Evaluation of Direct from Sample Metabolomics of Human Faeces using Rapid Evaporative Ionisation Mass Spectrometry (REIMS)


Published in:
Analytical Chemistry

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2019 American Chemical Society. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Evaluation of Direct from Sample Metabolomics of Human Faeces using Rapid Evaporative Ionisation Mass Spectrometry (REIMS)

Simon J.S. Cameron (1)†, James L. Alexander (1)†, Frances Bolt (1), Adam Burke (1), Hutan Ashrafian (1), Julian Teare (1), Julian R. Marchesi (1,2), James Kinross (1), Jia V. Li (1), Zoltán Takáts (1)*.

(1) Department of Surgery and Cancer, Imperial College London, London, SW7 2AZ, UK.

(2) School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK.

† Contributed equally.

* Corresponding Author: Professor Zoltan Takáts, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, South Kensington Campus, Sir Alexander Fleming Building, London, SW7 2AZ. Telephone: 020 7594 2760. Email: z.takats@imperial.ac.uk
Abstract

Mass spectrometry is a powerful tool in the investigation of the human faecal metabolome. However, current approaches require time-consuming sample preparation, chromatographic separations, and consequently long analytical run times. Rapid evaporative ionisation mass spectrometry (REIMS) is a method of ambient ionisation mass spectrometry and has been utilised in the metabolic profiling of a diverse range of biological materials, including human tissue, cell culture lines, and microorganisms. Here, we describe the use of an automated, high-throughput REIMS robotic platform for direct analysis of human faeces. Through the analysis of faecal samples from five healthy male participants, REIMS analytical parameters were optimised and used to assess the chemical information obtainable using REIMS. Within the faecal samples analysed, bile acids, including primary, secondary, and conjugate species were identified, and phospholipids of possible bacterial origin were detected. In addition, the effect of storage conditions and consecutive freeze/thaw cycles was determined. Within the REIMS mass spectra, the lower molecular weight metabolites, such as fatty acids, were shown to be significantly affected by storage conditions for prolonged periods at temperatures above -80°C, and consecutive freeze/thaw cycles. However, the complex lipid region was shown to be unaffected by these conditions. A further cohort of 50 faecal samples, collected from patients undergoing bariatric surgery, were analysed using the optimised REIMS parameters, and the complex lipid region mass spectra used for multivariate modelling. This analysis showed a predicted separation between pre- and post-surgery specimens, suggesting that REIMS analysis can detect biological differences, such as microbiome-level differences, which have traditionally been reliant upon methods utilising extensive sample preparations and chromatographic separations and/or DNA sequencing.
Introduction

Mass spectrometry is a powerful and widely used tool for the metabolic profiling of biological fluids and tissues. It is capable of generating extensive information on the metabolic composition of a sample, which allows for interrogation of metabolic function, host-microbiome dysbiosis, response to external stimuli and stress, and biomarker discovery for disease. As a result of its ease of sampling, the analysis of human faeces is commonplace for the study of the gut microbiome. It is also at the interface of the cross-talk between the gut microbiome and the host, which is involved in the regulation of a number of host metabolic pathways. However, many studies look solely at the taxonomic composition of the microbes in a gut microbiome, meaning insightful metabolomic data is overlooked. In the study of the human metabolome, samples such as urine and blood are frequently used; primarily as a result of their ease of collection and rich content of metabolites. However, alternative sample types show promise as a source of disease biomarkers, particularly those which are at the interface of disease onset, such as sputum or exhaled breath for lung cancer, and saliva for oral cancers. Human faeces has received particular attention in recent years as a source of disease biomarkers, primarily due to its ease and non-invasive method of collection, and localisation to affected gastrointestinal diseases sites, such as colorectal cancer and inflammatory bowel disease.

Mass spectrometric techniques and platforms which are commonly used in the metabolic profiling of human faeces include liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS). Both of these techniques require extensive sample preparation, such as the creation of faecal water samples, chemical derivatisation, or targeted metabolite extraction and/or removal. The requirement for sample preparation reduces analytical throughput, is typically user intensive and unpleasant, and can result in significant chemical changes, which may impair the ability to identify, for example, disease biomarkers. In addition, both methods are reliant on chromatographic separation before mass spectrometry analysis; which carries the benefit of allowing...
high levels of analytical resolution, but extends sample running time within the range of approximately
ten to 20 minutes.

