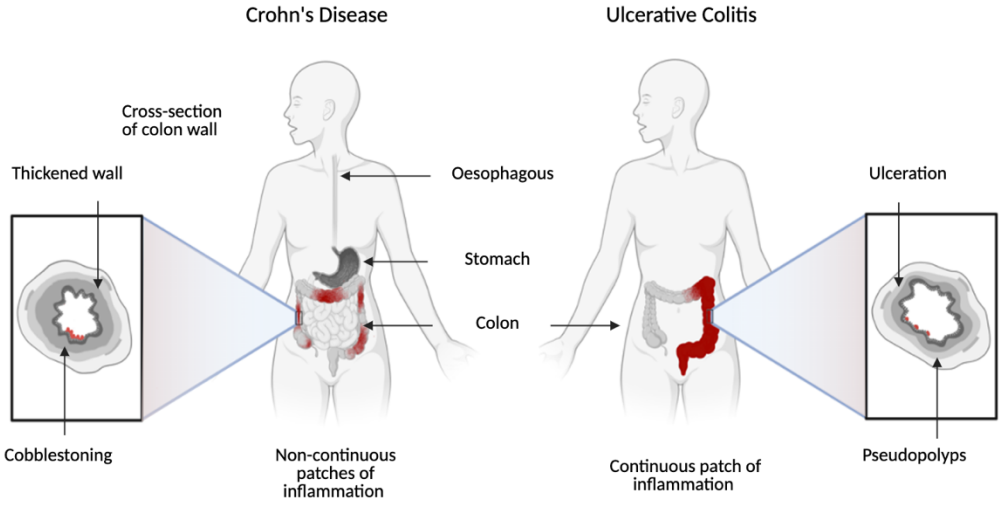


Figure 1. Differences between Crohn's disease (CD) and ulcerative colitis (UC). The anatomical and histological features differ between the two phenotypic forms of IBD. In CD (left), inflammation can occur throughout the gastrointestinal tract and is typically discontinuous. In contrast, UC (right) is restricted to the colon and rectum, with inflammation spreading continuously. Red areas in the schematic represent inflammation.

In line with current guidelines, diagnosis is given based on several factors, including clinical examination, ileocolonoscopy, histology, and blood tests. A full ileocolonoscopy assessment is important to determine the location and extent of inflammation presented. Key features of colonoscopy assessment include anatomical distribution, bowel wall thickness in parallel with vascularisation observations [19], and clinical manifestations. Moreover, during initial colonoscopy evaluation at least two biopsy samples should be taken at five separate sites for histological assessment. Diagnostic limitations still exist, however, with notable inconsistencies in disease interpretation [20]. For example, it has been shown that disease location not only determines subclassification but is also a predictor of disease treatment outcomes [21]. This is further complicated by large interindividual variability at time of diagnosis, due to the wide range of clinical and symptomatic presentations. As the specific diagnosis and classification of disease lead to different treatment and management interventions, an accurate diagnosis is vital in determining the direction of treatment.

Clinical outcomes of IBD can be assessed by several parameters. Mucosal healing, for example, is strongly associated with reduced disease activity, with evidence demonstrating that in early-stage patients mucosal healing is a predictor of remission [22]. Implementing disease evaluation methods into quantitative framework is important for therapeutic

monitoring and informing treatment choices (**Table 1**). This allows clinical response to be defined and compared throughout a time course. The most common index for CD assessment is the Crohn's Disease Activity Index (CDAI), a scoring system developed in 1976 [23] which categorises patients based on clinical outcomes such as abdominal pain, weight change, and complications. While this is the current gold standard for disease evaluation [24], limitations can be noted in the associated subjectivity resulting from the high dependency on patient answers. Furthermore, the large number of variables required for CDAI calculations bring complexities in interpretation, and therefore a simplified version of the CDAI, named the Harvey-Bradshaw Index, was developed, which uses five equally weighted variables to assess disease activity. Inclusion of endoscopic data was identified as an important factor for disease evaluation, which led to the creation of the Crohn's Disease Endoscopic Index of Severity (CDEIS). This scoring system is based on objective parameters from endoscopy assessment. The simple endoscopic score for CD (SES-CD) is another index that has been developed as a simpler and more routine alternative to CDEIS, and measures parameters such as ulcers and surface involvement. Disease activity in UC patients is predominantly assessed by two similar methods: The Ulcerative Colitis Disease Activity Index (UCDAI) and the Mayo score [25]. Both indices use an objective scoring system which include endoscopic data, and while this allows good reproducibility, neither tool has yet been validated. An additional assessment of UC activity that is used is the Truelove and Witts Severity Index classification, although this is much less common at the clinical level.

Table 1. Classification systems used to assess IBD disease activity.

Index	IBD Subclass	Number of Parameters Assessed	Classification	Reference
CDAI	CD	8	Remission: <150 Mild: 150-219 Moderate: 220-450 Severe: >450 Clinical response: Decrease >70	[23]
Harvey-Bradshaw Index	CD	5	Remission: <5 Mild: 5-7 Moderate: 8-16 Severe: >16	[26]
CDEIS	CD	4	Remission: <3 Mild: 3-8 Moderate: 9-12 Severe: >12	[27]
SES-CD	CD	4	Remission: 0-2 Mild: 3-6 Moderate: 7-15 Severe: >15	[27]
UCDAI	UC	4	Remission: 0 Mild: 1 Moderate: 2 Severe: 3	[25]
PUCAI	UC	6	Remission: <10 Mild: 10-34 Moderate: 35-64 Severe: >65	[28]
Mayo score	UC	4	Remission: 0-2 Mild: 3-5 Moderate: 6-10 Severe: >10	[25]
Truelove and Witts Severity Index	UC	6	Mild: <4 Moderate: 4-6 Severe: >6	[29]

While these classification systems provide a basis for clinicians to assess disease severity at different levels, several limitations are inherent in their application. With only a small number of parameters assessed by each of the current systems, an overarching simplicity undernote the calculated outcomes. Several studies have been carried out to investigate associations of disease activity classification systems with molecular evidence of disease

pathology, with mixed results. For example, one research group found no significant correlation between CDAI and numbers of lymphocytes [30]. This indicates that assessing disease activity by clinical parameters alone is not sufficient to provide an accurate evaluation of the true disease state, and therefore it is reasonable to suggest that molecular measurements at the gastrointestinal level could enhance such investigations. Interestingly, comparisons of the different systems are limited, with a general consensus that there are no major consistent correlations between clinical scales, which is likely due to the high subjectivity of the indices.

The onset of IBD is an integrative result of a complex network of factors, including genetics, environment, immune response, microbiota, and metabolite profile (**Figure 2**). This multi-directional relationship presents a complex nature of disease predisposition which together determines the overall trajectory of disease course [31].

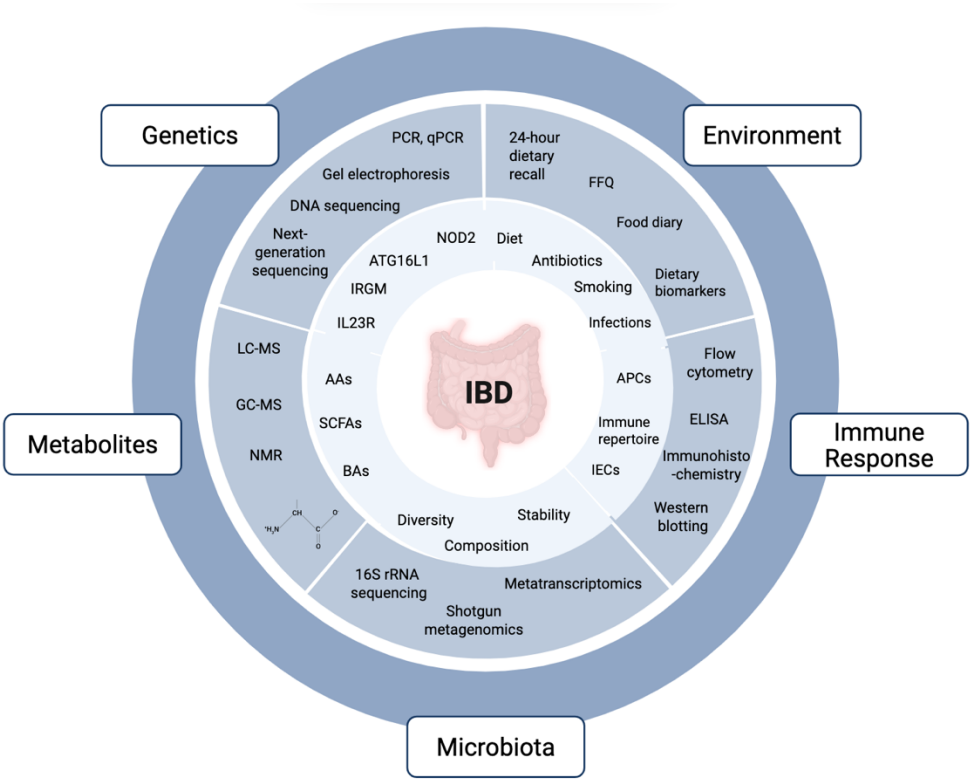


Figure 2. The multifaceted nature of IBD. Complex interactions between genetics, environment, immune response, microbiota, and metabolites result in the onset and development of IBD. 16S rRNA, 16S ribosomal ribonucleic acid; AA, amino acids; ATG161L, autophagy gene 161L; BA, bile acids; ELISA, enzyme-linked Immunosorbent Assay; FFQ, food frequency questionnaire; GC-MS, gas chromatography-mass spectrometry, IRGM, immunity-related GTPase family M; IL-23R, interleukin-23 receptor; LC-MS, liquid chromatography-mass spectrometry; NMR, nuclear magnetic resonance;

NOD2, Nucleotide Oligomerisation Domain containing protein 2; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; SCFAs, short-chain fatty acids.

1.1.1 IBD Risk Factors

IBD is a complex polygenic disorder, with an individual's genetic profile contributing to the risk of developing disease. Several lines of evidence, including familial association [32], twin studies [33], and identification of various susceptibility loci in patients [34], suggest that genetic factors are closely involved in IBD pathogenesis. Nucleotide Oligomerisation Domain containing protein 2 (NOD2) was the first gene to be linked to IBD in the Western population in 2001 by candidate gene analysis [35, 36], with a more recent study suggesting that up to 50% of CD patients carry a mutated form of the NOD2 gene [37]. NOD2 is a cytoplasmic protein expressed in Paneth cells and has important roles in mucosal homeostasis. The leucine-rich repeat (LRR) domain recognises and binds to its ligand muramyl dipeptide (MDP), an immunoreactive peptidoglycan present in bacterial cell walls. This interaction triggers a network of immune responses leading to the activation of nuclear factor kappa B (NF- κB) and production of pro-inflammatory cytokines. Mutations of the NOD2 gene have been associated with IBD susceptibility in European and North American populations. Genome-wide association studies (GWAS), observational research approaches used to detect disease-associated genetic variants, have identified further susceptibility genes with involvement across a wide range of areas including innate immunity, epithelial integrity, drug transport, and adhesion. To date, GWAS have identified 201 susceptibility loci for IBD [34, 38]. Those exhibiting the strongest associations are highlighted in **Table 2**.

Table 2. IBD Susceptibility Genes. NOD2, Nucleotide Oligomerisation Domain containing protein 2; ATG161L, autophagy gene 161L; IRGM, immunity-related GTPase family M; IL-23R, interleukin-23 receptor; TNFSF15, vascular endothelial growth inhibitor; CDH1, Cadherin 1.

Gene	Function of Protein	Mutations	Reference
NOD2	Production of pro-inflammatory cytokines via NF- κB pathway	Arg702Trp Gly908Arg Leu1007fsinsC	[39]
ATG16L1	Autophagy	rs2241879 rs2241880	[40]
IRGM	Autophagy	rs1000113 rs9637876 rs13361189	[41]
IL23R	Differentiation of Th1 and Th17 T-lymphocytes	rs1004819 rs11209032	[42]
TNFSF15	T cell activation and proliferation	rs6478109 rs7848647 rs10817678	[43]
CDH1	Involved in epithelial adherens junction	rs12597188	[44]

The genetic contribution of IBD risk is additionally indicated by patterns of familial predisposition: It is estimated that around 5.2% - 22.5% of IBD patients have an affected family member [45]. Variance in familial risk in IBD can also be noted between populations, for example white populations have the highest prevalence of family history [46].

Genetic involvement is confirmed by evidence from twin studies. One study found the concordance rate for CD in monozygotic twins to be 50%, in comparison to 4% in dizygotic twins [47]. Similar findings were observed in the study underpinning the genetic basis of UC but to a lesser extent, with the concordance rate in monozygotic twins 19% compared to 0% in dizygotic twins. This data ultimately shows that identical twins are more concordant than non-identical twins, highlighting the genetic basis of IBD. It is important to note that these findings are again dependent on population ethnic differences.

Genetic factors alone are not sufficient to account for the development of IBD, and diverse environmental factors also have a critical role in its increasing global incidence. The rate at which IBD prevalence has markedly risen in recent years, estimated at a 33.8% increase between 2006 and 2016 [48], surpasses the rate increase that could be explained by genetic drift. Influential environmental factors include smoking, geographical and social status,

pathogen infections, stress, antibiotic treatment, and diet. For example, evidence indicates that disease incidence is higher in developed countries [49]. Each individual factor has an important role in the predisposition, initiation, and outcome of disease. These interactions are complex and not of equal impact between disease type, however the reasons for this are unclear.

1.1.2 IBD Pathophysiology

The aetiology of IBD is yet to be fully resolved, however, it is widely accepted that exposure to a triggering environmental factor(s) leads to an aberrant immune response in genetically predisposed individuals. This immune response is directed towards the body's own gut microbiota, giving rise to a chronic, autoimmune reaction in the GIT (**Figure 3**). Both innate and adaptive immune responses contribute to the inflammatory process in IBD patients.

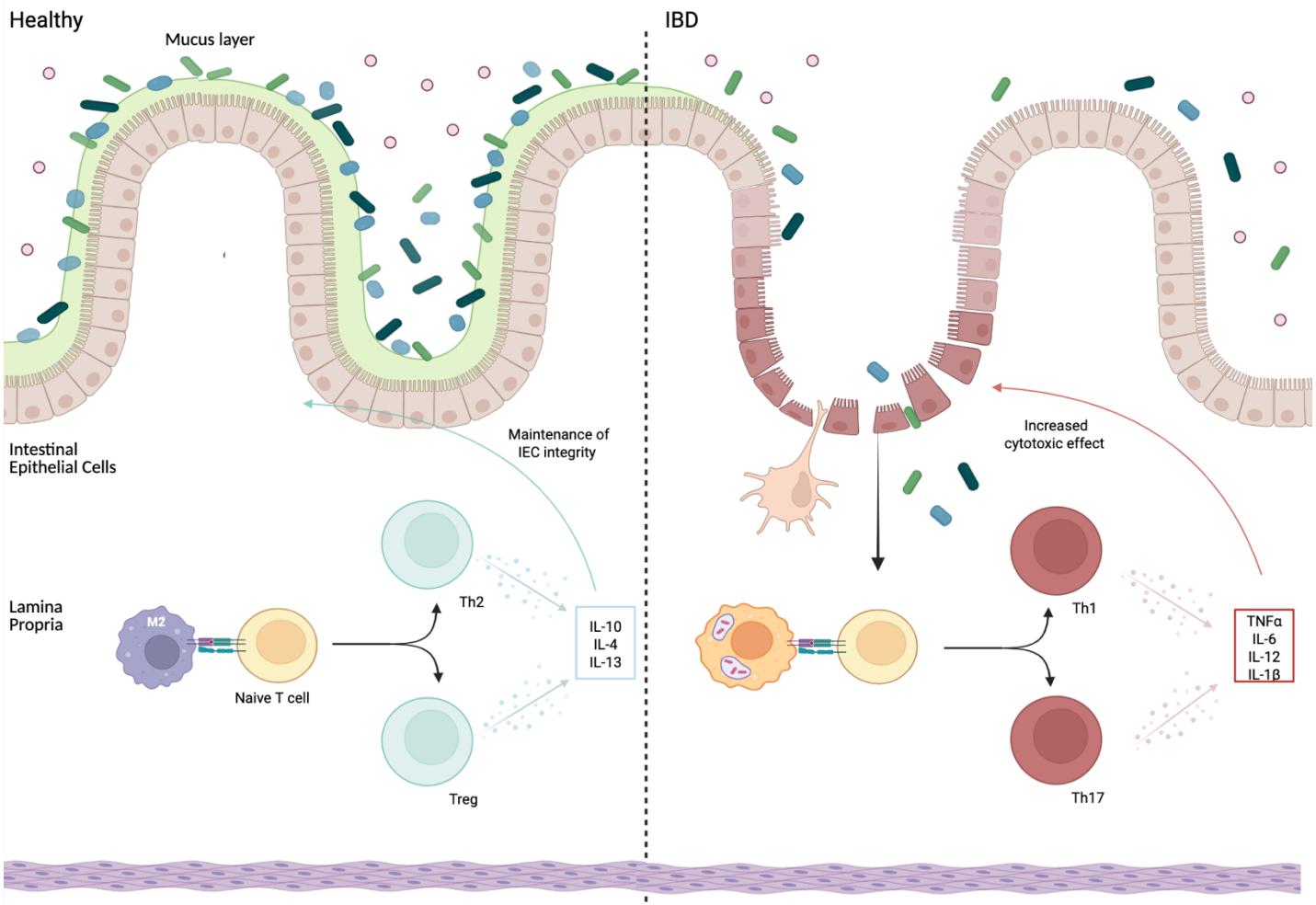


Figure 3. Overview of IBD Pathophysiology. Translocation of commensal bacteria through a degraded epithelial layer initiates an inflammatory cascade predominantly driven by the differentiation and effect of Th1 and Th17 cells. Increased pro-inflammatory signalling molecules feed back into the exacerbation of disease.

1.1.3 Immune Response

The innate and adaptive immune response both play a role in the pathophysiology of IBD. Impairment in the mechanisms which usually exist to protect the body and limit invasion of pathogenic factors initiates the immune response and subsequent inflammatory cascade. Understanding the cellular and molecular dynamics of the immune response to IBD requires a comprehensive toolkit of experimental techniques. For example, flow cytometry and its specialised form, fluorescence activated cell sorting (FACS), are commonly used to quantify immune cell subsets [50]. Visualisation of immune cells within the intestinal tissue can be achieved using immunohistochemistry (IHC) [51], which can reveal spatial information of immune markers. At the molecular level, qPCR [52] and bulk RNA and single-cell RNA sequencing [53] provide quantification of gene expression patterns. Insights into functional immune activity can be enabled by enzyme-linked immunosorbent assays (ELISA) and multiplex cytokine assays [54, 55], which can be applied for analysing both local and systemic immune responses. For example, a recent study integrated transcriptomics, qPCR, and ELISA to show that $IL1B^{+}$ macrophages and $CD14^{+}$ monocytes drive immune dysregulation in IBD [56]. Collectively, these techniques enable a multidimensional analysis of the immune response in IBD, providing an understanding of the stages of pathogenesis.

The intestinal epithelial layer represents an interface between the host and the luminal microenvironment, providing a physical and biochemical barrier to commensal and pathogenic organisms. The large source of potential stimuli that come into contact with the mucosal surface of the GIT, as given by its large surface area meeting the external environment, requires strict control to maintain homeostasis. Normal functioning of the GIT is dependent on the selective permeability of molecules through the intestinal epithelium. As such, the structure of the intestinal epithelium allows the absorption of water and nutrients, without permitting translocation of noxious substances. Together, the physical and biological functions of the epithelial barrier are critical in conserving host-microbe interactions. Breakdown of epithelial regulatory mechanisms is implicated in the development of various diseases, including IBD. Significantly, the loss of integrity of the epithelial barrier allows commensal bacteria to move into the intestinal lumen, where a further immune response is elicited. Loss of epithelial barrier function occurs through several methods.

Intestinal epithelial cells (IECs), including enterocytes, Paneth cells, goblet cells, and neuroendocrine cells, make up the epithelial layer and act as the first line of defence against pathogens in the gut. Several diverse regulatory mechanisms are given by the epithelial cells to maintain tissue homeostasis, in addition to its role in providing a physical barrier between the intestinal lumen and the extracellular milieu of the body. Enterocytes make up the majority of small IECs and maintain epithelial integrity, as determined by their structural properties and cell polarity [57]. The importance of enterocyte function in IBD is demonstrated by findings of increased enterocyte apoptosis in patients [58]. The same study found significant differences in the percentages of apoptotic enterocytes in inflamed intestinal areas compared to non-inflamed areas, highlighting their involvement in active disease pathogenesis.

Paneth cells, located in the small intestinal crypts, produce granules which hold antimicrobial peptides and immunomodulating proteins. These include α-defensins, lysozyme C, phospholipases, and C-type lectins, which have a critical role in host defence. The function of defensins, for example, is to inhibit bacterial action via membrane pore formation. The altered function of α-defensins in IBD, as shown by their attenuated expression in parallel with reduced antibacterial activity in patients [59] provides evidence signifying Paneth cell involvement in the disease process.

A variety of mucins and peptides with important roles in growth and repair are produced by Goblet cells [60]. Intestinal epithelial barrier integrity depends on mucus production; the colonic inner mucus layer provides a continuous protective coating over the gastric mucosal surfaces and modulates intestinal homeostasis. 29 core proteins form the mucus barrier, which aid in the limitation of pathogen exposure and influences multiple cell interactions and signalling pathways [61]. Of significant note, the secretion of mucin 2 (Muc2), the major component of mucus, is decreased in IBD patients [62]. Other prominent associations between mucus component composition and increased epithelial permeability in IBD patients include decreased glycosylation products [63] and trefoil factors [64]. Trefoil factors 1, 2, and 3 (TFF1-3) are peptides with important roles in the protection and repair of IECs; the process of restitution, for example is dependent on their function. Further supporting evidence is provided by histopathology analysis, revealing that the mucus layer is

significantly thinner in IBD patients [65]. One study found that the thickness of the mucus layer was decreased by an average of 144 and 135 µm between controls and CD and UC patients, respectively [66]. As the first anatomical point of contact for bacteria, such modulation of the mucosal layer resulting in decreased thickness and increased penetrability has important implications in disease onset and progression.

The epithelial lining is also precisely regulated by multiprotein complexes known as tight junctions (TJs), formed by assembly of the proteins claudin, occludin and junctional adhesion molecule (JAM) and connected to the actin cytoskeleton via zonula occludins [67]. The phosphorylation and expression of these TJ proteins influence the protective outcome of the barrier. The overall TJ structure contributes to epithelial barrier integrity by controlling paracellular transport of molecules between cells [68] and protecting against inflammatory stress stimuli. Alterations in gut junctional complexes are a key feature of both active and quiescent IBD, with evidence demonstrating that patients have decreased phosphorylation levels of TJ proteins [69]. Loss of the regulatory mechanisms provided by TJs which control paracellular permeability therefore damages barrier integrity and promotes exposure to harmful luminal contents such as microbial antigens, toxins, and dietary components, thereby triggering mucosal immune activation and inflammation.

The described mechanisms ultimately lead to the priming of naïve CD4+ T cells into specific inflammatory subgroups, characterised by the cytokine production profile [70]. The adaptive immune response is a system of specific regulatory mechanisms following antigen presentation, primarily driven towards the polarisation of T cells. In 1991, defined patterns of T cell profiles in response to different stimuli were established in humans [71]. Progression of the field has since provided an understanding of distinct T cell subsets and their roles in the immune response [72]. A key underlying theme is the categorisation of Th1 and Th17 subsets as pro-inflammatory, in contrast to anti-inflammatory Th2 and Treg subsets. As recognised in many diseases including IBD, this has important implications in inflammatory research. T cell differentiation models have become extensively studied hallmarks of IBD pathogenesis. In a healthy intestine there is a balance between pro-inflammatory Th1 and Th17 subtypes, and anti-inflammatory Th2 and Treg cells. It is understood that under conditions of IBD, this balance is offset [73]. Inflammation results from excessive production of pro-inflammatory cytokines and chemokines such as TNF-α,

IL-1β, and IL-6, in line with downregulation of protective pathways including IL-10 and TGF-β. Several adaptive immune mechanisms associated with the T cell differentiation paradigm have been reported in disease development. These intracellular signalling pathways give a diversity in cytokine response that requires strict control to maintain a delicate balance between protection and inflammatory mechanisms. In IBD, the altered T cell polarisation favouring pro-inflammatory responses drives the pathogenic cascade that contributes to the gastrointestinal inflammation. Moreover, while the Th1, Th17, Th2, and Treg subsets are well evidenced adaptive elements of IBD pathogenesis demonstrating strong association to disease, recent years has seen emergence of evidence suggesting the role of other T cell phenotypes in pathogenesis (Th9, Th22, Tfh, cytotoxic $CD8^{+}$). The current most widely accepted hypothesis of IBD pathogenesis can therefore be generally stated as excessive effector T cell function and/ or deficient regulatory T cell function.

As IBD is a chronic autoimmune disease, inflammation is persistent over long periods of time meaning that pro-inflammatory cytokines released in inflamed tissues further propagate the immune response, acting as a feedback loop to allow continuation of disease activity [74]. Such overactivation of the immune response causes additional breakdown of epithelial barrier function, which in turn increases translocation of bacteria and exacerbates the inflammatory response. This persistent cycle gives IBD its chronic nature, which results in tissue damage unless the cycle can be broken by successful intervention.

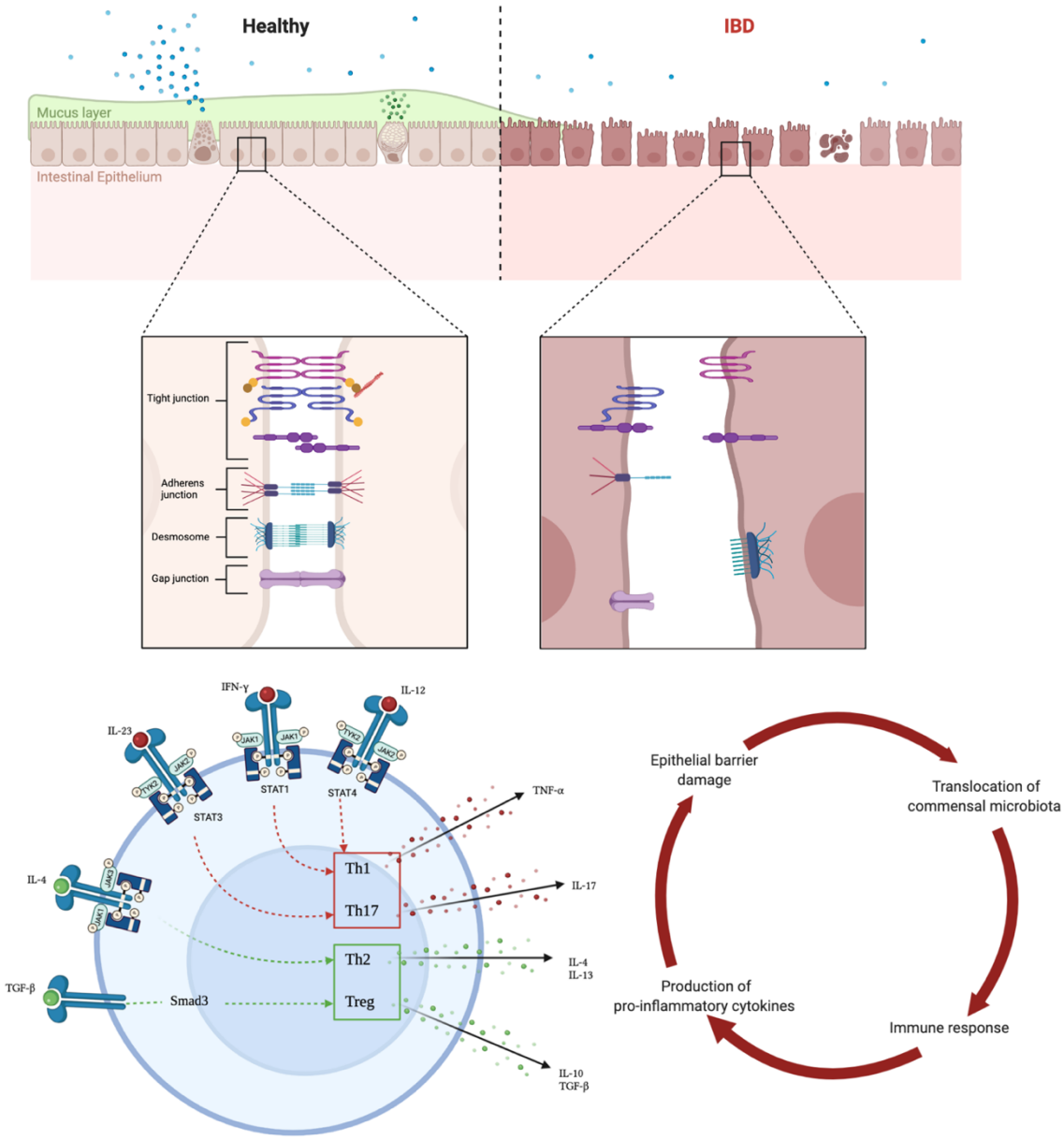


Figure 4. Immune mechanisms implicated in IBD pathogenesis**. (A)** Intestinal barrier dysfunction and downregulation of tight junction proteins in IBD. Schematic illustrating compromised intestinal epithelial integrity in IBD, which is indicated by mechanisms including increased enterocyte apoptosis, reduced numbers of granules containing anti-microbial peptides, decreased thickness of the mucus layer, altered enteroendocrine cell expression and hormone secretion, and decreased tight junction proteins (right panel). **(B)** Cytokine production and inflammation. An imbalance between pro-inflammatory (red) and anti-inflammatory (green) pathways results in an increase in pro-inflammatory cytokines. This immunological shift gives rise to an inflammatory state in the gastrointestinal tract. **(C)** Circular model of the chronic nature of IBD. The cycle of inflammation occurring in IBD persists due to the chronic progression and amplification of disease.

1.1.4 Role of the Microbiome in IBD

Changes to the structure, composition, and function of the gut microbiome is an integral part of IBD pathogenesis. This alteration of gut microbiota, termed dysbiosis, changes the intestinal homeostatic dynamic and is associated with intestinal inflammation. This can be attributed to the loss of the symbiotic microbiome-host relationship and therefore also the protective functions that come with it. To investigate the dynamic interactions between the microbiota and host in IBD, a range of high-resolution techniques can be used to profile the microbiome. One of the most widely used methods for profiling microbial communities is 16S ribosomal RNA (rRNA) gene sequencing, which leverages the highly conserved nature of the 16S rRNA gene of prokaryotes and can be used to compare bacteria present within a given sample [75]. In IBD research, this provides a foundational tool for identifying microbial signatures associated with aspects of disease, including inflammation, disease severity, and treatment response [76]. Shotgun metagenomics is a technique used to provide comprehensive insights into the full repertoire of microbial taxa using untargeted sequencing of genomic content and is frequently used in IBD research [77, 78]. Often used to complement metagenomics, metatranscriptomics captures the active gene expression of microbial communities through sequencing of total microbial RNA [79].

Not surprisingly, the microbiome has become a subject of considerable interest in IBD research. Accumulating evidence has shown an altered microbial composition in patients, and while no definitive uniform profile of disease has been identified, some common trends have emerged. There is a general observation of an increase in harmful species and decrease in protective species under disease conditions [80]. **Table 3** displays the primary microbial species alterations observed in IBD patients. These key changes to important microbial species demonstrate a link between microbial composition and disease state. It is difficult, however, to differentiate association from causation, and microbiome complexity in humans brings challenges in providing strong evidence for causation.

Table 3. Dysbiosis of gut microbial composition in IBD. Alterations in the composition of bacteria phyla and species in IBD patients. There is an overall increase in harmful species and a decrease in protective species.

Phylum	Species Increased in IBD	Species Decreased in IBD	Reference
Bacteroidetes		*Alistipes putredinis* *Bacteroides coprocola* *Bacteroides uniformis* *Bacteroides cellulosilyticu* *Bacteroides intestinalis* *Parabacteroides goldsteinii*	[81]
Firmicutes	*Ruminococcus gnavus* *Lactobacillus*	*Dialister invisus* *Faecalibacterium prausnitzii* *Eubacterium hallii*	[82] [83] [84] [85]
Actinobacteria		*Bifidobacterium adolescentis*	[86]
Proteobacteria	*Escherichia coli* *Klebsiella pneumoniae*		[87] [88]
Fusobacteria	*Fusobacterium*		[89]
Verrucomicrobia		*Akkermansia muciniphila*	[90]
Uroviricota	*Caudovirales*		[91]

Moreover, microbial compositional analysis reveals a decrease in the overall diversity of commensal bacteria in IBD patients. One study reported that on average, IBD patients have 25% fewer microbial genes than healthy individuals [92]. This is significant as the diverse community of microbes that make up the protective barrier determines the overall capacity to limit exposure of pathogens. Thus, a decreased diversity is associated with a reduced ability to regulate entry of harmful microorganisms. This is consistent with reports of increased epithelial barrier degradation in IBD patients. Furthermore, a correlation between loss of species diversity and the disease activity of CD patients confirms the involvement of the microbiota at the clinical level.

Differences in gut microbiota composition and function at the intra-individual level are found depending on the specific anatomical region of the GIT. This is due to location-dependent microbiological differences such as oxygen tension, pH, and digestion flow rate, which generate overall conditions suitable for different microorganisms. It is significant to note that a bidirectional interaction exists between microbial alterations and inflammation. The microbial composition is also altered by mucosal inflammation through various mechanisms, including oxidative stress [93]. Furthermore, it has been shown that colonic inflammation changes luminal bacterial gene expression in mice models of IBD [94]. Although microbial dysbiosis is evident as a central component of disease pathogenesis, these findings also reinforce the uncertainty around the chronology of IBD pathogenesis. This reintroduces the pathogenic feedback model in IBD, demonstrating an undetermined structure of cause and consequence in the sequence of events. While a precise causal relationship is unknown, the involvement of commensal microbiota in IBD is clear nonetheless and therefore continued focus on specific mechanistic roles will enable further progress in the field.

1.1.5 Metabolism and IBD

There is a crucial role of metabolic programming and specific metabolic pathways in the development and perpetuation of intestinal inflammation, with recent evidence demonstrating the metabolic nature of IBD [95]. It is shown that metabolic dysregulation is a central feature of IBD pathogenesis, influencing the immune response, epithelial integrity, and microbial interactions [96]. Understanding these biochemical processes helps to unravel disease mechanisms and provides a foundation for novel metabolism-targeted therapeutic strategies in IBD.

The interplay between immune activity and cellular metabolism is a key feature of IBD pathogenesis, highlighting how metabolic dysregulation actively drives disease progression. As noted earlier, chronic activation of the mucosal immune system is a hallmark of IBD, and this activation is accompanied by profound shifts in immune cell metabolism to support sustained inflammatory responses [97]. Under normal, homeostatic conditions, Tregs and M2 macrophages primarily rely on oxidative phosphorylation and fatty acid oxidation to

meet their energy needs [98]. These pathways support anti-inflammatory functions and maintain immune tolerance in the gut. However, in the inflamed gut environment characteristic of IBD, there is a metabolic reprogramming whereby immune cells shift toward a glycolysis-dominant profile, especially in effector T cells (Teffs) and M1 macrophages [99, 100]. Activated Teffs upregulate glucose transporter 1 (GLUT1) to enhance glucose uptake and fuel aerobic glycolysis, a process where glucose is fermented to lactate even in the presence of oxygen [101, 102]. This shift, known as the Warburg effect, not only meets the high energetic and biosynthetic demands of inflammation but also results in lactate accumulation and acidification of the local tissue environment [103]. These conditions further potentiate inflammation by promoting the expression of pro-inflammatory cytokines such as IL-17 [104]. Concurrently, oxidative phosphorylation becomes impaired in both immune and epithelial cells during inflammation. For epithelial cells, which are highly dependent on mitochondrial energy production, this results in a significant energy deficit [105]. As a consequence, essential functions such as nutrient absorption, mucosal barrier maintenance, and tissue repair are compromised, further exacerbating the cycle of inflammation and tissue damage seen in IBD.

1.2 Coeliac Disease

Coeliac disease (CoD) is a chronic autoimmune disorder of the small intestine, characterised by an abnormal immune response to gluten that leads to intestinal damage and nutrient malabsorption. The incidence of CoD has increased in recent years, with an estimated 1.4% of the global population affected [106]. Gluten is composed of two main proteins, gliadin and glutenin, which are found primarily in grains such as wheat, barley, and rye, and help provide the structure and functions of gluten, particularly in food production and baking. Gliadin, a prolamin rich in proline and glutamine responsible for the stretchiness of dough, is the main component implicated in the triggering of an immune response in CoD [107]. While beneficial for food structure, the gliadin proteins that make up gluten trigger an unregulated immune reaction in genetically predisposed individuals. Currently, the only effective treatment of CoD is the strict adherence to a gluten-free diet (GFD) [108].

CoD is a complex disorder with various clinical presentations based on histopathological findings, clinical presentation of disease, and response to treatment. Five types of subclassification exist: classical [109], non-classical [110], silent [111], refractory [112], and potential CoD [113] (**Table 4**). Each classification of disease represents a different inflammatory manifestation which can result in subtle differences in symptoms and response to treatment. Understanding the different subtypes is therefore essential for an accurate diagnosis, effective treatment strategies, and overall patient outcomes. Histopathological classification systems are applied to help indicate each clinical phenotype, including the Marsch, Marsch-Oberhuber, and Corazza systems [114].

Table 4. Clinical Classifications of Coeliac Disease.

Coeliac Disease Classification	Description	Symptoms	Reference
Classical	The most common form of CoD characterised by GI symptoms from small intestinal villous atrophy.	Abdominal pain Bloating Weight loss Chronic diarrhoea Malabsorption	[109]
Non-classical	Characterised by a potential lack of typical CoD-associated-GI symptoms and presence of extra-intestinal manifestations.	Anaemia Fatigue Weakness Osteoporosis Joint pain	[110]
Silent	Charactered by the presence of small intestinal damage and positive serological markers without noticeable symptoms.	No overt symptoms	[111]
Refractory	A severe form of CoD where symptoms persist despite strict adherence to a GFD	Abdominal pain Bloating Weight loss Chronic diarrhoea Severe malabsorption	[112]

	for at least 12 months.		
Potential	Characterised by positive serological markers without intestinal damage or symptoms at the time of diagnosis.	No overt symptoms	[113]

In addition to the existence of multiple disease subtypes, CoD can be challenging to diagnose due to symptom overlap with other gastrointestinal disorders such as non-coeliac gluten sensitivity [115] and wheat allergy [116]. Furthermore, there is a requirement for gluten to have been consumed regularly (typically at least one-three slices of bread or equivalent per day) for at least 6-8 weeks prior to diagnosis for an accurate outcome [117-119]. This can be challenging for many individuals, as they may experience severe symptoms when reintroducing gluten, or have anxiety and fear around the risk of long-term damage, which makes compliance difficult. The variability in the gluten immune response additionally means that some individuals may react strongly to small amounts of gluten, making it challenging to consume the recommended intake required for a diagnosis. As a result, CoD is commonly underdiagnosed or misdiagnosed [120-122], which brings further complications if left untreated, CoD can lead to long-term complications and nutrient deficiencies, including vitamin deficiencies and malignancies.

A combination of diagnostic tools is therefore required in combination with evidence of clinical manifestations to obtain an accurate diagnosis. Firstly, serological tests are used to measure levels of IgA anti-tissue transglutaminase antibodies (tTG) and anti-endomysial antibodies (EMA), which are produced in response to gluten-activated immune pathways. Although these measurements provide an insight into the inflammatory immune response, their measures are not specific to CoD. Therefore, if positive blood tests are obtained, endoscopic evaluation is required to examine small intestinal damage, with disease presence indicated by duodenal villous atrophy and crypt hyperplasia [123]. To confirm diagnosis, small intestinal mucosal biopsies are taken for histopathological assessment, where application of a classification system, such as the Marsh classification system, are then used to characterise disease, as described above. As per the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) guidelines, a biopsy is

required for a confirmed CoD diagnosis in adults, however in children, a biopsy may be skipped if all three of the following criteria are met: (1) the tTG-IgA levels are greater than 10 times the upper limit of normal, (2) a second antibody test such as EMA-IgA is positive, and (3) symptoms consistent with CoD are presented [124]. If any one of these criteria are missing, a small intestinal biopsy is also required to confirm a CoD diagnosis.

While advances have been made to improve the diagnosis of CoD in recent years, the challenges associated with the current process remain significant. This is highlighted by the prevalent diagnostic delay that has been revealed, with an average time to a CoD diagnosis in the UK of 13 years [125], which has a huge impact on quality of life [126]. There is a clear need for better diagnostic approaches for CoD patients, with a core issue identified in its early diagnosis. There is growing interest in the potential for earlier identification of disease during its prodromal phase, prior to the occurrence of overt mucosal damage or clinical symptoms [127, 128]. This preclinical window represents a critical frontier in CoD research, where predictive screening strategies could significantly reshape the course of disease. This potential is demonstrated in studies showing that antibody seroconversion can precede presentation of symptoms in genetically predisposed individuals. The prospect of integrating genetic risk profiling and novel biomarkers such as gut microbiome and metabolome signatures holds promise for determining disease presence prior to any pathophysiological presentation. Achieving this goal would not only allow for timely dietary intervention but could also prevent long-term complications and shift the clinical paradigm from reactive to proactive care.

1.2.1 CoD Risk Factors

Genetic predisposition is a key determinant in CoD, with Human Leukocyte Antigen (HLA)-DQ8 and HLA-DQ-2 established as primary genetic risk factors of CoD [129]. It has been shown that more than 90% of patients carry the DQ2 allele and the majority of the remaining 10% carry the DQ8 allele. Non-HLA genes have additionally been found to contribute to disease susceptibility, although at a much lower prevalence. Over 40 non-HLA risk loci have been identified by genome-wide association studies (GWAS), many of which have important implications in immune regulation pathways [130]. Examples include

cytotoxic T-lymphocyte antigen 4 (CTLA4) [131] and tumor necrosis factor alpha-induced protein 3 (TNFAIP3) [132].

While CoD is primarily a genetically driven disease and the presence of HLA-DQ8 and HLD-DQ-2 haplotypes is necessary for the development of CoD, it is not sufficient to cause disease on its own as 30-40 % of the general population carry them without developing CoD [133]. Gluten exposure is the essential environmental risk factor triggering onset of the disease; however, research has revealed that aspects of gluten exposure in early life may contribute to disease risk and severity, with evidence suggesting that the early introduction or the delayed introduction of gluten impacts CoD risk [134]. Uncertainty still surrounds this relationship, as other studies have demonstrated the absence of a strong correlation between early gluten consumption and disease onset in later life [135], and therefore further research is required to confirm this hypothesis. Certain infections have also been implicated in increasing the risk of CoD. For example, early childhood infections with enterovirus [136] and rotavirus [137] have been linked to an increased risk of CoD. Additional environmental risk factors for CoD have been suggested, including antibiotic use [138], breastfeeding [139], and pregnancy outcome [140], however conclusive evidence for these are limited. Therefore, no specific recommendations can currently be provided on optimal gluten introduction or breastfeeding duration for CoD prevention.

1.2.2 CoD Pathophysiology

The pathophysiology of CoD is charactered by the dysregulation of gluten processing and resulting activation of the immune response, leading to intestinal damage [141]. In healthy individuals, gluten processing follows a typical digestive pathway without triggering an immune response. This consists of the breakdown of gluten into smaller peptides in the stomach by pepsin and gastric acid, followed by further enzymatic breakdown by peptidases in the small intestine [142, 143]. In CoD, however, gluten processing is significantly affected. Partial breakdown of gluten by gastric acid and pepsin results in the inability of enzymes to fully degrade the gluten peptides, particularly gliadin, which resist further breakdown and accumulation in the lumen [144-146]. Gliadin is then able to stimulate zonulin, a protein which increases the permeability of the small intestine's epithelial layer and has therefore

also recently been investigated for use as a biomarker for the development of CoD [147]. The increased small intestinal barrier permeability allows gliadin peptides to translocate through the epithelial barrier into the lamina propria [148, 149].

Immune activation occurs following gliadin translocation into the lamina propria and ultimately results in the inflammation and intestinal damage that is characteristic of CoD [150]. The mechanisms by which this takes effect can be first explained by the deamidation of gliadin by tTG, increasing its immunogenicity. Antigen presenting cells (APCs) present deamidated gliadin to CD4+ T cells via HLA-DQ2 or HLA-DQ8, which triggers the production of pro-inflammatory cytokines such as IL-2, IL-6, IFN-γ, and TNF-α. Further elements of the immune response shown to be upregulated in CoD include the production of anti-tTG-IgA autoantibodies by B cells, endomysium, and gliadin peptides, and the attack of enterocytes by cytotoxic IELs driven by IL-15 activation.

The resulting inflammation leads to small intestinal epithelial damage, with key histological and structural features. Villous atrophy, defined as the flattening and loss of the villi lining the small intestine and resulting decreased absorptive surface area and subsequent malabsorption, is one of the key defining features of active CoD [151]. This is primarily assessed through duodenal biopsy via upper endoscopy and can be measured using the villous-to-crypt ratio (Vh:Cd ratio), where a normal Vh:Cd ratio is defined as ≥3:1, meaning that the villous height is at least three times the crypt depth. This ratio is reduced due to villous atrophy and informs categorisation according to the Marsh classification system. This measurement carries additional importance due to the finding that persistent villous atrophy can predict the development of complications and mortality in adult CoD patients [152]. However, a lack of correlation between the degree of villous atrophy and the severity of disease presentation brings challenges in using this feature for specific disease phenotyping and management. Crypt hyperplasia refers to the elongation and deepening of the intestinal crypts due to increased proliferation of intestinal epithelial cells and is suggested to reflect an attempt to compensate for villous loss [153]. The Vh:Cd ratio can also be used to measure crypt hyperplasia, with additional quantification provided by microscopic measurement of crypts under haematoxylin and eosin (H&E) staining, where a significant increase in crypt depth observed in CoD patients in comparison to healthy individuals [154]. Inflammation can additionally be measured by quantification of IELs, with

an increase in the number of cells observed in CoD patients [155, 156], even at early stages of pathogenesis [157]. Persistence of these inflammatory features of CoD lead to the resulting disease symptoms such as abdominal pain, diarrhoea, bloating, and weight loss.

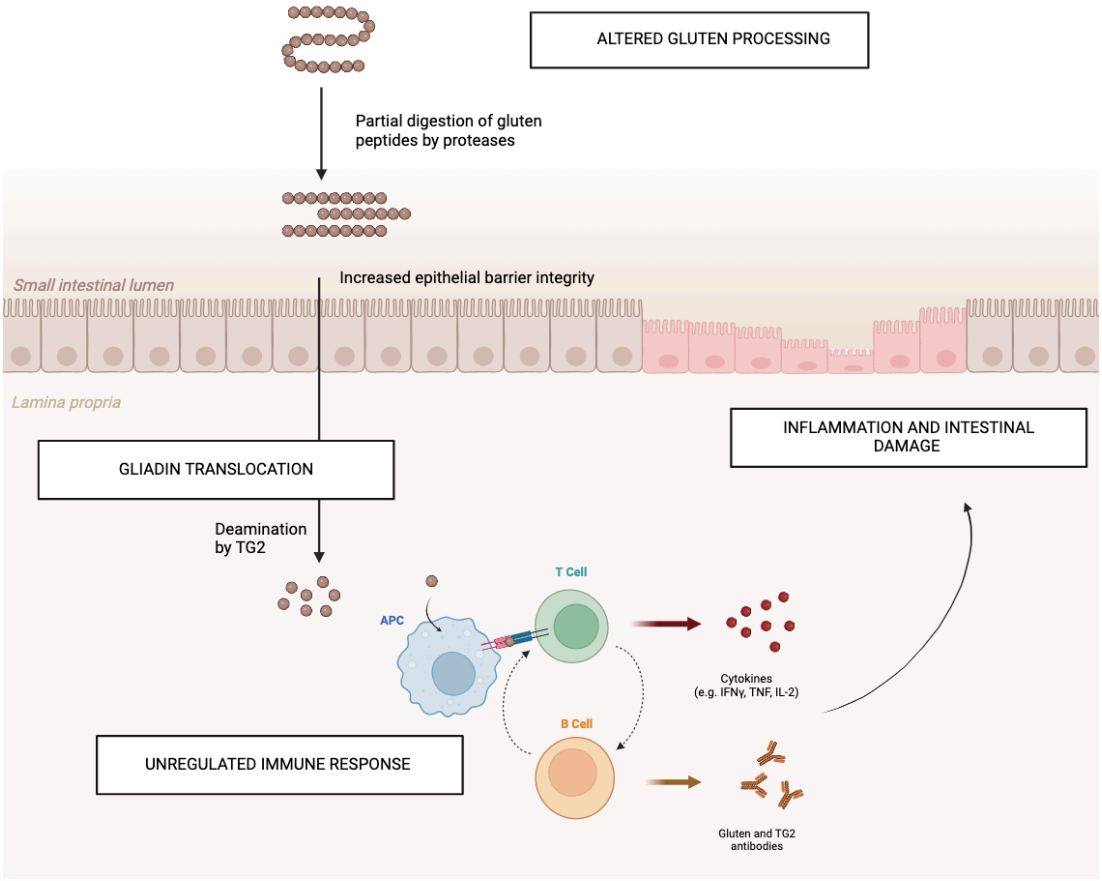


Figure 5. Pathophysiology of Coeliac Disease. Coeliac disease is characterised by an inappropriate immune response to dietary gluten, resulting in inflammation and damage to the small intestinal mucosa. In the small intestinal lumen, gluten proteins undergo partial digestion by proteases, producing gliadin peptides. In susceptible individuals, increased intestinal permeability allows translocation of gliadin peptides into the lamina propria. Here, gliadin is deamidated by tissue transglutaminase 2 (TG2). Antigen-presenting cells (APCs) present deamidated gliadin peptides to $CD4^+$ T cells, which become activated and secrete pro-inflammatory cytokines such as IFN-γ, TNF, and IL-2. These T cells also stimulate B cells to produce antibodies against gliadin and TG2. The combined effects of epithelial barrier disruption, cytokine-mediated inflammation, and autoantibody production lead to villous atrophy, crypt hyperplasia, and chronic inflammation.

1.2.3 Role of the Microbiome in CoD

Increasing attention has been given to the role of the gut microbiome in CoD, with distinct alterations noted in the microbial composition of CoD patients in comparisons to healthy controls. It is not yet certain whether this dysbiosis precedes disease symptoms or is a consequence of disease pathophysiology, however recent research has provided a strong case for the latter [158]. The microbial alterations characteristic of CoD may have important implications by influencing the development and severity of the disease through several potential mechanisms including modulation of the immune response, increasing intestinal barrier permeability, altered enzymatic modulation of gluten, and inflammatory environment regulation via altered short-chain fatty acid (SCFA) production.

Table 5. Dysbiosis of gut microbial composition in CoD. Alterations in the composition of bacteria phyla and species in CoD patients.

Phylum	**Species Increased in/ Associated with CoD**	**Species Decreased in CoD/ Used as a Potential Intervention**	**References**
Bacteroidetes	*Bacteroides fragilis* *Bacteroides dorei*	*Bacteroides ovatus* *Bacteroides vulgatus*	[159] [160]
Firmicutes	*Lactobacillus spp.* *Enterococcus spp.*	*Faecalibacterium prausnitzii* *Clostridium leptum*	[161] [162] [163] [164]
Actinobacteria	*Collinsella spp.* *Eggerthella lenta*	*Bifidobacterium spp.*	[165] [166] [167]
Proteobacteria	*Escherichia coli* *Klebsiella spp.*		[167] [168]
Fusobacteria	*Fusobacterium spp.*		[169]
Verrucomicrobia	*Akkermansia muciniphila*		[170]

1.2.4 Metabolism and Coeliac Disease

While the immunological and genetic underpinnings of CoD are well-documented, emerging evidence highlights the significant role of dysregulated metabolism in its pathogenesis. Altered energy metabolism [171], mitochondrial stress [172] and metabolic shifts in immune and epithelial cells [173] have been shown to contribute to the inflammatory environment in CoD. The small intestine is a metabolically active tissue that relies on tightly regulated energy metabolism for absorption, barrier function, and epithelial renewal. In CoD, villous atrophy and crypt hyperplasia profoundly disrupt this balance, leading to nutrient malabsorption and impaired metabolic function at the cellular level. Studies have shown that epithelial cells in CoD patients exhibit mitochondrial dysfunction**,** evidenced by reduced oxidative phosphorylation and increased oxidative stress [174]. This metabolic imbalance not only contributes to impaired epithelial integrity [175] but also perpetuates inflammation through the production of ROS and the activation of NF-κB [176]. The immune activation in response to gluten observed in CoD is a metabolically demanding process and requires a reprogramming of immune cell metabolism to support cytokine production, proliferation, and effector functions. Once activated by gliadin-derived peptides, $CD4^+$ T cells in the lamina propria exhibit a metabolic phenotype dominated by glycolysis [177, 178]. The upregulation of GLUT1 supports their rapid expansion and the production of pro-inflammatory cytokines such as IFN-γ which contribute to tissue damage, promote B cell activation, and support the generation of anti-tTG2 autoantibodies [179].

Several metabolites derived from host and microbial metabolism have been shown to influence disease inflammation and pathology via specific mechanisms. For example, glutamine, a critical fuel for enterocytes, is depleted in active CoD [180, 181], impairing epithelial regeneration and contributing to barrier dysfunction [182]. Abnormalities in lipid metabolism have also been observed, with some studies indicating altered expression of genes involved in fatty acid uptake and β-oxidation [183, 184].

Metabolism in relation to nutrition is a vital aspect of CoD research, with malnutrition observed in patients both at diagnosis and while under treatment. One recent study revealed that malnutrition was prevalent in 8.3% of CoD patients, with additional findings that these patients have a 108% higher risk of mortality [185]. The intestinal damage caused

by gluten-induced inflammation leads to malabsorption and impact systemic metabolic pathways, influencing overall health. It has been shown that fat and carbohydrate metabolism in CoD patients is altered in comparison to healthy individuals [186, 187]. This is due to the villous atrophy which occurs during disease pathogenesis, where the normal villous architecture becomes flattened and loss of absorptive area, resulting in reduced nutrient absorption, including carbohydrates. The efficiency of disaccharidase enzymes such as lactase, maltase, and sucrase in the brush border of enterocytes is also impaired [188, 189], and therefore complex carbohydrates are not adequately hydrolysed into monosaccharides. This subsequently leads to the accumulation of undigested carbohydrates in the intestinal lumen, which are fermented by gut bacteria, altering their metabolite production [190]. Fat metabolism can be compromised in CoD patients due to defects in bile acid reabsorption [191], resulting in deficiencies in fat-soluble vitamins (A, D, E, K) [192]. It is also important to note that a GFD, while effective at resolving intestinal inflammation, may itself influence systemic metabolism. Therefore, dietary management of coeliac disease must balance gluten elimination with the maintenance of metabolic health. Further investigation into the metabolic effects of both the disease itself and of a GFD are required to understand true impacts on metabolism and health.

1.3 Diet and Gut Health

The relationship between diet and gut health is profound, with the intake of food components having a direct impact on the microbiome, immune system, and overall health. A significant finding in dietary research is the observed loss of certain gut microbial species in humans, with the resulting loss in overall microbial diversity [6, 193]. This parallels trends of increasing non-communicable diseases, including IBD and CoD in a similar timeframe. Research into the role of diet in the predisposition of gut disease and in disease management is therefore crucial. Significantly, diet plays a key role in the pathogenesis of IBD, influencing disease risk, progression, and clinical outcome [194]. Dietary research into CoD is predominantly focuses on the dysregulated processing of gluten and the impacts of a GFD. Continued diet-focused investigation into gut health and disease is a powerful step towards improving patients' quality of life through disease management and also in the prevention of chronic diseases.

1.3.1 Diet and IBD

The association between diet and IBD can firstly be demonstrated by evidence linking disease incidence to the Western diet. This is supported by the increasing incidence and prevalence of IBD in conventionally low-incidence areas, such as Asia where a Western diet is becoming progressively adopted [195]. The Western diet refers to a modern nutritional lifestyle with characteristics relating to common food signatures and processing procedures, encompassing a diet high in fat, processed sugars/ sweeteners, and protein and low in fruits and vegetables. The changes associated with the Western diet have evolved in line with ancestorial adaptations associated with the agricultural and industrial revolution and impact risk of IBD to modern populations [196]. It is also noteworthy that human migration and evolution have shaped our digestive systems, as humans migrated from Africa and settled in diverse environments, adapting to diverse diets through natural selection. These adaptations occurred over thousands of years, allowing our alimentary canals and gut microbiota to co-evolve with regional food sources [197, 198]. In contrast, the rapid adoption of modern Western diets has outpaced these adaptations, potentially contributing to this rise of IBD. Several other dietary intake patterns have been associated with IBD risk and outcome (**Table 6**). In its broadest sense, these findings can be expressed through guidelines of eating a well-balanced diet ensuring a variety of lean proteins, healthy fats, fruits, and vegetables, with avoidance of processed foods.

Table 6. Impact of Specific Diets on IBD.

IBD Impact	Diet	Features	Reference
Increases risk/ exacerbates inflammation	Western	High content of protein, saturated fat; low fruit and vegetable content	[199]
	High salt	High salt content	[200]
	High fat	High fat content	[201]
Decreases risk/ promotes healing	Vegetarian	Elimination of meat, fish, and poultry	[202]
	Vegan	Elimination of animal products	[202]

Low FODMAP	Reduced consumption of short-chain carbohydrates	[203]
Paleo	Limited consumption of processed foods; high content of lean meats, fish, fruits, and vegetables	[204]
IBD Anti-inflammatory Diet (IBD-AID)	Elimination of trans-fats, wheat, corn, lactose, and sucrose	[205]
Mediterranean	Increased fruits, vegetables, and whole grains; decreased meat and dairy.	[206]
Specific Carbohydrate Diet	Elimination of grains; low in sugar and lactose	[207]
Low residue	Limited consumption of high-fibre foods, nuts, seeds, raw fruits, and vegetables	[208]

While identification of dietary trends impacting IBD have provided general suggestions for decreasing risk and improving symptoms, they are a long way from providing specific criteria in a therapeutic sense. Strong interindividual variability in response makes it difficult to identify specific appropriate dietary habits. Despite considerable research focusing on advancing knowledge in this field, specific data obtained from human studies are both limited and conflicting. This inconsistency is particularly evident when translating findings into concrete dietary recommendations, as studies often differ in patient populations, intervention types, and outcome measures. This limits their practical application in clinical guidance and reflects the great complexities of the relationship between diet and gastrointestinal health.

In a 2017 partnership collaboration involving multidisciplinary clinicians, patients, and organisations supporting patients to identify the top priorities of IBD research, the impact

of diet was recognised as one of the highest rated subjects requiring further research [209]. The results identified diet as the central topic of the third and seventh overall top priority questions, which address the role of diet in disease management and optimising treatment, respectively. This framework therefore places diet amongst the current most important research areas of IBD. Furthermore, it has been revealed that malnutrition is a disease-associated feature for up to 85% of IBD patients [210]. It is therefore vital to elucidate the mechanisms by which dietary factors impact disease state.

Several specific food components, including macronutrients, micronutrients, and food additives, have been suggested to impact IBD. Macronutrients are the nutrients providing energy that are required in large quantities, comprising three main groups, fat, protein, and carbohydrates. Carbohydrates serve as the primary energy source, with effects on gut health depending on the type and source. Complex carbohydrates including fruits, vegetables and grains provide dietary fibre and have been shown to support gut health [211]. On the other hand, refined carbohydrates are associated with gut dysbiosis and gut inflammation [212]. Fats play a crucial role in energy metabolism and cellular function, with their type and composition shown to significantly influence gut health and inflammation. For example, saturated fats, commonly found in processed foods, have been linked to an exacerbated inflammatory response [213], whereas monounsaturated fats and polyunsaturated fats have been associated with beneficial immune modulation and anti-inflammatory effects. Protein is essential for immune function, tissue repair, and maintaining muscle mass, making it an important dietary consideration for IBD. Micronutrients encompass the vitamins and minerals required by the body which are required only in small quantities, including iron, folate, vitamin B12, vitamin D, and zinc. It has been suggested that these micronutrients could play a key role in the management of IBD by supporting immune function, reducing inflammation, and preventing complications associated with malnutrition. It has been revealed that micronutrient deficiencies are observed in greater than 50% of IBD patients [214]. In addition to macro- and micro-nutrients, the relationship between diet and IBD incidence can has linked to the increasing use of food additives in processed foods. Despite legislations requiring an FDA classification of Generally Regarded as Safe (GRAS), evidence indicates that food additives may be key drivers of gastrointestinal inflammation [215]. As such, recent years have seen a growing

body of evidence indicating the involvement of common food additives in inflammation [216-218], which has propelled an interest in their role in IBD.

1.3.2 Nutritional Interventions for IBD

Standard treatment for IBD aims to supress the inflammatory immune response and control associated disease symptoms. As there is no cure for IBD, current treatments aim to induce and maximise time in complete remission. This is done using varying forms of medical strategies, including biologics and small molecule drugs [219], as well as dietary changes and nutritional therapies [220]. It can be noted that the choice of treatment is dependent on many factors, including disease severity and previous treatment responses, and different treatments may be required throughout the disease course. This is particularly important when considering the two stages of IBD, active and quiescent, as different strategies may be applied for treating a disease relapse compared to maintaining remission. One crucial component in effective disease management is the diet, which is key area of interest to patients, researchers, and clinicians. With the additional advantage of providing selective and non-invasive methods of patient management, this has placed an increasing interest in non-pharmacological approaches to managing IBD such as nutritional interventions. This makes dietary therapies an attractive option for the treatment of IBD patients.

Exclusive enteral nutrition (EEN) is a nutritionally complete liquid dietary therapy used as the primary treatment for paediatric CD patients in Europe [221, 222]. The liquid-only diet consists of a variable nutritional composition, including carbohydrates (22.8%-89.3%), protein (7.8%-30.1%), and fat (0%-52.5%) [223]. EEN formulas are also fortified with essential vitamins and minerals to provide the required micronutrients. Treatment consists of following the liquid-only diet for a specified period, which is typically 6-12 weeks, and has been shown to induce clinical remission in approximately 80% of patients [224]. The exact mechanisms by which EEN exerts its therapeutic effects are not yet fully understood; however, several potential mechanisms have been proposed. Microbial modulation and subsequent reduction in gut inflammation have been hypothesised as a consequence of EEN treatment [225, 226]. Further research has shown that gut microbiome and metabolome signatures can predict response to EEN treatment [227]. The anti-inflammatory nature of EEN is additionally supported by evidence demonstrating the

reduction in levels of inflammatory markers such as IL-6 [228]. Furthermore, it has been suggested that the consistent nutrient provision by EEN consumption may result in mucosal healing and restoration of epithelial barrier function [229, 230]. Despite its effectiveness, EEN is a very restrictive diet which presents several limitations and challenges. Adherence to the liquid diet can be difficult for patients, particularly with the associated length of treatment which is not always well tolerated due to taste and palatability preferences. While EEN is effective at inducing remission in the short-term, it's long-term use is constrained by these practical limitations, making sustained adherence challenging [231].

To overcome some of the limitations associated with EEN, researchers developed the Crohn's Disease Treatment with EATing (CD-TREAT) diet [232]. CD-TREAT is a solid food-based nutritional intervention designed to mimic the beneficial effects of EEN while allowing for a more varied and palatable diet. Based on the idea that specific food components influence the gut microbiome composition and gut inflammation, CD-TREAT replicates EEN by excluding dietary components such as gluten, lactose, and alcohol, and matching fibre and other macro- and micro- nutrients. The use of CD-TREAT as a therapeutic strategy for CD patients has shown to be successful, with the CD-TREAT diet showing improved tolerability and inducing similar effects on the microbiome and metabolome in comparison to EEN. Metabolomics applications have the potential to shed light on the mechanism of EEN by revealing how it influences host metabolism and gut microbial properties. Metabolites such as SCFAs and bile acids, for example, are key mediators in gut health and inflammation, both of which are central to EENs therapeutic effects. Profiling these molecules at different stages of treatment can allow identification of physiological changes and potential biomarkers associated with clinical outcomes. This molecular insight can help clarify how EEN exerts its anti-inflammatory effects, ultimately refining its application and enhancing its clinical use.

1.4 Metabolomics

Metabolomics, which can be broadly defined as the study of small molecule intermediates and end products in a biological system, allows the analysis of molecular patterns at a specific timepoint. This is particularly beneficial as the metabolite profile represents the endpoint of gene expression in the "central dogma of molecular biology." This flux of genetic information, first graphically demonstrated by Francis Crick in 1957, describes the path from DNA to the biological phenotype (**Figure 6**). While investigations into each of these stages is pivotal in elucidating disease processes, it is key to note that alterations in mRNA levels may not cause changes in protein levels, and similarly, changes in protein concentration are not guaranteed to cause associated alterations in protein function. Analysis of changes at the metabolite level may therefore provide a more accurate reflection of the true biochemical state at a specific timepoint. Thus, the metabolite profile is considered the closest representation of the phenotype.

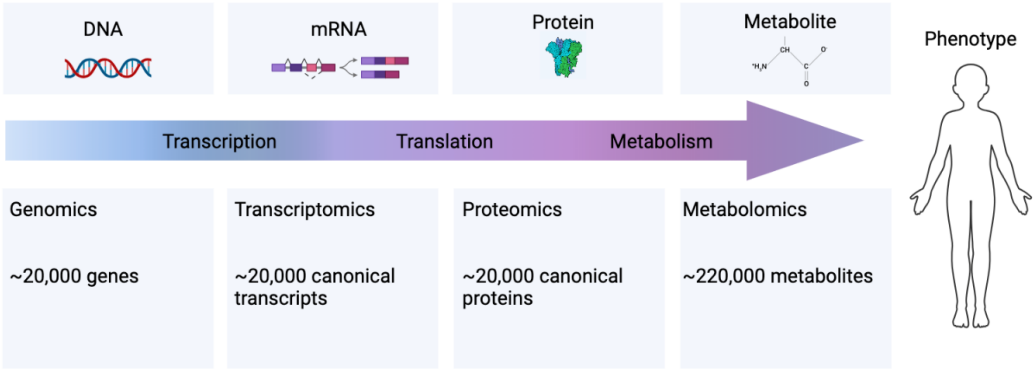


Figure 6. The Central Dogma of Molecular Biology. The flow of information from DNA to metabolites through transcription, translation, and metabolism, together forming the backbone of -omics fields.

The accurate analysis of metabolites in biological samples requires a reproducible and robust approach. A metabolomics workflow consisting of several stages is typically followed in such investigations, consisting of study design, sample collection, extraction, data acquisition, data analysis, and biological interpretation (**Figure 7**). It is important to recognise that metabolomics is fundamentally a hypothesis-generating approach and therefore the putative targets or metabolic signatures uncovered require rigorous validation in independent cohorts.

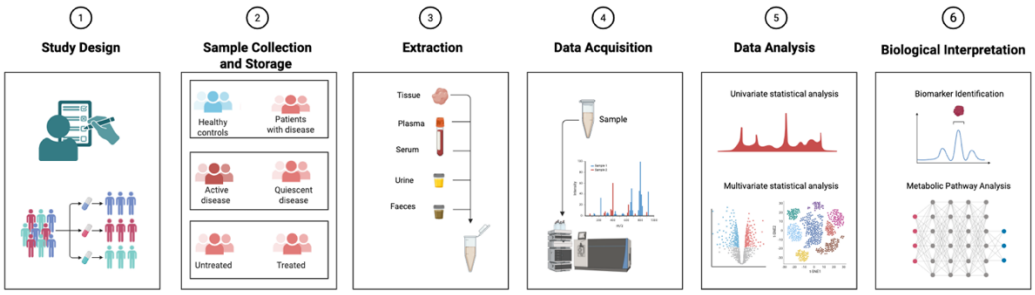


Figure 7. Metabolomics Workflow. Protocols for metabolomic analysis generally follow a method consisting of sample collection, extraction, data acquisition, data analysis, and biological interpretation.

1.4.1 Study Design

Designing a metabolomics study requires careful consideration of various factors to ensure that the data obtained are robust, reproducible, and meaningful. The study design should consider the biological questions being asked, the types of metabolites of interest, the methodologies used for metabolite measurement, and the statistical tools required for data analysis [233]. Defining the research objectives is an important starting point, as the hypothesis will guide sample selection and grouping, ensuring that experimental groups are well-defined and take into consideration factors such as age or disease status. Additionally, sample size is a critical factor to consider for a metabolomics experiment in ensuring reliability and statistical power. A power analysis helps determine the minimum number of samples required to detect significant differences [234, 235], with the appropriate sample ensuring that findings are robust and reproducible. This not only strengthens the study's conclusions but also supports the identification of reliable biomarkers and metabolic pathways with high confidence, leading to more impactful scientific insights.

1.4.2 Sample Collection and Storage

Sample collection from the appropriate patients and controls for the study is an important first step in the metabolomics workflow. During the design of metabolomics experiments there are numerous considerations for optimal analysis which set the trajectory of subsequent interpretations and conclusions. Firstly, understanding human metabolome

variation is crucial for sample collection and storage choices. Variables to be considered during sample collection include the gender and age of subjects, environmental factors such as diet and exercise, and time of sample collection. Metabolic variations are evident between gender and age groups, with one third of serum metabolites suggested to be different between females and males [236] and over 1500 ageing-related differences identified on The MetaboAge database [237]. Additionally, as circadian rhythmicity is known to influence metabolite profile, it is also important to note the time of day that samples are collected. Taking these steps to recognise potential variations in downstream metabolic output will allow for more accurate and representative data interpretations.

A range of biological samples can be used in metabolomics experiments; blood, urine, faeces, and tissue are common in IBD studies. The choice of sample is specific to the aims of the study and the biological system under investigation, which will ultimately influence interpretation of the data. Upon collection, most samples should be stored at -80°C to ensure sample stability [238]. There are certain recommendations for each sample, for example tissues should ideally be snap-frozen in liquid nitrogen immediately after collection and homogenised [239]. The timing of preparation stages is essential as delays in freezing can lead to changes in metabolism which will be reflected in the LC-MS analysis. Aliquoting samples into multiple replicates and thus minimising the number of freeze-thaw cycles is also recommended.

1.4.3 Metabolite Extraction

It is imperative that pre-analytical processing of samples is carried out efficiently as the quality of data obtained is largely dependent on this stage. Extraction methods aim to release the metabolites from the sample, providing a smaller, concentrated volume for analysis. Effective metabolite extraction is essential for elution of metabolites of interest. Liquid-liquid extractions (LLEs) are commonly used in metabolomics experiments, which utilise solvent solubility and immiscibility for partitioning. The sample size, extraction solvent, and reconstitution solvent used in the extraction method will influence the detected metabolite pool. However, it is important to recognise that no single metabolomics experiment can capture the full scope of the metabolome, estimated at over 220,000 metabolites. Each methodological decision inherently selects for a small fraction of

this total pool, with pre-selection bias inevitable due to factors such as solubility, stability, and ionisation efficiency. Therefore, careful optimisation and awareness of these limitations are essential for experimental study design and data interpretation.

1.4.4 Data acquisition

Acquiring data through metabolomics analysis can be done in two distinct approaches: untargeted metabolomics, a comprehensive exploration of the total number of measurable metabolites in a sample, and targeted metabolomics, the measurement of pre-defined molecules. Untargeted metabolomics, otherwise known as global metabolomics, aims to reproducibly analyse as many compounds as possible under specific analytical conditions [240]. Untargeted metabolomics studies are typically followed by targeted metabolomics to validate and precisely quantify selected metabolites of interest (**Figure 8**). These studies use optimised methods with reference standards to ensure accurate identification. In this way, untargeted metabolomics acts as a hypothesis generating approach, guiding the development of focused targeted assays to support biomarker validation and clinical translation.

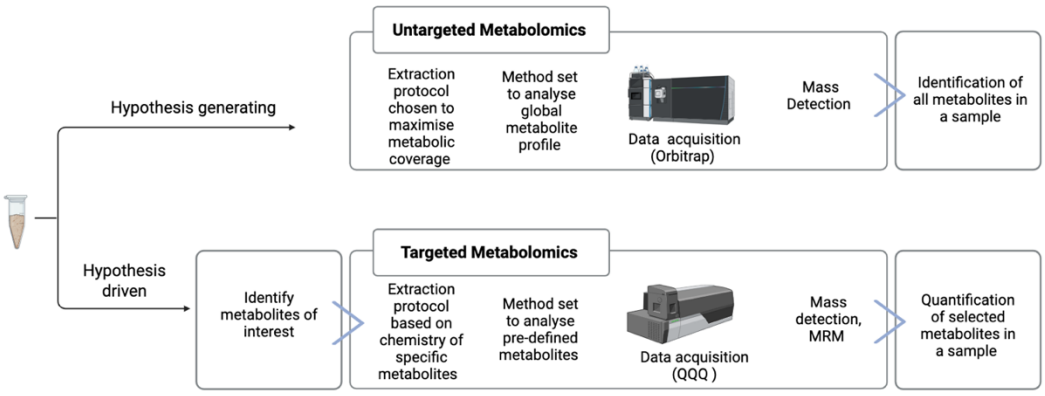


Figure 8. Metabolomics Data Acquisition. Untargeted metabolomics approaches generate hypotheses by identifying as many metabolites as possible in a sample, providing an indication of those which may be involved in disease. Absolute quantification of a pre-defined chemically characterised set of metabolites can be performed by targeted metabolomics to test such hypotheses. MRM, multiple reaction monitoring; QQQ, triple quadrupole.

Advances in technology have led to the development of methodologies to detect a broad range of metabolites in different biomatrices, thereby allowing in-depth study of the GIT. One of the main platforms used in the detection and quantification of metabolites is mass spectrometry (MS). MS, when used in a metabolomics context, is a high throughput analytical technology which measures the mass-to-charge ratio (m/z) of molecular ions. When MS is used in combination with separation techniques such as liquid chromatography (LC), quantitative analysis and identification of metabolic entities can be accurately performed. LC-MS is a common and robust method which can be applied to a large range of biological samples, and the high sensitivity of LC-MS analyses has made it a popular approach of choice among researchers.

LC separations are performed by passing a liquid mobile phase (the mixture that contains the sample) through a solid stationary phase. The rate at which the molecules travel through the column is dependent on their size and charge, therefore providing different elution conditions, allowing them to be separated over time. This separation step permits individual introduction of metabolites into the MS system. Following chromatographic separation, ionisation techniques are applied to the samples at the LC-MS interface to facilitate detection. The mass analyser can then separate ions according to their m/z which are subsequently detected and quantified. The working principle and instrument schematic of LC-MS, with the example of an Orbitrap mass analyser, is shown in **Figure 9**.

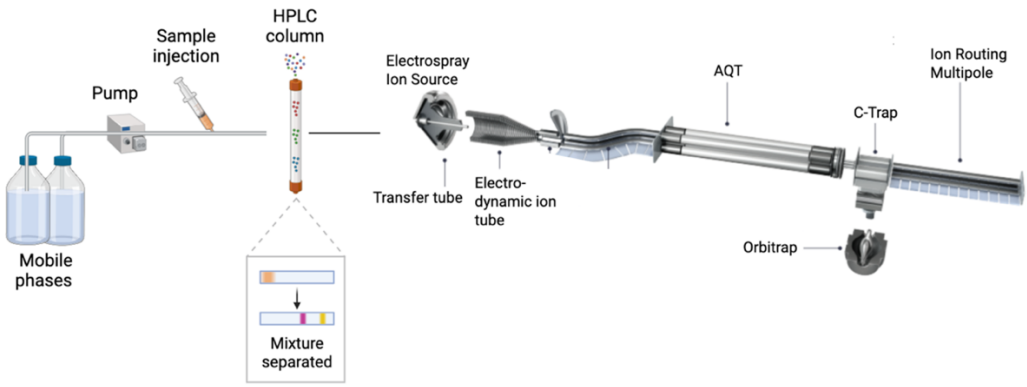


Figure 9. Schematic representation of the internal architecture of an Orbitrap LC-MS system. The diagram illustrates the integration of liquid chromatography with a mass spectrometer. In the LC module, analytes are separated based on their physicochemical properties as they pass through a chromatographic column. The eluent enters the ESI source, where analytes are converted into gas-phase ions. These ions are guided through the various ion optics and mass filtering components. AQT, quadrupole mass filter with Advanced Quadrupole Technology; HPLC, high performance liquid chromatography.

1.4.5 Data analysis

Metabolomics analyses produce a large volume of data from a range of variables and therefore appropriate data analysis is crucial for accurate interpretation of results. While the software used to analyse raw data is chosen independently by the specific laboratory, each processing solution similarly handles data preparation and presentation. Prior to statistical interpretation, raw data generated by mass spectrometry must be transformed into a structured format, typically a feature matrix. Raw files are first converted to an open format (e.g., mzML), which then undergo processes of peak detection, deconvolution, and retention time alignment [241]. This can be done using a selection of tools and software, including Compound Discoverer (ThermoScientific), R packages such as XCMS, MZmine, and MS-DIAL. Quality control (QC) samples are used throughout preprocessing approaches to monitor instrument performance [242]. Samples are often run across multiples batches or over extended periods of time in large-scale metabolomics experiments, thereby introducing systemic variation. QC samples are therefore also used for batch alignment and correction to mitigate these effects. Software such as Quality Control-based Metabolite eXpression Preprocessing (QC:MXP) [243] utilises QC samples for batch correction and normalisation across batches, assuming that any systemic variation observed in the QCs can be used to correct the features in the samples. This correction enhances data comparability and reproducibility for large and complex datasets. Ultimately, the preprocessing stage is a crucial determinant of data quality and biological interpretation.

Multivariate statistical analysis methods are used to communicate the large volume of MS data into an interpretable model. One powerful method of data dimensionality reduction common in metabolomics investigations is principal component analysis (PCA) [244]. PCA is an unsupervised learning method, through which the algorithms are trained on unlabelled datasets. A set of observations can be transformed into a set of variables termed principal components that have potential linear correlation. The first principal component (PC1) is the linear set of variables that explains the maximum variance, PC2 explains the second amount of variation, and so on. In a PCA plot, the x-axis is considered the most important dimension and the strongest contributor of variance, and as all axes are mutually orthogonal, there is no association of variance between the axes. PCA therefore provides a succession of principal components that correspond to the maximum axes of variation. In

this way, relevant information is extracted from large and complex datasets and separations between groups identified. In the context of disease-applied metabolomics, this is a valuable tool for assessing whether experimental classification, for example IBD patients versus healthy controls, show significant variance in metabolic profile.

Partial least squares (PLS) is a supervised multivariate statistical analysis method that is also used to interpret metabolomics data. PLS simultaneously models the relationship between multiple independent and dependent variables through latent components [245]. The primary strength of PLS lies in its ability to handle data with many collinear variables, which is common in metabolomics data. Two common variants of PLS are partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA), which build upon the basic PLS framework, but each is tailored for specific objectives. PLS-DA is primarily used for classification, where the goal is to classify samples into distinct groups, for example healthy vs. diseased or treated vs. untreated [246]. It is designed to maximise the separation between the groups based on the metabolites in the dataset. OPLS-DA is an enhanced version of PLS-DA that aims to improve the interpretability of the model by separating the variation into two components, the predictive component and the orthogonal component [247]. Each of these statistical techniques are crucial in metabolomics data processing as they enable the analysis and interpretation of complex and high-dimensional metabolomics data.

