Title
A Metabolomics Approach to Profiling Chemicals of the Human Intestinal Tract

Author
Folz, Jacob Sylvester

Publication Date
2021

Peer reviewed|Thesis/dissertation

A Metabolomics Approach to Profiling Chemicals of the Human Intestinal Tract

By

JACOB SYLVESTER FOLZ
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Food Science and Technology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Prof. Oliver Fiehn, Chair

_____

Prof. Carolyn Slupsky

_____

Prof. Alyson Mitchell

Committee in Charge

2021

Acknowledgements

I first and foremost thank Professor Oliver Fiehn who supported, instructed, and showed me how to perform science at the highest level and made this thesis work possible. I am grateful for the help and guidance of members of the Fiehn laboratory and graduate committee members who allowed me to earn my graduate degree. I cannot begin to express my thanks to my parents and grandparents who gave me the opportunity and freedom to pursue an education in science and have supported me at every step throughout my journey of education and life. I thank my siblings who support me unconditionally and to the many friends who have made this educational journey awesome. I thank all my teachers and professors who inspired me to think in a critical, scientific manner. I would like to specifically thank Mikki Brettingen for educational inspiration and Dr. Philipp Wiemann who convinced me that pursuit of a PhD was right for me. I also extend my sincere thanks to Dr. Clayton Bloszies, Dr. Cory Marx, Dr. Uri Keshet, Dr. Brian DeFelice, Dr. Megan Showalter, Dr. Shen Tong, Juan Montes, Sophia Pinton, Dr. Tyler Simons, Dr. Honglian Ye, Dr. Su Ling Zeng, Dr. Dari Shalon, Dr. Arpana Vaniya, Dr. Tobias Kind, Dr. Tomas Cajka, Luis Valdiviez, Kelly Paglia, and Jeannette Martins for emotional and/or scientific support throughout my graduate studies.

Thesis Abstract

The use of mass spectrometry-based metabolomics analysis is a powerful technique to provide a comprehensive measurement of the chemical profile of biological samples. By observing the chemical composition of biological samples, we can better understand biological systems. All biological samples are chemically complex, but the human gastrointestinal tract is one of the most chemically complex matrices currently studied. This dissertation focuses on the approach, execution, and results of mass spectrometry-based metabolomics analysis of biological samples, with a focus on human gastrointestinal tract samples and approaches to better understand of the metabolome of the human gut. This research tests the hypothesis that the chemical profile of the human intestinal tract varies spatially, temporally and has high interindividual variation.

Chapter one provides background information on mass spectrometry-based metabolomics methods. Experimental data are provided to emphasize important factors to consider when performing technically sound metabolomics analysis on different biological matrices. Bacteria, plant, blood, and stool experiments are reported with select publications included and methodological considerations for each sample type. Challenges in analysis presented by each matrix type are outlined and approaches to address these challenges are discussed with a focus on metabolite identification.

Chapter two presents a metabolomics analysis of time-series human small intestinal samples. Samples were collected from the human jejunum over the course of one day and analyzed using four LC-MS/MS analytical platforms. A total of 828 metabolites were annotated within these samples and correlation-based clustering proved to be an efficient

approach to connect metabolites with related biological function. This experiment reveals novel chemical and time-dynamic findings from within a human small intestine.

Chapter three discusses data collected during targeted bile acid analyses. First bile acid analysis of blood and stool samples from a flaxseed dietary intervention study are reported, and then a study using a novel sampling method which allows samples to be retrieved from the small and large intestine of humans *in vivo.* Samples collected from the intestinal tract of 15 subjects reveal spatial variation and inter-individual variation in bile acid metabolism within the human gastrointestinal tract.

Thesis Introduction

The human intestinal tract is a complex and dynamic system that performs functions essential for human life. The chemical profile of the human intestinal tract is understudied as it is a difficult region to sample *in vivo*. The question of what chemicals exist in the intestinal tract, and how these chemicals change over time and by intestinal tract region is addressed through my research. I hypothesized that chemical profiles are distinct between different intestinal regions and show high interindividual variation. This thesis work includes state of the art chemical analysis applied to investigating the entirety of small molecules in the human intestinal tract. The process of measuring all chemicals within samples allows discovery of unexpected biological trends that are otherwise missed if traditional targeted chemical analysis is performed. The field of comprehensive chemical profiling of biological samples is termed metabolomics. In this thesis the metabolome of interest is the complex chemical ensemble that travels through the human intestinal tract. I measured chemicals derived from food, human, and microbes from

samples of human intestinal tracts using mass spectrometry to reveal trends occurring within humans *in-vivo* that have never before been observed. Findings reported here contribute to the cumulative knowledge of the composition, temporal, and spatial chemical profile of the human intestinal tract.

Metabolomics as a field has evolved over the past two decades. Improvements have been made in performing accurate reproducible comprehensive chemical analysis for complex samples. Key technological advances in hardware and software have allowed for improved metabolomics analyses. Introduction of the orbitrap mass analyzer enables rapid collection of highly accurate mass spectra[1] and ultra-high performance liquid chromatography (UHPLC) provides fast and reproducible separation of chemical mixtures[2]. Software have improved alongside the analytical instrumentation to pull chemical information from these data in a progressively more automated manner[3–5]. Metabolomics methods are not standardized across the field and the optimal approach to measure the metabolome of samples is continuously changing and improving. In chapter one of this thesis work, examples are presented on how to best perform metabolomics analysis of bacteria, plants, serum, and feces. These methods and applications to different matrix types were compiled into methodological considerations to reliably measure the metabolome of human intestinal tract samples. Metabolomics workflows contain many steps. However, one step common to all comprehensive metabolomics analyses is metabolite identification. Metabolite annotation and structure identification is the key step of determining which signals from a mass spectrometer can reliably refer to specific known chemicals. This step requires careful consideration of data to match a signal to a known chemical. Yet, the individual considerations in compound identifications

are frequently poorly reported in metabolomics publications. Metabolite identification with respect to samples from human intestinal contents is essential to ensure that correct biological conclusions are made. Metabolite identification in MS based metabolomics workflows is thoroughly discussed at the end of chapter one.

Chapters two and three focus on the analyses of human gut intestinal tract samples. Food and diet shape the chemical profile of the intestinal tract and cause effects throughout the entire human body. The human gut is an essential component for human life. Energy is absorbed as energy rich chemicals from the diet which fuel the body and essential nutrients are taken up to maintain health. Microbes in the gut increase energy absorption efficiency[6], produce vitamins[7], and are intertwined in nearly all aspects of human health. Research efforts applied to the gut microbiota are continuously finding new facets of gut microbiota interaction with the human host. The function of gut microbes in the human GI tract *in vivo* has been investigated using human stool, and samples retrieved from the gut through invasive sampling methods that perturb the intestinal tract from its normal condition. During this thesis, two novel sampling methods were employed that retrieved samples from human intestinal tracts *in vivo* in a non-invasive manner. These studies gave novel insights into normal functioning of the human GI tract. The first study investigated a single subject sampled from the small intestine many times over the course of one day (chapter two), and the second study presents samples collected from the entirety of the small intestine of 15 subjects (chapter three). These studies revealed trends that were previously impossible to observe in live human subjects.

While this thesis focused on the analysis of upper GI tract samples, other biofluids are also regularly used as surrogates to give insight into gastrointestinal tract function. First,

blood chemicals are directly altered in response to specific foods through absorption into the bloodstream, or through interaction of dietary compounds with cells of different organs throughout the body. Second, human and animal stool samples are easily collected in non-invasive manner. Stool is used to gain insight into the function of the gastrointestinal tract by presenting all the waste and non-absorbed components after digestion. Here, I present metabolomics results of a dietary intervention study of 178 human subjects for who both serum and stool samples were collected. Samples for this study were collected before and after a period of six weeks of flaxseed consumption with no other dietary guidance. Metabolomics analysis of these samples led to biologically interesting findings. However, the originally hypothesized effect from flaxseed consumption was negated, as findings indicated no change in the proportion of bile acids in either stool or blood. This study and its details are reported in chapter one and three.

For this reason, I focused more on bile acid conversions in studies comparing the small and large intestines. Bile acids are functional molecules synthesized in the liver and secreted into the beginning of the small intestine as a component of bile. Bile serves roles in digestion, as a microbial regulator, and as a route for excreting waste products from the body. Here, I present novel findings in bile excretion and chemical composition in a study of the small intestine of a single subject sampled multiple times over eight hours, with carefully registered food and drink events. This experiment and its findings are reported in chapter two. Bile acids are transformed extensively by gut microbes and are reabsorbed back into circulation. Microbially modified bile acids are found in circulation and act as signaling molecules by binding human protein receptors making this chemical class an axis of communication between the gut microbiota and human body[8]. Bile acids

in the human and gut are regulated by a complex interaction of human, microbial and dietary signals. The small intestine represents the primary location where bile acids are in contact with gut microbes. Since the small intestine is difficult to study without using invasive procedures, a gap exists in the knowledge of the profile of bile acids across the small intestine of humans under normal conditions. Importantly, the thesis presented here now addresses this gap through bile acid analysis of samples collected throughout the small intestine and stool of 15 healthy volunteers. Trends were observed that have never been previously measured in humans *in vivo* and are presented in chapter three.

Metabolomics analysis reveals a snapshot of all the chemicals participating in metabolism for biological samples. This snapshot represents what is truly taking place at a specific time as opposed to other methods such as sequencing the genome, transcriptome, or proteome. These sequencing approaches provide information on the community and functional potential of a system, but not the actual properties of the environment itself. The chemicals from the environment must be measured to know what is *actually* occurring at the chemical level. In this way metabolomics analysis provides valuable information from the human intestinal tract by revealing the chemicals that may be absorbed, chemicals that may interact positively or negatively with intestinal epithelium, or chemicals involved in metabolism occurring within the intestinal tract. This thesis presents information to improve metabolomics analysis of human intestinal tract samples and reports novel observations from the metabolome of human intestinal tracts.

Chapter 1: Comprehensive chemical profiling of biological matrices using mass spectrometry

1.1 Abstract

Different biological matrices have dramatically different composition of macromolecules, and small metabolites. The ideal approach to profile chemicals using mass spectrometry depends on matrix type and must be considered when performing comprehensive metabolomics analyses. Using an incorrect workflow can lead to poor metabolome coverage, incorrect metabolite identifications, failure to detect target chemicals, or incorrect biological findings due to artifacts of measurement. Several matrix types will be discussed including blood plasma, serum, plant tissue, animal tissue, microbial cultures, and human stool. Sample preparation, data acquisition, data processing, and biological findings will be covered for these sample matrices, and important considerations that apply to analysis of gastrointestinal tract samples will be highlighted.

1.2 Overview of metabolomic analysis of biological samples using mass spectrometry

Biological samples are chemically complex and measuring the metabolome of any biological sample is a formidable task. The metabolome is considered here as the entirety of chemicals less than 1500 Daltons participating in a biological process within a system. The molecular mass of 1500 Daltons is used as a typical cutoff as it encompasses small metabolites while excluding macromolecules such as proteins and DNA[9]. The number of chemicals in biological samples is difficult to estimate from known biochemical pathways, as enzymes can be promiscuous and spontaneous chemical reactions also occur within

cells[10]. One approach is to use genomic reconstruction to identify all possible chemical pathways within an organism. However, this approach cannot elucidate the entirety of the metabolome since there are genes of unknown function in all living organisms, including the best-studied model organism, *Escherichia coli,* which has most of its genes linked to known functions[9,11]. The best way to measure the metabolome is to use analytical chemistry approaches. The most advanced and comprehensive chemical analyses use mass spectrometry (MS) and nuclear magnetic resonance (NMR). Both approaches have advantages and disadvantages. Advantages of NMR include high reproducibility, its quantitative nature, and its ability to measure concentrations in a large dynamic range, while MS has high sensitivity, low sample volume requirement, and ability to couple to many different chemical separation technologies (*e.g.,* gas chromatography, liquid chromatography, capillary electrophoresis, direct infusion, matrix assisted laser desorption). Both are fit for high throughput metabolomics analysis. Using either of these approaches to measure the metabolome of biological samples can link chemical signatures from data to known chemicals; unfortunately, a majority of signals from comprehensive metabolomics analyses can only be reported as unknown metabolites[12]. Comprehensive MS approaches were used for this thesis work.

Measuring samples using MS can be carried out by directly infusing samples into an MS instrument[13,14], which results in a chemical mass to charge (*m/z)* ratio being measured, and optionally chemical fragmentation (MS/MS) spectra being collected. This approach can be used to rapidly measure many samples, however with only *m/z* and MS/MS information, there are many isomers that cannot be distinguished, reducing the ability to identify specific chemicals. It is possible to address this hurdle by separating chemicals

before analyzing them using MS. This separation can be carried out using a chromatography method before MS which adds an additional dimension (retention time) to data collection. Using chromatography dramatically improves MS analysis since chemicals are separated in time leading to improved sensitivity, and the capacity to distinguish between more signals from chemicals of the metabolome. Liquid chromatography mass spectrometry (LC-MS) and gas chromatography mass spectrometry (GC-MS) are commonly used chromatography techniques. LC coupled with electrospray ionization (ESI) and mass spectrometry has become a widely used method for chemical profiling due to its reliable separation, broad range of chemical class coverage (dependent on the LC column), and low limits of detection. Tandem MS (MS/MS) spectra are acquired through fragmentation of one ion of interest and improves chemical identification through matching experimental MS/MS measured in samples to library MS/MS measured from authentic chemical standards. Annotation of metabolites using retention time, *m/z*, and MS/MS matching allows annotation of hundreds of metabolites within biological samples per LC-MS analysis. Many signals detected with MS cannot be confidently annotated as known chemicals resulting in hundreds to thousands of unknown features being detected during non-targeted analyses. When all this information is collected in a single analytical analysis for all chemicals, this is termed a non-targeted mass spectrometry approach. Information concerning metabolite identification in non-targeted metabolomics will be discussed thoroughly at the end of this chapter.

Each LC-MS analysis is best suited to analyze a certain category of metabolites. This is primarily based on the physical properties of the solid phase of the LC column. For

example, reverse phased (RP) columns are well suited to separate hydrophobic metabolites since hydrophobic metabolites interact with the column; hydrophilic interaction liquid chromatography (HILIC) is well adapted to separate hydrophilic metabolites. Another variable to consider using LC-MS is whether positive or negative ions will be measured. Some metabolites form only positive or negative ions, which leads to the necessity for acquiring data in both positive and negative ionization modes to maximize chemical coverage from a sample. Taking these variables into consideration, a combination of four different LC-MS analyses (RP-LC-ESI-MS/MS and HILIC-LC-ESI-MS/MS in positive and negative ESI modes) has the capacity to measure a large portion of the metabolome[15]. These analytical methods will be discussed in applications towards a variety of biological matrices. Each matrix has unique characteristics that require adaptation of these LC-MS/MS methods to maximize coverage and reproducibility in pursuit of accurately measuring the metabolome. There are valuable lessons to learn from analyzing metabolites from microbes, food, and humans individually. Non-targeted metabolomics profiling of microbes, food, and human samples individually will be reported, and factors important to consider when measuring a combination of all these matrixes simultaneously as in human intestinal tract contents will be emphasized.