In recent years, the field of ambient ionisation mass spectrometry has expanded rapidly, allowing mass
spectrometric analysis of a biological sample with no preparative steps and under normal atmospheric
conditions. One such method is rapid evaporative ionisation mass spectrometry (REIMS); which has
seen successful applications including the *in-situ*, real-time detection of intra-operative tumour
margins and as a platform for the species-level classification of clinically important microorganisms
directly from a culture plate. REIMS works by applying a radiofrequency electrical current directly
to a sample. The sample, as a result of its non-zero impedance, rapidly heats and gas phase ions of
metabolites and structural lipids are released into an analyte-containing vapour, which is
subsequently aspirated into a mass spectrometer, using the instrument’s operational vacuum, for
mass spectrometric analysis. Initial applications of REIMS utilised electrosurgical hand-pieces, such as
electrical diathermy surgical devices and bipolar surgical forceps, to direct the electrical current and
aspirate the analyte-containing vapour to a mass spectrometer. Recent advances have adapted
REIMS to a modified liquid handling platform, which allows for the high-throughput (approximately 3-
4,000 samples per 24 hour period) semi-automated analysis of samples with minimal user input.
Although initially developed to be suitable for implementation in a high-throughput clinical
microbiology diagnostic laboratory, the platform is highly versatile and capable of analysing a wide
range of biological samples using a universal set-up.

In this study, we report on the first application of REIMS to the metabolic profiling of human faecal
samples; showing its ability to detect bile acids and phospholipid markers of microbiome constituents.
Further, we utilise the optimised REIMS parameters identified to the analysis of faecal samples from
patients undergoing bariatric surgery, showing that REIMS possesses the analytical sensitivity to
detect biological changes resulting from a clinical intervention.
Experimental Section

Collection of Faecal Samples

For REIMS analysis of human faeces, a total of five healthy male participants, with no history of gastrointestinal disease, donated a stool sample in the morning using a FECOTAINER collection device. Full local ethical approval was obtained (Research Ethics Committee reference 14/EE/0024). Participant information is given in Table S1. Of the five stool samples, two were graded as Type 3, two as Type 5, and one Type 2; according to the Bristol Stool Scale\textsuperscript{23}. Within two hours of sample donation, faecal samples were separated into eleven equal aliquots and stored according to the study outline shown in Figure S1. In addition, 50 faecal samples were collected from patients undergoing bariatric surgery. Of these, 15 were collected from patients prior to undergoing surgery, 11 were collected 2-6 months post-surgery, and 24 were collected 1-2 years post-surgery. Patients were recruited from Charing Cross Hospital and Hammersmith Hospital (London, UK) under an ethics permission (Metabolic Phenotype and Modulation in Obese and Bariatric Surgical Patients 08/H0711/123).

REIMS Analysis of Faecal Samples

REIMS analysis was completed on all samples as previously described for the high-throughput, semi-automated analysis of microbial cultures on solid growth media\textsuperscript{21}. In brief, faecal samples were placed onto the base of a sterile petri dish on a Pickolo visualisation platform, which also acted as the return electrode for REIMS monopolar analysis, connected to a modified TECAN Freedom EVO One liquid handling platform\textsuperscript{21, 22}. For each analysis, an area of approximately 1 mm\textsuperscript{2} is analysed with the monopolar probe penetrating approximately 0.2 mm into the sample. The glass surface of the visualisation platform was coated with a conductive material that allowed the transfer of the electrical current through capacitive coupling. A radiofrequency electrical current was passed through a stainless steel monopolar probe with a sharp tip and the analyte vapour channelled to a Xevo G2-XS QToF mass spectrometer via PTFE tubing. Prior to entry into the mass spectrometer, the analyte
vapour was mixed with 2-propanol containing leucine encephalin at a concentration of 10 ng/mL, at a flow rate of 0.2 mL per minute, in a T-piece unit. The resulting mixture passed through the REIMS atmospheric interface chamber, as previously detailed, where solvent-ion clusters collide with a heated (700-800°C) coil before entry into the Stepwave® of the MS instrument. For each sample and/or analysis parameter, a total of five analytical repeats were acquired. For each analytical repeat, a single stainless steel monopolar probe was used. Probes were re-used after removal of any particulate matter and vortex cleaning in 2-propanol. Mass spectra were collected in negative ion and positive ion detection modes using instrument parameters in Table S2 and over the m/z range of 50 to 2500. For each day of operation, the mass spectrometer was calibrated using sodium formate according to the manufacturer’s instructions. For instrument cleaning, the disassembled T-piece unit was cleaned at the end of each running day through sonication for 10 minutes in 2-propanol. The standard laboratory routine for cleaning of the instrument was followed whereby the Stepwave® is removed after venting and sonicate cleaned in 50% methanol (v/v) made up using water for ten minutes and then in 100% methanol for a further ten minutes.