1.4.6 Biological interpretation

The biological interpretation of data from metabolomics experiments can be a challenging task, particularly in the case of untargeted methods. Metabolites do not act in isolation but are part of interconnected metabolic networks and therefore a change in one metabolite can affect the levels of many others [248]. Moreover, the functional relevance of metabolites in specific biological processes or disease pathology may not always be clear due to the high complexity of metabolic interactions and networks. To address these challenges, researchers utilise various strategies to aid biological interpretation through biochemical pathway analysis and network formations, which map observed metabolite changes to metabolic pathways. Metabolic networks can be built from a variety of methods and platforms utilising computational strategies. These tools include pathway analysis

environments including databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [249] and Human Metabolome Database (HMDB) [250], in addition to tools such as Cytoscape [251], which allow researchers to visualize and explore metabolic networks. Specialised metabolomics-focused platforms such as MetaboAnalyst [252], Mummichog [253], XCMS Online [254], and various programming packages, additionally offer capabilities for functional interpretation and biomarker identification.

Despite these advances, there are several limitations associated with biological interpretation of metabolomics data. Pathway analysis, in particular, presents challenges because it is largely adopted from genomics. In metabolomics analysis, many metabolites lack consistent or universally accepted identifiers in pathway databases, meaning that a substantial portion of detected metabolites may be excluded from pathway-based analysis. This can bias results towards well-characterised metabolic pathways and omit novel or poorly annotated metabolites that may be biologically significant. When pathway analysis is performed in conjunction with multi-omics integration, it provides a more holistic view of biological systems [255], yet the limitations of database coverage, annotation quality, and cross-platform integration continue to pose challenges for achieving a fully comprehensive understanding of disease mechanisms.

1.4.7 Application of Metabolomics to Gut Disease Research

Through metabolomic analysis, changes in metabolite levels can be detected between groups, which may reflect mechanisms of disease pathogenesis. Studying the human metabolome is therefore valuable for investigating mechanisms underlying disease development and progression. This is done through discrimination of metabolic profiles between disease and non-disease states as well as disease activity-based subtyping. Deeper analyses can then reveal the metabolites and associated pathways implicated in various stages of disease, which can lead to identification of novel prognostic and diagnostic biomarkers. Pathophysiology process elucidation can additionally be applied to therapeutic investigations, through which the mechanism of action for pharmaceutical drugs and disease interventions can be assessed. Furthermore, the determination of response to treatment is another fundamental application of metabolomics, an increasingly popular

subdivision of precision medicine termed pharmacometabolomics [256]. Identification of patients expected to respond to treatment and those who will likely show no response, or at risk of adverse drug reactions, brings benefits to patients and clinicians throughout the therapeutic pipeline.

The potential of metabolomics applications in IBD research has been shown by recent studies providing valuable insights into metabolic alterations associated with disease and treatment. Significant alterations in the faecal, urine, blood, and tissue metabolome of IBD patients have been described [257-260], reflecting shifts in metabolic processes underpinning disease progression. One consistently observed alterations in metabolomics research of IBD is the significant change in levels of amino acids [261]. For example, a reduction in levels of tryptophan have been reported in IBD patients in comparison to healthy controls due to increased catabolism via the kynurenine pathway [262, 263]. Metabolites produced by gut bacteria are an area of interest in IBD research due to the significant involvement of the gut microbiome in disease pathophysiology. For example, current research shows that levels of microbially-produced metabolites including SCFAs [264, 265] and bile acids [266, 267] are altered in IBD patients. Metabolomics research has additionally started to uncover a wide array of further metabolic alterations in IBD, spanning organic acids [268], lipids [269] and carbohydrates [270], reflecting the complex and multifactorial nature of disease. While significant progress has been made in identifying metabolic changes associated with IBD, it has been noted that inconsistent findings across studies currently limits a full understanding of the metabolic basis of disease and their role as potential biomarkers [271]. Continued metabolomics investigation in IBD will be crucial for determining specific metabolic roles and translating metabolic insights into reliable diagnostic tools and targeted therapies.

One of the main goals of metabolomics research is the identification of biomarkers that can diagnose, classify, or monitor disease. Although single biomarkers alone are unlikely to give an accurate diagnosis or classification of disease, there are several serological, faecal, and histologic markers which are currently used as an adjunct to image diagnostic measures. This can help differentiate between; presence and absence of disease, subclassification of disease, and between active or quiescent disease. Given diagnostic issues of patient heterogeneity and misalignments in clinical symptoms and disease activity, a combination

of disease assessment methods is recommended for physicians to provide an optimal evaluation.

Biomarkers have been successfully applied to the clinic for a variety of diseases including IBD and CoD. For example, C-reactive protein (CRP) is an acute phase molecule produced in the liver in response to inflammatory cytokines. CRP is the most widely used serum biomarker to assess inflammation in IBD, with endoscopic and histological evaluation correlating well with CRP measurements [272]. Measurements of CRP provide a guide to assess disease severity and extent of inflammation; however, CRP is not specific to IBD. Wherein elevated levels indicate an increased inflammatory state, this does not explicitly signify the presence of one disease, only a general state of inflammation. This underlines the requirement for the combined use of biomarkers with additional methods of disease assessment. Other serum biomarkers used in IBD evaluation include serum antibodies, cytokines, and serum amyloid A [273]. Faecal biomarkers are also used in IBD practice as measurements of inflammatory activity. Faecal calprotectin (FC), a calcium-binding protein released by neutrophils upon inflammatory stimulation, is extensively used as a biomarker for IBD [274]. FC release reflects the number of neutrophils involved in the inflammatory process and is therefore proportionate to the extent of inflammation. Similarly, however, FC levels are not disease specific, which carries limitations in confirming the precise cause of the inflammation [275].

The wealth of information given by metabolic profiling which links immune processes, gut microbial activity, and diet, places metabolomics as an ideal approach to both IBD and CoD research. Determining metabolic signatures of gut inflammation in patients with disease is therefore imperative in driving the research field forward.

1.5 Aims

Research knowledge around mechanisms of inflammatory diseases of the gut including IBD and CoD has expanded rapidly in recent years, however uncertainty still surrounds the specific biochemical events involved in disease processes, particularly at the metabolic level. Advances in mass spectrometry-based metabolomics has provided an invaluable approach for profiling metabolites and understanding metabolic changes in different

disease states and treatment responses. Application of LC-MS-based metabolomics approaches to the study of gastrointestinal disease through both untargeted and targeted methodologies can therefore provide crucial disease-specific insights, potential novel biomarkers, and therapeutic targets.

The overall aim of this thesis is to develop and employ advanced metabolomics methodologies to investigate molecular mechanisms of gastrointestinal disease, with a particular focus on how metabolic alterations contribute to IBD and CoD. Through the optimisation of LC-MS methods across stages of the analysis pipeline and specific to the biomatrix used, this research aims to provide comprehensive insights into the metabolic nature of these diseases using optimal and robust methodologies. Furthermore, the application of the developed methods to patient cohorts including large-scale clinical trials, allows for the investigation into real-life clinical populations and aims to facilitate our understanding of metabolic pathways that may drive disease mechanisms and inform treatment strategies for IBD and CoD. The specific aims of the chapters are further detailed below.

The second chapter investigates how metabolite extraction from faecal samples can be optimised for both untargeted and targeted metabolomics analyses. It also examines whether the optimised extraction protocol is suitable for analysing samples from patients with gastrointestinal disease, providing a methodological foundation for subsequent studies.

Building on this, Chapter 3 explores how the faecal metabolome of children with CoD can reveal mechanistic insights about disease and changes upon treatment with a gluten-free diet (GFD). It addresses the research question of what patterns emerge from a combined cross-sectional and prospective cohort analysis using untargeted and targeted LC-MS, providing the first comprehensive profiling of the faecal metabolome in this patient population.

Chapter 4 focuses on urinary metabolomics, investigating how an untargeted LC-MS method can be optimised for large-scale clinical studies. It evaluates the efficiency and

applicability of the optimised method in analysing urine samples from both IBD patients and healthy controls, representing the largest urinary metabolomics study of IBD to date.

The role of specific food additives in the context of IBD were examined in Chapter 5, investigating the mechanisms by which they promote or reduce gastrointestinal inflammation. This chapter also considers how current evidence can inform the design of dietary interventions to improve disease management.

1.6 References

1. Wiertsema, S.P., et al., *The Interplay between the Gut Microbiome and the Immune System in the Context of Infectious Diseases throughout Life and the Role of Nutrition in Optimizing Treatment Strategies.* Nutrients, 2021. **13**(3).
2. Murphy, K.G. and S.R. Bloom, *Gut hormones and the regulation of energy homeostasis.* Nature, 2006. **444**(7121): p. 854-859.
3. Osadchiy, V., C.R. Martin, and E.A. Mayer, *The Gut-Brain Axis and the Microbiome: Mechanisms and Clinical Implications.* Clin Gastroenterol Hepatol, 2019. **17**(2): p. 322-332.
4. Pabst, O., et al., *Gut–liver axis: barriers and functional circuits.* Nature Reviews Gastroenterology & Hepatology, 2023. **20**(7): p. 447-461.
5. Ferranti, E.P., et al., *20 things you didn't know about the human gut microbiome.* J Cardiovasc Nurs, 2014. **29**(6): p. 479-81.
6. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-30.
7. King, C.H., et al., *Baseline human gut microbiota profile in healthy people and standard reporting template.* PLOS ONE, 2019. **14**(9): p. e0206484.
8. Jalanka-Tuovinen, J., et al., *Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms.* PLoS One, 2011. **6**(7): p. e23035.
9. Xu, C., H. Zhu, and P. Qiu, *Aging progression of human gut microbiota.* BMC Microbiology, 2019. **19**(1): p. 236.
10. Lawley, T.D. and A.W. Walker, *Intestinal colonization resistance.* Immunology, 2013. **138**(1): p. 1-11.
11. Karczewski, J., et al., *Regulation of human epithelial tight junction proteins by Lactobacillus plantarum in vivo and protective effects on the epithelial barrier.* Am J Physiol Gastrointest Liver Physiol, 2010. **298**(6): p. G851-9.
12. Wiles, T.J., et al., *Host Gut Motility Promotes Competitive Exclusion within a Model Intestinal Microbiota.* PLOS Biology, 2016. **14**(7): p. e1002517.
13. Chen, C., X. Yang, and X. Shen, *Confirmed and Potential Roles of Bacterial T6SSs in the Intestinal Ecosystem.* Front Microbiol, 2019. **10**: p. 1484.

14. Coppell, K.J., et al., *Annual Incidence and Phenotypic Presentation of IBD in Southern New Zealand: An 18-Year Epidemiological Analysis.* Inflamm Intest Dis, 2018. **3**(1): p. 32-39.
15. Wang, R., et al., *Global, regional and national burden of inflammatory bowel disease in 204 countries and territories from 1990 to 2019: a systematic analysis based on the Global Burden of Disease Study 2019.* BMJ Open, 2023. **13**(3): p. e065186.
16. Santiago, M., et al., *What forecasting the prevalence of inflammatory bowel disease may tell us about its evolution on a national scale.* Therap Adv Gastroenterol, 2019. **12**: p. 1756284819860044.
17. Maher, M.M., *Inflammatory bowel disease: review and future view.* Front Biosci (Elite Ed), 2012. **4**(5): p. 1638-47.
18. Geboes, K., et al., *Indeterminate colitis: a review of the concept--what's in a name?* Inflamm Bowel Dis, 2008. **14**(6): p. 850-7.
19. Strobel, D., R.S. Goertz, and T. Bernatik, *Diagnostics in inflammatory bowel disease: ultrasound.* World J Gastroenterol, 2011. **17**(27): p. 3192-7.
20. Thia, K.T., et al., *Measurement of disease activity in ulcerative colitis: interobserver agreement and predictors of severity.* Inflamm Bowel Dis, 2011. **17**(6): p. 1257-64.
21. Singh, N., et al., *Multi-Center Experience of Vedolizumab Effectiveness in Pediatric Inflammatory Bowel Disease.* Inflamm Bowel Dis, 2016. **22**(9): p. 2121-6.
22. Baert, F., et al., *Mucosal healing predicts sustained clinical remission in patients with early-stage Crohn's disease.* Gastroenterology, 2010. **138**(2): p. 463-8; quiz e10-1.
23. Best, W.R., et al., *Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study.* Gastroenterology, 1976. **70**(3): p. 439-44.
24. Leake, I., *A score to settle—measuring Crohn's disease activity.* Nature Reviews Gastroenterology & Hepatology, 2013. **10**(10): p. 564-564.
25. Administration, F.a.D., *Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry*. 2016.
26. Harvey, R.F. and J.M. Bradshaw, *A SIMPLE INDEX OF CROHN'S-DISEASE ACTIVITY.* The Lancet, 1980. **315**(8167): p. 514.
27. Sipponen, T., et al., *Endoscopic evaluation of Crohn's disease activity: comparison of the CDEIS and the SES-CD.* Inflamm Bowel Dis, 2010. **16**(12): p. 2131-6.

28. Turner, D., et al., *Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study.* Gastroenterology, 2007. **133**(2): p. 423-32.
29. Truelove, S.C. and L.J. Witts, *Cortisone in ulcerative colitis; final report on a therapeutic trial.* Br Med J, 1955. **2**(4947): p. 1041-8.
30. Gordon, F.H., *Lymphocyte adhesion and activation in inflammatory bowel diseases. Studies relating to the effects of blockade of alpha-4 integrins by a monoclonal antibody. Doctoral thesis (Ph.D), UCL (University College London).* . 2002.
31. M'Koma, A.E., *The Multifactorial Etiopathogeneses Interplay of Inflammatory Bowel Disease: An Overview.* Gastrointestinal Disorders, 2019. **1**(1): p. 75-105.
32. Gupta, A., et al., *Familial aggregation of inflammatory bowel disease in patients with ulcerative colitis.* Intest Res, 2017. **15**(3): p. 388-394.
33. Gordon, H., et al., *UK IBD Twin Registry: Concordance and Environmental Risk Factors of Twins with IBD.* Dig Dis Sci, 2022. **67**(6): p. 2444-2450.
34. Liu, J.Z., et al., *Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations.* Nat Genet, 2015. **47**(9): p. 979-986.
35. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 603-6.
36. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 599-603.
37. Nelson, A., et al., *The Impact of NOD2 Genetic Variants on the Gut Mycobiota in Crohn's Disease Patients in Remission and in Individuals Without Gastrointestinal Inflammation.* J Crohns Colitis, 2021. **15**(5): p. 800-812.
38. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.
39. Yamamoto, S. and X. Ma, *Role of Nod2 in the development of Crohn's disease.* Microbes Infect, 2009. **11**(12): p. 912-8.
40. Salem, M., et al., *ATG16L1: A multifunctional susceptibility factor in Crohn disease.* Autophagy, 2015. **11**(4): p. 585-94.
41. Mehto, S., et al., *The Crohn's Disease Risk Factor IRGM Limits NLRP3 Inflammasome Activation by Impeding Its Assembly and by Mediating Its Selective Autophagy.* Mol Cell, 2019. **73**(3): p. 429-445.e7.

42. Peng, L.L., et al., *IL-23R mutation is associated with ulcerative colitis: A systemic review and meta-analysis.* Oncotarget, 2017. **8**(3): p. 4849-4863.
43. Richard, A.C., et al., *Reduced monocyte and macrophage TNFSF15/TL1A expression is associated with susceptibility to inflammatory bowel disease.* PLoS Genet, 2018. **14**(9): p. e1007458.
44. Muise, A.M., et al., *Polymorphisms in E-cadherin (CDH1) result in a mis-localised cytoplasmic protein that is associated with Crohn's disease.* Gut, 2009. **58**(8): p. 1121-7.
45. Halme, L., et al., *Family and twin studies in inflammatory bowel disease.* World J Gastroenterol, 2006. **12**(23): p. 3668-72.
46. Santos, M.P.C., C. Gomes, and J. Torres, *Familial and ethnic risk in inflammatory bowel disease.* Ann Gastroenterol, 2018. **31**(1): p. 14-23.
47. Halfvarson, J., et al., *Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics.* Gastroenterology, 2003. **124**(7): p. 1767-73.
48. Freeman, K., et al., *The incidence and prevalence of inflammatory bowel disease in UK primary care: a retrospective cohort study of the IQVIA Medical Research Database.* BMC Gastroenterol, 2021. **21**(1): p. 139.
49. Burisch, J., et al., *The burden of inflammatory bowel disease in Europe.* J Crohns Colitis, 2013. **7**(4): p. 322-37.
50. Smids, C., et al., *Intestinal T Cell Profiling in Inflammatory Bowel Disease: Linking T Cell Subsets to Disease Activity and Disease Course.* Journal of Crohn's and Colitis, 2018. **12**(4): p. 465-475.
51. Paridhi, et al., *Role of Immunohistochemistry and serology in subclassifying the Inflammatory Bowel Disease cases diagnosed as Inflammatory Bowel Diseases---unclassified on colonic biopsies.* Indian J Pathol Microbiol, 2022. **65**(3): p. 558-564.
52. Sezgin, E., et al., *Quantitative real-time PCR analysis of bacterial biomarkers enable fast and accurate monitoring in inflammatory bowel disease.* PeerJ, 2022. **10**: p. e14217.
53. Xu, L., et al., *Bulk and single-cell RNA sequencing reveal the roles of neutrophils in pediatric Crohn's disease.* Pediatric Research, 2025.
54. Li, X., L. Conklin, and P. Alex, *New serological biomarkers of inflammatory bowel disease.* World J Gastroenterol, 2008. **14**(33): p. 5115-24.

55. Silva, M.A., et al., *Cytokine Tissue Levels as Markers of Disease Activity in Pediatric Crohn Disease.* Pediatric Research, 2003. **54**(4): p. 456-461.
56. Liu, C., J. Liu, and Y. Yang, *Bulk and Single-Cell Transcriptomic Reveals Shared Key Genes and Patterns of Immune Dysregulation in Both Intestinal Inflammatory Disease and Sepsis.* J Cell Mol Med, 2025. **29**(4): p. e70415.
57. Roda, G., et al., *Intestinal epithelial cells in inflammatory bowel diseases.* World J Gastroenterol, 2010. **16**(34): p. 4264-71.
58. Di Sabatino, A., et al., *Increased enterocyte apoptosis in inflamed areas of Crohn's disease.* Dis Colon Rectum, 2003. **46**(11): p. 1498-507.
59. Wehkamp, J., et al., *Reduced Paneth cell alpha-defensins in ileal Crohn's disease.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 18129-34.
60. Gersemann, M., et al., *Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis.* Differentiation, 2009. **77**(1): p. 84-94.
61. Carraway, K.L., et al., *Cell signaling through membrane mucins.* Bioessays, 2003. **25**(1): p. 66-71.
62. Wibowo, A.A., et al., *Decreased expression of MUC2 due to a decrease in the expression of lectins and apoptotic defects in colitis patients.* Biochem Biophys Rep, 2019. **19**: p. 100655.
63. Theodoratou, E., et al., *The role of glycosylation in IBD.* Nature Reviews Gastroenterology & Hepatology, 2014. **11**(10): p. 588-600.
64. Aamann, L., E.M. Vestergaard, and H. Grønbæk, *Trefoil factors in inflammatory bowel disease.* World J Gastroenterol, 2014. **20**(12): p. 3223-30.
65. van der Post, S., et al., *Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis.* Gut, 2019. **68**(12): p. 2142-2151.
66. Fyderek, K., et al., *Mucosal bacterial microflora and mucus layer thickness in adolescents with inflammatory bowel disease.* World J Gastroenterol, 2009. **15**(42): p. 5287-94.
67. Förster, C., *Tight junctions and the modulation of barrier function in disease.* Histochem Cell Biol, 2008. **130**(1): p. 55-70.
68. Ulluwishewa, D., et al., *Regulation of tight junction permeability by intestinal bacteria and dietary components.* J Nutr, 2011. **141**(5): p. 769-76.

69. Furuse, M., et al., *Direct association of occludin with ZO-1 and its possible involvement in the localization of occludin at tight junctions.* J Cell Biol, 1994. **127**(6 Pt 1): p. 1617-26.
70. Gomez-Bris, R., et al., *CD4 T-Cell Subsets and the Pathophysiology of Inflammatory Bowel Disease.* Int J Mol Sci, 2023. **24**(3).
71. Del Prete, G.F., et al., *Purified protein derivative of Mycobacterium tuberculosis and excretory-secretory antigen(s) of Toxocara canis expand in vitro human T cells with stable and opposite (type 1 T helper or type 2 T helper) profile of cytokine production.* J Clin Invest, 1991. **88**(1): p. 346-50.
72. Zielinski, C.E., *T helper cell subsets: diversification of the field.* European Journal of Immunology, 2023. **53**(12): p. 2250218.
73. Imam, T., et al., *Effector T Helper Cell Subsets in Inflammatory Bowel Diseases.* Front Immunol, 2018. **9**: p. 1212.
74. Sonnenberg, A. and J.F. Collins, *Vicious circles in inflammatory bowel disease.* Inflamm Bowel Dis, 2006. **12**(10): p. 944-9.
75. Johnson, J.S., et al., *Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis.* Nature Communications, 2019. **10**(1): p. 5029.
76. Ashton, J.J., et al., *16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease.* Medicine (Baltimore), 2017. **96**(26): p. e7347.
77. Orejudo, M., et al., *P010 Faecal microbiota composition by shotgun metagenomic sequencing approach in a newly diagnosed cohort of inflammatory bowel disease patients: results from the IBDomics project.* Journal of Crohn's and Colitis, 2024. **18**(Supplement_1): p. i258-i259.
78. Sharpton, T., et al., *Development of Inflammatory Bowel Disease Is Linked to a Longitudinal Restructuring of the Gut Metagenome in Mice.* mSystems, 2017. **2**(5).
79. Ojala, T., E. Kankuri, and M. Kankainen, *Understanding human health through metatranscriptomics.* Trends in Molecular Medicine, 2023. **29**(5): p. 376-389.
80. Amos, G.C.A., et al., *Exploring how microbiome signatures change across inflammatory bowel disease conditions and disease locations.* Scientific Reports, 2021. **11**(1): p. 18699.
81. Nomura, K., et al., *Bacteroidetes Species Are Correlated with Disease Activity in Ulcerative Colitis.* J Clin Med, 2021. **10**(8).

82. Hall, A.B., et al., *A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients.* Genome Medicine, 2017. **9**(1): p. 103.
83. Joossens, M., et al., *Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives.* Gut, 2011. **60**(5): p. 631-7.
84. Sokol, H., et al., *Low counts of Faecalibacterium prausnitzii in colitis microbiota.* Inflamm Bowel Dis, 2009. **15**(8): p. 1183-9.
85. Vatn, S., et al., *Faecal microbiota signatures of IBD and their relation to diagnosis, disease phenotype, inflammation, treatment escalation and anti-TNF response in a European Multicentre Study (IBD-Character).* Scand J Gastroenterol, 2020. **55**(10): p. 1146-1156.
86. Scaldaferri, F., et al., *Gut microbial flora, prebiotics, and probiotics in IBD: their current usage and utility.* Biomed Res Int, 2013. **2013**: p. 435268.
87. Martin, H.M., et al., *Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer.* Gastroenterology, 2004. **127**(1): p. 80-93.
88. Ticer, T., et al., *KLEBSIELLA PNEUMONIAE IN THE COLONIC MUCUS LAYER INFLUENCES CLOSTRIDIODES DIFFICILE PATHOGENESIS.* Gastroenterology, 2022. **162**: p. S69.
89. Strauss, J., et al., *Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host.* Inflamm Bowel Dis, 2011. **17**(9): p. 1971-8.
90. Wang, L., et al., *A purified membrane protein from Akkermansia muciniphila or the pasteurised bacterium blunts colitis associated tumourigenesis by modulation of CD8(+) T cells in mice.* Gut, 2020. **69**(11): p. 1988-1997.
91. Norman, J.M., et al., *Disease-specific alterations in the enteric virome in inflammatory bowel disease.* Cell, 2015. **160**(3): p. 447-60.
92. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.
93. Alemany-Cosme, E., et al., *Oxidative Stress in the Pathogenesis of Crohn's Disease and the Interconnection with Immunological Response, Microbiota, External Environmental Factors, and Epigenetics.* Antioxidants (Basel), 2021. **10**(1).
94. Barnett, M.P., et al., *Changes in colon gene expression associated with increased colon inflammation in interleukin-10 gene-deficient mice inoculated with Enterococcus species.* BMC Immunol, 2010. **11**: p. 39.

95. Adolph, T.E., et al., *The metabolic nature of inflammatory bowel diseases.* Nature Reviews Gastroenterology & Hepatology, 2022. **19**(12): p. 753-767.
96. Zhang, S., et al., *Metabolic regulation of the Th17/Treg balance in inflammatory bowel disease.* Pharmacological Research, 2024. **203**: p. 107184.
97. Zaiatz Bittencourt, V., et al., *Targeting Immune Cell Metabolism in the Treatment of Inflammatory Bowel Disease.* Inflamm Bowel Dis, 2021. **27**(10): p. 1684-1693.
98. Batista-Gonzalez, A., et al., *New Insights on the Role of Lipid Metabolism in the Metabolic Reprogramming of Macrophages.* Front Immunol, 2019. **10**: p. 2993.
99. Xia, Y., et al., *Role of glycolysis in inflammatory bowel disease and its associated colorectal cancer.* Front Endocrinol (Lausanne), 2023. **14**: p. 1242991.
100. Zhou, Z., et al., *Increased stromal PFKFB3-mediated glycolysis in inflammatory bowel disease contributes to intestinal inflammation.* Front Immunol, 2022. **13**: p. 966067.
101. Macintyre, A.N., et al., *The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function.* Cell Metab, 2014. **20**(1): p. 61-72.
102. Lee, H., J.H. Jeon, and E.S. Kim, *Mitochondrial dysfunctions in T cells: focus on inflammatory bowel disease.* Front Immunol, 2023. **14**: p. 1219422.
103. Li, X., et al., *Lactate metabolism in human health and disease.* Signal Transduction and Targeted Therapy, 2022. **7**(1): p. 305.
104. Wang, B., et al., *Pharmacological modulation of mitochondrial function as novel strategies for treating intestinal inflammatory diseases and colorectal cancer.* Journal of Pharmaceutical Analysis, 2025. **15**(4): p. 101074.
105. Rath, E. and D. Haller, *Intestinal epithelial cell metabolism at the interface of microbial dysbiosis and tissue injury.* Mucosal Immunol, 2022. **15**(4): p. 595-604.
106. Singh, P., et al., *Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis.* Clin Gastroenterol Hepatol, 2018. **16**(6): p. 823-836.e2.
107. Barone, M.V., R. Troncone, and S. Auricchio, *Gliadin peptides as triggers of the proliferative and stress/innate immune response of the celiac small intestinal mucosa.* Int J Mol Sci, 2014. **15**(11): p. 20518-37.
108. Lebwohl, B., D.S. Sanders, and P.H.R. Green, *Coeliac disease.* The Lancet, 2018. **391**(10115): p. 70-81.
109. Green, P.H.R., *The many faces of celiac disease: Clinical presentation of celiac disease in the adult population.* Gastroenterology, 2005. **128**(4): p. S74-S78.

110. Radlovic, N., et al., *Clinical features of non-classical celiac disease in children and adolescents.* Srpski arhiv za celokupno lekarstvo, 2020. **149**: p. 70-70.
111. Çaltepe, G., *The hidden danger: Silent celiac disease.* Turk J Gastroenterol, 2018. **29**(5): p. 530-531.
112. Rubio-Tapia, A. and J.A. Murray, *Classification and management of refractory coeliac disease.* Gut, 2010. **59**(4): p. 547-57.
113. Newton, M., et al., *What are the clinical consequences of 'potential' coeliac disease?* Digestive and Liver Disease, 2023. **55**(4): p. 478-484.
114. Ludvigsson, J.F., et al., *The Oslo definitions for coeliac disease and related terms.* Gut, 2013. **62**(1): p. 43-52.
115. Barbaro, M.R., et al., *Recent advances in understanding non-celiac gluten sensitivity.* F1000Res, 2018. **7**.
116. Sharma, N., et al., *Pathogenesis of Celiac Disease and Other Gluten Related Disorders in Wheat and Strategies for Mitigating Them.* Front Nutr, 2020. **7**: p. 6.
117. Al-Toma, A., et al., *European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders.* United European Gastroenterol J, 2019. **7**(5): p. 583-613.
118. Bai, J.C. and C. Ciacci, *World Gastroenterology Organisation Global Guidelines: Celiac Disease February 2017.* J Clin Gastroenterol, 2017. **51**(9): p. 755-768.
119. Downey, L., et al., *Recognition, assessment, and management of coeliac disease: summary of updated NICE guidance.* Bmj, 2015. **351**: p. h4513.
120. Khan, A.S., et al., *The Role of Physicians' Factors in Underdiagnosis of Celiac Disease in the Eastern Province of Saudi Arabia.* Cureus, 2023. **15**(9): p. e44690.
121. Whitburn, J., et al., *Diagnosis of celiac disease is being missed in over 80% of children particularly in those from socioeconomically deprived backgrounds.* Eur J Pediatr, 2021. **180**(6): p. 1941-1946.
122. Kårhus, L.L., et al., *Symptoms and biomarkers associated with undiagnosed celiac seropositivity.* BMC Gastroenterology, 2021. **21**(1): p. 90.
123. Husby, S., et al., *European Society Paediatric Gastroenterology, Hepatology and Nutrition Guidelines for Diagnosing Coeliac Disease 2020.* Journal of Pediatric Gastroenterology and Nutrition, 2020. **70**(1): p. 141-156.
124. ESPGHAN. *New Guidelines for the Diagnosis*

of Paediatric Coeliac Disease. 2022; Available from: 2020_New_Guidelines_for_the_Diagnosis_of_Paediatric_Coeliac_Disease._ESPGHAN_Advice_Guide.pdf.

125. Gray, A.M. and I.N. Papanicolas, *Impact of symptoms on quality of life before and after diagnosis of coeliac disease: results from a UK population survey.* BMC Health Serv Res, 2010. **10**: p. 105.
126. Majsiak, E., et al., *The impact of symptoms on quality of life before and after diagnosis of coeliac disease: the results from a Polish population survey and comparison with the results from the United Kingdom.* BMC Gastroenterol, 2021. **21**(1): p. 99.
127. Kemppainen, K.M., et al., *Factors That Increase Risk of Celiac Disease Autoimmunity After a Gastrointestinal Infection in Early Life.* Clin Gastroenterol Hepatol, 2017. **15**(5): p. 694-702.e5.
128. Oliveira, D.R., et al., *HLA DQ2/DQ8 haplotypes and anti-transglutaminase antibodies as celiac disease markers in a pediatric population with type 1 diabetes mellitus.* Arch Endocrinol Metab, 2022. **66**(2): p. 229-236.
129. Aboulaghras, S., et al., *Meta-Analysis and Systematic Review of HLA DQ2/DQ8 in Adults with Celiac Disease.* Int J Mol Sci, 2023. **24**(2).
130. Sharma, A., et al., *Identification of Non-HLA Genes Associated with Celiac Disease and Country-Specific Differences in a Large, International Pediatric Cohort.* PLoS One, 2016. **11**(3): p. e0152476.
131. Hunt, K.A., et al., *A common CTLA4 haplotype associated with coeliac disease.* European Journal of Human Genetics, 2005. **13**(4): p. 440-444.
132. Trynka, G., et al., *Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling.* Gut, 2009. **58**(8): p. 1078-83.
133. Sciurti, M., et al., *Genetic susceptibilty and celiac disease: what role do HLA haplotypes play?* Acta Biomed, 2018. **89**(9-s): p. 17-21.
134. Logan, K., et al., *Early Gluten Introduction and Celiac Disease in the EAT Study: A Prespecified Analysis of the EAT Randomized Clinical Trial.* JAMA Pediatr, 2020. **174**(11): p. 1041-1047.
135. Crespo-Escobar, P., et al., *The role of gluten consumption at an early age in celiac disease development: a further analysis of the prospective PreventCD cohort study12.* The American Journal of Clinical Nutrition, 2017. **105**(4): p. 890-896.

136. Oikarinen, M., et al., *Enterovirus Infections Are Associated With the Development of Celiac Disease in a Birth Cohort Study.* Front Immunol, 2020. **11**: p. 604529.
137. Stene, L.C., et al., *Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study.* Am J Gastroenterol, 2006. **101**(10): p. 2333-40.
138. Dydensborg Sander, S., et al., *Association Between Antibiotics in the First Year of Life and Celiac Disease.* Gastroenterology, 2019. **156**(8): p. 2217-2229.
139. Silano, M., et al., *Infant feeding and risk of developing celiac disease: a systematic review.* BMJ Open, 2016. **6**(1): p. e009163.
140. Mårild, K., et al., *Pregnancy outcome and risk of celiac disease in offspring: a nationwide case-control study.* Gastroenterology, 2012. **142**(1): p. 39-45.e3.
141. Kagnoff, M.F., *Overview and pathogenesis of celiac disease.* Gastroenterology, 2005. **128**(4): p. S10-S18.
142. Gutiérrez, S., et al., *The human digestive tract has proteases capable of gluten hydrolysis.* Mol Metab, 2017. **6**(7): p. 693-702.
143. Fernández-Pérez, S., et al., *The Human Digestive Tract Is Capable of Degrading Gluten from Birth.* Int J Mol Sci, 2020. **21**(20).
144. Frazer, A.C., et al., *GLUTEN-INDUCED ENTEROPATHY THE EFFECT OF PARTIALLY DIGESTED GLUTEN.* The Lancet, 1959. **274**(7097): p. 252-255.
145. Cornell, H.J., *Partial In Vitro Digestion of Active Gliadin-Related Peptides in Celiac Disease.* Journal of Protein Chemistry, 1998. **17**(8): p. 739-744.
146. Kumar, J., et al., *Physiopathology and Management of Gluten-Induced Celiac Disease.* Journal of Food Science, 2017. **82**(2): p. 270-277.
147. DaFonte, T.M., et al., *Zonulin as a Biomarker for the Development of Celiac Disease.* Pediatrics, 2024. **153**(1).
148. Schumann, M., et al., *Mechanisms of epithelial translocation of the alpha(2)-gliadin-33mer in coeliac sprue.* Gut, 2008. **57**(6): p. 747-54.
149. Rauhavirta, T., et al., *Epithelial transport and deamidation of gliadin peptides: a role for coeliac disease patient immunoglobulin A.* Clin Exp Immunol, 2011. **164**(1): p. 127-36.
150. Voisine, J. and V. Abadie, *Interplay Between Gluten, HLA, Innate and Adaptive Immunity Orchestrates the Development of Coeliac Disease.* Front Immunol, 2021. **12**: p. 674313.

151. Elli, L., et al., *Small bowel villous atrophy: celiac disease and beyond.* Expert Rev Gastroenterol Hepatol, 2017. **11**(2): p. 125-138.
152. Schiepatti, A., et al., *Persistent villous atrophy predicts development of complications and mortality in adult patients with coeliac disease: a multicentre longitudinal cohort study and development of a score to identify high-risk patients.* Gut, 2023. **72**(11): p. 2095-2102.
153. Villanacci, V., et al., *Celiac disease: histology-differential diagnosis-complications. A practical approach.* Pathologica, 2020. **112**(3): p. 186-196.
154. Adelman, D.C., et al., *Measuring Change In Small Intestinal Histology In Patients With Celiac Disease.* Am J Gastroenterol, 2018. **113**(3): p. 339-347.
155. García-Hoz, C., et al., *Intraepithelial Lymphogram in the Diagnosis of Celiac Disease in Adult Patients: A Validation Cohort.* Nutrients, 2024. **16**(8).
156. Sánchez-Castañon, M., et al., *Intraepithelial lymphocytes subsets in different forms of celiac disease.* Auto Immun Highlights, 2016. **7**(1): p. 14.
157. Collin, P., P.J. Wahab, and J.A. Murray, *Intraepithelial lymphocytes and coeliac disease.* Best Practice & Research Clinical Gastroenterology, 2005. **19**(3): p. 341-350.
158. Zafeiropoulou, K., et al., *Alterations in Intestinal Microbiota of Children With Celiac Disease at the Time of Diagnosis and on a Gluten-free Diet.* Gastroenterology, 2020. **159**(6): p. 2039-2051.e20.
159. Sánchez, E., J.M. Laparra, and Y. Sanz, *Discerning the role of Bacteroides fragilis in celiac disease pathogenesis.* Appl Environ Microbiol, 2012. **78**(18): p. 6507-15.
160. Sánchez, E., et al., *Intestinal Bacteroides species associated with coeliac disease.* J Clin Pathol, 2010. **63**(12): p. 1105-11.
161. Akobeng, A.K., et al., *Role of the gut microbiota in the pathogenesis of coeliac disease and potential therapeutic implications.* Eur J Nutr, 2020. **59**(8): p. 3369-3390.
162. Olshan, K.L., et al., *Gut microbiota in Celiac Disease: microbes, metabolites, pathways and therapeutics.* Expert Rev Clin Immunol, 2020. **16**(11): p. 1075-1092.
163. Lupu, V.V., et al., *Advances in Understanding the Human Gut Microbiota and Its Implication in Pediatric Celiac Disease-A Narrative Review.* Nutrients, 2023. **15**(11).
164. Valitutti, F., S. Cucchiara, and A. Fasano, *Celiac Disease and the Microbiome.* Nutrients, 2019. **11**(10).

165. Gholam-Mostafaei, F.S., et al., *Gut microbiota, angiotensin-converting enzyme, celiac disease, and risk of COVID-19 infection: a review.* Gastroenterol Hepatol Bed Bench, 2021. **14**(Suppl1): p. S24-s31.

166. Slager, J., et al., *High-resolution analysis of the treated coeliac disease microbiome reveals increased inter-individual variability.* bioRxiv, 2024: p. 2024.03.08.584098.

167. Galipeau, H.J. and E.F. Verdu, *The double-edged sword of gut bacteria in celiac disease and implications for therapeutic potential.* Mucosal Immunology, 2022. **15**(2): p. 235-243.

168. Matera, M. and S. Guandalini *How the Microbiota May Affect Celiac Disease and What We Can Do*. Nutrients, 2024. **16**, DOI: 10.3390/nu16121882.

169. Di Biase, A.R., et al., *Gut microbiota signatures and clinical manifestations in celiac disease children at onset: a pilot study.* J Gastroenterol Hepatol, 2021. **36**(2): p. 446-454.

170. Nobel, Y.R., et al., *Lack of Effect of Gluten Challenge on Fecal Microbiome in Patients With Celiac Disease and Non-Celiac Gluten Sensitivity.* Clin Transl Gastroenterol, 2021. **12**(12): p. e00441.

171. Upadhyay, D., et al., *NMR based metabolic profiling of patients with potential celiac disease elucidating early biochemical changes of gluten-sensitivity: A pilot study.* Clinica Chimica Acta, 2022. **531**: p. 291-301.

172. Picca, A., et al., *Mitochondria and redox balance in coeliac disease: A case-control study.* Eur J Clin Invest, 2018. **48**(2).

173. McCreery, C.V., et al., *Investigating intestinal epithelium metabolic dysfunction in celiac disease using personalized genome-scale models.* BMC Medicine, 2025. **23**(1): p. 95.

174. Moretti, S., et al., *Oxidative stress as a biomarker for monitoring treated celiac disease.* Clin Transl Gastroenterol, 2018. **9**(6): p. 157.

175. Ferretti, G., et al., *Celiac disease, inflammation and oxidative damage: a nutrigenetic approach.* Nutrients, 2012. **4**(4): p. 243-57.

176. Fernandez-Jimenez, N., et al., *Coregulation and modulation of NFκB-related genes in celiac disease: uncovered aspects of gut mucosal inflammation.* Hum Mol Genet, 2014. **23**(5): p. 1298-310.

177. Christophersen, A., et al., *Therapeutic and Diagnostic Implications of T Cell Scarring in Celiac Disease and Beyond.* Trends in Molecular Medicine, 2019. **25**(10): p. 836-852.
178. Dumitru, C., A.M. Kabat, and K.J. Maloy, *Metabolic Adaptations of CD4+ T Cells in Inflammatory Disease.* Frontiers in Immunology, 2018. **Volume 9 - 2018**.
179. Palmer, C.S., et al., *Glucose Metabolism Regulates T Cell Activation, Differentiation, and Functions.* Frontiers in Immunology, 2015. **Volume 6 - 2015**.
180. Sevinc, E., et al., *Plasma glutamine and cystine are decreased and negatively correlated with endomysial antibody in children with celiac disease.* Asia Pac J Clin Nutr, 2016. **25**(3): p. 452-6.
181. Sevinc, E., et al., *AMINO ACID LEVELS IN CHILDREN WITH CELIAC DISEASE.* Nutr Hosp, 2015. **32**(1): p. 139-43.
182. Deters, B.J. and M. Saleem, *The role of glutamine in supporting gut health and neuropsychiatric factors.* Food Science and Human Wellness, 2021. **10**(2): p. 149-154.
183. Khalkhal, E., et al., *Screening of Altered Metabolites and Metabolic Pathways in Celiac Disease Using NMR Spectroscopy.* Biomed Res Int, 2021. **2021**: p. 1798783.
184. Calabrò, A., et al., *A metabolomic perspective on coeliac disease.* Autoimmune Dis, 2014. **2014**: p. 756138.
185. Bains, K., et al., *Prevalence and Impact of Malnutrition in Hospitalizations Among Celiac Diseases: A Nationwide Analysis.* Cureus, 2023. **15**(8): p. e44247.
186. Mędza, A. and A. Szlagatys-Sidorkiewicz, *Nutritional Status and Metabolism in Celiac Disease: Narrative Review.* J Clin Med, 2023. **12**(15).
187. Wallace Ross, C., *Intestinal absorption in coeliac disease, with some remarks on the effect of liver extracts upon carbohydrate metabolism.* Transactions of The Royal Society of Tropical Medicine and Hygiene, 1936. **30**(1): p. 33-50.
188. Nieminen, U., et al., *Duodenal disaccharidase activities in the follow-up of villous atrophy in coeliac disease.* Scand J Gastroenterol, 2001. **36**(5): p. 507-10.
189. Šuligoj, T., P.J. Ciclitira, and B. Božič, *Diagnostic and Research Aspects of Small Intestinal Disaccharidases in Coeliac Disease.* J Immunol Res, 2017. **2017**: p. 1042606.

190. Oliphant, K. and E. Allen-Vercoe, *Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health.* Microbiome, 2019. **7**(1): p. 91.
191. Vuoristo, M. and T.A. and Miettinen, *The Role of Fat and Bile Acid Malabsorption in Diarrhoea of Coeliac Disease.* Scandinavian Journal of Gastroenterology, 1987. **22**(3): p. 289-294.
192. Wierdsma, N.J., et al., *Vitamin and mineral deficiencies are highly prevalent in newly diagnosed celiac disease patients.* Nutrients, 2013. **5**(10): p. 3975-92.
193. Sprockett, D.D. and K.Z. Coyte, *When microbes go missing: Understanding the impact of diversity loss within the gut microbiome.* Cell Host & Microbe, 2023. **31**(8): p. 1249-1251.
194. Anneberg, O.M., et al., *The dietary inflammatory potential and its role in the risk and progression of inflammatory bowel disease: A systematic review.* Clin Nutr, 2025. **47**: p. 146-156.
195. Meng, M.J., et al., *Diet and the risk of inflammatory bowel disease: A retrospective cohort study in Taiwan.* Journal of the Formosan Medical Association, 2024.
196. Rizzello, F., et al., *Implications of the Westernized Diet in the Onset and Progression of IBD.* Nutrients, 2019. **11**(5).
197. Good, B.H. and L.B. Rosenfeld, *Eco-evolutionary feedbacks in the human gut microbiome.* Nature Communications, 2023. **14**(1): p. 7146.
198. Hartenstein, V. and P. Martinez, *Structure, development and evolution of the digestive system.* Cell Tissue Res, 2019. **377**(3): p. 289-292.
199. Clemente-Suárez, V.J., et al., *Global Impacts of Western Diet and Its Effects on Metabolism and Health: A Narrative Review.* Nutrients, 2023. **15**(12).
200. Gubatan, J., et al., *Dietary Exposures and Interventions in Inflammatory Bowel Disease: Current Evidence and Emerging Concepts.* Nutrients, 2023. **15**(3): p. 579.
201. Dang, Y., et al., *The Effects of a High-Fat Diet on Inflammatory Bowel Disease.* Biomolecules, 2023. **13**(6).
202. Jin, Z., et al., *Healthy Plant-Based Diet Is Associated With a Reduced Risk of Inflammatory Bowel Disease: A Large-Scale Prospective Analysis.* Mol Nutr Food Res, 2025: p. e70151.

203. Cox, S.R., et al., *Effects of Low FODMAP Diet on Symptoms, Fecal Microbiome, and Markers of Inflammation in Patients With Quiescent Inflammatory Bowel Disease in a Randomized Trial.* Gastroenterology, 2020. **158**(1): p. 176-188.e7.

204. Hou, J.K., D. Lee, and J. Lewis, *Diet and inflammatory bowel disease: review of patient-targeted recommendations.* Clin Gastroenterol Hepatol, 2014. **12**(10): p. 1592-600.

205. Shafiee, N.H., et al., *Anti-inflammatory diet and inflammatory bowel disease: what clinicians and patients should know?* Intest Res, 2021. **19**(2): p. 171-185.

206. Godny, L., et al., *Mechanistic Implications of the Mediterranean Diet in Patients With Newly Diagnosed Crohn's Disease: Multiomic Results From a Prospective Cohort.* Gastroenterology, 2025. **168**(5): p. 952-964.e2.

207. Kakodkar, S. and E.A. Mutlu, *Diet as a Therapeutic Option for Adult Inflammatory Bowel Disease.* Gastroenterol Clin North Am, 2017. **46**(4): p. 745-767.

208. Wang, J., et al., *THE PRACTICE AND PERCEPTION OF LOW RESIDUE DIETS IN ADULT PATIENTS WITH INFLAMMATORY BOWEL DISEASE.* Inflammatory Bowel Diseases, 2024. **30**(Supplement_1): p. S17-S18.

209. Hart, A.L., et al., *What Are the Top 10 Research Questions in the Treatment of Inflammatory Bowel Disease? A Priority Setting Partnership with the James Lind Alliance.* Journal of Crohn's and Colitis, 2017. **11**(2): p. 204-211.

210. Balestrieri, P., et al., *Nutritional Aspects in Inflammatory Bowel Diseases.* Nutrients, 2020. **12**(2).

211. Yao, T., M.H. Chen, and S.R. Lindemann, *Structurally complex carbohydrates maintain diversity in gut-derived microbial consortia under high dilution pressure.* FEMS Microbiol Ecol, 2020. **96**(9).

212. Seo, Y.S., et al., *Dietary Carbohydrate Constituents Related to Gut Dysbiosis and Health.* Microorganisms, 2020. **8**(3).

213. Basson, A.R., et al., *Regulation of Intestinal Inflammation by Dietary Fats.* Front Immunol, 2020. **11**: p. 604989.

214. Weisshof, R. and I. Chermesh, *Micronutrient deficiencies in inflammatory bowel disease.* Curr Opin Clin Nutr Metab Care, 2015. **18**(6): p. 576-81.

215. Zangara, M., N. Sangwan, and C. McDonald, *COMMON FOOD ADDITIVES ACCELERATE ONSET OF INFLAMMATORY BOWEL DISEASE IN MICE BY ALTERING*

MICROBIOME COMPOSITION AND HOST-MICROBE INTERACTION. Gastroenterology, 2021. **160**(3): p. S53.

216. Besedin, D., et al., *Food additives and their implication in inflammatory bowel disease and metabolic syndrome.* Clinical Nutrition ESPEN, 2024. **64**: p. 483-495.
217. Whelan, K., et al., *Ultra-processed foods and food additives in gut health and disease.* Nature Reviews Gastroenterology & Hepatology, 2024. **21**(6): p. 406-427.
218. Jarmakiewicz-Czaja, S., A. Sokal-Dembowska, and R. Filip, *Effects of Selected Food Additives on the Gut Microbiome and Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD).* Medicina, 2025. **61**(2): p. 192.
219. Xue, J.-C., et al., *Biological agents as attractive targets for inflammatory bowel disease therapeutics.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2025. **1871**(3): p. 167648.
220. Reznikov, E.A. and D.L. Suskind, *Current Nutritional Therapies in Inflammatory Bowel Disease: Improving Clinical Remission Rates and Sustainability of Long-Term Dietary Therapies.* Nutrients, 2023. **15**(3).
221. Ashton, J.J., J. Gavin, and R.M. Beattie, *Exclusive enteral nutrition in Crohn's disease: Evidence and practicalities.* Clinical Nutrition, 2019. **38**(1): p. 80-89.
222. van Rheenen, P.F., et al., *The Medical Management of Paediatric Crohn's Disease: an ECCO-ESPGHAN Guideline Update.* J Crohns Colitis, 2020.
223. Logan, M., et al., *Analysis of 61 exclusive enteral nutrition formulas used in the management of active Crohn's disease-new insights into dietary disease triggers.* Aliment Pharmacol Ther, 2020. **51**(10): p. 935-947.
224. Buchanan, E., et al., *The use of exclusive enteral nutrition for induction of remission in children with Crohn's disease demonstrates that disease phenotype does not influence clinical remission.* Aliment Pharmacol Ther, 2009. **30**(5): p. 501-7.
225. Levine, A. and E. Wine, *Effects of enteral nutrition on Crohn's disease: clues to the impact of diet on disease pathogenesis.* Inflamm Bowel Dis, 2013. **19**(6): p. 1322-9.
226. Day, A.S. and R.N. Lopez, *Exclusive enteral nutrition in children with Crohn's disease.* World J Gastroenterol, 2015. **21**(22): p. 6809-16.
227. Nichols, B., et al., *Gut metabolome and microbiota signatures predict response to treatment with exclusive enteral nutrition in a prospective study in children with active Crohn's disease.* Am J Clin Nutr, 2024. **119**(4): p. 885-895.

228. Geesala, R., et al., *Exclusive Enteral Nutrition Alleviates Th17-Mediated Inflammation via Eliminating Mechanical Stress-Induced Th17-Polarizing Cytokines in Crohn's-like Colitis.* Inflamm Bowel Dis, 2024. **30**(3): p. 429-440.
229. Borrelli, O., et al., *Polymeric diet alone versus corticosteroids in the treatment of active pediatric Crohn's disease: a randomized controlled open-label trial.* Clin Gastroenterol Hepatol, 2006. **4**(6): p. 744-53.
230. Pigneur, B., et al., *Mucosal Healing and Bacterial Composition in Response to Enteral Nutrition Vs Steroid-based Induction Therapy-A Randomised Prospective Clinical Trial in Children With Crohn's Disease.* J Crohns Colitis, 2019. **13**(7): p. 846-855.
231. de Sire, R., et al., *Exclusive Enteral Nutrition in Adult Crohn's Disease: an Overview of Clinical Practice and Perceived Barriers.* Clinical and Experimental Gastroenterology, 2021. **14**(null): p. 493-501.
232. Svolos, V., et al., *Treatment of Active Crohn's Disease With an Ordinary Food-based Diet That Replicates Exclusive Enteral Nutrition.* Gastroenterology, 2019. **156**(5): p. 1354-1367.e6.
233. Barnes, S., *Overview of Experimental Methods and Study Design in Metabolomics, and Statistical and Pathway Considerations.* Methods Mol Biol, 2020. **2104**: p. 1-10.
234. Blaise, B.J., et al., *Power Analysis and Sample Size Determination in Metabolic Phenotyping.* Analytical Chemistry, 2016. **88**(10): p. 5179-5188.
235. Billoir, E., V. Navratil, and B.J. Blaise, *Sample size calculation in metabolic phenotyping studies.* Briefings in Bioinformatics, 2015. **16**(5): p. 813-819.
236. Krumsiek, J., et al., *Gender-specific pathway differences in the human serum metabolome.* Metabolomics, 2015. **11**(6): p. 1815-1833.
237. Bucaciuc Mracica, T., et al., *MetaboAge DB: a repository of known ageing-related changes in the human metabolome.* Biogerontology, 2020. **21**(6): p. 763-771.
238. González-Domínguez, R., et al., *Recommendations and Best Practices for Standardizing the Pre-Analytical Processing of Blood and Urine Samples in Metabolomics.* Metabolites, 2020. **10**(6).
239. Smith, L., et al., *Important Considerations for Sample Collection in Metabolomics Studies with a Special Focus on Applications to Liver Functions.* Metabolites, 2020. **10**(3).

240. Souza, A.L. and G.J. Patti, *A Protocol for Untargeted Metabolomic Analysis: From Sample Preparation to Data Processing.* Methods Mol Biol, 2021. **2276**: p. 357-382.
241. Karaman, I., R. Climaco Pinto, and G. Graça, *Chapter Eight - Metabolomics Data Preprocessing: From Raw Data to Features for Statistical Analysis*, in *Comprehensive Analytical Chemistry*, J. Jaumot, C. Bedia, and R. Tauler, Editors. 2018, Elsevier. p. 197-225.
242. Broeckling, C.D., et al., *Current Practices in LC-MS Untargeted Metabolomics: A Scoping Review on the Use of Pooled Quality Control Samples.* Analytical Chemistry, 2023. **95**(51): p. 18645-18654.
243. DI, B. *QC:MXP Repeat Injection based Quality Control, Batch Correction, Exploration & Data Cleaning (Version 2) Zendono.* 2025; Available from: https://broadhurstdavid.github.io/QC-MXP/.
244. Nyamundanda, G., L. Brennan, and I.C. Gormley, *Probabilistic principal component analysis for metabolomic data.* BMC Bioinformatics, 2010. **11**(1): p. 571.
245. Andreella, A., et al., *Toward Power Analysis for Partial Least Squares-Based Methods.* Biom J, 2025. **67**(2): p. e70050.
246. Ruiz-Perez, D., et al., *So you think you can PLS-DA?* BMC Bioinformatics, 2020. **21**(1): p. 2.
247. Bylesjö, M., et al., *OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.
248. Jansen, J.J., et al., *Between Metabolite Relationships: an essential aspect of metabolic change.* Metabolomics, 2012. **8**(3): p. 422-432.
249. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
250. Wishart, D.S., et al., *HMDB 5.0: the Human Metabolome Database for 2022.* Nucleic acids research, 2022. **50**(D1): p. D622-D631.
251. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
252. Pang, Z., et al., *MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation.* Nucleic Acids Research, 2024. **52**(W1): p. W398-W406.
253. Li, S., et al., *Predicting network activity from high throughput metabolomics.* PLoS Comput Biol, 2013. **9**(7): p. e1003123.

254. Tautenhahn, R., et al., *XCMS Online: a web-based platform to process untargeted metabolomic data.* Anal Chem, 2012. **84**(11): p. 5035-9.
255. Gutierrez Reyes, C.D., et al., *Multi Omics Applications in Biological Systems.* Curr Issues Mol Biol, 2024. **46**(6): p. 5777-5793.
256. Prakash, C., P. Moran, and R. Mahar, *Pharmacometabolomics: An emerging platform for understanding the pathophysiological processes and therapeutic interventions.* International Journal of Pharmaceutics, 2025. **675**: p. 125554.
257. Vich Vila, A., et al., *Faecal metabolome and its determinants in inflammatory bowel disease.* Gut, 2023. **72**(8): p. 1472-1485.
258. Yamamoto, M., et al., *Urinary Metabolites Enable Differential Diagnosis and Therapeutic Monitoring of Pediatric Inflammatory Bowel Disease.* Metabolites, 2021. **11**(4).
259. Kim, J.E., et al., *Identifying robust biomarkers for the diagnosis and subtype distinction of inflammatory bowel disease through comprehensive serum metabolomic profiling.* Scientific Reports, 2025. **15**(1): p. 5661.
260. Santoru, M.L., et al., *Metabolic Alteration in Plasma and Biopsies From Patients With IBD.* Inflamm Bowel Dis, 2021. **27**(8): p. 1335-1345.
261. Jagt, J.Z., et al., *Fecal Amino Acid Analysis in Newly Diagnosed Pediatric Inflammatory Bowel Disease: A Multicenter Case-Control Study.* Inflammatory Bowel Diseases, 2022. **28**(5): p. 755-763.
262. Paydaş Hataysal, E., et al., *Impaired Kynurenine Pathway in Inflammatory Bowel Disease.* J Clin Med, 2024. **13**(20).
263. Nikolaus, S., et al., *Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases.* Gastroenterology, 2017. **153**(6): p. 1504-1516.e2.
264. Parada Venegas, D., et al., *Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases.* Front Immunol, 2019. **10**: p. 277.
265. Xu, H.-M., et al., *Characterization of short-chain fatty acids in patients with ulcerative colitis: a meta-analysis.* BMC Gastroenterology, 2022. **22**(1): p. 117.
266. Sommersberger, S., et al., *Altered fecal bile acid composition in active ulcerative colitis.* Lipids in Health and Disease, 2023. **22**(1): p. 199.
267. Vítek, L., *Bile Acid Malabsorption in Inflammatory Bowel Disease.* Inflammatory Bowel Diseases, 2015. **21**(2): p. 476-483.

268. Kaczmarczyk, O., et al., *Fecal Levels of Lactic, Succinic and Short-Chain Fatty Acids in Patients with Ulcerative Colitis and Crohn Disease: A Pilot Study.* J Clin Med, 2021. **10**(20).

269. Agouridis, A.P., M. Elisaf, and H.J. Milionis, *An overview of lipid abnormalities in patients with inflammatory bowel disease.* Ann Gastroenterol, 2011. **24**(3): p. 181-187.

270. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.* Genome Biology, 2012. **13**(9): p. R79.

271. Ning, L., et al., *Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts.* Nature Communications, 2023. **14**(1): p. 7135.

272. Solem, C.A., et al., *Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease.* Inflamm Bowel Dis, 2005. **11**(8): p. 707-12.

273. Chen, P., et al., *Serum Biomarkers for Inflammatory Bowel Disease.* Front Med (Lausanne), 2020. **7**: p. 123.

274. Banerjee, A., et al., *Faecal calprotectin for differentiating between irritable bowel syndrome and inflammatory bowel disease: a useful screen in daily gastroenterology practice.* Frontline Gastroenterol, 2015. **6**(1): p. 20-26.

275. Bermejo, F., et al., *Limitations of the determination of faecal calprotectin in patients with ulcerative colitis and inflammatory polyps.* Gastroenterol Hepatol, 2020. **43**(2): p. 73-78.

2.0 CHAPTER 2

An Optimised Monophasic Faecal Extraction Method for LC-MS Analysis and Its Application in Gastrointestinal Disease

Patricia E. Kelly [1,2] H Jene Ng [3], Gillian Farrell [1], Shona McKirdy [2,3], Richard K. Russell [2,4], Richard Hansen [2,5], Zahra Rattray [1], Konstantinos Gerasimidis [2,3] and Nicholas J. W. Rattray[1,2]*

The study was published in *Metabolites* and has 5 citations.

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G1 1XQ, UK.

[2] Department of Paediatric Gastroenterology Bacteria, Immunology, Nutrition, Gastroenterology and Omics Group, Royal Hospital for Children; University of Glasgow, Glasgow G12 8QQ, UK.

[3] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmiry, Glasgow G12 8QQ, UK.

[4] Royal Hospital for Children and Young People, 50 Little France Cresent, Edinburgh EH16 4TJ, UK.

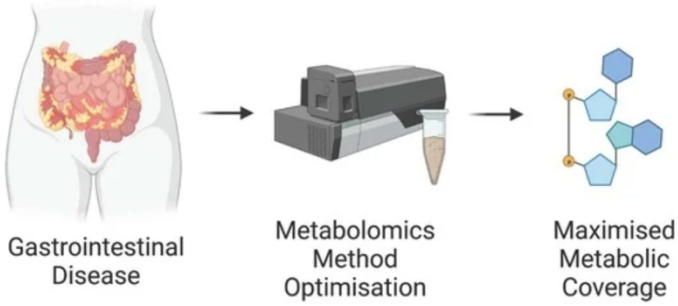
[5] Royal Hospital for Children, 1345 Govan Road, Glasgow G52 4TF, UK.

Author Contributions: Conceptualization, P.E.K., G.F., Z.R., K.G. and N.J.W.R.; methodology, P.E.K., H.J.N., S.M., Z.R., K.G. and N.J.W.R.; software, P.E.K., G.F. and N.J.W.R.; validation, P.E.K., G.F. and N.J.W.R.; formal analysis, P.E.K., H.J.N., S.M., Z.R., K.G. and N.J.W.R.; investigation, P.E.K., H.J.N., S.M. and N.J.W.R.; resources, R.K.R., R.H. and K.G.; data curation, P.E.K., G.F. and N.J.W.R.; writing—original draft preparation, P.E.K. and N.J.W.R.; writing—review and editing, P.E.K., H.J.N., G.F., S.M., R.K.R., R.H., Z.R., K.G. and N.J.W.R.; visualisation, P.E.K. and N.J.W.R.; supervision, N.J.W.R., Z.R. and K.G.; project administration, N.J.W.R.; funding acquisition, N.J.W.R., K.G. and Z.R. All authors have read and agreed to the published version of the manuscript.

2.1. Abstract

Liquid chromatography coupled with mass spectrometry (LC-MS) metabolomic approaches are widely used to investigate underlying pathogenesis of gastrointestinal disease and mechanism of action of treatments. However, there is an unmet requirement to assess faecal metabolite extraction methods for large-scale metabolomics studies. Current methods often rely on biphasic extractions using harmful halogenated solvents, making automation and large-scale studies challenging. The present study reports an optimised monophasic faecal extraction protocol that is suitable for untargeted and targeted LC-MS analyses. The impact of several experimental parameters, including sample weight, extraction solvent, cellular disruption method, and sample-to-solvent ratio, were investigated. It is suggested that a 50 mg freeze-dried faecal sample should be used in a methanol extraction (1:20) using bead beating as the means of cell disruption. This is revealed by a significant increase in number of metabolites detected, improved signal intensity, and wide metabolic coverage given by each of the above extraction parameters. Finally, we addressed the applicability of the method on faecal samples from patients with Crohn's disease (CD) and coeliac disease (CoD), two distinct chronic gastrointestinal diseases involving metabolic perturbations. Untargeted and targeted metabolomic analysis demonstrated the ability of the developed method to detect and stratify metabolites extracted from patient groups and healthy controls (HC), highlighting characteristic changes in the faecal metabolome according to disease. The method developed is, therefore, suitable for the analysis of patients with gastrointestinal disease and can be used to detect and distinguish differences in the metabolomes of CD, CoD, and HC.

Keywords: mass spectrometry; metabolite extraction; inflammatory bowel disease; Crohn's disease; coeliac disease



Graphical Abstract. Overview of experimental design for metabolomics method optimisation. LC-MS method development was carried out on samples from patients with gastrointestinal disease to maximise metabolite coverage.

2.2 Introduction

Metabolomics is a powerful tool for detecting small molecule cellular and microbial products. Through the reflection of active physiological mechanisms, metabolite characterisation and quantification can give critical insights into human health and disease. The large abundance and diversity of metabolites that are present in human faecal samples, as given by the identification of 6791 faecal metabolites on the Human Metabolome Database (HMBD) [1], provides an ideal target for metabolomic analysis [2] and, thus, allows for insights into the outcomes of gut-microbial interactions and dietary impacts on disease [3]. Accumulating evidence indicating the involvement of the gut metabolome in a multitude of diseases [4–6] has propelled an intense interest in the role of faecal metabolites under certain environments. The accurate quantification of metabolites in faecal samples, therefore, holds value in a wide range of research areas. A clear role of faecal metabolomics has been demonstrated in the field of gastrointestinal disease, including inflammatory bowel disease (IBD) [7] and coeliac disease (CoD) [8]. Although the aetiology of such diseases remains elusive, shifts in metabolic profile are associated with disease activity and may represent central components of pathogenesis [9–12]. Irrespective, detection of altered patient metabolites may help unravel underlying disease mechanisms or reveal new diagnostic or prognostic markers of clinical utility.

Liquid-chromatography mass spectrometry (LC-MS) is a popular metabolite analysis technique due to its high sensitivity and selectivity. Sample preparation and pre-treatment is a vital stage of the LC-MS workflow, providing the scaffolding to support metabolite

detection. The experimental framework therefore shapes the biological interpretation of a metabolomics study, and so it is crucial to consider best practices regarding specific study aims. Certain challenges are inherent in the sample preparation phase, such as the large physio-chemical variation of the target metabolite pool, technical and environmental variation, and the complex and heterogeneous nature of human faeces. This brings difficulties in standardising metabolomic methods, which is evident in the lack of "gold standard" metabolite extraction procedures. As the effective and reliable identification of metabolites is largely dependent on the extraction method used, it is imperative to consider sample preparation when comparing results between studies. To date, previous studies have addressed some of the challenges associated with metabolomic sample preparation [13–15]; however, these are mainly based on biphasic extraction protocols with limitations in scalability. While efficient biphasic extraction systems for faecal analysis contribute towards protocol standardisation, they are associated with complicated handling due to the requirement for phase separation. It can, therefore, be challenging to utilise two-phase protocols in large scale clinical studies, with further limitations in protocol automation. With the increasing demand for translating metabolomics data into meaningful clinical output, one major requirement for bridging the bench to bedside gap is the use of large population studies. It is, therefore, also important to optimise less-complex monophasic extraction protocols that can be used as an alternative to classical biphasic protocols for LC-MS analysis. Moreover, the applicability of metabolite extraction in the context of gastrointestinal disease requires further acknowledgement. Thus, the present study has the goal of advancing a method for monophasic metabolite extraction that can be easily implemented in large scale clinical studies investigating gastrointestinal disease. To the best of our knowledge, there is no current documentation on optimal extraction methods for IBD or CoD samples for LC-MS analysis. There is an important unmet requirement for the effect of faecal sample type to be explored, which is exemplified here in the comparison between gastrointestinal disease and the non-disease state.

Herein, we evaluate different faecal extraction methods for metabolomic measurements in human faecal samples from healthy individuals, Crohn's Disease (CD) and CoD patients. A range of trial experiments were performed to determine the optimal sample weight, extraction solvent, disruption method, and sample-to-solvent ratio using LC-MS. The overall aim of this study is twofold; firstly, to optimise metabolite extraction parameters for faecal

samples and secondly, to determine whether this optimised extraction protocol is suitable for analysis of samples from patients with gastrointestinal disease. To capture the large quantity of metabolites and ensure maximal coverage in the method development phase, untargeted metabolomic analysis was performed to assess each sample parameter. Targeted metabolomic analysis was subsequently applied to assess method suitability in patients with disease.

2.3 Materials and Methods

2.3.1 Ethics Statement

All participants and their carers provided written informed consent. The study was approved by the NHS West of Scotland Research Ethics Committee (14/WS/1004 for Crohn's disease patients and 11/WS/0006 for patients with coeliac disease). Ethical approval from the University of Strathclyde Departmental Ethics Committee (DEC) was not required for this study, as all research activities involved anonymised patient samples collected under NHS approval. All patient data were treated in accordance with data. protection regulations, anonymised prior to analyses, with confidentiality ensured.

2.3.2 Faecal Samples

Faecal samples were collected for metabolomic analysis within 2 h of passage, kept in anaerobic conditions (Anaerocult™ A) and inside an ice box with ice packs. The samples were transferred to the laboratory immediately, homogenised with mechanical kneading, and aliquots were kept at −80 °C until further processing. After metabolite extraction, the samples were again kept at −80 °C until LC-MS analysis. The samples were kept on ice during transportation.

2.3.3 Chemicals and Reagents

LC-MS grade methanol (MeOH), acetonitrile (ACN), chloroform ($CHCl_3$), and water (H_2O) were purchased from Fisher Scientific (Geel, Belgium). LC-MS grade formic acid was purchased from Thermo Scientific (Prague, Czech Republic).

2.3.4 Extraction Protocol

Freeze-dried faecal samples were added to the extraction solvent and the cells were disrupted using bead beating (FastPrep 24 MP Biomedicals), sonication, and freeze-thaw lysis methods. The samples were then centrifuged at 13,000× *g* for 15 min and the supernatant recovered. The samples were evaporated to dryness using a SpeedVac Savant SPD121P system (Thermo Scientific, Milford, UK) and stored at −80 °C until further processing. Reconstitution was performed in 250 μL 50/50 H_2O: acetonitrile (ACN), vortexed for 1 min and centrifuged at 15,000× *g* for 15 min, and aliquots transferred into glass vials for MS analysis. Quality control (QC) samples were prepared by pooling samples across all groups undergoing simultaneous analysis. Solvent blanks and QC samples were entered at the beginning of every analytical run, and after every five samples in each batch over the course of the study to assess background in the system and detect potential contaminations. Experimental details for each extraction parameter are shown (**Table 1**).

Table 1. Experimental conditions for each extraction parameter.

Experiment	Independent Variable	Sample Weight	Solvent Used	Cell Lysis Method
1	Sample weight	10 mg, 20 mg, 50 mg, 100 mg	MeOH	Bead beating (5 ms^{-1}, 60 s)
2	Extraction solvent	50 mg	MeOH, 1:1 MeOH/H_2O, 2:1 $CHCl_3$/MeOH	Bead beating (5 ms^{-1}, 60 s)
3	Cell lysis method	50 mg	MeOH	Bead beating (5 ms^{-1}, 60 s), sonication (40 kHz) freeze-thaw cycle (24 h)
4	Sample-to-solvent ratio	50 mg	MeOH	Bead beating (5 ms^{-1}, 60 s)

2.3.5 Untargeted LC-MS Metabolite Measurement

Untargeted metabolomic analysis was performed on an ultra-high performance liquid chromatography (UHPLC) system (ThermoFisher Scientific) coupled to an Orbitrap Exploris 240 (ThermoFisher Scientific) mass spectrometer. The LC-MS method was previously optimised on the Orbitrap system, with the settings transferred from the applied method [16]. Chromatographic separation was performed on a Vanquish Accucore C18 + UHPLC analytical column (ThermoScientific, 100 mm × 2.1 mm, 2.6 μM) at a flow rate of 400 μL min^{-1}. Mobile phase A was composed of 99.9% water + 0.1% formic acid and mobile B was composed of 99.9% MeOH + 0.1% formic acid. Electrospray ionisation (ESI) was used as the ionisation method, set at 3900 V and 2500 V for positive and negative mode, respectively. The elution gradient used can be found in Supplementary Information Table S1. The source-dependent parameters were operated under the following conditions: sheath gas, 40 Arb; auxiliary gas, 10 Arb; sweep gas, 1 Arb; ion transfer tube temperature, 300 °C; vaporiser temperature, 280 °C. Instrument calibration was performed using Pierce™ FlexMix™ calibration solution (Thermo Scientific) and ran under vendor recommended settings. MS data collection was performed in a top-5 data dependent acquisition mode (DDA) to give putative metabolite identification at MSI level 2.

2.3.6 Targeted LC-MS Metabolite Measurement

Targeted metabolomic analysis was performed on a UHPLC system coupled to a triple quadrupole mass spectrometer (Shimadzu 8060NX, Kyoto, Japan). The method used for metabolite detection and quantification was provided by the vendor; Primary Metabolites LC/MS/MS Method Package version 2.0 (Shimadzu, Kyoto, Japan). The method was designed to detect 97 metabolites. The list of 97 detected metabolites and associated parameters are shown in Supplementary Information Table S4. Chromatographic separation was performed on a pentafluorophenylpropyl (PFPP) + UHPLC analytical column (Merck, 150 mm × 2.1 mm, 3 μM) at a flow rate of 400 μL min^{-1}. Mobile phase A was composed of 99.9% water + 0.1% formic acid and mobile B was composed of 99.9% acetonitrile + 0.1% formic acid. Electrospray ionisation (ESI) was used as the ionisation method, set at 3900 V and 2500 V for positive and negative mode, respectively. The source-dependent parameters were operated under the following conditions: column oven temperature, 40 °C, nebulising gas flow rate, 3.0 L min^{-1}, drying gas flow rate, 10 L min^{-1},

desolvation line temperature, 250 °C, and block heater temperature, 400 °C. The elution gradient used can be found in Supplementary Information Table S2.

2.3.7 Method Application

We applied the method to three biological groups: CD patients, CoD patients, and HCs (Table 2). CD patients were undergoing varying forms of treatment and CoD patients were following a gluten-free diet. HCs were defined as individuals with the absence of gastrointestinal disease. Both untargeted and targeted metabolomic analyses were applied to the sample sets combined after randomisation.

This analysis was conducted as a subset study, using a smaller group of 20 patients selected from larger disease cohorts to assess method application. While the primary studies were powered to detect a difference in faecal calprotectin levels between intervention and control groups in the full cohort, no formal power calculation was performed specifically for this subset. This subset was chosen to balance sample availability, resource costs, and expected variability in metabolic measurements, while maintaining representation across the study arms.

Table 2. Table of patient demographics.

Variable	**HC** ***n* = 20**	**CD** ***n* = 20**	**CoD** ***n* = 20**
Gender			
Female (%)	45	40	60
Male (%)	55	60	40
Age (range)	6.6 (2.3–13.7)	12.3 (7.6–14.8)	9.2 (4.0–14.8)
BMI z-score	0.3	−0.7	0.2

HC, healthy control; CD, Crohn's disease, CoD, Coeliac disease.

2.3.8 Mass Spectrometry Data Processing

For the processing of untargeted metabolomics data, Thermo Scientific Xcalibur format raw data files (.RAW) were imported into Compound Discoverer software version 3.2 (Thermo Fisher Scientific, Waltham, MA, USA). Details of the workflow for analysis in Compound Discoverer is included in Supplementary Information Table S5. The targeted metabolomics data were converted from Shimazdu vendor format (.lcd) to mzML format. A data matrix of identified metabolites and associated peak areas was constructed and processed using R-Studio v 3.5.2 (RStudio, PBC MA, USA).

2.3.9 Data and Statistical Analysis

For untargeted analysis, principal component analysis (PCA) was performed using Compound Discoverer software 3.2 (Thermo Fisher Scientific, Waltham, MA, USA). For targeted analysis, PCA was performed using Lab Profiling Solutions software version 5.6 (Shimadzu, Kyoto, Japan) and R-Studio (RStudio, PBC, MA, USA). PCs were calculated using prcomp function and PCA scores plots were generated using the following packages in R: ggplot2, ggfortify, grid, and gridExtra. Differential analysis using volcano plots allowed significant differences between groups to be determined. Univariate statistical analyses were performed using unpaired *t*-test and one-way ANOVA, with the level of significance set at $p < 0.05$. Central network analysis was performed in R-studio (RStudio, PBC, MA, USA) using the igraph package.

2.3.10 Putative Metabolite Identification

Putative metabolite identification was performed by assigning likely metabolite identities to detected features in a metabolomics dataset based on accurate mass, fragmentation patterns, and database matches, without conforming the identity using authentic chemical standards. The inclusion criteria for putative metabolite identification were set and applied to refine the total number of features in the metabolomics dataset. Only features with a full mzCloud match and mass accuracy within 4 ppm were retained, and duplicate entries were removed. Contaminants were excluded by analysing blank samples interspersed throughout the analysis. Metabolite identification was performed both manually and using reference databases, with the Human Metabolome Database (HMDB) [1] serving as the primary source for metabolite identification.

2.4 Results

For method development, metabolites were measured in freeze-dried faecal samples obtained from healthy participants. The metabolic output was first measured by PCA to observe any differences in the overall metabolic signature obtained from each method. Further statistical analysis was performed for data quantification by calculating the number of *m*/*z* features, putatively identified metabolites, signal intensity, and metabolic coverage.

2.4.1. Analysis of Sample Weight

Positive ionisation mode was used for analysis of experimental parameters as previous investigations found that a significantly higher number of *m*/*z* features were detected in comparison to the negative ionisation mode. While examining the effect of sample weight, 10 mg samples were disregarded during the extraction process as the aliquots had very little extractable supernatant for subsequent processing. This was likely due to the sample being absorbed by the zirconium beads as the sample weight was too small for the solvent volume. During the reconstitution step, the 100 mg sample was also disregarded as there was too much particulate left undissolved. The metabolites were successfully extracted from 20 mg and 50 mg samples and measured using LC-MS and PCA demonstrated clear separation of the two sample weight groups (**Figure 1**). In this case, 50 mg samples showed a significantly higher mean number of *m*/*z* features and mean number of putatively identified metabolites in comparison to 20 mg samples. Furthermore, the mean signal intensity given by 50 mg samples (2.1×10^7) was significantly increased compared to 20 mg samples (1.2×10^7). As shown in Figure 1F, both sample weight displayed similar overall distributions of metabolite classes. It was furthermore demonstrated that 69.1% of detected metabolites were found at significantly increased levels in 50 mg samples compared to 20 mg samples (**Supplementary Information Figure S1**). A comparison of the total number of metabolites per chemical class from each sample weight is shown in **Figure 2**.

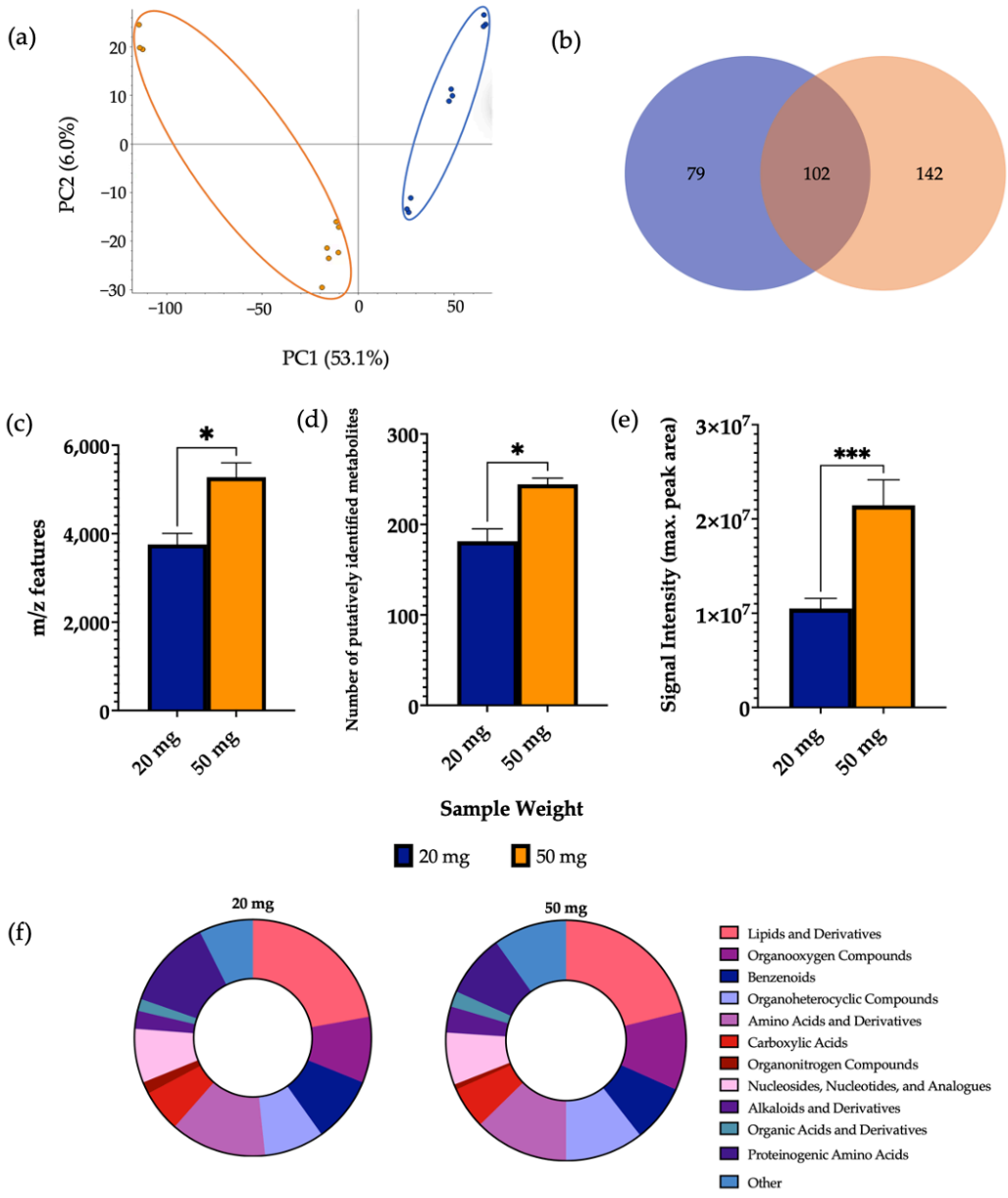


Figure 1. The effect of sample weight on features of metabolomic analysis. 1 µL of 20 mg and 50 mg sample was injected onto a C18 column (*n* = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of sample weight. PCA score plots demonstrating extracted faecal metabolites between different sample weights. Discrimination between 20 mg (blue) and 50 mg (orange) samples was characterised by a variability of 53.1%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m/z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each sample weight was assessed. (**f**) A metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in 20 mg and 50 mg samples. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using an unpaired *t*-test. * $p < 0.05$, *** $p < 0.001$.