1.3 Chemical profiling of pure microbial cultures

Microbes dominate the digestion of human food. Pure microbial cultures are widely used model organisms for a variety of applications. Pure cultures are made up of genetically identical single cell organisms which makes them a relatively simple biological matrix since there are relatively few enzymes present to perform chemical reactions. Two metabolomics analyses of pure microbial cultures will be discussed. The first is analysis

of a synthetic biological organism that contains the fewest possible genes required to maintain function as a replicating organism. The second example is a study of samples from the model organism *Escherichia coli*, and investigation of spontaneous chemical reactions occurring in samples of proteins that were digested to single amino acids.

1.3.1 Metabolomics analysis of the Minimal Genome Project

The most simple possible living organism maintaining the capacity to reproduce was constructed during the "minimal genome project"[16]. This microbial strain was designed from mycoplasmas (bacteria with the smallest known natural genome). Genes essential for cell replication were identified in multiple mycoplasma species, and these essential genes were synthesized into a fully synthetic genome. The genome was transplanted into a receptor vessel of a mycoplasma cell with all genetic material removed. The result is the smallest genome for a self-replicating cell. Through multiple rounds of editing and removal of non-essential genes, the strain JCVI-syn3.0 has a 531 kb genome, with 473 genes. Within this genome ~100 of the genes have loosely defined or unknown functions[17]. Metabolic pathways built based from genes of known function contain an estimated metabolome of 300 metabolites. These pathways encompass known biochemical pathways; however, enzymatic errors or spontaneous chemical reactions occurring to the known chemical pool were not considered. In an attempt to elucidate genes responsible for handling metabolites outside of the central chemical pathways, several gene knockout strains were constructed and grown. The buildup of chemicals was explored in these strains using a non-targeted metabolomics approach by analyzing the cell pellets of the minimal genome JCVI-syn3.0 wildtype strain, in addition to 8 gene knockout strains. In total, 170 metabolites were detected at higher abundance in the

bacterial cell pellet samples compared to blank samples of the growth media using a HILIC column for analysis.

Interestingly, the cell pellets analyzed in this study could not be seen with the naked eye. These small amounts of biological sample presented challenges as well as benefits during LC-MS/MS analysis. The challenge of achieving sufficient signal to measure the metabolites within cells was overcome by extracting all available sample, performing extra rinsing steps in tubes to improve metabolite recovery and concentrating sample in small volumes prior to analysis. The benefit of low metabolite concentration was very low technical variation in the analytical analysis across the study. Minimal ion suppression is the best explanation for this low technical variation and will be discussed at the end of this chapter. Low technical variation during measurement led to clear effects observed between the gene knockout treatment groups. The minimal genome cells are presented here as the biological sample with the simplest metabolome of any organism.

1.3.3 Metabolomics analysis of *E. coli* protein digests

This study began as a hypothesis driven targeted analysis of post-translational modification (PTM) of lysine in bacterial protein. Non-targeted LC-MS/MS analysis revealed spontaneous chemical formation had occurred during this analysis. This finding was compiled and is reported below to emphasize complications important to consider even when measuring relatively simple biological matrices.

Adapted from *"*Quantification of *N*^{6}-formylated lysine in bacterial protein digests by LC-MS/MS despite spontaneous formation and matrix effects*".* 2021. Jacob S. Folz, Jenelle A. Patterson, Andrew D. Hanson, Oliver Fiehn. *Rapid Communications in Mass Spectrometry, 35, e9019.*

PTMs of proteins, including acetylation and formylation, are important in regulating biological systems.[18,19] Measuring acetyl and formyl PTMs employs LC-MS/MS analysis via a proteomic approach (analyzing peptides)[19,20], or measuring modifications after digesting proteins to single amino acids.[21–23] Measuring *N*^{6}-formyl lysine in proteins is of increasing interest because this modified amino acid forms at locations in proteins that interfere with other PTMs (acetylation and methylation).[19] *N*^{6}-Formyl lysine levels increase in biological systems with more oxidation of DNA[24] or exposure to formaldehyde[21,25], which links *N*^{6}-formyl lysine to detrimental processes.

Acetylation and formylation of lysine can occur non-enzymatically.[26,27] Lysine is prone to spontaneous modification due to its reactive side-chain amino group. *N*^{6}-Formyl lysine has been reported to form spontaneously in the presence of formaldehyde[27], formic acid[20], aristolochic acid[22], 4-nitrophenyl formate[28], in beer samples[29] and would likely form in the presence of other formyl donors. Since these modified amino acids can be formed non-enzymatically, it is important to correct quantitative analyses for such spontaneous formation.

N^6-Formyl lysine was investigated in the current study to determine the extent of its spontaneous formation in run solvents and sample matrix of whole cell *Escherichia coli* protein digests. An untargeted metabolomics analysis was carried out to capture both the expected and unexpected metabolites that were formed.

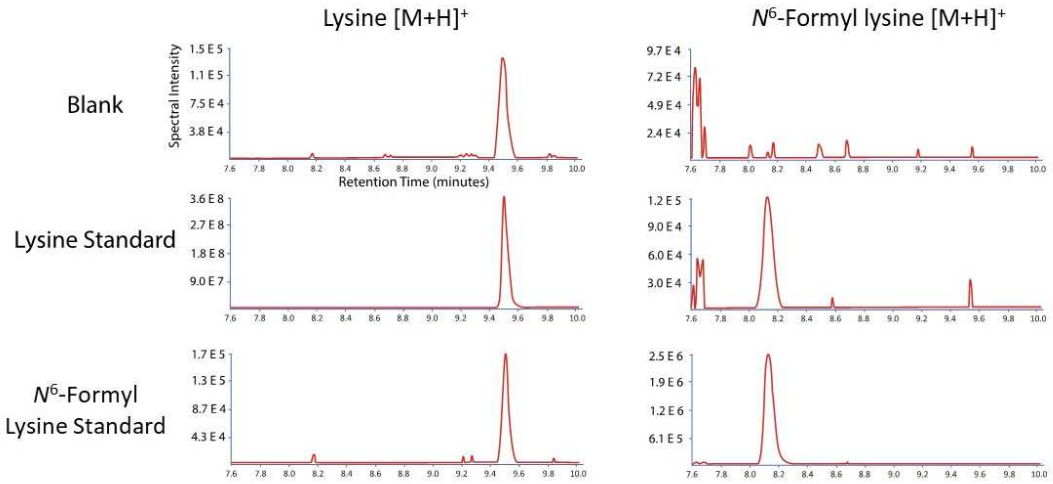


Figure 1.1. Extracted ion current chromatograms of lysine in the three left extracted ion current chromatograms ($[M+H]^+$ adduct at *m/z* 147.113) and N^6-formyl lysine in the right three extracted ion current chromatograms ($[M+H]^+$ adduct at *m/z* 175.108). The lysine standard was at a concentration of 10 µg/ml, and N^6-formyl lysine standard was at a concentration of 10 ng/ml. Blank indicates analysis of blank stock run solvent.

N^6-Formyl lysine was detected as a protonated adduct ion $[M+H]^+$ at *m/z* 175.107 in the extracted ion current chromatogram at 8.15 min retention time. This peak consistently appeared during analysis of an authentic lysine standard (Figure 1.1). We assessed the possibility of contamination of the lysine standard with N^6-formyl lysine by collecting the UHPLC fraction within 10 sec of the lysine signal at 9.5 min retention time (Figure 1.1). The collected fraction was reanalyzed and showed the same ratio of N^6-formyl lysine to lysine compared to the original authentic lysine standard. Since LC purification did not reduce the presence of N^6-formyl lysine, we hypothesized that lysine is spontaneously modified to N^6-formyl lysine during sample preparation or analysis. To test this

hypothesis, lysine and isotope-labeled lysine-D_8 were spiked at multiple concentrations into fully digested protein extracts or pure LC solvent, and analyzed using LC-MS/MS. N^6-Formyl lysine concentrations in pure LC solvent and in protein digest matrix increased linearly with the concentration of lysine spiked into samples (Figure 1.2). Formation of N^6-formyl lysine-D_8 was not observed in pure LC solvent (Figure 1.3) possibly due to an isotope effect discussed further below. However, we observed increased formation of N^6-formyl lysine-D_8 in protein digest matrix with increasing amounts of lysine-D_8 spiked standard (Figure 1.3). We identified N^6-formyl lysine-D_8 based on identical retention time and similarity of MS/MS product ions compared to N^6-formyl lysine (Figure 1.6). Interestingly, the increase of N^6-formyl lysine from total lysine precursor was also dependent on the presence or absence of sample matrix, with the rate of formation of N^6-formyl lysine > 6-times higher in sample matrix compared to pure LC solvent (Figure 1.2). These findings highlight the digested protein sample matrix as a major factor in spontaneous lysine formylation.

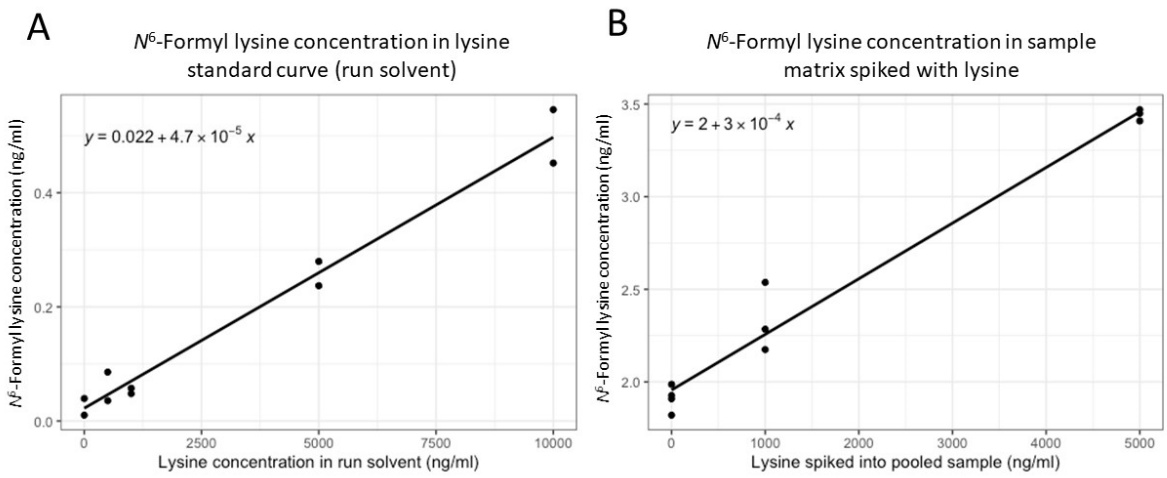


Figure 1.2. Linear models of A.) N^6-formyl lysine concentration dependent on concentration of lysine added to run solvent (both in ng/ml) and B.) N^6-formyl lysine concentration dependent on concentration of lysine spiked into sample matrix (both in ng/ml). Pearson correlation coefficient of both A and B are above 0.98.

N^6-Formyl lysine has been reported to form spontaneously in the presence of formyl donors including formaldehyde[27] and formic acid[28]. Since formaldehyde was used during growth of *E. coli* for protein collection and formic acid was used as an LC mobile phase modifier, it is possible that trace amounts of these chemicals are responsible for the conversion of lysine to N^6-formyl lysine. It was recently reported that during proteomic analysis, N^6-formyl lysine forms even when low concentrations of formic acid are present in resuspension solvent.[20] While we used formic acid as a mobile phase modifier for HILIC chromatography, it was previously reported that formic acid in LC-solvents had negligible impact on lysine formylation compared to sample resuspension conditions.[20] N^6-Formyl lysine also forms from lysine as a stable end product of the Maillard reaction.[26,30] Our extracts did not include any sugars, excluding this spontaneous reaction to formylate

lysine. We therefore concluded that additional formyl donors must be present in these protein digests that contributed to spontaneous formylation of lysine.

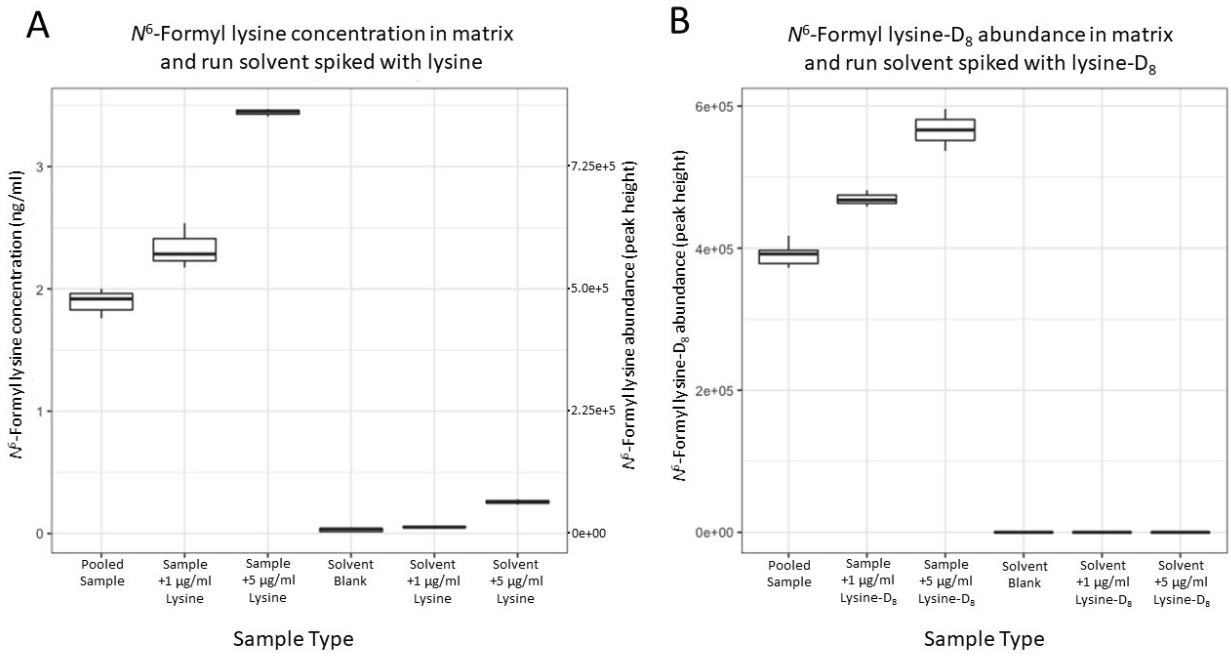


Figure 1.3. N^6-formyl lysine and N^6-formyl lysine-D_8 separated by sample type. A.) Concentration in ng/mL (on left y axis) and peak height intensity (right y axis) of N^6-formyl lysine separated by sample type of pooled sample matrix (Pooled Sample) or stock run solvent (Solvent) with and without added lysine (n=3). B.) Peak height intensity of N^6-formyl lysine-D_8 separated by sample type with and without added lysine-D_8 (n=3). Boxes represent +/- 1 interquartile range and line extends to the sample furthest from the median sample.