Analysis of Mass Spectrometry Data and Statistical Analysis
To correct for mass drift and remove background mass spectral signals, all .RAW data files were processed using MassLynx software (V4.1, Waters) through the Accurate Mass Measure tool employing the Automatic Peak Detection option, background subtraction, and with lock mass correction against leucine enkephalin (554.2615 for negative ion mode data and 556.2712 for positive ion mode data) with a mass window of +/- 0.500, and averaged over a total of ten scans. Resulting processed centroid mass spectra were used in subsequent intensity analysis calculations, identification of most intense spectral peaks, and m/z values for accurate mass identification of metabolites. For intensity analysis, mean and standard deviations were calculated from individual intensity values taken for five identified analytical repeats identified through total ion count chromatogram peaks.
Additionally, raw mass spectral data files were processed using the Offline Model Building (OMB)
software (Version 1.1.29.0, Waters) in order to perform background subtraction, mass drift correction
against leucine encephalin lock mass compound (negative mode m/z = 554.3), and mass binning to
0.1 Da within chosen restricted mass ranges. For statistical analysis, including principal component
analysis (PCA), hierarchical cluster analysis (HCA), and univariate statistical analysis, the
MetaboAnalyst 3.0²⁴ platform was used. After processing of mass spectra, a data matrix was exported
from the OMB software and uploaded to the MetaboAnalyst 3.0 online analysis pipeline where it
underwent additional processing for log transformation and Pareto scaling (mean-centred and divided
by the square root of standard deviation of each variable). For univariate statistical analysis, the non-
parametric Kruskal-Wallis test was used with a significance threshold, after false discovery rate
correction, of less than 0.05 used to identify significantly different mass bins. For PCA analysis, the
first and second principal components were used in two dimensional plots, with 95% confidence
regions displayed using colour shading. For HCA analysis, dendograms with heat maps were
constructed using the top 25 mass bins based on univariate statistical analysis and constructed using
Euclidean distance measure, and Ward clustering algorithm.

**Tentative Identification of Highest Intensity Phospholipids and Bile Acids**

Metabolite identification of mass spectral peaks was accomplished through two approaches. For the
highest intensity phospholipids within the 600 to 1000 m/z range, collision-induced dissociation
tandem mass spectrometry was conducted using the Xevo G2-XS QToF instrument. The peaks were
chosen from across the five initial samples used for heating power optimisation, using spectra
obtained using the optimised heating power of 35W. The five participants’ sample aliquot stored as
shown in Figure S1 were used in tandem mass spectrometry. Before fragmentation of the target m/z
peak, the resolution of the quadrupole was optimised using the constant signal of the leucine
enkephalin lock mass compound infused with 2-propanol. After optimisation to ensure that only one
mass spectral peak was visible after the quadrupole instrument was set to target 554.26 (in negative ion detection mode), each of the targeted phospholipid m/z values were in turn fragmented. The collision energy of the Xevo G2-XS QToF instrument was optimised for each targeted m/z peak until the parent ion had a relative intensity value of between one third and one half of the product ion with the highest relative intensity. The resulting fragmentation patterns were used to assign identifications according to head group and acyl chain combinations.

For tentative identification of metabolites of interest, such as bile acids (primary, secondary, and conjugates), fatty acids, low molecular weight metabolites, and phospholipids, accurate mass measures were used. A number of negative ionisation m/z values were identified in the acquired REIMS mass spectra. Accurate masses of candidate m/z peaks, to two decimal places after mass drift correction, were used to interrogate the Human Metabolome Database\textsuperscript{25} and the LIPI D MAPS database\textsuperscript{26}, with the highest ranked match, based on Delta value (< 0.01), after removal of specific drug- and diet-related metabolites, taken as the tentative identification shown in Tables S3 and S4 and Supplementary Data Matrix 1.