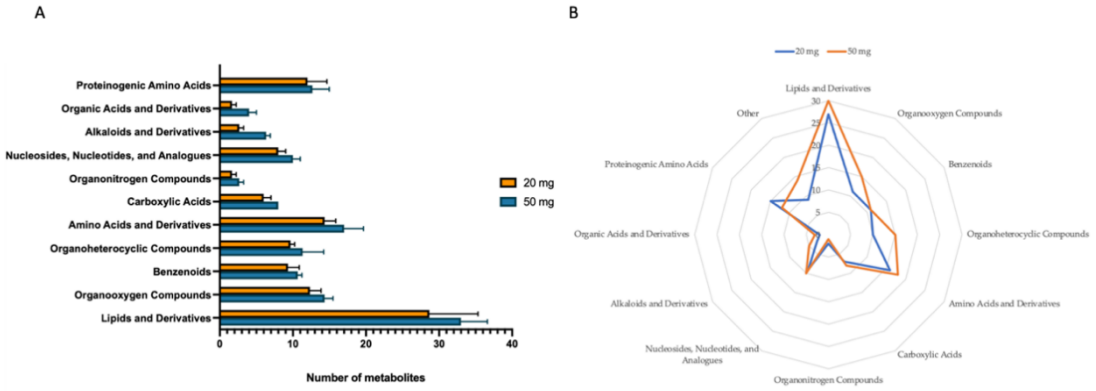


Figure 2. Untargeted metabolite class analysis of sample weight. **(A)** Comparison of the total number of metabolites identified by chemical class in 20 mg and 50 mg samples (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in 20 mg and 50 mg samples. Data were expressed as mean ± SEM and statistical significance was assessed using unpaired t-test.

2.4.2 Analysis of Extraction Solvent

PCA demonstrated a clear separation of the extraction solvents (**Figure 3**). Using 100% MeOH gave a significantly higher number of *m/z* features in comparison to 1:1 MeOH/H_2O and a significantly higher number of putatively identified metabolites than both MeOH/H_2O and 2:1 $CHCl_3$/MeOH. No significant differences were observed in the signal intensity between the extraction solvents. Differential analysis revealed a significant increase in the levels of 30.6% and 20.9% of metabolites detected using MeOH as the extraction solvent in comparison to MeOH/H_2O and $CHCl_3$/MeOH, respectively (**Supplementary Information Figure S2**). In this case, 32.0% of metabolites detected were found at significantly increased levels in $CHCl_3$/MeOH compared to MeOH/H_2O. MeOH extractions additionally had a significantly increased number of lipids compared to MeOH/H_2O extractions. Once more, all metabolite classes were detected from all extraction solvents, with a similar structure of metabolite classification. A comparison of the total number of metabolites per chemical class from each extraction solvent is shown in **Figure 4**.

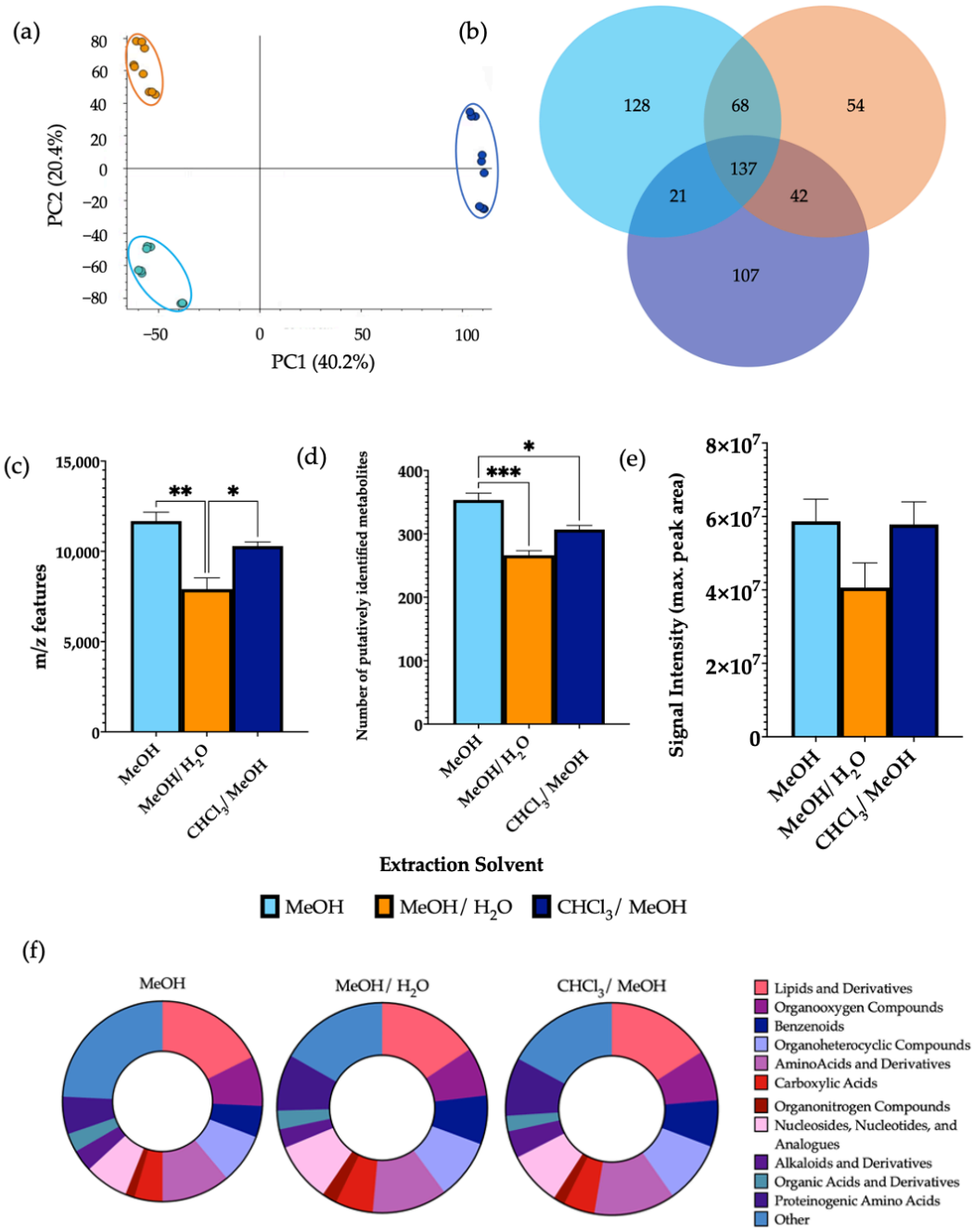


Figure 3. The effect of extraction solvents, MeOH, MeOH/H_2O, and $CHCl_3$/MeOH, on features of metabolomic analysis. 1 µL of each extraction sample was injected onto a C18 column (n = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of extraction solvent. PCA score plots demonstrating extracted faecal metabolites between different extraction solvents. Discrimination between extraction solvents MeOH (light blue), MeOH/H_2O (orange), and $CHCl_3$/MeOH (dark blue) was characterised by a variability of 40.2%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of m/z features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each extraction solvent was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using one-way ANOVA. *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

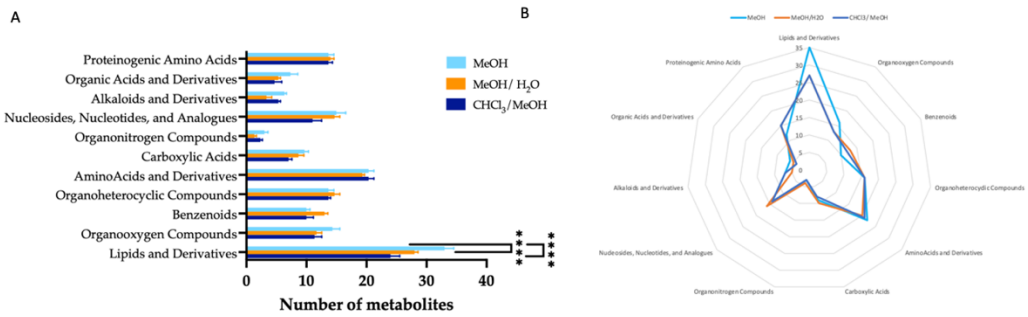


Figure 4. Untargeted metabolite class analysis of extraction solvent. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted with MeOH, MeOH/H_2O, and $CHCl_3$/ MeOH (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted with MeOH/ H_2O, MeOH, and $CHCl_3$/ MeOH. Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *p < 0.05, **** p < 0.0001.

2.4.3 Analysis of the Cellular Disruption Method

The choice of cellular disruption method affected the overall metabolic output, as shown by PCA which demonstrated a clear separation between the three groups (**Figure 5**). Bead beating extracted a significantly higher mean number of *m/z* features in comparison to freeze-thawing and a significantly higher number of putatively identified metabolites than both sonication and freeze-thawing. No significant differences were observed in the signal intensity between lysis methods. A significant increase in the levels of 29.5% and 48.4% of metabolites detected were found using bead beating as the method of cellular disruption compared to sonication and freeze-thawing, respectively (**Supplementary Information Figure S3).** Of the metabolites identified, 23.7% were found at significantly increased levels in sonicated samples in comparison to freeze-thawing. Each disruption method allowed for the measurement of metabolites from all classification groups. While similar patterns of metabolite classification are shown between methods, it was shown that bead beating led to detection of a significantly increased number of lipids compared to the other lysis techniques. A comparison of the total number of metabolites per chemical class using each cellular disruption method is shown in **Figure 6**.

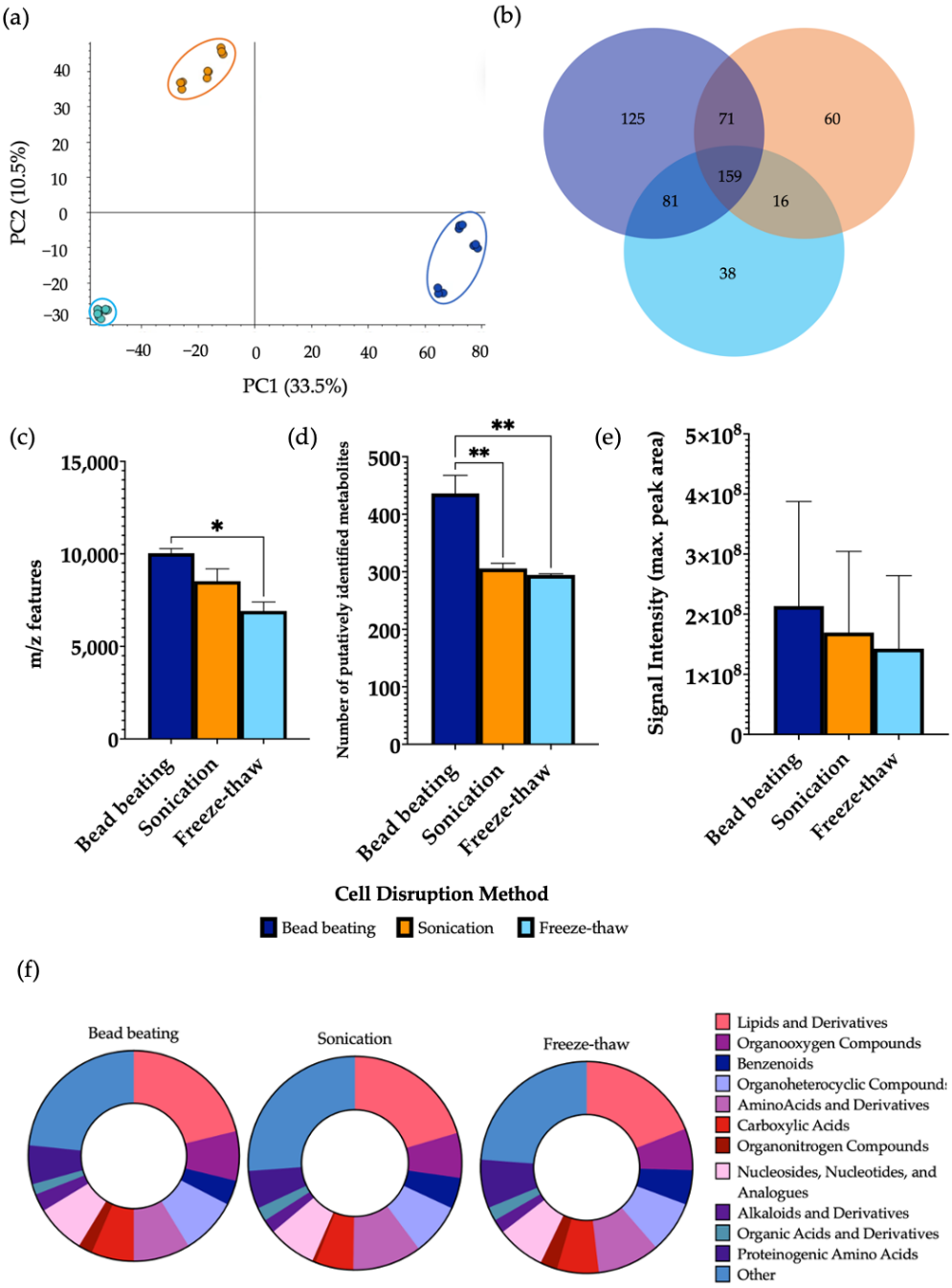


Figure 5. The effect of cellular disruption methods, bead beating, sonication, and freeze-thaw cycles, on features of metabolomic analysis. 1 µL of each extraction sample was injected onto a C18 column (*n* = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of disruption method. PCA score plots demonstrating extracted faecal metabolites between bead beating, sonication, and freeze-thaw cycles. Discrimination between extraction solvents A, bead beating (dark blue); B, sonication (orange) and C, freeze-thaw cycles (light blue) was characterised by a variability of 33.5%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m/z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each disruption method was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *$p < 0.05$, ** $p < 0.01$.

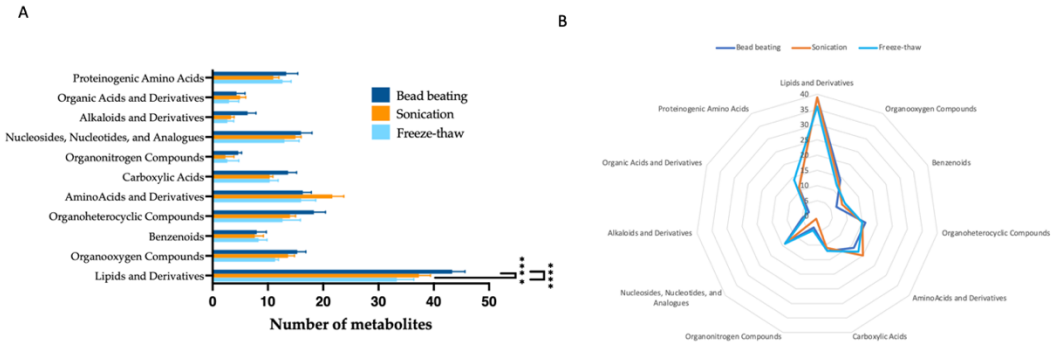


Figure 6. Untargeted metabolite class analysis of cellular disruption method. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted using bead beating, sonication, and freeze-thaw cycles (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted using bead beating, sonication, and freeze-thaw cycles. Data are expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA, **** $p < 0.0001$.

2.4.4 Analysis of Sample-to Solvent Ratio

A clear separation was observed between the three different sample-solvent ratios by PCA (**Figure 7**). Performing extractions using a ratio of 1:20 gave a significantly higher mean number of *m/z* features and putatively identified metabolites than ratios of 1:5 and 1:10. Furthermore, a significant increase in the signal intensity of samples of a 1:20 ratio was observed in comparison to the other groups. A significant increase in the levels of 70.0% and 66.7% of metabolites detected were found using a ratio of 1:20 in comparison to ratios of 1:5 and 1:10, respectively (**Supplementary Information Figure S4**). In this case, 43.5% of metabolites detected were found at significantly increased levels in samples extracted using a ratio of 1:10 compared to 1:5. Several metabolite classes were increased in extractions carried out using a ratio of 1:20 compared to the other groups. Additionally, the overall composition according to chemical class of each sample remained similar between each group. A comparison of the total number of metabolites per chemical class using each sample-to-solvent ratio is shown in **Figure 8**.

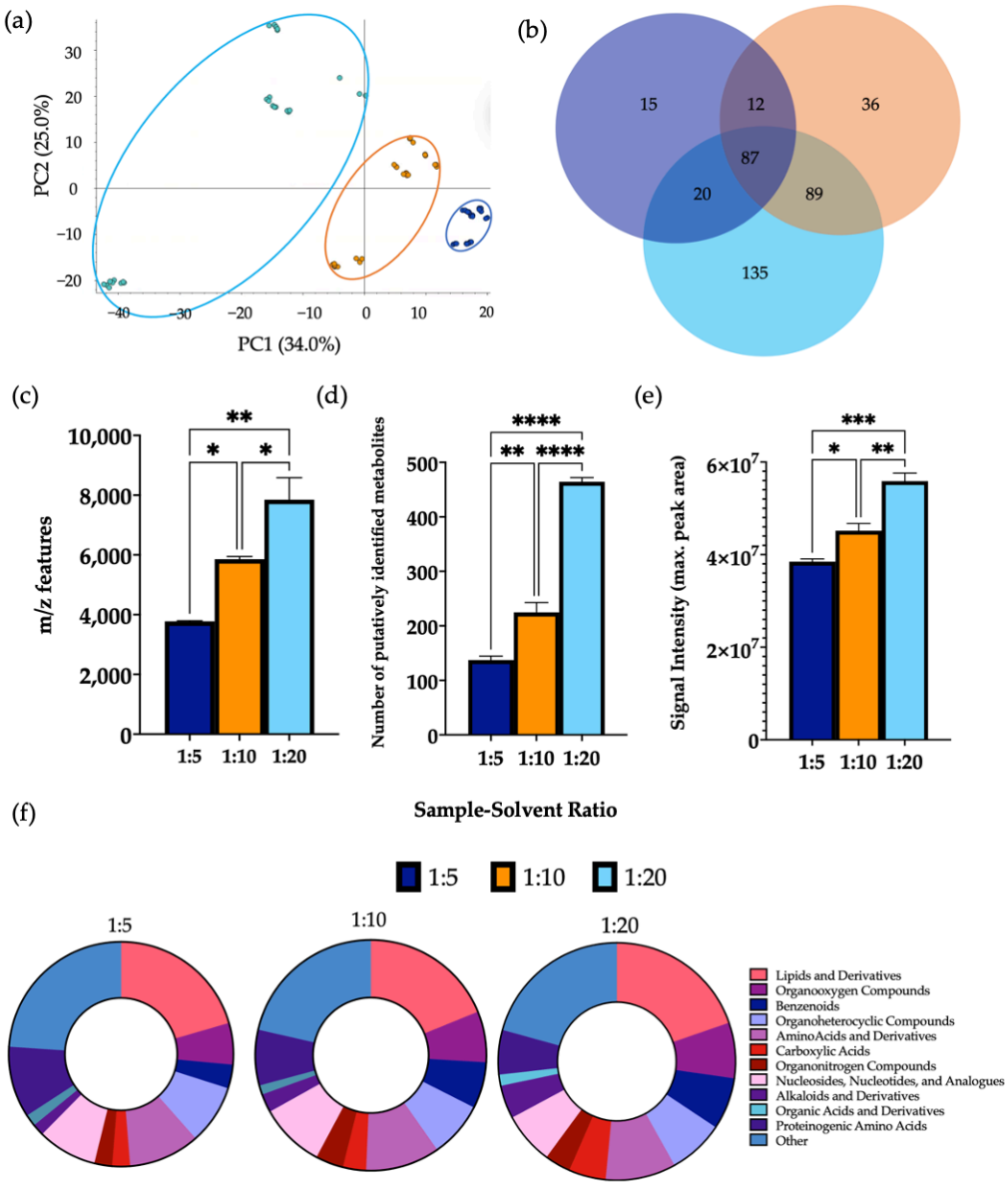


Figure 7. The effect of sample-solvent ratio on features of metabolomic analysis. 1 µL of each extraction sample was injected onto a C18 column (n = 3), performed in triplicate. (**a**) PCA of metabolomic profiles obtained as a function of sample-to-solvent ratio. PCA score plots demonstrating extracted faecal metabolites between different ratios. Discrimination between extraction solvents 1:5 (dark blue), 1:10 (orange) and 1:20 (light blue) was characterised by a variability of 33.3%. (**b**) A Venn diagram of the mean number of metabolites detected between each method. (**c**) The total number of *m/z* features and (**d**) total number of putatively identified metabolites were calculated in positive ionisation mode and (**e**) the overall mean signal intensity of each sample-to-solvent-ratio was assessed. (**f**) The metabolite class quantification demonstrating the faecal metabolome patterns according to chemical class in each extraction sample. The bar chart data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. *$p < 0.05$ ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

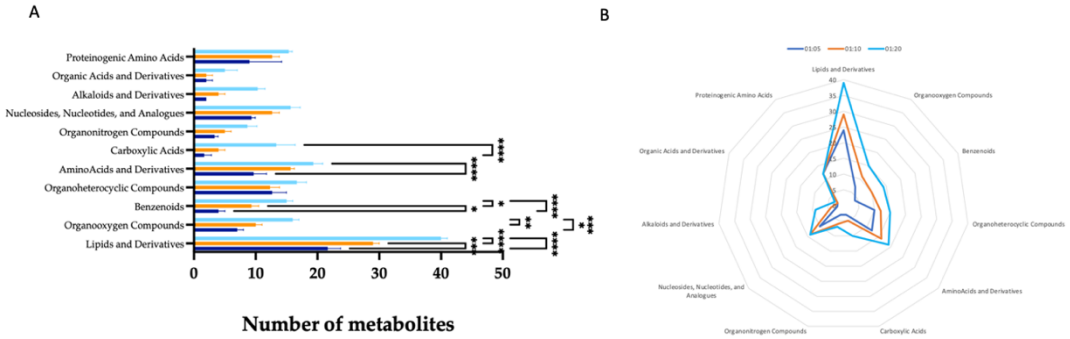


Figure 8. Untargeted metabolite class analysis of sample-to-solvent ratio. **(A)** Comparison of the total number of metabolites identified by chemical class in samples extracted using sample-to-solvent ratios of 1:5, 1:10, 1:20 (n=3), performed in triplicate. **(B)** Radar plot comparing the relative abundance of metabolite classes in samples extracted using 1:5, 1:10, 1:20. Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA., ** $p < 0.01$, **** $p < 0.0001$.

Through the exploration of the overall metabolite extraction efficiency through the optimisation process, it was observed that the number of putatively identified metabolites significantly increased throughout stages of method optimisation with the improvement of each individual extraction parameter (**Figure 9**).

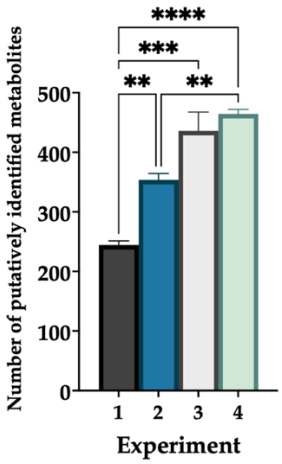


Figure 9. Comparison of individual optimization experiments. Total number of putatively identified metabolites given by optimal parameters of each experiment. Experiment 1, Analysis of Extraction Weight; Experiment 2, Analysis of Extraction Solvent; Experiment 3; Analysis of Cellular Disruption Method; Experiment 4, Analysis of Sample-to-Solvent Ratio. Data were expressed as mean ± SEM and statistical significance was assessed using a one-way ANOVA. **$p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

2.4.5 Applicability of the Method to Patients with Gastrointestinal Disease

To assess the applicability of the developed method, we applied the protocol to CD, CoD, and HC groups and compared the metabolic differences. In an untargeted analysis, PCA demonstrated a clear separation between CD samples and the other groups (**Figure 10**). A significant decrease in the levels of 72.3% of metabolites detected were found in CD samples compared to HCs, and 74.1% compared to CoD samples (**Supplementary Information Figure S5**). Of the metabolites detected, 27.1% were found to be at significantly decreased levels in CoD samples in comparison to HCs. Furthermore, targeted metabolomics analysis further confirmed the ability of the method to both detect and stratify metabolites extracted from faecal sample from patients with CD and CoD and healthy individuals. PCA showed characteristic changes in the faecal metabolome between each of the groups (**Figure 11**). In order to ensure the present method was effective in the specific context of gastrointestinal disease, we carried out further analysis investigating metabolites that are important in IBD. The metabolites that were putatively identified in the current method and throughout the literature in the context of IBD are compared (**Figure 12**).

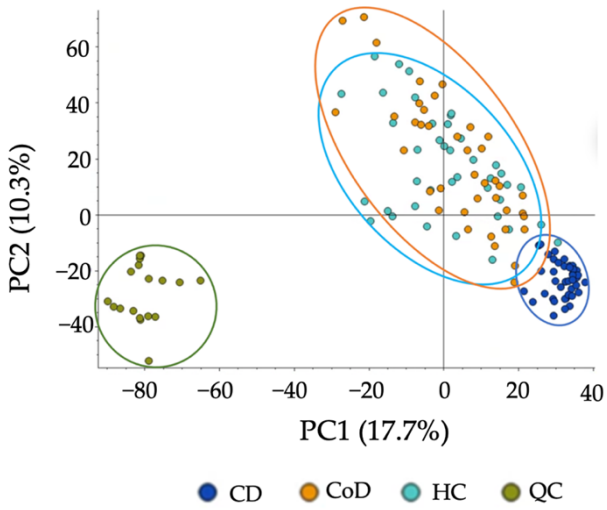


Figure 10. PCA of metabolomic profiles based on untargeted analysis of gastrointestinal disease. PCA score plots demonstrating extracted faecal metabolites between patient groups. Principle Component 1 directionality describes the variance between CD (dark blue), CoD (orange) and HC (light blue) and explains 17.7% of the total variance of the data. QCs are shown in green. The samples were performed in triplicate and are shown as individual datapoints to represent the variance in the dataset.

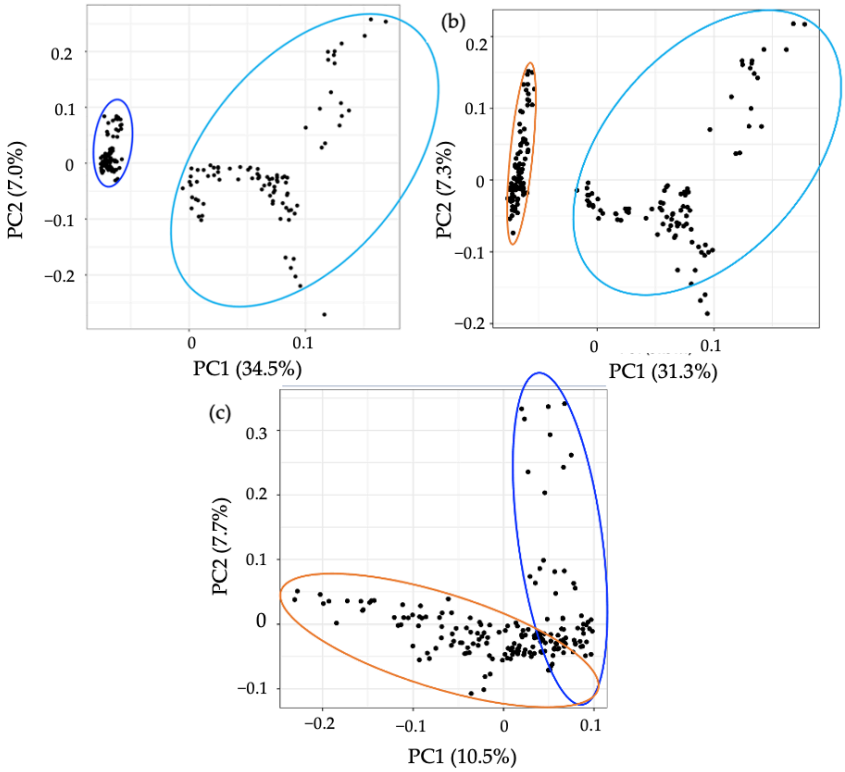


Figure 11. PCA of the metabolomic profiles based on targeted analysis of gastrointestinal disease. PCA score plots demonstrating extracted faecal metabolites between CD (dark blue), CoD (orange) and HC (light blue). The discrimination between (**a**) CD vs. HC, (**b**) CoD vs. HC, and (**c**) CD vs. Co was characterised by variabilities of 34.5%, 31.3%, and 10.5%, respectively. The samples were performed in triplicate and are shown as individual datapoints to represent the variance in the dataset.

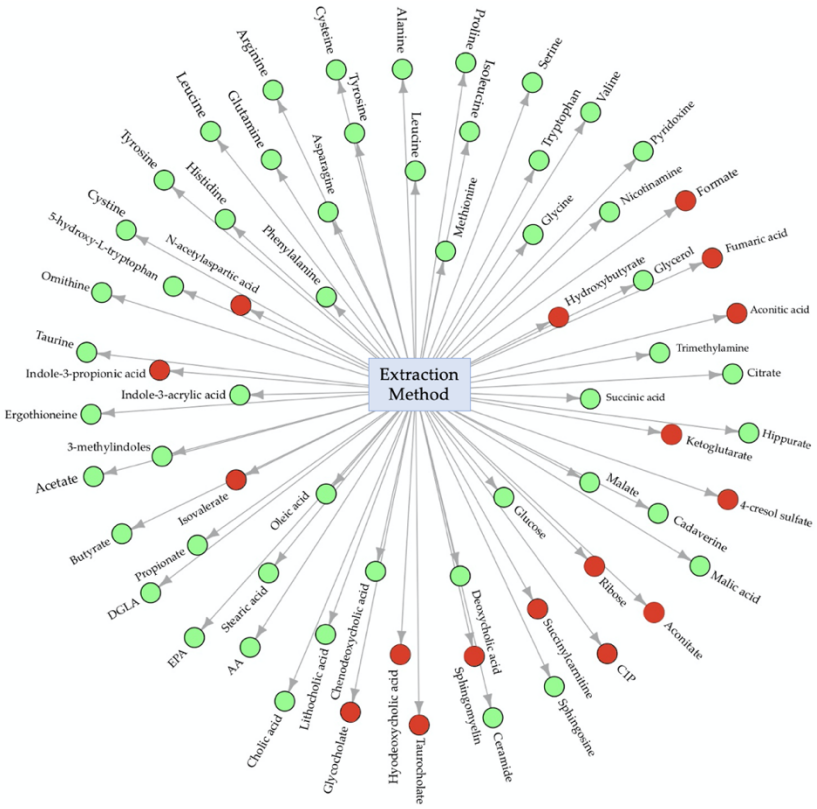


Figure 12. Central network analysis of developed metabolite extraction method. Circles shown in green represent metabolites successfully extracted using the developed method and circles shown in red represent metabolites not found using the developed method. C1P, Ceramide-1-phosphate; AA, Arachidonic acid; EPA, Eicosapentanenoic acid; DGLA, Dihomo-gamma linolenic acid.

2.5 Discussion

Since extraction methodology directly affects metabolite constitution within MS metabolomics experiments, it was important to optimise a range of experimental parameters and to document the chemical coverage in faecal samples. To this end, the present study first aimed to assess parameters of maximal metabolic LC-MS output, utilising an untargeted metabolomics approach to allow fingerprinting of the total metabolite profile in samples. The ideal extraction protocol was therefore one that elucidated the greatest number of metabolites whilst minimising interferences. As such, methods were evaluated by measuring the total number of metabolites detected using each protocol. Since we cannot assume that the number of features is equal to the number of correctly identified metabolites, due to unmatched features, blanks, and duplicate

readings, further refinement methods were applied to allow for a more accurate evaluation of the protocols. Additionally, the markedly different characteristics of metabolites in the faecal metabolite pool brings challenges in extracting all the metabolites present in each sample. For this reason, it was important to assess the number of metabolites belonging to different metabolic classes from each method to ensure maximum chemical coverage. Feature annotation was performed to quantify and compare metabolite classifications between the extraction methods. As a complete characterisation of the metabolome is not possible, a compromise will always exist in practice; however, the multi-parameter method used in the present study allows for the selection of the greatest metabolite signal and coverage.

Herein, we describe an optimised protocol for extraction of metabolites from human faecal samples, thus providing an efficient setup for subsequent metabolomic analysis. The method is recapitulated in the following stages: (1) 50 mg sample weighed out, (2) 1000 μL MeOH added to sample and cell lysed by bead beating, (3) samples evaporated to dryness under vacuum and stored at −80 °C until further processing, (4) reconstitution carried out in 50/50 ACN: H_2O, (5) LC-MS analysis using 1 μL injection volume (Supplementary Information Figure S12).

The metabolite extraction from 10 mg and 100 mg samples were unsuitable for metabolomic analysis and therefore not included in the results. This is important, as when run on the MS, sample particulate may crash the column and lead to instrument breakdown. The faeces weight-to-solvent ratio (100 μL of solvent for every 10 mg of sample) was, therefore, not sufficient for samples out with a 20–50 mg range. For this reason, we explored the impact of sample-solvent ratio on metabolic output in a further analysis. In consideration to this, for the assessment of sample weight, 20 mg and 50 mg samples were successfully extracted and metabolomic analysis was continued. A clear separation was shown by the PCA comparing 20 mg and 50 mg samples, indicating the different metabolite profiles given by the two groups. Further analysis showed that 50 mg samples additionally contained an increased number of *m*/*z* features, identified metabolites, and signal intensity—this result was to be expected due to the increased levels of biomass in the 50 mg samples. It was also important to investigate whether the observed differences in metabolite numbers were reflected in the overall metabolic coverage. Thus,

the detected metabolites were grouped according to their chemical classification, and calculation of the number of metabolites in each class was used as a measurement of metabolic coverage. This is essential for untargeted metabolomics experiments, as the analytical conditions should aim to detect a broad range of metabolites of different chemical properties that may be implicated in disease. As such, expansion of metabolic coverage is important to maximise information for hypothesis generation. From the classification analysis, it was revealed that metabolite class is conserved across sample weight. Using 50 mg faecal samples for metabolite extraction aligns with previously reported studies [3,17–19], in which 50 mg samples were also used as the starting point for sample preparation and subsequent analysis. Based on findings of increased metabolite numbers without compromising metabolic coverage or signal intensity, it is reasonable to suggest that 50 mg samples are optimal for use in faecal extraction protocols.

While investigation into extraction solvent was here carried out using MeOH, MeOH/H_2O, and $CHCl_3$/MeOH, it is worth mentioning that other solvents, such as ACN and isopropanol have previously been used in faecal extractions. However, due to limited clinical sample availability, the extraction solvents for this study were chosen based on a previous literature search. The results from this analysis showed a clear separation between protocols using MeOH, MeOH/H_2O, and $CHCl_3$/MeOH, with an increased number of *m*/*z* features and identified metabolites given by pure MeOH extractions. While it was shown that the number of lipids and derivatives were increased in the samples extracted using MeOH in comparison to the other groups, the overall metabolic coverage was very similar for all extraction solvents investigated. As maximal chemical coverage is largely maintained, it can again be noted that metabolite class is conserved across extraction solvents. As the use of pure MeOH increases overall metabolic features obtained from molecules across a wide range of different chemical properties, its use can therefore be recommended as the optimal solvent for faecal extraction. This result agrees with a recently reported study, where MeOH was chosen as the optimal solvent for the extraction of metabolites from human faecal samples in order to assess gut health [20]. Furthermore, MeOH has been found to be the optimal extraction solvent in a range of metabolomics studies, including the investigation of dietary influences in faecal samples [3], serum metabolite profiling [21], and *Blastocystis'* metabolism [22]. In comparison with one of the current most used extraction solvents, phosphate buffer saline (PBS) [23], the recognition of MeOH as an

efficient organic buffer and resultant choice in a range of sample preparation methods may be attributed to effective protein denaturation [24] and multi-polarity chemical capture [25].

Cell lysis is the process of breaking down the cell membrane to release contents contained inside the cell for molecular analysis. Bead beating, sonication, and cycles of freeze-thawing are common techniques used to disrupt the cell, and a sense of uncertainty resides about optimal methodological choice. The samples that underwent cell lysis using bead beating contained a significantly higher number of *m*/*z* features than freeze-thawing and a significantly increased number of identified metabolites than both those with sonication and freeze-thawing. Moreover, cell disruption by bead beating had a significantly increased number of lipids compared to both other methods. Overall, these findings indicate that bead beating was the most effective cell lysis method for extracting metabolites from human faecal samples. Additional studies have found analogous findings; for example, one study showed that bead beating was the best method for cell disruption and subsequent extraction of both polar and non-polar compounds from platelet samples, as given by optimal extraction efficiencies [26]. Bead beating has also previously been used as the cell lysis method of choice in the sample preparation of human faecal samples [27], as well as for gastrointestinal stromal tumour [28] and the characterisation of tissue samples [29].

Sample-to-solvent ratio, as aforementioned, is vital not only to maximise the data obtained, but also to ensure sufficient sample quality for LC-MS analysis so as not to cause blockage and instrument breakdown. This is particularly important for complex biomatrices such as faeces, which are composed of an abundance of organic and cellular material. The sample-to-solvent ratio, therefore, must allow extraction of large metabolite numbers that are compatible with LC-MS systems. Therefore, the metabolic output resulting from sample-to-solvent ratios of 1:5, 1:10, and 1:20 were assessed. Different metabolite quantification analyses identified a higher number of *m*/*z* features, identified metabolites, and signal intensity were given by samples using a sample-solvent ratio of 1:20 compared to the other tested ratios. Over 300 *m*/*z* features were detected and putatively identified using the optimal procedure with a 1:20 sample-solvent ratio, which holds great promise for maximising capture of biological information in future metabolomic studies. It is important to note that this work is part of an ongoing effort to document the metabolites putatively

identified in faecal samples, which will in future will be built upon by the creation of a standards library and the additional use of pure standards. Putative metabolite identification at MSI level 2 without the use of internal standards is, however, a current limitation of the present study, and the resulting lack of validation techniques must also be highlighted. Nonetheless, this work algins with the reporting standards of chemical analysis [30] and will be extended in future in order to increase the confidence of identification and validity of findings.

While contradictive reports are found regarding metabolite extraction procedures, it is important to bring to light methods that are suitable in specific contexts to continue the drive towards standardisation. The use of biphasic extraction protocols is common in metabolomics sample preparation; however, method advancement must also reflect amenability to study design. A considerable amount of research [29,31-33] suggests the importance of single-phase extraction procedures that can be used as simple, fast, and scalable alternatives to some of the more extensive approaches, giving impetus for investigating the optimal monophasic extraction protocol for human faecal samples. Rapid and easy-to-use methods can greatly simplify metabolite extraction and thereby improve scalability and application in large clinical studies. In this sense, single-phase methods are advantageous as the single layer can easily be removed, minimising the risk of sample loss and contamination [34,35]. This is of paramount importance for large studies as well as those with limited sample amounts. Moreover, the method developed in this study uses fewer toxic chemicals and can, therefore, be deemed as more friendly to both the operator and environment [36,37]. However, it must also be noted that while monophasic protocols provide simple and scalable extractions, consideration must also be given to the potential trade-off regarding metabolome coverage in comparison to biphasic methods. Improvements to the automation and scaling of extraction methods for large studies using monophasic methods should be conducted without significantly reducing the metabolome coverage. Extraction methods utilising biphasic partitioning are advantageous in their ability to separately recover polar and non-polar metabolites, ensuring coverage across the polarity scale. While contradictive reports have previously been noted regarding the comparative coverage of monophasic and biphasic approaches [13,37], recent research has provided evidence to suggest the differences in coverage between the two approaches may not be significant. For example, recent studies have demonstrated that single-step sample

preparation methods showed metabolome coverage and signal intensities equivalent to or greater than biphasic methods [33,38,39]. Careful consideration is required when implementing metabolite extraction methods to fit the specific study aims; however, in addressing the requirement for simple and rapid extraction methods for large-scale studies, it can be suggested that monophasic methods may be implemented as the best compromise for both scalability and coverage.

Finally, we demonstrated the applicability of the method on samples from patients with two forms of gastrointestinal disease involving metabolic and microbial perturbation, CD and CoD [40,41]. The developed method successfully detected and differentiated metabolic patterns of each group with a wide coverage. The method demonstrates strong cross-platform compatibility, with successful method application using two distinct analytical platforms, Orbitrap 240 LC-MS (ThermoFisher Scientific) and targeted triple-quadrupole (Shimadzu, Kyoto, Japan). This is valuable for future use of the method in laboratories using different technologies for metabolomic analysis.

In summary, the untargeted and targeted LC-MS analyses of different extraction factors provide insights into specific methods which give the strongest metabolic output. Optimised sample pre-treatment and extraction methods ultimately improve protocol efficiency while simultaneously enhancing the MS signal obtained [42]. Each small parameter change may cause a small increase in the efficiency of LC-MS characteristics and so when combined, the accumulated difference in the overall protocol can result in a large improvement to the number and coverage of metabolites detected (Supplementary Information Figure S9). Furthermore, reproducibility of the method and the instrument are increased by documenting and working towards method standardisation. As the results from this study bring together some of the parameters of faecal metabolite extraction in agreement with existing studies, this supports evidence of an optimised and reproducible protocol that can be applied in a vast array of research and clinical settings. Moreover, the method covers a wide range of metabolites of different physiochemical properties to increase the capture of biological compounds. As an extension, employing the method to patients with gastrointestinal disease expands the protocol applicability to different sample types. This method addresses the requirement for affordable, reproducible, and environmentally friendly metabolite extraction protocols. Thus, the method described build

on the foundations of protocol standardisation, allowing for improved comparisons of future metabolomics studies using faecal samples.

2.6 Conclusion

Based on a series of optimisation experiments, we describe a protocol to extract metabolites from faecal samples for metabolomic analysis using an LC-MS system. We recommend the use of 50 mg freeze-dried faecal samples in a 1000 μL MeOH and bead beating extraction, as given by a reproducible increased metabolite measurement. The optimised faecal extraction method described here can be used for metabolomics investigations of a wide array of applications, with strong evidence for its suitability in studies of gastrointestinal disease. This contributes towards standardising a framework of sample preparation, allowing for easier and more accurate comparisons between studies.

Funding: This study was supported by Shimadzu UK and the University of Strathclyde through joint contribution to P.E.K.'s studentship. The clinical studies were funded by the Glasgow Children Hospital Charity and the Nutricia Research Foundation. Z.R. want to acknowledge the EPSRC Multiscale Metrology Suite (EP/V028960/1).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of NHS West of Scotland Research Ethics Committee (Ref: 11/WS/0006) for the study in patients with coeliac disease and the Yorkhill Research Ethics Committee (Reb: 05/S0708/66) for the study in patients with Crohn's disease.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: N.J.W.R. want to acknowledge the Strathclyde Centre of Molecular Bioscience (www.scmb.strath.ac.uk) for access to LCMS instrumentation. We also acknowledge the help and advice given by the team at Shimadzu UK—Christine Heinz, Emily Armitage, Sean Devers, Chris Titman, Jonathan McGeehan and Stephen Brookes.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

2.7 References

1. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res*. **2022**, *50*, D622–D631. https://doi.org/10.1093/nar/gkab1062.
2. Zierer, J.; Jackson, M.A.; Kastenmüller, G.; Mangino, M.; Long, T.; Telenti, A.; Mohney, R.P.; Small, K.S.; Bell, J.T.; Steves, C.J.; et al. The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet*. **2018**, *50*, 790–795. https://doi.org/10.1038/s41588-018-0135-7.
3. Erben, V.; Poschet, G.; Schrotz-King, P.; Brenner, H. Evaluation of different stool extraction methods for metabolomics measurements in human faecal samples. *BMJ NPH* **2020**, *4*, 374–384. https://doi.org/10.1136/bmjnph-2020-000202.
4. Coker, O.O.; Liu, C.; Wu, W.K.K.; Wong, S.H.; Jia, W.; Sung, J.J.Y.; Yu, J. Altered gut metabolites and microbiota interactions are implicated in colorectal carcinogenesis and can be non-invasive diagnostic biomarkers. *Microbiome* **2022**, *10*, 1–12. https://doi.org/10.1186/s40168-021-01208-5.
5. Yu, D.; Du, J.; Pu, X.; Zheng, L.; Chen, S.; Wang, N.; Li, J.; Chen, S.; Pan, S.; Shen, B. The Gut Microbiome and Metabolites Are Altered and Interrelated in Patients with Rheumatoid Arthritis. *Front. Cell. Infect Microbiol*. **2022**, *11*, 763507. https://doi.org/10.3389/fcimb.2021.763507.
6. Fang, Q.; Liu, N.; Zheng, B.; Guo, F.; Zeng, X.; Huang, X.; Ouyang, D. Roles of Gut Microbial Metabolites in Diabetic Kidney Disease. *Front. Endocrinol*. **2021**, *12*, 636175. https://doi.org/10.3389/fendo.2021.636175.
7. Zheng, L.; Wen, X.L.; Duan, S.L. Role of metabolites derived from gut microbiota in inflammatory bowel disease. *World J. Clin. Cases* **2022**, *10*, 2660–2677. https://doi.org/10.12998/wjcc.v10.i9.2660.
8. Di Cagno, R.; Rizzello, C.G.; Gagliardi, F.; Ricciuti, P.; Ndagijimana, M.; Francavilla, R.; Guerzoni, ME.; Crecchio, C.; Gobbetti, M.; De Angelis, M. Different fecal microbiotas and volatile organic compounds in treated and untreated children with celiac disease. *Appl. Environ. Microbiol*. **2009**, *75*, 3963–3971. https://doi.org/10.1128/AEM.02793-08.
9. Metwaly, A.; Dunkel, A.; Waldschmitt, N.; Raj, A.C.D.; Lagkouvardos, I.; Corraliza, A.M.; Mayorgas, A.; Martinez-Medina, M.; Reiter, S.; Schloter, M.; et al. Integrated

microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. *Nat. Commun.* **2020**, *11*, 4322. https://doi.org/10.1038/s41467-020-17956-1.

10. Heinken, A.; Hertel, J.; Thiele, I. Metabolic modelling reveals broad changes in gut microbial metabolism in inflammatory bowel disease patients with dysbiosis. *npj Syst. Biol. Appl.* **2021**, *7*, 19. https://doi.org/10.1038/s41540-021-00178-6.
11. Khalkhal, E.; Rezaei-Tavirani, M.; Fathi, F.; Nobakht, M.; Gh, BF.; Taherkhani, A.; Rostami-Nejad, M.; Asri, N.; Haidari, M.H. Screening of Altered Metabolites and Metabolic Pathways in Celiac Disease Using NMR Spectroscopy. *Biomed. Res. Int*. **2021**, 1798783. https://doi.org/10.1155/2021/1798783.
12. Martín-Masot, R.; Mota-Martorell, N.; Jové, M.; Maldonado, J.; Pamplona, R.; Nestares, T. Alterations in one-carbon metabolism in celiac disease. *Nutrients* **2020**, *12*, 1–14. https://doi.org/10.3390/nu12123723.
13. Hosseinkhani, F.; Dubbelman, A.C.; Karu, N.; Harms, A.C.; Hankemeier, T. Towards standards for human fecal sample preparation in targeted and untargeted lc-hrms studies. *Metabolites* **2021**, *11*, 364. https://doi.org/10.3390/metabo11060364.
14. Moosmang, S.; Pitscheider, M.; Sturm, S.; Seger, C.; Tilg, H.; Halabalaki, M.; Stuppner, H. Metabolomic analysis—Addressing NMR and LC-MS related problems in human feces sample preparation. *Clin. Chim. Acta* **2019**, *489*, 169–176. https://doi.org/10.1016/j.cca.2017.10.029.
15. Nandania, J.; Peddinti, G.; Pessia, A.; Kokkonen, M.; Velagapudi, V. Validation and automation of a high-throughput multitargeted method for semiquantification of endogenous metabolites from different biological matrices using tandem mass spectrometry. *Metabolites* **2018**, *8*, 44. https://doi.org/10.3390/metabo8030044.
16. Berardi, D.; Hunter, Y.; van den Driest, L.; Farrell, G.; Rattray, N.J.W.; Rattray, Z. The Differential Metabolic Signature of Breast Cancer Cellular Response to Olaparib Treatment. *Cancers* **2022**, *14*, 3661. https://doi.org/10.3390/cancers14153661.
17. Wu, J.; An, Y.; Yao, J.; Wang, Y.; Tang, H. An optimised sample preparation method for NMR-based faecal metabonomic analysis. *Analyst* **2010**, *135*, 1023–1030. https://doi.org/10.1039/b927543f.
18. Gholib, G.; Heistermann, M.; Agil, M.; Supriatna, I.; Purwantara, B.; Nugraha, T.P.; Engelhardt, A. Comparison of fecal preservation and extraction methods for steroid hormone metabolite analysis in wild crested macaques. *Primates* **2018**, *59*, 281–292. https://doi.org/10.1007/s10329-018-0653-z.

19. Zhao, X.; Zhang, Z.; Hu, B.; Huang, W.; Yuan, C.; Zou, L. Response of gut microbiota to metabolite changes induced by endurance exercise. *Front. Micro*. **2018**, *9*, 765. https://doi.org/10.3389/fmicb.2018.00765.
20. De Zawadzki, A.; Thiele, M.; Suvitaival, T.; Wretlind, A.; Kim, M.; Ali, M.; Bjerre, A. F.; Stahr, K.; Mattila, I.; Hansen, T.; et al. High-Throughput UHPLC-MS to Screen Metabolites in Feces for Gut Metabolic Health. *Metabolites* **2022**, *12*, 211. https://doi.org/10.3390/metabo12030211.
21. Want, E.J.; O'Maille, G.; Smith, C.A.; Brandon, T.R.; Uritboonthai, W.; Qin, C.; Trauger, S.A.; Siuzdak, G. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem*. **2006**, *78*, 743–752. https://doi.org/10.1021/ac051312t.
22. Newton, J.M.; Betts, E.L.; Yiangou, L.; Roldan, J.O.; Tsaousis, A.D.; Thompson, G.S. Establishing a metabolite extraction method to study the metabolome of blastocystis using nmr. *Molecules* **2011**, *26*, 3285. https://doi.org/10.3390/molecules26113285.
23. Deda, O.; Gika, H.G.; Wilson, I.D.; Theodoridis, G.A. An overview of fecal sample preparation for global metabolic profiling. *J. Pharm. Biomedical.* **2015**, *113*, 137–150. https://doi.org/10.1016/j.jpba.2015.02.006.
24. Zeng, M.; Cao, H. Fast quantification of short chain fatty acids and ketone bodies by liquid chromatography-tandem massspectrometry after facile derivatization coupled with liquid-liquid extraction. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci*. **2018**, *1083*, 137–145. https://doi.org/10.1016/j.jchromb.2018.02.040.
25. Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-massspectrometry. *Anal. Bioanal. Chem*. **2012**, *403*, 1523–1548. https://doi.org/10.1007/s00216-012-6039-y.
26. Fu, X.; Calderón, C.; Harm, T.; Gawaz, M.; Lämmerhofer, M. Advanced unified monophasic lipid extraction protocol with wide coverage on the polarity scale optimized for large-scale untargeted clinical lipidomics analysis of platelets. *Anal. Chim. Acta* **2022**, *1221*, 340155. https://doi.org/10.1016/j.aca.2022.340155.
27. Cheng, K.; Brunius, C.; Fristedt, R.; Landberg, R. An LC-QToF MS based method for untargeted metabolomics of human fecal samples. *Metabolomics* **2020**, *16*, 1–8. https://doi.org/10.1007/s11306-020-01669-z.
28. Macioszek, S.; Dudzik, D.; Jacyna, J.; Wozniak, A.; Schöffski, P.; Markuszewski, M.J. A robust method for sample preparation of gastrointestinal stromal tumour for LC/MS

untargeted metabolomics. *Metabolites* **2021**, *11*, 554. https://doi.org/10.3390/metabo11080554.

29. Southam, A.D.; Pursell, H.; Frigerio, G.; Jankevics, A.; Weber, R.J.M.; Dunn, W.B. Characterization of Monophasic Solvent-Based Tissue Extractions for the Detection of Polar Metabolites and Lipids Applying Ultrahigh-Performance Liquid Chromatography-Mass Spectrometry Clinical Metabolic Phenotyping Assays. *J. Proteome Res*. **2021**, *20*, 831–840. https://doi.org/10.1021/acs.jproteome.0c00660.
30. 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.; et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221. https://doi.org/10.1007/s11306-007-0082-2.
31. Medina, J; van der Velpen, V.; Teav, T.; Guitton, Y.; Gallart-Ayala, H.; Ivanisevic, J. Single-step extraction coupled with targeted hilic-ms/ms approach for comprehensive analysis of human plasma lipidome and polar metabolome. *Metabolites* **2020**, *10*, 1–17. https://doi.org/10.3390/metabo10120495.
32. Alshehry, Z.H.; Barlow, C.K.; Weir, J.M.; Zhou, Y.; McConville, M.J.; Meikle, P.J. An efficient single phase method for the extraction of plasma lipids. *Metabolites* **2015**, *5*, 389–403. https://doi.org/10.3390/metabo5020389.
33. Peterson, A.L.; Walker, A.K.; Sloan, E.K.; Creek, D.J. Optimized Method for Untargeted Metabolomics Analysis of MDA-MB-231 Breast Cancer Cells. *Metabolites* **2016**, *6*, 30. https://doi.org/10.3390/metabo6040030.
34. Lydic, T.A.; Busik, J.v; Reid, G.E. A monophasic extraction strategy for the simultaneous lipidome analysis of polar and nonpolar retina lipids. *J. Lipid Res.* **2014**, *55*, 1797–1809. https://doi.org/10.1194/jlr.D050302.
35. Merrill, A.H.; Sullards, M.C.; Allegood, J.C.; Kelly, S.; Wang, E. Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* **2005**, *36*, 207–224. https://doi.org/10.1016/j.ymeth.2005.01.009.
36. Southam, A.D.; Haglington, L.D.; Najdekr, L.; Jankevics, A.; Weber, R.J.M.; Dunn, W.B. Assessment of human plasma and urine sample preparation for reproducible and high-throughput UHPLC-MS clinical metabolic phenotyping. *Analyst* **2020**, *145*, 6511–6523. https://doi.org/10.1039/D0AN01319f.

37. Joshi, D.R.; Adhikari, N. An Overview on Common Organic Solvents and Their Toxicity. *J. Pharm. Res. Int.* **2019**, *28*, 1–18. https://doi.org/10.9734/jpri/2019/v28i330203.
38. Wong, M.W.K.; Braidy, N.; Pickford, R.; Sachdev, P. S.; Poljak, A. Comparison of single phase and biphasic extraction protocols for lipidomic studies using human plasma. *Front. Neurol.* **2019**, *10*, 879. https://doi.org/10.3389/fneur.2019.00879.
39. Calderón, C.; Sanwald, C.; Schlotterbeck, J.; Drotleff, B.; Lämmerhofer, M. Comparison of simple monophasic versus classical biphasic extraction protocols for comprehensive UHPLC-MS/MS lipidomic analysis of Hela cells. *Anal. Chim. Acta* **2019**, *1048*, 66–74. https://doi.org/10.1016/j.aca.2018.10.035.
40. Zafeiropoulou, K.; Nichols, B.; Mackinder, M.; Biskou, O.; Rizou, E.; Karanikolou, A.; Clark, C.; Buchanan, E.; Cardigan, T.; Duncan, H.; et al. Alterations in Intestinal Microbiota of Children with Celiac Disease at the Time of Diagnosis and on a Gluten-free Diet. *Gastroenterology* **2020**, *159*, 2039–2051. e20. https://doi.org/10.1053/j.gastro.2020.08.007.
41. Quince, C.; Ijaz, U.Z.; Loman, N.; Eren, A.M.; Saulnier, D.; Russell, J.; Haig, S.; Calus, S.; Quick, J.; Barclay, A.; et al. Extensive Modulation of the Fecal Metagenome in Children with Crohn's Disease During Exclusive Enteral Nutrition. *Am. J. Gastroenterol*. **2015**, *110*, 1718–1729. https://doi.org/10.1038/ajg.2015.357.
42. Zheng, X.; Yu, J.; Cairns, T.C.; Zhang, L.; Zhang, Z.; Zhang, Q.; Zheng, P.; Sun, J.; Ma, Y. Comprehensive Improvement of Sample Preparation Methodologies Facilitates Dynamic Metabolomics of Aspergillus niger. *Biotechnol. J.* **2019**, *14*, e1800315. https://doi.org/10.1002/biot.201800315.

3.0 CHAPTER 3

Metabolic Dysregulation Driven by Coeliac Disease is Ameliorated by a Gluten Free Diet

Patricia Kelly[1], Gillian Farrell[1], Richard K. Russell[2], Richard Hansen[3], Paraic McGrogan[4], Christine Edwards[5], David Broadhurst[6], Konstantinos Gerasimidis[5], Nicholas J.W. Rattray[1,7].

To be submitted to Clinical Nutrition

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G4 0RE, UK.

[2] Royal Hospital for Children and Young People, 50 Little France Crescent, Edinburgh EH16 4TJ, UK.

[3] Royal Hospital for Children, 1345 Govan Road, Glasgow G52 4TF, UK.

[4] Paediatric Gastroenterology Department, The Royal Hospital for Children, Glasgow, UK.

[5] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow G12 8QQ, UK.

[6] Centre for Integrative Metabolomics & Computational Biology, Edith Cowan University, Joondalup, Australia.

[7] NHS Lanarkshire, UK.

Author Contributions: Conceptualization, P.E.K., G.F., R.K.R., R.H K.G., P.M., C.E., and N.J.W.R.; methodology, P.E.K., S.M., K.G. and N.J.W.R.; software, P.E.K., G.F. and N.J.W.R.; validation, P.E.K., G.F. and N.J.W.R.; formal analysis, P.E.K., H.J.N., K.G. and N.J.W.R.; investigation, P.E.K., and N.J.W.R.; resources, R.K.R., R.H. and K.G.; data curation, P.E.K., D.B., G.F. and N.J.W.R.; writing—original draft preparation, P.E.K. and N.J.W.R.; writing—review and editing, P.E.K., H.J.N., R.K.R., R.H., P.M., C.E., K.G. and N.J.W.R.; visualisation, P.E.K. and N.J.W.R.; supervision, N.J.W.R., and K.G.; project administration, N.J.W.R.; funding acquisition, N.J.W.R and K.G. All authors have read and agreed to the published version of the manuscript.

3.1 Abstract

Coeliac disease (CoD) is an autoimmune disorder triggered by gluten ingestion, and strict adherence to a gluten-free diet (GFD) remains the only effective treatment. To investigate the metabolic impact of CoD and dietary intervention, we characterised the faecal metabolome of children with untreated CoD (UCD), treated CoD on a GFD (TCD), healthy controls (HC), and unaffected siblings using untargeted LC-MS. Across 143 participants, 1,749 metabolite features were detected, and multivariate analysis revealed distinct clustering between groups. Comparison of UCD and HC identified 58 significantly altered metabolites, including elevations in bile acid derivatives, acylcarnitines, and amino acid dipeptides. Treatment with a GFD led to partial restoration of the metabolome, with 27 treatment-responsive metabolites altered between UCD and TCD, primarily amino acid dipeptides and oligopeptides, alongside changes in purine and phenolic metabolites. However, several metabolite classes such as bile acids, sterols, and microbial-derived metabolites remained persistently altered in both UCD and TCD compared with HC, indicating a core CoD-specific signature independent of diet. Additional diet-driven changes were identified in TCD patients, including shifts in amino acid derivatives and purine metabolism, reflecting secondary effects of dietary exclusion. Collectively, our findings demonstrate that paediatric CoD is characterised by distinct metabolic signatures across CoD-specific, treatment-responsive, and diet-driven metabolite groups. These insights highlight the potential of faecal metabolomics as a non-invasive tool for biomarker discovery, dietary monitoring, and identification of patients at risk of non-responsive disease.

3.2 Introduction

Coeliac disease (CoD) is a multifactorial autoimmune disorder of the small intestine which is caused by ingestion of gluten in genetically predisposed individuals. Disease incidence has increased in recent years, with 1.4% of the global population currently estimated to be living with the condition [1]. CoD results in small intestine villi atrophy and inflammation causing nutrient malabsorption abdominal pain, diarrhoea, bloating, and weight loss.

CoD can be challenging to diagnose due to symptom overlap, particularly in patients with silent disease phenotype who do not have overt gastrointestinal symptoms. A combination of diagnostic tools is therefore required along with evidence of clinical manifestations to obtain an accurate diagnosis. Firstly, serological tests are used to measure levels of IgA anti-tissue transglutaminase antibodies (tTG) and anti-endomysial antibodies (EMA), which are produced in response to gluten-activated immune pathways. Although these measurements provide an insight into the inflammatory immune response, their measures are not specific to CoD. Therefore, if positive blood tests are obtained, endoscopic evaluation is required to examine small intestinal damage, with disease presence indicated by duodenal villous atrophy and crypt hyperplasia [2]. To confirm diagnosis, small intestinal mucosal biopsies are taken for histopathological assessment, where application of a classification system, such as the Marsh classification system, are then used to characterise disease, with stages ranging from normal mucosa to severe villous atrophy. As per the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) guidelines, a biopsy is required for a confirmed CoD diagnosis in adults, however in children, a biopsy may be skipped if all three of the following criteria are met: (1) the tTG-IgA levels are greater than 10 times the upper limit of normal, (2) a second antibody test such as EMA-IgA is positive, and (3) symptoms consistent with CoD are presented [3]. If any one of these criteria are missing, a small intestinal biopsy is also required to confirm a CoD diagnosis.

The only effective treatment at present is the strict adherence to a gluten-free diet (GFD) which can often be challenging to adhere to. The currently limited therapeutic options highlight the need for an improved understanding of disease pathogenesis. Identification of novel disease biomarkers and mechanistic insights will help to improve patient outcomes

through earlier diagnoses, improved disease and treatment monitoring, and prediction of complications.

It is now recognised with increasing evidence that the gut microbiota plays a crucial role in CoD pathophysiology [4, 5]. A recent study observed distinct changes in the gut microbiota in children with CoD compared to healthy children [6]. Most notably, a microbial signature specific to CoD was identified comprising eleven operational taxonomic units, with a distinct altered microbial composition found following a GFD. Furthermore, the predictive ability of the gut microbiota in the diagnosis of CoD has been revealed. Recent studies have shown that the gut microbiota may serve as non-invasive diagnostic or early predictive biomarkers for CoD, achieving up to 82% accuracy [7] and identifying microbial and metabolite shifts 18 months before disease onset in at-risk infants [8, 9]. In line with evidence supporting an altered microbiome as part of CoD pathophysiology, study of the metabolome can be utilised to facilitate disease understanding and interactions at the level of metabolism. Through the detection and measurement of small molecules, metabolomics is a valuable tool to enable understanding of the underlying molecular mechanisms governing the onset and progression of disease. Progress has been made to establish the molecular effects of CoD, with metabolomics studies beginning to characterise the CoD metabolome [10-12]. While there is clear evidence for disease-associated metabolic alterations, the determination of a metabolic fingerprint for CoD is still in the early phases. Liquid chromatography mass spectrometry (LC-MS) approaches are beneficial for the analysis of metabolism due to their high sensitivity and selectivity for detecting molecules in a range of applications. Importantly, stool samples are an advantageous non-invasive biomatrix for studying gastrointestinal disease, providing a window into gut microbial and metabolic function. Using faecal metabolomics to investigate disease therefore has the potential to unravel complex relationships between microbial metabolism, the immune response, and the diet, three major pillars of CoD. This will help not only in the understanding of disease mechanisms and dietary impact but also holds promise in complementary diagnostic and prognostic tools through biomarker development.

The aim of the present study is to characterise the faecal metabolome of children with CoD. The utilisation of both cross-sectional and prospective cohorts allowed an in-depth analysis of the metabolic impact of disease and treatment. Comparisons were performed between

untreated CoD patients at the point of disease diagnosis (UCD), patients with established CoD and on a GFD (TCD), healthy siblings of TCD patients, and healthy controls (HC). UCD patients were examined at the time of diagnosis and additionally at 6- and 12- month follow-up time points after treatment with a GFD. Analysis was further extended to utilise faecal gluten immunogenic peptide (GIP), a sensitive and specific biomarker of GFD compliance. This study provides the first combined cross-sectional and prospective cohort analysis of the faecal metabolome of CoD patients using untargeted LC-MS.

3.3 Materials and Methods

3.3.1 Ethics Statement

The study was approved by the West of Scotland Research Ethics Committee (reference no.11/WS/0006). All participants and their carers provided written consent.

3.3.2 Subjects

Faecal samples were collected from children with CoD receiving care at the Royal Hospital for Children in Glasgow for metabolomics analysis. For participant selection, children with a confirmed diagnosis of CoD were recruited from annual clinic appointments, while newly diagnosed CoD cases were referred by primary healthcare services. For the control group, healthy children displaying no clinical symptoms and who tested negative for tissue transglutaminase IgA were recruited through advertisement methods, and unaffected siblings of children diagnosed with CoD were also recruited in the same way. Tissue transglutaminase IgA antibody (tTG) was measured in UCD children at the time of their diagnosis and at follow up time points upon the initiation of a GFD and in TCD children during their clinic visit. In both groups, a healthy status denoted children who did not have regular medical consultations, were not on routine medication, and had no history of GI disorders. Inclusion was extended to all eligible children, except those who met exclusion criteria such as recent antibiotic use, regular prebiotic/ probiotic use, or presence of other comorbidities. For this piece of work, no additional tests beyond metabolomics analyses were undertaken.

Confirmation of a CoD diagnosis was given by small bowel biopsy, following the UK guidelines that were effective at the time of recruitment [13]. Assessment of GI symptoms was carried out using the PedsQL-GS questionnaire (version 1) [14], a symptom measurement scale in which a higher the PedsQL-GS score is associated with lower levels of GI symptoms. Compliance to GFD was assessed by measuring faecal GIP levels (iVYLISA; Biomedal, Seville, Spain) [15].

3.3.3 Sample Collection and Storage

Complete faecal samples were collected and stored under cold anaerobic conditions (Anaerocult A; Merck, Darmstadt, Germany) within 2 hours. Samples were homogenized and stored at -80 °C until further processing. A single faecal sample was obtained from all groups except the UCD cohort, from which three samples were collected: a baseline sample prior to diagnostic endoscopy while patients were consuming a gluten-containing diet, and follow-up samples at 6 and 12 months after commencing a gluten-free diet (GFD).

3.3.4 Chemicals and Reagents

LC-MS grade methanol (MeOH), acetonitrile (ACN) and water (H_2O) were purchased from Fisher Scientific (Geel, Belgium). LC-MS grade formic acid (FA) was purchased from Thermo Scientific (Czech Republic).

3.3.5 Faecal Metabolite Extraction Protocol

LC-MS grade methanol, acetonitrile and water were purchased from Fisher Scientific (Geel, Belgium). LC-MS grade formic acid was purchased from Thermo Scientific (Czech Republic). Metabolite extraction was performed using a previously optimised method for LC-MS analysis of faecal samples [16]. In brief, freeze-dried faecal samples were added to methanol and cells were lysed using bead beating at 5 ms^{-1} for 60s (FastPrep 24 MP Biomedicals). Samples were then centrifuged at 13,000 g for 15 minutes and the supernatant recovered. Samples were dried using a SpeedVac system and stored at -80 °C until further processing. Reconstitution was performed in 250 µL 50/50 H_2O: acetonitrile, vortexed for 1 minute and centrifuged at 15,000 g for 15 minutes, and aliquots transferred into glass vials for MS analysis.

3.3.6 Untargeted LC-MS Metabolite Measurement

Untargeted metabolomics analysis was performed on a binary Vanquish ultra-high performance liquid chromatography UHPLC system (ThermoFisher Scientific, Bremen, Germany) coupled to an Orbitrap Exploris 240 (ThermoFisher Scientific, Bremen, Germany) orbitrap based mass spectrometer. Chromatographic separation was performed on a Vanquish Accucore C18 UHPLC analytical column (ThermoScientific, 100 mm x 2.1 mm, 2.6 µM) at a flow rate of 400 µL min $^{-1}$. Mobile phase A was composed of 99.9% water + 0.1% formic acid and mobile B was composed of 99.9% MeOH + 0.1% formic acid. Electrospray ionization was used as the ionization method, set at 3900 V.

Quality control samples were prepared by pooling samples across all groups undergoing analysis. Solvent blanks and quality control samples were entered at the beginning of every analytical run and after every five samples in each batch over the course of the study to assess background in the system and detect potential contaminations. Instrument performance was assessed using ^{13}C-labelled L-glutamine as an internal standard. 400 ng/mL ^{13}C-glutamine was spiked into each sample and the signal stability, and any retention time drift was measured.

3.3.7 Targeted LC-MS Metabolite Measurement

Targeted metabolomic analysis was performed on a UHPLC system coupled to a Shimadzu 8060NX triple quadrupole mass spectrometer (Shimadzu Corp, Kyoto, Japan). The method used for metabolite detection and quantification was provided by the vendor; Primary Metabolites LC/MS/MS Method Package version 2.0 (Shimadzu Corp, Kyoto, Japan). Chromatographic separation was performed on a pentafluorophenylpropyl (PFPP) + UHPLC analytical column (Merck, 150 mm × 2.1 mm, 3 µM) at a flow rate of 400 µL min $^{-1}$. Mobile phase A was composed of 99.9% water + 0.1% formic acid and mobile B was composed of 99.9% acetonitrile + 0.1% formic acid, with ESI used as the ionisation method.

3.3.8 Mass Spectrometry Data Processing

For the processing of untargeted metabolomics data, Thermo Scientific Xcalibur format raw data files (.RAW) were uploaded and processed using Compound Discoverer 3.2. Data was filtered by removing duplicate compound detections and setting the mass deviation to

within 5 ppm. A data matrix of identified metabolites and associated peak areas was constructed.

3.3.9 Batch Alignment

Samples were run in two batches and aligned using Quality Control-based Metabolite eXpression Preprocessing (QC:MXP) software version 1.1.0 hosted in Matlab [17]. Metabolites with more than 20% missing values were excluded. Configuration settings were as follows: {"LogTransform": true, "RemoveZeros": true, "OutlierScope": "Local", "OutlierMethod": "Linear", "OutlierCI": 0.95, "OutlierPostHoc": "MPV", "QCRSCmode": "Spline", "QCRSCgammaRange": "1:0.5:4", "QCRSCcvMethod": "5-Fold", "QCRSCmcReps": 5, "QCRSCtype": "Subtract", "BlankRatioMethod": "QC", "RelativeLOD": 1.5, "StatsParametric": true, "ParallelProcess": true}. Missing values were imputed using k-Nearest Neighbours (KNN).

3.3.10 Tissue Transglutaminase (tTG) Antibody Measurement

Serological measurement of tissue transglutaminase (tTG) antibodies was performed as part of routine clinical care. Samples were analysed by the NHS Greater Glasgow and Clyde (GGC) Diagnostic Laboratory, Immunology Department, using their standard accredited protocol, conducted according to established laboratory procedures and quality control

3.3.11 Data and Statistical Analysis

Identification of metabolite changes were based on univariate and multivariant analysis. Data and statistical analysis were performed primarily using MetaboAnalyst 5.0. Orthogonal partial least squares discriminant analysis (OPLS-DA) was used to reveal significant differences between the experimental groups and understand clustering patterns. Method validity was assessed using the goodness of fit and predictive ability of the OPLS-DA models evaluated by R^2 and Q^2 values, respectively, with a good quality of model defined by parameter values greater than 0.5. Differential analysis using volcano plots allowed significant metabolite differences between groups to be determined. Univariate statistical analyses were additionally performed using unpaired t-tests and one-way ANOVA, with significance determined by false discovery rate (FDR)-adjusted p-values < 0.05 to account for multiple testing.

3.3.12 Receiver Operating Characteristic (ROC) Analysis

Receiver Operating Characteristic (ROC) analyses were performed to assess the discriminatory performance of selected metabolites using the pROC R package. To evaluate the combined predictive power of the metabolite panel, least absolute shrinkage and selection operator (LASSO) logistic regression was applied with stratified 5-fold cross-validation using the glmnet package.

3.3.13 Metabolomics Pathway Analysis

Pathway analysis was performed in MetaboAnalyst 5.0 by matching identified metabolites to the KEGG database ($p < 0.05$).

3.3.14 Overall Experimental Workflow

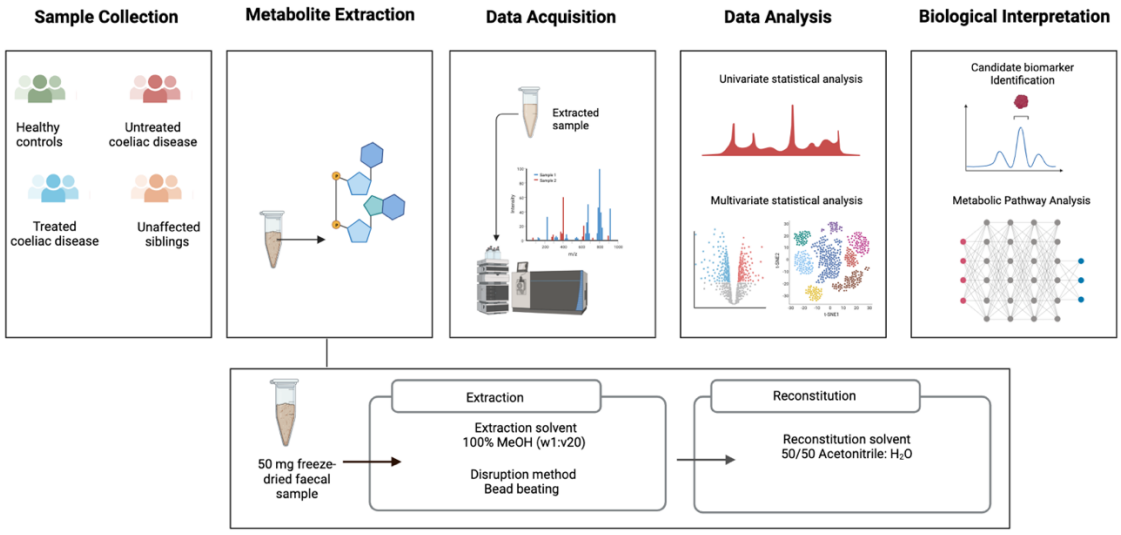


Figure 1. Overall experimental workflow following stages of sample collection, metabolite extraction using optimised method, data acquisition and analysis, and biological interpretation.

3.4. Results

3.4.1 Demographics and Clinical Parameters

This study investigated a total of 143 participants aged between two and fifteen, including 56 HCs, 27 UCD children, 40 TCD children on a GFD, and 20 unaffected siblings of TCD children (**Table 1**). Patients in the TCD group were children with pre-established disease, with a mean time since diagnosis of 4.6 years, and recommended to follow a GFD as a treatment strategy. There were no significant differences in the mean age, gender, weight, and BMI among the study groups ($p < 0.05$). UCD children had a higher incidence of gastrointestinal issues in comparison to both the TCD group and HCs. This aligns with results from the Paediatric Quality of Life Inventory Gastrointestinal Symptoms (PedsQL-GS), showing that the TCD group additionally presented a lower PedsQL-GS score than HCs. As expected, tTG levels were shown to decrease throughout the duration of a GFD. In the prospective cohort of new-onset CoD patients, analysis of the clinical parameters revealed that tTG levels decreased with GFD treatment, paralleled by an increase in PedsQL-GS score.

Table 1. Patient demographics of cross-sectional and prospective cohorts.

Cross-Sectional Study					Prospective Study		
Variable	HC n = 56	Siblings n = 20	UCD group n = 27	TCD group n = 40	At diagnosis n = 12	GFD 6 months n = 12	GFD 12 months n = 12
Age (y)	7.8 (0.41)	9.3 (0.62)	9.6 (0.60)	9.2 (0.48)	9.3 (0.92)	9.9 (0.92)	10.5 (0.91)
Gender (F/M)	30/26	10/10	14/13	23/17	6/6	6/6	6/6
Weight z-score	0.14 (0.15)	0.34 (0.26)	-0.19 (0.27)	0.15 (0.18)	-0.30 (0.36)	-0.35 (0.33)	-0.29 (0.32)
Height z-score	0.27 (0.15)	0.55 (0.28)	-0.18 (0.20)	-0.05 (0.16)	-0.02 (0.27)	-0.02 (0.27)	-0.06 (0.26)
BMI (kg/m^2)	16.8 (0.46)	17.3 (0.68)	17.2 (0.75)	17.5 (0.41)	16.3 (0.69)	16.4 (0.72)	16.8 (0.79)

BMI z-score	0.05 (0.14)	0.08 (0.24)	-0.15 (0.25)	0.26 (0.17)	-0.46 (0.36)	-0.53 (0.35)	-0.41 (0.35)
tTG (U/mL)	_	_	75.3 (11.21)	11.7 (4.41)	105.7 (14.37)	23.9 (11.97)	7.7 (2.08)
PedsQL-GS score	100.0 (2.32)	_	56.3 (7.28)	77.7 (3.95)	51.2 (6.60)	65.97 (5.66)	71.2 (8.73)

Values are mean (SEM) unless otherwise stated. HC, healthy control; UCD, Untreated coeliac disease, TCD, Treated coeliac disease.

3.4.2 Faecal Metabolome Profiling

In total, 1749 unique m/z features were detected through untargeted LC-MS analysis at MSI level 2. To explore overall metabolic variation across study groups, we performed Principal Coordinates–Canonical Variate Analysis (PC-CVA) on the faecal metabolomics dataset. In the cross-sectional cohort, this approach revealed clear group-level clustering between healthy controls (HC), untreated coeliac disease (UCD), and treated coeliac disease (TCD) individuals (**Figure 2A**). In the prospective cohort, global metabolite changes across the treatment period were observed, with PC-CVA plot illustrating metabolic trajectories over 0, 6, and 12 months of treatment with a gluten-free diet (**Figure 2B**). PC-CVA model showing metabolite trajectories across treatment duration, comparing UCD patients at time of diagnosis and after following a GFD for 6 months (n = 40) and 12 months (n = 42).