Interestingly, we saw correlations of non-annotated peaks with increasing amounts of spiked lysine into sample matrix. Specifically, an unknown compound was detected at 12 sec HILIC retention time later than N^6-formyl lysine elution. When comparing the MS/MS spectrum of this peak to N^6-formyl lysine, N^6-acetyl lysine and N^2-acetyl lysine we found

highly similar product ion ratios between N^2- and N^6-modifications, thus we annotated this unknown peak as N^2-formyl lysine (Figure 1.6;Figure 1.7). For example, product ions *m/z* 112.075 and 126.091 indicate the exact difference of a CH_2-group between N-formyl and N-acetyl lysine species. We subsequently confirmed the annotation of N^2-formyl-lysine using an authentic chemical standard that showed the same retention time and MS/MS spectrum compared to the previously unknown peak. The MS/MS spectrum is now publicly available at the Mass Bank of North America (splash10-003r-4900000000-5ab794b9b8a6a334bbd0). In the same manner as N^6-formyl lysine, N^2-formyl lysine was spontaneously formed with increasing amounts of spiked lysine in protein digest matrix (Figure 1.4). Similarly, we identified N^2-formyl lysine-D_8 (Figure 1.6) that increased to a similar extent as N^2-formyl lysine. N^2-formyl lysine formed to a lesser extent compared to N^6-formyl lysine, which represents the difference in reactivity between the N^2- and N^6-positions of lysine.[28]

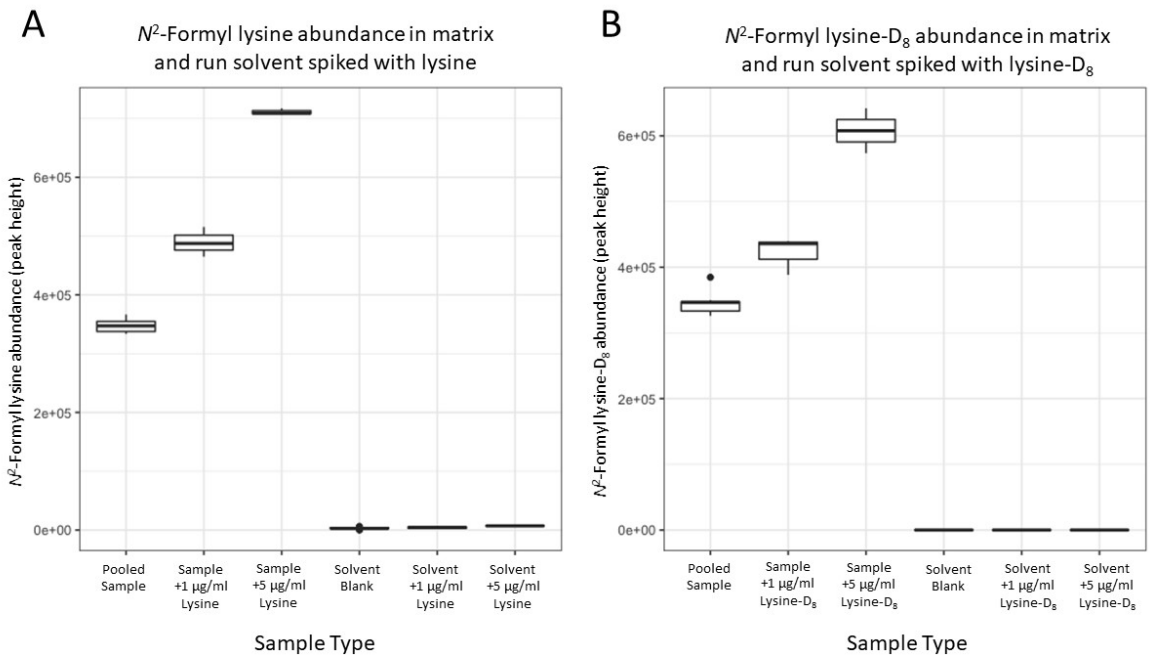


Figure 1.4. N^2-formyl lysine and N^2-formyl lysine-D_8 abundance separated by sample type. A.) Peak height intensity of N^2-formyl lysine separated by sample type of pooled sample matrix (Pooled Sample) or stock run solvent (Solvent) with and without added lysine (n=3). B.) Abundance of N^2-formyl lysine-D_8 separated by sample type with and without added lysine-D_8. Boxes represent +/- 1 interquartile range and line extends to the sample furthest from the median sample, unless sample was further than 1.5 interquartile ranges away from the median in which case the sample is plotted as a single point.

The most dramatic effect observed with respect to formylation and acetylation of lysine is the difference between sample matrix and run solvent, however there was also an effect of deuterium labeling. N^6-Formyl lysine formed in run solvent with addition of lysine, but there was no signal of N^6-formyl lysine-D_8 in run solvent with addition of lysine-D_8 (Figure 1.3). In run solvent there are trace signals (peaks with signal/noise < 2) for N^2-formyl

lysine with addition of 5 µg/mL lysine, but no signal for N^2-formyl lysine-D_8 with addition of 5 µg/mL lysine-D^8 (Figure 1.4). We expect that these lysine-D_8 adducts form in pure run solvent, but the signals are below the limit of detection. An explanation for a lower rate of reaction for deuterated lysine compared to non-deuterated lysine is the deuterium isotope effect, which can lead to different reaction rates.[31] An isotope effect would also explain the slightly lower abundance of N^6-formyl lysine-D_8 and N^2-formyl lysine-D_8 being formed in sample matrix with addition of lysine-D^8 compared to their non-isotope labeled counterparts (Figure 1.3;Figure 1.4).

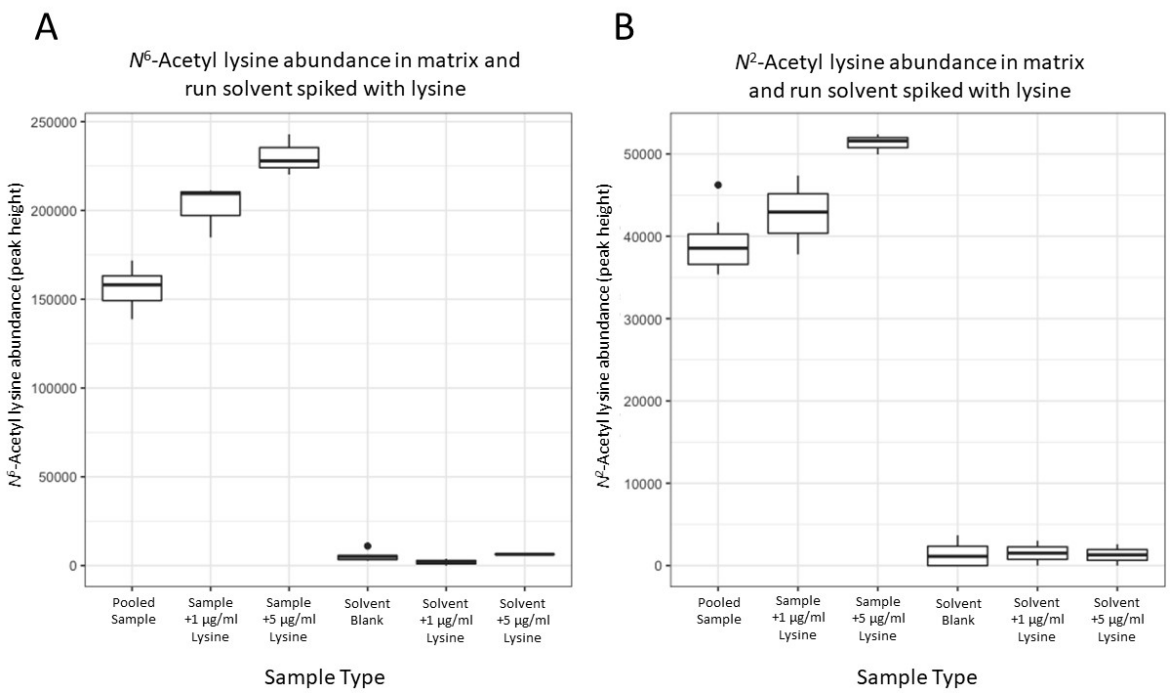


Figure 1.5. N^6-Acetyl lysine and N^2-acetyl lysine abundance by sample type. A.) Peak height intensity of N^6-acetyl lysine separated by sample type of pooled sample matrix (Pooled Sample) or stock run solvent (Solvent) with and without added lysine (n=3). B. Peak height intensity of N^2-acetyl lysine separated by sample type with and without added lysine. Boxes represent +/- 1 interquartile range and line extends to the sample furthest from the median sample, unless sample was further than 1.5 interquartile ranges away from the median in which case the sample is plotted as a single point.

Together these data support the hypothesis that N^6-formyl lysine and N^2-formyl lysine are formed spontaneously from lysine during analysis of digested protein extracts. Furthermore, we show that the formation of these adducts occurred at or after sample resuspension during the analytical workflow. This finding is in agreement with a previous proteomic study that concluded that N^6-formyl lysine was not formed during sample protein extraction or digestion.[22] This was tested by performing sample preparation with and without strong antioxidant protection to exclude formaldehyde exposure.[22] Similarly,

ascorbate could be used as an antioxidant to inhibit formaldehyde induced formylation of lysine[27]. Yet, our study is the first to use isotopically labeled lysine to investigate N^6-formyl lysine formation at or after the point of resuspension in an LC-MS/MS workflow. Interestingly, we also found spontaneous formation of increased amounts of N^2- and N^6-acetyl lysine (Figure 1.7) when lysine was spiked into sample matrix (Figure 1.5). Such spontaneous formation has been previously reported in mitochondrial matrix[32], but our experiments indicated that lysine acetylation can be also expected as a chemical artifact due to acyl donors present in the complex protein digest samples, especially because acetic acid was not used at any point during our sample preparation or analysis. These results also suggest that other amino acids may become formylated or acetylated during analysis since these adducts can form at the at the N^2 position which is present in all free amino acids. Unfortunately, the exact nature of the chemical donor for either formylation or acetylation in the protein digest extracts remain elusive. Our data show that this phenomenon exists and that proper negative controls must be included to correct for spontaneous chemical damage reactions in any study of biologically driven acetylations and formylations.

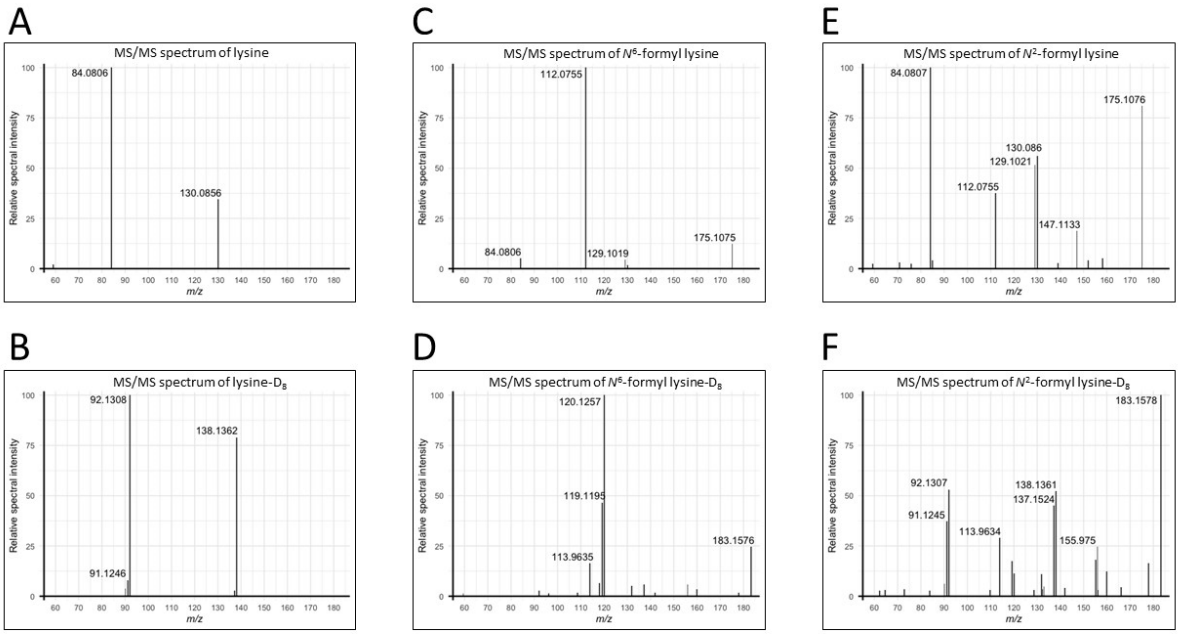


Figure 1.6. Experimental MS/MS fragmentation spectra and retention time (RT) in minutes for A.) Lysine (RT): 9.50 B.) Lysine-D_8 RT: 9.50 C.) N^6-Formyl lysine RT: 8.13 D.) N^6-Formyl lysine-D_8 RT: 8.13 E.) N^2-Formyl lysine RT: 8.33 F.) N^2-Formyl lysine-D_8 RT: 8.33.

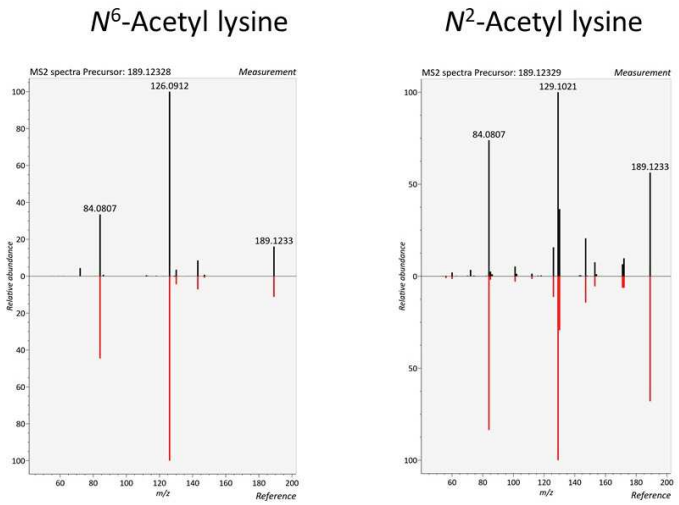


Figure 1.7. MS/MS spectra gathered from experimental acetylated lysine residues shown in top panels (ions marked in black), compared to library MS/MS matches in bottom (ions marked in red). N^6-Acetyl lysine was matched to retention time (8.00 minutes) and a MS/MS library, while N^2-acetyl lysine (RT: 8.21 minutes) was matched to library MS/MS spectrum of N^2-acetyl lysine.