\textbf{Safety Considerations}

In this work, all samples were treated as potential Hazard Group 2 material and were therefore manipulated within containment level 2 facilities. All REIMS analysis was handled within a class 2 biological safety cabinet. All solvents, such as 2-propanol and methanol, were handled according to the material safety data sheet provided by their respective manufacturer.
Results and Discussion

Faecal samples were successfully collected from five male participants and analysed using an automated REIMS monopolar electrode platform, Figure 1. The automated set-up allowed sample analysis without pre-treatment nor extraction in approximately ten seconds. As five replicate sampling points were taken, this allowed sample analysis to be completed in approximately one minute. Initial inspection of the resulting mass spectra shows a dominance of fatty acids including saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids above 13 carbons in length. In the higher mass range, between 600 and 1200 m/z, there is a lower intensity of complex lipids, typically detected using REIMS, including glycerophospholipids and diglycerides.

Optimisation of Heating Power and Analytical Set-Up

Signal intensities for each heating power within the fatty acid (50 to 500 m/z), and complex lipid (600 to 1000 m/z) regions typically seen in REIMS mass spectra, alongside a measure of instrument noise (50 to 51 m/z), are given in Figure 2a. Unsupervised principal component analysis (PCA) was completed using sample classes according to heating power, Figure 2b, and participant, Figure 2c. Within both PCA plots, high variability is seen between mass spectra acquired using a heating power of 25W or lower and 40W or higher. However, such a trend is not visible with spectra acquired using a heating power of either 30W or 35W. A similar trend is evident within signal intensities as was observed with PCA modelling. Mean signal intensities for heating powers below 30W show a trend towards decreased fatty acid and complex lipid region intensities, and 40W and above a trend towards increased fatty acid, but decreased complex lipid region signals, with an overall positive correlation between spectral noise and heating power observed. As a result of these evident trends, an optimum heating power of 35W was chosen as a universal parameter suitable for the analysis of all stool types. Following heating power optimisation, a total of 169 features were given tentative identifications, across the entire mass spectral range analysed for all five participants. This figure however, represents
only a fraction of the total features detected in the REIMS mass spectra. Representative mass spectra for analysis of all five participant samples, after background subtraction and mass drift correction, are given in Supplementary Figures S2 and S3. To determine the requirement for analytical replicates of a sample, the heterogeneity present within samples was explored for REIMS analysis. As traditional mass spectrometry approaches require sample preparation and extraction, homogenisation of the sample is typical. With REIMS analysis however, individual analytical sampling points across the sample are used. Supplementary Figure S4 shows PCA modelling of the five analytical points for the optimised 35W heating power. There appears to be minimal heterogeneity across the five analytical repeats, particularly within the 600 to 1000 m/z range, although the analytical repeats from one participant’s sample show a lower degree of clustering. Heterogeneity within a faecal sample is not unexpected and therefore, the composite mass spectra of five analytical repeats from different spatial regions of each sample were used for subsequent REIMS analysis.

**Comparison of Negative and Positive Ion Detection Mode REIMS Mass Spectra**

Initial optimisation of faecal analysis using REIMS was completed in negative ion detection mode as previous studies have shown this mode provides the richest mass spectra with regards to both signal intensity and diversity of detectable ion species\(^{19,27}\). Positive ion detection mode REIMS data was also collected on the five samples used in heating power optimisation using 35W. A representative mass spectrum is shown in Figure 1c alongside mass spectra for all five participants in Supplementary Figure S3. Principal component analysis of positive ion detection mode, Supplementary Figure S5, shows that participant differences are not observable in either the 50 to 500 m/z nor 600 to 1000 m/z mass ranges, unlike those which are evident in negative ion detection mode, Figures 3 and 4. From visual inspection of mass spectra obtained in both ion detection modes, it appears that negative ion spectra contain a high level of complex lipids, typically found within the 600 to 800 m/z range, whilst positive ion spectra contain likely complex lipid features across a wider mass range of 600 to 1000 m/z. At the lower mass range, negative mode spectra have easily distinguished (un/mono/poly-saturated) fatty


acid ions which are not easily distinguishable in positive ion spectra due to the higher level of analytical noise. Analysis of mass spectra in both ion detection modes indicates higher levels of analytical noise are present within the positive mode spectra which may explain the inability to separate participants in PCA. Thus, for further analysis of faecal samples, negative ion detection mass spectra were used.