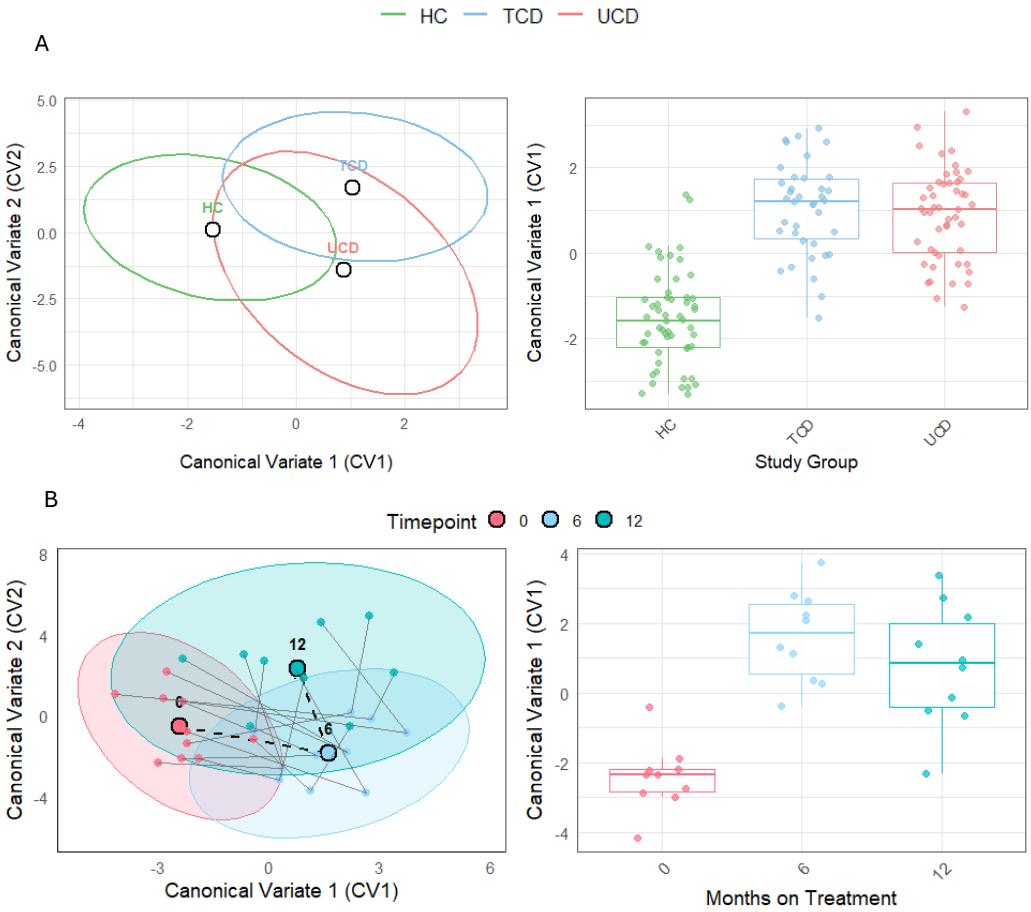


Figure 2. Metabolome profile for the cross-sectional (A) and prospective (B) cohorts. Principal Coordinates–Canonical Variate Analysis (PC-CVA) of the faecal metabolome across study groups (TCD, UCD, and HC) and throughout treatment timepoints (before GFD, 6 months on GFD, and 12 months on GFD) show group centroids and 95% confidence ellipses based on canonical variates. Boxplots show comparison of the CV1 values between groups.

In contrast to the widespread differences observed in the untargeted analysis, the targeted profiling revealed limited changes (**Figure 3**). A more detailed view of amino acid quantification is provided in **Supplementary Information Table S1**, which illustrates the largely comparable levels of most amino acids despite subtle fluctuations at different disease and treatment strategies.

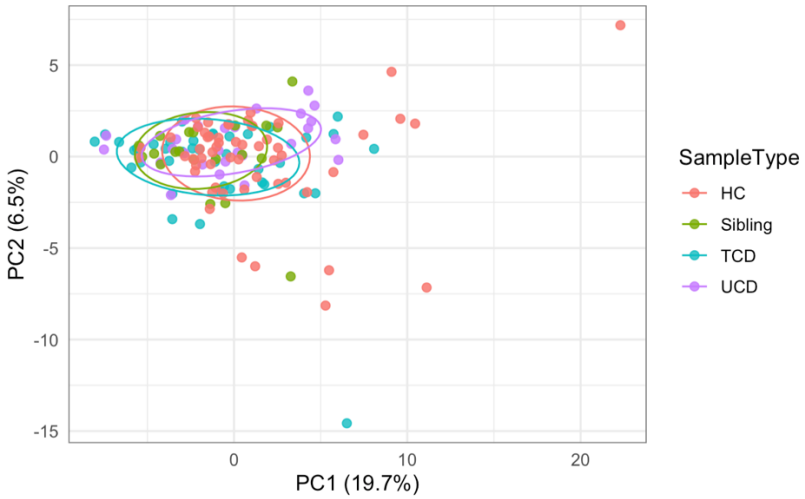


Figure 3. Principal Component Analysis (PCA) of targeted amino acid profiles across study groups. Each point represents an individual sample, and ellipses indicate the 95% confidence interval for each group.

3.4.3 Differential Analysis in Metabolome between Untreated CoD and HCs

Multivariate analysis using orthogonal partial least squares discriminant analysis (OPLS-DA) revealed partial separation of the global metabolome between UCD patients and HCs (R^2Y= 0.555, Q^2= 0.267; **Figure 4A**), identifying key metabolites driving class discrimination based on their variable importance in projection (VIP) scores (**Supplementary Information Figure S1**). Subsequent univariate analysis identified 58 metabolites to be significantly altered between the two groups, with 31 metabolites increased and 27 decreased relative to controls (**Figure 4B**). The top differentiating metabolites are illustrated in Figure 3C through representative boxplots (**Figure 4C**).

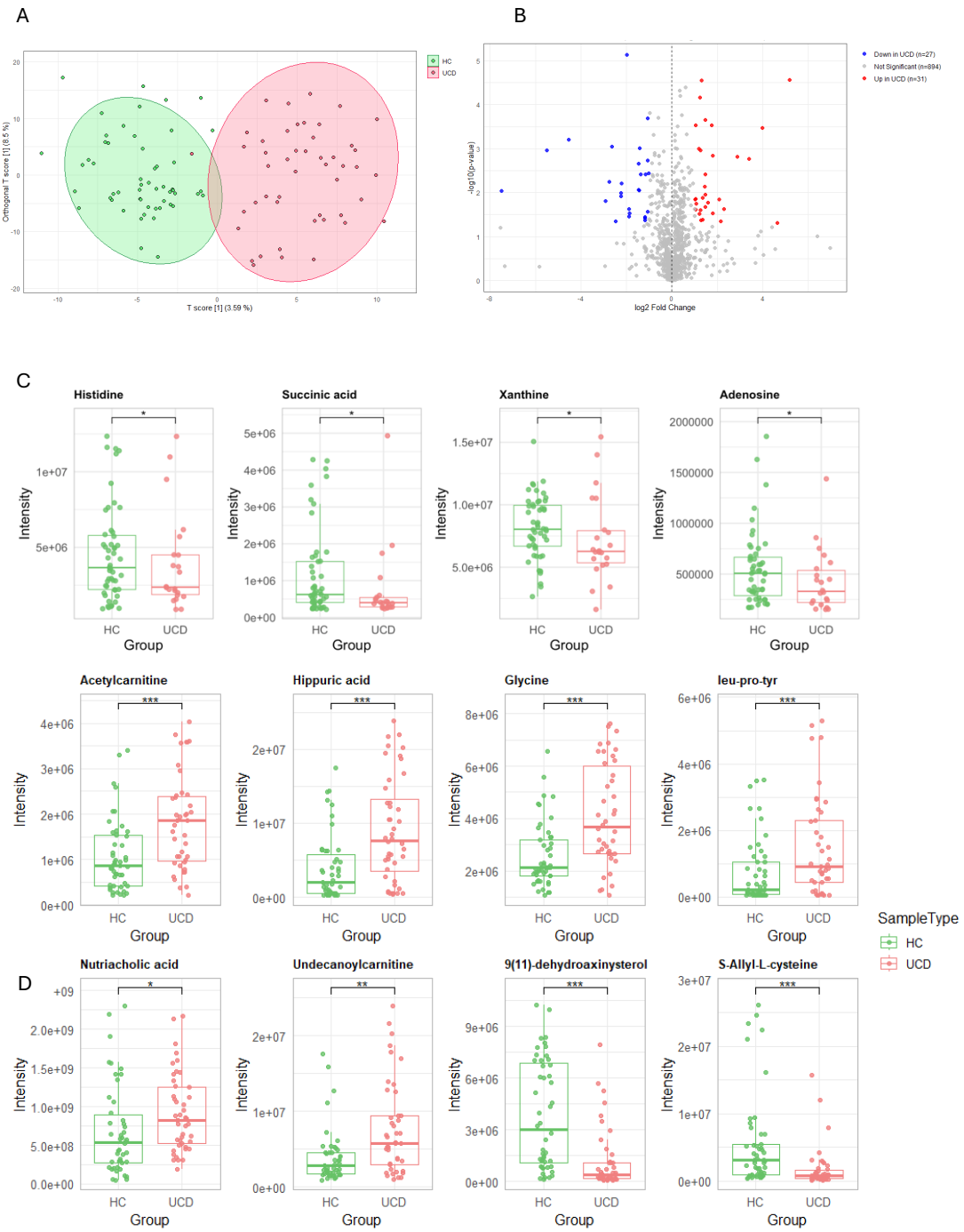


Figure 4. Metabolome profile of children with newly diagnosed coeliac disease compared to healthy controls (n = 82). (A) Scores plot of the orthogonal partial least square discriminant analysis (OPLS-DA) model with R^2Y= 0.555, Q^2= 0.267. (B) Volcano plot of significantly differential faecal metabolites comparing children with newly diagnosed coeliac disease compared to healthy controls, $p < 0.05$, fold change = 2. Box and whisker plots of the top significantly differential faecal metabolites from (C) the untargeted analysis and (D) the targeted analysis.

3.4.4 Differential Analysis in Metabolome Between Patients on Recommendation to a GFD with Untreated CoD or HCs

To investigate the impact that a GFD has on the faecal metabolome, we compared metabolite profiles between TCD patients with both UCD patients and HC groups. We first explored the effects of a GFD on the metabolome of CoD patients, with VIP scores from the OPLS-DA model, highlighting the metabolites that most strongly contributed to distinguishing TCD from UCD patients (**Supplementary Figure S2**). This was complemented by pairwise analysis identifying 27 significantly altered metabolites (11 decreased and 22 increased), relative to UCD. The metabolites which are altered only between TCD and UCD groups, reflecting the metabolic changes upon a GFD in individuals with CoD, are shown in **Figure 5**.

Next, metabolite profiles were compared between TCD patients and HCs. The corresponding VIP plot identified metabolites contributing to this difference, with univariate analysis identifying 24 metabolites altered between the two groups (OPLS-DA and VIP plot shown in **Supplementary Figure S3**). This reveals persistent metabolic alterations in TCD patients compared with HCs. Of the metabolites altered between TCD and HC, 11 (46%) were also significantly altered between TCD and UCD groups, indicating that treatment with a GFD influences these metabolites independently of disease status (**Figure 6**).

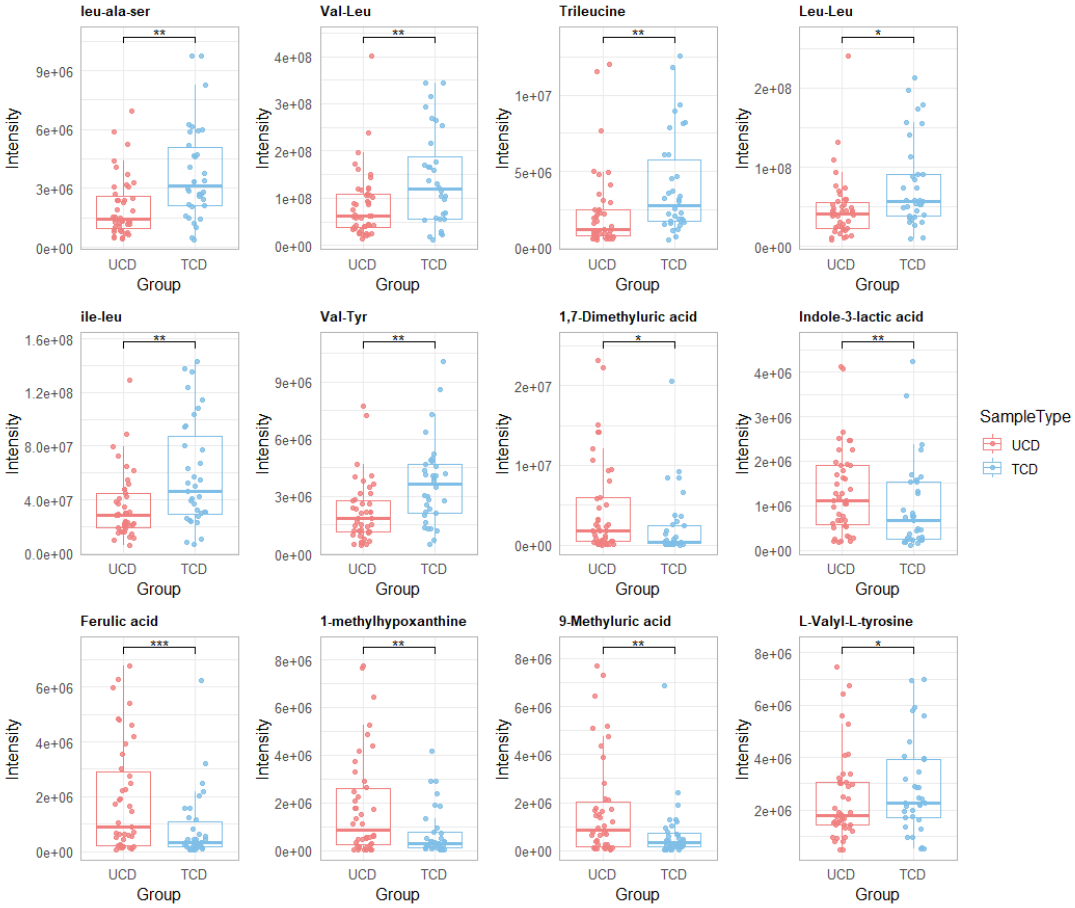


Figure 5. Treatment-responsive faecal metabolites in patients with CoD. Boxplots showing the relative intensities of metabolites significantly altered between untreated coeliac disease (UCD, red) and treated coeliac disease (TCD, blue). Statistical significance is indicated by *p < 0.05, **p < 0.01, ***p < 0.001.

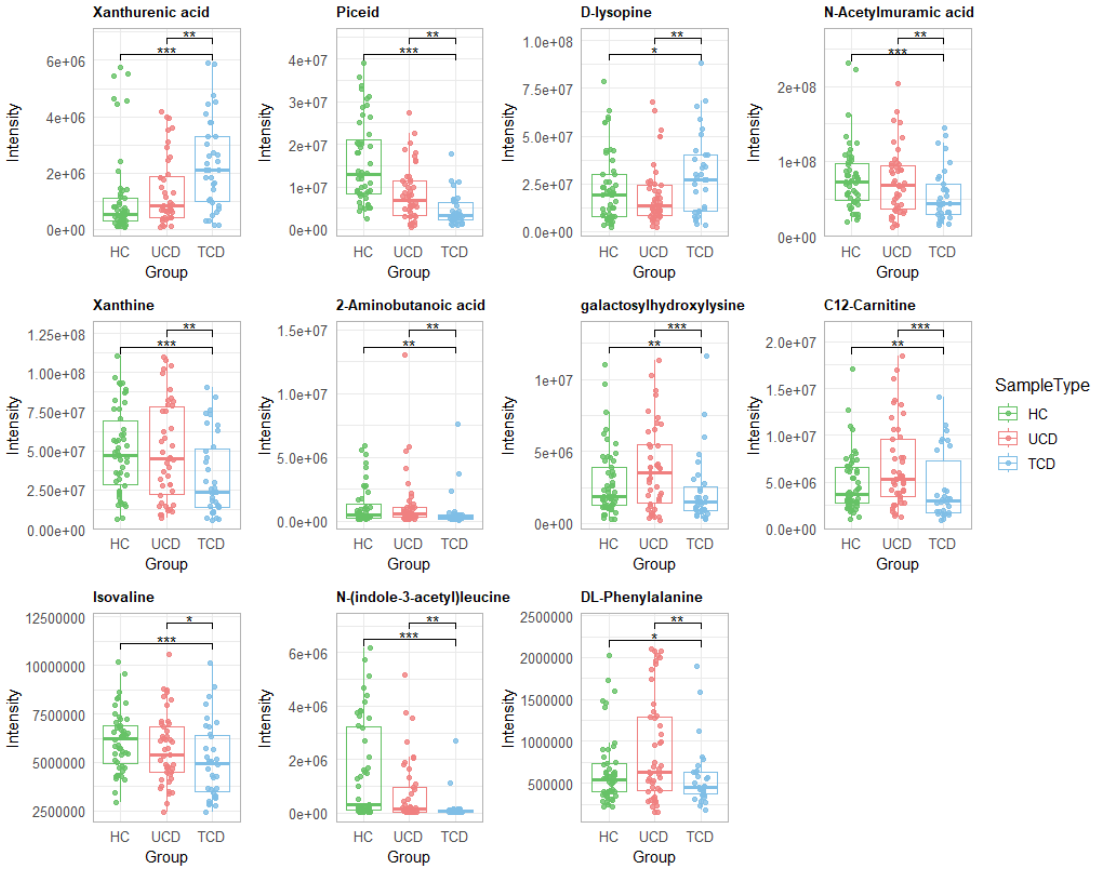


Figure 6. Treatment responsive metabolites in all participants. Boxplots showing the relative intensities of metabolites significantly altered in both untreated coeliac disease (UCD, red) and HCs (green) compared with treated coeliac disease (TCD, blue). These metabolites reflect changes driven by a GFD rather than CoD status. Statistical significance is indicated by *p < 0.05, **p < 0.01, ***p < 0.001.

3.4.5 Coeliac Disease-specific Metabolome Signature

To identify metabolite features specific to CoD, we identified the metabolites that are significantly altered between both HCs and UCD as well as between HCs and TCD (**Figure 7**). These metabolites are therefore potential disease-specific biomarkers. Receiver operating characteristic (ROC) analysis was performed to characterise the predictive value of the most influential disease metabolites, with results showing an area under the curve (AUC) value of 0.885.

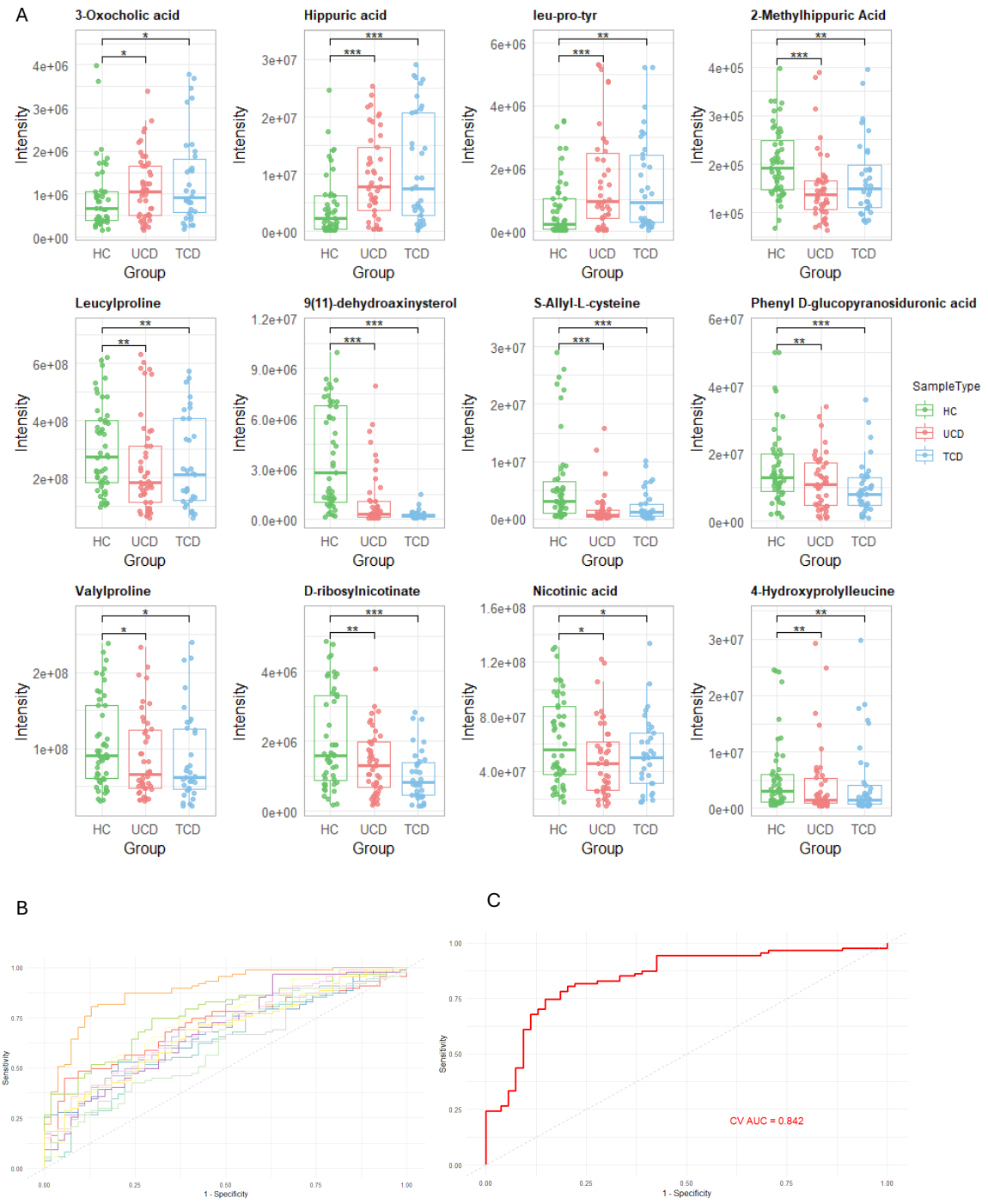


Figure 7. Most influential metabolites. (A) Boxplots of CoD-specific metabolites. (B) Receiver operating characteristic (ROC) curve using the top 10 metabolites from the VIP plot. (C) Combined ROC.

3.4.6 Discriminant Analysis in Metabolome Profile in New-onset CoD After Recommendation to a GFD

Analysis of the prospective cohort was performed, whereby samples were taken and analysed at three timepoints: (1) at time of disease diagnosis, (2) 6 months and (3) 12 months following a GFD. 15 and 23% of patients had detectable levels of faecal GIP at 6 and 12 month follow ups respectively, indicating recent consumption of gluten. GIP positive patients were removed from the analysis, and therefore the results show the true metabolic impact of following a GFD.

In this cohort, 10 and 9 metabolites were significantly altered at 6 and 12 months after initiation of a GFD, respectively (**Figure 8**). 7 metabolites were only increased at 6 months on a GFD, while 6 metabolites were only increased at 12 months on a GFD. Compared with CoD diagnosis, monolinolenin, N-indole-3-acetyl-leucine, and 9(11)-dehydroaxinysterol were consistently altered in both groups. These findings align with trends observed in the cross-sectional cohort, where 10 out of the 16 metabolites were also altered in this cohort (**Supplementary Table S2**). tTG and PedsQL-GS levels in coeliac disease patients throughout the duration of a GFD are shown in **Supplementary Figure S4.**

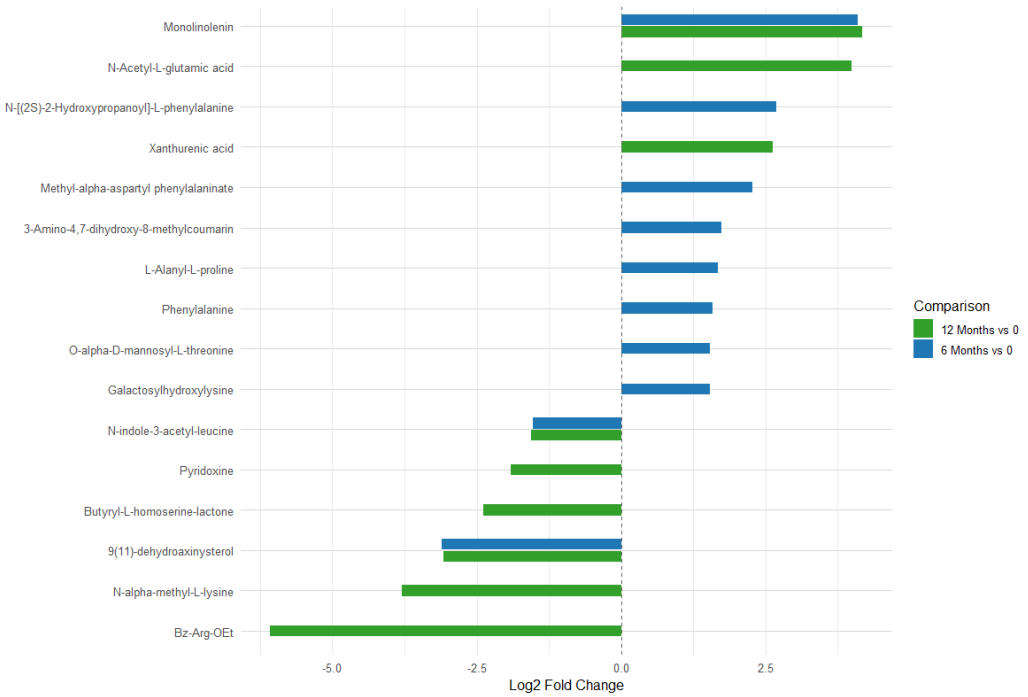


Figure 8. Statistically significant differences (log2 fold change) in metabolite levels between coeliac disease diagnosis and follow-up time points on a gluten-free diet.

3.4.7 Comparison in Metabolome Between Treated Patients with CoD and Their Unaffected Siblings

The faecal metabolome between treated CoD patients and their unaffected siblings was compared to distinguish metabolic alterations specifically associated with CoD from those related to shared genetic risk (**Figure 9**). 23 metabolites differed significantly between TCD and their unaffected siblings (10 decreased and 13 increased). 19 metabolites were significantly altered between the unaffected siblings of TCD children and HCs (7 decreased and 12 increased). OPLS-DA showed separation in both groups (TCD vs siblings and HC vs siblings).

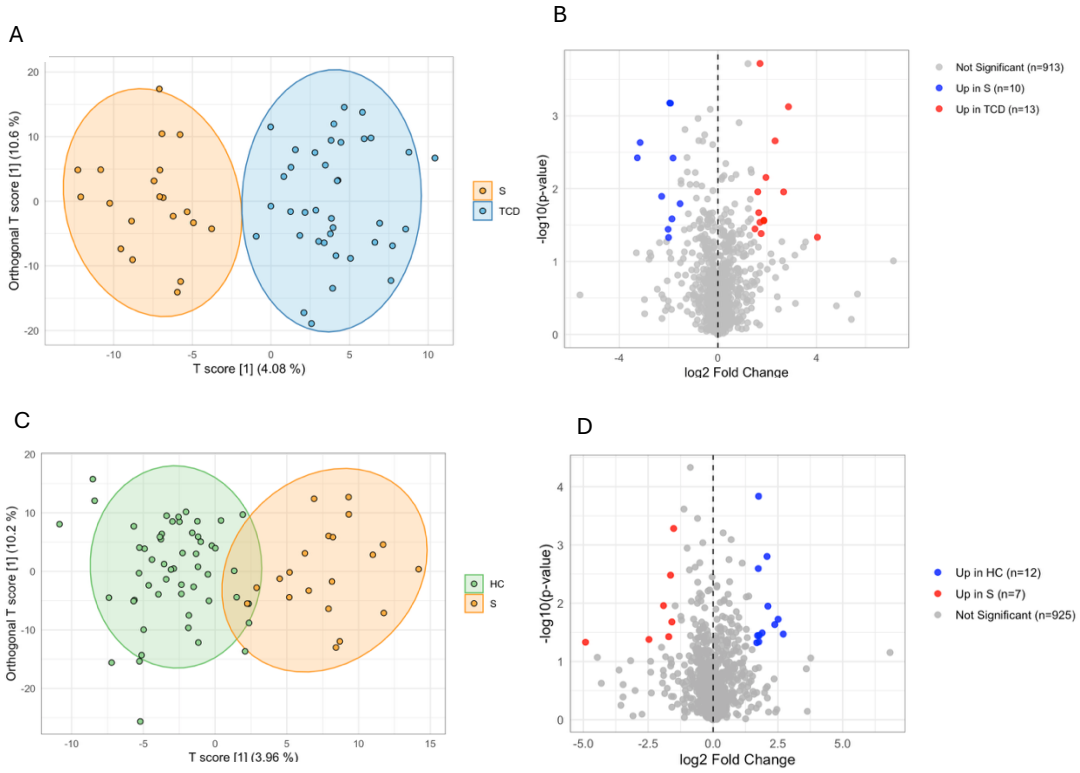


Figure 9. Metabolomics analysis comparing treated coeliac disease (TCD) and healthy controls (HC) with unaffected siblings. (A) OPLS-DA scores plot for TCD vs siblings (ellipses = 95% confidence intervals). (B) Volcano plot of differential metabolite features for TCD vs siblings ($p < 0.05$, fold change > 1.5, log2FC = TCD/S). (C) OPLS-DA scores plot for HC vs siblings (ellipses = 95% confidence intervals). (D) Volcano plot of differential faecal metabolite features for HC vs siblings ($p < 0.05$, fold change > 1.5, log2FC = HC/S).

3.5. Discussion

While previous studies have characterised microbial alterations in CoD, the corresponding functional consequences on the gut metabolome remain largely unexplored. In this study, we profiled the faecal metabolome of children with new-onset and treated CoD, alongside healthy controls and first-degree relatives using untargeted LC-MS, to identify metabolites that may reflect disease processes or treatment effects. The majority of CoD metabolomics work to date has focussed on serum or plasma samples [11, 20-23], which are critical for understanding systemic disease mechanisms. However, there is a growing need to investigate intestinal disorders using faecal samples, given their non-invasive nature and their ability to provide direct insights into gut environment. Faecal samples reflect both the composition and functional activity of the gut microbiota, as well as microbial-derived metabolites that influence host physiology. This is particularly pertinent in CoD, where microbial dysbiosis has been increasingly implicated in disease pathogenesis.

Our analysis revealed distinct clustering between HCs, UCD patients, and TCD patients, highlighting the presence of disease and treatment associated metabolic signatures. Consistent with the absence of pronounced microbial dysbiosis reported previously [6, 24], UCD patients showed largely unchanged global faecal metabolome profiles compared with HCs, yet specific metabolites were differentially altered between the groups. UCD patients exhibited subtle alterations in amino acid, bile acid, and lipid metabolites compared with HCs, suggesting early shifts in nutrient absorption and microbial metabolism. Notably, UCD patients showed increased levels of acylcarnitines (acetylcarnitine, undecanoylcarnitine), bile acid derivatives (nutriacholic acid), amino acids and dipeptides (glycine, leucyl-prolyl-tyrosine), and microbial metabolites.

Importantly, three major patterns of metabolite profiles were identified, analogous to the microbial groups previously described in this cohort. Firstly, a distinct panel of 12 CoD-specific, non-treatment responsive metabolites were identified, reflecting core disease-associated metabolic signatures that are not fully corrected by dietary treatment. These metabolites spanned several chemical classes, including the disruption of bile acid metabolism, with 3-oxocholic acid elevated in both untreated and treated patient groups. Amino acid related metabolites showed widespread alterations, including leu-pro-tyr,

leucylproline, valylproline, 4-hydroxyprolylleucine, S-allyl-L-cysteine, and nicotinic acid, indicating persistent dysregulation of protein turnover and amino acid catabolism. Microbial co-metabolites, such as hippuric acid, 2-methylhippuric acid, and phenyl-D-glucopyranosiduronic acid were also consistently perturbed, pointing to altered host–microbiota metabolic interactions, and changes in sterol metabolism were evident through the persistent alteration of 9(11)-dehydroaxinysterol. Receiver operating characteristic (ROC) analysis confirmed the diagnostic potential of this 12-metabolite panel, with the top features achieving an AUC of 0.885. Collectively, these findings indicate a panel of disease-specific metabolites, serving as candidate biomarkers of CoD independent of gluten exposure.

Another group of metabolites identified in this study comprises treatment-responsive metabolites associated with new-onset CoD. These metabolites are altered in UCD patients relative to HCs but normalise following treatment with a GFD, reflecting dietary modulation of disease-associated pathways. Comparison of TCD and UCD patients identified 27 metabolites significantly altered by dietary intervention. Most of these changes involved amino acid dipeptides and oligopeptides, including Leu-Ala-Ser, Val-Leu, Trileucine, Ile-Leu, Leu-Leu, Val-Tyr, and L-Valyl-L-tyrosine, which were consistently increased in TCD relative to UCD. These findings suggest enhanced protein catabolism and remodelling of amino acid metabolism during dietary recovery. Indole-3-lactic acid, a tryptophan-derived microbial product, and 1,7-dimethyluric acid, a purine metabolite, were additionally reduced in TCD patients, reflecting shifts in gut microbial activity and nucleotide metabolism with gluten exclusion.

The third group comprises treatment-dependent, non-disease-specific metabolites that are primarily influenced by dietary intervention rather than disease status. Comparison of TCD patients with HCs revealed 24 metabolites significantly altered following GFD adherence, of which nearly half (46%) overlapped with metabolites also altered between TCD and UCD patients. This suggests that many of these changes occur independently of CoD itself and are likely driven by the GFD. Among the most prominent alterations were amino acid–related metabolites. Levels of 2-aminobutanoic acid, DL-phenylalanine, and N-(indole-3-acetyl) leucine were significantly reduced in TCD patients, suggesting remodelling of amino acid metabolism during dietary adaptation. In particular, the reduction in indole-derived

compounds highlights shifts in microbial tryptophan catabolism, consistent with dietary modulation of host-microbe interactions. Isovaline, a non-proteinogenic amino acid, was also altered, further supporting broad impacts of GFD on nitrogen and amino acid metabolism.

In parallel, several metabolites linked to microbial and host co-metabolism were altered by a GFD. Xanthurenic acid and xanthine, both intermediates of tryptophan and purine metabolism, remained perturbed in TCD compared with HCs, pointing to persistent alterations in nucleotide turnover and kynurenine pathway activity. Similarly, changes in N-acetylmuramic acid, a bacterial cell wall component, indicate modulation of gut microbial composition under dietary treatment. Additional alterations were observed in energy and lipid-related metabolites. C12-carnitine, a medium-chain acylcarnitine, was reduced in TCD, suggesting changes in mitochondrial β-oxidation and fatty acid handling with dietary restriction. The decreased abundance of galactosylhydroxylysine, a collagen-derived metabolite, may reflect shifts in extracellular matrix remodelling or microbial proteolysis. These findings underscore the importance of considering diet-driven metabolic effects when interpreting faecal metabolomic profiles, particularly in treated CoD patients, where residual changes may not reflect disease activity but rather secondary dietary consequences. Persistent disturbances despite adherence may signal ongoing inflammation or incomplete remission, underscoring the need to map key metabolic changes to inflammatory markers and clinical outcomes. This is particularly important in non-responsive or refractory CoD, where patients face higher risks of complications, including intestinal lymphoma and adenocarcinoma [32] and may require additional drug treatments, emphasising the value of early identification of this patient subgroup.

To gain further insights into CoD, investigation was carried out into the metabolome of unaffected siblings by comparing them to the TCD group. Unaffected siblings are an informative reference group as they share genetic susceptibility and often similar environmental exposures, but do not develop disease. By comparing their metabolomic profiles to those of treated CoD patients, we can identify metabolic alterations that are associated with disease, rather than genetic susceptibility or lifestyle factors.

The outcomes observed from analysis of the prospective cohort build upon the findings obtained from the cross-sectional study. Looking into the effect of a GFD across 6-months, it was shown that aspartic acid, malic acid, and serine metabolites increased, while only ornithine was shown at higher levels after 12-months of treatment. This co-occurred with a decrease in tTG levels and an increase in PedsQL-GS score, indicating improvement of GI symptoms upon initiation of a GFD. While it is interesting to note that metabolic alteration was more profound at 6 months than at 12 months on a GFD, the same finding were revealed in the microbiome analysis of the same CoD study cohort [6]. The authors assumed that this can likely be explained by loss of adherence to GFD with time, as supported by the change in GIP levels, which further emphasises the difficulties associated with sustaining complete exclusion of gluten in the diet. These findings do however contrast those of a different study suggesting the metabolome fingerprint returns to normal after 12 months [21] as we did not observe complete metabolic restoration after this time.

This study has various limitations; firstly, there was a relatively small size included in the prospective cohort, partly due to unavailability of follow-up measurements for some of the patients. This is particularly relevant for analysis of faecal samples which are inherently complex and interindividual metabolite matrices, highly influenced by factors such as geographical location (e.g., regional dietary patterns or food availability) and cultural dietary practices. While untargeted metabolomics is unable to capture the complete metabolite pool in each sample, we utilised a previously optimised method which maximised metabolite detection and coverage. A further challenge relates to the inherent ambiguity in interpreting faecal metabolomic data. Reductions in the stool abundance of a given metabolite may reflect decreased microbial or host production, but it could equally indicate increased utilisation or absorption in the body. Conversely, elevated faecal levels may arise from enhanced production, reduced utilisation, impaired absorption, or dietary changes. These factors underscore the importance of cautious interpretation, and integration with dietary records, microbial profiling, and host biomarker measurements is therefore important to distinguish relevant biological effects. Future work should work on building the current findings to provide a deeper understanding of the intricate interplay between the microbiome, metabolome, and disease to derive clinical relevance of CoD through omics. Integrating metabolomic and microbiome data will offer deeper insights into disease

pathogenesis and its downstream consequences, ultimately guiding strategies to improve patient outcomes.

3.6. Conclusion

This study provides a comprehensive characterisation of the faecal metabolome in children with coeliac disease (CoD), revealing a distinct metabolic fingerprint involving perturbations in amino acid, bile acid, lipid, and microbial metabolites. While adherence to a gluten-free diet (GFD) improved clinical parameters and partially restored the metabolome, a subset of metabolites remained persistently altered, indicating core dysregulation not fully corrected by diet and potentially linked to long-term mucosal injury, host-microbiota interactions, or dietary restriction. By distinguishing CoD-specific metabolites (e.g., bile acids, sterols, and microbial co-metabolites), treatment-responsive metabolites (predominantly amino acid dipeptides and oligopeptides), and diet-driven metabolites (notably amino acid derivatives and purine-related compounds), we reveal distinct metabolic signatures associated with disease, recovery, and dietary effects. These findings underscore the value of faecal metabolomics as a non-invasive tool for biomarker discovery, disease monitoring, and the identification of non-responsive or refractory cases, with future integration of metabolomic, microbiome, and clinical data needed to guide improved diagnostic and therapeutic strategies.

Funding: This study was supported by Shimadzu UK and the University of Strathclyde through joint contribution to P.E.K.'s studentship. The clinical studies were funded by the Glasgow Children Hospital Charity and the Nutricia Research Foundation. Z.R. want to acknowledge the EPSRC Multiscale Metrology Suite (EP/V028960/1).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of NHS West of Scotland Research Ethics Committee (Ref: 11/WS/0006) for the study in patients with coeliac disease and the Yorkhill Research Ethics Committee (Reb: 05/S0708/66) for the study in patients with Crohn's disease.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: N.J.W.R. want to acknowledge the Strathclyde Centre of Molecular Bioscience (www.scmb.strath.ac.uk) for access to LCMS instrumentation. We also acknowledge the help and advice given by the team at Shimadzu UK—Christine Heinz, Emily Armitage, Sean Devers, Chris Titman, Jonathan McGeehan and Stephen Brookes.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

3.7. References

1. Singh, P., et al., *Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis.* Clin Gastroenterol Hepatol, 2018. 16(6): p. 823-836.e2.
2. Husby, S., et al., *European Society Paediatric Gastroenterology, Hepatology and Nutrition Guidelines for Diagnosing Coeliac Disease 2020.* J Pediatr Gastroenterol Nutr, 2020. 70(1): p. 141-156.
3. Mearin, M.L., et al., *ESPGHAN Position Paper on Management and Follow-up of Children and Adolescents With Celiac Disease.* J Pediatr Gastroenterol Nutr, 2022. 75(3): p. 369-386.
4. Di Biase, A.R., et al., *Gut microbiota signatures and clinical manifestations in celiac disease children at onset: a pilot study.* J Gastroenterol Hepatol, 2021. 36(2): p. 446-454.
5. Olivares, M., et al., *Gut microbiota trajectory in early life may predict development of celiac disease.* Microbiome, 2018. 6(1): p. 36.
6. Zafeiropoulou, K., et al., *Alterations in Intestinal Microbiota of Children With Celiac Disease at the Time of Diagnosis and on a Gluten-free Diet.* Gastroenterology, 2020. 159(6): p. 2039-2051.e20.
7. El Mouzan, M., A. Assiri, and A. Al Sarkhy, *Gut microbiota predicts the diagnosis of celiac disease in Saudi children.* World J Gastroenterol, 2023. 29(13): p. 1994-2000.
8. Leonard, M.M., et al., *Microbiome signatures of progression toward celiac disease onset in at-risk children in a longitudinal prospective cohort study.* Proc Natl Acad Sci U S A, 2021. 118(29).
9. Girdhar, K., et al., *Dynamics of the gut microbiome, IgA response, and plasma metabolome in the development of pediatric celiac disease.* Microbiome, 2023. 11(1): p. 9.
10. Kirchberg, F.F., et al., *Investigating the early metabolic fingerprint of celiac disease - a prospective approach.* J Autoimmun, 2016. 72: p. 95-101.
11. Upadhyay, D., et al., *Abnormalities in metabolic pathways in celiac disease investigated by the metabolic profiling of small intestinal mucosa, blood plasma and urine by NMR spectroscopy.* NMR Biomed, 2020. 33(8): p. e4305.
12. Upadhyay, D., et al., *NMR based metabolic profiling of patients with potential celiac disease elucidating early biochemical changes of gluten-sensitivity: A pilot study.* Clin Chim Acta, 2022. 531: p. 291-301.

13. Murch, S., et al., *Joint BSPGHAN and Coeliac UK guidelines for the diagnosis and management of coeliac disease in children.* Archives of Disease in Childhood, 2013. 98(10): p. 806-811.

14. Varni, J.W., et al., *PedsQL gastrointestinal symptoms module: feasibility, reliability, and validity.* J Pediatr Gastroenterol Nutr, 2014. 59(3): p. 347-55.

15. Gerasimidis, K., et al., *Comparison of Clinical Methods With the Faecal Gluten Immunogenic Peptide to Assess Gluten Intake in Coeliac Disease.* J Pediatr Gastroenterol Nutr, 2018. 67(3): p. 356-360.

16. Kelly, P.E., et al., *An Optimised Monophasic Faecal Extraction Method for LC-MS Analysis and Its Application in Gastrointestinal Disease.* Metabolites, 2022. 12(11).

17. Broadhurst, D.I., *QC:MXP Repeat Injection based Quality Control, Batch Correction, Exploration & Data Cleaning (version 1) Zendono. https://zenodo.org/doi/10.5281/zenodo.11101541. Retrieved from https://github.com/broadhurstdavid/QC-MXP.* 2024.

18. Vich Vila, A., et al., *Faecal metabolome and its determinants in inflammatory bowel disease.* Gut, 2023. 72(8): p. 1472-1485.

19. Wu, L., et al., *Altered Gut Microbial Metabolites in Amnestic Mild Cognitive Impairment and Alzheimer's Disease: Signals in Host-Microbe Interplay.* Nutrients, 2021. 13(1).

20. Solakivi, T., et al., *Serum fatty acid profile in celiac disease patients before and after a gluten-free diet.* Scand J Gastroenterol, 2009. 44(7): p. 826-30.

21. Bertini, I., et al., *The metabonomic signature of celiac disease.* J Proteome Res, 2009. 8(1): p. 170-7.

22. Lewis, N.R., et al., *Cholesterol profile in people with newly diagnosed coeliac disease: a comparison with the general population and changes following treatment.* Br J Nutr, 2009. 102(4): p. 509-13.

23. Jamnik, J., D.J. Jenkins, and A. El-Sohemy, *Biomarkers of cardiometabolic health and nutritional status in individuals with positive celiac disease serology.* Nutr Health, 2018. 24(1): p. 37-45.

24. Aboulaghras, S., et al., *Meta-Analysis and Systematic Review of HLA DQ2/DQ8 in Adults with Celiac Disease.* Int J Mol Sci, 2023. 24(2).

25. Bortoluzzi, C., S.J. Rochell, and T.J. Applegate, *Threonine, arginine, and glutamine: Influences on intestinal physiology, immunology, and microbiology in broilers.* Poult Sci, 2018. 97(3): p. 937-945.

26. Mao, X., et al., *Specific roles of threonine in intestinal mucosal integrity and barrier function.* Front Biosci (Elite Ed), 2011. 3(4): p. 1192-200.

27. Ruth, M.R. and C.J. Field, *The immune modifying effects of amino acids on gut-associated lymphoid tissue.* Journal of Animal Science and Biotechnology, 2013. 4(1): p. 27.

28. Torinsson Naluai, Å., et al., *Altered peripheral amino acid profile indicate a systemic impact of active celiac disease and a possible role of amino acids in disease pathogenesis.* PLoS One, 2018. 13(3): p. e0193764.

29. Sevinc, E., et al., *AMINO ACID LEVELS IN CHILDREN WITH CELIAC DISEASE.* Nutr Hosp, 2015. 32(1): p. 139-43.

30. Caminero, A., et al., *Duodenal Bacteria From Patients With Celiac Disease and Healthy Subjects Distinctly Affect Gluten Breakdown and Immunogenicity.* Gastroenterology, 2016. 151(4): p. 670-83.

31. Gambini, J., et al., *Properties of Resveratrol: In Vitro and In Vivo Studies about Metabolism, Bioavailability, and Biological Effects in Animal Models and Humans.* Oxid Med Cell Longev, 2015. 2015: p. 837042.

32. Demiroren, K., *Possible relationship between refractory celiac disease and malignancies.* World J Clin Oncol, 2022. 13(3): p. 200-208.

4.0 CHAPTER 4

Optimisation of a Dilute-and-Shoot UHPLC-MS Method for High-Throughput Urinary Metabolomics

Patricia E. Kelly [1], Gillian Farrell [1], Konstantinos Gkikas[2], Caroline Kerbiriou[2], Bernadette White[2], Konstantinos Gerasimidis[2], Nicholas J. W. Rattray[1].

To be submitted to Analytical Chemistry

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G4 0RE, UK.

[2] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow G12 8QQ, UK.

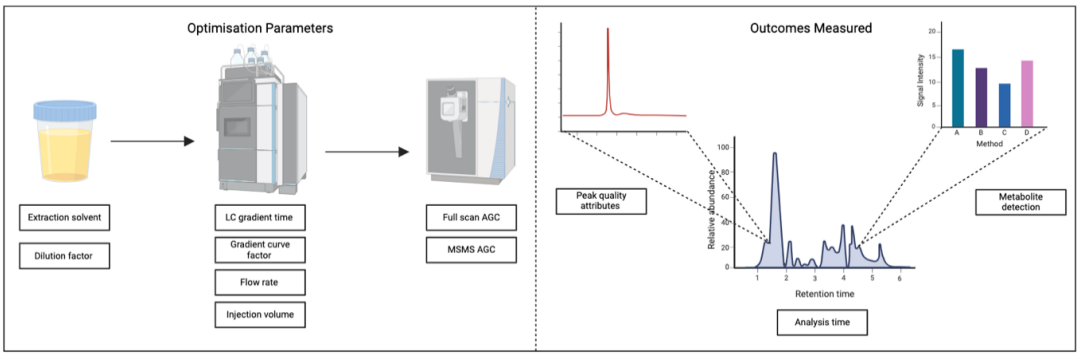
Author Contributions: Conceptualization, P.E.K., G.F., K.G., C.K., B.W., K.G. and N.J.W.R.; methodology, P.E.K., K.G. and N.J.W.R.; software, P.E.K., G.F. and N.J.W.R.; validation, P.E.K., G.F. and N.J.W.R.; formal analysis, P.E.K., K.G., C.K., B.W., K.G. and N.J.W.R.; investigation, P.E.K., K.G., C.K., B.W., and N.J.W.R.; resources, K.G.; data curation, P.E.K., G.F. and N.J.W.R.; writing—original draft preparation, P.E.K. and N.J.W.R.; writing—review and editing, P.E.K., K.G. and N.J.W.R.; visualisation, P.E.K. and N.J.W.R.; supervision, N.J.W.R., K.G.; project administration, N.J.W.R.; funding acquisition, N.J.W.R and K.G. All authors have read and agreed to the published version of the manuscript.

4.1 Abstract

Urinary metabolomics using liquid chromatography-mass spectrometry (LC-MS) holds great potential for biomarker development and clinical application for a variety of diseases. Large-scale clinical studies uniquely require a rapid and high-throughput method which can be easily scaled for the analysis of large sample numbers in a minimal timeframe. While several extraction and LC-MS methods currently exist for urine samples, there is a lack of standardisation and method recommendations for larger studies and clinical trials.

In the current study, we herein optimised a method for urinary metabolite extraction and untargeted ultra-high-performance LC-MS (UHPLC-MS) analysis for large scale studies. Eight UHPLC-MS parameters were optimised based on the quantification of the following outcomes: the number of metabolites measured, peak quality, creatinine performance, and peak quality factors (PQFs). The extraction solvent, dilution factor, chromatography gradient, injection volume, flow rate, gradient curve factor, full scan AGC, and MSMS AGC were all considered to complete a comprehensive method optimisation. A rapid chromatographic separation with a 6.5-minute data collection time on a C18 column was developed in positive ionisation mode, using water + 0.1% FA (v/v) and acetonitrile + 0.1% FA (v/v) as the mobile phases. The total run time of the optimised method, including a cleaning step, was 10 minutes.

To demonstrate the effectiveness of the optimised method, 1094 urine samples from healthy controls and inflammatory bowel disease (IBD) patients were analysed, where over 1500 metabolites were putatively identified with an analysis time of 182.3 hours, saving over 91 hours of total LC-MS instrument running time in comparison to the previously used method.



Graphical Abstract. Overview of Experimental Design for Untargeted Urinary Metabolomics Optimisation. Eight parameters were optimised across the protocol, including sample preparation, LC and MS analytical conditions of the experimental pipeline. Outcomes were measured by peak quality attributes, analysis time, and metabolite detection. AGC, automatic gain control; MSMS, tandem mass spectrometry.

4.2 Introduction

Urine is a valuable sample biomatrix for disease research due to the ease and non-invasive nature of collection, making it preferable for large-scale studies [1]. The large quantity of urine produced by humans allows for multiple tests to be conducted on the same sample, thereby enabling easier replication and validation and ensuring data consistency across experimental techniques. Furthermore, due to the frequent production of urine, longitudinal studies can be performed more easily, for example to monitor response to treatment or disease progression over time [2]. Urine contains a diverse range of biological molecules which reflect metabolic and physiological changes and can therefore act as biomarkers for a multitude of diseases. Urinary biomarkers have been successfully established for a number of diseases and have had significant impacts on clinical care. For example, urinary albumin is a recognised biomarker of kidney damage [3], with the albumin-to-creatinine ratio (ACR) routinely used to diagnose and monitor disease such as chronic kidney disease (CKD) [4]. Urinary biomarkers have also shown great promise in cancer research, with prostate cancer antigen 3 (PCA3) now used as a biomarker for the diagnosis of prostate cancer [5, 6], and in many other fields including metabolic disease [7] and neurology research [8]. Among the approaches used for biomarker discovery, metabolomics, the study of small molecules and products of metabolism, has more recently emerged as a tool with great potential to give unique disease insights. 2-hydroxyglutarate,

for example, is used as a metabolite biomarker for certain types of brain tumour, particularly gliomas [9, 10] and acute myeloid leukaemia [11].

The identification of urinary biomarkers and their translation into clinical use has been enabled by advances in the analytical technologies that are used to detect biomarkers. Liquid-chromatography mass spectrometry (LC-MS) is a powerful analytical tool for biomarker discovery and disease mechanism studies, particularly through untargeted approaches, due to its high sensitivity and selectivity and ability to analyse a comprehensive range of chemical molecules found in urine [12, 13]. Large-scale clinical studies routinely utilise untargeted LC-MS approaches for the identification, validation, and clinical application of biomarkers. The use of a method which both maximises chemical coverage and minimises analysis time and sample deterioration, particularly for the analysis of sensitive metabolites, is therefore vital for application to larger studies.

The pipeline of urinary metabolomics analysis involves several steps, including metabolite extraction, UHPLC-MS data acquisition, and data analysis. Optimisation of the associated method parameters is crucial for accurately detecting the metabolites of interest for the specific aims of a study, which is particularly important in untargeted metabolomics studies where the aim is to maximise coverage of the large range of chemically diverse compounds found in complex biofluids such as urine. Crucially, for large scale metabolomics studies, fast sample preparation and analysis is essential to prevent sample degradation. A rapid extraction and LC-MS method for urinary metabolite detection is therefore optimal for large scale studies to reduce the overall run time and therefore minimise sample degradation across the study duration.

The “dilute-and-shoot” urine extraction method, which involves diluting the collected sample with a chosen solvent and directly injecting it into the LC-MS system [14, 15], offers numerous benefits in large-scale clinical studies, where the focus is on high-throughput and cost-effective analysis of a large number of samples [16]. In comparison to alternative methods, including solid phase extraction (SPE), advantages of the dilute-and-shoot method include a simplified workflow with a reduced processing time, minimal sample handling and sample loss, and ease of automation and scalability [17, 18]. As the need for large-scale clinical trials investigating urinary metabolomics grows, the dilute-and-shoot approach will

remain essential in biomarker discovery and monitoring. It is therefore vital to work towards the standardisation of an optimal LC-MS method using the dilute-and-shoot extraction approach.

In this investigation, we aimed to optimise an untargeted UHPLC-MS method for urinary metabolomics analysis for large-scale studies while reserving the simplicity of the dilute-and-shoot method. Parameter optimisation was performed across stages of the analysis pipeline, including metabolite extraction, LC, and MS analysis to maximise metabolite identification in urine samples with a reduced preparation and analysis time. To demonstrate the efficiency and applicability of the optimised method, 1094 urine samples from inflammatory bowel disease (IBD) patients and healthy controls were analysed, concluding the largest global IBD urinary metabolomics study performed to date.

4.3 Experimental

4.3.1 Ethics Statement

All participants and their carers provided written informed consent. Approval for the iPENS and CD-TREAT clinical trial studies were granted by the West of Scotland Research Ethics Committee (REC reference: 17/WS/0119, 19/WS/0163, respectively) and NHS Research and Development office. For ENIP the original approval was granted by the College of Medical, Veterinary and Life Sciences Ethics Committee, University of Glasgow (Project number: 200220086). All studies were registered on https://clinicaltrials.gov (identifier NCT number: IPENS, NCT04225689; CD-TREAT, NCT03171246; ENIP, NCT06828094).

4.3.2 Study Design

Individual parameter optimisation was carried out sequentially to improve the performance of the overall method across the experimental pipeline, including the sample preparation protocol and UHPLC-MS method. The best performing parameters were selected using multiple outcome measures, including peak quality attributes, analysis time, and metabolite detection.

4.3.3 Human Urine Sample Collection

Urine samples were collected from healthy individuals and IBD patients aged 6-17 and stored at -80 °C until processing. For the clinical study and application, 1094 urine samples were collected and analysed using the optimised method.

4.3.4 Chemicals and Reagents

LC-MS grade water (H_2O) and acetonitrile (ACN) were purchased from Fisher Scientific (Geel, Belgium). LC-MS grade formic acid was purchased from Thermo Scientific (Prague, Czech Republic).

4.3.5 Urine Sample Preparation

Urine samples were aliquoted into a 96-well plate and diluted with solvent at differing ratios according to the method optimisation stage. The samples were then mixed with a pipette to ensure solvent distribution, and the plate was submitted for UHPLC-MS analysis. ^{13}C labelled tryptophan was used as an internal standard for assessing instrument stability and experimental conditions.

4.3.6 Untargeted LC-MS Metabolite Measurement

Untargeted metabolomic analysis was performed on an ultra-high performance liquid chromatography (UHPLC) system (ThermoFisher Scientific) coupled to an Orbitrap Exploris 240 (ThermoFisher Scientific) mass spectrometer. Chromatographic separation was performed on a Vanquish Accucore C18 + UHPLC analytical column (ThermoScientific, 100 mm × 2.1 mm, 2.6 μM). Mobile phase A was composed of 99.9% water + 0.1% formic acid and mobile B was composed of 99.9% ACN + 0.1% formic acid. Electrospray ionisation (ESI) was used as the ionisation method in positive mode (3900 V). The elution gradient used can be found in Supplementary Information Table S1. The source-dependent parameters were operated under the following conditions: sheath gas, 40 Arb; auxiliary gas, 10 Arb; sweep gas, 1 Arb; ion transfer tube temperature, 300 °C; vaporiser temperature, 280 °C. Instrument calibration was performed using Pierce™ FlexMix™ calibration solution (Thermo Scientific) and ran under vendor recommended settings. MS data collection was

performed in data dependent acquisition mode (DDA) to give putative metabolite identification at MSI level 2.

4.3.7 Feature Annotation and Metabolite Identification

Feature annotation and metabolite identification at MSI Level 2 were performed using Compound Discoverer version 3.3 and metabolites were matched to *mz*Cloud [19] and ChemSpider [20] databases. Data processing involved feature filtration according to pre-defined criteria (**Supplementary Information Table S1**) and peak quality factor (PQF) quantification was subsequently performed.

4.3.8 Statistical Analysis

Chromatographic data processing and visualisation was carried out using FreeStyle software (Thermo Fisher Scientific). Mass spectrometry data processing and analysis was performed using Compound Discoverer software 3.3 (Thermo Fisher Scientific, Waltham, MA, USA), with processing settings displayed in **Supplementary Information Table S2**. Statistical analysis was additionally performed using Prism software 10.4.1.

4.4 Results and Discussion

Optimisation of the sample preparation protocol and UHPLC-MS method focused on eight key parameters to enhance the overall performance and decrease the analysis time of the method. This was carried out sequentially, by firstly optimising the sample preparation method, followed by the liquid chromatography parameters, and finishing with systematic adjustments to mass spectrometry settings. Together, this study analyses sixty parameter comparisons via LC-MS which were analysed in triplicate in Compound Discoverer to make a total of 180 comparisons.

Table 1. Overall Summary of LC-MS protocol parameter optimisation. AGC, automatic gain control.

Protocol Section	Parameter	Number of Parameter Settings Tested
Metabolite Extraction	Extraction solvent	10
	Dilution factor	4
Liquid Chromatography	Chromatography gradient time	2
	Injection volume	4
	Flow rate	5
	Gradient curve	3
Mass Spectrometry	Full scan AGC	16
	MSMS scan AGC	16

The primary metrics used to evaluate parameter performance were the number of metabolites detected after post-acquisition filtration processing and their peak ratings. Additionally, the peak performance of creatinine was evaluated as a metric for evaluating optimisation parameters. Creatinine is a well-characterised stable and endogenous marker present in all urine samples that is commonly used as a reference standard for data normalisation [21]. Evaluation of its peak performance therefore aids in the assessment of the UHPLC-MS method. Peak quality factors (PQFs) were used to assess peak performance, including the zig-zag index, full width at half maximum to base width (FWHM2Base), jaggedness, and modality factors of individual peaks. Visual representation of each of the PQFs evaluated is provided in **Figure 1**, and an overview of outcome parameters is provided in **Supplementary Information Table S3**. For all PQF settings, a low/ zero value refers to a high-quality peak profile, whereas a higher value suggests a poorer peak profile.

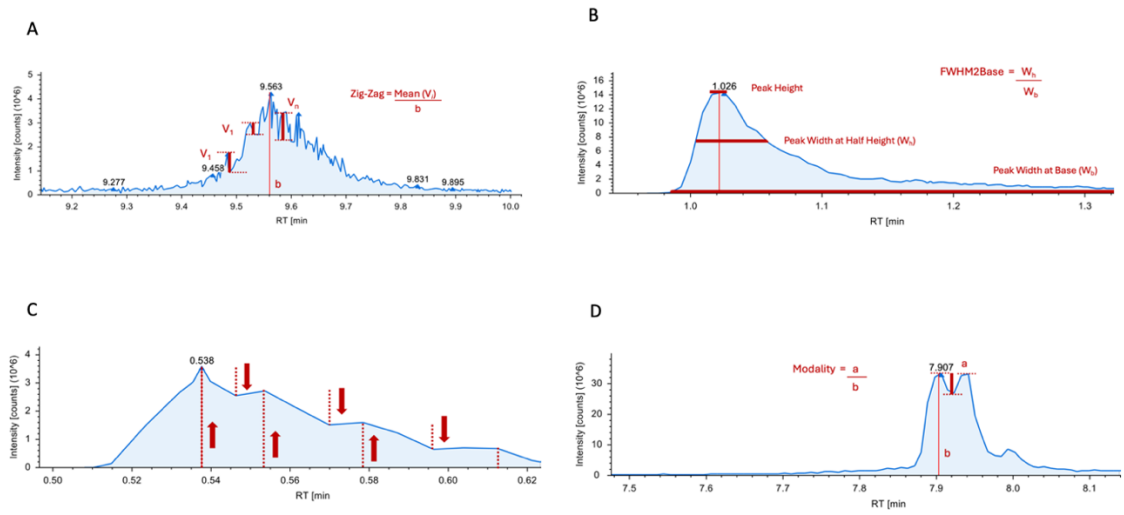


Figure 1. Peak Quality Factor (PQF) metric description. The **(A)** zig-zag quality factor, **(B)** FWHM2Base, **(C)** jaggedness, and **(D)** modality quality factors are described, using example peaks from acquired data [22].

4.4.1 Sample Preparation Optimisation

Sample preparation is a key determinant of a metabolomics experiment. The 'dilute-and-shoot' method was chosen as the urine preparation protocol due to the ease and speed of the preparation protocol, which is a priority for large scale clinical studies. The solvent used for urine metabolite extraction and the dilution factor were investigated in the present study.

4.4.2 Extraction Solvent

For untargeted LC-MS analysis of urine samples using the 'dilute-and-shoot' approach, the choice of solvent is crucial to ensure maximal metabolite detection of a wide range of chemical classes. Here, ten solvents/ solvent combinations were tested to select the best performing solvent. This was first assessed by analysis of the chromatography of each of the solvents (**Supplementary Information Figure S1)** and further method performance was measured by quantifying the total number of metabolites detected, their peak rating, and assessment of creatinine as a reference standard (**Figure 2**).

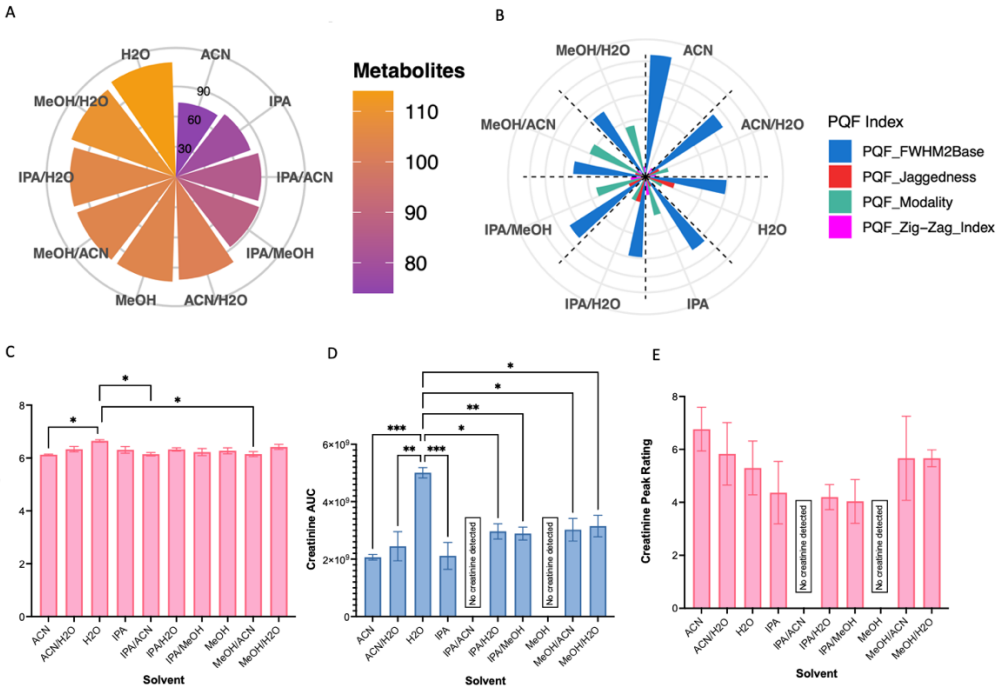


Figure 2. The effect of the extraction solvent on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** the peak performance of creatinine, as measured by the quantification of peak quality factors (PQF)s, Zigzag, FWHM2Base, Jaggedness, and Modality indices, **(C)** average peak rating of all metabolites, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated creatinine peak rating. Creatinine was not detected when IPA/H_2O or MeOH were used as the extraction solvents.

Visual inspection of the solvent chromatograms revealed that H_2O was the only solvent showing successful resolution of peaks between 0.5-2.5 minutes in the method. It was also shown that a 100% H_2O solvent extraction revealed the highest number of detected metabolites in comparison to all other solvents tested (114), with a significant increase notable when compared to five of the other solvents tested. A significantly higher average peak rating of all metabolites was additionally observed in comparison to three of the other solvents tested. When looking at the values of the creatinine peak PQFs of the different solvents/solvent combinations, there were varied results, however; no significant differences were observed between any of the groups. It can be noted that creatinine was not detected when IPA/H_2O or MeOH were used as the extraction solvents and therefore PQF analysis was unable to be carried out for these groups. When selecting high quality data in regard to PQF analysis, lower PQF values (e.g., PQF index = 1) correlate to higher quality peaks and higher PQF values (e.g., PQF = 10) correlate to lower quality peaks. With the exception of peak jaggedness, H_2O showed low/ zero creatinine PQF values, indicating a high-quality peak shape and resolution. Parameter data is displayed in more detail in

Supplementary Information Figure S2 and a comparison of the chromatograms obtained from the analysis of different extraction solvents is shown in **Supplementary Information Figure S3.** As demonstrated by an optimal extraction performance and peak performance, H_2O was therefore selected as the solvent for untargeted LC-MS analysis, and further parameter optimisation was subsequently performed using a H_2O as the extraction solvent. This is in alignment with other LC-MS methods that have also used a H_2O extraction for a variety of applications of urinary metabolomics [14, 23]. While it can be noted that there are some studies which suggest that other solvents, for example the use of 25/75 H_2O/ACN [24] result in a greater number of metabolite features extracted, sample preparation and treatment prior to LC-MS analysis differs between these studies, and therefore we highlight that the H_2O extraction was found to be optimal for a dilute-and-shoot preparation method.

4.4.3 Dilution Factor

The dilution factor used for the H_2O extraction was subsequently optimised, with the aim of simultaneously maintaining detection of low abundance metabolites, preventing instrument overload, and minimising matrix effects. **Figure 3** displays the peak performance characteristics that were obtained during optimisation. The number of metabolites detected was significantly higher when a 1:1 dilution was used in comparison to both a 1:5 and a 1:10 dilution. Results additionally showed that there were no significant differences observed in the average peak rating between the different dilution factors tested. The AUC for the creatinine peak was significantly highest for the 1:1 dilution in comparison to all other parameters tested, with a decreasing trend observed as the dilution factor increased. However, the opposite trend was observed for the peak rating of creatinine, where we see the highest average peak rating for the highest dilution factor used. Creatinine PQF analysis showed no significant differences between the dilution factors for FWHM2Base, and only a 1:1 urine: H_2O ratio gave a non-zero value for jaggedness and modality factors, indicating a lower peak quality. Through consideration of all evaluated outcomes, a dilution factor of 1:5 appeared to provide an optimal balance between metabolite detection and peak quality and was therefore selected to be used for the extraction protocol.

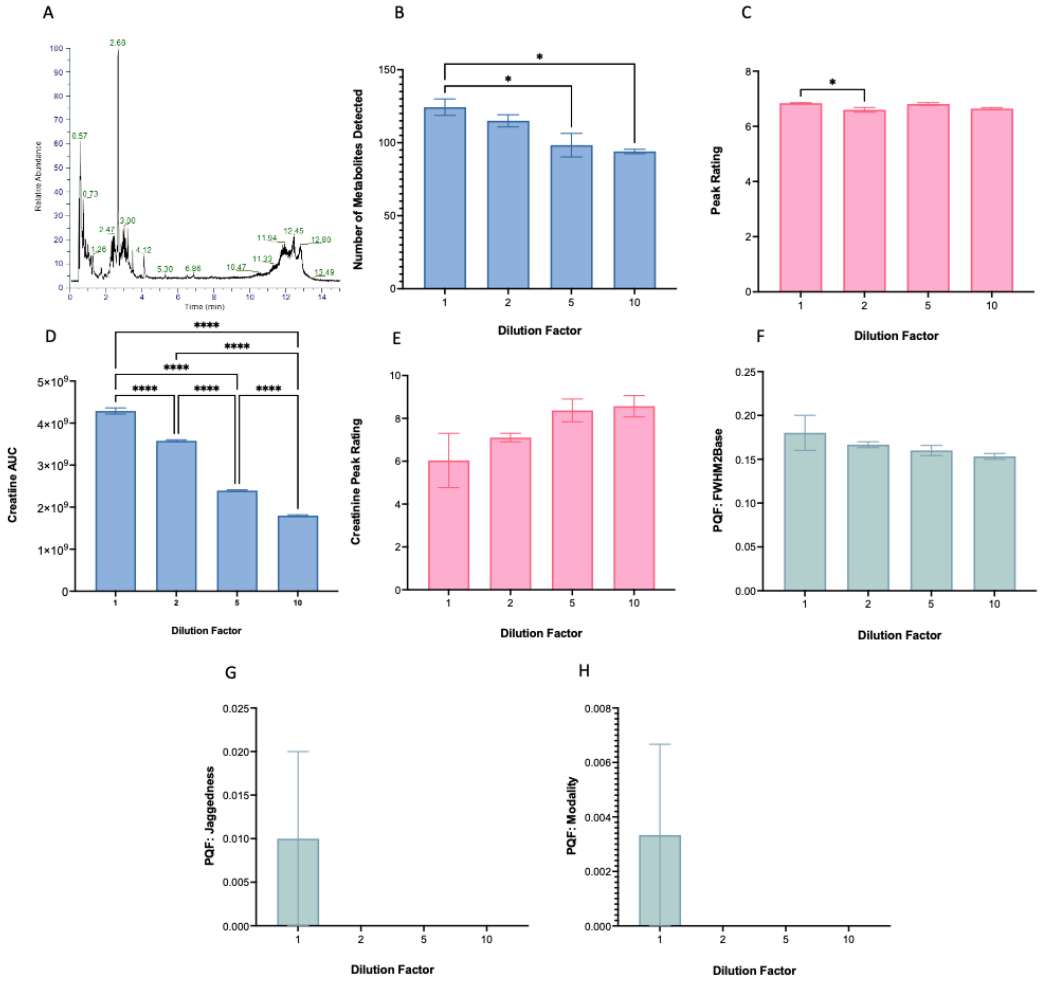


Figure 3. The effect of the dilution factor on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zig-zag indices are not shown as all parameters tested gave a zero value.

4.4.4 Liquid Chromatography Parameter Optimisation

Following sample preparation, the urine samples were subjected to separation by liquid chromatography, which required the selection of several settings. The chromatography run time, flow rate, and injection volume were individually optimised to define an effective method for untargeted urine metabolomics studies.

4.4.5 Chromatography Gradient Time

A short chromatography method, without significantly compromising metabolite detection, is desirable for use in large-scale clinical studies by enabling high-throughput analysis. In this study, we aimed to reduce the run time of UHPLC-MS urine analysis by shortening the time for chromatographic separation.

The chromatography was optimised to give a method with a data collection time of 6.5 mins (70-1000 scan range), with a 3.5-minute cleaning step to give a total run time of 10 minutes. The gradient was designed to maximise metabolite detection throughout the analysis time and avoid empty chromatographic space. It is essential during method optimisation to ensure that a shorter chromatographic analysis time doesn't significantly sacrifice metabolite detection, and therefore the two methods were compared in regard to the number of metabolites identified by each method and their PQFs (**Figure 4**). There was no significant difference in the number of metabolites detected or the peak quality between the two methods, confirming that the 6.5-minute method provides a shorter overall analysis time without sacrificing metabolite detection, which is an essential consideration for clinical metabolomics studies. The AUC for creatinine was shown to be significantly lower in the optimised method, however the inverse relationship was observed for the peak quality of creatinine, although the differential peak quality did not reach significance. No significant differences were observed for the PQF rating of either FWHM2Base, jaggedness, or modality, however the optimised method gave a zero value for the latter, indicating improved peak shape of the new shorter method.

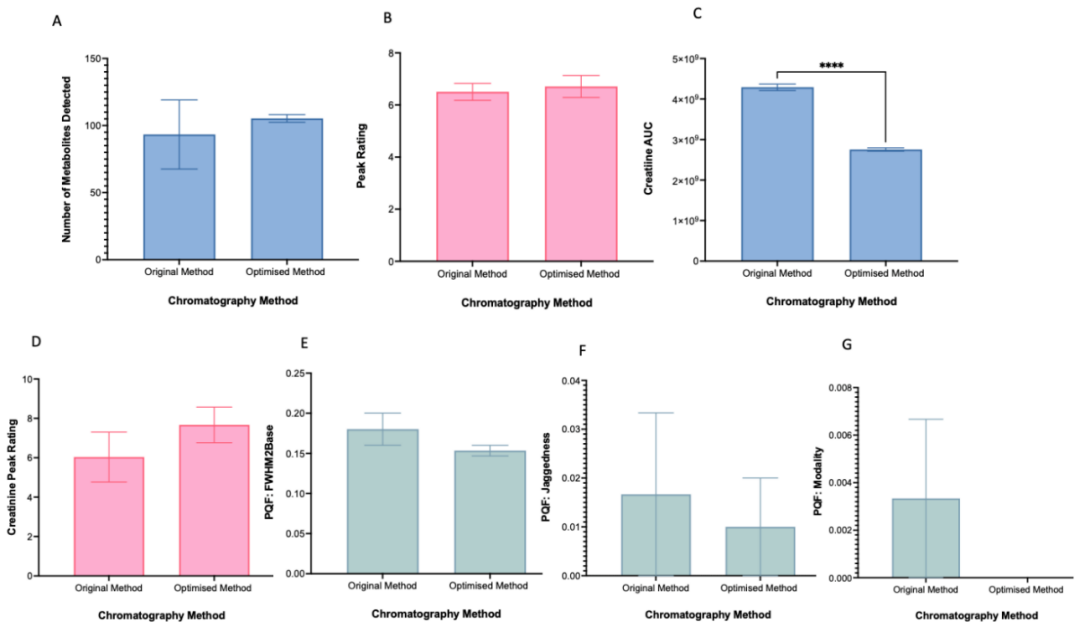


Figure 4. The effect of chromatography gradient elution time. The effects of the chromatography analysis time on untargeted urinary metabolomics were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** FWHM2Base, **(F)** Jaggedness, and **(G)** Modality indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

4.4.6 Injection Volume

The injection volume used in an LC-MS run is a critical method parameter directly affecting the analytical performance of an experiment. The specific amount of analyte injected into the column affects the peak shape and resolution of the data obtained, and a balance is required to determine a large enough signal intensity for optimal detection of metabolites, including low-abundance metabolites, and preventing column overload. A variety of injection volumes ranging from 0.5-5µL were compared in the current study to determine the optimal volume for untargeted urine metabolomics using the Orbitrap-240 system, as shown in **Figure 5**. Examination of method parameters revealed that an injection volume of 5 µL resulted in the greatest number of metabolites detected, with no significant differences in their average peak rating. However, when a 5 µL injection volume was used, the creatinine peak rating was significantly lower than a 1 µL and 2 µL injection volume. While metabolite quantification would initially suggest that a 5 µL injection volume may be optimal for the UHPLC-MS method, the poor peak quality of creatinine, as demonstrated by

the peak broadening in the creatinine chromatogram and large creatinine AUC, led to the selection of a 2 µL as the optimal injection volume. This additionally reduces the risk of column overload, which is an important consideration for studies analysing a large number of samples. Several published studies have also utilised an injection volume of 2 µL in urinary mass spectrometric analysis [25-27] which support the results obtained from the present analysis. A comparison of the chromatograms obtained from the analysis of different injection volumes is shown in **Supplementary Information Figure S4.**

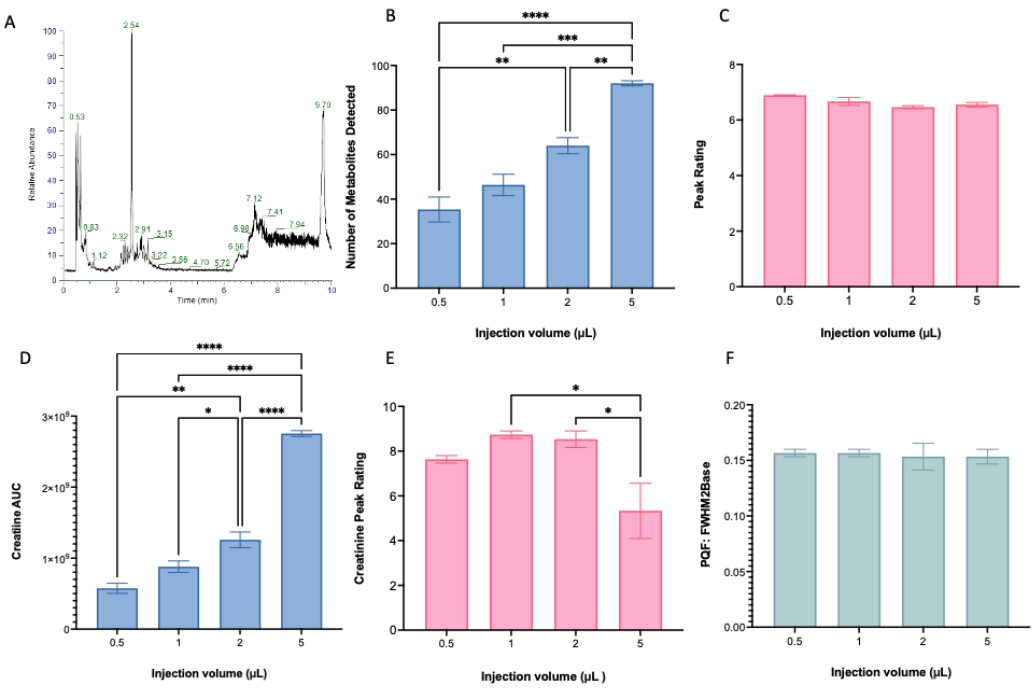


Figure 5. The effect of injection volume on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by **(F)** the FWHM2Base. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zigzag, jaggedness, and modality indices are not shown as all parameters tested gave a zero value.