The non-targeted metabolomics data analysis used in this study was integral in making these discoveries. Targeted measurement of N^6-formyl lysine was important in making the preliminary observation of spontaneous formation of N^6-formyl lysine, but measurement of the deuterium labeled standards, deuterium labeled spontaneous products, and additional unexpected chemicals (acetyl-lysines, and deuterium labeled acetyl-lysines) was allowed through the non-targeted LC-MS/MS data collection technique.

1.3.4 Takeaways from analyzing pure bacterial samples

Analysis of these pure cultures of single celled organisms show that even simple biological samples possess complex metabolomes. These samples were extremely

uniform between different treatment groups which allowed for reproducible data collection and data processing. The low abundance of metabolites in JCVIsyn3.0 samples lead to very little technical variation. Low sample concentration is expected to have resulted in low ion suppression effects and thus contributed to low technical variation. Minimizing ion-suppression is desirable and will be thoroughly discussed at the end of this chapter. In total 170 unique metabolites were annotated and hundreds of other “chromatographic features” were detected which suggests hundreds more metabolites being created within this organism. This observation of a chemically complex sample matrix from one simple organism sets the expectation high for the chemical complexity of samples containing diverse mixtures of bacteria.

Measurement of the spontaneous formation N^6-formyl lysine reveals that chemicals can form in samples spontaneously. These results highlight the importance of using non-targeted metabolomics methods that can detect unexpected trends in chemicals that were not hypothesized to be important. This study also highlights the importance of protecting samples from conditions that can lead to degradation of chemicals, and formation of new chemicals. To address spontaneous chemical formation samples should be stored at low temperatures, protected from light, and proper negative controls should be included to track chemical degradation and formation for reactive metabolites.

1.4 Chemical profiling of plant-based foods

When measuring the metabolome of human intestinal samples, food is a major source of metabolites. Plant based foods contain tens of thousands of metabolites which can be absorbed during digestion, transformed by microbes, or transformed by human enzymes.

Plant metabolites are thus prevalent and important to consider when investigating the metabolome of the human GI tract. In an ideal situation, all metabolites consumed in a particular food are known and then tracked through the digestive tract. Knowing exactly what metabolites enter the intestinal tract would help discriminate whether metabolites are derived from food, human, or microbes, primarily because we do not have comprehensive libraries of metabolites from food (whether plant, animal or derived from other sources), nor do we know all of the metabolic pathways that human endogenous bacteria use during fermentation. The reason to be further considered here is the difficulty in performing comprehensive metabolite analysis of diverse plant-based foods.

Metabolomics analysis of plant-based foods is hindered by extreme variation between samples. Plants vary profoundly in metabolite makeup, and total metabolite concentration. Samples discussed with respect to pure microbial cultures were homogenous and had little variation in the metabolome between samples and treatment groups. Homogeneity in sample sets facilitates data collection and data processing, yet metabolomics experiments do not always contain chemically similar matrices. For example, measuring multiple types of plant-based foods in the same metabolomics experiment shows dramatic differences between plant types. This variation makes data processing more prone to incorrect metabolite annotations which is further discussed below.

Measurement of 10 different types of commonly eaten foods was carried out using several metabolomics analytical platforms. I chose the sample types to encompass a range of fresh and dried foods. The samples analyzed were fresh foods (soybean, kale, blueberry, and blackberry), dried foods (tea leaves, coffee beans, biliberry powder), and processed

foods (chocolate, wine). Each food was homogenized without further treatment or drying and 100mg (100uL for wine) of each was extracted using methanol and water. Each sample was analyzed using non-targeted RP and HILIC LC-MS/MS methods. The profound difference in the concentration of metabolites within the different food types caused some foods (specifically foods with high water content like the fresh foods and wine) to have extremely few metabolites detected, while other foods (the dried foods and chocolate) had metabolite concentrations that saturated the chromatography column and prevented reliable peaks to be measured. To accurately and thoroughly measure metabolites from all food types, each food could be individually optimized for the amount of sample to be analyzed. After acquiring reliable data, the data processing presents an additional hurdle for documenting metabolites from different foods. Data processing software programs perform peak alignment and gap filling in non-targeted data when analyzing many samples together. This leads to chromatographic features at a specific *m/z* and retention time to be annotated as the same metabolite under the assumption of homogeneity in the chemical makeup of all sample types in a given study. Yet, in comparative studies like this with different foods, and later when analyzing gut contents at different time points or from different subjects, the homogeneity of sample types is not given. When analyzing different types of samples, such as different plant-based foods, there may be distinct chemical isomers that co-elute chromatographically and lead to false conclusions that the same chemicals present. When data from all plant types are aligned, only one metabolite annotation is chosen to correspond to all the aligned chromatographic features even if the features are from different unique chemicals. In these cases, metabolites can be incorrectly assigned. To avoid mis-annotation of

metabolites due to similar peaks appearing in multiple plant types, the data from each plant type would need to be processed individually. Data processing to achieve accurate metabolomics datasets is time consuming and processing each food type individually was not feasible for completion within this thesis work. For these reasons, high throughput analysis of the 10 different foods considered was not further pursued for the effort of building a library of food metabolites that could be tracked through the GI tract.

1.4.2 Takeaways from metabolomics analysis of plant-based foods

If a single extraction and analytical analysis is used to measure different matrices, the method may not result in reliable or trustworthy metabolome coverage across all sample types. Analysis of plant samples with dramatically different total metabolite concentrations will also lead to data collection difficulties including ion suppression, and difficulty in achieving reliable metabolite annotations. The primary interest of this research is to profile chemicals of the human GI tract, so the analysis of individual foods was not further pursued. Fortunately, there are databases such as “foodb” and “COCONUT” that contain tens of thousands of food and plant related metabolites that can be matched with metabolites found in human GI tract samples and provide some insight into the origin of a metabolite.

1.5 Chemical profiling of mammal serum and blood plasma

Serum and blood plasma are frequently collected from mammals (humans, mice, non-human primates, pigs, etc.) for various research goals. It is popular to collect this biofluid as it can usually be collected without sacrificing or causing biologically significant harm to the study subjects. Furthermore, serum and plasma circulate through the entire body

which makes it possible to gain insights into different aspects of physiology of an organism without sampling each organ or tissue individually. Blood can also provide information relevant to the GI tract. Metabolites from food are found circulating in the bloodstream after being absorbed during digestion. Blood plasma gives insight into aspects of the digestive tract and is a useful matrix to study when considering the effect of diet and gut microbiota intervention studies (*e.g.* intestinal microbiota transplantation and antibiotic treatment).

Metabolites in the blood give insight into the type of food an individual consumes. Some plants contain unique metabolites that can pinpoint a specific food or type of food that a person recently consumed. For example, secoisolariciresinol diglucoside is a chemical found at very high concentration in flaxseed and leads to biomarkers in blood that suggests flaxseed consumption. Secoisolariciresinol diglucoside is deglycosylated by gut microbes and absorbed into the bloodstream[33]. Plant metabolites like this are powerful in identifying specific foods consumed, but endogenous human metabolites also change depending on the food consumed[34]. For example blood plasma levels of amino acids, choline, creatinine, pyruvate, and succinic acid have been shown to change in response to different meals[34]. These metabolites are more difficult to interpret because they could be from food directly or endogenously regulated due to diet. It is therefore important to consider both endogenous and exogenous metabolites when looking at how diet affects metabolism of the host.

Two diet related studies are included in this chapter. The first study is an experiment investigating how different ratios of dietary sodium and potassium and the gut microbiota influence circulating blood lipids in rats. The second study is an investigation of how

consumption of 10 grams of ground flaxseed per day affects the chemical profile of serum in 178 postmenopausal women. For this dietary intervention study non-targeted metabolomics analysis of stool was also conducted and reported in the following section. Targeted analysis of bile acids for this flaxseed experiment is reported in chapter three.

1.5.1 Interaction of Gut Microbiota and High-Sodium, Low-Potassium Diet in Altering Plasma Triglyceride Profiles Revealed by Lipidomics Analysis

Adapted from "Interaction of Gut Microbiota and High-Sodium, Low-Potassium Diet in Altering Plasma Triglyceride Profiles Revealed by Lipidomics Analysis"Jacob Folz, Young Taek Oh, Ivana Blaženović, Joyce Richey, Oliver Fiehn and Jang H. Youn. 2019. *Molecular Nutrition and Food Research*, 63, 1900752

1.5.2.1 Introduction

Cardiovascular disease (CVD) and type 2 diabetes are leading causes of mortality and morbidity and are major public health issues in the United States and worldwide. Diet has profound effects on the pathogenesis and progression of these diseases. Macronutrients in the diet, such as fat, carbohydrate, and protein, have been extensively studied as important environmental factors for increased prevalence of cardiovascular (CV) and metabolic diseases. In addition, high-sodium (Na^+) and low-potassium (K^+) (HNaLK) intake have been shown to increase blood pressure, endothelial dysfunction, and CVD morbidity and mortality[35]. HNaLK intake were also associated with metabolic syndrome[36], a cluster of CV risk factors.

Dietary components may directly affect the host by altering circulating levels of energy substrates, hormones, electrolytes, and/or signaling for organ interactions. However,

recent studies have established that the gut microbiota plays major roles in host metabolic and CV health. Ample evidence shows that gut bacteria interacting with dietary components mediate many dietary effects on host metabolic and CV systems[37,38]. Recently, we reported interactions of the gut microbiota and a HNaLK diet in altering circulating biogenic amines, providing novel insights into the roles of the gut microbiota in regulating biogenic amines and possibly mediating the CV detrimental effects of HNaLK intake[39]. In the present study, we expanded our previous study to investigate the interaction of gut microbiota and dietary minerals (HNaLK) in altering plasma lipid profiles. Dyslipidemia is associated with insulin resistance, type 2 diabetes, and CVD; altered lipid profiles play major roles in diet effects on metabolic and CV health[40]. The gut microbiota was shown to significantly impact host lipid metabolism or circulating lipid profiles[41]. In the present study, we tested the hypothesis that HNaLK diets, by interacting with the gut microbiota, significantly impact circulating lipid profiles. This is an important study, as positive results would provide novel insights into mechanisms by which HNaLK diets lead to metabolic or CV diseases.

Circulating lipid profiles, determined in rats by an integrated lipidomics analysis and analyzed by various statistical and bioinformatics tools, showed significant interactions between gut bacteria and an HNaLK diet in altering plasma lipids, particularly triglyceride (TG) profiles. Interestingly, we identified two distinct groups of TG species in plasma differentially regulated by the gut microbiota. One group represented TGs composed of very-long-chain polyunsaturated fatty acids (PUFAs), previously shown to be associated with reduced diabetes risk.[9] This group was regulated by the gut microbiota and the

HNaLK diet in a manner suggesting this may be a novel pathway by which HNaLK intake increases metabolic or CV risk.

1.5.2.2 Results

In our earlier study, [5] we reported that food intake and weight gain were altered neither by the HNaLK diet nor by the antibiotic treatment under present experimental conditions. In addition, despite marked differences in Na^+ or K^+ intake, plasma Na^+ and K^+ concentrations were not different between the control and the HNaLK diet groups, thanks to efficient homeostatic mechanisms for the regulation of extracellular Na^+ and K^+ levels.

Profiling of plasma lipids on reversed phase liquid chromatography detected a total of 2,873 features, including 328 known lipids. The known lipids included different species of acylcarnitine (*n*=7), cholesterol ester (CE; *n*=11), ceramide (*n*= 17), diacylglyceride (DG; *n*=9), phosphatidylcholine ester (*n*=11), FFA (*n*=16), galactosylceramide (*n*=4), lysophosphatidylcholine (LPC; *n*=18), lysophosphatidylethanolamine (*n* = 5), phosphatidylcholine (PC; *n* = 67), phosphatidylethanolamine (*n* = 24), phosphatidylinositol (*n* = 14), sphingomyelin (*n* = 24), and TG (*n* = 101). PCA analysis showed that rat plasma samples were clustered together with no apparent outliers and clearly separated from the cluster of quality controls of human-plasma samples. However, no separation was observed among the 4 experimental groups.

We first examined the effect of antibiotic treatment on lipids by combining the data with different diets (*n* = 14 each for untreated vs. antibiotic-treated rats). ChemRICH analysis identified several lipid clusters that were significantly altered by the antibiotic treatment (Figure 1.8,Table 1.1). The most impressive effects were seen with unsaturated TGs (*P*

= 5.3E-15): 60 TG species (of 101 detected) identified as a cluster. Among them, 17 species (28%) significantly decreased and six increased with antibiotic treatment (Table 1.1). Another dramatically regulated lipid class was unsaturated PCs: 20 unsaturated PCs (of 67 detected) were significantly higher in antibiotic-treated rats compared to untreated rats (*P* = 8.0E-12). In addition, 15 unsaturated FFAs were identified as a cluster that showed significant effects to increase with antibiotic treatment (*P* = 2.7E-5). ChemRICH additionally calculated statistically significant increase in carnitines, CEs, LPCs, and DGs, and significant decreases in PEs and unsaturated ceramides. Changes in PCs, FFAs, PEs, and CEs were chemical classes leading to several significantly impacted groups/functions identified by LION/web analysis in which regulation of glycerophosphocholines were the most strongly changed lipid with antibiotic treatment.