**Effect of Sample Storage Conditions on REIMS Mass Spectra**

Following optimisation of the sample heating power, the effect of storage duration and condition on the mass spectra generated by REIMS was determined. REIMS analysis was completed on samples after normalisation to room temperature for one hour using the previously identified optimum universal heating power of 35W. Unsupervised PCA analysis was completed on the eight-different storage length and condition combinations and compared to mass spectra obtained from fresh faecal samples. Figure 3a shows that overall, no significant differences are evident as a result of any storage condition and time duration combination. Indeed, hierarchical cluster analysis based on the top 25 mass bins, ranked by the $P$ value of analysis of variance tests between participants, shows that participant differences are clear across all sample storage condition and duration combinations, Figure 3b. The contributing mass bins to this clear separation are primarily within the REIMS complex lipid region (600 to 1000 $m/z$). Analysis of signal intensities between storage conditions as compared to fresh faecal samples, Supplementary Figure S6, show that for all conditions, fatty acid and low molecular weight metabolites regions are significantly higher than that of the typically complex lipid region. In regard to the effect of storage conditions on signal intensities, it is clear that the complex lipid region is not significantly different. However, within the 50 to 500 $m/z$ range (typically fatty acids and lower molecular weight metabolites) there is a significant increase for all storage conditions at 72 hours, with the largest increase seen at room temperature. This increase is in line with previous work on the metabolic profiling, albeit using 1H nuclear magnetic resonance spectroscopy, of human faecal samples which found that storage at room temperature substantially affected fatty acids and low
molecular weight metabolites, but that storage at 4°C or below reduced this effect\textsuperscript{28} - likely as a result of reduced microbial metabolism of complex lipids.

\textbf{Effect of Freeze/Thaw Cycles on REIMS Mass Spectra}

In addition to storage condition and length, the effect of consecutive freeze/thaw cycles on the REIMS mass spectra was assessed. For this assessment, one portion of participants' faecal samples was frozen at -80°C for two hours, thawed at room temperature for one hour and analysed using the high-throughput, monopolar REIMS platform as previously described. For each sample, a total of six freeze/thaw cycles, with REIMS analysis after each, was completed. Unsupervised PCA modelling shows a clear trend towards separation of samples, from all participants, according to freeze/thaw cycle, Figure 3c. Univariate, non-parametric statistical analysis using Kruskal-Wallis one-way analysis of variance identified all significant, FDR-corrected $P$ value below 0.001, mass bins, between 50 to 2500 m/z, different between freeze/thaw cycles, Supplementary Figure S7. The majority of significantly different m/z bins are below 300 m/z or above 1100 m/z suggesting that molecular species such as fatty acids, and dimers of diglyceride and glycerophospholipid which are of high intensity within these respective ranges are more susceptible to freeze/thaw-induced breakdown. The stability of the 600 to 1000 m/z range was confirmed through unsupervised PCA modelling, Figure 3d, which shows no significant separation of samples based on freeze/thaw cycles. Additionally, intensity analysis shows significant increases in both noise (50 to 51 m/z) and the 50 to 500 m/z range, particularly after three of more freeze/thaw cycles. Conversely, the complex lipid region of 600 to 1000 m/z range shows no significant differences in signal intensity of any freeze/thaw cycle compared to fresh faecal samples. Unsupervised PCA modelling using this restricted mass range, Figure 4a, shows that clear and significant separation is evident between participants within this mass range, using mass spectra from all six freeze/thaw cycles; with reduced separation evident between participants using the 50 to 500 m/z range evident. This difference in separation may be indicative of two factors. Firstly, it may suggest that individual differences between participants are found within the biochemical
information from the complex lipid region (600 to 1000 m/z) rather than the fatty acid and lower molecular weight metabolite region (50 to 500 m/z). Arguably, the complex lipid region holds more scope for biochemical differences to emerge as a result of the wide range of detectable combinations of head groups and acyl chains. Secondly, it may support the conclusions drawn from Figure 3 that the complex lipid region is more stable through successive freeze/thaw cycles than the lower mass range and is therefore, better able to preserve individual differences that are detectable through REIMS. This would further support the benefit of using REIMS for direct-from-sample metabolomics of faeces over traditional mass spectrometry approaches as it does not require the storage and bulk processing of samples that is common in techniques such as LC-MS and GC-MS.