4.4.7 Flow Rate

The rate at which the solvent passes through the system also affects chromatographic and ionisation performance and is therefore a crucial parameter to be optimised as part of method development. A range of flow rates were applied to the method with multiple outcomes compared (**Figure 6**). The number of metabolites detected was significantly

higher at 0.4 mL/min when compared to all other flow rates investigated. For the average peak quality of all putatively identified metabolites, an increasing trend was observed, with 0.4 mL/min showing a significantly higher peak quality in comparison to the lower flow rates, but lower when compared to the higher flow rates. When looking into detection quality of creatinine, we can see that 0.4 mL/min shows the highest AUC and a significantly higher peak rating than 0.25 mL/min. The jaggedness was the only PQF that demonstrated differential significance, with 0.25 mL/min showing a higher value than all other flow rates. Due to the significantly increased metabolite detection, high performing average peak and creatinine peak data, and prevention of column overload, 0.4 mL/min was chosen as the optimal flow rate parameter for the current method. This view is supported by a range of studies noting the use of a 0.4 mL/min flow rate for LC-MS methods analysing urine samples [28-30]. A comparison of the chromatograms obtained from the analysis of different flow rates is shown in **Supplementary Information Figure S5.**

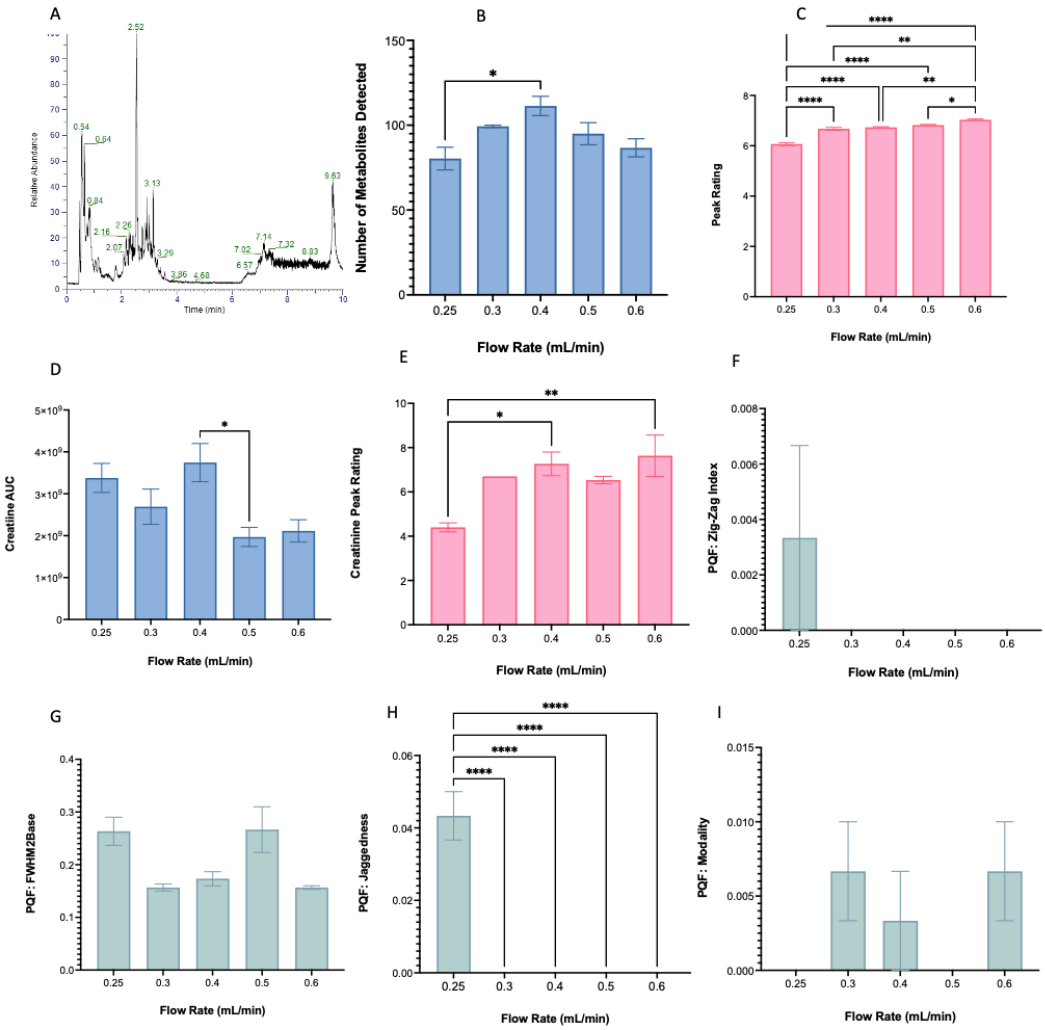


Figure 6. The impact of flow rate on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** Zigzag, **(G)** FWHM2Base, **(H)** Jaggedness, and **(I)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method.

4.4.8 Gradient Curve

During the chromatographic separation process, the gradient of the mobile phase is represented by the composition change of solvent A (in this case, water + 0.1% FA) and solvent B (acetonitrile + 0.1% FA) over time. The gradient curve refers to the curve shape parameter, defining how the solvent gradient progresses throughout the method, and the

rate of solvent composition change. The value of the gradient curve is therefore important to optimise in an LC-MS method, as it directly impacts the retention time and separation efficiency of detected metabolites. A gradient curve of 3, 5, and 7 were applied to the method and analysed (**Figure 7**). There were no significant differences observed in any of the outcomes measured for the gradient curve analysis, with a value of three, five, and seven showing a similar metabolite detection and peak performance. Selection for this parameter was therefore made based on chromatographic performance, with a gradient curve of seven chosen due to an overall higher signal intensity of the metabolites detected and better resolution of peaks from visual inspection. Details of the final chromatography elution gradient, including each of the liquid-chromatography method parameters optimised, are shown in **Supplementary Information Table S4** and a comparison of the chromatograms obtained from the analysis of different gradient curves shown in **Supplementary Information Figure S6.**

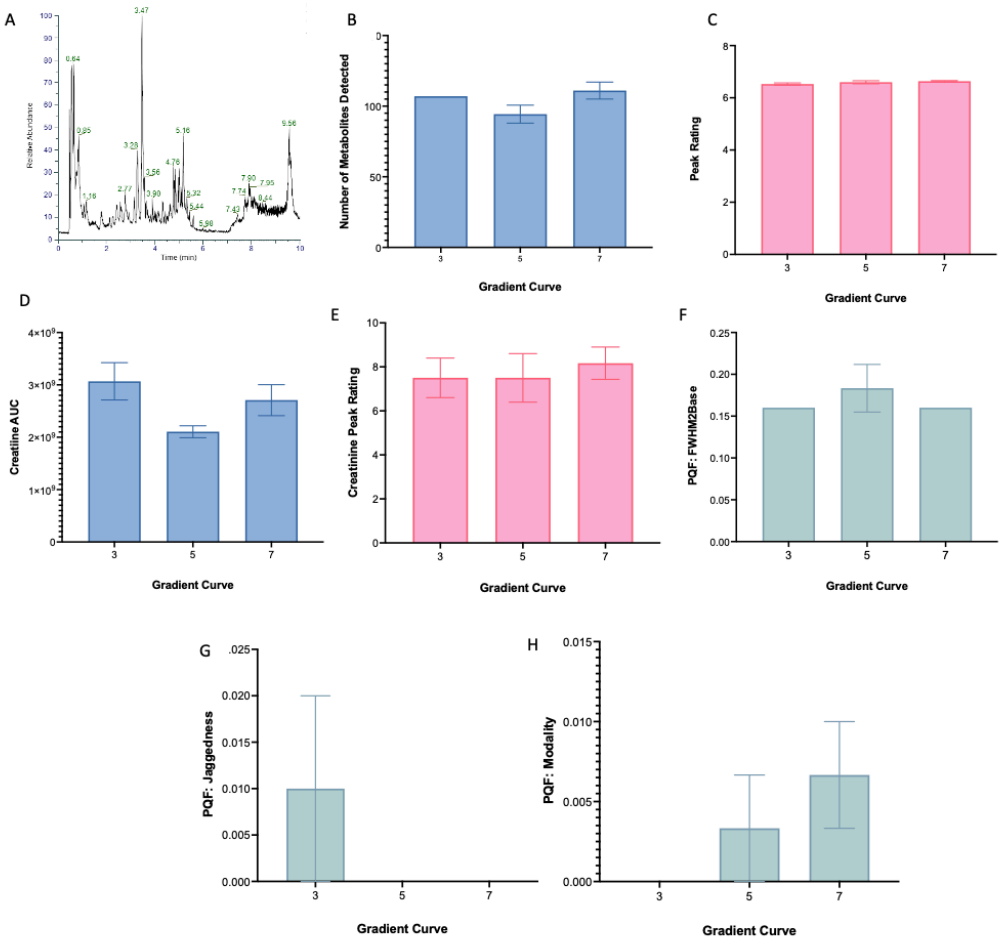


Figure 7. The impact of gradient curve on untargeted urinary metabolomics. Outcomes were assessed by **(A)** chromatographic visualisation, **(B)** the number of metabolites detected, **(C)** their average peak rating, **(D)** the area under the curve (AUC) of the detected creatinine peak and **(E)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. The chromatogram presented in (A) represents the optimised method parameter selected for the method. The zig-zag indices are not shown as all parameters tested gave a zero value.

4.4.9. Mass Spectrometer Parameter Optimisation

4.4.10 Automatic Gain Control (AGC)

The automatic gain control (AGC) is used for controlling the number of ions that are accumulated during a survey scan in the orbitrap analyser. A higher target value refers to a longer accumulation time and therefore implicates an increased analysis time with improved sensitivity, and vice versa. This accumulation time, also known as transient time, is a critical factor in determining the resolution of the mass analyser. Longer transient times allow more precise measurement of ion oscillations, increasing resolution and enabling more accurate assignment of m/z values to detected ions. However, this sacrifices the cost of throughput, as longer transients reduce the number of spectra required per unit time, linking AGC settings closely with both data quality and acquisition speed. The maximum ion injection time setting controls the maximum time the instrument will spend accumulating ions for each scan, which requires a balance between ensuring enough time for appropriate ion accumulation and speed of overall analysis. Optimisation of the specific AGC target value and maximum ion injection time settings is a powerful tool to control the analytical output and peak parameters. We therefore performed method optimisation of AGC parameters for both full and MSMS scans.

The full scan and MSMS AGC parameters were optimised, with all potential setting combinations on the Orbitrap 240 analysed and compared (**Figure 8**). This included four AGC target values, expressed as percentage AGC, and four maximum injection time values. Analysis was further extended to optimisation of MSMS AGC parameters, which similarly involved investigation into all potential settings on the Orbitrap 240 (four AGC target values and four maximum injection time values) for the tandem mass spectrometry settings. Detailed parameter analysis of the full scan and MSMS AGC optimisation are displayed in **Figures 9-10**.

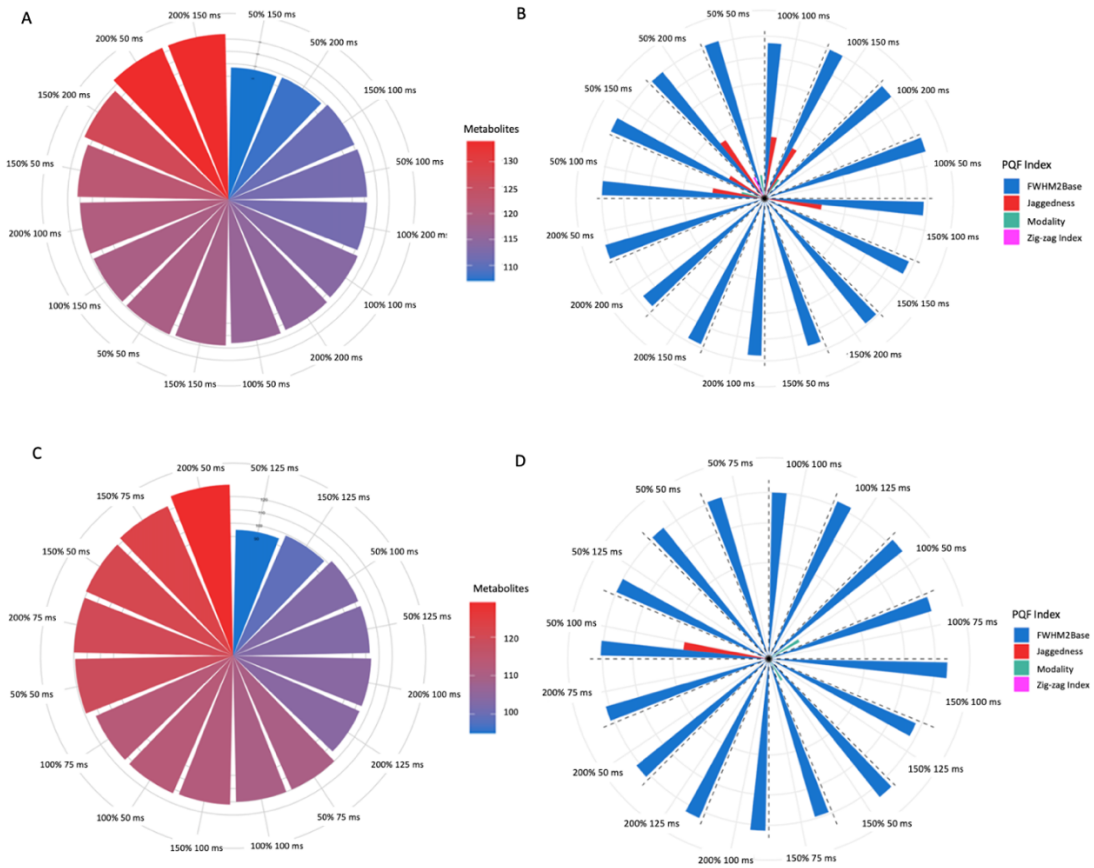


Figure 8. Comparison of AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected per full scan AGC parameter, **(B)** the associated peak performance of creatinine, as measured by the quantification of peak quality factors (PQF)s, Zigzag, FWHM2Base, Jaggedness, and Modality indices, **(C)** the number of metabolites detected per MSMS scan AGC parameter, and **(D)** peak performance characteristics for the MSMS scan settings. The zig-zag indices are not shown as all parameters tested gave a zero value.

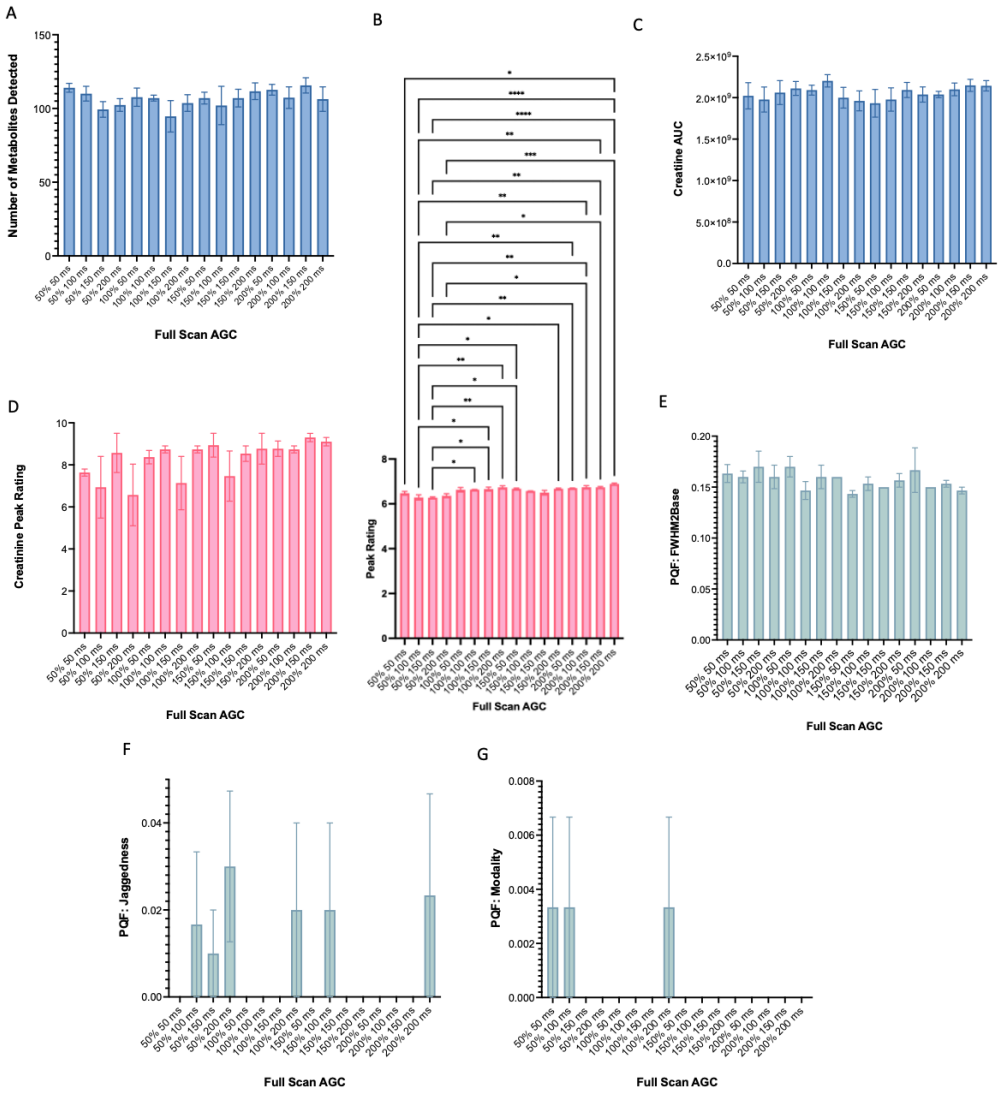


Figure 9. Comparison of full scan AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** FWHM2Base, **(F)** Jaggedness, and **(G)** Modality indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

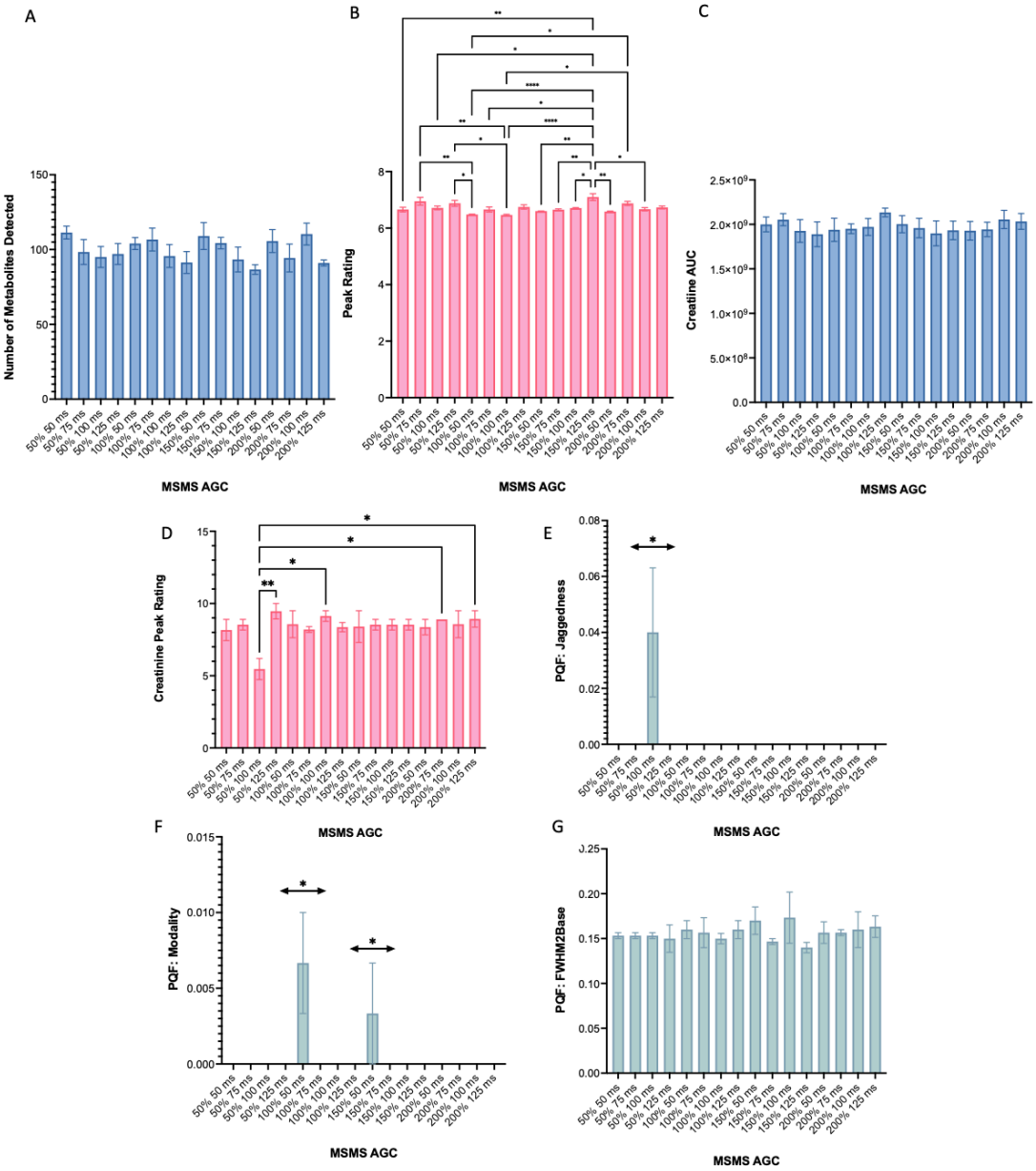


Figure 10. Comparison of MSMS scan AGC parameters and their impact on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance of creatinine was further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** Jaggedness, **(F)** Modality, and **(G)** FWHM2Base indices. The zig-zag indices are not shown as all parameters tested gave a zero value.

Notably, no significant differences were observed in the number of metabolites detected, the creatine AUC and peak rating, or PQFs for the full scan AGC analysis, however, we can comment on some data trends. When comparing the four AGC parameters, 50% and 100% target values showed a more variable peak performance when looking at the PQF analysis, which may be explained by insufficient ion accumulation at these lower values. On the other hand, the highest AGC target of 200% resulted in a slight increase in signal intensity but did not significantly improve the number of metabolites detected. Alongside analysis of the chromatograms obtained from each parameter, the decision was made to select an AGC target of 150%, as driven by its peak performance, signal intensity, and metabolite detection. Similarly, the choice of a maximum injection time of 100 ms provided a favourable balance between ion collection and spectral acquisition speed. Overall, a full scan AGC target of 150% with a 100 ms maximum injection time was selected to be the most effective configuration for this method.

For the MSMS scan analysis, it was observed that increasing the AGC target above 100% resulted in a slight increase in signal intensity, however it did not significantly enhance the overall number of metabolites detected. This suggests that ion saturation may occur at higher MSMS AGC settings [31]. A 100% AGC target is therefore suggested for optimal ion accumulation without the increased risk of saturation. A maximum injection speed of 50 ms maintained high PQFs, particularly with respect to the zigzag, jaggedness, and modality metrics which consistently showed zero values. An MSMS AGC target of 100% with a 50 ms maximum injection time was therefore selected to be the most effective configuration, which completed the parameter optimisation to finalise the method. The final optimised method for untargeted urinary UHPLC-MS analysis includes a streamlined and shortened preparation and analysis process with improved metabolite detection and peak performance metrics. A chromatographic overview of the two methods is shown in **Figure 11**.

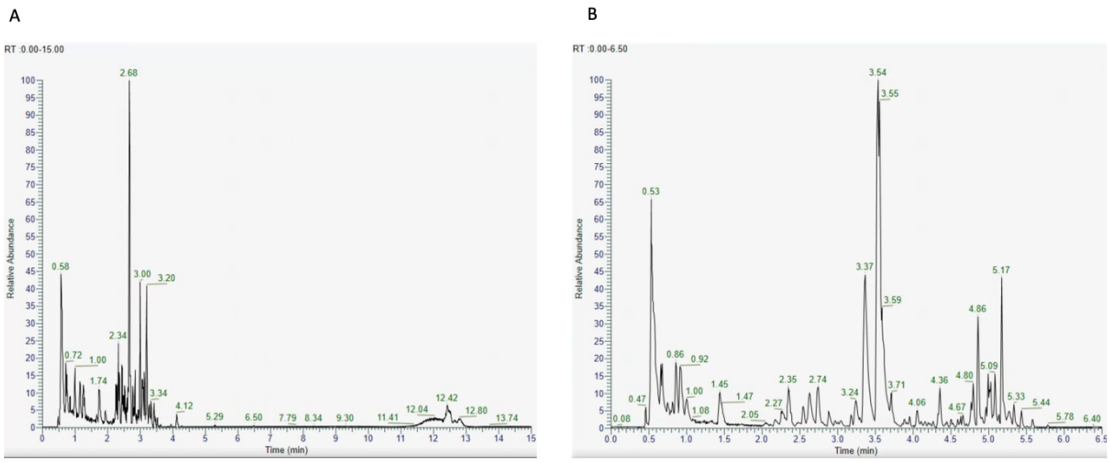


Figure 11. Method Optimisation Overview. Comparison of **(A)** the original 15-minute method and **(B)** shortened 10-minute chromatography gradient.

4.4.11 Application of the Method to a Clinical Trial

The optimised method was subsequently applied to a clinical study which analysed 1094 urine samples from healthy individuals and IBD patients, making this the world's largest study to date investigating urinary metabolomics of IBD. The whole trial consisted of three different patient cohorts: The ENIP (Exclusive or partial enteral nutrition in healthy individuals) study investigated the effect of enteral nutrition, a nutritionally complete liquid diet used as the primary treatment for patients with paediatric Crohn's disease [32], and explored the impact on healthy individuals, and involved the analysis of 107 urine samples. The CD-TREAT (Crohn's Disease Treatment-with-EATing) diet is a prescriptive and personalised diet which recreates Exclusive Enteral Nutrition (EEN). The optimised urine method was applied to investigate disease outcomes and prediction response of CD-TREAT using UHPLC-MS, where 125 urine samples were analysed. The intensive Post Exclusive Enteral Nutrition Study (iPENS) study, a randomised trial to evaluate CD-TREAT diet as a food reintroduction regime in children and young adults with Crohn's disease, looked into metabolic signatures in different biomatrices including urine. 862 urine samples were analysed as part of the iPENS project using the optimised method. In total, 1094 urine samples were analysed throughout the clinical trial using the method presented in this paper, as displayed in **Figure 12**.

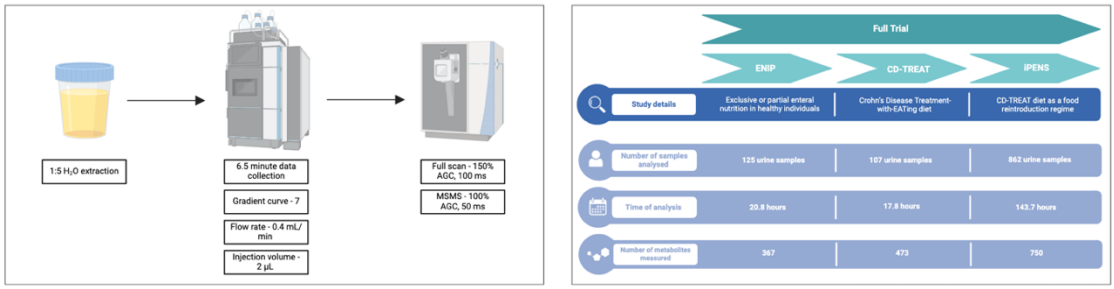


Figure 12. Final optimised workflow for the analysis of urine samples using untargeted UHPLC-MS and application of the optimised method to a clinical trial for urine metabolomics analysis. Schematic representation of the optimised method, which includes the selection of specific sample preparation and LC-MS parameters to ensure the comprehensive and reproducible profiling of urinary metabolites. Illustration of the clinical trial workflow, consisting of three independent studies which aimed to measure the global urine metabolic profile. The total analysis time of the urine samples for all three studies was 182.3 hours and over 1500 metabolites were putatively identified.

In comparison to the previous method which had a run time of 15 minutes and equating to 273.5 hours of analysis time, the optimised 10-minute method allowed all samples to be run in 182.3 hours, saving a total of over 91 hours of LC-MS analysis time throughout the clinical trial and therefore demonstrating an improvement in method throughput, without sacrificing data quality. This has important implications for future large-scale studies, where time of analysis is an essential factor to prevent sample degradation during the duration of sample preparation and analysis. Many existing large scale urinary metabolomics studies to date have used longer chromatography elution gradients. For example, the analysis of urine samples from 348 healthy children and 315 adults characterising age- and sex-dependent metabolic variations used an 18-minute chromatography method [33], a 28.5-minute elution gradient was used to identify urinary biomarkers of type 2 diabetes patients [34], and a 29-minute elution gradient was used for the comprehensive profiling of the normal human metabolome using 663 urine samples [35]. This demonstrates the varied application of urinary LC-MS studies and the potential for method standardisation, efficiency improvement, and cost reduction.

4.4.12 Comparison of Urine Normalisation Strategies for Adjusting Urine Dilution

Normalisation is essential in urinary metabolomics to correct for differential sample dilution, particularly in untargeted LC-MS workflows. Creatinine normalisation and probabilistic quotient normalisation (PQN) are two widely used approaches for adjusting urine data post-acquisition. Creatinine normalisation corrects for dilution based on the concentration of urinary creatinine, while PQN adjusts each sample based on the overall distribution of metabolites relative to quality control (QC) reference samples. In this study, we compared both approaches alongside non-normalised data to evaluate their impact on data quality. To assess the global impact of normalisation on data structure, principal component analysis (PCA) was performed on each dataset. Minimal differences were observed in the PCA plots between the three conditions (**Supplementary Figure S7**), suggesting that PCA-based evaluation alone is insufficient to differentiate between data quality of the normalisation methods.

The coefficient of variation (CV) was calculated for each metabolite across QC samples to evaluate the precision of each normalisation strategy. A CV threshold of ≤ 20% was applied, and the number of metabolites that met the cut-off using each normalisation method were quantified (**Figure 13**). PQN normalisation showed the highest number of metabolites that met this criterion (n = 320), followed by the non-normalised data (n = 308). In contrast, creatinine normalisation resulted in substantially fewer metabolites below the threshold (n = 138). These findings suggest that PQN normalisation offers superior performance in terms of analytical precision as measured by CV. Detailed CV values for the detected metabolites (MSI level 2) using each of the normalisation methods are provided in **Supplementary Information Table S5**.

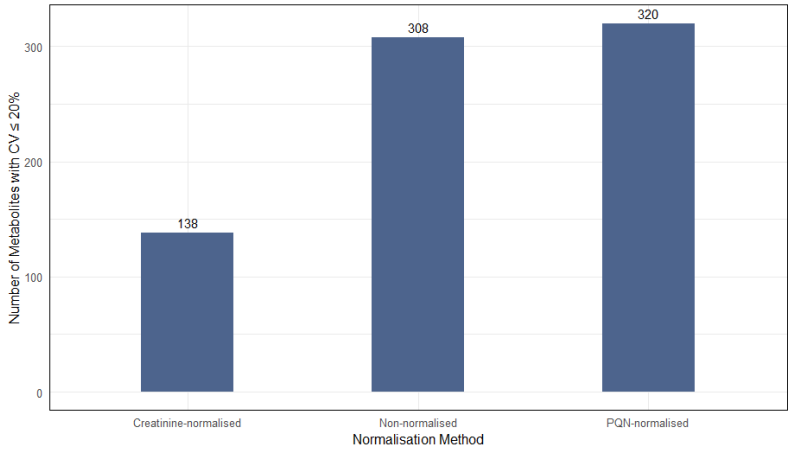


Figure 13. Number of metabolites with coefficient of variation (CV) ≤ 20% across three normalisation methods. The bar plot shows the total number of metabolites meeting a CV threshold of ≤ 20% for non-normalised data, creatinine normalised data, and probabilistic quotient normalisation (PQN) normalised data.

While post-acquisition normalisation strategies such as creatinine adjustment and PQN are valuable for mitigating dilution effects, they also have inherent limitations. These include their reliance on assumptions about the stability of reference metabolites or the representativeness of QC samples. For this reason, pre-acquisition methods such as specific gravity or osmolality measurements can be employed to correct for urine dilution at the point of sample preparation prior to LC-MS analysis. Such approaches directly account for sample concentration before analysis, providing a more physiologically meaningful normalisation framework. Nonetheless, post-acquisition normalisation remains a practical and widely adopted solution in large-scale studies, where pre-acquisition measurements may not be feasible due to time constraints.

4.5 Conclusions

The main purpose of this study was to develop a rapid and reliable method for large-scale comprehensive LC-MS metabolic profiling. We present a 10-minute UHPLC-MS method, comprised of a 6.5-minute data collection time and a 3.5-minute cleaning step, based on urine samples extracted using the dilute-and-shoot method. Optimisation of eight individual method parameters were sequentially carried out to provide an overall method which effectively detects a large number of metabolites of high data quality, as determined by a range of peak performance quality factors. Additionally, probabilistic quotient normalisation (PQN) was identified as the most effective post-acquisition strategy for

correcting urine dilution, compared to creatinine or no-normalisation. This study successfully investigated the possibility to reduce the overall UHPLC-MS method analysis time for untargeted urinary metabolomics, which has positive implications for high-throughput and large-scale clinical trials, as demonstrated through the application of the optimised method to large urinary metabolomics study.

Funding: This study was supported by Shimadzu UK and the University of Strathclyde through joint contribution to P.E.K.'s studentship.

Institutional Review Board Statement: The iPENS and CD-TREAT studies were approved by the Institutional Review Board (or Ethics Committee) of NHS West of Scotland Research Ethics Committee (Ref: 11/WS/0006, REC reference: 17/WS/0119, 19/WS/0163) and NHS Research and Development office. For ENIP the original approval was granted by the College of Medical, Veterinary and Life Sciences Ethics Committee, University of Glasgow (Project number: 200220086). All studies were registered on https://clinicaltrials.gov (identifier NCT number: IPENS, NCT04225689; CD-TREAT, NCT03171246; ENIP, NCT06828094).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: N.J.W.R. want to acknowledge the Strathclyde Centre of Molecular Bioscience (www.scmb.strath.ac.uk) for access to LCMS instrumentation.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

4.6 References

1. Thomas, C.E., et al., *Urine collection and processing for protein biomarker discovery and quantification.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(4): p. 953-9.
2. Hon, C.-Y. and N. Motiwala *Biological Monitoring via Urine Samples to Assess Healthcare Workers' Exposure to Hazardous Drugs: A Scoping Review*. Applied Sciences, 2022. **12**, DOI: 10.3390/app122111170.
3. Lezaic, V., *Albuminuria as a Biomarker of the Renal Disease*, in *Biomarkers in Kidney Disease*, V.B. Patel, Editor. 2015, Springer Netherlands: Dordrecht. p. 1-18.
4. Chen, T.K., D.H. Knicely, and M.E. Grams, *Chronic Kidney Disease Diagnosis and Management: A Review.* Jama, 2019. **322**(13): p. 1294-1304.
5. Merola, R., et al., *PCA3 in prostate cancer and tumor aggressiveness detection on 407 high-risk patients: a National Cancer Institute experience.* Journal of Experimental & Clinical Cancer Research, 2015. **34**(1): p. 15.
6. Opoku Mensah, B., et al., *Urinary PCA3 a Superior Diagnostic Biomarker for Prostate Cancer among Ghanaian Men.* Dis Markers, 2022. **2022**: p. 1686991.
7. Kabasawa, K., et al., *Association of metabolic syndrome traits with urinary biomarkers in Japanese adults.* Diabetol Metab Syndr, 2022. **14**(1): p. 9.
8. Duggins-Warf, M., et al., *Disease specific urinary biomarkers in the central nervous system.* Scientific Reports, 2023. **13**(1): p. 19244.
9. Fathi, A.T., et al., *Elevation of Urinary 2-Hydroxyglutarate in IDH-Mutant Glioma.* Oncologist, 2016. **21**(2): p. 214-9.
10. Riviere-Cazaux, C., et al., *Cerebrospinal fluid 2-hydroxyglutarate as a monitoring biomarker for IDH-mutant gliomas.* Neuro-Oncology Advances, 2023. **5**(1): p. vdad061.
11. Wang, J.-H., et al., *Prognostic significance of 2-hydroxyglutarate levels in acute myeloid leukemia in China.* Proceedings of the National Academy of Sciences, 2013. **110**(42): p. 17017-17022.
12. Joshi, N., et al., *Recent progress in mass spectrometry-based urinary proteomics.* Clinical Proteomics, 2024. **21**(1): p. 14.
13. van Duijl, T.T., et al., *Multiplex LC-MS/MS Testing for Early Detection of Kidney Injury: A Next-Generation Alternative to Conventional Immunoassays?* The Journal of Applied Laboratory Medicine, 2022. **7**(4): p. 923-930.

14. Zhao, B., et al., *A dilute-and-shoot liquid chromatography–tandem mass spectrometry method for urinary 18-hydroxycortisol quantification and its application in establishing reference intervals.* Journal of Clinical Laboratory Analysis, 2022. **36**(8): p. e24580.
15. deventer, K., et al., *Dilute-and-shoot-liquid chromatography-mass spectrometry for urine analysis in doping control and analytical toxicology.* TrAC Trends in Analytical Chemistry, 2013. **55**.
16. Gracia-Marín, E., et al., *Dilute-and-shoot approach for the high-throughput LC-MS/MS determination of illicit drugs in the field of wastewater-based epidemiology.* Water Research, 2024. **259**: p. 121864.
17. Tamama, K., *Dilute and shoot approach for toxicology testing.* Front Chem, 2023. **11**: p. 1278313.
18. Castilla-Fernández, D., et al., *Dilute-and-shoot versus clean-up approaches: A comprehensive evaluation for the determination of mycotoxins in nuts by UHPLC-MS/MS.* LWT, 2022. **169**: p. 113976.
19. *Advanced Mass Spectral Database*. Accessed 01.2025; Available from: https://www.mzcloud.org/.
20. Chemistry, R.S.o. *ChemSpider*. 2025; Available from: https://www.chemspider.com/?gad_source=1&gclid=Cj0KCQiAwOe8BhCCARIsAGKeD56dnavRm_pXFE8dcHA5ggaOcPB5JPS_T0IfOneyhPiL6qmtrAFY5nMaAnqiEALw_wcB.
21. Hou, H., et al., *LC-MS-MS Measurements of Urinary Creatinine and the Application of Creatinine Normalization Technique on Cotinine in Smokers' 24 Hour Urine.* J Anal Methods Chem, 2012. **2012**: p. 245415.
22. Scientific, T. *Compound Discoverer User Guide for LC Studies Software Version 3.3 SP2*. 2023; Available from: https://assets.thermofisher.com/TFS-Assets/CMD/manuals/XCALI-98478-Compound-Discoverer-User-Guide-LC-Studies-XCALI98478-en.pdf.
23. Cao, Z., E. Kaleta, and P. Wang, *Simultaneous Quantitation of 78 Drugs and Metabolites in Urine with a Dilute-And-Shoot LC-MS-MS Assay.* Journal of analytical toxicology, 2015. **39**.

24. Hemmer, S., et al., *Impact of four different extraction methods and three different reconstitution solvents on the untargeted metabolomics analysis of human and rat urine samples.* Journal of Chromatography A, 2024. **1725**: p. 464930.
25. Scheijen, J.L., et al., *L(+) and D(-) lactate are increased in plasma and urine samples of type 2 diabetes as measured by a simultaneous quantification of L(+) and D(-) lactate by reversed-phase liquid chromatography tandem mass spectrometry.* Exp Diabetes Res, 2012. **2012**: p. 234812.
26. Wei, B., et al., *A high-throughput robotic sample preparation system and HPLC-MS/MS for measuring urinary anatabine, anabasine, nicotine and major nicotine metabolites.* Clin Chim Acta, 2014. **436**: p. 290-7.
27. Al-Saffar, Y., N.N. Stephanson, and O. Beck, *Multicomponent LC-MS/MS screening method for detection of new psychoactive drugs, legal highs, in urine-experience from the Swedish population.* J Chromatogr B Analyt Technol Biomed Life Sci, 2013. **930**: p. 112-20.
28. Liu, R., et al., *Simultaneous determination of corynoline and acetylcorynoline in human urine by LC-MS/MS and its application to a urinary excretion study.* J Chromatogr B Analyt Technol Biomed Life Sci, 2016. **1014**: p. 83-9.
29. Leninskii, M.A., et al., *Determination of 15 Functional State Biomarkers in Human Urine by High-Performance Liquid Chromatography with Tandem Mass Spectrometric Detection.* Journal of Analytical Chemistry, 2023. **78**(10): p. 1344-1354.
30. Ming, D.S. and J. Heathcote, *A rapid and accurate UPLC/MS/MS method for the determination of benzodiazepines in human urine.* J Chromatogr B Analyt Technol Biomed Life Sci, 2011. **879**(5-6): p. 421-8.
31. Hall, A.B., et al., *Extending the dynamic range of the ion trap by differential mobility filtration.* J Am Soc Mass Spectrom, 2013. **24**(9): p. 1428-36.
32. Connors, J., et al., *Exclusive Enteral Nutrition Therapy in Paediatric Crohn's Disease Results in Long-term Avoidance of Corticosteroids: Results of a Propensity-score Matched Cohort Analysis.* J Crohns Colitis, 2017. **11**(9): p. 1063-1070.
33. Liu, X., et al., *Characterization of LC-MS based urine metabolomics in healthy children and adults.* PeerJ, 2022. **10**: p. e13545.

34. Chen, C.-J., et al., *Identification of Urinary Metabolite Biomarkers of Type 2 Diabetes Nephropathy Using an Untargeted Metabolomic Approach.* Journal of Proteome Research, 2018. **17**(11): p. 3997-4007.

35. Xu, J., et al. *A Comprehensive 2D-LC/MS/MS Profile of the Normal Human Urinary Metabolome*. Diagnostics, 2022. **12**, DOI: 10.3390/diagnostics12092184.

5.0 CHAPTER 5

Understanding Food Additives in Inflammatory Bowel Disease: Challenging Perceptions to Improve Gastrointestinal Health

Patricia Kelly[1], Zahra Rattray[1], Konstantinos Gerasimidis[2], Nicholas JW Rattray[1]

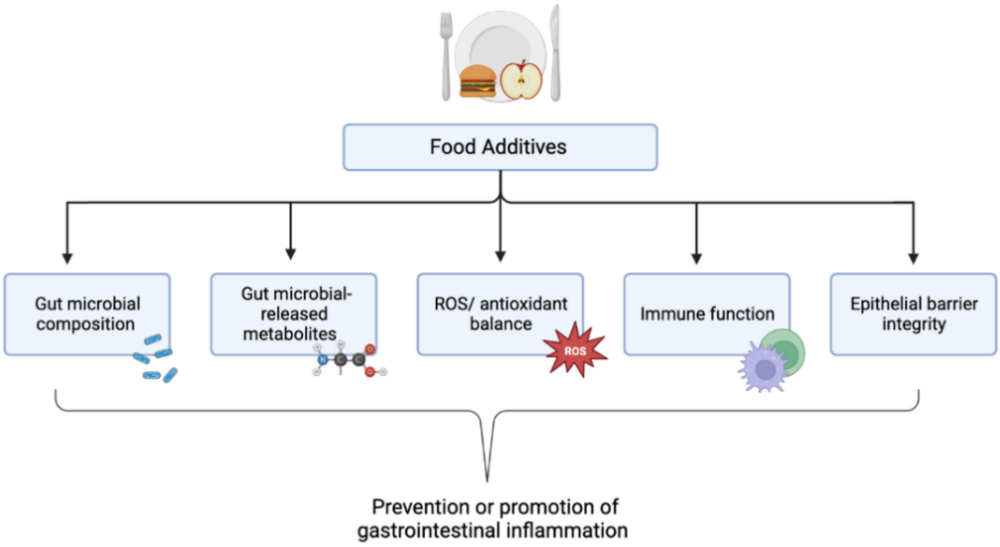
[1] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, UK, G4 0RE

[2] Human Nutrition, School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow, UK, G12 8QQ

Author Contributions: Conceptualization, P.E.K., N.J.W.R.; methodology, P.E.K., N.J.W.R., investigation, P.E.K., N.J.W.R.; writing—original draft preparation, P.E.K.; writing—review and editing, P.E.K., Z.R K.G. and N.J.W.R.; visualisation, P.E.K. and N.J.W.R.; supervision, N.J.W.R., K.G.; project administration, N.J.W.R.; funding acquisition, N.J.W.R and K.G. All authors have read and agreed to the published version of the manuscript.

5.1 Abstract

Current nutritional recommendations for inflammatory bowel disease (IBD) patients are limited due to challenges in translating preclinical nutritional research into successful clinical trials. The increasing body of evidence associating food additives to the western diet-driven surge of IBD may have contributed to the adoption of highly restrictive elimination diets, which are often not practical in the long-term. However, a strong body of emerging evidence suggests that many food additives have beneficial effects with therapeutic potential for IBD. This review explores the physiological mechanisms of food additives shown to specifically affect gastrointestinal (GI) inflammation from both a pro- and anti-inflammatory viewpoint, thus providing a comprehensive review of additives in relation to their role in IBD. We highlight that existing research into the effects of food additives on gut health is predominantly based on evidence from in vitro and in vivo studies, with limited studies carried out in humans. While the potential mechanisms of certain food additives are known, randomised controlled trials (RCTs) in humans are required to confirm their potential implications in IBD. In this review we demonstrate that food additives have widespread roles in preclinical models of IBD of both beneficial and harmful nature and their effects should therefore not be generalised in human health. Pending further human studies, the roles of food additives may bring new perspectives on diet in IBD and contribute to the continually changing dynamics of food additive frameworks within the food industry.



Graphical Abstract. Mechanisms of food additives in the prevention or promotion of gastrointestinal inflammation. Food additives have been shown to influence gut microbial composition and their released metabolites, reactive oxygen species (ROS) and antioxidant balance, immune function, and epithelial barrier integrity.

5.2 Introduction

Intestinal inflammation and the functions of the digestive tract are strongly influenced by diet, rendering it a significant modifiable risk factor in disease [1]. Common themes have emerged from studies linking diet to gastrointestinal (GI) diseases in recent years, and of significant importance is the association with the Western diet, a diet characterised by low intake of fruits and vegetables and high sugar, salt and saturated fat content. The increased adoption of Western dietary habits has paralleled the rise in dietary-associated intestinal disorders, including inflammatory bowel disease (IBD) [2]. IBD is a multifaceted disease with a complex pathogenesis involving genetic and environmental factors. The aetiology of IBD remains unclear, however gut microbial dysbiosis [3-6], epithelial barrier dysfunction [7-9], and pro-inflammatory immune activation [10, 11] have been evidenced as important mechanisms of disease initiation and progression. In 2017, 6.8 million individuals were estimated to be affected by IBD globally [12], representing a 20% increase in cases since 2004, which is projected to rise at an annual rate of 5% [13].

An aspect of the Western diet that has received increasing interest is the impact of food additives and their role in disease pathophysiology [14-16]. Food additives are substances that are added to enhance one or more properties of a food product and with improved marketability through enhanced taste, visual appearance and increased shelf life, the use of additives is increasingly popular within the food industry. The demand for food additives continues to rise, with the global market value sized at $107 billion in 2022 and projected to be worth $176.79 billion by 2028 [17]. In Europe, substances approved for use as food additives are given specified Europe numbers (E numbers), which can be located on food labels. Mounting evidence has linked food additives to GI inflammation and disease; with the rise in food additives in recent years being observed alongside an increased incidence of diet-related diseases [13, 14]. The role of diet and food additives is therefore an important consideration for both development and management of IBD. Nutritional epidemiological studies have revealed associations between dietary patterns and IBD risk [18-20], however direct links with specific food additives are less characterised. When considering the role of food additives on inflammation, researchers aim to understand additives that initiate and exacerbate disease state, with current reviews considering food additives with therapeutic relevance being limited in number. This is important as there remains an unmet need for patient dietary advice and therapies, predominantly due to the lack of translation from preclinical studies to human randomised control trials (RCTs).

This review provides a comprehensive analysis of both inflammatory and therapeutic food additives in the context of IBD. The food additives discussed in this review are authorised for use in the European Union (EU). We present current evidence of the mechanisms of food additives which specifically promote or reduce gastrointestinal inflammation and discuss future directions to inform the design of dietary interventions in the management of IBD.

5.3 Methods

A PRISMA designed literature search was performed in PubMed and Google Scholar and articles were screened for eligibility (**Figure 1)**. Search terms included “food additives,” “food colourants “preservatives,” “antioxidants,” “sweeteners,” “emulsifiers,” “thickeners,” “stabilisers,” “Inflammatory Bowel Disease,” “gastrointestinal inflammation,” “Crohn’s Disease,” “Ulcerative Colitis.”

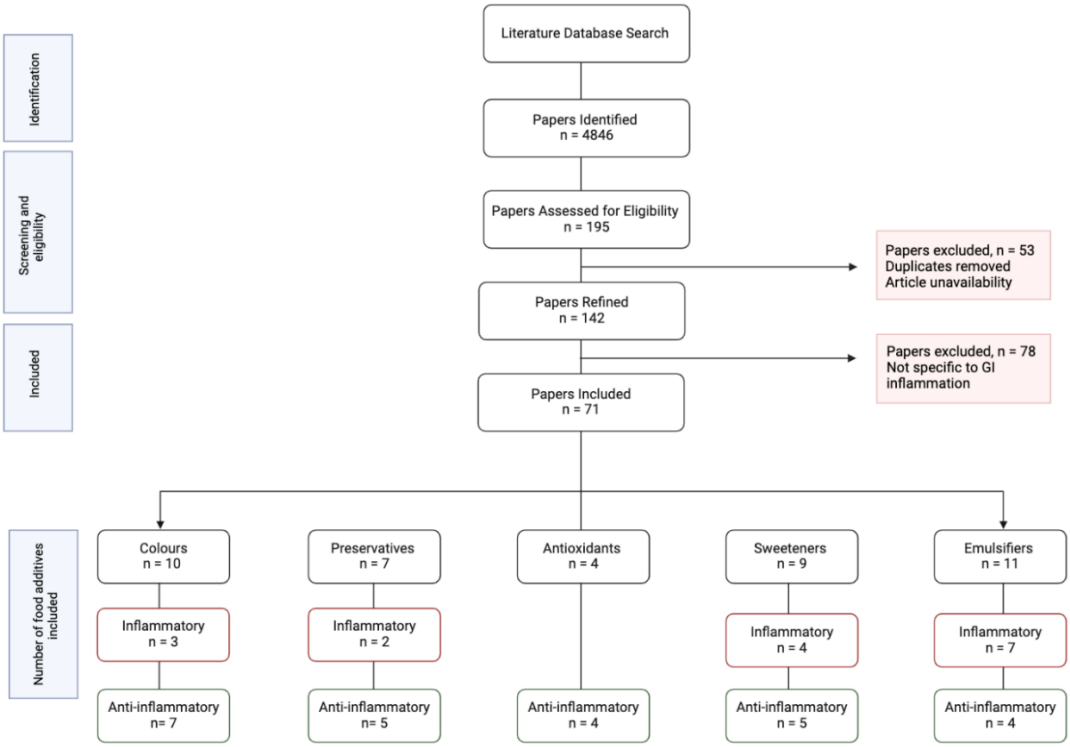


Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram of literature database search. Selection criteria including paper identification, screening and eligibility, excluding duplicate papers, unavailable papers, and those unspecific to GI inflammation, and final selection of articles included in the review, according to food additive classification.

5.4 Food Additives

Diet may represent a non-pharmacologic strategy to manage intestinal inflammation, however, due to the diversity of food product availability and interindividual differences in the GI tract, elucidating the role of diet remains a complex factor to establish.

Food additives can be categorised according to their functional role in a food product, including colourants, preservatives, sweeteners, and emulsifiers. Controlled safety assessments determine an accepted daily intake (ADI) for each additive, which is based on the lowest no observed adverse effect level (NOAEL) obtained from longitudinal in vivo studies [21]. ADI levels are evaluated by regulatory authorities, the European Food Safety Authority (EFSA-EU) and the Joint Expert Committee on Food Additives (JECFA) and World Health Organisation (WHO) internationally. In parallel with regulatory assessment, analytical methodologies have been developed to accurately measure the presence of food additives in complex food matrices. Among these, liquid chromatography coupled with mass spectrometry (LC-MS) has emerged as a widely accepted technique, which enables the simultaneous detection and quantification of multiple additives.

Over 300 food additives are currently authorised for consumption in Europe, and a growing number of research studies have explored their physiological roles in the context of gastrointestinal health. Based on existing evidence, food additives can be differentiated according to their effect on intestinal inflammatory state, with experimental and clinical data showing harmful effects of some food additives on intestinal physiology, while others exhibit therapeutic properties (**Figure 2**).

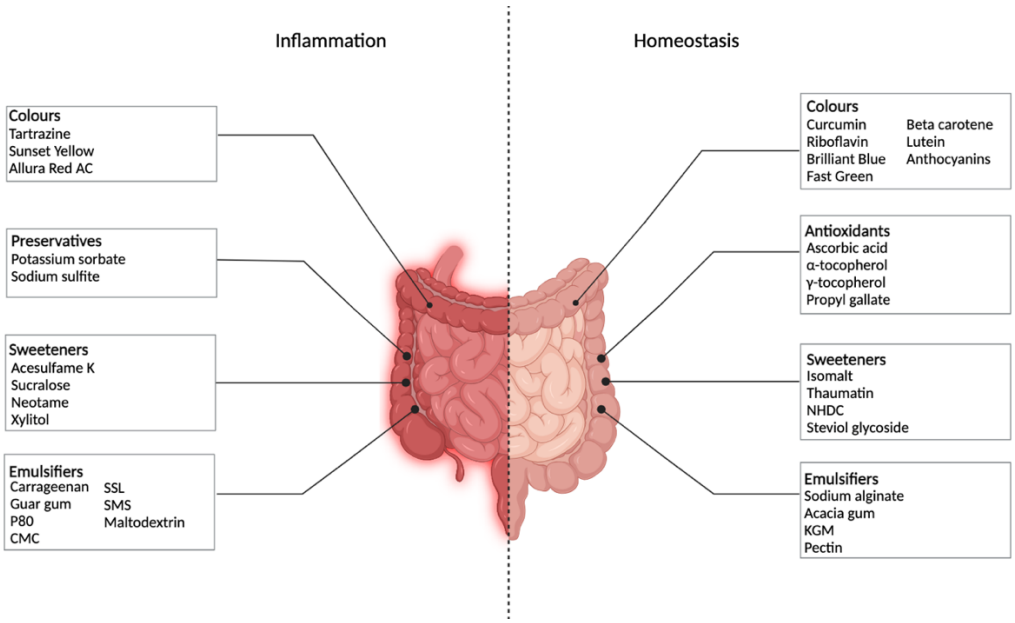


Figure 2. Food additives implicated in gastrointestinal inflammation. Food additives, grouped according to their functional class, (colours, preservatives, antioxidants, sweeteners, and emulsifiers) shown to either enhance or mitigate gastrointestinal inflammation. P80; polysorbate 80, CMC; carboxymethylcellulose, SSL; sodium stearoyl lactylate, SMS; sorbitan monostearate, KGM; konjac glucomannan.

5.5 Food Colours

Colour is a property that motivates consumers to purchase certain foods, thus driving manufacturing processes to enhance products through colour additives [22]. Food dyes of natural or synthetic origin drive a global market of $2.5 billion in 2018, projected to reach $4.77 billion by 2028 [23]. A multitude of studies have demonstrated the impact of different food dyes on intestinal pathophysiology, with different colourants being implicated in inflammatory processes. **Figure 3** shows the variety of food colorants shown to impact intestinal homeostasis and **Supplementary Table 1** documents food colours suggested to exert anti/inflammatory or therapeutic effects on IBD.

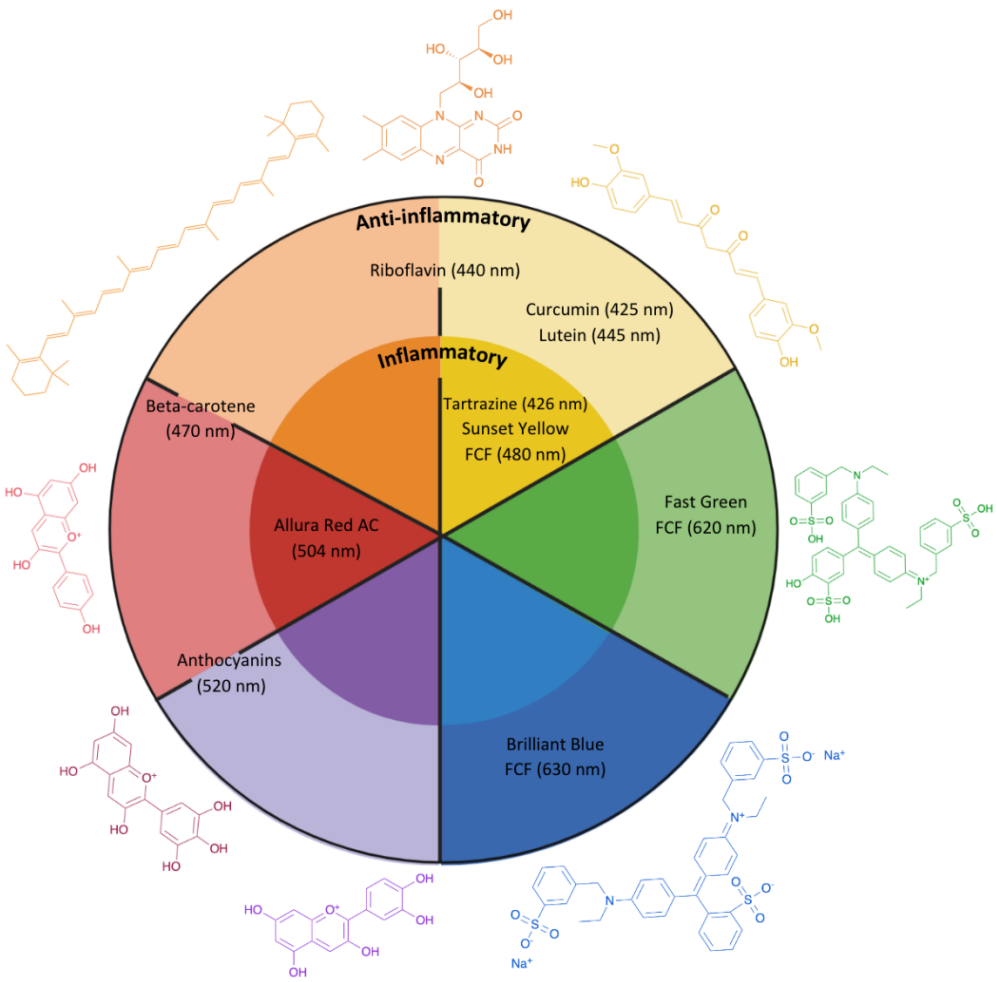


Figure 3. Food colourants impacting gastrointestinal inflammation. The inner circle represents food colours that demonstrate inflammatory properties; outer circle represents food colours that demonstrate therapeutic properties, in the context of GI health. Food additives are shown with their maximum absorbance wavelengths: Riboflavin: 440 nm [24], Curcumin: 425 nm [25], Lutein: 445 nm [26], Fast green FCF: 620 nm [27], Brilliant blue FCF: 630 nm [28], Anthocyanins: 520 nm [29], Beta-carotene: 470 nm [30], Tartrazine: 426 nm [31], Sunset yellow FCF: 480 nm [32], Allura red AC: 504 nm [33].

Beyond classification by colour, food colorants are grouped according to their origin and chemical structure. Azo dyes, characterised by ≥1-N = N- linkage, are used in the food, pharmaceutical, textile, and cosmetics industries [34, 35]. While safety reports have deemed azo dyes such as tartrazine, sunset yellow FCF, and Allura Red AC as safe for human consumption within defined ADI limits, they have been linked to concerns about their carcinogenicity [36, 37]. Their carcinogenicity is largely due to degradation by reduction reactions in intestinal [38] and skin [39] microbiota, resulting in the production of aromatic amines, which are carcinogenic [40].

Food colourants that have demonstrated intestinal inflammatory outcomes include the azo dyes tartrazine (E102), sunset yellow FCF (E110), and Allura Red AC (E129). Tartrazine is a synthetic yellow dye commonly used in confectionary. It has been demonstrated that low tartrazine intake (1.4 mg/kg bodyweight /day) is associated with an increased IL-6 and TNF-α levels (proinflammatory), with a decrease in activities of the antioxidant enzymes catalase (CAT), superoxide dismutase (SOD), and glutathione peroxidase (GSH-Px) [41]. Furthermore, tartrazine intake resulted in altered intestinal microbiota, as characterised by reductions in *Bacteroides*, *Clostridium*, and *Bacillus*. These observations were further supported by investigations demonstrating the induction of colonic DNA damage in ddYmice [42], toxic histological changes in rat gastric mucosa [43], and increased intestinal lymphocytes and eosinophils [44] in response to tartrazine consumption. Sunset yellow, another example of an azo dye, has also shown to exercise profound effects on intestinal inflammation. Its intake has been associated with an increased disease activity index and increased levels of IL-1β and TNF-α in dextran sulfate sodium (DSS)-induced colitis in addition to inhibited growth of murine intestinal organoids [45]. Recently it was shown that prolonged consumption of Allura Red AC caused colitis in C57BL/6 mice, which manifested as damage of the intestinal epithelial barrier through dysregulation of the myosin light chain kinase (MLCK) pathway, an effect which is dependent on the activity of colonic serotonin [46].

In addition to the well documented colour additives with inflammatory potential, data from several studies have identified anti-inflammatory properties of certain food colours. Curcumin (E100), a natural phenolic compound obtained from plants, mediates protective effects via inhibition of the mitogen-activated protein kinase (MAPK) signalling pathway, resulting in the downregulation of pro-inflammatory cytokines including TNF-α [47]. This view is supported by the alleviation of symptoms and intestinal mucosal damage in mice models of IBD upon exposure to the food additive [48]. Additional studies have shown analogous findings, including the downregulation of phosphoinositide 3-kinase (PI3K), extracellular signal-regulated kinase 1 (ERK1), fibronectin 1 (FN1), and TNF superfamily member 1 (TNFS12) [49], and decreases in genes associated with oxidative stress [50]. Importantly, this animal data is corroborated by a number of clinical studies showing the anti-inflammatory potential of curcumin for IBD [51-55]. Riboflavin (E101), vitamin B2, is a yellow-orange food colourant. Daily supplementation of 100 mg riboflavin for three weeks resulted in a reduction in serum IL-2 in CD patients with low FC [56]. In CD patients with high FC, C-reactive protein

(CRP) decreased, while elevated levels of plasma free thiols, indicative of reduced inflammation and oxidative stress were observed. Brilliant Blue (E133) and Fast Green FCF (E143) are selective inhibitors of Pannexin 1 (Panx1) [57], an ATP release channel, which when blocked results in a decrease in the inflammatory response. Therefore, the inhibition of Panx1 via Brilliant Blue or Fast Green can achieve an anti-inflammatory effect. β-carotene (E160a), a precursor to vitamin A, is a naturally occurring orange pigment promoting levels of the tight junction proteins claudin-1 and occludin, and regulating the toll-like receptor 4 (TLR4) signalling pathway in HT-29 cells [58]. Another carotenoid, lutein (E161b), decreases inflammation in DSS-induced colitis via a reduction in IL-6 levels, serum amyloid A (SAA), and myeloperoxidase (MPO), when given in combination with fucoidan [59]. This is confirmed by findings in patients with UC, in which lutein intake was associated with reduced faecal blood and mucus [60]. However, a different study found that lutein is depleted during exclusive enteral nutrition (EEN), a dietary therapy for paediatric CD patients [61] which highlights the current knowledge gap on food additives particularly in the translation from animal models to humans. Anthocyanins (E163) are a group of food colours for which there is increasing evidence for the improvement of intestinal inflammation upon consumption, including mechanisms of enhanced barrier function [62], tight junction regulation [63], and microbial regulation [64]. In Caco-2 cells, anthocyanin supplementation increased GLP-2 and MUC2 levels and prevented TNF-α-induced monolayer permeabilization. Furthermore, a 20 mg/ kg bw/ day dose of bilberry anthocyanin extract, given to female Sprague-Dawley (SD) rats, promoted intestinal barrier function via microbial modulation. When a 50- 200 mg/ kg bw/day anthocyanin dose was given to male pathogen-free C57BL/6J mice, SCFA-producing bacteria (e.g., Ruminococcaceae, Muribaculaceae, Akkermansia) were enriched, alongside increased intestinal tight junction mRNA expression levels, and decreased intestinal permeability. Food colours and their effects on gut inflammation are summarised in **Supplementary Table 1**.

5.6 Preservatives

Preservatives are substances added to foods to minimise or prevent product deterioration caused by microbial growth, oxidation, or physical factors such as temperature and light. Preservatives provide benefits to overall product quality by increasing shelf-life, maintaining taste and appearance, and reducing cost, leading to a global market value of $2.77 billion in

2021 [65]. Various physiological effects to been noted upon their use in different models, some of which have demonstrated harmful effects on the GI system. For example, potassium sorbate (E202) negatively influences the gut microbial composition by decreasing the abundance of specific genera in zebrafish [66]. Similar findings were demonstrated in experimental mouse models, with alterations in the abundance of *Bacteroidetes*, *Verrucomicrobia*, and *Proteobacteria,* contributing to an overall decreased microbial diversity [67, 68]. Sodium sulfite (E221) has also been raised as a safety concern for GI inflammation, due to its microbial diversity diminishing effects including significant reductions in *Faecalibacterium prausnitzii* [69] and *Lactobacillus* species [70]. More recently, it was shown that sodium sulfite exerted cytotoxicity and cell death of rat gastric mucosal cells accompanied by an increase in oxidative stress markers [71].

Research into additional preservatives have generated conflicting results deriving differing outcomes on gut inflammation. Benzoic acid (E210) supplementation improved GI development in weaner pigs, with increased nitrogen digestibility, villous height, and microbial diversity [72]. Moreover, benzoic acid enhanced gut barrier function by stimulating insulin-like growth factor 1 (IGF-1), glucagon-like peptide 2 (GLP-2), zonula occludens-1 (ZO-1), and occludin expression via mammalian target of rapamycin (mTOR) and nuclear factor erythroid 2–related factor 2 (Nrf2) signalling pathways [73]. However, a benzoic acid exclusion diet is a mainstream treatment for orofacial granulomatosis [74], a common reported co-morbidity associated with CD [75]. Evidence indicating the physiological effects of sodium benzoate (E211) are disparate; however, results from experimental IBD models suggest potential anti-inflammatory effects in the gut. In an acetic acid-induced rat model of UC, sodium benzoate decreased levels of MPO and increased levels of GSH, with an associated improvement of disease activity [76]. Detailed examination of the immunological effects of sodium benzoate on THP-1 cells showed inhibition of IL-6 and IL-1β, concluding a major attenuating effect exerted by the additive on the immune response [77]. While controversy surrounds the use of sulphur dioxide (E220) as a food additive, recent data has revealed the reversal of several inflammatory factors that arose upon induction of experimental colitis, including NF-κB activation and increased oxidative stress [78]. From these findings it was suggested that the sulphur dioxide/ glutamate oxaloacetate transaminase pathway may be implicated in IBD. Sodium propionate (E281) and calcium propionate (E282) are both commonly used preservatives in the food industry, with

inconsistencies in the literature regarding their impact on gut function. One study showed that sodium propionate improved experimental colitis, predominantly through restoring intestinal barrier function and decreasing inflammation [79]. This was achieved through inhibiting activation of the STAT3 signalling pathway via downregulation of IL-6, IL-1β, TNF-α, and MPO and increasing SOD and CAT levels. Furthermore, this view is supported by reduced pro-inflammatory and oxidative stress markers upon sodium propionate supplementation [80]. Calcium propionate was shown to reduce plasma IFN-γ and calprotectin in DSS-induced colitis, which resulted in the improved histological scores and overall attenuation of colitis [81]. In contrast, propionate induced virulent properties of CD-associated *Escherichia coli* [82, 83].

Antioxidants are a classification of preservatives, used to prevent or limit food product deterioration. Ascorbic acid (E300), known as vitamin C, is used frequently in the food industry due to its antioxidant properties. One study observed a decrease in inducible nitric oxide synthase (iNOS) and cyclooxygenase-2 (COX-2) proteins in DSS-induced colitis upon ascorbic acid consumption [84]. This effect may be attributed to the inhibition of the nuclear factor kappa B (NF-κB) pathway. Furthermore, levels of pro-inflammatory cytokines decreased with ascorbic acid exposure alongside MPO and malondialdehyde (MDA) activities. These changes paralleled an increase in superoxide dismutase (SOD) and glutathione peroxidase (GPx) activity, evidencing improved physiological effects of IBD resulting from ascorbic acid. This is supported by a further study demonstrating mucosal barrier repair through regulation of tight junction proteins [85]. Following combination treatment of ascorbic acid and vitamin D, ZO-1 mRNA and neurogenic locus notch homolog protein 1 (Notch-1) levels increased, while claudin-2 expression decreased. α- and γ-tocopherol (E307 and E308, respectively) are forms of vitamin E which exhibit intestinal restorative functions: by altering gut bacteria and promoting epithelial barrier integrity, α- and γ- tocopherol can improve colitis in experimental mice models [86]. These results were in accordance with a study carried out in humans to investigate the effects of using d-α-tocopherol as a treatment for mild and moderately active UC [87]. It was found that the average disease activity index (DAI) score was significantly lower after twelve weeks post treatment compared to measurements at baseline, however, no objective inflammatory markers were measured to support this. This effect may be explained by the inhibition of NF-κB, a finding previously observed in a human monocytic cell line [88]. The antioxidant

and food additive propyl gallate (E310) demonstrates GI healing properties via several mechanisms. Levels of ROS and of TNF-α, IL-6, IL-1β, and IFN-γ decreased in mice with DSS-colitis treated with propyl gallate [89] . Treatment also positively affected clinical manifestations, including improvement of DAI score, body weight, and colon length. The effects of preservatives suggested to exert inflammatory or therapeutic effects on IBD are documented in **Supplementary Table 2**.

5.7 Sweeteners

The global market for food sweeteners has been valued at $79 billion in 2021 [90] and is driven by the increasing demand for low sugar and sugar-free products (a response to the obesity pandemic and regulations around sugar tax) several artificial sweeteners have been introduced in food industry. The long-term impact of sweeteners on human health is currently poorly understood, particularly in the long-term. Considerable interest and research have begun to focus on unravelling their effects, and a number of sweeteners have been identified as having a harmful effect to the GI tract, including acesulfame potassium (Ace-K, E950), no-calorie sweetener which increases small intestinal injury in mice [91]. It was shown that expression of pro-inflammatory cytokines and lymphocyte migration to the mucosa increased when mice consumed Ace-K, with decreased expression of glucagon-like peptide-1 and 2 receptors (GLP-1R/ GLP-R2). Microbial dysbiosis was also noted upon treatment with Ace-K. Similarly, sucralose (E955) has been found to result in an alteration in the gut microbiome of male C57BL/6 mice with results showing alterations in several bacterial phylum [92]. Bacterial dysbiosis is therefore one of the proposed mechanisms contributing to the sucralose-induced enhanced susceptibility to DSS-induced colitis. Other mechanisms include a rise in pro-inflammatory cytokines, which, in combination with increased signal transducer and activator of transcription/vascular endothelial growth factor (STAT3/VEGF) signalling, also promotes colitis-associated colorectal cancer [93]. Neotame (E961) has also displayed adverse effects on the gut, including reduction in α and β diversity and escalations of fatty acids and lipids [94]. Alterations in gut microbial composition are additionally observed in response to xylitol (E967), increasing risk of intestinal inflammation and IBD [95].

Conversely, several studies have emphasised beneficial effects of sweeteners in IBD. The bifidogenic properties of isomalt (E953) [96] denotes a capability of promoting a healthy colonic mucosal environment. Isomalt, having shown to exercise a profound growth of *Bifidobacterium*, is therefore regarded as a prebiotic carbohydrate. Thaumatin (E957) is a plant-derived sweetener that is 100,000 times sweeter than sucrose. It has been shown that thaumatin contributes to restoration of microbial balance by increasing intestinal *Butyricioccus* [97], a bacteria genus that has anti-inflammatory properties and has been found to exist at significantly lower levels in IBD patients in comparison to healthy controls [98]. Also influencing gut microbiome composition is the additive neohesperidin dihydrochalcone (NHDC, E959), an intense sweetener derived from citrus. It was reported that treatment with NHDC resulted in an increase in the levels of *Lactobacillus* [99]. This observation was additionally associated with anti-inflammatory effects, including the promotion of oxidative phosphorylation. Another important group of sweeteners that may contribute to intestinal healing are steviol glycosides (E960), the most abundant of which is stevioside, shown to exert several anti-inflammatory properties. One of the proposed anti-inflammatory mechanisms of stevioside is the reduction in TNF-α, IL-6, COX-2, and iNOS via inhibition of the NF-κB and MAPK pathways, which was highlighted in the decreased of p38, ERK, and June N-terminal kinase (JNK) phosphorylation in colonic tissue of DSS-induced UC upon consumption [100]. Furthermore, this study demonstrated that stevioside promoted SOD, catalase, and glutathione s-transferase. Another study considered the relationship between erythritol (E968), a sugar alcohol, and inflammation [101]. Findings from this study revealed that erythritol increased levels of butyric acid, which has also previously demonstrated anti-inflammatory properties [102]. Findings on the impact of sweeteners on intestinal inflammation are summarised in **Supplementary Table 3**.

5.8 Emulsifiers, Stabilisers, and Thickeners

Emulsifiers are food additives widely used in food manufacturing (estimated global market size valued at $7.87 billion in 2023 [103]) to aid mixing of otherwise immiscible substances. While emulsifiers have been long used in the food industry, their effects on gut health have only more recently been proposed. Emulsifiers are often categorised alongside thickeners and stabilisers within the food industry and are therefore also discussed together in the present review. A number of studies have indicated both harmful and beneficial emulsifiers,

stabilisers, and thickeners in the context of GI health (**Supplementary Table 4)**. Carrageenan (E407) is a natural polysaccharide extracted from red algae commonly used as a stabiliser or thickener. Increasing evidence suggests that carrageenan is associated with intestinal damage. One study found that a high carrageenan diet (given in drinking water at 50g/L) caused ulcers in guinea pig colons [104] with further studies in rats revealing that consumption of the emulsifier at the same concentration resulted in increased intestinal permeability [105], epithelial cell loss [106], and diarrhoea [107]. While these studies collectively demonstrate the potential effects of carrageenan, it is important to note that the doses used were high, stated at a maximum 2 g/kg bw/day (in comparison to the ADI for humans which is 75 mg/kg bw/day [108]). In vitro experiments have supported these findings, showing increased expression of pro-inflammatory molecules [109] and reduced sulfatase enzyme activity resulting from carrageenan exposure in human intestinal cells [110]. Importantly, a randomised clinical trial in twelve patients receiving either carrageenan-containing capsules or a placebo [111] showed that three patients receiving the carrageenan capsule relapsed, in comparison to none of the patients in the placebo group reporting disease relapse. There is also significant evidence to suggest that guar gum (E412), a soluble fibre extracted from guar beans, is detrimental to gut health. Mice receiving a diet containing guar gum presented severe colonic inflammation, as shown by colon thickening and increased levels of IL-1β, effects which were not observed in control mice [112]. Recently Polysorbate 80 (P80, E436) and carboxymethylcellulose (CMC, E466) have received attention due to their frequent use in the Western diet. Preclinical evidence supports the notion that both P80 and CMC induce intestinal inflammation via several mechanisms. Microbial dysregulation, including a decrease in *Streptococcus* and *Faecalibacterium,* has been noted upon consumption of these additives [113], with CMC inducing a more aggressive form of colitis [114]. CMC exposure also damages the intestinal epithelial barrier and increases expression of pro-inflammatory cytokines including TNF-α [115]. Additional mechanisms of action for P80 have been proposed, including reduced mucin thickness due to decreased Mucin2 (Muc2) RNA expression, thereby leading to increased intestinal permeability and increased microbial translocation across the epithelial barrier [116]. Sodium stearoyl lactylate (SSL, E481) is a dietary emulsifier commonly used in bread products. To better understand the impact of SSL, it was added to an in vitro model and microbial analysis was performed, with findings that the abundance of *Clostridia* and bacterial butyrate producers were inhibited [117]. Microbial dysbiosis was also observed from exposure to sorbitan

monostearate (SMS, E491) [113]. Experimental data pertaining to the effects of maltodextrin (E1400), a polysaccharide derived from starch hydrolysis, indicates an increased severity of colitis resulting from reduced levels of mucus producing cells, altered microbiome, and crypt hyperplasia [118].

Research into the possible beneficial role of food additives has led to the recognition of certain emulsifiers with beneficial properties that may be utilised in the context of IBD and intestinal inflammatory disorders. Evaluation of sodium alginate (E401), the sodium salt of alginic acid, revealed its therapeutic potential in experimental UC [119]. Oral administration of sodium alginate in drinking water of 2,4,6-Trinitrobenzene sulfonic acid (TNBS)-induced mice resulted in a significant decrease in pro-inflammatory cytokines and matrix metalloproteinase-2 (MMP-2) activity. This was reinforced by findings from a similar study showing that DSS- and TNBS-induced colitis was diminished after sodium alginate treatment [120]. The anti-inflammatory activity of acacia gum (E414) was demonstrated in a study where supplementation decreased the abundance of *clostridium histolyticum* and increased abundances of *Bifidobacteria* [121]. Konjac glucomannan (KGM, E425) is a soluble fibre extracted from the Amorphophallus konjac plant. In addition to its role as a stabiliser and substitute for gelatine, it also has potential roles as an anti-inflammatory agent: treatment with KGM ameliorated DSS-induced colitis and repaired the intestinal epithelial barrier, an outcome suggested to be predominantly attributed to prebiotic effects [122]. Other associated changes noted upon KGM treatment include the reduction of pro-inflammatory factors, increased weight, and alteration of gut microbial composition. A major influence on gut inflammation has also been established by pectin (E440), a common fibre found in fruits including improvement of experimental colitis in mice fed with pectin [123]. This was characterised by a reduction in inflammatory cytokine levels and histological scores, in comparison to the control group of mice. To further facilitate research on pectin, another study set out to characterise the immune response to pectin treatment using an IL-10 deficient mouse model of colitis [124]. Observed effects included an increase in NLR family CARD domain-containing protein 4 (NLRC4) and a decrease in colonic IL-1β and *Verrucomicrovia* abundance. The authors therefore concluded that pectin exhibits anti-inflammatory properties via microbial and immune restoration.

Overall, the predominant mechanisms of food additive action include changes to gut microbial composition and function, altered ROS/ antioxidant balance, immune function, and epithelial barrier integrity. These mechanisms are similar to those governing the pathophysiology of IBD and therapeutic targets and are documented in **Figure 4**.