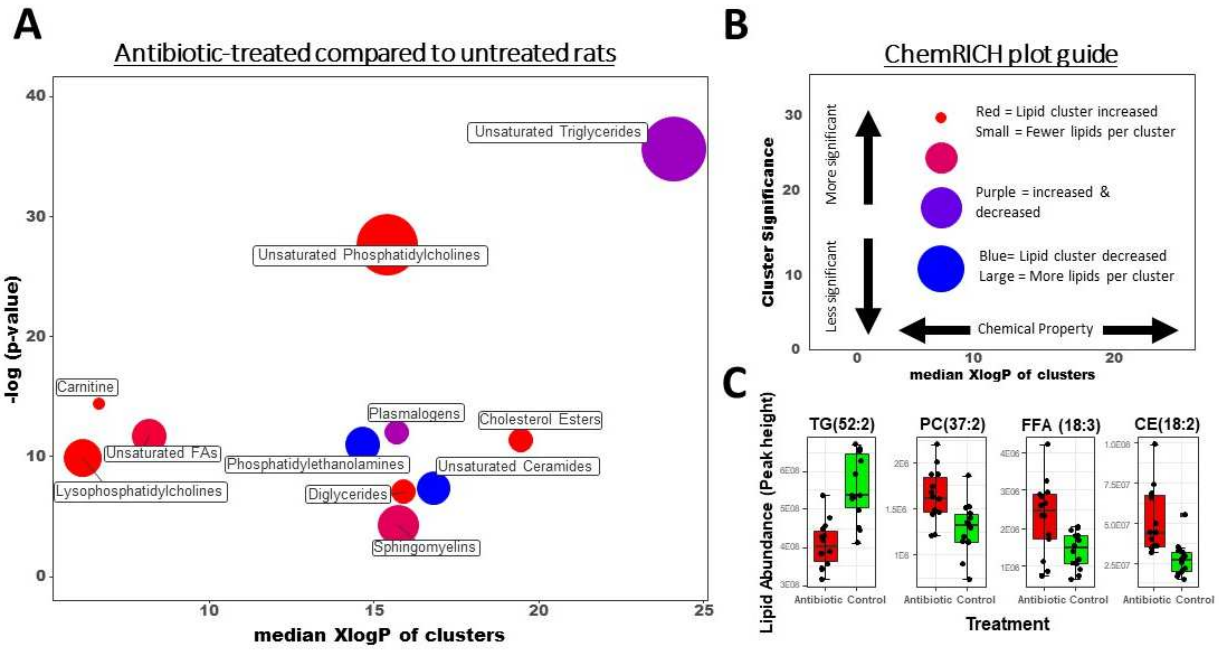


Figure 1.8. ChemRICH plot for comparison of antibiotic-treated to untreated rats (A), guidance as to how to read a ChemRICH plot (B), representative boxplots for some significant lipid classes (C). (A)The lipidomic dataset generated in this study was analyzed using ChemRICH to identify chemical classes that were significantly regulated by antibiotic treatment (n = 14). (B) Circles represent chemical classes named with text nearest to each circle. Circle size represents relative number of lipids in each cluster. The y-axis position indicates significance level calculated using Kolmogorov–Smirnov-test with more significantly altered clusters positioned higher on the y coordinate. The x-axis position represents the average intrinsic chemical property of XlogP for the cluster of lipids. Circle color indicates whether the chemical class was higher (red), lower (blue) or had a mixed change (purple) in antibiotic-treated animals compared to untreated animals. (C) Boxplots of lipid peak heights for examples of antibiotic compared to control groups (n=14). Boxes represent distance between quartiles 1 and

3 and error bars extend up to 1.5 inter quartile ranges (IQR) from the upper (third quartile) or lower (first quartile) portion of boxes. Samples greater than 1.5 IQR from the first or third quartile are plotted as individual points outside of error bars. PC, phosphatidylcholine; CE, Cholesterol Ester.

Table 1.1. Lipid clusters with a significant antibiotic-treatment effect determined by ChemRICH. The antibiotic-treated group was compared to the untreated group (n = 14 each). Calculations for significance were performed with ChemRICH software. Δ indicates the number of changed lipids in the cluster. FDR, False detection rate.

Cluster name	**Size**	**p-values**	**FDR**	**Δ**	**↑**	**↓**	**Δ/Size**
Unsaturated Triglycerides	60	3.3E-16	5.3E-15	23	6	17	0.4
Unsaturated Phosphatidylcholines	53	1.0E-12	8.0E-12	20	20	0	0.4
Carnitines	5	5.5E-07	2.9E-06	4	4	0	0.8
Plasmalogens	8	6.1E-06	2.5E-05	3	1	2	0.4
Unsaturated FFA	15	8.4E-06	2.7E-05	7	6	1	0.5
Cholesterol Esters	8	1.1E-05	3.1E-05	3	3	0	0.4
Phosphatidylethanolamines	15	1.6E-05	3.8E-05	8	0	8	0.5
Lysophosphatidylcholines	18	5.2E-05	1.0E-04	6	6	0	0.3
Unsaturated Ceramides	14	6.4E-04	1.1E-03	3	0	3	0.2
Diglycerides	8	0.00086	0.0014	2	2	0	0.2
Sphingomyelins	21	0.014	0.02	7	5	2	0.3

We next examined whether the HNaLK diet had significant effects on plasma lipids by combining data from untreated and antibiotic-treated rats (n = 14 each for the control and the HNaLK diet group). ChemRICH analysis detected no lipid clusters that showed significant effects of the HNaLK (vs. control) diet. However, when ChemRICH analysis for antibiotic effects was performed separately in each diet group, the results depended on the diet; the antibiotic effects on TGs were more significant (more TG species affected) with the control than with the HNaLK diet (Table 1.2). In addition, the effect of antibiotic treatment to decrease PEs was observed only with the control diet, whereas the effect on carnitine was observed only with the HNaLK diet. These data suggest some interaction between gut bacteria and the HNaLK diet in altering plasma lipids. This interaction was also apparent in PLS-DA analysis, which identified 10 TGs as highly ranked VIP lipids, as well as two LPCs, two PCs, and CE 18:2 in differentiating the 4 experimental groups. LION/web analysis showed a significant impact of diet regardless of antibiotic treatment and identified several TG-focused groups (i.e., headgroup with neutral charge, glycerolipids, lipid storage, lipid droplet, and triacylglycerols) that significantly changed with the HNaLK diet.

Table 1.2. Lipid clusters with a significant antibiotic-treatment effect separated by diet. Antibiotic effects were determined using ChemRICH to compare antibiotic-treated to untreated groups in each diet group (n = 14 for each diet group; n = 7 each for antibiotic-treated and untreated groups). Δ indicates the number of changed lipids in the cluster. FDR, False detection rate.

Control Diet							
Cluster name	**Size**	**p-values**	**FDR**	**Δ**	**↑**	**↓**	**Δ/Size**
Unsaturated Triglycerides	60	5.5E-13	8.8E-12	17	0	17	0.3
Unsaturated Phosphatidylcholines	53	0.00013	0.001	4	4	0	0.08
Phosphatidylethanolamines	15	0.00041	0.0022	4	0	4	0.3
Cholesterol Esters	8	0.0048	0.019	3	3	0	0.4
Lysophosphatidylcholines	18	0.021	0.066	3	3	0	0.2
High Sodium Low Potassium Diet							
Cluster name	**Size**	**p-values**	**FDR**	**Δ**	**↑**	**↓**	**Δ/Size**
Unsaturated Phosphatidylcholines	53	0.000015	0.00024	7	7	0	0.1
Unsaturated Triglycerides	60	0.00012	0.00095	5	1	4	0.08
Carnitine	5	0.00095	0.0051	4	4	0	0.8
Lysophosphatidylcholines	18	0.0034	0.014	5	5	0	0.3
Cholesterol Esters	8	0.0094	0.03	2	2	0	0.2
Unsaturated FFA	15	0.022	0.058	2	2	0	0.1

In all analyses (ChemRICH, LION, and PLS-DA), TG was the lipid class most impacted by the diet or antibiotic treatment. Antibiotic treatment decreased most TG species, but

a group of TG species were increased by antibiotic treatment (Figure 1.9 A). Interestingly, these TG species showed a tendency to decrease with the HNaLK diet (Figure 1.9 B). Although many of these individual changes did not gain statistical significance due to large variations, this association was significant; 16 TG species were increased > 2-fold by antibiotic treatment. When we selected TG species (n = 13) altered more than 40% by the diet, they were all among the 16 TG species altered by antibiotic treatment (Table 1.3). The probability of this association is calculated to be 6.9E-14 (= ${}_{88}C_3/{}_{101}C_{16}$). Thus, the TGs increased by gut-bacteria depletion were the same TGs affected by the diet to the greatest extent. We define this group of unsaturated TGs as cluster 2 (n = 13) to distinguish it from the group of unsaturated TGs (cluster 1, n = 23) that were significantly decreased by antibiotic treatment ($P < 0.05$). TG clusters 1 and 2 exhibited significantly different acyl-chain carbon numbers and double-bond content (Figure 1.10); TG cluster 2 is characterized by greater carbon number and double-bond content, compared to cluster 1. In addition, cluster 2 includes TGs comprising very-long-chain PUFAs, such as 20:4, 20:5 (eicosapentaenoic acid [EPA]), 22:4, 22:5, and 22:6 (docosahexaenoic acid [DHA]), previously identified to be associated with reduced diabetes risk.[9] In the present study, we show for the first time that gut bacteria and the HNaLK diet both modulated these TGs.

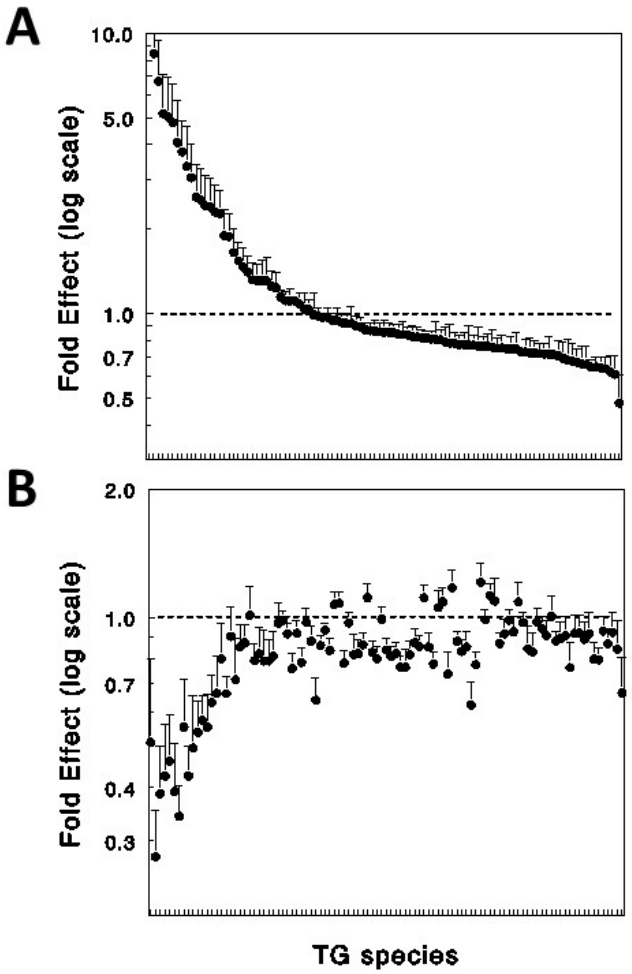


Figure 1.9. Effects of antibiotic treatment (A) and HNaLK diet (B) on individual plasma TG species detected by lipidomics analysis. Data are means of lipid abundance ± S.E.M. (n = 14), expressed as a fold effect compared to the untreated group (A) or the control diet group (B). Marks on the x-axis represent individual TG species, sorted by the size of the fold effect of antibiotic treatment; individual TGs are located in the same positions on the x-axis in both graphs.

Table 1.3. TG species altered both by antibiotic treatment and by HNaLK diet. Fold change was determined by taking the antibiotic treatment group over the untreated group (Antibiotic effect), and the HNaLK diet over the control diet (Diet effect) (n = 14). TG, triglyceride. * and ** indicate two separate peaks in the LC-MS chromatogram.

TG Species	**Antibiotic effect**		**Diet effect**	
	Fold change	*P (unadjusted)*	Fold change	*P (unadjusted)*
TG(54:9)	10.1	0.059	0.28	0.469
TG 58:11; TG(18:2_20:4_20:5)	8.5	0.080	0.34	0.214
TG 60:12; TG(18:2_20:4_22:6)	6.7	0.044	0.39	0.242
TG 58:8; TG(18:1_18:2_22:5)*	5.2	0.042	0.39	0.239
TG(60:11)	5.0	0.051	0.43	0.295
TG 54:8; TG(16:1_18:2_20:5)	4.8	0.044	0.43	0.189
TG(56:9)**	4.1	0.076	0.46	0.156
TG 58:7; TG(18:1_18:2_22:4)	3.8	0.021	0.50	0.272
TG(58:9)	3.3	0.082	0.51	0.200
TG 56:9; TG(16:1_18:2_22:6)**	3.1	0.047	0.54	0.196
TG 54:7; TG(16:0_18:2_20:5)	2.6	0.059	0.55	0.213
TG(58:10)	2.5	0.048	0.55	0.230
TG(58:8)*	2.4	0.045	0.57	0.175

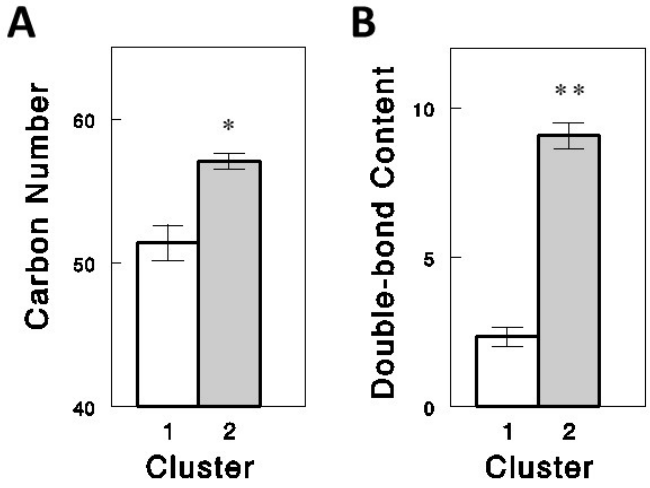


Figure 1.10. Average carbon number (A) and double-bond count (B) of cluster 1 (n = 23) and cluster 2 (n = 13) TGs. Cluster 2 TGs had significantly more carbons and double-bonds than cluster 1 TGs. Cluster 1 TGs are unsaturated TGs that were significantly decreased by antibiotic treatment ($P < 0.05$). TGs in cluster 2 appear in Table 3. Data are means ± S.E.M. (n = 23 for cluster 1 and 13 for cluster 2). *, $P < 0.001$; **, $P < 0.0001$ vs. cluster 1.

Figure 1.11 shows the effects of antibiotic treatment and/or the HNaLK diet on relative levels of TGs in clusters identified as described above. The data from individual TG species were averaged after normalizing to their own control group (i.e., control diet without antibiotic treatment). TG clusters 1 and 2 showed different patterns of regulation by the HNaLK diet and antibiotic treatment; cluster 1 was downregulated by the diet and antibiotic treatment whereas cluster 2 was downregulated by the diet but upregulated by antibiotic treatment. Unsaturated FFAs showed small but significant effects of antibiotic treatment ($P < 0.05$).