Identification of High-Intensity Phospholipids, Bile Acids, and Other Metabolites

In addition to the identification of biological differences between participants’ faecal samples, and the assessment of signal stability as a result of storage conditions, metabolites commonly targeted using hyphenated mass spectrometry methods were detected. Here, we have tentatively identified a range of bile acids, including primary, secondary, and conjugate species, using accurate mass database matches; as shown in the annotated spectra of a selected participant in Figure 5, and Table S3. Bile acids are metabolised by the gut microbiota and have been linked to a range of gastrointestinal disorders and diseases, including inflammatory bowel disease\(^29\) and liver cirrhosis\(^30\). In this study, we have shown for the first time that REIMS can be used for the direct detection of liver-produced bile acids, including chenodeoxycholic acid and taurodeoxycholic acid, and products derived through their bacterial metabolism such as lithocholic acid. Bile acids which are metabolised by the gut microbiota have been linked to a range of gastrointestinal disorders and diseases, including inflammatory bowel disease\(^29\) and liver cirrhosis\(^30\). Currently, HPLC-MS is the main method used to assess bile acids from faecal samples. Due to the lengthy duration of analysis, and the requirement for highly-skilled users, such an approach is unlikely to find utility in a clinical environment. In contrast, our high-throughput, automated REIMS platform described herein could enable timely bile acid profiling and thus aid clinical
decisions. In addition to bacterial-derived metabolites of bile acids, we have also identified a number of bacteria-associated metabolites in faecal samples including cresol, butanal, mandelic acid, and urocanic acid.

In addition to bile acids, phospholipids, which constituted a large proportion of the REIMS mass spectra generated, were identified directly from faecal samples using tandem mass spectrometry, Table 1. The dominant phospholipid class identified using this method, phosphatidylglycerol (PG), is the main lipid component, alongside phosphatidylethanolamine (PE), of the bacterial inner cell membrane. Using REIMS, we have previously shown that these lipid classes hold substantial species-level discriminatory power between clinically important microorganisms\textsuperscript{21,22} and could thus serve as biomarkers for constituents of the faecal microbiome. Supporting this, three of the four identified PG lipids contained odd-chain, 15:0, acyl chains, which are indicative of bacterial lipids\textsuperscript{31}, although, it should be noted that odd-chain acyl chains and free fatty acids are also associated with ruminant metabolism\textsuperscript{32}, and thus may be indicative of dietary intake. Nonetheless, REIMS could serve as a tool for the rapid identification of pathobiont species directly from sample. In addition, REIMS may also allow for a rapid and non-invasive method for the screening of faecal samples to monitor shifts in metabolic content which may be diagnostic of disease status and/or onset. In addition to those features identified using tandem mass spectrometry, a total of 169 features were tentatively identified, Supplementary Data Matrix 1, across a wide range of classes, using accurate mass measurements. This data shows the wide range of compounds which can be detected using REIMS, including amino acids and peptides, in human faeces.

### Application of Optimised REIMS Parameters to Clinical Faecal Samples

Bariatric surgery is performed on patients with severe obesity to bring about profound weight loss, and can lead to improvements in comorbidities such as type two diabetes mellitus, hypertension, and obstructive sleep apnoea\textsuperscript{33}. Although the marked effect of bariatric surgery on the human gut
microbiome has been extensively studied\textsuperscript{34, 35}, there is limited understanding of how the faecal metabolome is altered. NMR-based spectroscopy of serum shows that the wider human metabolome is affected by bariatric surgery\textsuperscript{36}, suggesting that concomitant changes to the faecal metabolome are likely and these should be detectable using REIMS.