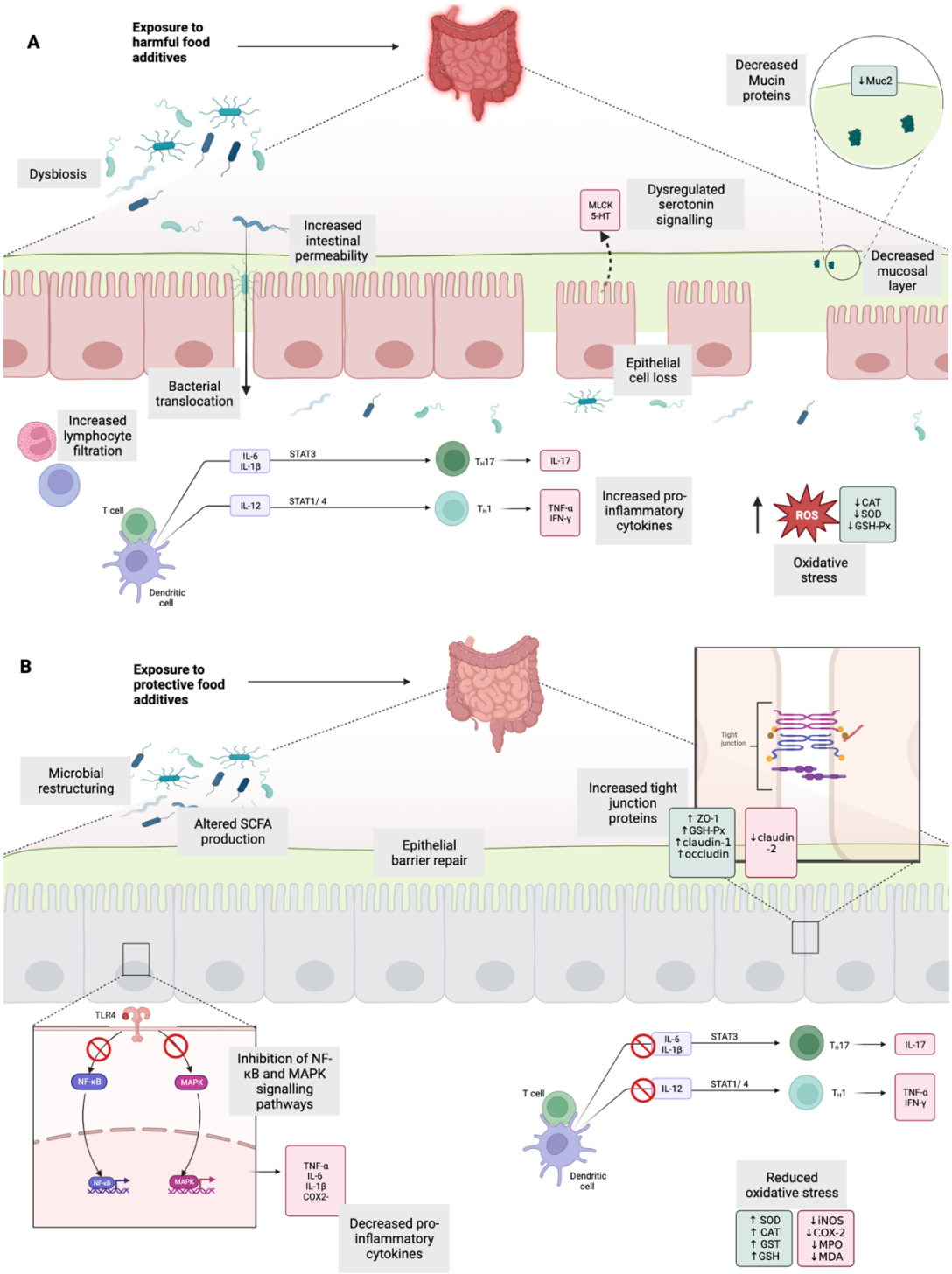


Figure 4. Overview of mechanistic effects of food additives on the GI system. (**A**) Mechanisms of food additives promoting intestinal inflammation. Food additives which exert a pro-inflammatory effect disrupt epithelial barrier integrity, for example through increased intestinal permeability, epithelial cell loss, and decreased mucin production, promoting translocation of bacteria into the intestinal lumen, where an adaptive immune response is elicited. (**B**) Mechanisms of food additives promoting intestinal healing. Food additives which exert an anti-inflammatory effect restore intestinal homeostasis via

epithelial barrier and microbial restructuring, which is associated with a downregulated inflammatory immune response, e.g., decreased production of pro-inflammatory cytokines and ROS. Muc2: mucin2; MLCK: mysosin light-chain kinase; 5-HT: 5-hydroxytrypamine; STAT3: signal transducer and activator of transcription; CAT: catalase; SOD: superoxide dismutase; GSH-Px: glutathione peroxidase; ZO-1: zonula occludens 1; SCFA: short chain fatty acid; COX-2: cyclooxygenase-2; MAPK: mitogen-activated protein kinases; GST: glutathione S-transferase; GSH: glutathione; iNOS: inducible nitric oxide synthase; myeloperoxidase; MDA: malondialdehyde.

5.9 Applications in Dietary Management

Considering the impact of a single food additive on GI inflammation, the total food additive consumption in an individual's diet has significant potential to alter the intestinal environment and physiology. In addition to the exclusion of food additives that exacerbate inflammation, the therapeutic potential demonstrated by certain food additives provides scope for their use in ameliorating disease. This can be achieved through increased intake or supplementation, for example as pre- and pro- biotics. While the majority of clinical trials with pre- and pro- biotics have shown no success to date, as more research is conducted alongside advancing characterisation of what constitutes a healthy human microbiome, this may allow development of novel strategies to substantiate potential health benefits. Further prospective uses include the utilisation of the molecule in the development of a targeted therapy, and as adjuvants to other therapeutic approaches. Each of the mechanisms discussed in the current review represents a potential strategy or target for treatment and future investigation of food additives may drive mechanistic insights of disease pathophysiology and treatment actions. It is crucial to note that current research on the impact and mechanisms of food additives in humans is not yet at a place where we can put this into clinical practice as extensive further research is required before implementation of dietary guidelines. Here, we summarise the food additives that have preclinical evidence of benefit to gut inflammation and propose them to be further tested with RCTs. Potential considerations for food additives in RCTs are shown (**Figure 5**).

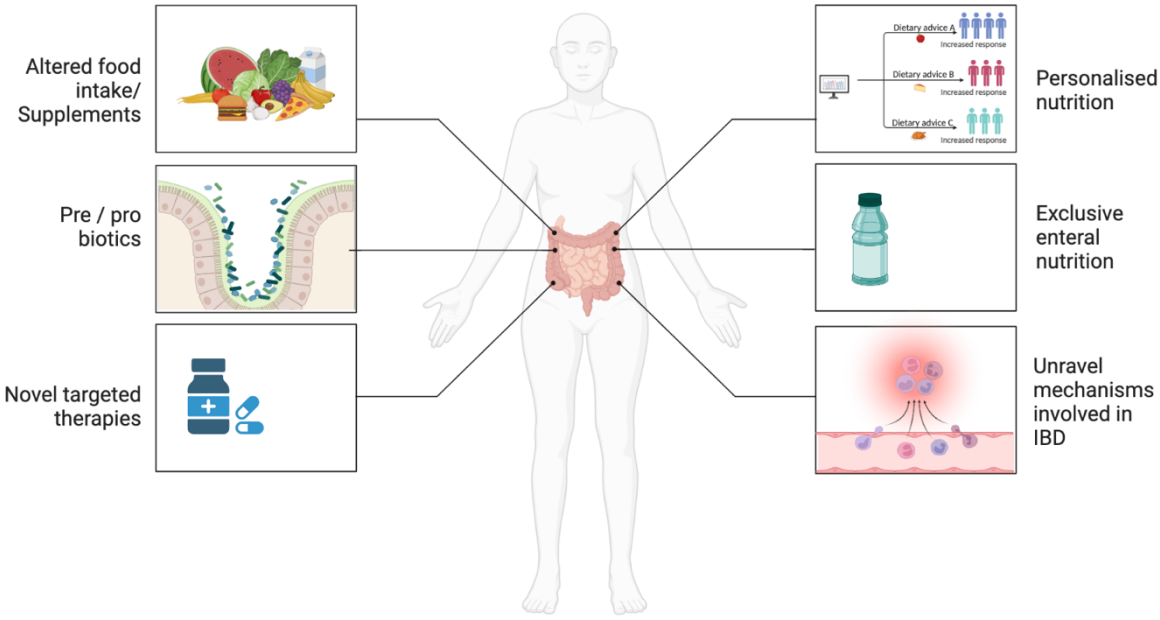


Figure 5. Potential opportunities for utilising mechanistic knowledge of food additives, upon further human research and controlled trials. Future applications may include altered intake of food additives, pre/ probiotics, novel targeted therapies, personalised nutrition, exclusive enteral nutrition, and unravelling mechanisms of dietary-associated disease.

In addition to the limited completed clinical trials, as discussed in this review within the context of each additive, there are a number of further ongoing trials investigating their effects in relation to different aspects of gut health (**Table 1**). Clinical trials displayed are those listed by the Clinical Trials database [125] and the International Clinical Trials Registry Platform [126]. While results are not yet available on their outcomes, they will expand knowledge of their mechanistic and clinical effects in humans and provide more data on food additives and begin to bridge the translational gap between animal and human studies.

Table 1. Overview of food additive clinical trials in GI health. Clinical trial status information was last verified in September 2025.

Clinical Trial ID	Title	Additive(s)	Population	Outcomes Measured	Status
NCT05743374	Micronutrient and Additive Modifications may Optimise Diet to Health (Mammoth)	Emulsifying agents within the E400 group	70 UC patients, elimination diet (n=35), normal diet (n=35)	Disease activity (diarrhoea frequency, blood in stools, abdominal pain, CRP, and calprotectin) and microbial analysis (dysbiosis index)	Recruiting
NCT04046913	The ADDapt Diet in Reducing Crohn's Disease Inflammation	Food additives associated with the Western diet	Mildly active CD patients	CD activity, health-related quality of life, gut bacteria, gut permeability, gut inflammation, and dietary intake	Active
NCT05852587	Xylitol Use for Decolonization of C. Difficile in Patients With IBD	Xylitol	72 IBD patients, xylitol (n=36), placebo (n=36)	Decolonization of C. Difficile, disease activity, and development of CDI	Not yet recruiting
NCT05849012	A Pilot Study Examining Low Sulfur Diet as Treatment for	Sulfur	CD patients	Small intestinal bacterial overgrowth, and intestinal permeability	Recruiting

	Persistent Symptoms in Quiescent Crohn's Disease				
NCT03500653	Curcumin Supplementation as an Add on Treatment for Patients With Inflammatory Bowel Diseases Treated With Vedolizumab	Curcumin	IBD patients (n=84)	CDAI, FC, and CRP.	Unknown
NCT02683733	Bio-enhanced Curcumin as an Add-On Treatment in Mild to Moderate Ulcerative Colitis	Curcumin	Mild to moderate UC patients (n=50)	Time to induction of clinical and endoscopic remission	Unknown
NCT02683759	Bio-enhanced Curcumin as an Add-on Treatment	Curcumin	UC patients in remission (n=50)	Maintenance of clinical and endoscopic remission	Unknown

	in Maintaining Remission of Ulcerative Colitis				
NCT05803811	Effect of Colon Delivered Vitamin B2 on Gut Microbiota and Related Health Biomarkers in Healthy Older Adults	Riboflavin	Healthy older adults (n=348)	Faecal microbial composition and diversity, fatty acid content, intestinal inflammation, intestinal barrier integrity.	Recruiting
NCT00275418	Beta Carotene From Natural Source for Patients With Non-Active Crohn's Disease	Beta-carotene	CD patients in remission (n=300)	CDAI score	Unknown
NCT04000139	Anthocyanin Rich Extract (ACRE) in Patients With Ulcerative Colitis (ACRE)	Anthocyanin	UC patients (n=48)	Clinical response, clinical remission, rectal bleeding, stool frequency, FC.	Completed, results not available

ACTRN12619001099112	The role of a low emulsifier diet in treating intestinal inflammation in patients with Crohn's disease	Emulsifiers	40 patients with mild CD	Change in Harvey-Bradshaw Crohn's Disease Activity Index, faecal calprotectin, and intestinal wall thickness.	Recruiting

5.9.1 Exclusive Enteral Nutrition

Exclusive enteral nutrition (EEN) is an entirely liquid formulated diet which provides all nutritional requirements, typically given for up to 8 weeks orally or through a nasogastric tube [127]. With improved endoscopic, and histological remission in comparison to corticosteroids [128-130], in addition to benefits of minimal side effects [131-133], EEN is currently used as the first line of treatment for adolescent CD patients in Europe [134, 135]. The mechanisms by which EEN induces remission and improves GI inflammation remain unclear; however, microbial and metabolic alteration [136-138] are associated with EEN forming the subject of a multitude of clinical trials exploring molecular mechanisms and clinical outcomes (**Table 2**).

Table 2. Overview of enteral nutrition clinical trials in GI health, CD-TREAT; Crohn's Disease Treatment-with-EATing, PEN; partial enteral nutrition, SEN; standard enteral nutrition, EEN; exclusive enteral nutrition. Clinical trial status information was last verified in September 2025.

Clinical Trial ID	Title	Diet	Population	Outcomes Measured
NCT04225689	The Intensive Post Exclusive Enteral Nutrition Study (iPENS)	CD-TREAT	60 adult CD patients, CD-TREAT diet (n=30), unrestricted diet (n=30)	Disease activity, anthropometric measures, FC levels, and quality of life
NCT04859088	Biologics and Partial Enteral Nutrition Study (BIOPIC)	PEN	80 adult CD patients	Remission rates, bacteria and metabolite changes
NCT02341248	Bacteria & Inflammation in the Gut (BIG) Study (BIG)	SEN	42 paediatric CD patients	Bacterial composition, CD faecal and urine biomarkers, bacterial metabolites.
NCT02426567	The Impact of "Crohn's Disease-TReatment-with-EATing" Diet and Exclusive Enteral Nutrition on Healthy Gut Bacteria [139]	CD-TREAT	CD adult CD patients	Gut microbiota composition and metabolic activity
NCT03171246	CD-TREAT Diet: a Novel Therapy for Active Luminal Crohn's Disease	CD-TREAT	10 adult and 10 paediatric CD patients	Blood and faecal inflammatory markers, disease activity, faecal bacteria and metabolites
NCT02521064	Effects of Exclusive Enteral Nutrition on the Microbiome in Pediatric Inflammatory Bowel Disease	EEN	Paediatric CD patients	Clinical, biochemical, and microbial changes

Different compositions of EEN have been formulated by manufacturers and from their compositional analysis , various levels of food additives have been identified as ingredients

[140]. The presence of food additives in EEN shown to be effective in inducing remission in paediatric IBD patients denotes the importance of these substances in nutritional therapies. However, the food additives identified in EEN, at therapeutic levels consumed, are unlikely to impact on GI inflammation. The main conclusion of this study is that preclinical data does not translate to humans. Based on the evidence presented in the current review, both inflammatory and anti-inflammatory additives, in relation to IBD, can be found in EEN (**Figure 6**).

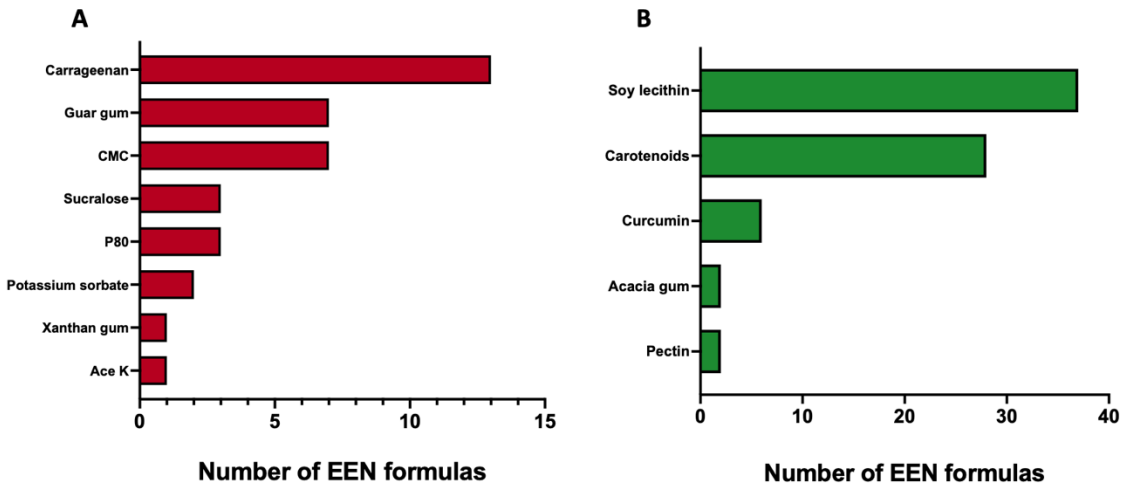


Figure 6. Food additives present in EEN formulas, adapted from Logan M et al., 2020 [140]. (**A**) Food additives in EEN formulas with inflammatory potential; (**B**) Food additives in EEN formulas with therapeutic potential.

5.10 Discussion

As dietary constituents significantly affect intestinal physiology, it is vital to understand the mechanisms of specific components, particularly for improving therapeutic outcomes for patients with GI inflammation such as IBD. Food additives, seen in an expansive variety of products, have become a key focus of dietary research; however, the true impact of food additives on human health remains unclear. In this review we provide a detailed list of food additives with evidence of promoting intestinal inflammation along with a list of food additives which may promote intestinal healing. By addressing the conflicting evidence of food additives on gut inflammation with different models and doses, we highlight existing gaps in our understanding of their mechanisms and future areas of investigation. Continued

emphasis on the molecular outcomes of food additives will promote renewed interest into dietary safety outcomes and build upon existing dietary knowledge.

Importantly, food additive research to date has predominantly been conducted using animal or cell models, with very limited human data. Available clinical data is discussed within the context of specific food additives in this review; however, it is clear that the evidence does not provide strong enough conclusions to make dietary recommendations. While in vitro and preclinical models provide useful insights into chemical-driven inflammatory mechanisms, it is unclear how these findings can be extrapolated to humans, apparent from inconsistencies between animal and human data and lack of successful clinical trials to-date. A reason for contrasting findings may be the difference in food additive doses used in research studies. Animal studies often use larger doses than the maximum ADI defined by regulatory bodies, overall, not reflecting human consumption and rendering comparison of physiological outcomes challenging. Therefore, further longitudinal RCTs with appropriate physiological and clinical outcomes are required.

Knowledge of food additives worsening disease is undeniably crucial to progressing dietary management, however it is also important not to overlook food additives that can provide therapeutic effects. Viewing diet holistically allows a more representative outlook of diet in human health and is crucial for recognising the multifaceted impact of diet on the gut. IBD, associated with reduced quality of life, health costs, and undesirable adverse effects associated with existing treatments, presents an unmet need for more effective management strategies. Utilising food additives in restorative and therapeutic strategies is a desirable approach, as they are frequently consumed in diet and may be better tolerated with less side effects than traditional medications. Dietary advice, novel target development, and nutraceutical adjuncts are suggestions of how food additives can be utilised to improve IBD management strategies in the future, however given the current lack of clinical evidence, considerable research is needed to substantiate their health benefits in humans prior to making dietary recommendations. A crucial aspect deserving attention is the duration of additive exposure: The impact of short-term consumption on human health may differ significantly to long-term exposure, and inconsistencies in the duration of existing studies and additive doses used underlines a considerable challenge in deriving human application. For example, one study looking into the short-term exposure of CMC

[141] addresses the need for additional studies investigating the impact of long-term consumption. Furthermore, to achieve the desired effects of dietary modulation, one must also consider a molecule's metabolism upon digestion: if the additive is degraded or absorbed or modified before it reaches the target site, the desired effects may not be attained or exacerbated by the new version or product of food additive metabolism. Consideration of food additive metabolism is necessary to employ food molecules in a therapeutic sense. In this way, metabolomics approaches can be employed as a powerful tool to investigate the metabolic fate of food additives under different conditions and may further progress dietary strategies. Advances in metabolomics tools and food biomarker approaches are integral to this goal [142, 143].

Fundamental to understanding food additives in human health is the recognition that diet may play a differing role in triggering IBD compared to managing the disease, and due to the chronic relapsing nature of IBD, this also includes impacts at different disease stages, i.e., active disease vs. remission. It is yet to be elucidated whether food additives are equally important for disease development and management, an issue which has been recently highlighted in the literature [144]. For future research, the presence of food additives in EEN, which is used frequently in IBD patients worldwide, provides a widow of opportunity for nutrition research. It is well evidenced that the use of EEN is effective at inducing remission in children with CD throughout the course of treatment; however, after reinduction of a solid-based food diet, it is frequently observed that inflammation increases and disease relapses. It has been suggested that there is a critical time period after returning to a solid food diet which may determine the subsequent trajectory of disease state. As such, the specific dietary components consumed during this period are of great significance. Consideration of the food additives that are taken by an individual post-EEN treatment as part of the reintroduction diet provides one example of how data can be better utilised to provide information on the role of food additives in IBD. While little evidence is currently available around the food reintroduction phase post-EEN, the Intensive Post Exclusive Enteral Nutrition Study (iPENS) is collecting data to assess gut inflammation at this stage [145], results from which may help progress our understanding of the role of diet in disease management.

In this paper, we focus on the documented influences of food additives that impact the GI system. This is relevant for improving knowledge and future management of diseases of

intestinal inflammation such as IBD, however this does not capture all food additives that are consumed by humans and their effects may also be more complicated when one considers their impact beyond the GI system. Although not the focus of this review, there is a requirement for collated knowledge of food additives and their systemic effects prior to dietary suggestions. This captures a further limitation of in vivo models and lack of human evidence, as a full understanding of their role and potential side effects remains unknown. Additionally, there are still gaps in our understanding of the role of food additives in combination with other additives and ingredients, which may further enhance or mitigate their individual effects. It is clear that collaborative research is needed to establish effects in humans and assess suitability for disease management in humans.

In addition to utilising mechanistic food additive knowledge within the realms of currently available food products, there is impetus to suggest a role within the changing dynamics of the food industry. The requirement for safer alternatives in food products is clear; however, several challenges exist in the road towards improved food additive utilisation for human health. Currently, one of the biggest limitations in enabling the incorporation of therapeutic food additives in products is improving their chemical performance to maximise properties such stability and shelf-life. Optimising food additive choice based on chemistry and stability remains challenging task, for example, as synthetic food additives tend to be more stable than natural alternatives and so they are easier to process at various temperatures and solubilities [146-148]. However, the use of additives in the food industry is continually changing and increased understanding of their physiological mechanisms ensure that safety guidelines are updated in line with new evidence of their effects, thereby providing opportunities for product changes. Optimising their use simultaneously for human health and product stability is an open challenge and we have begun to see a change in the direction of food manufacturing as a result. For example, titanium dioxide was previously used as a food additive in a white food colourant, however upon recent safety re-evaluations, the EFSA deemed that it could no longer be considered safe due to genotoxicity concerns [149-151]. This opened a gap for white colourants within the food industry which, through increased efforts and improved formulations, was subsequently filled with safer alternatives. Similarly, the safety of the food additive sodium tetraborate, also known as borax, was reassessed in 2010, as potential impacts on reproductive health and increased risk of liver cancer resulted it its ban in Europe [152, 153]. This model of continual reassessment which brings in stricter

and safer regulations may provide a shift in dietary components to include more natural and less harmful alternatives. This provides a window of opportunity for food additive research to be utilised in both the food and pharmaceutical industries. Improved and standardised methods of food additive analysis will allow the progression of dietary knowledge for healthy populations through maintenance of gut homeostasis and patients of gastrointestinal disease through restoration of disrupted pathways. In consideration to these challenges and requirements, we here set out perspectives on future research (**Figure 7**).

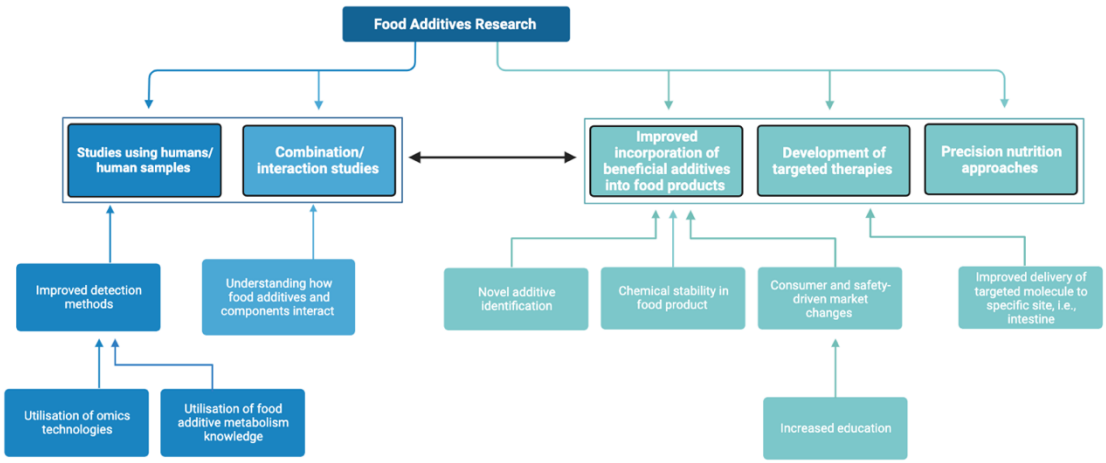


Figure 7. Roadmap to an evidence-driven future of improved food additive utilisation within the food and pharmaceutical industries.

5.11 Conclusions

Food additives are ubiquitous in the Western diet with important physiological influences on gut health. We here demonstrate that the roles of food additives can be of either an inflammatory or anti-inflammatory nature depending on their mechanism, which may bring important implications for the treatment and management of IBD. However, the proposed mechanisms are predominantly derived from preclinical models with undefined translation to human health, and therefore further research exploring food additive interactions and outcomes using controlled clinical trials are required to gain a better understanding of their true effects. The continued movement towards improved health outcomes through food additives represents a growing opportunity for IBD management.

5.12 References

1. Reddavide, R., et al., *The role of diet in the prevention and treatment of Inflammatory Bowel Diseases.* Acta Biomed, 2018. **89**(9-s): p. 60-75.
2. Chiba, M., K. Nakane, and M. Komatsu, *Westernized Diet is the Most Ubiquitous Environmental Factor in Inflammatory Bowel Disease.* Perm J, 2019. **23**: p. 18-107.
3. Alam, M.T., et al., *Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels.* Gut Pathogens, 2020. **12**(1): p. 1.
4. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.* Genome Biology, 2012. **13**(9): p. R79.
5. Magro, D.O., et al., *Remission in Crohn's disease is accompanied by alterations in the gut microbiota and mucins production.* Scientific Reports, 2019. **9**(1): p. 13263.
6. Gao, S., et al., *Comprehensive microbiota alterations in IBD and improved diagnostic accuracy for IBD using multi-kingdom microbial features.* The FASEB Journal, 2021. **35**(S1).
7. Lechuga, S. and A.I. Ivanov, *Disruption of the epithelial barrier during intestinal inflammation: Quest for new molecules and mechanisms.* Biochim Biophys Acta Mol Cell Res, 2017. **1864**(7): p. 1183-1194.
8. McGuckin, M.A., et al., *Intestinal barrier dysfunction in inflammatory bowel diseases.* Inflamm Bowel Dis, 2009. **15**(1): p. 100-13.
9. Gitter, A.H., et al., *Epithelial barrier defects in ulcerative colitis: characterization and quantification by electrophysiological imaging.* Gastroenterology, 2001. **121**(6): p. 1320-8.
10. Xavier, R.J. and D.K. Podolsky, *Unravelling the pathogenesis of inflammatory bowel disease.* Nature, 2007. **448**(7152): p. 427-34.
11. Ghoreschi, K., et al., *Generation of pathogenic T(H)17 cells in the absence of TGF-β signalling.* Nature, 2010. **467**(7318): p. 967-71.
12. *The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet Gastroenterol Hepatol, 2020. **5**(1): p. 17-30.
13. Santiago, M., et al., *What forecasting the prevalence of inflammatory bowel disease may tell us about its evolution on a national scale.* Therap Adv Gastroenterol, 2019. **12**: p. 1756284819860044.

14. Raoul, P., et al., *Food Additives, a Key Environmental Factor in the Development of IBD through Gut Dysbiosis.* Microorganisms, 2022. **10**(1).
15. Bancil, A.S., et al., *Food Additive Emulsifiers and Their Impact on Gut Microbiome, Permeability, and Inflammation: Mechanistic Insights in Inflammatory Bowel Disease.* J Crohns Colitis, 2021. **15**(6): p. 1068-1079.
16. Jarmakiewicz-Czaja, S., D. Piątek, and R. Filip, *The impact of selected food additives on the gastrointestinal tract in the example of nonspecific inflammatory bowel diseases.* Arch Med Sci, 2022. **18**(5): p. 1286-1296.
17. *https://www.precedenceresearch.com/food-additive-market.*
18. Jin, Z.-Q., et al., *A meta-analysis of dietary carbohydrate intake and inflammatory bowel disease risk: evidence from 15 epidemiology studies.* Revista Espanola de Enfermadades Digestivas (REED), 2019. **111**: p. 5+.
19. Nie, J.Y. and Q. Zhao, *Beverage consumption and risk of ulcerative colitis: Systematic review and meta-analysis of epidemiological studies.* Medicine (Baltimore), 2017. **96**(49): p. e9070.
20. de Castro, M.M., et al., *Role of diet and nutrition in inflammatory bowel disease.* World J Exp Med, 2021. **11**(1): p. 1-16.
21. Toropov, A.A., et al., *CORAL: model for no observed adverse effect level (NOAEL).* Mol Divers, 2015. **19**(3): p. 563-75.
22. Spence, C., *On the psychological impact of food colour.* Flavour, 2015. **4**(1): p. 21.
23. *The global food colors market size was USD 2.55 billion in 2018 and is projected to reach USD 4.77 billion by 2026, exhibiting a CAGR of 8.19% during the forecast period.... Read More at:- https://www.fortunebusinessinsights.com/food-colors-market-102644*.
24. Bartzatt, R. and T. Wol, *Detection and Assay of Vitamin B-2 (Riboflavin) in Alkaline Borate Buffer with UV/Visible Spectrophotometry.* International Scholarly Research Notices, 2014. **2014**: p. 453085.
25. Van Nong, H., et al., *Fabrication and vibration characterization of curcumin extracted from turmeric (Curcuma longa) rhizomes of the northern Vietnam.* Springerplus, 2016. **5**(1): p. 1147.
26. Bernstein, P.S., et al., *Lutein, zeaxanthin, and meso-zeaxanthin: The basic and clinical science underlying carotenoid-based nutritional interventions against ocular disease.* Prog Retin Eye Res, 2016. **50**: p. 34-66.

27. Smirnova, S.V. and V.V. Apyari, *Aqueous Two-Phase Systems Based on Cationic and Anionic Surfactants Mixture for Rapid Extraction and Colorimetric Determination of Synthetic Food Dyes.* Sensors (Basel), 2023. **23**(7).
28. Denish, P.R., et al., *Discovery of a natural cyan blue: A unique food-sourced anthocyanin could replace synthetic brilliant blue.* Science Advances, 2021. **7**(15): p. eabe7871.
29. F Rsc, S., et al., *Anthocyanin Profiling Using UV-Vis Spectroscopy and Liquid Chromatography Mass Spectrometry.* Journal of AOAC International, 2019. **103**.
30. Popescu, M., et al., *Different spectrophotometric methods for simultaneous quantification of lycopene and β-carotene from a binary mixture.* LWT, 2022. **160**: p. 113238.
31. Gobara, M. and M. Elsayed, *Preparation and Characterization of PANI/TiO2 Composite for Photocatalytic Degradation of Tartrazine Dye.* The International Conference on Chemical and Environmental Engineering, 2016. **8**: p. 370-389.
32. Rigoni, R., et al., *DISCOLORING OF CANDY INDUSTRY WASTEWATERS BY HETEROGENEOUS PHOTOCATALYSIS PROCESS.* 2023.
33. Rovina, K., S. Siddiquee, and S.M. Shaarani, *Extraction, Analytical and Advanced Methods for Detection of Allura Red AC (E129) in Food and Beverages Products.* Front Microbiol, 2016. **7**: p. 798.
34. Rattray, N.J.W., et al., *Fluorescent probe for detection of bacteria: conformational trigger upon bacterial reduction of an azo bridge.* Chemical Communications, 2012. **48**(51): p. 6393-6395.
35. Rattray, N.J.W., et al., *Chemical and bacterial reduction of azo-probes: monitoring a conformational change using fluorescence spectroscopy.* Tetrahedron, 2013. **69**(13): p. 2758-2766.
36. Golka, K., S. Kopps, and Z.W. Myslak, *Carcinogenicity of azo colorants: influence of solubility and bioavailability.* Toxicology Letters, 2004. **151**(1): p. 203-210.
37. Monisha, B., et al., *Sensing of azo toxic dyes using nanomaterials and its health effects - A review.* Chemosphere, 2023. **313**: p. 137614.
38. Xu, H., et al., *Sudan azo dyes and Para Red degradation by prevalent bacteria of the human gastrointestinal tract.* Anaerobe, 2010. **16**(2): p. 114-9.
39. Stingley, R.L., et al., *Metabolism of azo dyes by human skin microbiota.* J Med Microbiol, 2010. **59**(Pt 1): p. 108-114.

40. Feng, J., C.E. Cerniglia, and H. Chen, *Toxicological significance of azo dye metabolism by human intestinal microbiota.* Front Biosci (Elite Ed), 2012. **4**(2): p. 568-86.
41. Wu, L., et al., *Impacts of an azo food dye tartrazine uptake on intestinal barrier, oxidative stress, inflammatory response and intestinal microbiome in crucian carp (Carassius auratus).* Ecotoxicol Environ Saf, 2021. **223**: p. 112551.
42. Sasaki, Y.F., et al., *The comet assay with 8 mouse organs: results with 39 currently used food additives.* Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2002. **519**(1): p. 103-119.
43. Ghonimi, W., *Histological Changes of Selected Westar Rat Tissues Following the Ingestion of Tartrazine With Special Emphasis on the Protective Effect of Royal Jelly and Cod Liveroil.* Journal of Cytology & Histology, 2015. **06**.
44. Moutinho, I.L., L.C. Bertges, and R.V. Assis, *Prolonged use of the food dye tartrazine (FD&C yellow no 5) and its effects on the gastric mucosa of Wistar rats.* Braz J Biol, 2007. **67**(1): p. 141-5.
45. Kong, X., et al., *Effects of sunset yellow on proliferation and differentiation of intestinal epithelial cells in murine intestinal organoids.* J Appl Toxicol, 2021. **41**(6): p. 953-963.
46. Kwon, Y.H., et al., *Chronic exposure to synthetic food colorant Allura Red AC promotes susceptibility to experimental colitis via intestinal serotonin in mice.* Nat Commun, 2022. **13**(1): p. 7617.
47. Camacho-Barquero, L., et al., *Curcumin, a Curcuma longa constituent, acts on MAPK p38 pathway modulating COX-2 and iNOS expression in chronic experimental colitis.* Int Immunopharmacol, 2007. **7**(3): p. 333-42.
48. Wei, C., et al., *Curcumin ameliorates DSS-induced colitis in mice by regulating the Treg/Th17 signaling pathway.* Mol Med Rep, 2021. **23**(1).
49. Cooney, J.M., et al., *A combined omics approach to evaluate the effects of dietary curcumin on colon inflammation in the Mdr1a(-/-) mouse model of inflammatory bowel disease.* J Nutr Biochem, 2016. **27**: p. 181-92.
50. Nones, K., et al., *The effects of dietary curcumin and rutin on colonic inflammation and gene expression in multidrug resistance gene-deficient (mdr1a-/-) mice, a model of inflammatory bowel diseases.* Br J Nutr, 2009. **101**(2): p. 169-81.

51. Holt, P.R., S. Katz, and R. Kirshoff, *Curcumin therapy in inflammatory bowel disease: a pilot study.* Dig Dis Sci, 2005. **50**(11): p. 2191-3.
52. Lang, A., et al., *Curcumin in Combination With Mesalamine Induces Remission in Patients With Mild-to-Moderate Ulcerative Colitis in a Randomized Controlled Trial.* Clin Gastroenterol Hepatol, 2015. **13**(8): p. 1444-9.e1.
53. Masoodi, M., et al., *The efficacy of curcuminoids in improvement of ulcerative colitis symptoms and patients' self-reported well-being: A randomized double-blind controlled trial.* J Cell Biochem, 2018. **119**(11): p. 9552-9559.
54. Singla, V., et al., *Induction with NCB-02 (curcumin) enema for mild-to-moderate distal ulcerative colitis - a randomized, placebo-controlled, pilot study.* J Crohns Colitis, 2014. **8**(3): p. 208-14.
55. Hanai, H., et al., *Curcumin maintenance therapy for ulcerative colitis: randomized, multicenter, double-blind, placebo-controlled trial.* Clin Gastroenterol Hepatol, 2006. **4**(12): p. 1502-6.
56. von Martels, J.Z.H., et al., *Riboflavin Supplementation in Patients with Crohn's Disease [the RISE-UP study].* J Crohns Colitis, 2020. **14**(5): p. 595-607.
57. Wang, J., D.G. Jackson, and G. Dahl, *The food dye FD&C Blue No. 1 is a selective inhibitor of the ATP release channel Panx1.* J Gen Physiol, 2013. **141**(5): p. 649-56.
58. Cheng, J., et al., *The Role of β-Carotene in Colonic Inflammation and Intestinal Barrier Integrity.* Front Nutr, 2021. **8**: p. 723480.
59. Lee, K. and W.H. Yoon, *Preventive effects of co-treatment with fucoidan and lutein on the development of inflammatory bowel disease in DSS mouse model.* Natural Product Sciences, 2011. **17**: p. 234-238.
60. Głąbska, D., et al., *Lycopene, Lutein and Zeaxanthin May Reduce Faecal Blood, Mucus and Pus but not Abdominal Pain in Individuals with Ulcerative Colitis.* Nutrients, 2016. **8**(10).
61. Gerasimidis, K., et al., *Impact of exclusive enteral nutrition on body composition and circulating micronutrients in plasma and erythrocytes of children with active Crohn's disease.* Inflamm Bowel Dis, 2012. **18**(9): p. 1672-81.
62. Li, J., et al., *Bilberry anthocyanin extract promotes intestinal barrier function and inhibits digestive enzyme activity by regulating the gut microbiota in aging rats.* Food Funct, 2019. **10**(1): p. 333-343.

63. Cremonini, E., et al., *Anthocyanins protect the gastrointestinal tract from high fat diet-induced alterations in redox signaling, barrier integrity and dysbiosis.* Redox Biol, 2019. **26**: p. 101269.
64. Tian, B., et al., *Lycium ruthenicum Anthocyanins Attenuate High-Fat Diet-Induced Colonic Barrier Dysfunction and Inflammation in Mice by Modulating the Gut Microbiota.* Mol Nutr Food Res, 2021. **65**(8): p. e2000745.
65. https://www.grandviewresearch.com/industry-analysis/food-preservatives-market.
66. Peng, Q., et al., *Potassium sorbate suppresses intestinal microbial activity and triggers immune regulation in zebrafish (Danio rerio).* Food Funct, 2019. **10**(11): p. 7164-7173.
67. Nagpal, R., N. Indugu, and P. Singh, *Distinct Gut Microbiota Signatures in Mice Treated with Commonly Used Food Preservatives.* Microorganisms, 2021. **9**(11).
68. Hrncirova, L., et al., *Food Preservatives Induce Proteobacteria Dysbiosis in Human-Microbiota Associated Nod2-Deficient Mice.* Microorganisms, 2019. **7**(10).
69. Jimenez Loayza, J.J., et al., *P837 The common food additives sodium sulfite and polysorbate 80 have a profound inhibitory effect on the commensal, anti-inflammatory bacterium Faecalibacterium prausnitzii: the ENIGMA study.* Journal of Crohn's and Colitis, 2019. **13**(Supplement_1): p. S542-S543.
70. Irwin, S.V., et al., *Sulfites inhibit the growth of four species of beneficial gut bacteria at concentrations regarded as safe for food.* PLoS One, 2017. **12**(10): p. e0186629.
71. Oshimo, M., et al., *Sodium sulfite causes gastric mucosal cell death by inducing oxidative stress.* Free Radic Res, 2021. **55**(6): p. 731-743.
72. Halas, D., et al., *Dietary supplementation with benzoic acid improves apparent ileal digestibility of total nitrogen and increases villous height and caecal microbial diversity in weaner pigs.* Animal Feed Science and Technology, 2010. **160**(3): p. 137-147.
73. Diao, H., et al., *Effects of benzoic Acid and thymol on growth performance and gut characteristics of weaned piglets.* Asian-Australas J Anim Sci, 2015. **28**(6): p. 827-39.
74. White, A., et al., *Improvement in orofacial granulomatosis on a cinnamon- and benzoate-free diet.* Inflamm Bowel Dis, 2006. **12**(6): p. 508-14.
75. Kim, S.K. and E.S. Lee, *Orofacial granulomatosis associated with Crohn's disease.* Ann Dermatol, 2010. **22**(2): p. 203-5.

76. Walia, D., et al., *Exploring the therapeutic potential of sodium benzoate in acetic acid-induced ulcerative colitis in rats.* J Basic Clin Physiol Pharmacol, 2019. **30**(5).
77. *https://discovery.ucl.ac.uk/id/eprint/10045000/.*
78. Banerjee, S., et al., *Sulphur dioxide ameliorates colitis related pathophysiology and inflammation.* Toxicology, 2019. **412**: p. 63-78.
79. Tong, L.C., et al., *Propionate Ameliorates Dextran Sodium Sulfate-Induced Colitis by Improving Intestinal Barrier Function and Reducing Inflammation and Oxidative Stress.* Front Pharmacol, 2016. **7**: p. 253.
80. Filippone, A., et al., *The Anti-Inflammatory and Antioxidant Effects of Sodium Propionate.* Int J Mol Sci, 2020. **21**(8).
81. Ghadimi, D., et al., *Calcium propionate alleviates DSS-induced colitis and influences lipid profile, Interferon gamma, peptidoglycan recognition protein 3 and calprotectin in one-year-old female and male C57BL/6 mice.* Scientific Journal of Immunology and Immunotherapy, 2017. **1**: p. 07-016.
82. Pobeguts, O.V., et al., *Propionate Induces Virulent Properties of Crohn's Disease-Associated Escherichia coli.* Front Microbiol, 2020. **11**: p. 1460.
83. Ormsby, M.J., et al., *Inflammation associated ethanolamine facilitates infection by Crohn's disease-linked adherent-invasive Escherichia coli.* EBioMedicine, 2019. **43**: p. 325-332.
84. Yan, H., et al., *Ascorbic acid ameliorates oxidative stress and inflammation in dextran sulfate sodium-induced ulcerative colitis in mice.* Int J Clin Exp Med, 2015. **8**(11): p. 20245-53.
85. Qiu, F., et al., *Combined effect of vitamin C and vitamin D(3) on intestinal epithelial barrier by regulating Notch signaling pathway.* Nutr Metab (Lond), 2021. **18**(1): p. 49.
86. Liu, K.Y., et al., *Vitamin E alpha- and gamma-tocopherol mitigate colitis, protect intestinal barrier function and modulate the gut microbiota in mice.* Free Radic Biol Med, 2021. **163**: p. 180-189.
87. Mirbagheri, S.A., et al., *Rectal administration of d-alpha tocopherol for active ulcerative colitis: a preliminary report.* World J Gastroenterol, 2008. **14**(39): p. 5990-5.

88. Blackwell, T.S., et al., *In vivo antioxidant treatment suppresses nuclear factor-kappa B activation and neutrophilic lung inflammation.* J Immunol, 1996. **157**(4): p. 1630-7.

89. Chougule, P.R., et al., *Effects of Ethyl Gallate and Propyl Gallate on Dextran Sulfate Sodium (DSS)-Induced Ulcerative Colitis in C57BL/6J Mice: Preventive and Protective*. 2023, Research Square.

90. *https://www.grandviewresearch.com/industry-analysis/sweeteners-market-report.*

91. Hanawa, Y., et al., *Acesulfame potassium induces dysbiosis and intestinal injury with enhanced lymphocyte migration to intestinal mucosa.* J Gastroenterol Hepatol, 2021. **36**(11): p. 3140-3148.

92. Bian, X., et al., *Gut Microbiome Response to Sucralose and Its Potential Role in Inducing Liver Inflammation in Mice.* Front Physiol, 2017. **8**: p. 487.

93. Li, X., et al., *Sucralose Promotes Colitis-Associated Colorectal Cancer Risk in a Murine Model Along With Changes in Microbiota.* Front Oncol, 2020. **10**: p. 710.

94. Chi, L., et al., *Effects of the Artificial Sweetener Neotame on the Gut Microbiome and Fecal Metabolites in Mice.* Molecules, 2018. **23**(2).

95. Xiang, S., et al., *Xylitol enhances synthesis of propionate in the colon via cross-feeding of gut microbiota.* Microbiome, 2021. **9**(1): p. 62.

96. Gostner, A., et al., *Effect of isomalt consumption on faecal microflora and colonic metabolism in healthy volunteers.* Br J Nutr, 2006. **95**(1): p. 40-50.

97. *https://ruor.uottawa.ca/handle/10393/42287.*

98. Eeckhaut, V., et al., *Butyricicoccus pullicaecorum in inflammatory bowel disease.* Gut, 2013. **62**(12): p. 1745-52.

99. Daly, K., et al., *Bacterial sensing underlies artificial sweetener-induced growth of gut Lactobacillus.* Environ Microbiol, 2016. **18**(7): p. 2159-71.

100. Alavala, S., et al., *Stevioside, a diterpenoid glycoside, shows anti-inflammatory property against Dextran Sulphate Sodium-induced ulcerative colitis in mice.* Eur J Pharmacol, 2019. **855**: p. 192-201.

101. Mahalak, K.K., et al., *Impact of Steviol Glycosides and Erythritol on the Human and Cebus apella Gut Microbiome.* J Agric Food Chem, 2020. **68**(46): p. 13093-13101.

102. Canani, R.B., et al., *Potential beneficial effects of butyrate in intestinal and extraintestinal diseases.* World J Gastroenterol, 2011. **17**(12): p. 1519-28.

103. *https://www.precedenceresearch.com/emulsifiers-market.*

104. Watt, J. and R. Marcus, *Carrageenan-induced ulceration of the large intestine in the guinea pig.* Gut, 1971. **12**(2): p. 164-71.

105. Delahunty, T., L. Recher, and D. Hollander, *Intestinal permeability changes in rodents: a possible mechanism for degraded carrageenan-induced colitis.* Food Chem Toxicol, 1987. **25**(2): p. 113-8.

106. Wilcox, D.K., J. Higgins, and T.A. Bertram, *Colonic epithelial cell proliferation in a rat model of nongenotoxin-induced colonic neoplasia.* Lab Invest, 1992. **67**(3): p. 405-11.

107. Grasso, P., et al., *Studies on degraded carrageenan in rats and guinea-pigs.* Food and Cosmetics Toxicology, 1975. **13**(2): p. 195-201.

108. Additives, E.Panel o.F., et al., *Re-evaluation of carrageenan (E 407) and processed Eucheuma seaweed (E 407a) as food additives.* EFSA Journal, 2018. **16**(4): p. e05238.

109. Borthakur, A., et al., *Prolongation of carrageenan-induced inflammation in human colonic epithelial cells by activation of an NFκB-BCL10 loop.* Biochim Biophys Acta, 2012. **1822**(8): p. 1300-7.

110. Yang, B., et al., *Exposure to common food additive carrageenan leads to reduced sulfatase activity and increase in sulfated glycosaminoglycans in human epithelial cells.* Biochimie, 2012. **94**(6): p. 1309-16.

111. Bhattacharyya, S., et al., *A randomized trial of the effects of the no-carrageenan diet on ulcerative colitis disease activity.* Nutr Healthy Aging, 2017. **4**(2): p. 181-192.

112. Nair, D.V.T., et al., *Food Additive Guar Gum Aggravates Colonic Inflammation in Experimental Models of Inflammatory Bowel Disease.* Curr Dev Nutr, 2021. **5**(Suppl 2): p. 1142.

113. Naimi, S., et al., *Direct impact of commonly used dietary emulsifiers on human gut microbiota.* Microbiome, 2021. **9**(1): p. 66.

114. Rousta, E., et al., *The Emulsifier Carboxymethylcellulose Induces More Aggressive Colitis in Humanized Mice with Inflammatory Bowel Disease Microbiota Than Polysorbate-80.* Nutrients, 2021. **13**(10).

115. Martino, J.V., J. Van Limbergen, and L.E. Cahill, *The Role of Carrageenan and Carboxymethylcellulose in the Development of Intestinal Inflammation.* Front Pediatr, 2017. **5**: p. 96.

116. Jin, G., et al., *Maternal Emulsifier P80 Intake Induces Gut Dysbiosis in Offspring and Increases Their Susceptibility to Colitis in Adulthood.* mSystems, 2021. **6**(2).
117. Elmén, L., et al., *Dietary Emulsifier Sodium Stearoyl Lactylate Alters Gut Microbiota in vitro and Inhibits Bacterial Butyrate Producers.* Front Microbiol, 2020. **11**: p. 892.
118. Zangara, M.T., et al., *Maltodextrin Consumption Impairs the Intestinal Mucus Barrier and Accelerates Colitis Through Direct Actions on the Epithelium.* Front Immunol, 2022. **13**: p. 841188.
119. Mirshafiey, A., et al., *Sodium alginate as a novel therapeutic option in experimental colitis.* Scand J Immunol, 2005. **61**(4): p. 316-21.
120. Yamamoto, A., et al., *Effect of sodium alginate on dextran sulfate sodium- and 2,4,6-trinitrobenzene sulfonic acid-induced experimental colitis in mice.* Pharmacology, 2013. **92**(1-2): p. 108-16.
121. Rawi, M.H., et al., *Manipulation of Gut Microbiota Using Acacia Gum Polysaccharide.* ACS Omega, 2021. **6**(28): p. 17782-17797.
122. Changchien, C.H., C.H. Wang, and H.L. Chen, *Konjac glucomannan polysaccharide and inulin oligosaccharide ameliorate dextran sodium sulfate-induced colitis and alterations in fecal microbiota and short-chain fatty acids in C57BL/6J mice.* Biomedicine (Taipei), 2021. **11**(3): p. 23-30.
123. Ishisono, K., et al., *Dietary Fiber Pectin Ameliorates Experimental Colitis in a Neutral Sugar Side Chain-Dependent Manner.* Front Immunol, 2019. **10**: p. 2979.
124. Singh, V., B.S. Yeoh, and M. Vijay-Kumar, *Fermentable Fiber Pectin Improves Intestinal Inflammation by Modulating Gut Microbial Metabolites and Inflammasome Activity.* Curr Dev Nutr, 2020. **4**(Suppl 2): p. 1535.
125. ClinicalTrials.gov, *Clinical Trials Database* 2024.
126. Organisation, W.H., *International Clinical Trials Registry Platform (ICTRP).* 2024.
127. Critch, J., et al., *Use of enteral nutrition for the control of intestinal inflammation in pediatric Crohn disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(2): p. 298-305.
128. Yu, Y., K.C. Chen, and J. Chen, *Exclusive enteral nutrition versus corticosteroids for treatment of pediatric Crohn's disease: a meta-analysis.* World J Pediatr, 2019. **15**(1): p. 26-36.
129. Heuschkel, R.B., et al., *Enteral nutrition and corticosteroids in the treatment of acute Crohn's disease in children.* J Pediatr Gastroenterol Nutr, 2000. **31**(1): p. 8-15.

130. Soo, J., et al., *Use of exclusive enteral nutrition is just as effective as corticosteroids in newly diagnosed pediatric Crohn's disease.* Dig Dis Sci, 2013. **58**(12): p. 3584-91.
131. Luo, J., et al., *Global attitudes on and the status of enteral nutrition therapy for pediatric inflammatory bowel disease.* Front Med (Lausanne), 2022. **9**: p. 1036793.
132. Rolandsdotter, H., et al., *Exclusive Enteral Nutrition: Clinical Effects and Changes in Mucosal Cytokine Profile in Pediatric New Inflammatory Bowel Disease.* Nutrients, 2019. **11**(2).
133. Liu, J., et al., *Benefits of Exclusive Enteral Nutrition in Adults With Complex Active Crohn's Disease: A Case Series of 13 Consecutive Patients.* Crohn's & Colitis 360, 2019. **1**(3).
134. Ruemmele, F.M., et al., *Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease.* J Crohns Colitis, 2014. **8**(10): p. 1179-207.
135. van Rheenen, P.F., et al., *The Medical Management of Paediatric Crohn's Disease: an ECCO-ESPGHAN Guideline Update.* J Crohns Colitis, 2020.
136. Diederen, K., et al., *Exclusive enteral nutrition mediates gut microbial and metabolic changes that are associated with remission in children with Crohn's disease.* Sci Rep, 2020. **10**(1): p. 18879.
137. Day, A.S., *The impact of exclusive enteral nutrition on the intestinal microbiota in inflammatory bowel disease.* AIMS Microbiol, 2018. **4**(4): p. 584-593.
138. Bajaj, A., et al., *Exclusive Enteral Nutrition Mediates Beneficial Gut Microbiome Enrichment in Acute Severe Colitis.* Inflammatory Bowel Diseases, 2023.
139. *OP018. The impact of 'Crohn's Disease-TReatment-with-EATing' diet (CD-TREAT diet) and exclusive enteral nutrition on healthy gut bacteria metabolism.* Journal of Crohn's and Colitis, 2016. **10**(suppl_1): p. S14-S15.
140. Logan, M., et al., *Analysis of 61 exclusive enteral nutrition formulas used in the management of active Crohn's disease-new insights into dietary disease triggers.* Aliment Pharmacol Ther, 2020. **51**(10): p. 935-947.
141. Chassaing, B., et al., *Randomized Controlled-Feeding Study of Dietary Emulsifier Carboxymethylcellulose Reveals Detrimental Impacts on the Gut Microbiota and Metabolome.* Gastroenterology, 2022. **162**(3): p. 743-756.
142. West, K.A., et al., *foodMASST a mass spectrometry search tool for foods and beverages.* NPJ Sci Food, 2022. **6**(1): p. 22.

143. Di Ottavio, F., et al., *A UHPLC-HRMS based metabolomics and chemoinformatics approach to chemically distinguish 'super foods' from a variety of plant-based foods.* Food Chem, 2020. **313**: p. 126071.
144. Gkikas, K., et al., *Take-Home Messages from 20 Years of Progress in Dietary Therapy of Inflammatory Bowel Disease.* Annals of Nutrition and Metabolism, 2023. **79**(6): p. 476-484.
145. ClinicalTrials.gov. *The Intensive Post Exclusive Enteral Nutrition Study (iPENS)*. 2020; Available from: https://classic.clinicaltrials.gov/ct2/show/NCT04225689.
146. Downham, A. and P. Collins, *Coloring our foods in the last and next millenium.* International Journal of Food Science & Technology, 2000. **35**: p. 5-22.
147. Calvo, C. and A. Salvador, *Use of natural colorants in food gels. Influence of composition of gels on their colour and study of their stability during storage.* Food Hydrocolloids - FOOD HYDROCOLLOID, 2000. **14**: p. 439-443.
148. Delgado-Vargas, F., A.R. Jiménez, and O. Paredes-López, *Natural pigments: carotenoids, anthocyanins, and betalains--characteristics, biosynthesis, processing, and stability.* Crit Rev Food Sci Nutr, 2000. **40**(3): p. 173-289.
149. Kirkland, D., et al., *A weight of evidence review of the genotoxicity of titanium dioxide (TiO(2)).* Regul Toxicol Pharmacol, 2022. **136**: p. 105263.
150. Suzuki, T., et al., *Genotoxicity assessment of titanium dioxide nanoparticle accumulation of 90 days in the liver of gpt delta transgenic mice.* Genes and Environment, 2020. **42**(1): p. 7.
151. Chen, T., J. Yan, and Y. Li, *Genotoxicity of titanium dioxide nanoparticles.* J Food Drug Anal, 2014. **22**(1): p. 95-104.
152. Aguilar, F., et al., *Scientific Opinion on the re-evaluation of boric acid (E 284) and sodium tetraborate (borax) (E 285) as food additives EFSA Panel on Food Additives and Nutrient Sources added to Food (ANS).* EFSA journal, 2013. **11**: p. 3407.
153. Lopez, A., et al., *Evolution of Food Safety Features and Volatile Profile in White Sturgeon Caviar Treated with Different Formulations of Salt and Preservatives during a Long-Term Storage Time.* Foods, 2021. **10**(4).

6.0 CHAPTER 6

DISCUSSION

Current investigative, diagnostic, and monitoring methods for chronic inflammatory diseases of the gut, including IBD and CoD, rely on clinical symptoms and invasive biomarkers. While such measures are widely used, they are limited by poor specificity, variability in patient presentation, and frequent delays in achieving a definitive diagnosis [1-3]. In recent years, LC-MS-based metabolomics has emerged as a powerful complementary approach, capable of capturing disease-associated metabolic interactions and offering new insights into pathophysiology, biomarker discovery, and therapeutic response [4-7]. By profiling small molecules that represent the downstream products of cellular metabolism, metabolomics enables the characterisation of biochemical perturbations more directly linked to disease processes. Importantly, this analysis can be performed on non-invasive biomatrices such as stool and urine, which are particularly valuable for longitudinal monitoring and large-scale clinical studies. The use of stool and urine as biomatrices therefore not only facilitates biomarker discovery but also supports the development of non-invasive tools for disease diagnosis, patient stratification, and treatment monitoring.

However, despite the promise of LC-MS for advancing gut metabolomics, progress is hindered by the lack of standardised methodologies across laboratories [8, 9], which limit reproducibility, comparability of findings, and ultimately the translation of biomarkers into clinical practice. Addressing these challenges is essential for improving data reproducibility and clinical translation. In this thesis, an in-depth multi-parameter optimisation of LC-MS-based metabolomics methods designed for the analysis of human stool and urine samples was conducted. The developed methods were subsequently applied to characterise the metabolic phenotype of gastrointestinal disease and includes a clinical study investigating mechanisms of CoD and the analysis of the largest urinary metabolomics investigation of IBD.

A powerful contribution of this thesis is the establishment of robust, high-throughput LC-MS pipelines for non-invasive biomarkers. The optimisation of a faecal extraction method identified a reproducible protocol that enabled consistent metabolite detection and

stratification of patient groups. The use of 50 mg freeze-dried faecal samples in a 1000 μL MeOH and bead beating extraction can be recommended as an optimal extraction strategy for metabolomics analysis, which maximises metabolite coverage and reproducibility while maintaining compatibility with both untargeted and targeted workflows. Importantly, the utility of this optimised approach was demonstrated through its application to a clinical cohort of paediatric CoD patients, where it facilitated the characterisation of disease-associated metabolic alterations and the identification of treatment responsive metabolic signatures following adherence to a GFD. While previous studies have focused largely on serum or plasma metabolomics, this work highlights the value of faecal sample analysis, which captures host-microbe metabolic interactions and directly reflect the gut environment. Our analyses revealed three distinct patterns of metabolic signatures. Firstly, we observed a panel of 12 CoD-associated metabolites that were unaffected by treatment, with persistent disruptions across bile acids and amino acid derivatives. A group of treatment-responsive metabolites were also identified, including amino acid dipeptides and indole- and purine- related metabolites, which normalised following dietary treatment. Finally, a third group of metabolites emerged that were non-disease-specific and shaped primarily by the GFD itself, such as indole-derived compounds, purine intermediates, and acylcarnitines, reflecting the broader metabolic consequence of dietary intervention. Longitudinal analysis extended these findings, showing that the most pronounced shifts occurred within 6 months of dietary treatment, coinciding with improved tTG and PedsQL-GS scores, but were less evident at 12 months. The methodological advances established in this work therefore not only strengthen the pipeline for biomarker discovery but also show potential for their application in non-invasive tests, disease monitoring, and the stratification of patients at risk of refractory disease. Such insights may ultimately guide personalised medicine and nutrition, as well as the development of adjunctive or alternative therapies.

Beyond the methodological and translational significance of these findings, this work also offers novel biological insights into CoD pathophysiology. The identification of CoD-specific, treatment-unresponsive metabolites is a particularly important finding, as it suggests that fundamental disturbances in amino acid turnover and bile acid metabolism persist despite strict adherence to a GFD. This challenges the prevailing assumption that dietary exclusion alone is sufficient to restore intestinal homeostasis and instead indicates that metabolic

dysregulation may represent an intrinsic feature of CoD pathophysiology. Such persistent alterations could contribute to ongoing low-grade inflammation, incomplete mucosal healing, or long-term complications, thereby offering mechanistic explanations for why some patients remain symptomatic or progress to refractory disease. Equally important are the treatment-responsive group of metabolites, which demonstrate that faecal metabolomics can capture early biochemical improvements following dietary therapy. Moreover, the detection of diet-driven, non-disease-specific metabolites highlights the broader metabolic footprint of a GFD, offering important context for interpreting metabolomics data in treated CoD patients and pointing to possible unintended nutritional or microbial consequences of long-term dietary restriction.

The approach to developing a metabolomics workflow was broadened to urine, expanding the potential of non-invasive biomarker discovery across multiple biological matrices. A rapid chromatographic separation with a 6.5-minute data collection time was developed, as determined by a range of peak performance quality factors. The potential of scaling the LC-MS protocol was again demonstrated within a clinical trial setting, illustrating the practicality of applying metabolomics at scale. The reduction in overall LC-MS analysis time for untargeted metabolomics highlights the potential to streamline workflows, enabling high-throughput, cost-effective implementation in large-scale clinical trials and enhancing the feasibility of metabolomics as a routine tool for patient stratification and disease monitoring.

The scope of this thesis was extended to a dietary context through a critical evaluation of food additives as potential modulators of gut inflammation, situating the findings within the wider framework of diet-gut interactions and their implications for IBD management. The effect of food additives on gut health is an important consideration for both the development and management of IBD, and in this work we reframe the traditional view of food additives by understanding the mechanisms of additives in the context of IBD. Importantly, we demonstrate that the roles of food additives can be of either an inflammatory or anti-inflammatory nature depending on their mechanism, which may bring important implications for the treatment and management of IBD. This is critical as there remains an unmet need for patient dietary advice and therapies, predominantly due to the lack of translation from preclinical studies to human RCTs [10]. Furthermore, we address the

requirement for safer alternatives in food products and the challenges that exist to achieve this and set out perspectives on future research. Ultimately, by addressing the conflicting evidence of food additives on gut inflammation with different models and doses, we highlight existing gaps in our understanding of their mechanisms and future areas of investigation.

This section of work highlights an important issue in gastrointestinal disease research, which is the complex relationship between diet and human health [11-13]. Dietary intake influences both host metabolism and the gut microbiome, making it a major determinant of metabolic readouts in biomatrices such as stool and urine. In the context, diet represents both a variable of interest and a potential cofounder. For example, in the CoD cohort, adherence to a GFD was not only the primary therapeutic intervention but also a significant driver of metabolic changes, complicating the separation of treatment effects from disease mechanisms. Similarly, in the IBD studies, differences in dietary intake may contribute to metabolic heterogeneity across patient groups, underscoring the importance of integrating metadate with metabolomics analysis. The review of food additives further reinforces this point by illustrating how individual dietary components can exert inflammatory or therapeutic effects, thereby shaping metabolic outcomes. Addressing the diet in greater detail in studies of gut disease is therefore essential for advancing metabolomics-based biomarker discovery and for ensuring that identified signatures are truly reflective of disease rather than dietary intervention.

Several strengths underpin this work. First, the systematic optimisation of multiple method parameters across stool and urine matrices represents a comprehensive evaluation of LC-MS workflows in gastrointestinal disease. The combined optimisation and application of these workflows help bridge the gap between methodological development and clinical utility, as demonstrated by their use in large-scale clinical cohorts. Furthermore, the integration of clinical metadata, objective adherence markers, and metabolic outcomes exemplifies the multi-layered approach necessary for translational success. This is evident in the CoD cohort, where treatment-responsive metabolic signatures were characterised in parallel with improvements in serological markers (tTG), quality of life measures (PedsQL-GS), and objective measures of gluten intake (gluten immunogenic peptide). The ability to link metabolomic findings with patient-reported and clinical outcome strengthens the

clinical relevance of the results. Importantly, this work also extends beyond single matrix analyses, with stool and urine samples optimised and prepared for metabolomics, setting the foundation for future multi-omics integration. Finally, the inclusion of a detailed insights into the role of food additives in gastrointestinal health broadens the scope of the thesis, highlighting the importance of diet and potential therapeutic strategies in disease management. These strengths position this work at the intersection of methodological innovation and translational application, demonstrating the potential of LC-MS metabolomics to move towards clinical impact in gut health research.

The limitations of this thesis must also be acknowledged. Metabolite identification confidence of the features identified in this work were at MSI level 2 [14], underscoring the need for expanded in-house reference libraries. The overall size for the prospective CoD cohort limited the statistical power of the longitudinal analysis, highlighting the importance of validation in larger prospective studies. Additionally, method development in this work was carried out on a ThermoFisher Orbitrap 240 system; however, to enhance the broader applicability of the protocol, it is important to evaluate whether the optimised protocols can be reproduced across different LC-MS platforms and laboratory settings. A further limitation is the absence of direct clinical validation of the identified metabolic signatures, which is essential to establish their robustness, reproducibility, and added diagnostic or prognostic value across diverse patient populations and clinical settings [15]. Therefore, while this thesis demonstrates proof-of-concept applications in IBD and CoD, the translation of metabolomic biomarkers into routine clinical practice requires rigorous testing, integration with established clinical endpoints, and demonstration of diagnostic or prognostic utility. Without this level of clinical validation, the biomarkers identified remain exploratory. To progress towards clinical adoption, candidate biomarkers must meet several critical requirements [16]. Biomarkers must demonstrate robustness and reliability, being detectable despite potential dietary or environmental confounders. Equally important are stability in biofluids and analytical performance, such as accuracy, precision, sensitivity, and specificity under laboratory conditions. This includes disease specificity, as it is crucial that a biomarker can distinguish a disease such as IBD from other conditions that also present with gut inflammation. This is important as generic markers of intestinal inflammation (e.g., faecal calprotectin and C-reactive protein (CRP)) often lack discriminatory power and can be

elevated across multiple diseases [17]. A successful biomarker must therefore be unique to the disease of interest rather than reflecting non-specific inflammation.

The progression of this research can be viewed as a pipeline of interconnected methodological and clinical contributions. The large clinical study cohorts recruited patients between 2020 and 2024, which generated a substantial collection of faecal, urine, plasma, and dietary samples, together with detailed clinical and nutritional metadata, creating a unique source for metabolomics research. The early stages of this thesis centred on optimising faecal extraction protocols, where a reproducible extraction method was developed to establish a foundation for standardised LC-MS analysis of gastrointestinal disease. Building on this methodological framework, subsequent work applied the optimised approach to clinical cohorts on paediatric CoD patients. At the same time, work was conducted on the optimisation of a rapid dilute-and-shoot LC-MS workflow that enabled high-throughput application in large-scale clinical studies. These datasets will ultimately be integrated across multiple biomatrices and combined with microbiome and immunological data, in addition to dietary and clinical metadata to advance a multi-omics framework for understanding gastrointestinal disease mechanisms and guiding personalised nutritional interventions.

Conclusions

This thesis demonstrated how carefully optimised LC-MS methods can deliver biological insights that extend beyond descriptive profiling, providing a foundation for translational research in gastrointestinal disease. The development of robust workflows for stool and urine provided reproducible pipelines that enabled large-scale analyses. By applying these optimised methods to IBD and CoD patient cohorts, this work highlights the value of mass spectrometry-based metabolomics in advancing mechanistic understanding of gastrointestinal disease, identifying specific metabolites implicated in pathophysiology and treatment response. These studies illustrate how accelerated LC-MS protocols can make high-throughout, clinically embedded metabolomics achievable, a prerequisite for integration into patient stratification and monitoring strategies. Together with continued collaborative clinical efforts and technological advancements, this work contributes to the

establishment of standardised LC-MS frameworks, paving the way for robust biomarker discovery, mechanistic insight, and clinical application in gastrointestinal health and disease.

6.1 References

1. Mosli, M.H., et al., *C-Reactive Protein, Fecal Calprotectin, and Stool Lactoferrin for Detection of Endoscopic Activity in Symptomatic Inflammatory Bowel Disease Patients: A Systematic Review and Meta-Analysis.* Am J Gastroenterol, 2015. **110**(6): p. 802-19; quiz 820.
2. Cross, E., et al., *Diagnostic delay in adult inflammatory bowel disease: A systematic review.* Indian J Gastroenterol, 2023. **42**(1): p. 40-52.
3. Norström, F., et al., *Delay to celiac disease diagnosis and its implications for health-related quality of life.* BMC Gastroenterol, 2011. **11**: p. 118.
4. Füreder, J., et al., *Metabolomics-enabled biomarker discovery in breast cancer research.* Trends in Endocrinology & Metabolism, 2025.
5. Zhang, D., et al., *LC-MS/MS based metabolomics and proteomics reveal candidate biomarkers and molecular mechanism of early IgA nephropathy.* Clinical Proteomics, 2022. **19**(1): p. 51.
6. Torres, C.L., et al., *Untargeted LC-HRMS metabolomics reveals candidate biomarkers for mucopolysaccharidoses.* Clin Chim Acta, 2023. **541**: p. 117250.
7. Vo, D.K. and K.T.L. Trinh, *Emerging Biomarkers in Metabolomics: Advancements in Precision Health and Disease Diagnosis.* Int J Mol Sci, 2024. **25**(23).
8. Fux, E., et al., *A global perspective on the status of clinical metabolomics in laboratory medicine – a survey by the IFCC metabolomics working group.* Clinical Chemistry and Laboratory Medicine (CCLM), 2024. **62**(10): p. 1950-1961.
9. Begou, O., et al., *Hyphenated MS-based targeted approaches in metabolomics.* Analyst, 2017. **142**(17): p. 3079-3100.
10. Halmos, E.P. and P.R. Gibson, *Dietary management of IBD--insights and advice.* Nat Rev Gastroenterol Hepatol, 2015. **12**(3): p. 133-46.
11. Zhang, P., *Influence of Foods and Nutrition on the Gut Microbiome and Implications for Intestinal Health.* Int J Mol Sci, 2022. **23**(17).
12. Ross, F.C., et al., *The interplay between diet and the gut microbiome: implications for health and disease.* Nature Reviews Microbiology, 2024. **22**(11): p. 671-686.
13. Tang, Z.Z., et al., *Multi-Omic Analysis of the Microbiome and Metabolome in Healthy Subjects Reveals Microbiome-Dependent Relationships Between Diet and Metabolites.* Front Genet, 2019. **10**: p. 454.

14. Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics, 2007. **3**(3): p. 211-221.
15. Trezzi, J.P., K. Hiller, and B. Mollenhauer, *The importance of an independent validation cohort for metabolomics biomarker studies.* Mov Disord, 2018. **33**(5): p. 856.
16. Zander, H. and J. Engelbergs, *Requirements for regulatory acceptance of biomarkers.* Allergo Journal International, 2024. **33**(8): p. 309-312.
17. Vermeire, S., G. Van Assche, and P. Rutgeerts, *Laboratory markers in IBD: useful, magic, or unnecessary toys?* Gut, 2006. **55**(3): p. 426-31.

7.0 APPENDICES

7.1 APPENDIX 1

SUPPLEMENTARY INFORMATION FOR

An Optimised Monophasic Faecal Extraction Method for LC-MS Analysis and Its Application in Gastrointestinal Disease

Patricia E. Kelly [1,2] H Jene Ng [3], Gillian Farrell [1], Shona McKirdy [2,3], Richard K. Russell [2,4], Richard Hansen [2,5], Zahra Rattray [1], Konstantinos Gerasimidis [2,3] and Nicholas J. W. Rattray[1,2]*

Published in Metabolites

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G1 1XQ, UK.

[2] Department of Paediatric Gastroenterology Bacteria, Immunology, Nutrition, Gastroenterology and Omics Group, Royal Hospital for Children; University of Glasgow, Glasgow G12 8QQ, UK.

[3] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow G12 8QQ, UK.

[4] Royal Hospital for Children and Young People, 50 Little France Cresent, Edinburgh EH16 4TJ, UK.

[5] Royal Hospital for Children, 1345 Govan Road, Glasgow G52 4TF, UK.

Table S1. Untargeted metabolomics experiment elution gradient. Mobile phase A, 99.9% water + 0.1% formic acid; Mobile B, 99.99% MeOH + 0.1% formic acid.

Time (min)	Mobile Phase A (%)	Mobile Phase B (%)	Flow rate (mL/min)
0.0	99.0	1.0	0.4
0.5	99.0	1.0	0.4
2.0	50.0	50.0	0.4
10.5	1.0	99.0	0.4
11.0	1.0	99.0	0.4
11.5	99.0	1.0	0.4
14.9	99.0	1.0	0.4
15.0	99.0	1.0	0.4

Table S2. Targeted metabolomics experiment elution gradient. Mobile phase A, 99.9% H_2O + 0.1% formic acid; Mobile phase B, 99.9% Acetonitrile + 0.1% formic acid.

Time (min)	Mobile Phase A (%)	Mobile Phase B (%)	Flow rate (mL/min)
0	100	0	0.4
2	100	0	0.4
5	75	25	0.4
11	65	35	0.4
15	5	95	0.4
20	5	95	0.4
20.1	100	0	0.4
20.5	100	0	0.4

Table S3. Overview of Untargeted Metabolite Identification Levels

	Number of Metabolites
MSI Identification Level 2	424
MSI Identification Level 3	267

Table S4. List of metabolites included in targeted metabolomics method.

Name	Molecular Formula	Classification	Precursor m/z	Product m/z	Retention Time	Ref.(1) Precursor m/z	Ref.(1) Product m/z	Target Q1 Pre Bias	Target Collision Energy	Target Q3 Pre Bias	Ref.(1) Q1 Pre Bias	Ref.(1) Collision Energy	Ref.(1) Q3 Pre Bias
2-Aminobutyric acid	$C_4H_9NO_2$	Organic acid	104.1	58.05	2.831	104.1	41.05	-26	-12	-11	-26	-26	-17
2-Ketoglutaric acid	C_5H_6O5	Organic acid	144.9	101.1	2.317	144.9	57.05	23	12	18	23	13	21
4-Aminobutyric acid	$C_4H_9NO_2$	Organic acid	104.1	87.05	3.69	104.1	45.1	-28	-14	-17	-28	-22	-18
4-Hydroxyproline	$C_5H_9NO_3$	Amino acid	132.1	86.05	1.991	132.1	68.05	-10	-15	-18	-10	-22	-13
Acetylcarnitine	$C_9H_{17}NO_4$	Peptide	204.1	85.05	8.929	204.1	60.1	-16	-22	-18	-16	-16	-12
Acetylcholine	$C_7H_{16}NO_2$	Lipid	147.1	87.05	9.165	147.1	88.05	-12	-16	-17	-12	-16	-17
Aconitic acid	$C_6H_6O_6$	Choline	172.9	85.05	3.536	172.9	129.1	14	14	16	14	13	12
Adenine	$C_5H_5N_5$	Organic acid	136	119.05	6.46	136	65	-10	-26	-23	-10	-41	-13

Adenosine	$C_{10}H_{13}N_5O_4$	Purine base	268.1	136.05	6.764	268.1	119	-21	-18	-15	-21	-47	-23
Adenosine 3',5'-cyclic monophosphate	$C_{10}H_{12}N_5O_6P$	Nucleoside	330	136.05	6.179	330	119.1	-26	-26	-30	-26	-54	-23
Adenosine monophosphate	$C_{10}H_{14}N_5O_7P$	Nucleotide	348	136.05	2.969	348	97.1	-13	-20	-28	-13	-31	-20
Adenylsuccinic acid	$C_{14}H_{18}N_5O_{11}P$	Nucleotide	464.1	252.1	6.183	464.1	162	-18	-21	-18	-18	-47	-17
Alanine	$C_3H_7NO_2$	Organic acid	157	97.1	1.927	157	42.05	-22	-12	-18			
Allantoin	$C_4H_6N_4O_3$	Amino acid	175.1	70.1	3.365	175.1	60.1	18	15	18	18	10	15
Arginine	$C_6H_{14}N_4O_2$	Purine derivative	291	70.1	3.057	291	116.05	-13	-23	-13	-13	-16	-12
Argininosuccinic acid	$C_{10}H_{18}N_4O_6$	Amino acid	133.1	87.15	1.953	133.1	28.05	-24	-35	-14	-24	-21	-25
Asparagine	$C_4H_8N_2O_3$	Organic acid	134	74.05	1.953	134	88.1	-20	-12	-18	-20	-29	-30

Aspartic acid	$C_4H_7NO_4$	Amino acid	134	74.05	1.953	134	88.1	-30	-15	-14	-30	-13	-17
Asymmetric dimethylarginine	$C_8H_{18}N_4O_2$	Amino acid	203.1	70.1	7.207	203.1	46.1	-17	-25	-13	-17	-17	-19
Carnitine	$C_7H_{16}NO_3$	Amino acid	162.1	103.05	5.284	162.1	60.1	-13	-18	-22	-13	-17	-12
Carnosine	$C_9H_{14}N_4O_3$	Amino acid derivative	227.1	110.05	5.365	227.1	156.05	-18	-24	-23	-18	-16	-17
Cholic acid	$C_{24}H_{40}O_5$	Peptide	407.2	343.15	14.051	407.2	345.25	13	34	24	13	32	24
Choline	$C_5H_{14}NO$	Organic acid	104.1	60.05	4.436	104.1	45.1	-27	-22	-11	-27	-23	-18
Citicoline	$C_{14}H_{26}N_4O_{11}P_2$	Choline	489.1	184.1	2.045	489.1	264.05	-20	-43	-20	-20	-25	-30
Citric acid	$C_6H_8O_7$	Nucleotide	191.2	111.1	3.209	191.2	87.05	12	13	21	12	20	16
Citrulline	$C_6H_{13}N_3O_3$	Organic acid	176.1	70.05	2.321	176.1	159.05	-12	-25	-14	-12	-14	-18

Creatine	$C_4H_9N_3O_2$	Amino acid	132.1	44.05	3.431	132.1	90.05	-11	-22	-18	-11	-15	-18
Creatinine	$C_4H_7N_3O$	Organic acid	114.1	44.05	4.82			-10	-19	-18			
Cystathionine	$C_7H_{14}N_2O_4S$	Lactam	223	88.05	2.028	223	134	-17	-27	-18	-17	-15	-15
Cysteamine	C_2H_7NS	Amino acid	78.1	61.05	3.98			-19	-13	-25			
Cysteine	$C_3H_7NO_2S$	Aminothiol	122	76.05	2.148	122	59	-29	-16	-16	-29	-25	-23
Cystine	$C_6H_{12}N_2O_4S_2$	Amino acid	241	151.95	1.908	241	73.9	-19	-14	-17	-19	-29	-15
Cytidine	$C_9H_{13}N_3O_5$	Amino acid	244.1	112.05	6.393	244.1	95	-19	-13	-23	-19	-42	-19
Cytidine 3',5'-cyclic monophosphate	$C_9H_{12}N_3O_7P$	Nucleoside	306	112.1	4.093			-11	-22	-22			
Cytidine monophosphate	$C_9H_{14}N_3O_8P$	Nucleotide	324	112.05	2.26	324	95	-26	-14	-23	-26	-54	-19

Cytosine	$C_4H_5N_3O$	Nucleotide	112	95.1	4.044			-30	-23	-20			
Dimethylglycine	$C_4H_9NO_2$	Amino acid	104.1	58.05	2.189	104.1	44.05	-12	-16	-11	-12	-38	-17
Dopa	$C_9H_{11}NO_4$	Amino acid	198.1	152.1	6.278			-30	-14	-11			
Dopamine	$C_8H_{11}NO_2$	Amino acid derivative	154.1	91.05	8.078	154.1	137.05	-13	-27	-19	-13	-15	-15
Epinephrine	$C_9H_{13}NO_3$	Catecholamine	184.1	166.1	7.164	184.1	77	-15	-12	-19	-15	-44	-15
FAD	$C_{27}H_{33}P_2N_9O_{15}$	Catecholamine	786.15	136.1	6.213	786.15	348.1	-32	-47	-28	-32	-23	-26
FMN	$C_{17}H_{21}N_4O_9P$	Coenzyme	455	97	6.193	455	78.9	24	28	18	24	38	14
Fumaric acid	$C_4H_4O_4$	Coenzyme	115	71.1	4.571	115	26.95	12	11	12	12	14	26
Glutamic acid	$C_5H_9NO_4$	Organic acid	147.9	84.1	2.253	147.9	56.1	-11	-17	-17	-11	-30	-23

Glutamine	$C_5H_{10}N_2O_3$	Amino acid	147.1	84.15	2.073	147.1	130.1	-11	-18	-17	-11	-16	-27
Glutathione	$C_{10}H_{17}N_3O_6S$	Amino acid	308	179.1	4.543			-25	-13	-13			
Glycine	$C_2H_5NO_2$	Peptide	75.9	30.15	2.029			-17	-11	-30			
Guanine	$C_5H_5N_5O$	Amino acid	150	133	5.623	150	66.1	17	19		17	30	
Guanosine	$C_{10}H_{13}N_5O_5$	Purine base	284	152	6.187	284	135	-22	-12	-17	-22	-39	-15
Guanosine 3',5'-cyclic monophosphate	$C_{10}H_{12}N_5O_7P$	Nucleoside	346	152.05	5.393	346	135.05	-28	-22	-17	-28	-48	-27
Guanosine monophosphate	$C_{10}H_{14}N_5O_8P$	Nucleotide	364	152.05	2.552	364	135	-30	-17	-17	-30	-49	-27
Histamine	$C_5H_9N_3$	Nucleotide	112.1	95.05	5.803	112.1	41.05	-30	-17	-20	-30	-29	-17
Histidine	$C_6H_9N_3O_2$	Amino acid derivative	155.9	110.1	2.901	155.9	56.1	-18	-15	-23	-18	-35	-22

Homocysteine	$C_4H_9NO_2S$	Amino acid	136	90.1	3.188	136	56.1	-10	-13	-18	-10	-22	-21
Homocystine	$C_8H_{16}N_2O_4S_2$	Amino acid	3	1	269	136.05	4.321	-21	-11	-15	-21	-34	-19
Hypoxanthine	$C_5H_4N_4O$	Amino acid	137	55.05	4.251	137	110	-10	-32	-22	-10	-22	-23
Inosine	$C_{10}H_{12}N_4O_5$	Purine derivative	269.1	137.05	6.211	269.1	118.95	-23	-10	-15	-23	-41	-24
Isocitric acid	$C_6H_8O_7$	Nucleoside	191.2	111.1	2.358	191.2	73	12	15	20	12	24	27
Isoleucine	$C_6H_{13}NO_2$	Organic acid	132.1	86.2	7.241	132.1	69.15	-30	-12	-17	-30	-19	-14
Kynurenine	$C_{10}H_{12}N_2O_3$	Amino acid	209.1	192.05	8.34	209.1	94.1	-18	-11	-22	-18	-14	-19
Lactic acid	$C_3H_6O_3$	Amino acid derivative	89.3	89.05	2.795			10	7	17			

Leucine	$C_6H_{13}NO_2$	Organic acid	132.1	86.05	7.52	132.1	30.05	-30	-12	-17	-30	-18	-29
Lysine	$C_6H_{14}N_2O_2$	Amino acid	147.1	84.1	2.894			-11	-18	-18			
Malic acid	$C_4H_6O_5$	Amino acid	133.1	114.95	2.358	133.1	71.15	18	17	24	18	17	26
Methionine	$C_5H_{11}NO_2S$	Organic acid	149.9	56.1	5.304	149.9	104.1	-11	-18	-11	-11	-14	-21
Methionine sulfone	$C_5H_{11}NO_4S$	Amino acid	180	79.2	2.184			19	15	14			
Methionine sulfoxide	$C_5H_{11}NO_3S$	Amino acid	166	74.1	2.206	166	55.95	-12	-14	-15	-12	-25	-22
NAD	$C_{21}H_{27}N_7O_{14}P_2$	Coenzyme	663.1	541.05	3.882	663.1	540.1	26	17	38	26	17	26
Niacinamide	$C_6H_6N_2O$	Vitamin	123.1	80.05	5.344	123.1	53.1	-10	-23	-16	-10	-31	-21
Nicotinic acid	$C_6H_5NO_2$	Organic acid	124.05	80.05	4.08	124.05	78.05	-17	-22	-15	-17	-24	-15
Norepinephrine	$C_8H_{11}NO_3$	Catecholamine	170.1	152.15	4.988	170.1	107.1	-14	-10	-17	-14	-21	-22

Ophthalmic acid	$C_{11}H_{19}N_3O_6$	Organic acid	290.1	58.1	5.35	290.1	161.1	-24	-23	-23	-24	-13	-18
Ornitine	$C_5H_{12}N_2O_2$	Amino acid	133.1	70.1	2.679	133.1	116.05	-10	-18	-14	-10	-15	-24
Orotic acid	$C_5H_4N_2O_4$	Organic acid	155	111.1	2.588	155	42.1	17	13	22	17	22	15
Oxidized glutathione	$C_{20}H_{32}N_6O_{12}S_2$	Peptide	611.1	306	6.253	611.1	143.05	24	24	20	24	48	28
Pantothenic acid	$C_9H_{17}NO_5$	Organic acid	220.1	90.15	6.249	220.1	72.05	-18	-15	-18	-18	-23	-14
Phenylalanine	$C_9H_{11}NO_2$	Amino acid	166.1	120.1	8.068	166.1	103.1	-12	-15	-24	-12	-29	-20
Proline	$C_5H_9NO_2$	Amino acid	116.1	70.15	2.609	116.1	28.05	-30	-18	-14	-30	-35	-30
Pyruvic acid	$C_3H_4O_3$	Organic acid	86.9	87.05	2.585	86.9	42.95	12	7	16	12	12	15
S-Adenosylhomocysteine	$C_{14}H_{20}N_6O_5S$	Amino acid derivative	385.1	134	8.197	385.1	136.05	-15	-21	-15	-15	-21	-29

S-Adenosylmethionine	$C_{15}H_{22}N_6O_5S$	Amino acid derivative	399.1	250.05	6.939	399.1	136.1	-15	-16	-18	-15	-30	-29
Serine	$C_3H_7NO_3$	Amino acid	105.9	60.1	1.96			-25	-12	-11			
Serotonin	$C_{10}H_{12}N_2O$	Amino acid derivative	177.1	160.1	10.527	177.1	77.05	-15	-13	-18	-15	-49	-14
Succinic acid	$C_4H_6O_4$	Organic acid	117.3	73	4.055	117.3	99.05	13	13	28	13	14	19
Symmetric dimethylarginine	$C_8H_{18}N_4O_2$	Amino acid	203.1	70.15	6.817	203.1	71.1	-17	-27	-13	-17	-27	-14
Taurocholic acid	$C_{26}H_{45}NO_7S$	Organic acid	514.2	107.1	7.781	514.2	124.05	30	55	20	30	55	25
Threonine	$C_4H_9NO_3$	Amino acid	120.1	74.15	2.133	120.1	56.05	-27	-13	-14	-27	-17	-11

Thymidine	$C_{10}H_{14}N_2O_5$	Nucleoside	243.1	127.1	6.175			-18	-12	-14			
Thymidine monophosphate	$C_{10}H_{13}N_2O_8P_2$	Nucleotide	322.9	81.1	3.07	322.9	207.1	-25	-22	-16	-25	-9	-15
Thymine	$C_5H_6N_2O_2$	Pyrimidine base	127.1	54.05	5.448	127.1	110.05	-11	-29	-23	-11	-8	-21
Tryptophan	$C_{11}H_{12}N_2O_2$	Amino acid	205.1	188.15	10.092	205.1	146.1	-16	-12	-23	-16	-18	-17
Tyrosine	$C_9H_{11}NO_3$	Amino acid	182.1	136.1	6.694	182.1	91.1	-14	-15	-27	-14	-30	-18
Uracil	$C_4H_4N_2O_2$	Nucleoside	113	70	2.986			-20	-17	-13			
Uric acid	$C_5H_4N_4O_3$	Organic acid	167.1	123.95	3.159	167.1	96.2	12	19	24	12	19	17
Uridine	$C_9H_{12}N_2O_6$	Nucleoside	245	113.05	4.444			-19	-10	-23			
Valine	$C_9H_{12}N_2O_6$	Amino acid	118.1	72.15	4.761	118.1	55.05	-27	-13	-15	-27	-24	-11
Xanthine	$C_5H_4N_4O_2$	Purine base	151	108	4.093	151	42	17	20	20	17	21	15

Table S5. Parameters of Compound Discoverer workflow.