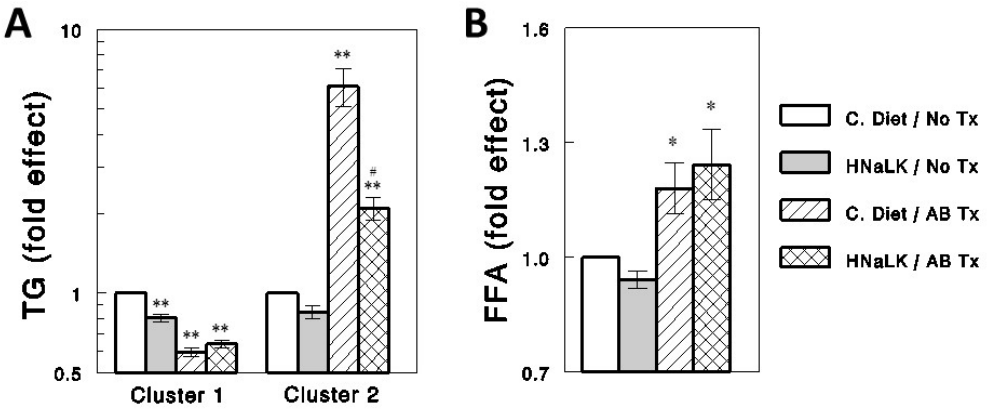


Figure 1.11. Effect of antibiotic treatment and/or HNaLK diet on relative levels of unsaturated TGs (A) and FFAs (B) in clusters identified by the lipidomics analysis. Data from individual unsaturated TG and FFA species were averaged for each group after normalizing to their own control group (i.e., control diet without antibiotic treatment). Cluster 1 TGs are unsaturated TGs that were significantly decreased by antibiotic treatment ($P < 0.05$). TGs in cluster 2 appear in Table 3. Data are means ± S.E.M. ($n = 23$, 13, and 15 for TG cluster 1, TG cluster 2, and unsaturated FFA, respectively), expressed as a fold effect. *, $P < 0.05$; **, $P < 0.001$ vs. C. Diet / No Tx; #, $P < 0.001$ vs. C. Diet / AB Tx. C. Diet, control diet; No Tx, no treatment; AB Tx, antibiotic treatment.

Total plasma TG levels, measured by an enzymatic colorimetric assay, showed patterns of changes similar to TG cluster 1 (Figure 1.12 A). The antibiotic effect to decrease total plasma TG levels was statistically significant ($P < 0.03$). This was expected because the majority of TG species detected by the lipidomics analysis decreased with antibiotic treatment (Figure 1.9). Total FFA levels measured by an enzymatic colorimetric assay did not show any significant treatment effects (Figure 1.12 B) different from the pattern observed with unsaturated FFAs detected by lipidomics analysis (Figure 1.11 B). It should be noted that our lipidomics analysis did not detect all major FFA species. For example, saturated FFAs palmitate (16:0) and stearate (18:0) were not reported due to high

background levels of these lipids, and the possibility exists that these FFA species may behave differently from unsaturated FFAs detected by the lipidomics analysis. The ratio of total FFA to total TG was greater in antibiotic-treated than in untreated rats (Figure 1.12 C), suggesting increased TG hydrolysis to FFA.

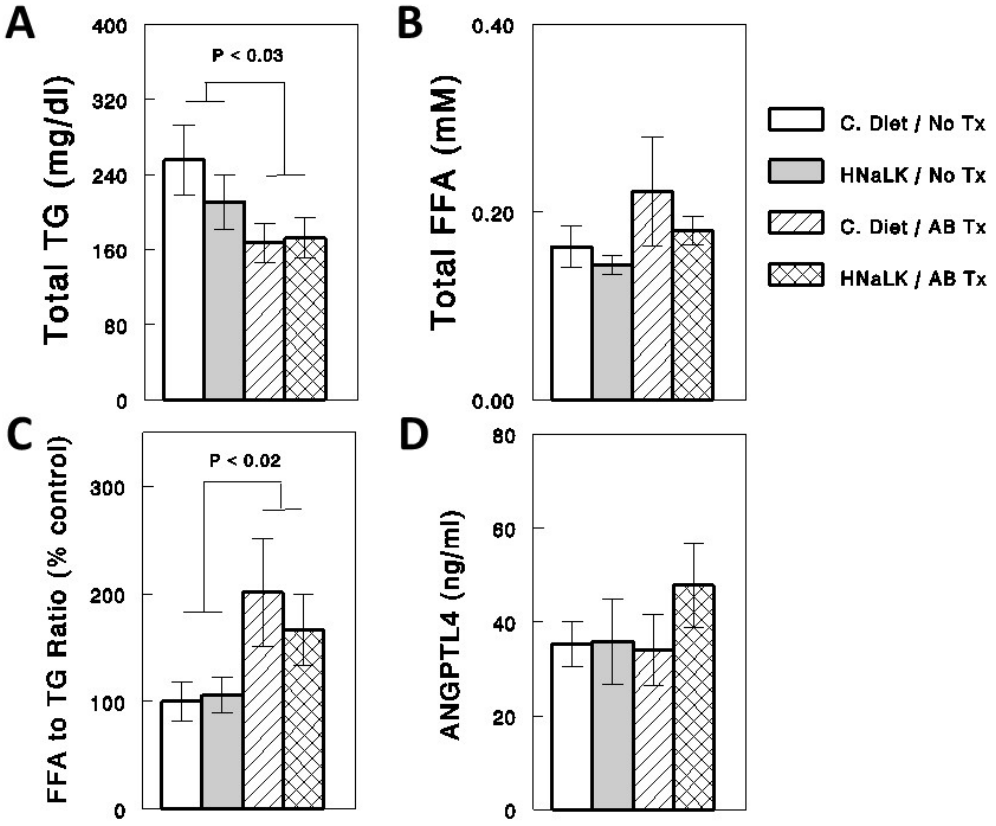


Figure 1.12. Plasma total TG (A) and total FFA (B) levels measured by enzymatic calorimetric assays, FFA-to-TG ratios (C), and plasma ANGPTL4 levels (D). Data are means ± S.E.M. (n = 7 for each group). C. Diet, control diet; No Tx, no antibiotic treatment; AB Tx, antibiotic treatment.

Previous studies suggested that gut bacteria regulate circulating TG hydrolyis to FFA by acting on intestinal cells to increase circulating ANGPTL4, an inhibitor of lipoprotein lipase.[22] We examined whether the apparent changes in TG hydrolysis to FFA in

antibiotic-treated rats were associated with changes in circulating ANGPTL4. Circulating ANGPTL4 levels did not change by antibiotic treatment or by the HNaLK diet (Figure 1.12 D), suggesting that ANGPTL4 might not be responsible for the changes in the ratio of FFA to TG.

1.5.2.3 Discussion

Targeting plasma lipids in the present lipidomics approach (combined by various statistical and bioinformatics analyses), we found significant interactions between gut bacteria and the HNaLK diet in modulating circulating lipids. In addition, our analysis (LION/web) showed significant effects of the HNaLK diet, regardless of antibiotic treatment, on several TG-focused lipid groups (e.g., glycerolipids, lipid storage, and lipid droplet). Importantly, we identified two distinct TG groups that were differentially regulated by antibiotic treatment. One group (cluster 1, n = 23), representing the majority of TG species detected, was downregulated, whereas the other group (cluster 2, n = 13) was upregulated by antibiotic treatment. Interestingly, cluster 2 TGs were also regulated by the diet; the same TG species were regulated by antibiotic treatment and by the diet. Cluster 2 TGs exhibited greater carbon-chain length and double-bond content compared to cluster 1 TGs, and included TGs composed of very-long-chain PUFAs, previously associated with reduced diabetes risk[42]. Total plasma TG levels, determined by an enzymatic colorimetric assay, showed patterns similar to those of cluster 1 TGs, expected from this group representing the majority of detected TGs. Taken together, these data suggest HNaLK diets may interact with gut bacteria to modulate circulating TG profiles.

Previous studies reported heterogeneity of circulating TGs in their regulation and association with risks of diabetes and CVD. Using lipidomics analysis, Rhee *et al.*

demonstrated that individual TGs responded differentially to insulin action or insulin resistance[42]. In addition, TGs of lower carbon number and double-bond content were associated with increased diabetes risk, whereas TGs of higher carbon number and double-bond content were associated with decreased diabetes risk. Rhee *et al.* suggested that using this information on heterogeneous associations of individual TGs with diabetes risk would improve risk assessments, compared to the classical approach of using TG as a single component. Arora *et al.*[43] showed that Roux-en-Y gastric-bypass surgery in insulin-resistant morbidly obese subjects caused diabetes remission, and those in remission 2 years after the surgery were distinguished from those not in remission by presurgery levels of TGs composed of long-chain FFAs, consistent with reduced diabetes risk associated with these TGs. Schwab *et al.*[44] reported that diet-induced weight loss reduced TGs containing saturated and short-chain fatty acids, which was accompanied by improved insulin sensitivity, but not those composed of long-chain PUFAs, consistent with differential regulation of distinct TG groups. The present data add to the evidence for differential regulation of TGs, especially between those composed of very-long-chain PUFAs and others composed of FFAs with lower chain length and double-bond content. More importantly, the present study shows for the first time that these distinct groups of TGs are regulated by interaction of gut bacteria and HNaLK diet.

TG cluster 2 identified in the present study included all those TGs (i.e., 56:9, 58:10, and 60:12) associated with decreased diabetes risk in the Rhee *et al.* study,[42] suggesting that TG cluster 2 may represent those associated with decreased diabetes or CVD risk. Cluster 2 TGs, whose 3 acyl chains were identified (Table 1.3), are characterized to contain very-long-chain PUFAs, such as 20:4, 20:5 (EPA), 22:4, 22:5, and 22:6 (DHA).

EPA and DHA are major constituents of fish oils known to mediate CV and metabolic beneficial effects of fish oils[45]. Because our animals consumed the same amounts of diets of identical compositions, the changes in these TGs may reflect those in the endogenous synthesis of very-long-chain omega-3 PUFAs. Thus, the present data may indicate regulation of synthesis of very-long-chain omega-3 PUFAs by gut bacteria and the HNaLK diet.

Very-long-chain omega-3 PUFAs, such as EPA and DHA, can be synthesized from α-linolenic acid (18:3), a major dietary omega-3 fatty acid derived from vegetable oils[46]. Although the endogenous capacity to synthesize these omega-3 PUFAs may be low in humans[47], it has been suggested to be associated with risks of metabolic syndrome, diabetes, and CVD[48]. Enzymes involved in the synthesis of very-long-chain omega-3 PUFAs are elongase and desaturases (∆-5 and ∆-6). In the randomized Finnish Diabetes Prevention Study on 407 overweight, middle-aged people with impaired glucose tolerance, very-long-chain omega-3 PUFA levels and ∆-5 desaturase activity were associated with lower incidence of type 2 diabetes during a follow-up of 11 years[49]. In a cross-sectional study of 1975 Tunis adults, PUFA levels and ∆-5 desaturase activity decreased in patients with metabolic syndrome[50]. Taken together, the present data suggest the intriguing possibility that gut bacteria and an HNaLK diet may interact to regulate the synthesis of very-long-chain omega-3 PUFAs, thereby impacting biomarkers of CVD.

The present study showed that unsaturated TGs and PCs were the lipid classes most impacted by gut bacteria (Table 1.1). This is consistent with a previous study by Velagapudi et al.[41] in germ-free mice, demonstrating that TGs and PCs are circulating

lipids that are the most affected by gut microbiota. However, the directions of changes observed without vs. with gut bacteria in this mouse study were opposite to those in the present rat study. This discrepancy may arise from differences in experimental conditions, and in particular, the fasting state; mice were studied after a 4-h fast in the Velagapudi study, whereas rats were studied immediately after feeding in the present study. Fasting may alter relative contributions of circulating chylomicrons and lipoprotein particles to circulating lipids, which may be relevant because the gut microbiota have been shown to have differential effects on circulating levels of chylomicrons and VLDL (or its synthesis)[41]. A more recent lipidomics study employing germ-free and antibiotic-treated mice (and their controls) demonstrated differential effects of the gut microbiota on circulating FFAs and PCs (and other glycerophospholipids) that depended on saturation levels of their fatty acids[51], which is consistent with the opposite effects of gut bacteria on TGs composed of FFAs with low (cluster 1) *vs.* high (cluster 2) saturation levels (Figure 1.11).

Antibiotic treatment, which depletes gut bacteria,[39] increased cluster 2 TGs associated with decreased diabetes risk, as discussed above, and decreased cluster 1 TGs associated with increased diabetes risk. These distinct effects of antibiotic treatment are expected to decrease diabetes risk in an additive manner. Ample evidence shows that the gut microbiota plays a major role in diet-induced obesity or metabolic syndrome[37,52]. Gut-bacteria regulation of circulating TGs may be an important mechanism underlying this relationship. In contrast, the "unhealthy" HNaLK diet (vs. control, low Na^+, high K^+ diet) decreased in both TG groups. It is rather unexpected that this diet decreased cluster 1 TGs associated with increased diabetes risk, as this effect would decrease diabetes

and CVD risks. However, it also decreased cluster 2 TGs composed of very-long-chain PUFAs, associated with reduced diabetes risk; the significance of this effect remains to be evaluated. It may be important to determine whether it is high Na^+ or low K^+ that exerts these effects. Both high Na^+ and low K^+ intake have been implicated in increased CV and metabolic risk[35,36]. Although these effects may be attributed to their effects on volume regulation or blood pressure, altered lipids may also be involved. If so, we may have discovered a novel mechanism by which high Na^+ and/or low K^+ intake may lead to CV or metabolic abnormalities.

The present study had several limitations. We assumed that the effects of antibiotic treatment were due to gut bacteria depletion. Our earlier study showed that, although the same antibiotic treatment resulted in > 95% reduction in overall bacterial biomass, certain gut bacteria increased their abundance after antibiotic treatment[39]. Therefore, we cannot attribute all effects of antibiotic treatment to gut-bacteria depletion. Also, we cannot exclude the possibility that some effects might arise from direct, off-target effects of antibiotics on lipid metabolism. In contrast, the same study showed that, although the HNaLK diet did not significantly alter diversity or composition of gut bacteria, it significantly altered specific species of gut bacteria, such as Porphyromonadaceae and Prevotellaceae, consistent with an apparent interaction of the diet with gut bacteria in altering circulating lipids. Whether any of these gut bacteria were responsible for some of the HNaLK effects on circulating lipids remains to be tested. Another limitation of the present study was that the sample size was small (n = 7 for each group), many large changes induced by treatments, especially by diet (Table 1.3), did not gain statistical significance. Also, the present exploratory study was not designed to elucidate the

mechanisms underlying the effects of antibiotic treatment and HNaLK diet on distinct groups of circulating TGs. Elucidating such mechanisms would lead to novel ways (e.g., using pro- and prebiotics) to manipulate circulating levels of TGs, especially those associated with diabetes and CVD risks (*i.e.*, cluster 2), and possibly reduce diabetes and CVD risks.

Despite these limitations, our analysis identified two TG clusters that are differentially regulated by the diet and gut microbiota. One cluster appears to represent the traditional group of TGs associated with increased risks of diabetes or CVD, whereas the other cluster appears to represent TGs associated with decreased risk of diabetes, suggested by a previous lipidomic study[42]. Thus, this study highlights the importance of lipidomic studies by identifying distinct groups of lipids with differential regulation, which would facilitate future studies delineating their individual roles in the pathogenesis of diabetes or CVD.