To show that REIMS is capable of detecting biological differences using the optimised parameters identified herein, a cohort of 50 faecal samples from patients pre-, two to six months post-, and one to two years post-bariatric surgery were analysed. REIMS analysis was carried out on these samples, and following data pre-processing as previously described, the 50 to 1200 m/z range at 0.1 Da bins was used for statistical analysis. From both unsupervised and supervised multivariate modelling, Figure 6, a clear separation is evident between the pre- and one to two years post-surgery faecal samples, with those collected one to six months post-surgery showing an intermediary distribution between the two time points. Analysis of the features contributing towards this observed separation, Supplementary Figure S7, show that the primary drivers are present within the 600 to 1200 m/z range, with a high number of important features between 1000 and 1200 m/z. Furthermore, univariate analysis of variance tests also shows that the most significantly different features between the three time points are above 600 m/z, albeit with a substantial number below this cut-off. Supplementary Table S4 shows tentative identifications of significant features driving the observed separation in Figure 6. Based on accurate mass determination, these tentative identifications support the observation that there are significant changes in faecal lipid composition, namely phospholipids including PC(44:3), PE(46:0), PC(42:0), PE(P-38:0), and PE(46:1) and triglycerides TG(54:3), TG(58:0), and TG(52:4) as a result of bariatric surgery. Unsurprisingly, alterations in lipid concentrations following bariatric surgery, as a result of reduced dietary intake, are common\textsuperscript{37, 38}. Indeed, dyslipidaemia, a condition linked to cardiovascular disease and stroke, resulting in abnormal composition of lipids within the body, is often resolved following bariatric surgery\textsuperscript{39}. Thus, REIMS is
capable of detecting important biological changes following a clinical intervention - proving its utility

as a rapid, direct-from-sample metabolic profiling platform for human faeces.
Conclusions

As an alternative to traditional metabolic profiling approaches which have linked mass spectrometry with a pre-analysis chromatographic separation, REIMS offers a platform for the high-throughput, automated, and low-cost analysis of biological material. As well as the benefit of faster analysis, resulting from the removal of the requirement for sample preparation, REIMS allows the analysis of a sample, such as faeces, in its natural state. These properties present a number of novel application areas in which this proof-of-concept work could be extended. For example, the biochemical analysis of faeces, including metabolite profiling, has been suggested as a potential approach for either whole or targeted population disease screening programmes, such as for colorectal cancer. Due to its high-throughput, automated, and low-cost nature, REIMS may provide a practical alternative to existing methods such as Cologuard\(^4\), which can be time-consuming and costly. This ability would allow for the screening of a large number of patient samples, and would not be limited to one, or a low number of target compounds. Although this work utilised the REIMS analysis of human faeces, a wide-range of biofluids could also be analysed using the technology, such as sputum or urine, allowing a single platform to be used for the screening of a diverse range of diseases e.g. cystic fibrosis and urinary tract infections. One of the main constituents of human faeces is microbial, and although the gut microbiome is an essential component of human physiology, pathogenic microorganisms, such as *Clostridiodes difficile*, are capable of causing disease. Current techniques require culturing of a sample to identify the causative organism of an infection, which is labour intensive and time-consuming. As we have suggested, REIMS may be capable of detecting bacterial phospholipids, which might provide a novel diagnostic method that allows direct-from-sample detection of pathogens. This would however, require additional work to determine the level of taxonomic specificity which could be achieved.
Supporting Information

Detailed information regarding bacterial and fungal species analysed in this study, and on additional data analyses completed as referenced within this text is given within the supporting information accompanying this manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

Author Contributions

Study was planned by SJSC, JLA, JVL, and ZT. Experiments were conducted by SJSC, JA, HA, and JVL. Data was analysed and interpreted by SJSC, JLA, and ZT. Technical assistance was provided by FB, AB, JT, JRM, and JK. The manuscript was written by SJSC, JLA, and ZT with input from all authors. All authors have given approval to the final version of the manuscript.

Conflict of Interest Statement

This work was funded and technically supported by the Waters Corporation and funded by the Biotechnology and Biological Sciences Research Council under grant BB/L020858/1 and European Research Council under contract number 617896. Z.T. provides remunerated consultancy to the Waters Corporation. The work detailed in this manuscript does not promote any available commercial product from Waters Corporation.

Acknowledgments

The authors would like to thank Imperial College Healthcare Trust for access to samples and to the participants/patients who agreed to donate samples which were utilised in this work. We also acknowledge support from Imperial College NIHR Biomedical Research Centre and Bowel and Cancer Research. This work was funded and technically supported by the Waters Corporation and funded by the Biotechnology and Biological Sciences Research Council under grant BB/L020858/1 and European Research Council under contract number 617896. This article is independent research funded by the
NIHR BRC, and the views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR, or the Department of Health.
References


Phospholipids identified through interrogation of LIPID MAPS database for possible species for highest intensity m/z bins, followed by tandem mass spectrometry fragmentation with accurate mass identification of potential acyl chains.