Workflow Node	Workflow Parameter	Workflow Information
Input files		.raw data
Select Spectra	Spectrum Properties Filter	Lower RT limit: 0 Upper RT limit: 0
	Scan Event Filters	Polarity Mode: Is +
Align Retention Times	General Settings	Alignment Model: Adaptive curve Maximum Shift [min]: 0.3 Mass Tolerance: 2 ppm
Detect Compounds	General Settings	Mass Tolerance: 2 ppm Intensity Tolerance [%]: 30 S/N Threshold: 5 Min. Peak Intensity: 500 000 Ions: [2M + FA + H]-1; [2M + H]+1; [2M + K]+1; [2M + Na]+1; [2M - H]-1; [M + 2H]+2; [M + Cl]-1; [M + FA - H]-1; [M + H]+1; [M + H + K]+2; 2M + H + MeOH]+1; [M + H + Na]+2; [M + H – H_2O]+1; [M + K]+1; [M + Na]+1; [M - 2H]-2; [M + 2H + K]-1; [M -H]-1; [M -H_2O]-1 Min. Element Counts: C H Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O15 P5 S5
Group Compounds	Compound Consolidation	Mass Tolerance: 2 ppm RT Tolerance [min]: 0.2
	Fragment Data Selection	Preferred Ions: [M - H] +1; [M-H] -1

Search mzCloud	General Settings	Compound Classes: Endogenous Metabolites, Excipients/ Additives/ Colourants, Extractables/ Leachables, Natural Products/ Medicines, Natural Toxins, Personal Care Products/ Cosmetics, Small Molecule Chemicals, Steroids/ Vitamins/ Hormones, Therapeutic/ Prescription Drugs Precursor Mass Tolerance: 10 ppm FT Fragment Mass Tolerance: 10 ppm Library: Autoprocessed, Reference Post. Processing: Recalibrated Annotation Matching. Fragments: True
	DDA Search	Identity Search: Cosine Match Activation Type: True Match Activation Energy: Match with Tolerance Activation Energy: 20 Apply Intensity Threshold: True Similarity Search: None Match Factor Threshold: 60
	DIA Search	Use DIA Scans: False Max. Isolation Width [Da]: 500 Match Activation Type: False Match Activation Energy: Any

		Activation Energy Tolerance: 100 Apply Intensity Threshold: False Match Factor Threshold: 20
Predict Compositions	Prediction Settings	Mass Tolerance: 2 ppm Min. Element Counts: C. H Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O15 P5 S5 Min. RDBE: 0 Max. RDBE: 40 Min. H/C: 0.1 Max. H/C: 4 Max.# Candidates: 10 Max.# Internal Candidates: 200
	Pattern Matching	Intensity Tolerance [%]: 30 Intensity Threshold [%]: 0.1 S/N Threshold: 3 Min. Spectral Fit [%]: 30 Min. Pattern Cov. [%]: 90 Use Dynamic Recalibration: True
	Fragments Matching	Use Fragments: True Mass Tolerance: 2 ppm S/N Threshold: 5
Map to Metabolika Pathways	Search Settings	Metabolika pathways: All Search Mode: By Formula or Mass
	By Mass Search Settings	Mass Tolerance: 2 ppm

	By Formula Search Settings	Max. # of Predicted Compositions to be searched per Compound: 3
	Display Settings	Max. # of Pathways in 'Pathways' column: 20
Apply mzLogic	General Settings	FT Fragment Mass Tolerance: 10 ppm IT Fragment Mass Tolerance: 0.4 Da Max. # Compounds: 0 Max. # mzCloud Similarity Results to consider per Compound: 10 Match Factor Threshold: 30
Assign Compound Annotations	General Settings	Mass Tolerance: 2 ppm
	Data Sources	Data source #1: mzCloud Search Data source #2: Predicted Compositions Data source #3: MassList Search Data source #4: ChemSpider Search Data source #5: Metabolika Search
	Sorting Rules	Use mzLogic: True Use Spectral Distance: True SFit Threshold: 20 SFit Range: 20
Fill Gaps	General Settings	Mass Tolerance: 2 ppm S/N Threshold: 5 Use Real Peak Detection: True
Apply QC Correction	General Settings	Min. QC Coverage [%]: 30

		Max. QC Area. RSD [%]: 30 Max. Corrected QC Area RSD [%]: 25 Max. # Files Between QC Files: 15
Mark Background Compounds	General Settings	Max. Sample/ Blank: 5 Max. Blank/ Sample: 0 Hide Background: True
Differential Analysis	General Settings	Log10 Transform Values: True

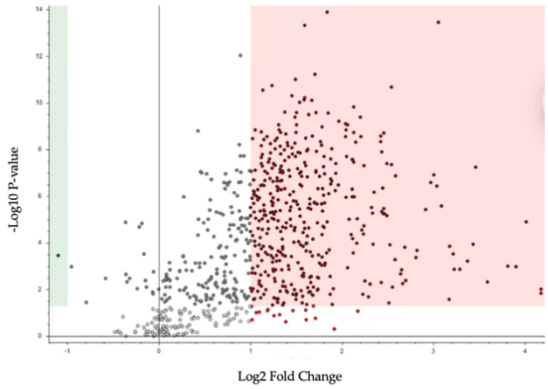


Figure S1. Untargeted differential analysis of sample weight showing volcano plot of altered metabolites, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased in 50 mg samples compared to 20 mg samples are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

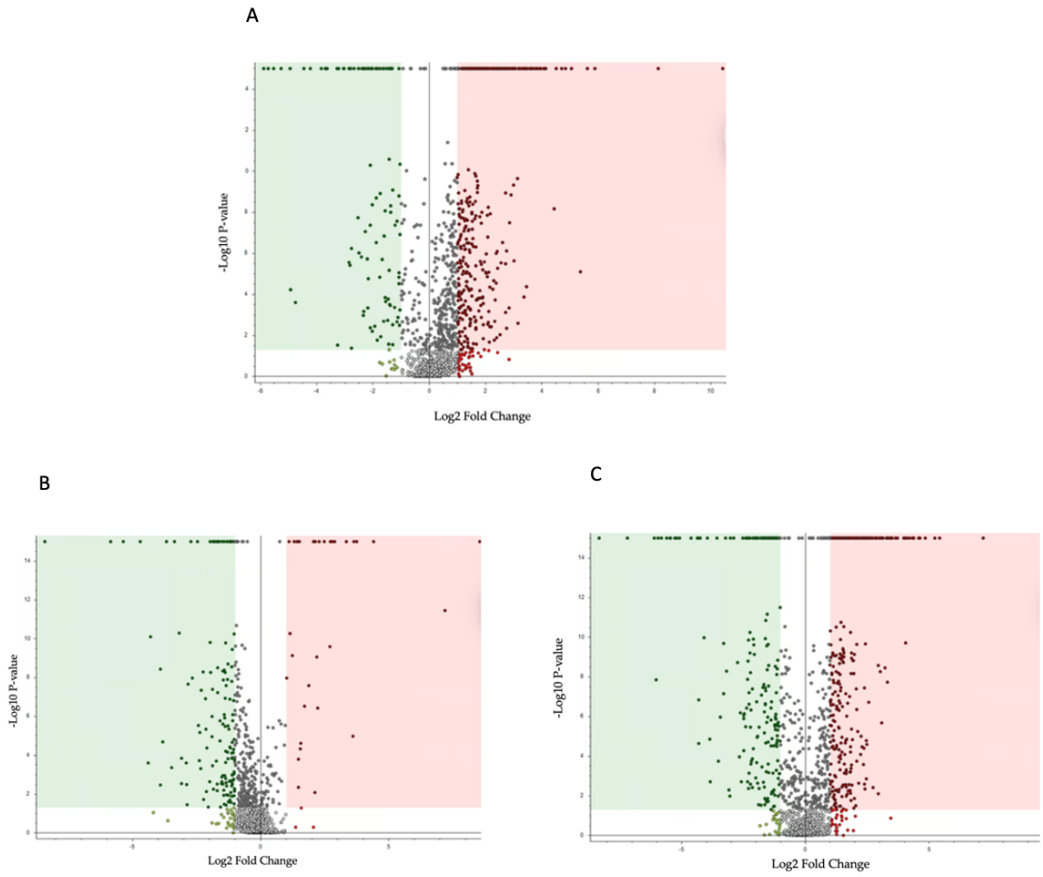


Figure S2. Untargeted differential analysis of extraction solvent showing volcano plot of altered metabolites between **(A)** MeOH vs. MeOH/ H_2O, **(B)** $CHCl_3$/ MeOH vs. MeOH/ H_2O, and **(C)** $CHCl_3$/ MeOH vs MeOH, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

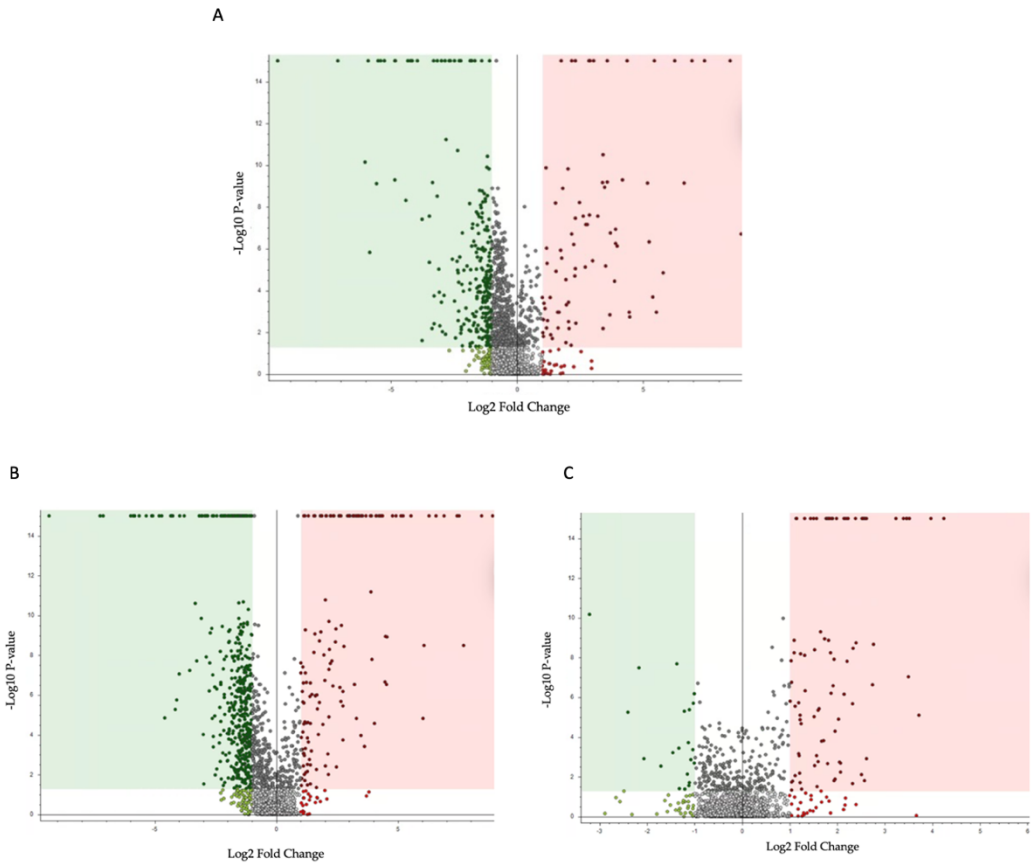


Figure S3. Untargeted differential analysis of extraction solvent showing the volcano plot of altered metabolites between **(A)** sonication vs. bead beating, **(B)** freeze-thaw vs. bead beating and **(C)** freeze-thaw vs. sonication, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite levels were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

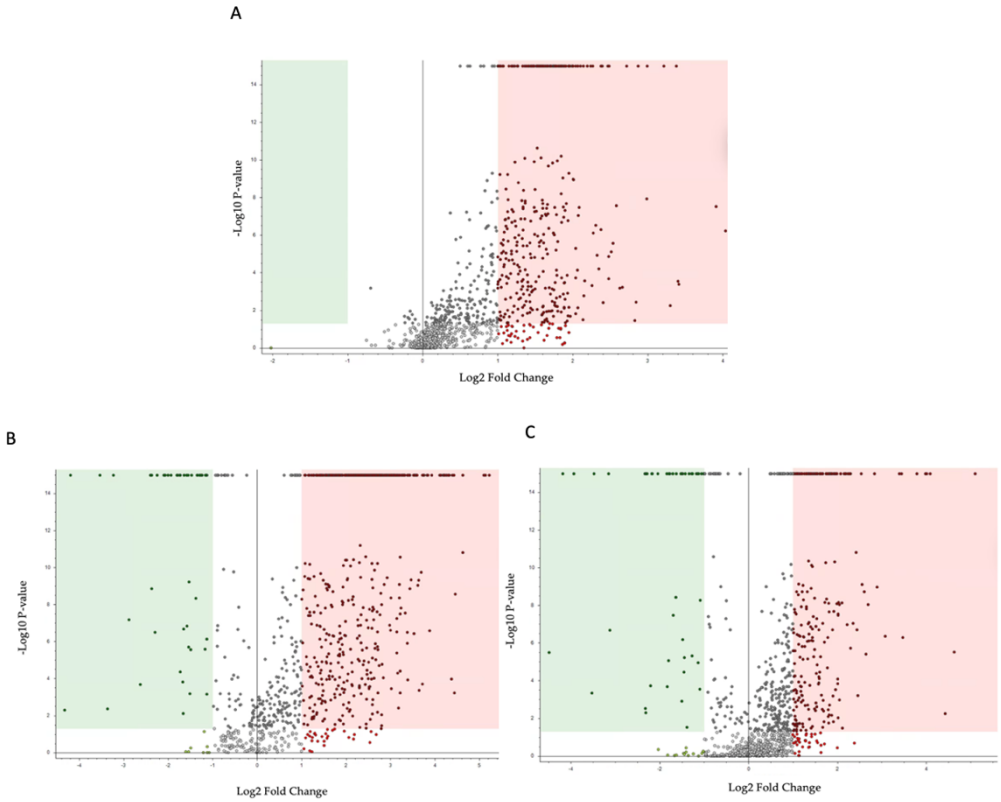


Figure S4. Untargeted differential analysis of sample-solvent ratio showing volcano plot of altered metabolites between **(A)** 1:10 vs. 1:5, **(B)** 1:20 vs. 1:5 and **(C)** 1:20: vs. 1:10, plotted as log2 fold change vs -log10P. Metabolites that are significantly increased are highlighted red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

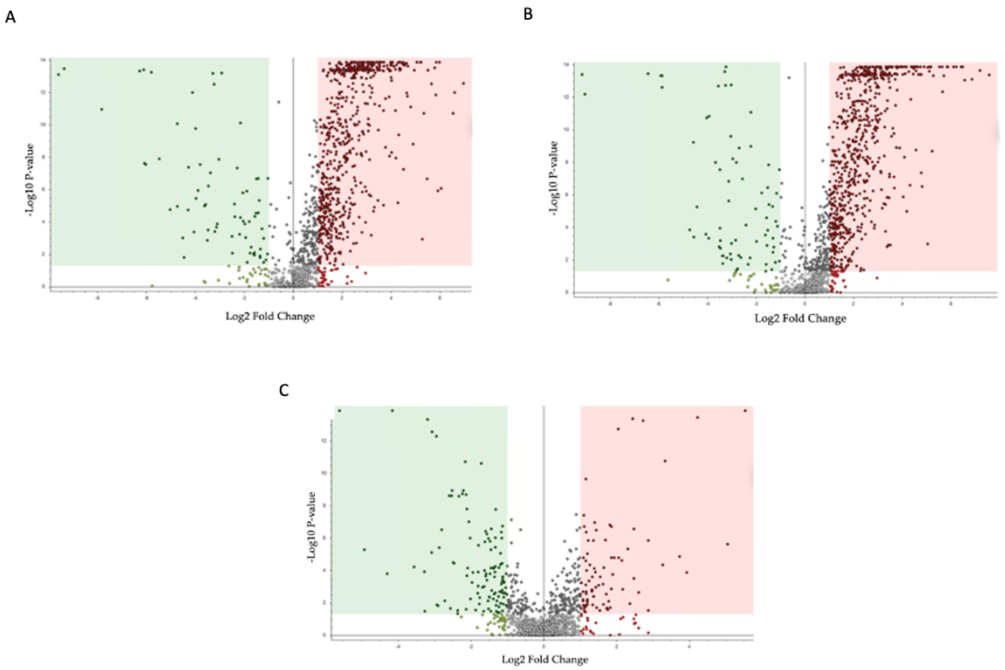


Figure S5. Untargeted differential analysis of cell lysis techniques. Volcano plot of **(A)** HC vs. CD; **(B)** CoD vs. CD **(C)** HC vs. CoD, for all patients. Log2 fold change vs. -log10P. Metabolites. Metabolites that are significantly increased are highlighted in red and those that are significantly decreased are shown in green. Differences in metabolite level were defined by a log2 fold change of 1 and the significance level was set at $p < 0.05$.

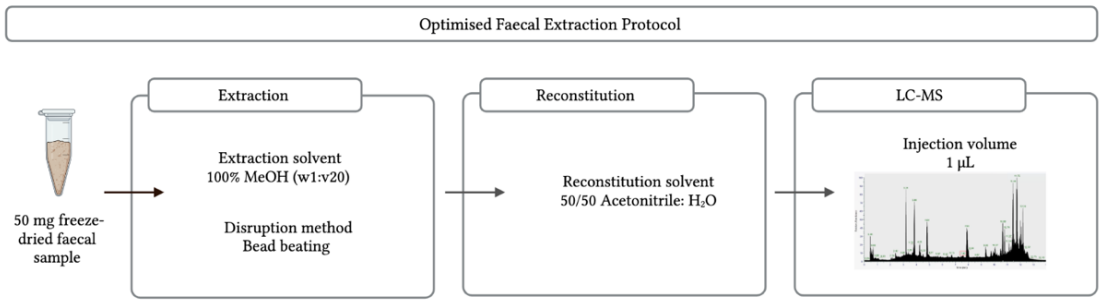


Figure S6. Summary of the developed methodology pipeline. Multi-parameter analysis showed that 50 mg samples give the strongest MS output, and from the extraction solvents analysed, MeOH is the most effective. Additionally, cellular metabolite release is optimal using bead beating as the cell lysis method. Combining optimised parameters provides an experimental protocol for faecal metabolite extraction that can be used for metabolomic analysis.

7.2 APPENDIX 2

SUPPLEMENTARY INFORMATION FOR

Metabolic Dysregulation Driven by Coeliac Disease is Ameliorated by a Gluten Free Diet

Patricia Kelly[1], Gillian Farrell[1], Richard K. Russell[2], Richard Hansen[3], Paraic McGrogan[4], Christine Edwards[5], David Broadhurst[6], Konstantinos Gerasimidis[5], Nicholas J.W. Rattray[1,7].

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G4 0RE, UK.

[2] Royal Hospital for Children and Young People, 50 Little France Crescent, Edinburgh EH16 4TJ, UK.

[3] Royal Hospital for Children, 1345 Govan Road, Glasgow G52 4TF, UK.

[4] Paediatric Gastroenterology Department, The Royal Hospital for Children, Glasgow, UK.

[5] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow G12 8QQ, UK.

[6] Centre for Integrative Metabolomics & Computational Biology, Edith Cowan University, Joondalup, Australia.

[7] NHS Lanarkshire, UK.

Table S1. Quantification of amino acids in coeliac disease. Values given as mean + SEM.

Cross-Sectional Study					Prospective Study		
Amino Acid	HC	Siblings	UCD group	TCD group	At diagnosis	GFD 6 months	GFD 12 months
Arginine	15880824 (2783467)	17862251 (7134413)	12618665 (3334124)	15138808 (3627211)	12618662 (3334124)	14344558 (4790192)	8122521 (1389777)
Aspartic acid	6020308 (1228639)	4753226 (941970)	4073034 (350347)	4606635 (643270)	4073033 (350347)	6781616 (1731701)	5503273 (1132431)
Cysteine	62677 (3399)	50678 (6704)	52651 (5191)	48658 (4457)	52650 (5191)	72165 (11768)	63446 (8108)
Glutamine	12890843 (2014647)	13382870 (2505781)	17629133 (3265021)	19524465 (3555358)	17629132 (3265021)	23579664 (1380777)	31457463 (1008043)
Glutamic acid	65249021 (5413875)	43002359 (6105563)	65055397 (7473774)	52732104 (5755336)	65055397 (7473774)	83828599 (7108565)	74262372 (9327554)
Histidine	5389946 (606537)	3736512 (673736)	3335733 (607412)	5712237 (909728)	33357323 (607412)	3031059 (748452)	4812482 (1034765)
Isoleucine	218747074 (16085824)	181640008 (22272777)	213281425 (22736024)	176809912 (17610276)	213281425 (22736024)	289045232 (43685756)	207824946 (25056848)
Leucine	237864329 (16652380)	207588367 (26119520)	268481534 (28643176)	213674263 (20859591)	268481533 (28643176)	305754777 (44888353)	248942258 (32062411)
Methionine	40725928	13930688	38792363	28148542	38792363	60845889	48326590

	(5928567)	(3293281)	(7098365)	(5498065)	(7098365)	(1302908)	(9035288)
Phenylalanine	123360137 (2741135)	111269182 (3842277)	114555219 (5335509)	107096546 (4877491)	114555219 (5335509)	128529134 (6610297)	118851281 (4594655)
Proline	70793444 (4722047)	53973367 (8754071)	75503546 (7689958)	55970888 (5867712)	75503547 (7689958)	89600396 (7690204)	81434395 (9433919)
Serine	15221234 (1539879)	10725674 (1695949)	12058714 (1402536)	11722395 (1468232)	12058714 (1402536)	19284265 (3462843)	16775588 (3284471)
Threonine	18745909 (1095634)	13209649 (1361778)	15397570 (1552669)	13671551 (1347115)	15397570 (1552669)	21216819 (3286542)	18373006 (2530913)
Tryptophan	55332794 (4098539)	38245241 (4511184)	52320598 (5509960)	43923439 (4915765)	52320598 (5509960)	56084436 (9626850)	49229377 (7121586)
Tyrosine	72999839 (2596951)	63460507 (4544420)	67343207 (3915537)	61188332 (3898725)	67343207 (3915537)	74703583 (4850878)	70546699 (3524076)
Valine	198951468 (7415409)	168702666 (10479670)	186973563 (12538238)	161505064 (10704812)	186973563 (12538238)	223150679 (16974042)	191712923 (13431142)

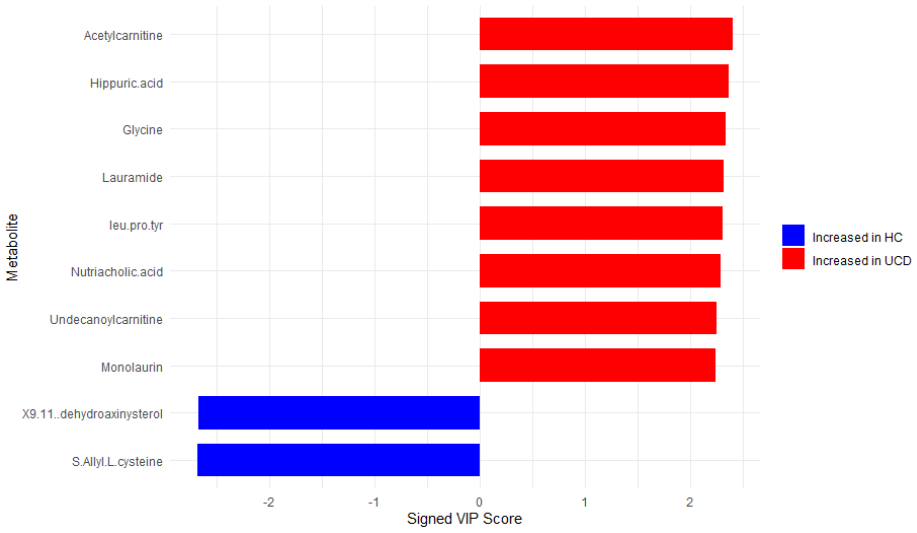
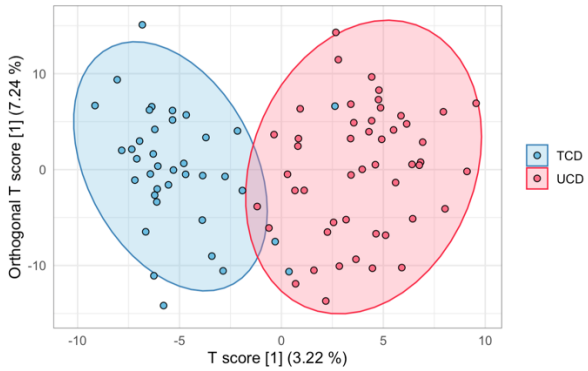


Figure S1. Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing UCD vs. HC.

A



B

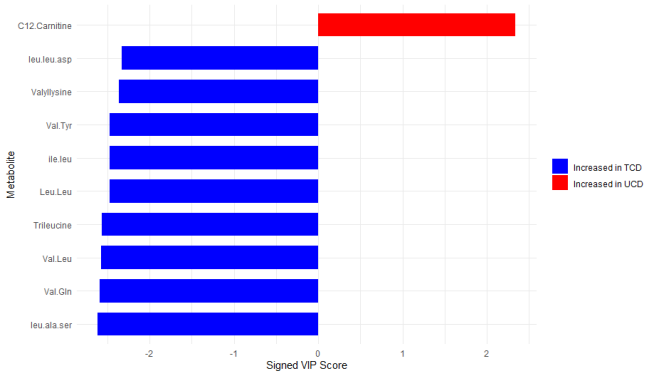


Figure S2. Orthogonal partial least square discriminant analysis (OPLS-DA) and variable importance for treated vs untreated coeliac disease (TCD vs UCD). (A) OPLS-DA scores plot comparing TCD vs UCD with points representing individual samples and ellipses showing 95% confidence intervals, R2Y = 0.767, Q2 = 0.0963. (B) Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing TCD vs. UCD.

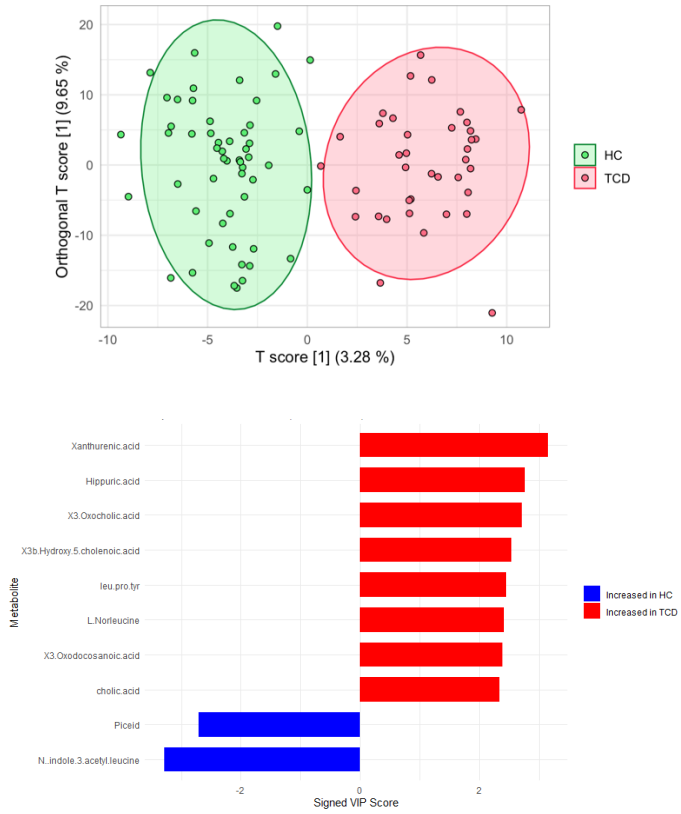


Figure S3. Orthogonal partial least square discriminant analysis (OPLS-DA) and variable importance for treated coeliac disease (TCD) vs HCs. (A) OPLS-DA scores plot comparing TCD vs HC with points representing individual samples and ellipses showing 95% confidence intervals, R2Y = 0.843, Q2 = 0.548. (B) Variable importance in projection (VIP) plot showing the top differential metabolites from the orthogonal partial least square discriminant analysis (OPLS-DA) model comparing TCD vs. HC.

Table S2. Top differentially abundant metabolites identified in the prospective cohort, with corresponding VIP scores in the cross-sectional cohort.

Variable	Direction	VIP Score
Monolinolenin	Increased in TCD	1.055896
N-Acetyl-L-glutamic acid	Low model contribution	VIP < 1
N-[(2S)-2-Hydroxypropanoyl]-L-phenylalanine	Increased in TCD	1.465013
Xanthurenic acid	Increased in TCD	1.956463
Methyl-alpha-aspartyl phenylalaninate	Low model contribution	VIP < 1
3-Amino-4,7-dihydroxy-8-methylcoumarin	Low model contribution	VIP < 1
L-Alanyl-L-proline	Low model contribution	VIP < 1
Phenylalanine	Increased in TCD	1.009252
O-alpha-D-mannosyl-L-threonine	Low model contribution	VIP < 1
Galactosylhydroxylysine	Increased in UCD	2.110364
N-indole-3-acetyl-leucine	Increased in UCD	2.199974
Pyridoxine	Increased in UCD	1.177892
Butyryl-L-homoserine-lactone	Increased in UCD	1.268518
9(11)-dehydroaxinysterol	Increased in UCD	2.040323
N-alpha-methyl-L-lysine	Low model contribution	VIP < 1
Bz-Arg-OEt	Increased in UCD	1.140965

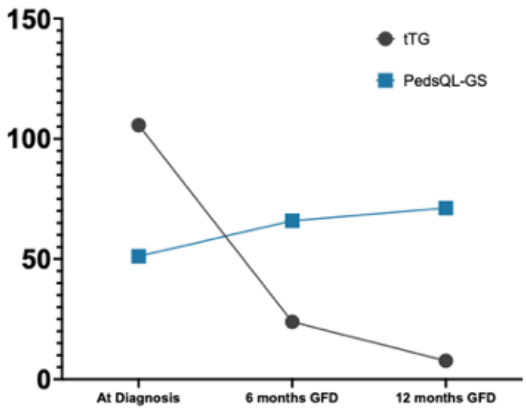


Figure S4. tTG and PedsQL-GS levels in coeliac disease patients throughout 6 and 12 months on a GFD.

7.3 APPENDIX 3

SUPPLEMENTARY INFORMATION FOR

Optimisation of a Dilute-and-Shoot UHPLC-MS Method for High-Throughput Urinary Metabolomics

Patricia E. Kelly [1], Gillian Farrell [1], Konstantinos Gkikas[2], Caroline Kerbiriou[2], Bernadette White[2], Konstantinos Gerasimidis[2], Nicholas J. W. Rattray[1].

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, Glasgow G4 0RE, UK.

[2] School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow G12 8QQ, UK.

Table S1. Compound Discoverer metabolite filter settings.

Compound Discoverer Filter	Setting
Name	Is not blank
Annotated Δ Mass [ppm]	Is between -5.00 and 5.00
Annotation Source	Has status full match in source Predicted Compositions
Annotation Source	Has status full match in source ChemSpider Search
MS2	Is not equal to No MS2

Table S2. Compound Discoverer workflow settings.

Workflow Node	Workflow Parameter	Workflow Information
Input files		.raw data
Select Spectra	Spectrum Properties Filter	Lower RT limit: 0 Upper RT limit: 0
	Scan Event Filters	Polarity Mode: Is +
Align Retention Times	General Settings	Alignment Model: Adaptive curve Maximum Shift [min]: 0.3 Mass Tolerance: 2 ppm
Detect Compounds	General Settings	Mass Tolerance: 2 ppm Intensity Tolerance [%]: 30 S/N Threshold: 5 Min. Peak Intensity: 2 000 000 Ions: [2M + FA + H]-1; [2M + H]+1; [2M + K]+1; [2M + Na]+1; [2M - H]-1; [M + 2H]+2; [M + Cl]-1; [M + FA - H]-1; [M + H]+1; [M + H + K]+2; 2M + H + MeOH]+1; [M + H + Na]+2; [M + H – H_2O]+1; [M + K]+1; [M +

		Na]+1; [M - 2H]-2; [M + 2H + K]-1; [M -H]-1; [M -H_2O]-1 Min. Element Counts: C H Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O15 P5 S5
Group Compounds	Compound Consolidation	Mass Tolerance: 2 ppm RT Tolerance [min]: 0.2
	Fragment Data Selection	Preferred Ions: [M - H] +1; [M-H] -1
Search mzCloud	General Settings	Compound Classes: Endogenous Metabolites, Excipients/ Additives/ Colourants, Extractables/ Leachables, Natural Products/ Medicines, Natural Toxins, Personal Care Products/ Cosmetics, Small Molecule Chemicals, Steroids/ Vitamins/ Hormones, Therapeutic/ Prescription Drugs Precursor Mass Tolerance: 10 ppm FT Fragment Mass Tolerance: 10 ppm Library: Autoprocessed, Reference Post. Processing: Recalibrated Annotation Matching. Fragments: True
	DDA Search	Identity Search: Cosine Match Activation Type: True

		Match Activation Energy: Match with Tolerance Activation Energy: 20 Apply Intensity Threshold: True Similarity Search: None Match Factor Threshold: 60
	DIA Search	Use DIA Scans: False Max. Isolation Width [Da]: 500 Match Activation Type: False Match Activation Energy: Any Activation Energy Tolerance: 100 Apply Intensity Threshold: False Match Factor Threshold: 20
Predict Compositions	Prediction Settings	Mass Tolerance: 2 ppm Min. Element Counts: C. H Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O15 P5 S5 Min. RDBE: 0 Max. RDBE: 40 Min. H/C: 0.1 Max. H/C: 4 Max.# Candidates: 10 Max.# Internal Candidates: 200
	Pattern Matching	Intensity Tolerance [%]: 30 Intensity Threshold [%]: 0.1 S/N Threshold: 3 Min. Spectral Fit [%]: 30 Min. Pattern Cov. [%]: 90

		Use Dynamic Recalibration: True
	Fragments Matching	Use Fragments: True Mass Tolerance: 2 ppm S/N Threshold: 5
Map to Metabolika Pathways	Search Settings	Metabolika pathways: All Search Mode: By Formula or Mass
	By Mass Search Settings	Mass Tolerance: 2 ppm
	By Formula Search Settings	Max. # of Predicted Compositions to be searched per Compound: 3
	Display Settings	Max. # of Pathways in 'Pathways' column: 20
Apply mzLogic	General Settings	FT Fragment Mass Tolerance: 10 ppm IT Fragment Mass Tolerance: 0.4 Da Max. # Compounds: 0 Max. # mzCloud Similarity Results to consider per Compound: 10 Match Factor Threshold: 30
Assign Compound Annotations	General Settings	Mass Tolerance: 2 ppm
	Data Sources	Data source #1: mzCloud Search Data source #2: Predicted Compositions Data source #3: MassList Search Data source #4: ChemSpider Search

		Data source #5: Metabolika Search
	Sorting Rules	Use mzLogic: True Use Spectral Distance: True SFit Threshold: 20 SFit Range: 20
Fill Gaps	General Settings	Mass Tolerance: 2 ppm S/N Threshold: 5 Use Real Peak Detection: True
Apply QC Correction	General Settings	Min. QC Coverage [%]: 30 Max. QC Area. RSD [%]: 30 Max. Corrected QC Area RSD [%]: 25 Max. # Files Between QC Files: 15
Mark Background Compounds	General Settings	Max. Sample/ Blank: 5 Max. Blank/ Sample: 0 Hide Background: True
Differential Analysis	General Settings	Log10 Transform Values: True

Table S3. Peak Quality Factor Descriptions.

Peak Quality Factor (PQF)	Description
Zig-zag	The zig-zag index measures the vertical peak smoothness, which is quantified by calculating the number and magnitude of vertical fluctuations, i.e., zig-zags in the signal intensity profile.
FWHM2Base	The FWHM2Base is defined as the ratio of the FWHM to the full width at baseline, thereby providing a metric for quantifying peak symmetry, resolution, and tailing.
Jaggedness	A quantification of the smoothness of a peak as measured by the irregularity or variability of the signal intensity.
Modality	The modality index indicates the number of peak apexs and infers the peak resolution and elution properties.

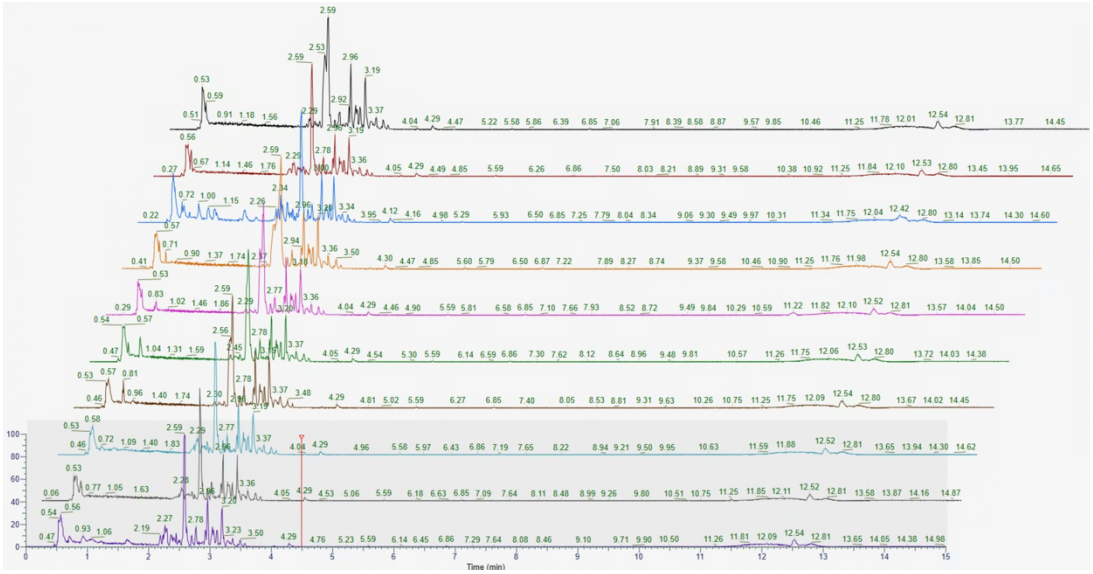


Figure S1. Comparison of chromatograms obtained from untargeted LC-MS analysis using different solvent systems for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – Acetonitrile (1.90E9), red - Acetonitrile/ H_2O (2.5E9), blue – H_2O (2.3E9), orange – IPA (1.81E9), pink – IPA/ ACN (2.22E9), green – IPA/ H_2O (2.07E9), brown – IPA/MeOH (2.05E9), light blue – MeOH (2.39E9), grey – MeOH/ ACN (2.44E9), purple – MeOH/H_2O (2.22E9).

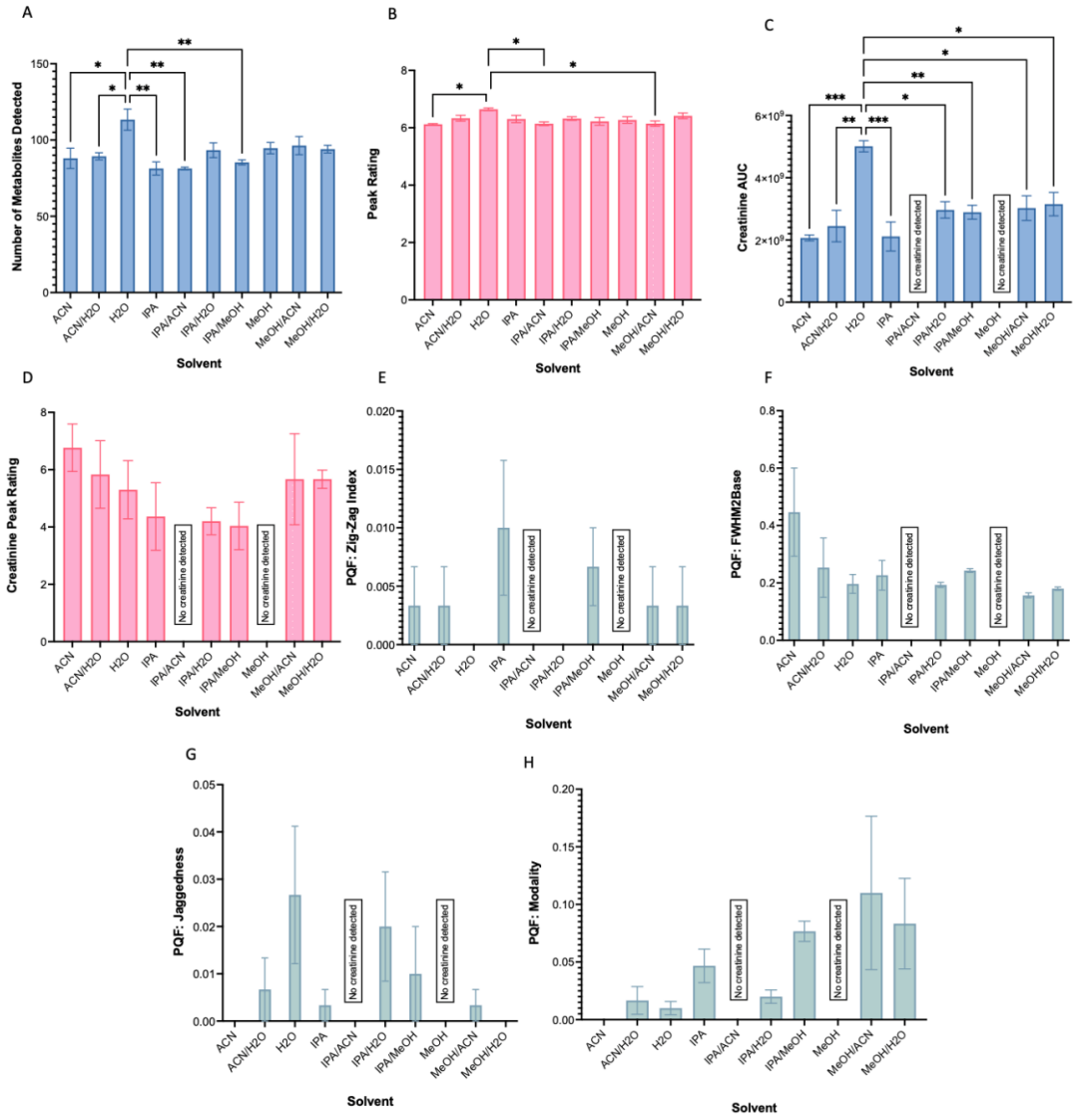


Figure S2. The effect of the extraction solvent on untargeted urinary metabolomics. Outcomes were assessed by **(A)** the number of metabolites detected, **(B)** their average peak rating, **(C)** the area under the curve (AUC) of the detected creatinine peak and **(D)** the associated peak rating. The peak performance creatinine of creatinine further evaluated, as measured by the quantification of peak quality factors (PQF)s, **(E)** Zigzag, **(F)** FWHM2Base, **(G)** Jaggedness, and **(H)** Modality indices. *Creatinine was not detected when IPA/H_2O or MeOH were used as the extraction solvents.

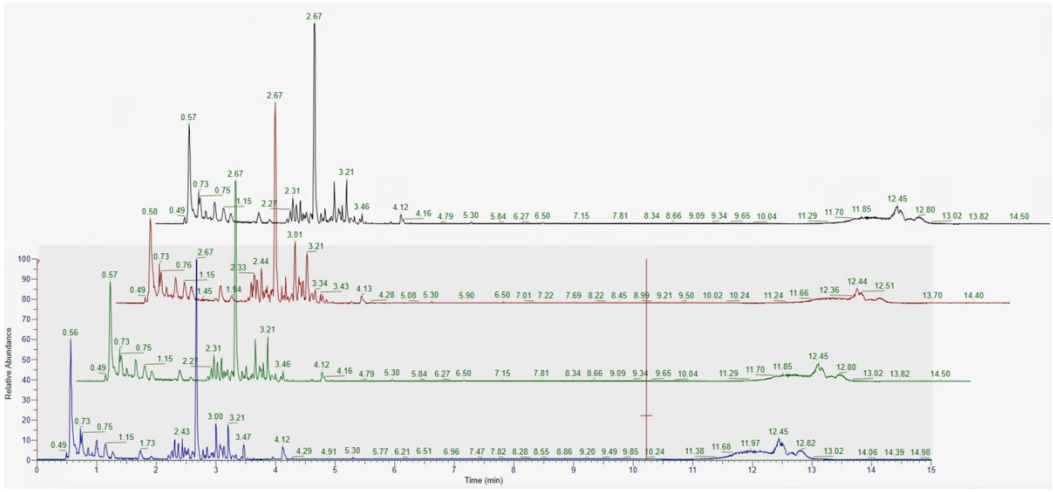


Figure S3. Comparison of chromatograms obtained from untargeted LC-MS analysis using different extraction solvent dilution factors for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – dilution factor 1 (1.7E9), red – dilution factor 2 (2.21E9), green – dilution facctor 5 (1.7E9), blue – dilution factor 10 (1.34E9).

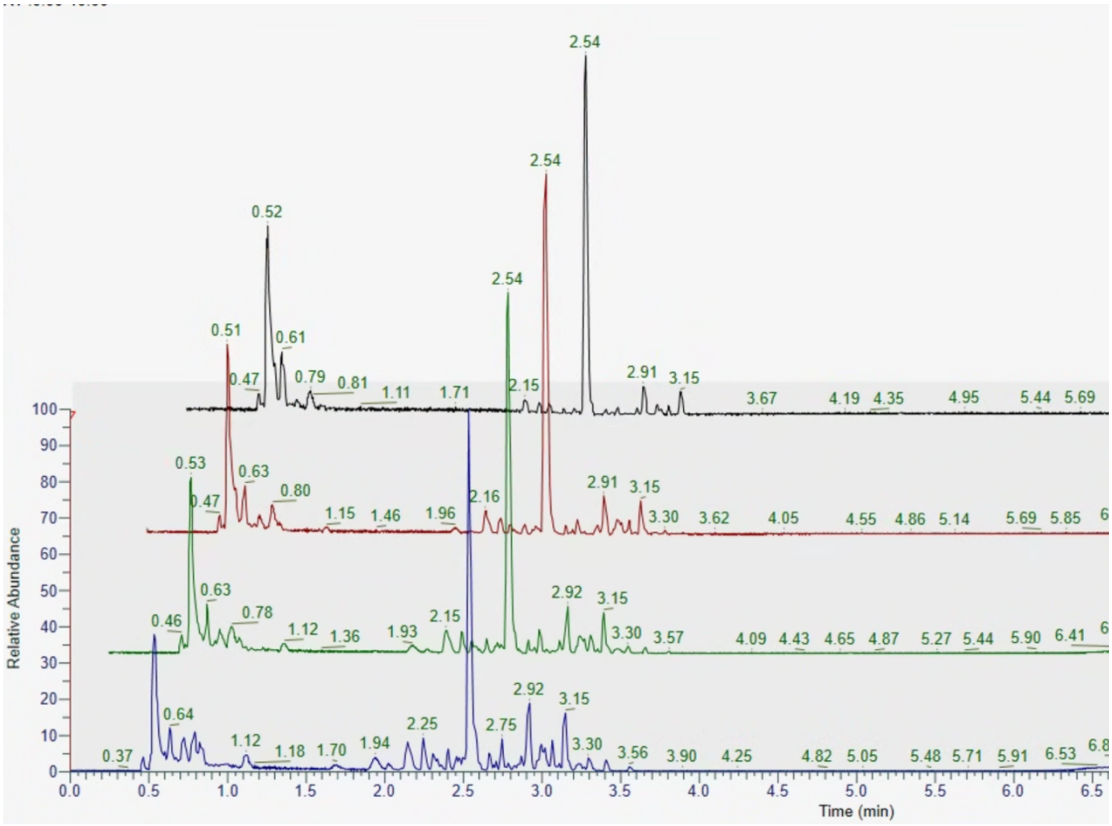


Figure S4. Comparison of chromatograms obtained from untargeted LC-MS analysis using different injection volumes for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – 0.5 (5.09E8), red – 1 (8.3E8), green – 2 (1.31E9), blue – 5 (2.07E9).

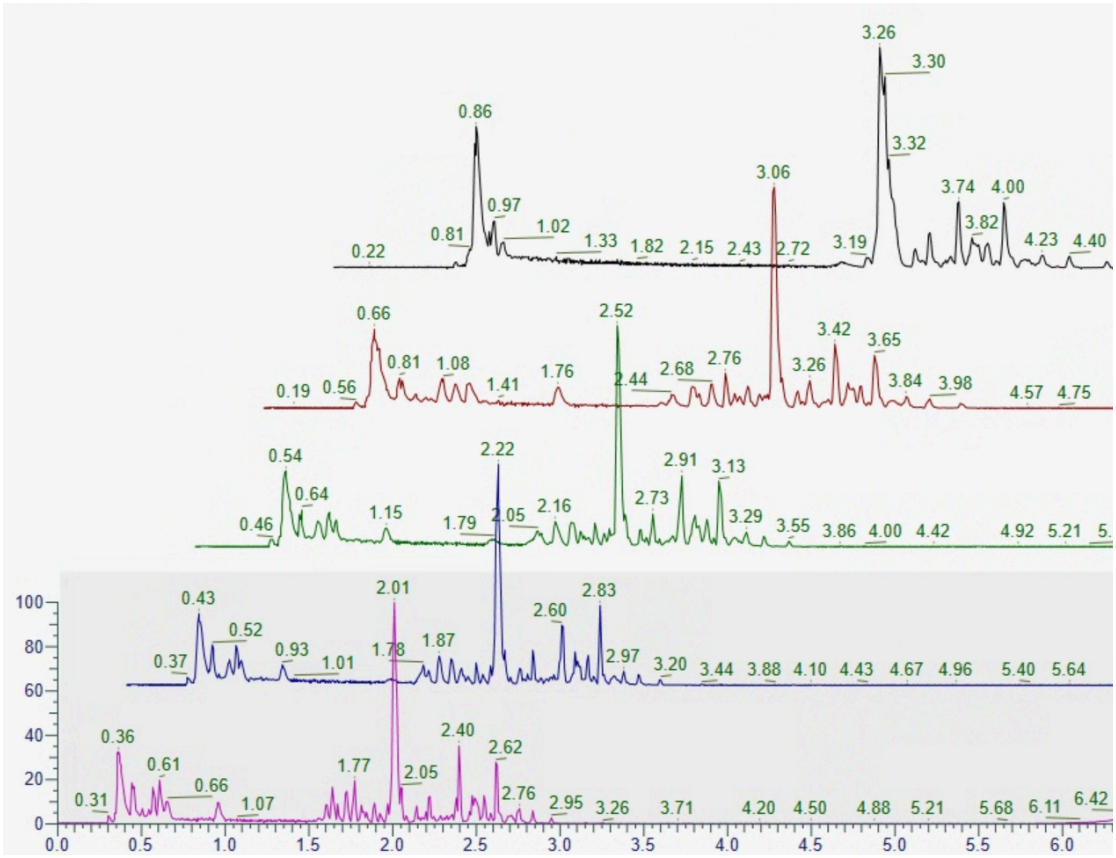


Figure S5. Comparison of chromatograms obtained from untargeted LC-MS analysis using different flow rates for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – 0.25 mL/min (1.86E9), red – 0.3 mL/min (2.42E9), green – 0.4 mL/min (2.31E9), blue – 0.5 mL/min (2.35E9), pink – 0.6 mL/min (2.30E9).

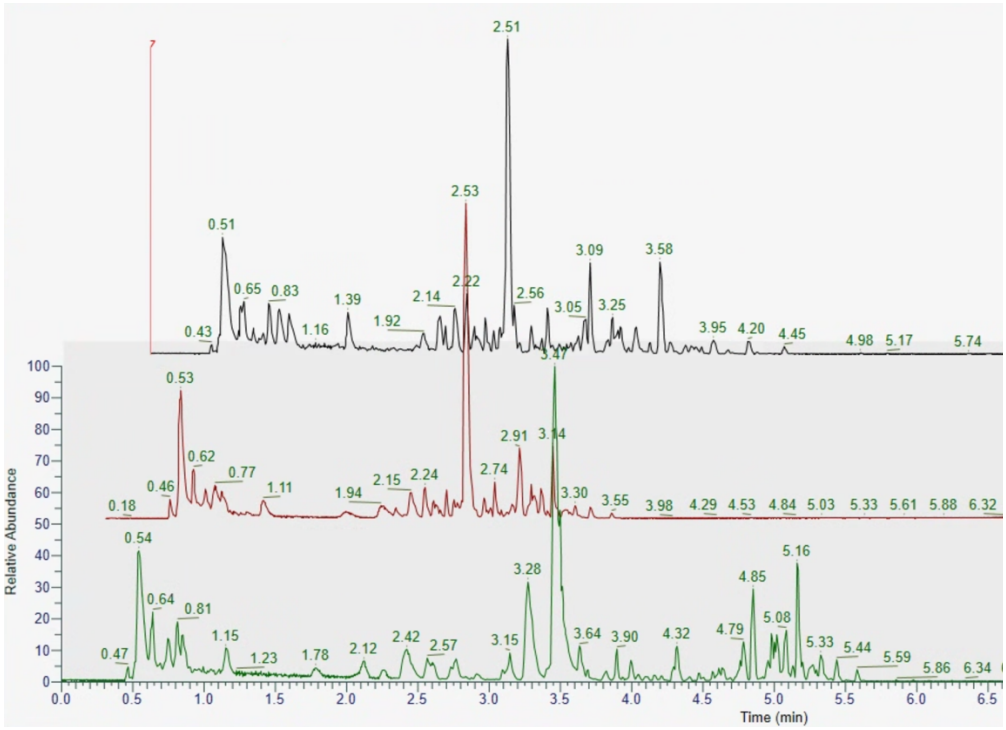


Figure S6. Comparison of chromatograms obtained from untargeted LC-MS analysis using different gradient curve parameter values for method optimisation. Each colour represents a solvent used for extraction with signal intensities shown. The colours are represented by the following solvents: black – gradient curve of 3 (2.17E9), red – 5 (1.96E9), green – 7 (1.84E9).

Table S4. Untargeted metabolomics experiment elution gradient. Mobile phase A, 99.9% water + 0.1% formic acid, mobile phase B, 99.9% ACN + 0.1% formic acid.

Time (min)	Mobile Phase A (%)	Mobile Phase B (%)	Flow rate (mL/min)	Gradient curve
0.0	99.0	1.0	0.4	7
0.5	99.0	1.0	0.4	7
4.0	50.0	50.0	0.4	7
6.0	5.0	55.0	0.4	7
6.01	1.0	99.0	0.4	7
8.0	1.0	99.0	0.4	7
9.5	99.0	1.0	0.4	7
10.0	99.0	1.0	0.4	7

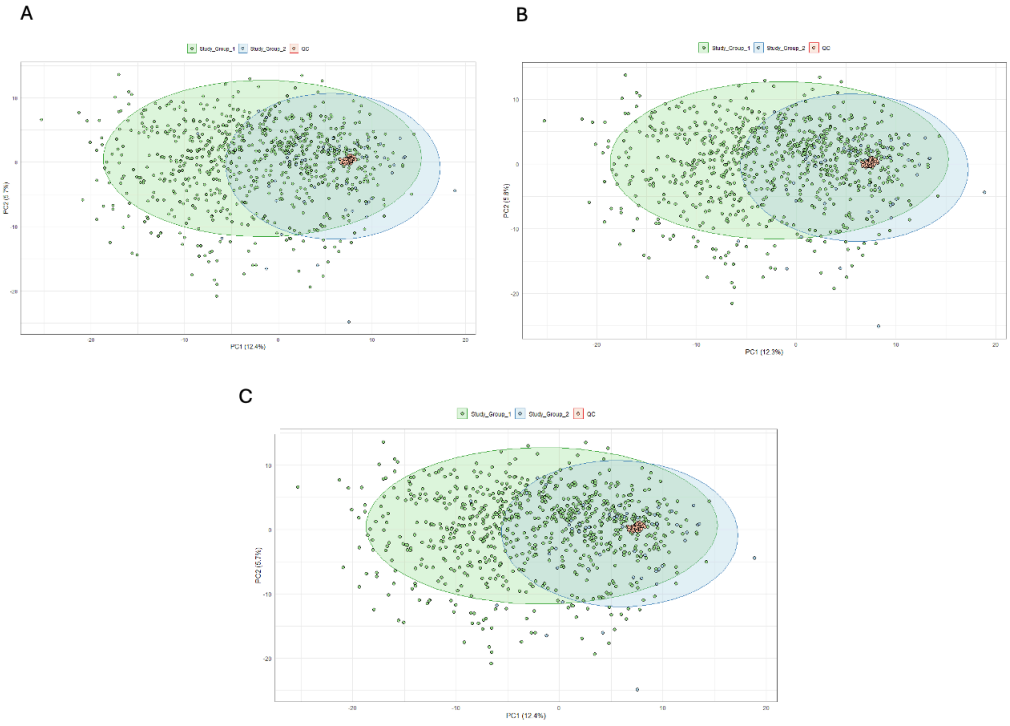


Figure S7. Principal component analysis (PCA) plots comparing three normalisation strategies, evaluating two study groups. (A) Non-normalised, (B) Creatinine normalised, and (C) PQN normalised urine data.

Table S5. Coefficient of Variation (CV) values for metabolites detected in three normalisation methods. Metabolites were detected at MSI level 2.

Metabolite	Coefficient of Variation (CV)		
	Non-Normalised	Creatinine Normalised	PQN Normalised
Acadesine	0.331688	0.560431	0.370701
Acetaminophen glucuronide	0.164351	0.223531	0.104566
Acetanilide	0.248759	1.711	2.520153
Acetoacetic acid	0.155875	0.803121	2.009837
Acetyl-L-carnitine	0.342468	0.300348	0.298355
Acetylagmatine	0.391684	0.587913	0.416512
Acronycidine	0.2768	0.26706	0.250245
Adenosine	1.25527	1.349067	1.266701
Aderbasib	0.424737	0.510881	0.40471
Ala-pro	0.31781	0.295325	0.268673
alangiside	0.256917	0.296218	0.246416
Alanyltyrosine	0.340473	0.280605	0.292433
alloxydim	0.327958	0.367294	0.301518
Allyl cyclohexanevalerate	0.211445	0.293483	0.278732
Allyl undecylenate	0.185418	0.153912	0.110492
alpha-CEHC	0.150504	0.52575	1.339858
Aminoadipic acid	0.497858	0.423307	0.491964
Amoxicillin	0.150426	0.285977	0.13401
Amylbenzene	0.190241	0.198508	0.124037
Anhydrovitamin A	0.192672	0.196269	0.19329
Ankorine	0.318115	0.360032	0.305704
anticapsin	0.220638	0.140762	0.136709
Apiin	0.184699	0.182665	0.103117
Aprobarbital	0.714812	2.368973	3.805627
Arginyltyrosine	0.992742	1.043739	1.058323
Artesunate	0.154425	0.206209	0.088057
Ascorbic acid	0.676673	0.651875	0.631369
asp-gln	0.420086	0.35792	0.367315
asp-leu	0.206243	0.387374	0.227824
Asp-tyr	0.380151	0.315801	0.331717
Aspartame	0.147943	0.199428	0.107807
Aspartyl-L-proline	0.253005	0.336906	0.230723
Aspartylphenylalanine	0.144494	0.198865	0.102862
Asymmetric dimethylarginine	1.375601	1.462349	1.701033
Azathioprine	0.141684	0.271926	0.120996
Bakankosin	0.183495	0.220842	0.120613
Benzaldehyde	0.159799	0.17218	0.329652
Benzenepentol	0.223226	0.23625	0.190758
Benzyl 6-O-D-xylopyranosyl-D-glucopyranoside	0.186646	0.232944	0.128437
Bicyclomycin	0.47775	0.420024	0.423532
Bilirubin	0.270926	0.339281	0.384622

Binapacryl	0.393696	0.456469	0.37949
Biopterin	0.495612	0.653828	0.512979
Biotin	0.209777	0.151497	0.14408
Biotin sulfone	0.559814	0.57346	0.542461
bispyribac	0.870502	0.922125	0.868396
Boschnaloside	0.149748	0.20617	0.090828
Buchananine	0.53707	0.431021	0.477016
Bufadienolide	0.280447	0.345598	0.270061
Buflomedil	0.338015	0.379196	0.326198
Butenylcarnitine	0.283405	0.194321	0.225307
butilfenin	0.157854	0.19916	0.072822
butoctamide	0.609991	0.689398	0.642857
Butyl 3-(2-butoxy-2-oxoethyl)-3-hydroxy-4-oxo-2-oxetanecarboxylate	0.191025	0.223066	0.16371
C12-Carnitine	0.610126	0.603067	0.578048
Capric acid	0.286486	0.310978	0.38058
Caprolactam	0.18946	1.590292	2.806307
Capryloylglycine	0.162916	0.816068	1.555328
carbapenem MM22383	0.166343	0.189836	0.112196
carbazeran	0.150181	0.202463	0.083905
Carbidopa	0.154507	0.274143	0.126903
Carbinoxamine	0.481381	0.458593	0.674575
Carboxy-ibuprofen	0.766323	0.882086	0.778611
Carboxytolbutamide	0.189998	0.322531	0.184449
Carminomycin I	0.288822	0.490485	0.840031
Carvone	0.275154	0.34754	0.27638
Cassaidine	0.170115	0.171355	0.164251
Cervonoyl ethanolamide	0.149239	0.208235	0.175742
Cetirizine	0.175663	0.184145	0.12115
Cetraxate	0.259837	0.248949	0.191821
Chlorpheniramine	0.435135	0.337113	0.433725
Ciclopirox	0.480207	1.704539	3.007247
Cinnamoylglycine	0.149903	0.195858	0.118189
cis-trihomoaconitic acid	0.204978	0.304273	0.159975
Citric acid	0.468642	0.703948	0.509385
Coronatine	0.29638	0.340257	0.424193
Corticosterone	0.429358	0.432366	0.409621
Cortisol	0.178864	0.236692	0.131256
Creatine	0.203362	0.104461	0.142706
Creatinine	0.205989	0.205989	0.408276
crinamidine	0.350479	0.360186	0.3418
Crotonic acid	1.772563	1.740595	1.810214
Cyclamic acid	0.366809	0.291482	0.317842
cyclic Melatonin	0.181304	0.178561	0.127048
Cyclo(leucylleucine)	0.339559	0.437208	0.317552
Cyclohexylamine	0.335535	1.149178	2.412831

cypendazole	0.47187	0.40928	0.470699
Cyprodenate	0.158089	0.202698	0.100317
cys-his	0.210014	0.224411	0.162586
Cytarabine	0.256925	0.348519	0.228832
D-(+)-Pyroglutamic Acid	0.235877	0.516011	1.867827
D-(+)-Tryptophan	0.155632	0.301537	0.159441
D-Lysyl-L-valyl-D-valyl-L-allothreonine	0.201977	1.356479	2.08465
Decanohydroxamic acid	0.14702	0.212356	0.168328
Decanoylcarnitine	0.305246	0.327253	0.280795
delcorine	0.309998	0.344542	0.361682
Demethylalangiside	0.217644	0.231229	0.149617
deoxyhypusine	0.321101	0.230469	0.267052
Desloratadine	0.169599	1.285405	2.294232
Dhurrin	0.843351	0.777095	0.847057
Diaminopimelic acid	0.508944	0.425101	0.466074
diethyl oxalpropionate	0.159229	0.219939	0.077385
Diethylcarbamazine N-oxide	0.249947	0.221475	0.218731
dihomomethionine	0.201114	0.20775	0.151185
Dihydrocaffeic acid 3-O-glucuronide	0.261502	0.407591	0.605558
Dihydroconiferin	0.165359	0.219242	0.097977
Dihydrothymine	0.204331	0.066222	0.249492
Dihyroxy-1H-indole glucuronide I	0.388787	0.392393	0.344731
Dinoseb	0.353421	0.404443	0.316515
Dipivefrin	0.210077	0.26613	0.166644
DL-Carnitine	0.244206	0.102423	0.184625
DL-Aminocaprylic acid	0.167564	0.191454	0.11783
Docosahexaenoic acid ethyl ester	0.211067	0.23706	0.176818
Dodecanedioic acid	0.148114	0.181343	0.065507
Dopamine	2.489068	2.848741	3.345976
Dynone	0.490969	0.415169	0.475163
Ecgonine	0.154566	0.213666	0.353273
Ecgonine methyl ester	0.175907	0.362642	0.799798
Ectoine	0.318361	0.213948	0.261204
Elacytarabine	0.489785	0.473824	0.588573
Epinephrine	0.147356	0.177008	0.06129
Epinephrine glucuronide	1.886223	2.002812	1.924165
Epinephrine sulfate	0.413029	0.439102	0.394113
Epithienamycin E	0.849786	0.744306	0.804779
Ergonovine	0.213346	0.179154	0.15519
Esmolol	0.246455	0.217002	0.193084
Eterobarb	0.161949	0.203795	0.09059
Ethychlozate	0.166104	0.214088	0.252722
Ethyl {2-[(2E)-5-(hexopyranosyloxy)-2-penten-1-yl]-3-oxocyclopentyl}acetate	0.276538	0.343872	0.489263
Ethyl butylacetylaminopropionate	0.157822	0.222262	0.248788
Ethyl methylphenylglycidate	0.419673	0.538436	0.422145

Ethylbenzene	0.160197	0.204977	0.09234
Ethylenediamine-N,N'-diacetic acid	0.297495	0.294168	0.259281
Fenethylline	0.156587	0.192454	0.082502
Fenitropan	0.158359	0.193679	0.072068
Ferulic acid	0.184779	0.218278	0.488299
Fesoterodine	0.187629	0.163314	0.112165
Fexofenadine	0.159324	0.197178	0.147096
fosfocreatinine	0.244949	0.174223	0.207969
Fraxin	0.18396	0.282202	0.406463
Furmecyclox	0.285387	0.329559	0.266969
Fusarochromanone	0.203979	0.179944	0.108891
Gabexate	0.193036	0.182213	0.174989
Galantamine	0.241581	0.162279	0.229246
Gentisuric acid	0.483957	0.41269	0.439795
gibberellin A(3) O-D-glucoside	0.264389	0.391938	0.287664
glaucarubinone	0.279087	0.270198	0.254165
Gln-Gln	0.537328	0.492761	0.510629
gln-phe	0.276119	0.264919	0.443854
Glucosylgalactosyl hydroxylysine	0.337886	0.247756	0.332777
Glutarylcarnitine	0.345741	0.507643	0.386649
Gly-DL-Phe	0.255917	0.444448	0.278562
Glycinexylidide	0.20543	0.269173	0.16598
Glycodiazine	0.183288	0.278163	0.152087
Glycyl-L-asparaginyl-D-leucine	0.182829	0.193662	0.117845
Glycylglutamine	0.433104	0.345923	0.388222
Glycylproline	0.296133	0.237142	0.243368
Guanadrel	0.238543	0.309154	0.194018
Guvacoline	0.177359	0.364491	2.119097
Gynocardin	0.460024	0.660554	0.483235
Hawkinsin	0.627212	0.536181	0.602509
HC Blue 1	0.162341	0.246389	0.127131
HC blue 2	0.487953	0.441689	0.45446
Heliotron	0.729528	0.688243	0.7117
Heptylbenzene	0.280753	0.23201	0.266212
hercynylcysteine sulfoxide	0.296864	0.307135	0.276555
Hesperetin 7-O-glucuronide	0.290197	0.248552	0.228185
Hexadecanedioic acid mono-L-carnitine ester	0.739949	0.643359	0.677168
Hexadienic acid	0.275793	1.459118	2.639353
hexahomomethionine	0.287371	0.31689	0.243606
Hexanoylcarnitine	0.298564	0.34525	0.276389
Hippuric acid	0.142487	0.204384	0.077657
His-pro	0.373188	0.419367	0.449158
Histamine	0.289122	0.191359	0.229764
Histidinol	0.247367	0.384297	0.675695
Homo-L-arginine	0.1958	0.136096	0.173443
Hostmaniane	0.161938	0.446571	0.636659

Hydroxyphenylacetylglycine	0.28416	0.314002	0.27093
Hymexazol N-glucoside	0.532662	0.595535	0.516619
Hymexazol O-glucoside	0.380657	0.323573	0.338524
Hypusine	0.268522	0.182997	0.198424
Imidazolelactic acid	0.218811	0.091507	0.296764
Indicine	0.472134	0.432033	0.535858
indol-2-one	0.167712	0.300046	0.140461
Indole-3-acetic-acid-O-glucuronide	0.147009	0.205831	0.104993
Indole-3-lactic acid	0.154059	0.20254	0.111446
Indoleacetyl glutamic acid	0.163304	0.196731	0.100034
Indoleacetyl glutamine	0.14718	0.200903	0.105389
Indoleacetylaspartate	0.165199	0.177339	0.056579
Indoxyl-D-glucuronide	0.171099	0.275101	0.131407
Ingenol mebutate	0.228816	0.319057	0.22163
Integerrimine	0.159928	0.184717	0.107912
Isoamylamine	0.45038	1.967554	3.277011
Isocromadurine	0.18373	0.175259	0.104885
Isoetharine	0.194218	0.266554	0.155272
Isofraxidin	0.229667	0.290378	0.445235
Isohomovanillic acid	0.189483	0.205123	0.132522
Isopropyl D-galactopyranoside	0.281203	0.622895	1.212605
Istamycin C1	0.159864	0.590179	1.041499
Jasmonic acid	0.826192	2.410488	3.711872
L-(-)-Methionine	0.193416	0.253963	0.229807
L-(+)-Citrulline	0.290849	0.193134	0.237147
L-dihydroanticapsin	0.76921	0.714401	0.738968
L-Dopa	0.635251	0.54845	0.597785
L-Fucose	0.394307	0.32819	0.335018
L-Glutamine	0.16719	0.28502	0.160159
L-Hexanoylcarnitine	0.179523	0.214972	0.115176
L-Histidine	0.299082	0.357517	0.720725
L-Kynurenine	0.241936	0.173924	0.279605
L-Lysine	0.253939	0.259965	0.692792
L-Norleucine	0.523615	1.325822	1.933269
L-Phenylalanine	0.223008	0.138711	0.165768
L-Prolyl-4-hydroxy-L-prolin	0.486566	0.400441	0.439451
L-Tyrosine	0.67572	0.649253	0.653036
L-Tyrosyl-L-prolyl-L-tryptophyl-L-threonine	0.27627	0.293397	0.202258
L-Valine	0.251274	2.326997	3.576211
L-Glutamyl-L-valine	0.242976	0.150224	0.191378
Lacosamide	0.362168	0.358232	0.329935
Lenticin	0.16022	0.209501	0.108546
leu-gln	0.942871	0.887265	0.91859
Leu-Val	0.378883	0.296812	0.339353
Leucylasparagine	0.423208	0.352995	0.407638
Leucylphenylalanine	0.163508	0.209994	0.07569

Leucylproline	0.148412	0.235863	0.094682
Levetiracetam	2.381772	2.229663	2.246061
Levodropropizine	0.322456	0.361108	0.372482
Lidocaine	0.138593	0.208654	0.109192
Lincomycin	0.18303	0.245908	0.113663
Lorbamate	0.151322	0.197298	0.10497
Lovastatin	0.151619	0.183767	0.10349
lys-leu	0.396623	0.492484	0.394544
Lysylvaline	0.186148	0.280986	0.166515
Maculosin	0.165839	0.166447	0.091521
Malonylcarnitine	0.412071	0.328398	0.370772
Mercaptopurine	0.319764	0.226919	0.257342
merimepodib	0.4767	0.511171	0.461507
Metanephrine	0.159321	0.184378	0.083612
Metaproterenol	0.202729	0.188581	0.139127
Metharbital	0.304391	0.178655	0.246544
Methotrexate	0.171295	0.226753	0.11013
Methyl (3aR,5Z,9E,11aS)-10-methyl-3-methylene-2-oxo-2,3,3a,4,7,8,11,11a-octahydrocyclodeca[b]furan-6-carboxylate	0.15818	0.246242	0.263893
Methyl {2-[(2E)-5-(hexopyranosyloxy)-2-penten-1-yl]-3-oxocyclopentyl}acetate	0.150116	0.221298	0.082561
methyl 1-methyl-1,2,5,6-tetrahydropyridine-3-carboxylate hydrobromide	0.428445	0.364538	0.542968
methyl 2-(benzoylamino)acetate	0.151156	0.201633	0.070434
Methyl 2-deoxy-3-O-(4-deoxy-4-methyl-D-glucopyranuronosyl)-2-[(Z)-(1-hydroxyethylidene)amino]-D-galactopyranoside	0.539998	0.610709	0.673654
Methyl 2,3-dihydro-3-hydroxy-2-oxo-1H-indole-3-acetate	0.286039	0.471623	0.853315
Methyl O-sulfo-L-tyrosinate	0.36435	0.535876	0.412776
Methylimidazoleacetic acid	0.206398	0.254213	1.507307
Methylone	0.16921	0.161331	0.071626
Methylphenidate	0.154032	0.421352	0.617855
Metronidazole	0.184129	0.330791	0.187088
Miglustat	0.227881	0.369665	0.224221
Miraxanthin-II	0.219989	0.197496	0.222078
Monobutyl phthalate	0.160886	0.225703	0.132681
Monocrotaline	0.185007	0.230135	0.114397
Mycalamide A	0.148155	0.186483	0.099777
Myriocin	0.151595	0.195201	0.113412
Myxalamid A	0.252481	0.284929	0.208852
Myxochelin A	0.514659	0.57637	0.83752
N-(2-Amino-3-phenylpropanoyl)glutamine	0.145898	0.190044	0.077515
N-(2-Furylmethyl)-7-(D-glucopyranosyl)-7H-purin-6-amine	1.248388	1.390662	1.258077
N-(2,3,4-Trimethoxybenzoyl)glycine	0.545433	0.451145	0.536334