1.5.3 Metabolomics analysis of human serum from flaxseed dietary intervention

Flaxseed has been investigated as a dietary supplement for many years due to suggested health impacts such as reduced risk of cancers, atherosclerosis and diabetes[53–55]. Health effects reported from flaxseed supplementation studies, specifically in postmenopausal women include decreased total inflammation measured by biomarker C-reactive protein (CRP) levels[56], decreased total and low density lipoprotein cholesterol (LDL-C)[57,58], and decreased intensity of menopause symptoms[59,60]. The possible health effects of flaxseed consumption are important to consider for postmenopausal women who are at increased risk of chronic inflammation related diseases[61]. The evidence for health effects of flaxseed

consumption on humans is variable and contradictory. As stated, studies have shown that postmenopausal women supplemented with flaxseed or flaxseed extracts have shown decreased LDL-C, and CRP, while other studies have found that flaxseed supplementation has no effect on LDL-C[62] or CRP[63]. The health effects observed from flaxseed supplementation, in addition to discrepant findings in relation to flaxseed consumption have spurred continued research efforts to investigate the mechanism of action of flaxseed on humans. Bioactive components of flaxseed proposed to cause health effects include high dietary fiber content[64], high content of alpha-linolenic acid[65], and high concentration of lignans[66]. Flaxseed is the richest source of lignan in food, commonly used to increase dietary lignan, and frequently studied for health impacts of lignan[67].

Lignans are plant metabolites transformed by the gut microbiota to create enterolignans, which are absorbed by the gastrointestinal (GI) tract[68]. Inter-subject variation in the microbiota leads to highly variable capacity to produce enterolignans from lignan precursors[69,70], which is thought to be responsible for discrepancies between some flaxseed consumption studies[71]. Enterolignans are bioactive compounds with estrogen receptor binding capacity[72], antioxidant activity[73], and ability to regulate immune signaling[74]. Enterolignans have been investigated as a route of flaxseed bioactivity, however these microbe-produced metabolites alone do not fully explain the health effects observed from flaxseed[62]. An additional mechanism by which flaxseed may impart health effects is through regulation of bile acid metabolism[64,75,76]. The hypothesized effect and results of flaxseed consumption on bile acid metabolism is reported in chapter three.

A randomized crossover flaxseed intervention was conducted in a cohort of 179 healthy postmenopausal women from New York. Serum was collected from each subject before and after a 6-week intervention period when 10 grams per day of ground flaxseed was consumed (Figure 1.13). Study participants were provided with ground flaxseed. Subjects maintained their usual diet throughout the study with the exception of flaxseed supplementation.

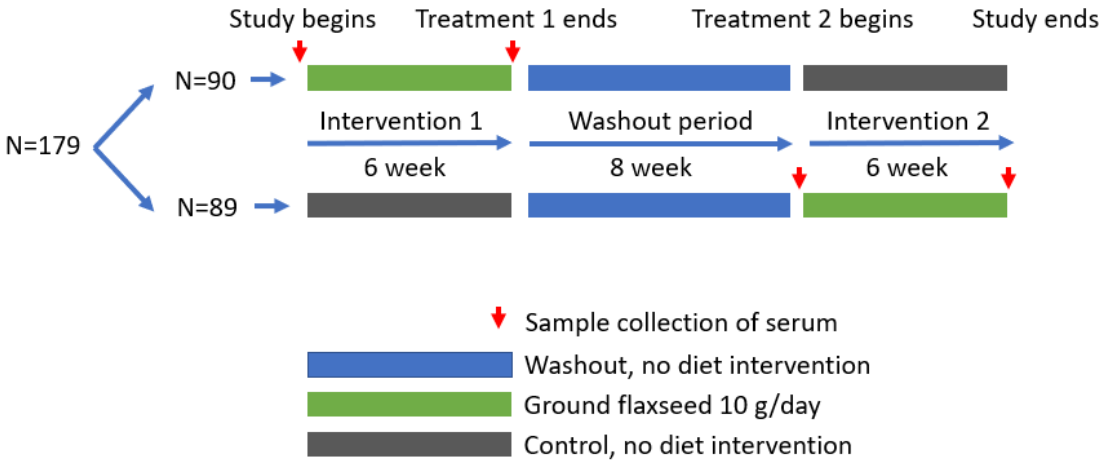


Figure 1.13. Flaxseed dietary intervention study design. The study timeline shows intervention periods in green (flaxseed supplementation) control diet periods in grey (no flaxseed supplementation), and washout periods in blue (no flaxseed consumption). Sample collections of serum for metabolomics analysis are shown with red arrows.

Metabolomics analysis of serum collected before and after flaxseed consumption, resulted in a total of 463 metabolites that could be annotated. Statistical analysis using a paired T-test revealed that seven of these metabolites were significantly different pre vs. post flaxseed intervention (FDR adjusted p value < 0.05). Four of these metabolites are

closely related to nitric oxide signaling[77] : (1) dimethylguanidino valeric acid (DMGV), found at lower levels post vs. pre-intervention, (2) dimethyl-arginine, observed at higher levels post vs. pre intervention, (3) arginine, also at higher levels post vs. pre intervention, and (4) citrulline, detected at lower concentrations post vs pre intervention. Citrulline is a product of nitric oxide synthesis, and lower citrulline after the dietary intervention suggest that nitric oxide signaling may be inhibited. Additionally, dimethylarginine is a known inhibitor of nitric oxide synthesis, and dimethylarginine was found at higher levels post- versus pre-flaxseed intervention, again indicating a stronger inhibition of nitric oxide signaling after dietary intervention, possibly caused by inhibition of nitric oxide synthase. Other significantly different metabolites included enterolactone glucuronide, which is a well-known biomarker of lignan consumption derived from flaxseed, and N-methylalanine which was increased post vs. pre flaxseed consumption. N-methylalanine has been reported as a microbially produced metabolite likely suggesting regulation of the gut microbiota by flaxseed and absorption through the intestinal tract.

Metabolomics analysis of these data resulted in two unconventional routes of metabolite annotation. The first of these was the annotation of enterolactone glucuronide. A chemical standard for enterolactone glucuronide was not available for analysis, and no publicly available library MS/MS spectra exists for this metabolite. The chemical standard for enterolactone, however, was analyzed using a HILIC method and has a retention time of 1.1 minutes. The MS/MS spectrum was also recorded for enterolactone. Enterolactone was not found in serum, as expected, since enterolactone glucuronide is the primary circulating form of enterolactone. However, there was a strong MS/MS spectrum match to enterolactone at a retention time of 6.15 minutes. Based on this spectral match, the

mass of enterolactone glucuronide was searched within the serum data and masses that matched two of the adducts of enterolactone glucuronide appeared with an identical retention time of 6.15 minutes. These results suggest that the original enterolactone MS/MS match at 6.15 minutes is an in-source fragmentation product of enterolactone glucuronide. In source fragmentation products are discussed in the final sections of this chapter. Based on these criteria enterolactone glucuronide was confidently annotated in this dataset, which serves as an important metabolite based on previous experiments of flaxseed. The presence of enterolactone glucuronide in serum was on average 17 times higher in the post-flax-seed intervention compared to the pre-flax-seed intervention serum samples, which is the most strongly influenced metabolite from these analyses.

The second unconventional metabolite annotation was the result of investigating statistically significant chromatographic features annotated as “unknowns”, meaning that no automated annotation was possible for these features when matching MS/MS spectra against comprehensive mass spectral libraries. An unknown chromatographic peak with RT of 7.32 and *m/z* of 202.1185 was statistically significantly decreased in post-intervention samples compared to pre-intervention samples. Information from this unknown was input to MS-FINDER[78], a software that predicts chemical structure based on MS1, and MS/MS information. The top result revealed an estimated structure for this unknown feature to be alpha-keto-dimethyl-delta-N, N-Dimethylguanidino valeric acid (DMGV), for which there were no publicly available library MS/MS spectra. The chemical structure of DMGV was submitted to SciFinder to find any previous mention of DMGV in literature. Previous studies had measured DMGV and collected MS/MS spectra from a chemical standard of DMGV[77,79]. The unknown spectrum under investigation and the

MS/MS spectrum collected from DMGV in this paper were very similar. The MS/MS spectrum of DMGV as previously published was in table form but not published in a public database, which is why automated annotation of this peak did not occur. The MS/MS spectrum for DMGV was transferred from the publication to the MassBank of North America so that future studies will now be able to annotate this metabolite using an automated approach. The finding that DMGV was decreased in response to the flaxseed consumption was integral to support the regulation of nitric oxide signaling by flaxseed consumption.

1.5.4 Takeaways from blood plasma and serum studies

Serum and blood plasma provide insight into biological effects of diet. These studies highlight the importance of considering both food related metabolites, as well as endogenously regulated metabolites in response to diet. Blood samples in the studies presented here revealed metabolites impacted by diet and linked to microbial metabolism in the intestinal tract , such as enterolactone glucuronide and N-methylalanine. Blood samples provide an extremely valuable aspect to investigate nutrition, since dietary or microbial metabolites that appear in the bloodstream represent metabolites that may have a direct impact on human health. This information is not included when analyzing only samples collected from stool or the intestinal tract. However, analysis of blood or urine samples do not provide information on where in the gut a metabolite is produced.

Another takeaway from analysis of blood plasma and serum is the necessity to consider well known phase I and phase II metabolism reactions that modify food metabolites. An enticing approach to nutritional metabolomics is finding metabolites present in plant

samples and finding the same metabolites in blood circulation. Popular examples of metabolites absorbed directly from food include resveratrol from wine[80], (-)-epichatechin from cocoa and tea[81], and secoisolariciresinol from flaxseed[82]. When searching for bioactive plant metabolites in the blood of humans, these metabolites are rarely in their non-transformed state but are modified by liver enzymes to their glucuronide forms, sulfate forms, or modified through other phase II transformation pathways[81,83]. Compounds that are absorbed into blood can be discovered by linking metabolites that appear in blood after food or beverage consumption. For example, enterolactone glucuronide is not present in flaxseed; however, the metabolite is directly linked to flaxseed consumption by microbial conversion, absorption, and liver conjugation. This example shows the necessity to consider products of phase I and phase II metabolism when looking for dietary metabolites in the bloodstream. As presented in the next chapter, these modified plant metabolites are also excreted in bile which contributes to the metabolome of the intestinal tract.

1.6 Chemical Profiling of Stool Samples

The GI tract is home to a large and diverse community of microbes, and also contains recently consumed foods, epithelium cells sloughed from the intestinal tract, and excretions into the GI tract including saliva, stomach juice, bile, and pancreatic juice. This mixture combines to form a complex assortment of chemicals as compared to any of the matrices discussed thus far. Stool is the most often used matrix used to investigate the human digestive tract. Stool, however, does not accurately represent what is occurring in the upper intestinal tract. Stool also does not portray information on the chemical profile at different regions of the intestinal tract. Stool is a non-invasive sample to collect, which

makes it a popular matrix to make educated guesses as to what is occurring in the intestinal tract. Analysis of stool will be considered in this section with respect to the flaxseed intervention study introduced in the previous section.

1.6.1 Metabolomics analysis of stool samples from flaxseed dietary intervention

Stool samples collected before and after the flaxseed dietary intervention introduced previously (Figure 1.13) were analyzed using non-targeted metabolomics analysis with a HILIC column. This analysis showed an increase in the known flaxseed metabolite secoisolariciresinol as well as enterolactone in the post-intervention samples compared to pre-intervention samples. Additionally, there were three other annotated metabolites with significantly (FDR adjusted p <0.05 from paired T-test) different abundances between pre- and post-intervention samples, which included tocopherol acetate, 2-aminophenol, and 2-ketoisocaproic acid. Tocopherol acetate is a form of vitamin E and was significantly higher post intervention compared to pre intervention. Vitamin E is present in a variety of foods, including flaxseed[84]. The difference in tocopherol acetate between pre- and post-intervention may be related to tocopherols from flaxseed being transformed in the gut to tocopherol acetate, or from other foods that contain tocopherol acetate consumed at different rates between pre- and post-intervention timepoints. 2-aminophenol is a known bacterial metabolite suggesting that the flaxseed intervention caused a shift in the gut microbiota of test subjects. The metabolite 2-ketoisocaproic acid is an intermediate in amino acid metabolism which may also be due to changed microbial metabolism. Connection to 16S rRNA sequencing performed on these same samples did not reveal clear connections between microbial and metabolite relative abundance.

Metabolomics analysis of stool and serum for this flaxseed intervention study were analogous. Non-targeted LC-MS/MS and GC-MS analysis led to annotation of of 463 unique metabolites in serum, and 616 unique metabolites in stool. Of the known metabolites annotated in stool and serum, 281 metabolites were the same, leaving 182 metabolites measured only in serum, and 434 metabolites measured only in stool. The higher number of annotated features in stool agrees with the higher number of features detected in the stool samples compared to serum samples during data processing (data not shown). Stool also presented a higher variance across samples, which is further discussed in chapter three with respect to targeted bile acid analysis performed on serum and stool from this study.

1.6.2 Takeaways from metabolomics analysis of stool samples

Based on the type, number and variation of chemicals detected in the stool, stool samples are most similar to upper intestinal tract samples considering all sample types reported above. This makes stool a valuable matrix to consider when considering the optimal analytical approach to measure the metabolome of samples from the lumen of the human intestinal tract. Stool samples are much more variable compared to the variation found in samples like blood plasma. The bloodstream is tightly regulated to maintain homeostasis throughout the body, while stool does not have similarly tight regulation. Stool chemical variation is also highly dependent on diet, and in the study performed here, subjects did not have a controlled diet, causing variation due to different foods consumed. To address this variation, parameters were optimized to avoid aligning peaks with different retention times, and critical consideration of the possibility of aligned peaks representing different

metabolites was necessary during metabolite identification. Data processing considerations for metabolite identification are further discussed in the next section.

1.7 Considerations for metabolomics analysis of human intestinal tract samples

The quality of data resulting from comprehensive chemical analysis of biological samples depends on many factors. Experimental design and sample collection determines whether biologically relevant conclusions can be made assuming perfectly analyzed chemicals. Reproducible sample preparation prior to analytical analysis is important to avoid technical error. Data collection using MS is important with respect to sensitivity, and more importantly reproducibility to produce quality comprehensive chemical analysis. Sensitivity of the instrument affects the number of metabolites that will be detected. Reproducibility in MS signal intensity is another point where technical variation can be introduced. Reproducibility in retention time is critical for performing accurate peak alignment between samples and for accurate automated data processing steps. During sample preparation and data collection the amount of sample that is analyzed determines the level of ion-suppression during data acquisition. Data processing is the final step towards forming an accurate and reliable metabolomics dataset. Most steps of MS data processing can be automated including peak picking, peak alignment, and preliminary annotations of metabolites through use of accurate *m/z,* retention time, and MS/MS spectral matching. Thousands of chromatographic peaks appear from biological samples and automated data processing software inevitably reports false positive metabolite annotations. Manual curation of metabolite annotations and critical consideration of all lines of evidence from raw MS data is the only approach to produce an accurate and reliable metabolomics dataset. Considerations for metabolomics analysis of samples

from the lumen of the human intestinal tract will be presented with a focus on ion suppression and metabolite identification.