<table>
<thead>
<tr>
<th>Lipid Candidate</th>
<th>MS/MS Fatty Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z Bin</td>
<td>ID</td>
</tr>
<tr>
<td>653.51</td>
<td>DG(39:6)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>691.49</td>
<td>PG(P-31:0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>717.51</td>
<td>PG(P-33:1)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>731.52</td>
<td>PG(O-34:1)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. REIMS Set-Up for Preparation-Free Faecal Analysis and Representative Mass Spectra

A modified TECAN Freedom Evoware 75 robotic platform was used for automated REIMS analysis of faecal samples. For analysis (a) a modified robotic tip with a sharp monopolar electrode makes contact with the faecal sample and generates a radiofrequency electrical current. Due to the non-zero impedance of the sample, this results in rapid heating of the sample and the generation of an aerosol containing gas phase ions. This aerosol is transferred directly to a mass spectrometer via an atmospheric interface where it combines with a solvent matrix and collides with a heated collision surface, resulting in the formation of individual ion species which enter the ion guide of the instrument. REIMS analysis allows for the detection of both (b) negative ion and (c) positive ion molecular species across the mass range of 50 to 2500 m/z. Mass spectra shown after background subtraction and lock mass correction using MassLynx software.
Figure 2. Optimisation of Heating Power

Heating power optimisation was completed using power levels between 10W and 50W, increasing in 5W increments. From both (a) signal intensity of three regions of interest and principal component analysis over the 50 to 2500 m/z range grouped by (b) heating power, and (c) participant shows an optimum value of 35W. Shaded colour areas represent 95% confidence intervals of significant separation. Error bars show one standard deviation around the mean.
Figure 3. Effect of Storage Condition and Freeze/Thaw Cycles on REIMS Mass Spectra

(a) Principal component analysis over the 50 to 2500 m/z range shows no significant separation of storage conditions, with (b) hierarchical cluster analysis shows participant differences using the 25 most significant mass bins regardless of storage conditions. PCA plots within the (c) 50 to 500 m/z range and (d) 600 to 1000 m/z range show significant effects of successive freeze/thaw cycles within the lower molecular weight region, but not within the complex lipid region. Shaded colour areas represent 95% confidence intervals of significant separation. Intensity analysis and univariate analysis is given in Supplementary Figures S6 and S7.
Figure 4. Maintenance of Participant Separation Following Successive Freeze/Thaw Cycles

Unsupervised PCA plots of participant samples after six freeze/thaw cycles using the restricted mass ranges of (a) 600 to 1000 m/z and (b) 50 to 500 m/z. Shaded areas indicate 95% confidence range of groupings. Cycle Key: ▲ = 1; ■ = 2; ◆ = 3; ● = 4; ✤ = 5; ▼ = 6;
Figure 5. Identification of Bile Acids and High Intensity Phospholipids in REIMS Mass Spectra

Annotated mass spectra for one participant of (a) bile acids tentatively identified through accurate mass interrogation of reference databases, and of (b) high intensity phospholipids identified through tandem mass spectrometry fragmentation and subsequent accurate mass identification of fatty acid tails. Metabolite identifications outside of these displayed mass ranges are given in Supplementary Data Matrix 1.
Figure 6. Multivariate Modelling of Bariatric Faecal Samples

Modelling of the 50 to 1200 m/z range using (a) principal component modelling and (b) partial least square discriminant analysis shows that there is a significant separation between time points. Shaded areas show 95% confidence intervals of each grouping. Performance scores for PLS-DA model in (b) using ten-fold cross-validation, are Accuracy = 0.70; R2 = 0.80; Q2 = 0.55. PCA loading and importance feature plots are show in Supplementary Figure S8. Univariate analysis using (c) one-way Kruskal-Wallis shows location of significantly different features between the three groups using an FDR corrected P value threshold of less than 0.05. Tentative identification of compounds driving separation in Supplementary Table S4.
The first application of rapid evaporative ionisation mass spectrometry, requiring no sample preparation, to the analysis of the human faecal metabolome.