N-(3-Carboxypropanoyl)-5-hydroxynorvaline	0.204847	0.305035	0.159285
N-(4-Heptanyl)-1,3-benzodioxole-5-carboxamide	0.151245	0.171053	0.093215
N-(N-(3-Amino-3-carboxypropyl)-3-amino-3-carboxypropyl)azetidine-2-carboxylic acid	0.156409	0.274021	0.146954
N-[(1S)-4-Carbamimidamido-1-carboxybutyl]asparaginylaspartic acid	0.199077	0.286247	0.152527
N-[(5S)-5-Amino-5-carboxypentanoyl]cysteinyl-D-valine	0.243922	0.248057	0.179768
N-[3-Carboxy-2-(carboxymethyl)-2-hydroxypropanoyl]glutamic acid	0.67086	0.879566	0.722492
N-[4-(5-Amino-2,2-dimethyl-4-oxo-3,4-dihydro-2H-chromen-6-yl)-1-hydroxy-4-oxo-2-butanyl]acetamide	0.484051	0.710285	1.165616
N-[4'-hydroxy-(E)-cinnamoyl]-L-aspartic acid	0.192549	0.297121	0.290482
N-{3-[(4-Acetamidobutyl)amino]propyl}acetamide	0.363407	0.590073	0.399255
N-a-Acetyl-L-arginine	0.281418	0.200804	0.22753
N-Acetyl-1-aspartylglutamic acid	0.27831	0.387451	0.276376
N-Acetyl-D-lactosamine	0.252648	0.255438	0.321597
N-Acetyl-L-aspartic acid	0.438187	0.564806	0.444441
N-Acetyl-L-carnosine	0.337344	0.265729	0.28774
N-Acetyl-L-cysteine	0.284255	0.302959	0.263248
N-Acetyl-L-glutamic acid	0.283404	0.431851	0.298402
N-Acetyl-L-histidine	0.263607	0.167046	0.635833
N-Acetyl-L-leucine	0.148272	0.201481	0.192337
N-Acetyl-L-phenylalanine	0.222097	0.214958	0.158961
N-Acetyl-L-tyrosine	0.146561	0.226592	0.102458
N-acetyl-S-(N-allylthiocarbamoyl)-L-cysteine	0.157649	0.192294	0.090869
N-Acetyl-S-2-hydroxyethyl-L-cysteine	0.355982	0.321149	0.315127
N-Acetyl-D-glucosamine	0.521256	0.541227	0.96582
N-Acetylglutamine	0.399395	0.375086	0.364154
N-Acetylhistamine	0.277219	0.173322	0.21545
N-Acetylleucylleucine	0.230877	0.165707	0.153911
N-Acetylneuraminic acid	0.52682	3.481264	4.566382
N-Acetylputrescine	0.25504	0.147585	0.339673
N-Acetylserotonin	0.167842	0.29944	0.138583
N-Acetyltryptophan	0.140351	0.176916	0.084884
N-Acetylvanilalanine	0.191411	0.223194	0.145415
N-Formyl-L-methionine	0.368457	0.372596	0.327426
N-lauroylglycine	0.189277	0.23854	0.150151
N-Methylhydantoin	0.513828	2.040917	3.34829
N-Methyltryptamine	0.167143	0.285684	0.140871
N-Nitrosodibutylamine	0.227034	0.248912	0.198799
N-Phenyl-D-glucopyranosylamine	0.204544	0.180672	0.141962
N-Phenylacetylglutamic acid	0.149707	0.190315	0.103635
N-Phenylacetylglutamine	0.14445	0.189237	0.109358
N-Phenylacetylphenylalanine	0.15486	0.168287	0.0908
N-Propionylmethionine	0.175014	0.180023	0.104914

N-Undecanoylglycine	0.233495	2.57094	4.182982
N,N-Bis(2-hydroxyethyl)dodecanamide	1.089717	4.574854	6.352043
N~2~-(Carboxymethyl)arginine	0.296249	0.206247	0.252059
N~6~-Octanoyllysine	0.184044	0.204302	0.127928
N2-Acetylornithine	0.37063	0.259675	0.317642
N2-Methylguanosine	0.647697	0.629113	0.635784
N4-Acetylcytidine	0.174486	0.184393	0.122642
N6-Acetyl-L-lysine	0.323827	0.215362	0.280305
N6-Me-Adenosine	0.226859	0.234606	0.1859
N6-threonylcarbamoyladenosine	0.152505	0.202709	0.116409
N8-Acetylspermidine	0.340977	0.210551	0.27856
Nadolol	0.20111	0.238502	0.216754
Nalidixic acid	0.176426	0.193687	0.132602
NAPQI	0.164552	0.23256	0.10309
Naringenin 5-O-glucuronide	0.169561	0.202194	0.151012
Naringeninchalcone	0.236606	0.201699	0.281431
Neopterin	0.451713	0.408871	0.410897
Nicotine	0.283076	0.246242	0.34983
Nicotinic acid	0.20623	0.416905	0.860414
Nicotinuric acid	0.139317	0.233614	0.082143
Nigakilactone N	0.392138	0.481155	0.424544
Nikethamide	0.201938	0.252871	0.192334
Nipradilol	0.22576	0.225399	0.165233
Nitrendipine	0.177238	0.186977	0.099717
Nitrosoguvacoline	0.24415	0.143701	0.193018
Nivalenol	0.172226	0.222967	0.10488
Nonanoylcarnitine	0.233895	5.075334	6.13448
Norepinephrine sulfate	0.295327	0.471569	0.313336
Norlidocaine	0.229815	0.32373	0.210802
NP-001346	0.158958	0.229062	0.124838
NP-008998	0.354106	1.386893	2.041588
NP-015114	0.227729	0.343428	0.218186
NP-022229	0.331123	0.35676	0.302823
nylon cyclic dimer	0.314564	0.214933	0.276207
O-(4,8-dimethylnonanoyl)carnitine	0.410737	0.461419	0.399308
O-heptanoylcarnitine	0.144191	0.192052	0.117882
O-ureido-D-serine	0.330518	0.252138	0.448852
Ondansetron	0.179822	0.243813	0.127262
ophthalmic acid	0.238684	0.150508	0.183106
ornithinoalanine	0.242802	0.309088	0.199134
Oxamniquine	0.266799	0.343666	0.235679
Oxepanone	0.202331	0.329517	0.284829
Oxirane, 2-(6-heptenyl)-3-(2,4-pentadiynyl)-	0.963779	0.957275	0.944684
Oxprenolol	0.158189	0.264899	0.127755
p-Cresol	0.164323	0.238606	0.413346
Panthenol	0.182662	0.25095	0.149079

Pantothenic acid	0.194529	0.354304	0.195236
Paracetamol	0.157065	0.287579	0.254364
Paracetamol-cysteine	0.196659	0.152786	0.144436
Paucin	0.198271	0.306611	0.198089
pC-HSL	0.270358	0.386706	0.273914
PEG n6	0.110793	1.136993	2.076918
PEG n7	0.163204	0.538931	0.974217
PEG n8	0.165874	0.482854	0.82985
Peimine	0.149702	0.215384	0.104081
Pentadecanoylcarnitine	0.193626	0.229951	0.212569
pentahomomethionine	0.1745	0.2415	0.142768
pentigetide	0.23274	0.213495	0.169501
Perindopril	0.186217	0.248316	0.139696
Phendimetrazine	0.177205	0.297735	0.387348
Phenol	0.145026	0.192206	0.096181
Phenylacetylglutamine	0.652851	0.707634	0.996101
Phenylglucuronide	0.178633	0.299124	0.154602
Pilocarpine	0.171152	0.249467	0.113951
Pindolol	0.223284	0.189929	0.164897
Pipecolic acid	0.560622	0.715232	1.240568
Pirfenidone	0.304433	0.40387	0.770898
Prednisolone	0.172187	0.245322	0.136126
pretyrosine	0.351671	0.446817	0.341285
Prilocaine	0.141257	0.22835	0.107283
Primidolol	0.140712	0.219276	0.068338
Proacaciberin	0.206209	0.235968	0.148355
prohydrojasmon	0.180909	0.214508	0.131466
Proline	0.335728	0.328209	0.54103
promolate	0.232785	0.249983	0.172716
Proparacaine	0.23695	0.244662	0.242576
Propionylcarnitine	0.311119	0.199369	0.263527
Pulcherriminic acid	0.151327	0.201255	0.106939
Pyridoxal	0.214617	0.19916	0.262255
Pyridoxamine	0.375743	0.853103	1.510623
Pyridoxine	0.359944	1.206134	2.313478
Pyrogallol-2-O-glucuronide	0.629119	0.612531	0.609958
Pyrraline	0.187039	0.352668	0.19961
Quinoline	0.147601	0.203286	0.102019
Rehmaionoside C	0.143138	0.220702	0.132793
Reproterol	0.562148	0.685498	0.640339
Resorcinol diglycidyl ether	0.187084	0.18282	0.10774
Retinoyl b-glucuronide	0.363992	0.473172	0.390087
Riboflavin reduced	0.170393	0.218638	0.230438
Ricinine	0.206369	0.511904	1.034023
Riddelliine	0.161288	0.211871	0.105205
Rimexolone	0.167803	0.184154	0.178869

Ritalinic acid	0.14142	0.204418	0.096882
Rosmarinine	0.200348	0.214554	0.128489
S-(Hydroxyphenylacetothiohydroximoyl)-L-cysteine	0.408751	0.453504	0.420195
S-{[(1aR,7aS,10aS,10bR)-1a,5-Dimethyl-9-oxo-1a,2,3,6,7,7a,8,9,10a,10b-decahydrooxireno[9,10]cyclodeca[1,2-b]furan-8-yl]methyl}-L-cysteine	0.465386	0.485365	0.482862
S-3-oxodecanoyl cysteamine	0.338514	0.390078	0.32874
S-Adenosylhomocysteine	0.602834	0.831539	0.643527
S-Allylcysteine	0.162633	0.188669	0.103062
Salbutamol	0.190686	0.212488	0.125309
Salbutamol 4-O-sulfate	0.517906	0.468755	0.56292
salicyluric D-glucuronide	0.219252	0.29954	0.172677
Salidroside	0.502534	0.530941	0.549261
salinosporamide B	0.163227	0.160661	0.09026
Salsolinol	0.194174	0.189226	0.213657
Sebacic acid	0.156706	0.220601	0.080734
Sedanolide	0.139127	0.820194	1.358763
Senkyunolide	0.85224	1.293216	2.577889
Sertraline	0.341986	1.318373	2.418345
Sinapinic acid	0.175468	0.254403	0.117383
sinapoyltartronic acid	0.314356	0.215867	0.272873
Solasodine	0.150339	0.200107	0.087041
Sorbitan, monododecanoate	0.210312	0.236926	0.198781
Spermidine	0.400294	0.340448	0.361949
Spermine	0.511136	0.484167	0.572616
streptobiosamine	0.215078	0.257041	0.164968
Suberic acid	0.14676	0.744518	3.657796
Suberylglycine	0.144507	0.210205	0.098062
Succinyladenosine	0.182323	0.352256	0.194563
Succinylcarnitine	0.197236	0.194728	0.114244
Sulcatol	0.214047	0.243421	0.260388
Sulfamethoxazole	0.354594	0.605785	0.713066
Sulfisoxazole	0.165422	0.32267	0.157768
Sulforaphane	0.180415	0.976828	2.238646
Sulforaphane-N-acetylcysteine	0.168102	0.168937	0.090055
Sulfurol propionate	0.348165	0.334826	0.327847
Sulpiride	0.701935	0.81657	0.704828
Taurocholic acid	0.312682	0.36701	0.3232
Tazobactam	0.79147	0.935795	0.813885
Terbutaline	0.152427	0.184747	0.098546
terpendole K	0.321131	0.542516	0.978596
Tetradecanedioic acid	0.17236	0.198547	0.128988
Tetraglycol	0.216286	1.326633	2.600929
tetrahomomethionine	0.28373	0.36305	0.252061
Tetrahydrocortisone	0.157787	0.192819	0.122029
Tetrahydrodeoxycorticosterone	0.157561	0.191159	0.144828

tetrahydrothiophene	0.292711	0.376661	0.396093
Theobromine	0.390348	0.568461	0.475733
Theodrenaline	0.256128	0.283413	0.292816
threonylphenylalanine	0.155584	0.200906	0.082986
thymol sulfate	0.170225	0.303033	0.146923
Tiglylcarnitine	0.143955	0.223793	0.118754
Tilisolol	0.297271	0.317945	0.285394
Tixocortol pivalate	0.546773	0.591653	0.551898
Tomatidine	0.155223	0.202701	0.116408
Tranexamic Acid	0.224919	0.199024	0.151521
trans-3-Hexenoic acid	0.176523	0.297176	0.546455
trans-3-Hydroxycotinine glucuronide	0.186818	0.230659	0.13969
trans-3-Indoleacrylic acid	0.148308	0.278071	0.138858
trans-Aconitic acid	0.497782	0.735006	0.534719
trans-Cinnamaldehyde	0.519516	0.567103	0.688743
trans-Zeatin	0.285318	0.419152	0.28294
Trepibutone	0.765782	0.798173	0.764142
Triazolealanine	0.40605	2.407128	3.400803
Triethyl citrate	0.224506	0.23993	0.181852
trilobolide	0.150284	0.195618	0.066166
Trilostane	0.225365	0.807984	1.523081
Tripropionin	0.154557	0.215207	0.10837
Tris(2-chloroethyl) phosphate	0.596277	3.360629	5.340446
Troxipide	0.171437	0.209572	0.109803
trp-asn	0.26009	0.318562	0.318903
trp-gln	1.251359	1.202536	1.410074
Tyramine	0.216763	0.173506	0.158297
Tyramine-O-sulfate	0.328387	0.236152	0.26304
Tyrosylleucine	0.171313	0.303673	0.161411
Ubiquinones	0.143381	0.196457	0.098909
Undecylenic acid	0.415197	0.466046	0.531941
Uric acid	0.384076	0.464273	0.362455
Urolithin A-3-O-glucuronide	0.233669	0.249989	0.227947
Valerylglycine	0.181983	1.076733	2.068156
Valproic acid glucuronide	0.29027	0.342244	0.292613
Valyl-4-hydroxyproline	0.173664	0.194153	0.081477
Valylproline	0.200726	0.375608	0.202452
Valylvaline	0.284595	0.204871	0.228331
Viloxazine	0.224811	0.241033	0.297489
Voglibose	0.772722	0.708369	0.723647
Volkenin	0.184812	0.249742	0.121438
Xamoterol	0.193878	0.177453	0.117697
Xanthine	0.63263	0.881734	0.675956
Xanthurenic acid	0.204458	0.313938	0.191564
Zizyphine A	0.347744	0.363505	0.397917
Aspartylphenylalanine	0.142998	0.219259	0.10682

Aminobutyryl-lysine	0.481968	0.830146	1.452028
Glu-Ala	0.310242	0.415253	0.291309
Glu-Gly	0.3239	0.260512	0.263579
(-)-Caryophyllene oxide	0.254561	0.264222	0.27309
(-)-Epicatechin 7-O-glucuronide	0.195062	0.210161	0.108122
(-)-Slaframine	0.171492	0.202027	0.120046
(+)-ar-Turmerone	1.456936	1.411757	1.404034
(1R,2R,3S,6S,7R,9S,10S,11S,13R,14R)-11-(1-Hydroxy-2-propanyl)-3,7,10-trimethyl-15-oxapentacyclopentadecane-2,6,9,11,13,14-hexol	0.347949	0.401549	0.348237
4-(hydroxymethyl)-4-cyclohexene-1,2,3-triol]	0.229718	0.310433	0.410693
(1S,3R,4R,5R)-1,3,4-trihydroxy-5-{[(2E)-3-(4-hydroxy-3-methoxyphenyl)prop-2-enoyl]oxy}cyclohexane-1-carboxylic acid	0.157461	0.220712	0.117491
(1S,4S)-menthone-8-thioacetate	0.359417	0.37052	0.317329
(1S,5R,9R,13R)-1,5,9-trimethyl-11,14,15,16-tetraoxatetracyclohexadecan-10-one	0.547882	0.665347	0.790285
(1S,5R)-5-Isopropenyl-2-methyl-2-cyclohexen-1-yl Î²-D-glucopyranoside	0.366548	0.454958	0.452968
(2E)-3-Methyl-4-(sulfooxy)-2-butenoic acid	1.756889	2.583507	3.248098
(2E)-N-(4-Amino-2-hydroxybutyl)-3-(3,4-dihydroxyphenyl)acrylamide	0.170207	0.192946	0.114848
(2R,3S,4S,5R,6S)-2-(Hydroxymethyl)-6-[4-(3-hydroxyprop-1-enyl)-2,6-dimethoxyphenoxy]oxane-3,4,5-triol	0.158707	0.225907	0.09472
(2S,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolane-2-carboxamide	0.249847	0.187143	0.195592
(2S,4R,9aR)-4-(2-Acetyl-4-oxo-3(4H)-quinazolinyl)-dimethyl-1,9a-dihydro-3H-spiro[furan-2,9-imidazo[1,2-a]indole]-3,5(2H,4H)-dione	1.432741	1.334294	1.627176
(2S,4R,9aS)-4-(2-Acetyl-4-oxo-3(4H)-quinazolinyl)-hydroxy-2-dimethyl-1-dihydro-3H-spiro[furan-2,9-imidazo[1,2-a]indole]-3,5(2H,4H)-dione	0.216387	0.230962	0.149094
(2S,4S)-hypoglycin B	0.19171	0.369998	0.206855
(2S,8R)-2-Amino-8-hydroxydecanoic acid	0.170426	0.222746	0.115551
(2S)-3-(1H-Imidazol-4-yl)-2-({[(3S,4S,5R)-2,3,4-trihydroxy-5-(hydroxymethyl)tetrahydro-2-furanyl]methyl}amino)propanoic acid	0.417627	0.368444	0.376403
(2S)-3-Phenyl-2-({[(3S,4S,5R)-2,3,4-trihydroxy-5-(hydroxymethyl)tetrahydro-2-furanyl]methyl}amino)propanoic acid	0.163067	0.235149	0.090232
(2S)-4-Methyl-2-({[(3S,4S,5R)-2,3,4-trihydroxy-5-(hydroxymethyl)tetrahydro-2-furanyl]methyl}amino)pentanoic acid	0.291007	0.21632	0.238382
(2Z)-12-Hydroxy-2-(hydroxymethyl)-1-methoxy-11-methyl-7-methylene-6-oxo-5,14-dioxatricyclo[9.2.1.0~4,8~]tetradec-2-en-9-yl (2Z)-2-methyl-2-butenoate	0.554219	0.646187	0.567906
(2Z)-2-(2-Ethoxy-2-oxoethylidene)succinic acid	0.169328	0.198687	0.090592
(2Z)-2-Benzylideneheptyl hydrogen sulfate	0.465283	0.719191	1.230672

(3E,5S,6S,7S,9R,11E,13E,15R,16R)-16-Ethyl-15-(hydroxymethyl)-5,7,9-trimethyl-2,10-dioxooxacyclohexadeca-3,11,13-trien-6-yl 3,4,6-trideoxy-3-(dimethylamino)-D-xylo-hexopyranoside	0.178163	0.204171	0.225226
(3R,5aS,6S,10aR)-6-Hydroxy-3-(hydroxymethyl)-2-methyl-10a-(methylsulfanyl)-3-sulfanyl-2,3,5a,6,10,10a-hexahydropyrazino[1,2-a]indole-1,4-dione	0.217555	0.310916	0.5248
(3R,5aS,6S,10aR)-6-Hydroxy-3-(hydroxymethyl)-2-methyl-3,10a-disulfanyl-2,3,5a,6,10,10a-hexahydropyrazino[1,2-a]indole-1,4-dione	0.181181	0.18985	0.117874
(3s,6R,7S)-6,7-Dihydroxy-8-methyl-8-azabicyclo[3.2.1]oct-3-yl (2E)-2-methyl-2-butenoate	0.169978	0.203964	0.105052
(3S,9S,14aR)-3,6-Dimethyl-9-{6-[(2S)-2-oxiranyl]-6-oxohexyl}decahydropyrrolo[1,2-a][1,4,7,10]tetraazacyclododecine-1,4,7,10-tetrone	0.174336	0.212615	0.102292
(3S)-3-{(Z)-[(3S)-3-{(Z)-[(3R)-3-Amino-1-hydroxy-4-methylpentylidene]amino}-1-hydroxybutylidene]amino}-5-methylhexanoic acid	0.142592	0.215664	0.10452
4,15-Diacetoxy-3-hydroxy-12,13-epoxytrichothec-9-en-8-yl propionate	0.155358	0.220004	0.438429
(4S)-4-{[2-O-(L-Arabinofuranosyl)-L-arabinofuranosyl]oxy}proline	0.630996	0.600658	0.607729
(5Z,8Z)-5,8-Tetradecadienoic acid	0.184557	0.22928	0.168216
(6abeta)-Tazettine	0.4766	0.471912	0.596421
(6R,8Z)-6-Hydroxy-3-oxo-8-tetradecenoic acid	0.701005	0.790497	0.743858
(6R)-5-Acetamido-4-O-acetyl-3,5-dideoxy-6-[(1R)-1,2,3-trihydroxypropyl]-L-threo-hex-2-ulopyranosonic acid	0.540722	0.771716	0.578806
(8S,9Z)-9-Heptadecene-4,6-diyne-1,8-diol	0.184325	0.214268	0.124491
(Carbamoylamino)(4-hydroxyphenyl)acetic acid	0.760532	0.871514	0.815612
(E)-1,3-Tridecadiene-5,7,9,11-tetrayne	0.372078	0.270637	0.422015
(E)-3-O-Methyl entacapone	0.328584	0.323197	0.271197
(R)-1-Aminopropan-2-yl phosphate	0.334285	1.032856	1.898265
(S)-2-hydrazino-3-(4-hydroxy-3-methoxyphenyl)-2-methylpropionic acid	0.159402	0.204186	0.109865
(Z,5Z)-5-{[(1Z,2S)-6-Amino-1-{[(1Z,2R)-1-{[(1R)-1-carboxyethyl]imino}-1-hydroxy-2-propanyl]imino}-1-hydroxy-2-hexanyl]imino}-N-[(2S)-2-amino-1-hydroxypropylidene]-5-hydroxy-L-norvaline	0.188247	0.228683	0.22372
(Z)-desulfoglucotropeolin	0.220312	0.330183	0.506384
[(1S,2R,4S,5R)-5-ethenyl-1-azabicyclo[2.2.2]octan-2-yl](6-methoxyquinolin-4-yl)methanol	0.278387	0.355977	0.277265
[3-(Hydroxymethyl)-3-methyl-2-oxobicyclo[2.2.1]hept-1-yl]methyl hexopyranoside	0.179269	0.253483	0.155587
[4,6-Dihydroxy-2-methoxy-3-(3-methyl-2-buten-1-yl)phenyl]acetic acid	0.183173	0.245038	0.150582
[7-Hydroxy-1-(4-hydroxy-3-methoxyphenyl)-3-(hydroxymethyl)-6-methoxy-1,2,3,4-tetrahydro-2-naphthalenyl]methyl pentofuranoside	0.381939	0.400269	0.381955

{(1R,2R)-2-[(2Z)-5-(D-glucopyranosyloxy)pent-2-en-1-yl]-3-oxocyclopentyl}acetic acid	0.148836	0.223791	0.116434
{2-[2-(Isobutyryloxy)-4-methylphenyl]-2-oxiranyl}methyl 2-methylbutanoate	0.154156	0.225723	0.107515
1-(3-Furyl)-1,4-pentanediol	0.508593	1.430635	2.158423
1-(4-Hydroxy-3-methoxyphenyl)-3,5-hexadecanedione	0.155108	0.251377	0.107449
1-(D-Ribofuranosyl)-1,3,4,7-tetrahydro-2H-1,3-diazepin-2-one	0.312326	0.331567	0.287692
1-[(5-Amino-5-carboxypentyl)amino]-1-deoxyfructose	0.387269	0.423362	0.597084
1-[3,4-Dihydroxy-5-(hydroxymethyl)oxolan-2-yl]-5-hydroxyimidazole-4-carboxamide	0.469088	0.423151	0.420527
1-Adamantanamine	0.222371	0.385806	0.567985
1-Hexyl-2-methylbenzene	0.219908	0.188479	0.172605
1-hydroxyhexanoylglycine	0.148515	0.236306	0.143552
1-Isothiocyanato-2-(methylthio)ethane	0.224669	0.382016	0.659646
1-Methyladenine	0.355671	0.283926	0.306878
1-Methylguanine	0.184886	0.348895	0.199224
1-Methylinosine	0.490488	0.436427	0.461799
1-Methyluric acid	0.289896	0.202773	0.229528
1-O-3,7,12-Trihydroxy-24-oxocholan-24-yl]-D-galactopyranose	0.231936	0.303844	0.241729
1,2-Benzisothiazolin-3-one	0.316986	0.367303	0.322575
1,2,3,4,Tetrahydro-1,5,7-trimethylnapthalene	0.184287	0.173624	0.090317
1,2,4-Trimethoxy-5-propenylbenzene	0.380795	0.402225	0.522571
1,3-Dimethyluric acid	0.15584	0.278945	0.121609
1,3,5-Heptatriene, (E,E)-	0.491579	0.506304	0.4778
1,3,7-trimethyl-2,3,6,7-tetrahydro-1H-purine-2,6-dione	0.142649	0.22335	0.107944
1,4-Dihydroxy-7-isopropylidene-1,4-dimethyloctahydro-6(1H)-azulenone	0.411179	0.433516	0.380939
1,5-Anhydro-1-(2,4,6-trihydroxyphenyl)hexitol	0.182413	0.191995	0.102707
1,5-Anhydro-1-{2-(3,4-dihydroxytetrahydro-2-furanyl)-4,7-dihydroxy-5-[(2E)-3-(4-hydroxyphenyl)-2-propenoyl]-6-oxo-6,7-dihydro-1H-indol-7-yl}hexitol	0.703824	0.662692	0.819274
1,7-Dihydroxy-12-methyl-13-vinyl-2,10-dioxatetracyclo[5.4.1.1~8,11~.0~4,12~]tridecan-9-one	0.199029	0.266665	0.162491
1,7-Dihydroxy-6,6-dimethyl-3,5,5a,6,7,8,9a,9b-octahydronaphtho[1,2-c]furan-9(1H)-one	0.338206	0.373766	0.345525
1,7-Dimethyluric acid	0.169119	0.315508	0.169573
11-Aminoundecanoic acid	0.153241	0.197521	0.077423
17alpha-Hydroxyprogesterone	0.31664	0.331576	0.280233
2-({6-O-[(2R,3R,4R)-3,4-Dihydroxy-4-(hydroxymethyl)tetrahydro-2-furanyl]-D-glucopyranosyl}oxy)-2-methylbutanenitrile	0.214494	0.216444	0.155469

2-(1-Hydroxy-2,2,4,6-tetramethyl-3-oxo-2,3-dihydro-1H-inden-5-yl)ethyl hexopyranoside	0.172566	0.224579	0.117283
2-(1,2-Dihydroxy-2-propanyl)-6,10-dimethylspiro[4.5]dec-6-en-8-one	0.353895	0.406896	0.403242
2-(2-Carboxyethyl)-4-methyl-5-pentyl-3-furoic acid	0.144226	0.223139	0.133948
2-(4-Methyl-5-thiazolyl)ethyl decanoate	0.218399	0.237679	0.190798
2-(4-Methylthiazol-5-yl)ethyl butyrate	0.300344	0.750074	2.089478
2-(6,10-Dimethyl-8-oxospiro[4.5]dec-6-en-2-yl)-2-hydroxypropyl hexopyranoside	0.198571	0.252302	0.211591
2-(acetylamino)-3-(1H-indol-3-yl)propanoic acid	0.149953	0.283641	0.593806
2-(Dimethylamino)-5,6-dimethylpyrimidin-4-ol	0.252255	0.167045	0.197579
2-[4-(3-Hydroxypropyl)-2-methoxyphenoxy]-1,3-propanediol	0.6058	0.809456	2.365916
2-Acetamido-2-deoxy-3-O-(6-deoxy-L-galactopyranosyl)-D-glucose	0.199538	0.224343	0.140353
2-Acetamido-2-deoxy-D-galactopyranosyl-(1->4)-[6-deoxy-L-galactopyranosyl-(1->5)]-4-C-methyl-D-arabinitol	0.242385	0.259127	0.288263
2-Acrylamido-2-methyl-1-propane sulfonic acid	0.452646	0.673284	0.488075
2-Amino-2-deoxy-D-gluconic acid	0.512598	0.563203	0.498325
2-Amino-3-hydroxy-3-phenylpropanoic acid	0.209674	0.322165	0.572911
2-Amino-5-[2-(4-formylphenyl)hydrazino]-5-oxopentanoic acid	0.273086	0.372795	0.613044
2-Amino-6-[(E)-(5-amino-5-carboxy-2-hydroxypentylidene)amino]-5-hydroxyhexanoic acid	0.360357	0.271052	0.307605
2-Aminoadenosine	0.157398	0.226053	0.106028
2-aminophenol sulphate	0.484521	0.394291	0.420015
2-Butenedioic acid (2E)-, 1,4-diethyl ester	0.177766	0.272696	0.136496
2-Butyl-5-ethyl-4-methyloxazole	0.164809	0.165623	0.121344
2-Hexenoylcarnitine	0.175886	0.160825	0.12889
2-Hydroxydecanedioic acid	0.167621	0.223469	0.091037
2-Hydroxyhippuric acid	0.154246	0.228926	0.18942
2-Isocapryloyl-3R-hydroxymethyl-butyrolactone	0.196097	0.565906	0.966476
2-Methoxy-1,3-benzenediol	0.273715	4.247725	5.343949
2-Methyl-3-phenylpropyl hydrogen sulfate	0.1516	0.197112	0.079854
2-Methylbutyroylcarnitine	0.147951	0.239368	0.122285
2-Methylhippuric acid	0.149366	0.198553	0.257756
2-nonenoylglycine	0.169759	0.186457	0.105281
2-octenoylglycine	0.138892	0.175479	0.139017
2-oxo-10-methylthiodecanoic acid	0.282385	0.274556	0.226472
2-Oxo-6-pentyltetrahydro-2H-pyran-3-carboxylic acid	0.141398	0.248805	0.359122
2-oxo-8-methylthiooctanoic acid	0.181049	0.176821	0.082944
2-Oxoarginine	0.323422	0.260614	0.273092
2-Phenylethyl 6-O-D-xylopyranosyl-D-glucopyranoside	0.183369	0.206561	0.105387
2-Phenylethyl D-glucopyranoside	0.149423	0.216648	0.100674
2,3-Diaminosalicylic acid	0.173798	0.144962	0.164676

2,3-Dimethoxy-5-methyl-6-(3-methyl-2-buten-1-yl)-1,4-benzenediol	0.525022	0.540159	0.472497
2,3-Methyleneglutaric acid	0.174119	0.199796	0.097909
2,3,4,5-tetrahydrodipicolinic acid	0.38128	0.351942	0.345284
2,3,4,5,6-Pentahydroxy-N-(2-hydroxyethyl)hexanamide	0.242924	0.329846	0.207426
2,3,4,9-Tetrahydro-1H-carboline-3-carboxylic acid	0.156567	0.262321	0.126225
2,4-diacetamido-2,4,6-trideoxy-L-altrose	0.156606	0.227991	0.142852
2,4-Diamino-6-nitrotoluene	0.443782	5.477325	6.553687
2,4-Dihydroxy-2H-1,4-benzoxazin-3(4H)-one	0.311841	0.395709	0.562996
2,4-Undecadien-1-al	0.190266	0.199006	0.124952
2,8-Dihydroxyquinoline-beta-D-glucuronide	0.196727	0.322235	0.206232
20 Dihydrocortisol	0.18455	0.192592	0.12144
3-(1-carboxyvinyloxy)anthranilic acid	0.150293	0.2165	0.105591
3-(2-methylpropyl)-octahydropyrrolo[1,2-a]pyrazine-1,4-dione	0.136496	0.192154	0.138914
3-(3,4-Dimethoxyphenyl)-2-propenoic acid	0.161798	0.188551	0.08955
3-(6-Amino-1H-purin-1-yl)-1-propanol	0.165556	0.192334	0.094933
3-(6-hydroxyindol-3-yl)lactic acid	0.145131	0.200803	0.072638
3-(Hydroxymethyl)-1-oxo-1H-isochromen-6-yl hexopyranosiduronic acid	0.279528	0.388844	0.585382
3-[(1S)-2-Cyclohexen-1-yl]-L-alanine	0.125672	1.088026	2.109898
3-[(2-Carboxy-2-hydroxyethyl)dithio]-L-alanine	0.376044	0.327648	0.326819
3-[(2Z)-1-Hydroxy-2-buten-2-yl]pentanedioic acid	0.152326	0.212775	0.15662
3-[(3-Hydroxyheptanoyl)oxy]-4-(trimethylammonio)butanoate	0.162271	0.25487	0.115731
3-[(3-Hydroxynonanoyl)oxy]-4-(trimethylammonio)butanoate	0.171243	0.238181	0.121037
3-[(4-hydroxyphenyl)methyl]-octahydropyrrolo[1,2-a]pyrazine-1,4-dione	0.381629	0.582612	0.763403
3-[(6-Oxodecanoyl)oxy]-4-(trimethylammonio)butanoate	0.793184	0.697618	0.777755
3-{[(2E)-4-Methoxy-4-oxo-2-butenoyl]amino}-L-alanyl-L-leucine	0.159157	0.20074	0.110046
3-Acetyl-6-hydroxy-4a,5-dimethyl-4a,5,6,7,8,8a-hexahydro-2(1H)-naphthalenone	0.292538	0.530848	0.75668
3-Hydroxycarbofuran	0.151437	0.186843	0.08417
3-hydroxydecanoyl carnitine	0.174724	0.219724	0.117799
3-Hydroxydodecanedioic acid	0.153818	0.222291	0.095541
3-hydroxydodecanoyl carnitine	0.255821	0.261539	0.192456
3-Hydroxyhexanoylcarnitine	0.148089	0.275076	0.150865
3-hydroxyoctanoyl carnitine	0.289159	0.336023	0.272159
3-Hydroxysebacic acid	0.163365	0.281898	0.144306
3-Hydroxytetradecanedioic acid	0.154293	0.218632	0.105431
3-Ketocarbofuran	0.16722	0.19949	0.091528
3-Mercaptohexyl butyrate	0.314869	0.212759	0.261729
3-Methoxyestra-1,3,5(10),16-tetraene	0.171769	0.176199	0.102384
3-Methoxytyrosine	0.387678	0.334298	0.372071

3-Methylcrotonylglycine	0.152271	0.256121	0.196175
3-Methylglutarylcarnitine	0.224688	0.389234	0.235568
3-Methylhistamine	0.378316	0.301151	0.450531
3-O-(L-olivosyl)oleandolide	0.35047	0.377737	0.387308
3-oxo-C12-HSL	0.751008	0.686687	0.842353
3-Oxoglutaric acid	0.548103	0.797772	0.58469
3-Oxotetradecanoic acid	0.17157	0.164323	0.092652
3-Phenylpent-4-enal	0.267103	0.282215	0.295089
3-Ureidopropionic acid	0.388325	0.326225	0.330726
3,10-Diamino-14-methyl-5,6,6a,6b,7,9-hexahydro-4H-pyrimido[1,4]diazepinopyrrolo[1,2-f]pteridine-1,12-dione	1.283515	1.315622	1.337591
3,12,13-Trihydroxy-11-methyl-6-methylene-16-oxo-15-oxapentacycloheptadecane-9-carboxylic acid	0.385887	0.535815	0.48349
3,4-Dimethyl-5-pentyl-2-furanpropanoic acid	0.428036	0.428799	0.398135
3,4-Methyleneazelaic acid	0.167778	0.454175	0.8604
3,4-Methylenesebacic acid	0.490227	0.528759	0.48135
3,7-Dimethyl-2,6-octadienal	0.337801	0.581575	1.107551
3'-Amino-3'-deoxythimidine glucuronide	0.174598	0.233698	0.140336
3',5,7-Trihydroxy-4'-methoxyflavanone	0.182911	0.17229	0.131611
4-(1-Carboxyethyl)-3-hydroxy-2-(3-methyl-2-buten-1-yl)phenyl hexopyranosiduronic acid	0.227434	0.222286	0.204948
4-(5-Hydroxy-2-methyl-2-azabicyclo[2.2.2]oct-5-yl)-3-methylbutanoic acid	0.307806	0.327221	0.309732
4-[(2-Isopropyl-5-methylcyclohexyl)oxy]-4-oxobutanoic acid	0.383146	0.439739	0.392837
4-[(2-Methyl-3-furyl)thio]-5-nonanone	0.299962	0.388813	0.287204
4-[2-Amino-3-(D-glucopyranosyloxy)phenyl]-4-oxobutanoic acid	0.407868	0.413303	0.607887
4-Acetamidobutanoic acid	0.272663	0.144693	0.216044
4-Acetoxy-2-hexyltetrahydrofuran	0.250128	0.396294	0.536649
4-Aminohippuric acid	0.223261	0.401759	0.245998
4-Guanidinobutyric acid	0.260028	0.186348	0.205836
4-hydroxy-4-(indol-3-ylmethyl)glutamic acid	0.153439	0.216014	0.096546
4-Hydroxy-5-(2-hydroxy-2-propanyl)-2-methylbicyclo[3.1.0]hex-2-yl hexopyranoside	0.159323	0.227224	0.107927
4-Hydroxy-5-methoxy-4-[2-methyl-3-(3-methylbut-2-en-1-yl)oxiran-2-yl]-1-oxaspiro[2.5]octan-6-one	0.415667	0.494686	0.427246
4-Hydroxy-8-(D-talopyranosyloxy)-2-quinolinecarboxylic acid	1.194884	1.290445	1.196596
4-Hydroxycyclohexylcarboxylic acid	0.330318	0.842766	2.403774
4-Hydroxyprolyltryptophan	0.21241	0.320296	0.355715
4-Imino-1-(D-ribofuranosyl)-1,4-dihydro-3-pyridinecarboxylic acid	0.626393	0.58948	0.605776
4-Methoxycinnamaldehyde	0.156663	0.179693	0.083338
4-Methyl-2-propyltetrahydro-2H-pyran-4-yl acetate	0.270894	0.253541	0.296031
4-Methylesculetin	0.170987	0.200403	0.095914
4-Methylumbelliferone hydrate	0.181344	0.168817	0.114255

4-O-(4-Deoxy-L-threo-hex-4-enopyranuronosyl)-D-galactopyranuronic acid	0.168893	0.258251	0.123337
4-O-(L-Araf)-cis-L-Hyp	0.395347	0.54945	0.407396
4-O-D-Glucopyranosylmoranoline	0.198436	0.303968	0.165604
4-Octylphenol	0.169418	0.195086	0.113232
4-Phenyl-3-buten-2-one	0.253315	0.292761	0.421501
4-Phenylbutyric acid	0.169971	0.178328	0.22612
4-Pyridoxic acid	0.289085	0.18074	0.236838
4-Trimethylammoniobutanoic acid	0.378977	0.33956	0.333835
4,4-Thiobis(2-butanone)	0.156203	0.219842	0.114154
4,5-Dihydroxy-3-oxo-1-cyclohexene-1-carboxylic acid	0.306024	0.483998	0.334126
4,7-Dimethoxy-3-oxo-3,4-dihydro-2H-1,4-benzoxazin-2-yl D-glucopyranoside	0.206043	0.22405	0.1391
4'-Methoxyacetophenone	0.180561	0.168285	0.115596
4a,5-dihydroriboflavin	0.292451	0.376518	0.296891
5-(2,3-Dihydroxy-3-methylbutyl)-4-[(3,3-dimethyl-2-oxiranyl)acetyl]-3,4-dihydroxy-2-(3-methylbutanoyl)-2-cyclopenten-1-one	0.340216	0.425656	0.33139
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-3'-O-methyl-4'-O-glucuronide	0.20333	0.276096	0.242057
5-[(Z)-(4-Ethyl-3-methyl-5-oxo-1,5-dihydro-2H-pyrrol-2-ylidene)methyl]-5-methoxy-3-methyl-4-vinyl-1,5-dihydro-2H-pyrrol-2-one	0.171866	0.184286	0.116829
5-Acetylamino-6-amino-3-methyluracil	0.338826	0.265549	0.320089
5-aminosalicyluric acid	0.412906	0.364822	0.356586
5-Butyloxazole	0.146182	0.244515	0.110065
5-Hydantoinpropionic acid	0.382711	0.514114	0.375364
5-Hydroxy-3-(4-hydroxyphenyl)-6-methoxy-7-[(2S,4S,5S)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxychromen-4-one	0.206416	0.238639	0.194096
5-Hydroxy-L-tryptophan	0.205894	0.157593	0.151588
5-Hydroxyindole-3-acetic acid	0.152693	0.21621	0.122772
5-Hydroxyindoleacetaldehyde	0.4765	4.849618	5.860421
5-Hydroxyomeprazole	0.159076	0.226776	0.115586
5-Methoxybenzimidazole	0.361537	0.512448	0.500621
5-Methylangelicin	0.315496	0.261319	0.267641
5-Methyltetrahydrofolic acid	0.40529	0.457839	0.372005
5-methylthioribose	0.394543	0.344147	0.356216
5-Nitro-o-toluidine	0.298825	0.169064	0.227641
5-Sulfanyl-L-histidine	0.216084	0.268094	0.167938
5,5-Dihydroxy-tetramethyl-4,5-dihydro-2H,3H-spiro[furan-2,6-[7]oxabicyclo[3.2.1]oct[3]en]-one	0.137773	0.216802	0.091906
5,7,11-Trihydroxy-7-(methoxymethyl)-2-methyl-10H-spiro[9-oxatricyclo[6.3.1.0~1,5]dodecane-6,3oxetane]10-dione	0.214861	0.269513	0.148408
5,8,12-Trihydroxy-2-oxododecanoic acid	0.188282	0.239948	0.117795
5'-O-beta-D-Glucosylpyridoxine	0.187962	0.2812	0.133289

6 Hydroxycortisol	0.145399	0.207275	0.090997
6-[(1E)-3,4-Dihydroxy-3-methyl-1-buten-1-yl]-7-methoxy-2H-chromen-2-one	0.287639	0.334916	0.344308
6-[(1R,2S)-1,2-Dihydroxypropyl]-3,4-dihydro-2,4-pteridinediol	0.233104	0.153092	0.176159
6-Amino-2-methyl-2-heptanyl (1,3-dimethyl-2,6-dioxo-1,2,3,6-tetrahydro-7H-purin-7-yl)acetate	0.33222	0.393949	0.350302
6-Hydroxy-5-methoxyindole glucuronide	0.151974	0.208701	0.097173
6-Hydroxy-5-methyl-4,11-dioxoundecanoic acid	0.375479	1.894208	3.091779
6-Hydroxy-8-methyl-8-azabicyclo[3.2.1]octan-3-yl 3-hydroxy-2-phenylpropanoate	0.180543	0.23578	0.149031
6-Hydroxypentadecanedioic acid	0.519186	0.580059	0.593053
6-Sulfatoxymelatonin	0.216703	0.374362	0.203563
6-Thiouric acid	0.406417	0.576783	0.424949
7-(3-Amino-3-carboxypropyl)-4,6-dimethyl-3-(D-ribofuranosyl)-3,4-dihydro-9H-imidazo[1,2-a]purin-9-one	0.287992	0.29128	0.261668
7-[1-Formyl-6-hydroxy-6-(hydroxymethyl)bicyclo[3.2.1]oct-2-yl]-3a,7-dimethyl-3-oxooctahydro-2-benzofuran-1-yl hexopyranoside	0.285725	0.360454	0.267243
7-Isopropyl-4a-methyl-1-methylene-1,2,3,4,4a,9,10,10a-octahydrophenanthrene	0.271392	0.287626	0.247027
7-Mercaptoheptanoylthreonine	0.890246	0.925024	0.890357
7-Methylxanthine	0.211842	0.2442	0.138329
7,8-Dihydrobiopterin	0.359871	0.582503	0.401795
7alpha,17beta-Dihydroxyandrost-4-en-3-one	0.227127	0.224643	0.237475
7C-aglycone	0.165707	0.19427	0.103244
8-(3-Furyl)-5-hydroxy-1,1,5a,7a,11b-pentamethyldecahydrooxireno[4,4a]isochromeno[6,5-g][2]benzoxepine-3,10,12(1H,4H,10aH)-trione	0.34269	0.393447	0.417367
8-{[2-Hydroxy-2-(4-hydroxyphenyl)ethyl](methyl)amino}-1,3,7-trimethyl-3,7-dihydro-1H-purine-2,6-dione	0.19955	0.254938	0.12742
8-Epideoxyloganic acid	0.183087	0.227335	0.138661
8-Hydroxy-4-methoxy-7-methyl-7,8-dihydro-5H-furo[2,3-g]isochromen-5-one	0.179834	0.245997	0.144333
8-Hydroxy-5,6-octadienoic acid	0.165926	0.698742	1.209775
8,8a-Diepiswainsonine	0.152113	0.231649	0.074541
9-(D-glucosyl)dihydrozeatin	0.221436	0.21907	0.169796
9-Decenoylcarnitine	0.153543	0.175939	0.129682
9-Methyluric acid	0.58836	0.538194	0.552006
9-O-Demethyl-2-hydroxyhomolycorine	0.438753	0.428139	0.416089
9,10,13-TriHOME	0.084229	3.084001	4.752036
9,12,13-TriHOME	0.076221	2.964374	4.577394

7.4 APPENDIX 4

SUPPLEMENTARY INFORMATION FOR

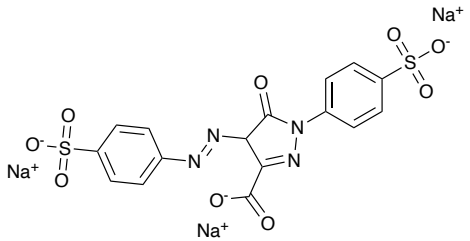
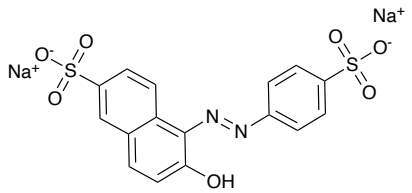
Understanding Food Additives in Inflammatory Bowel Disease: Challenging Perceptions to Improve Gastrointestinal Health

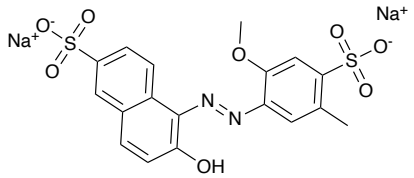
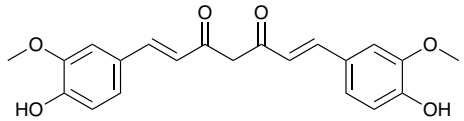
Patricia Kelly[1], Zahra Rattray[1], Konstantinos Gerasimidis[2], Nicholas JW Rattray[1]

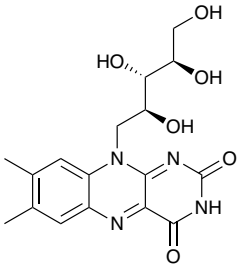
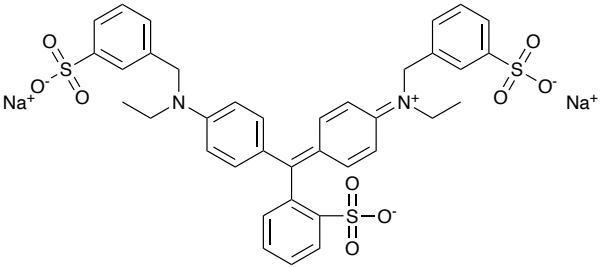
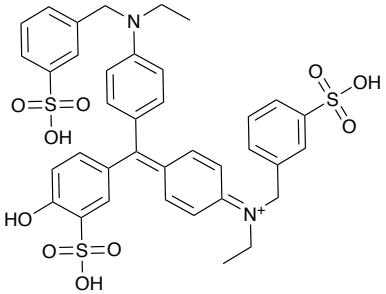
[1] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, UK, G4 0RE

[2] Human Nutrition, School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow Royal Infirmary, Glasgow, UK, G12 8QQ

Table S1. Effects of colours suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory food colours, green E numbers represent anti-inflammatory food colours. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

E Number	Additive	Chemical Structure	Main Findings	Disease Model	Dose Used	ADI (mg/kg bw/day)
E102	Tartrazine (Yellow No. 5)		Tartrazine induced microbial dysbiosis and adverse intestinal changes [1]	Crucian carp	1.4 – 10 mg/kg bw/ day for 60 days	10
E110	Sunset Yellow FCF (Yellow No.6)		Sunset yellow inhibited organoid growth and increased levels of TNF-α and IL-1β **[2]**	Murine intestinal organoids	40 mg/kg bw/day for 7 days	4

E129	Allura Red AC (Red 40)		Allura red AC consumption induces colitis [3]	C57BL/6 mice	7 mg/kg bw/ day for 12 weeks	7
E100	Curcumin (Natural Yellow)		Curcumin decreased TNF-α, MPO, COX-2, and iNOS [4]	TNBS-induced colitis	50-100 mg/kg bw/day for 2 weeks	3
			Curcumin restored balance of cytokine involved in the Treg/T17 pathway and decreased DAI [5]	DSS-induced colitis	100 mg/kg bw/day for 7 days	3
			Curcumin reduced colonic injury and decreased inflammatory markers **[6]**	Mdr1a (-/-) mouse model	Feed + 0.2% curcumin for 12 weeks	3
			Curcumin downregulated pro-inflammatory pathways and reduced histopathological inflammation [7]	Mdr1a (-/-) mouse model	Feed + 0.2% curcumin for 17 weeks	3

E101	Riboflavin (Vitamin B2)		Riboflavin consumption decreased serum inflammatory factors [8]	Patients with CD	100 mg for three weeks	0.5
E133	Brilliant Blue FCF (Blue 1)		Brilliant blue selectively inhibits Panx1 **[9]**	Oocytes	IC_{50} of 0.27 µM	6
E143	Fast Green FCF (Food Green 3)		Panx1 channels, involved in ATP release, are inhibited by fast green FCF [9]	Oocytes	IC_{50} of 3 µM	25

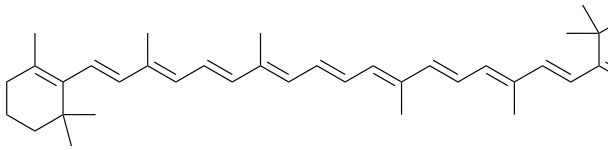
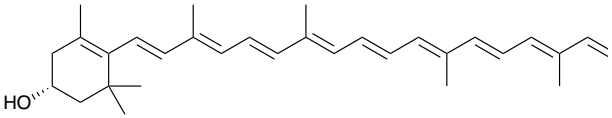
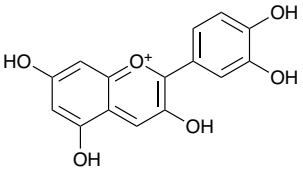
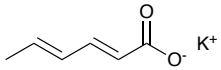
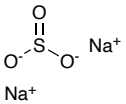
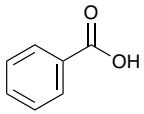
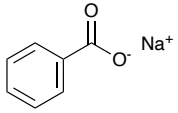
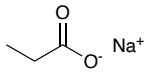
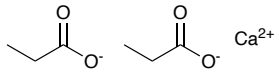
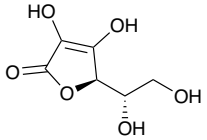
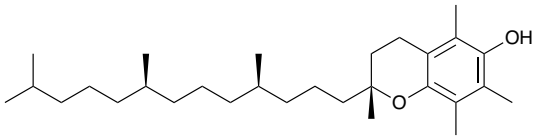
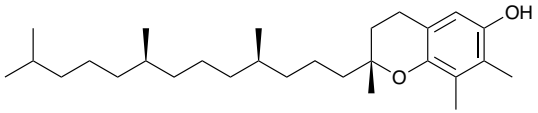
E160a	Beta-carotene (Food Orange 5)		β-Carotene decreased pro-inflammatory cytokine levels and enhanced tight junction protein levels [10]	HT-29 cells	20 mg/ kg bw/day for 28 days	5
E161b	Lutein		Lutein reduced reactive oxygen species and nitric oxide production in vitro and decreased disease activity index in vivo [11]	Male Swiss mouse model of UC	24.6 mg/kg bw/ day for 7 days	1
E163	Anthocyanins, e.g., Cyanidin		Anthocyanins protected against colonic damage through restoring IL-10 and decreasing NO, MPO, IL-12, TNF-α and IFN-γ levels [12]	TNBS-induced colitis	10-40 mg/kg bw/ day for 6 days	2.5

Table S2. Effects of preservatives suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory preservatives, green E numbers represent anti-inflammatory preservatives. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

E Number	Additive	Chemical Structure	Main Findings	Disease Model	Dose Used	ADI (mg/kg bw/day)
E202	Potassium sorbate		Potassium sorbate decreased TNF-α and IL-1β [13]	Zebrafish	0.1-1 g/L for 2 weeks	11
E221	Sodium sulfite		Sodium sulfite inhibited growth of *Faecalibacterium prausnitzii,* a protective microbial species **[14]**	*Faecalibacterium prausnitzii* culture	0.1% (wt/vol) in media	0.7
E210	Benzoic acid		Benzoic acid consumption resulted in	Female weaner pigs	5g/kg for 2 weeks	20

			intestinal healing [15]			
E211	Sodium benzoate		High doses of sodium benzoate reduced MPO and GSH levels [16]	Acetic acid-induced UC.	400-800 mg/kg i.p. for 7 days	5
E220	Sulphur dioxide	O=S=O	Sulphur dioxide decreased NF-κB and inflammasome activation [17]	TNBS-induced colitis	0.18-0.54 mmol/kg for 72 hours	0.7
E281	Sodium propionate		Sodium propionate inhibited inflammatory factors indicative of intestinal inflammation [18]	DSS-induced colitis	1% (w/v) in drinking water for 14 days	20
			Sodium propionate	J774-A1 cell line	0.1-10 mM for 24 hours	20

			decreased iNOS, COX-2 and inflammatory markers [19]			
E282	Calcium propionate		Calcium propionate decreased IFN-γ and calprotectin and increased PGlyRP3 **[20]**	DSS-induced colitis	3.85% (w/v) in drinking water for 7 days	1
E300	Ascorbic acid		Ascorbic acid decreased inflammatory and oxidative stress markers [21]	DSS-induced colitis	100 mg/kg bw/ day for 7 days	1.3 (RDI)
			Ascorbic acid supplementation increased levels of ZO-1 mRNA in guinea pigs and	Guinea pig and SW480 cells.	10-200 mg/kg bw/day for 4 days	1.3 (RDI)

			increased expression of Notch 1 in SW480 cells [22]			
E307	α-tocopherol		α-tocopherol improved intestinal barrier function [23]	Caco-2 cells	5 mg/kg bw/day	2
			DAI score significantly decreased after 12 weeks of α-tocopherol administration, with 64% achieving remission [24]	Human trial	8000 U/d enema for 12 weeks	2
E308	γ- tocopherol		γ- tocopherol restored microbial balance in mice	DSS-induced colitis and Caco-2 cells	5 mg/kg bw/day	2

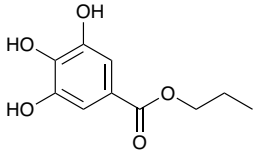
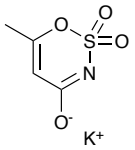
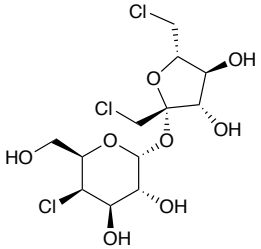
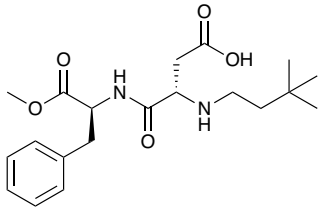
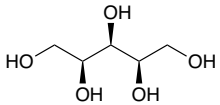
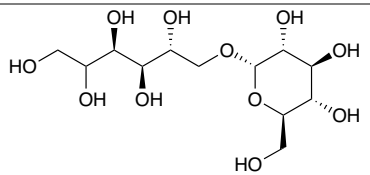
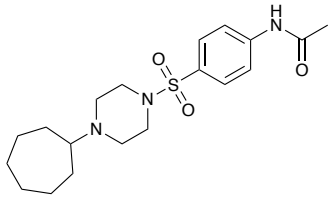
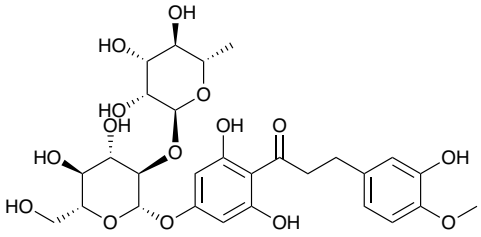
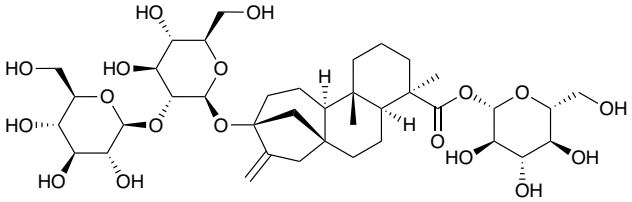
			induced with colitis and improved intestinal barrier function in Caco-2 cells [23]			
E310	Propyl gallate		Propyl gallate reduced pro-inflammatory cytokines and oxidative stress markers, reducing colitis severity **[25]**	DSS-induced colitis	50 mg/kg bw/ day for 21 days	0.5

Table S3. Effects of sweeteners suggested to exert inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory sweeteners, green E numbers represent anti-inflammatory sweeteners. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

E Number	Additive	Chemical Structure	Main Findings	Disease Model	Dose Used	ADI (mg/kg bw/day)
E950	Acesulfame potassium		Ace-K amplified the expression of pro-inflammatory cytokines and decreased the expression of GLP1R and GLP2R [26]	C57BL/6J mice	150 mg/kg bw/day for 8 weeks	9
E955	Sucralose		Sucralose caused microbial dysbiosis, increasing risk of intestinal inflammation **[27]**	C57BL/6 male mice	5 mg/kg bw/day for 6 months	15
			Sucralose increased expression of pro-inflammatory cytokines and exacerbates DSS-induced colorectal tumours [28]	DSS-induced colitis	1.5 mg/mL in drinking water for 6 weeks	15

E961	Neotame		Neotame induced gut microbial dysbiosis with alterations in α and β diversity [29]	CD-1 mice	0.75 mg/kg bw/day for 4 weeks	2
E967	Xylitol		Xylitol increased SCFA production, with significant shift in propionate **[30]**	C57BL/6 mice and in vitro colon model (CDMN)	2.17-5.43 g/kg bw/day in food for 3 months	428 (RDI)
E953	Isomalt		Isomalt caused a protective shift in gut microbial composition [31]	Healthy human individuals	30 g/ day for four weeks	25
E957	Thaumatin		Thaumatin increases the total biomass of the microbiome	Human faecal samples		1.1

E959	NHDC		NHDC induced microbial shifts towards a protective phenotype [32]	Suckling Landrace X Large White piglets	0.5 mM in growth media	5
E960	Stevioside		Stevioside decreased levels of TNF-α and IL-6 in RAW264.7 cells and improved inflammation in DSS-induced mice models [33]	DSS-induced colitis and RAW264.7 cells	50-100 mg/ kg bw for 12 days	4
			Steviol glycoside supplementation resulted in changes to the gut microbiome [34]	Human faecal smaples and Cebus apella model	6.2 mg/kg bw/day for 2 weeks	4

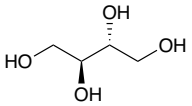
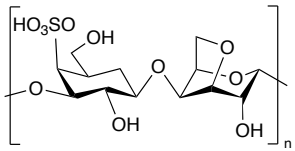
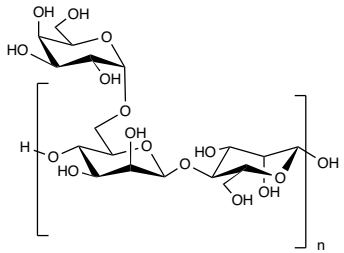
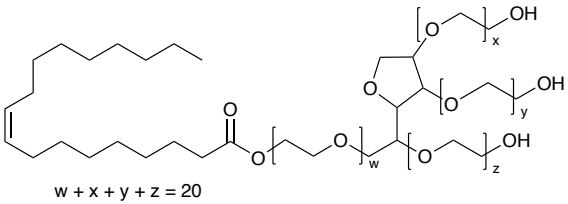
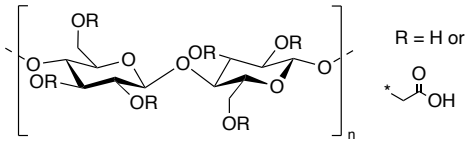
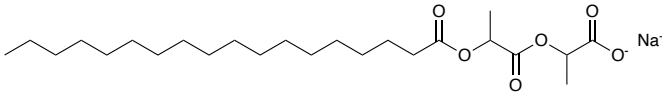
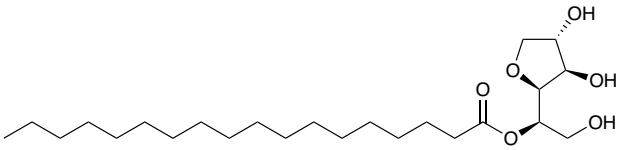
E968	Erythritol		Erythritol supplementation resulted in changes to the gut microbiome [34]	Human faecal smaples and Cebus apella model	6.2 mg/kg bw/day for 2 weeks	500
			Erythritol increased SCFAs, ILC3 and ILC2 [35]	C57BL/6J mice	5% in drinking water for 12 weeks	500

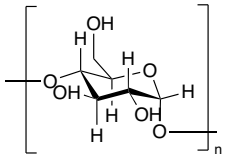
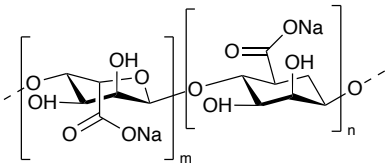
Table S4. Effects of emulsifiers, thickeners, and stabilisers exerting inflammatory or therapeutic effects on IBD, red E numbers represent inflammatory additives, green E numbers represent anti-inflammatory additives. For additives with an unspecified ADI, the average daily intake or recommended daily intake values are given.

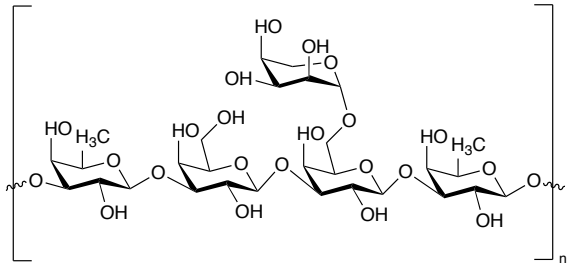
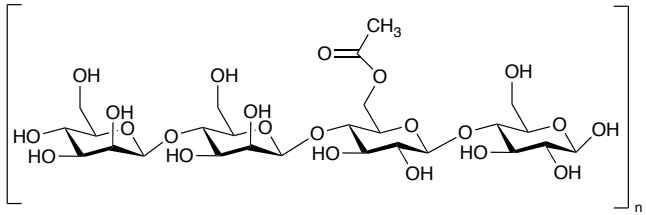
E Number	Additive	Chemical Structure	Main Findings	Disease Model	Dose Used	ADI (mg/ kg bw/day)
E407	Carrageenan		Carrageenan stimulated inflammtory features comparable to UC [36]	Guinea pigs	5% aqueous solution in drinking water for 45 days	75
			Carrageenan increased PEG-900 absorption [37]	Male Sprague-Dawley rats and weanling guinea pigs	50 g/L in drinking water for 4 weeks	75
			Carrageenan consumption caused severe diarrhoea in rats and colonic ulcerations in guinea pigs [38]	Rats and guinea pigs	0.25-5% solution in drinking water for 12 weeks	75

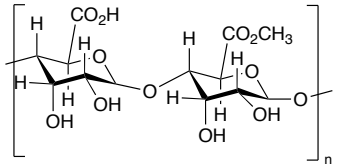
			Carrageenan increased activation of NFκB and BCL10 [39]	NCM460 and HT29 cells	1 µg/mL for 60 hours	75
			Carrageenan exposure inhibited sulfatase activity and increased glycosaminoglycans [40]	NCM460, T84, and CaCO2 cells.	1 µg/mL for 4 days	75
E412	Guar gum		Guar gum exacerbated inflammatory signs including increased levels of Lcn2, IL-1β KC, and SAA **[41]**	DSS-induced colitis, IL-10R neutralizatio n.	7.5% (w/w) in diet for 7 days	6.1 (RDI)

E436	Polysorbate 80		P80 induced a detrimental shift in the gut microbiota [42]	Human faecal samples	0.1% added to bioreactor medium for 216 hours	25
			P80 consumption resulted in inflammation and damage to the intestinal barrier in offspring [43]	C57BL/6J mice	1% (w/v) in drinking water for 3 weeks	25
E466	Carboxymethyl cellulose		CMC induced a detrimental shift in the gut microbiota [42]	Human faecal samples	0.1% added to bioreactor medium for 216 hours	900
			CMC treatment increased Lcn2 and inflammatory cytokine expression, in	Faecal transplant	1% in drinking water for four weeks	900

			comparison to P80 and controls [44]	colonised ex-GF IL10$^{-/-}$ mice		
			CMC increased levels of flagellin and other inflammatory bacteria [45]	C57BL/6 mice	100 mg daily for three weeks	900
E481	Sodium stearoyl lactylate		SSL increased levels of pro-inflammatory microbial communities [46]	Human faecal samples	0.025% (w/v)	22
E491	Sorbitan monostearate		SMS impacted bacterial composition and function in a non-reversible manner and	Human faecal samples	0.1% added to bioreactor medium for 216 hours	25

			increased LPS levels [42]			
E1400	Maltodextrin		Maltodextrin exacerbated colitis and decreased gut microbial diversity [47]	IL10KO mice	1% (w/w) in food for 11 weeks	428.6 (RDI)
E401	Sodium alginate		Sodium alginate administration decreased inflammatory markers and increased protective markers, reducing colonic damage score [48]	Acetic acid-induced colitis	0.5% (w/v) in drinking water for one week	0.4 (RDI)
			Sodium alginate treatment improved	DSS- and TNBS-	500-1000 mg/kg/day for 7 days	0.4 (RDI)

			colitis and repaired goblet cell damage [49]	induced colitis		
E414	Acacia gum		Gum Arabic demonstrated protective effects via reduced colonic fibrosis and *TGFβ1* expression [50]	DSS-induced colitis	140 g/L in drinking water	6.1 (RDI)
			Acacia gum altered gut bacteria towards an anti-inflammatory pDFhenotype [51]	In vitro colon model with human faecal microbiota	0.1% added to culture vessels for 24 hours	6.1 (RDI)
E425	Konjac glucomannan		KGM decreased levels of pro-inflammatory cytokines and leukocyte infiltration [52]	DSS-induced colitis	2% (w/w) in food for 29 days	7.1 (RDI)

E440	Pectin		Mice fed with orange pectin prior to induction of colitis had reduced signs of inflammation [53]	DSS- and TNBS-induced colitis	5% dietary pectin for 14 days	7.4 (RDI)

References for Chapter 5 Supplementary Information

1. Wu, L., et al., *Impacts of an azo food dye tartrazine uptake on intestinal barrier, oxidative stress, inflammatory response and intestinal microbiome in crucian carp (Carassius auratus).* Ecotoxicol Environ Saf, 2021. **223**: p. 112551.
2. Kong, X., et al., *Effects of sunset yellow on proliferation and differentiation of intestinal epithelial cells in murine intestinal organoids.* J Appl Toxicol, 2021. **41**(6): p. 953-963.
3. Kwon, Y.H., et al., *Chronic exposure to synthetic food colorant Allura Red AC promotes susceptibility to experimental colitis via intestinal serotonin in mice.* Nat Commun, 2022. **13**(1): p. 7617.
4. Camacho-Barquero, L., et al., *Curcumin, a Curcuma longa constituent, acts on MAPK p38 pathway modulating COX-2 and iNOS expression in chronic experimental colitis.* Int Immunopharmacol, 2007. **7**(3): p. 333-42.
5. Wei, C., et al., *Curcumin ameliorates DSS-induced colitis in mice by regulating the Treg/Th17 signaling pathway.* Mol Med Rep, 2021. **23**(1).
6. Cooney, J.M., et al., *A combined omics approach to evaluate the effects of dietary curcumin on colon inflammation in the Mdr1a(-/-) mouse model of inflammatory bowel disease.* J Nutr Biochem, 2016. **27**: p. 181-92.
7. Nones, K., et al., *The effects of dietary curcumin and rutin on colonic inflammation and gene expression in multidrug resistance gene-deficient (mdr1a-/-) mice, a model of inflammatory bowel diseases.* Br J Nutr, 2009. **101**(2): p. 169-81.
8. von Martels, J.Z.H., et al., *Riboflavin Supplementation in Patients with Crohn's Disease [the RISE-UP study].* J Crohns Colitis, 2020. **14**(5): p. 595-607.
9. Wang, J., D.G. Jackson, and G. Dahl, *The food dye FD&C Blue No. 1 is a selective inhibitor of the ATP release channel Panx1.* J Gen Physiol, 2013. **141**(5): p. 649-56.
10. Cheng, J., et al., *The Role of β-Carotene in Colonic Inflammation and Intestinal Barrier Integrity.* Front Nutr, 2021. **8**: p. 723480.
11. Meurer, M.C., et al., *Hydroalcoholic extract of Tagetes erecta L. flowers, rich in the carotenoid lutein, attenuates inflammatory cytokine secretion and improves the oxidative stress in an animal model of ulcerative colitis.* Nutr Res, 2019. **66**: p. 95-106.

12. Wu, L.H., et al., *Protective Effect of Anthocyanins Extract from Blueberry on TNBS-Induced IBD Model of Mice.* Evid Based Complement Alternat Med, 2011. **2011**: p. 525462.
13. Peng, Q., et al., *Potassium sorbate suppresses intestinal microbial activity and triggers immune regulation in zebrafish (Danio rerio).* Food Funct, 2019. **10**(11): p. 7164-7173.
14. Jimenez Loayza, J.J., et al., *P837 The common food additives sodium sulfite and polysorbate 80 have a profound inhibitory effect on the commensal, anti-inflammatory bacterium Faecalibacterium prausnitzii: the ENIGMA study.* Journal of Crohn's and Colitis, 2019. **13**(Supplement_1): p. S542-S543.
15. Halas, D., et al., *Dietary supplementation with benzoic acid improves apparent ileal digestibility of total nitrogen and increases villous height and caecal microbial diversity in weaner pigs.* Animal Feed Science and Technology, 2010. **160**(3): p. 137-147.
16. Walia, D., et al., *Exploring the therapeutic potential of sodium benzoate in acetic acid-induced ulcerative colitis in rats.* J Basic Clin Physiol Pharmacol, 2019. **30**(5).
17. Banerjee, S., et al., *Sulphur dioxide ameliorates colitis related pathophysiology and inflammation.* Toxicology, 2019. **412**: p. 63-78.
18. Tong, L.C., et al., *Propionate Ameliorates Dextran Sodium Sulfate-Induced Colitis by Improving Intestinal Barrier Function and Reducing Inflammation and Oxidative Stress.* Front Pharmacol, 2016. **7**: p. 253.
19. Filippone, A., et al., *The Anti-Inflammatory and Antioxidant Effects of Sodium Propionate.* Int J Mol Sci, 2020. **21**(8).
20. Ghadimi, D., et al., *Calcium propionate alleviates DSS-induced colitis and influences lipid profile, Interferon gamma, peptidoglycan recognition protein 3 and calprotectin in one-year-old female and male C57BL/6 mice.* Scientific Journal of Immunology and Immunotherapy, 2017. **1**: p. 07-016.
21. Yan, H., et al., *Ascorbic acid ameliorates oxidative stress and inflammation in dextran sulfate sodium-induced ulcerative colitis in mice.* Int J Clin Exp Med, 2015. **8**(11): p. 20245-53.
22. Qiu, F., et al., *Combined effect of vitamin C and vitamin D(3) on intestinal epithelial barrier by regulating Notch signaling pathway.* Nutr Metab (Lond), 2021. **18**(1): p. 49.

23. Liu, K.Y., et al., *Vitamin E alpha- and gamma-tocopherol mitigate colitis, protect intestinal barrier function and modulate the gut microbiota in mice.* Free Radic Biol Med, 2021. **163**: p. 180-189.
24. Mirbagheri, S.A., et al., *Rectal administration of d-alpha tocopherol for active ulcerative colitis: a preliminary report.* World J Gastroenterol, 2008. **14**(39): p. 5990-5.
25. Chougule, P.R., et al., *Effects of Ethyl Gallate and Propyl Gallate on Dextran Sulfate Sodium (DSS)-Induced Ulcerative Colitis in C57BL/6J Mice: Preventive and Protective*. 2023, Research Square.
26. Hanawa, Y., et al., *Acesulfame potassium induces dysbiosis and intestinal injury with enhanced lymphocyte migration to intestinal mucosa.* J Gastroenterol Hepatol, 2021. **36**(11): p. 3140-3148.
27. Bian, X., et al., *Gut Microbiome Response to Sucralose and Its Potential Role in Inducing Liver Inflammation in Mice.* Front Physiol, 2017. **8**: p. 487.
28. Li, X., et al., *Sucralose Promotes Colitis-Associated Colorectal Cancer Risk in a Murine Model Along With Changes in Microbiota.* Front Oncol, 2020. **10**: p. 710.
29. Chi, L., et al., *Effects of the Artificial Sweetener Neotame on the Gut Microbiome and Fecal Metabolites in Mice.* Molecules, 2018. **23**(2).
30. Xiang, S., et al., *Xylitol enhances synthesis of propionate in the colon via cross-feeding of gut microbiota.* Microbiome, 2021. **9**(1): p. 62.
31. Gostner, A., et al., *Effect of isomalt consumption on faecal microflora and colonic metabolism in healthy volunteers.* Br J Nutr, 2006. **95**(1): p. 40-50.
32. Daly, K., et al., *Bacterial sensing underlies artificial sweetener-induced growth of gut Lactobacillus.* Environ Microbiol, 2016. **18**(7): p. 2159-71.
33. Alavala, S., et al., *Stevioside, a diterpenoid glycoside, shows anti-inflammatory property against Dextran Sulphate Sodium-induced ulcerative colitis in mice.* Eur J Pharmacol, 2019. **855**: p. 192-201.
34. Mahalak, K.K., et al., *Impact of Steviol Glycosides and Erythritol on the Human and Cebus apella Gut Microbiome.* J Agric Food Chem, 2020. **68**(46): p. 13093-13101.
35. Kawano, R., et al., *Erythritol Ameliorates Small Intestinal Inflammation Induced by High-Fat Diets and Improves Glucose Tolerance.* Int J Mol Sci, 2021. **22**(11).
36. Watt, J. and R. Marcus, *Carrageenan-induced ulceration of the large intestine in the guinea pig.* Gut, 1971. **12**(2): p. 164-71.

37. Delahunty, T., L. Recher, and D. Hollander, *Intestinal permeability changes in rodents: a possible mechanism for degraded carrageenan-induced colitis.* Food Chem Toxicol, 1987. **25**(2): p. 113-8.
38. Grasso, P., et al., *Studies on degraded carrageenan in rats and guinea-pigs.* Food and Cosmetics Toxicology, 1975. **13**(2): p. 195-201.
39. Borthakur, A., et al., *Prolongation of carrageenan-induced inflammation in human colonic epithelial cells by activation of an NFκB-BCL10 loop.* Biochim Biophys Acta, 2012. **1822**(8): p. 1300-7.
40. Yang, B., et al., *Exposure to common food additive carrageenan leads to reduced sulfatase activity and increase in sulfated glycosaminoglycans in human epithelial cells.* Biochimie, 2012. **94**(6): p. 1309-16.
41. Nair, D.V.T., et al., *Food Additive Guar Gum Aggravates Colonic Inflammation in Experimental Models of Inflammatory Bowel Disease.* Curr Dev Nutr, 2021. **5**(Suppl 2): p. 1142.
42. Naimi, S., et al., *Direct impact of commonly used dietary emulsifiers on human gut microbiota.* Microbiome, 2021. **9**(1): p. 66.
43. Jin, G., et al., *Maternal Emulsifier P80 Intake Induces Gut Dysbiosis in Offspring and Increases Their Susceptibility to Colitis in Adulthood.* mSystems, 2021. **6**(2).
44. Rousta, E., et al., *The Emulsifier Carboxymethylcellulose Induces More Aggressive Colitis in Humanized Mice with Inflammatory Bowel Disease Microbiota Than Polysorbate-80.* Nutrients, 2021. **13**(10).
45. Chassaing, B., et al., *Dietary emulsifiers directly alter human microbiota composition and gene expression ex vivo potentiating intestinal inflammation.* Gut, 2017. **66**(8): p. 1414-1427.
46. Elmén, L., et al., *Dietary Emulsifier Sodium Stearoyl Lactylate Alters Gut Microbiota in vitro and Inhibits Bacterial Butyrate Producers.* Front Microbiol, 2020. **11**: p. 892.
47. Zangara, M.T., et al., *Maltodextrin Consumption Impairs the Intestinal Mucus Barrier and Accelerates Colitis Through Direct Actions on the Epithelium.* Front Immunol, 2022. **13**: p. 841188.
48. Mirshafiey, A., et al., *Sodium alginate as a novel therapeutic option in experimental colitis.* Scand J Immunol, 2005. **61**(4): p. 316-21.

49. Yamamoto, A., et al., *Effect of sodium alginate on dextran sulfate sodium- and 2,4,6-trinitrobenzene sulfonic acid-induced experimental colitis in mice.* Pharmacology, 2013. **92**(1-2): p. 108-16.
50. Al-Araimi, A., et al., *Gum Arabic Supplementation Suppresses Colonic Fibrosis After Acute Colitis by Reducing Transforming Growth Factor β1 Expression.* J Med Food, 2021. **24**(12): p. 1255-1263.
51. Rawi, M.H., et al., *Manipulation of Gut Microbiota Using Acacia Gum Polysaccharide.* ACS Omega, 2021. **6**(28): p. 17782-17797.
52. Changchien, C.H., C.H. Wang, and H.L. Chen, *Konjac glucomannan polysaccharide and inulin oligosaccharide ameliorate dextran sodium sulfate-induced colitis and alterations in fecal microbiota and short-chain fatty acids in C57BL/6J mice.* Biomedicine (Taipei), 2021. **11**(3): p. 23-30.
53. Ishisono, K., et al., *Dietary Fiber Pectin Ameliorates Experimental Colitis in a Neutral Sugar Side Chain-Dependent Manner.* Front Immunol, 2019. **10**: p. 2979.