1.7.1 Ion suppression in LC-MS/MS analysis of GI tract samples

Ion suppression is a concern in LC-MS experiments because the recorded ion abundance of a specific chemical can be impacted by the presence and abundance of other co-eluting chemicals, or the total amount of chemicals being ionized at a specific time[85,86]. Ion suppression is a type of matrix effect that is most impactful when there is a high concentration of metabolites flowing through the ionization source. Ion suppression is the result of a decrease in the efficiency of ionization of metabolites being measured. Ion suppression causes issues when comparing the ion intensity of metabolites between different samples. Different samples have different total ion profiles at specific retention times that can lead to different ion intensities measured, even if the actual concentration of a specific metabolite is identical between samples. Ion suppression is an inherent characteristic of electrospray ionization. Ion suppression can be corrected for through use of isotopically labeled standards for the specific chemicals being analyzed. By taking the ratio of the non-labeled chemical of interest to the isotope labeled chemical and comparing this ratio to analysis of known concentrations of chemical standards, the ionization efficiency can be accounted for. The ratio of non-labeled to labeled signal takes advantage of the same proportion of suppression occurring to both chemical species. This procedure of accounting for ion suppression can be performed for the number of isotopically standard and non-labeled standard pairs included in an analysis, and this technique is classically used in targeted metabolomics. However, when performing non-targeted metabolomics this technique faces severe limits because (a) many chemicals

that will be measured are not known *a priori*, and because (b) many compounds are not commercially available as isotopically labeled standards. Furthermore, using potentially hundreds of isotopically labeled compounds is more costly and takes significant time to correctly find and align all pairs of isotopically labeled and non-labeled chemical standards Alternatively, a smaller number of isotope labeled standards was used in my non-targeted metabolomics studies presented here for quality control and to assess the degree of ion-suppression effects during sample analysis.

Due to the shortcomings of correcting for ion suppression in non-targeted metabolomics experiments, it is best to mitigate the effect of ion suppression effects as much as possible. The most effective approach to reduce ion suppression is to analyze samples with relatively low total ion abundance since this effect is most influential at high total ion abundances. Low ion intensity flow can be achieved by analyzing a smaller amount of sample. The improved reproducibility of peak height intensity in isotope labeled internal standards was observed during analysis of the minimal genome JCVI-syn3A samples. These samples had a relatively low cell count per sample causing low intensity of total ions and low variation as compared to stool samples which had higher ion abundance and increased intensity variation in internal standards. GI tract lumen samples have high variation similar to stool. Variation in the lumen of the GI tract is due to temporal variation based on fasted/ fed state, in addition to different chemical profiles of diets consumed during sample collection. For example, some samples have high concentrations of lipids if a high fat meal was consumed before sample collection, but at different times or locations there are no dietary lipids present. There is a concern under these circumstances because the total ion abundance is very different between samples leading

to effects of ion suppression. For this reason, it is important to determine the optimal amount of sample to analyze to result in minimal ion suppression, while still measuring low abundant ions.

To find the optimal sample injection volume, many samples were screened to encompass the variability of samples. The highest abundant samples were then optimized to be at the upper limit of the desired total ion abundance. The injection volume was optimized in this fashion, and as result, 3 µL of resuspended sample (corresponding to 0.018 µL of gut lumen sample) was found to result in sufficient signal, without resulting in excessive ion suppression. Ion suppression effects were assessed by analyzing ion intensities of deuterium labeled internal standards D3-1-methylnicotinamide, and D9-choline which are present at the same time as highly abundant bile acids during HILIC analysis. In a select GI tract sample set there were four samples with very high bile acid concentration, and the other 16 samples had very little bile acid signals. D3-1-methylnicotinamide and D9-choline decreased by ~10-20% due to ion suppression. Bile acids in GI tract samples are one of the most highly abundant and variable metabolites observed, thus a maximum of ~20% technical variation at time points of high ion current is present in data collected from GI tract samples. These time points of high ion currents represented a small fraction of the total chromatographic time. Therefore, variation due to ion suppression was considered to be much less dramatic for a majority of the detected metabolites than for bile acids. While the degree of signal intensity variation due to ion suppression might be even further reduced if less sample was analyzed, low abundant metabolites would then be difficult to detect. Hence, when weighing metabolome comprehensiveness versus potential ion suppression effects, this observed degree of variation was accepted for the

benefit of increasing the signals of low abundant metabolites. Effects of ion suppression may be further reduced through data normalization such as SERRF[87] but should still be considered when reporting statistically significant results.

1.7.2 Justifiable evidence-based metabolite annotations in LC-MS/MS data

Data processing of non-targeted LC-MS/MS metabolomics experiments can be automated with accurate results to produce a list of chromatographic features detected in samples. Such automation requires high data quality with minimal retention time shifts or *m/z* shifts. Annotation of these peaks to specific metabolites is the next step in the metabolomics workflow. Metabolite annotation in non-targeted LC-MS/MS metabolomics cannot currently be automated to give sufficiently accurate metabolite annotations to result in trustworthy biological conclusions. Evidence including *m/z*, retention time and MS/MS spectra must be thoughtfully considered to accurately annotate features to known metabolites. Further evidence such as biological likelihood, and *in silico* modeled evidence can also be used to reduce incorrect annotations.

The first piece of evidence that should be considered in every metabolite annotation is the *m/z* measured through accurate mass, high resolution mass spectrometry. The *m/z* from experimental data should match within a defined window, generally accepted as <5 ppm or <3 mDa, to the theoretical *m/z* for a metabolite. The theoretical *m/z* can be calculated based on the chemical formula and given charge/ adduct information from known metabolites. Experimental *m/z* values should match within a defined window which depends on the type of mass analyzer being used. One way that an appropriate window can be estimated is by analyzing known chemicals and observing the differences in

theoretical and experimental *m/z* values. With sufficient measurements, a 95% confidence interval can be calculated for the *m/z* differences and this used as a window for peak annotation. For the case of an orbital ion trap operated with 60,000 resolving power, the difference between theoretical and experimental *m/z* values should be less than 1.5 mDa for small molecules <300 Da. This value may differ depending on the calibration of the instrument. A *m/z* match alone is insufficient to reliably annotate a metabolite. Even for small metabolites there are many possible isomers that can match a single *m/z* measurement, so additional evidence is needed to annotate features with reasonable confidence.

Another piece of evidence acquired during LC-MS/MS analysis is fragmentation spectra. During analysis chemicals are fragmented, usually by colliding ions with gas, resulting in fragmentation spectra (MS/MS spectra). Chemicals fragment reproducibly for a given type of mass spectrometer and under identical instrument parameters. This means that MS/MS spectrum from a chemical standard will fragment the same as the same chemical in a biological sample. Matching MS/MS spectra acquired from biological samples to a library of known MS/MS compounds provides a valuable piece of evidence towards annotating a feature in experimental data to a known metabolite. Algorithms are used to calculate MS/MS matches between experimental spectra and library MS/MS spectra. Many algorithms exist to calculate how well two spectra match; however, none of them reliably convey the quality of evidence that an MS/MS match contributes towards annotating a feature. Algorithms give high match scores to two spectra sometimes even if the two spectra do not match well. This is commonly observed when matching complex biological samples to large MS/MS libraries[88]. For this reason, manual assessments of

each MS/MS match should be performed. Furthermore, different metabolites can have the same fragmentation spectra. One example of this is leucine and isoleucine producing identical MS/MS spectra. Although these are two unique chemicals, they cannot be distinguished based on their fragmentation spectra but require retention time information as well. This is a common occurrence, and it is frequently observed that two or more unique metabolites from MS/MS libraries will have strong matches to a spectrum obtained in a biological sample. To identify situations where one experimental MS/MS spectra matches multiple library MS/MS spectra all possible library MS/MS matches to an experimental MS/MS spectrum should be manually reviewed. Unique library metabolite matches to an experimental MS/MS spectrum provides a strong line of evidence and increases confidence in annotating a feature to a metabolite. Unfortunately, it is still possible that other chemicals that are not present in MS/MS libraries may produce that same spectrum. Thus, consideration of false positives in MS/MS matching is an important consideration, and additional evidence is important to further increase confidence in metabolite annotations.

Retention time is therefore another crucial piece of information in metabolite identification. Retention time is the amount of time it takes for a chemical to pass through the LC system and be measured by MS. The retention time for a given chemical is reproducible when LC conditions are kept the same. This information is valuable since chemical standards can be analyzed and the retention time recorded in a retention time library. It is important to consider experiment specific RT deviation from a retention time library. Retention time on a single instrument can shift over time due to column age, column lot number, mobile phase batch, ambient temperature, and other experiment specific conditions. For GC, a

“retention index” is used to normalize retention time using internal standards to account for RT shifts between experiments. An LC specific retention index is not available, but the RT library can be “corrected” in an analogous manner using the 21 internal standards analyzed alongside RT library and experimental samples as reported previously[15]. In this thesis work no correction was performed if deviation of retention times for internal standards in experimental data was <0.01 minutes from the library retention times, but with RT deviation >0.01 minutes, a 3rd degree polynomial was fit between library and experiment specific retention times for internal standards and used to correct the library retention times for each specific experiment. Matches in *m/z* and retention time to a known chemical provides a valuable piece of evidence for metabolite annotation. Retention time as evidence for feature annotation also has shortcomings. One shortcoming is that different metabolites can present the same retention time and if they are isomers then their *m/z*-RT information is identical. For example, the isomers 3-methyluridine and 5-methyluridine have the same *m/z* and the same retention time when using the HILIC LC-MS/MS method used for analysis in this chapter. This fact highlights the need to use further evidence to annotate features in LC-MS/MS data.

Retention time can be used to prevent mis-annotation of metabolites. For example, if a metabolite with a known retention time is annotated based on MS/MS spectral matching to a library spectrum, but the retention time does not match the known retention time for that metabolite then the MS/MS match would be nullified. Unfortunately, only a small portion of metabolites have known retention times since it is costly to purchase and analyze chemical standards. It is however possible to predict retention times using machine learning methods if a sufficiently large retention time library is available. Using

a library of over 900 metabolites with known retention times for the aforementioned HILIC method, it is possible to predict other metabolite retention times with accuracy of approximately one minute[89]. This approach was recently published, and the scripts are freely available[89]. While predicted retention times are not sufficient to confirm metabolite annotations, they are highly useful to nullify poor metabolite annotations. When the predicted retention time is far away from the experimental retention time for a metabolite annotation (greater than 1 minute when using this HILIC method), confidence in the annotation is reduced. In this manner, false positives can be flagged. When I investigated such apparently poor retention time/ MS-MS annotations, I commonly found these poor-quality annotations to be caused by in-source fragmentation of a different metabolite such as the glucuronide food compound derivative explained above.

In-source fragmentation occurs during electrospray ionization (ESI). In-source fragmentation is the process of metabolite fragmentation in the ESI source before entering the MS. Commonly, the non-fragmented and fragmented ions from a given metabolite will both be detected, leading to two chromatographic features being generated from a single metabolite. In-source fragments are a common occurrence in LC-ESI-MS/MS analyses and frequently cause incorrect annotation of metabolites[90]. In-source fragments must be identified and filtered out of the dataset to avoid reporting incorrect metabolites which could lead to incorrect conclusions made in an experiment. There is not currently an automated method to accurately identify and filter out in-source fragments from non-targeted metabolomics data. There is software that facilitates the identification of features that are in-source fragments, but manual curation of these matches is still necessary. The primary method to identify in-source fragments is based

on retention time and correlation in intensity between features. The retention time of an in-source fragment and its non-fragmented parent metabolite are nearly identical because they are both derived from the same metabolite eluting from the LC column. The intensity of in-source fragment and its parent metabolite also correlate nearly perfectly (correlation coefficient of 1.0) across samples since the parent metabolite fragments at the same proportion in all samples. These two characteristics can be used to flag possible in-source fragments. However, adduct ions also form in the ion source from a single metabolite and show similar relationships. Adduct ions form when different charged metabolites associate with the parent metabolite. For example, ions can be formed from a metabolite with a proton (H+) or a metabolite with an ammonium (NH4+) ion. These two adducts will have the same retention time and a different *m/z*. Adducts also usually correlate very strongly. Two features that have the same retention time and correlation very strongly could be in-source fragments (for example a metabolite with loss of NH3), or it could represent a pair of two adducts (for example [M+H]+, and [M+NH4]+ species). In both these examples, the mass difference between the two features would be exactly the same. In this case software is unable to distinguish between in-source fragments and adducts. To identify in-source fragment pairs in a dataset, the correlation of peak intensities across samples should be checked for features at similar retention times. These features should then be investigated to determine whether there is an in-source fragment present, or the highly correlating features are adducts.

The biological likelihood of a metabolite being detected in a sample also contributes to the confidence of metabolite annotations. This line of evidence is of particular importance when using large retention time and MS/MS libraries. The prevalence of isomers in

biological samples cause false positive metabolite annotations. Examples of clearly false positive metabolite annotations are some rarely observed drugs supported by only a single line of evidence (*e.g.,* only a *m/z* – retention time match). Many drug chemicals are present in retention time and MS/MS libraries, including drugs that have been prohibited by the FDA for decades. When these synthetic drugs are annotated in a sample without a biologically likely cause, it is much more likely that the feature represents an isomer of this drug and is in fact a false positive metabolite annotation. Conversely, if a peak appears at very high abundance in a sample matrix where the metabolite is biologically expected to be present, for example for urobilin in GI tract samples, such annotations have increased confidence for the accuracy of compound identifications. Unfortunately, using biological likelihood assessments is less useful when investigating GI tract samples due to the high diversity of food and drug exposure in humans. At least for blood, use of *a priori* knowledge is a useful approach to generate the most accurate possible metabolomics dataset[91].

In summary, automation of LC-MS/MS non-targeted data processing performs many steps of the data processing workflow accurately and reproducibly. However, automation of feature annotation cannot yet be trusted to give correct results but requires expert curation. For example, in a dataset from GI tract samples (discussed in the next chapter), there were a total of 665 features annotated to known metabolites. After manual curation, considering all the evidence discussed above, 382 (57.4%) of the annotations showed sufficient evidence to provide a confident metabolite annotation. This process is unfortunately dependent upon human input, and thus is not broadly reproducible and depends upon the expertise of the data curator. Overall, therefore, automation of

metabolite identification will overcome the challenge of feature annotation, but manual curation by a knowledgeable data curator is currently required to result in an accurate and trustworthy metabolomics dataset. For GC-MS data reported in this thesis, all metabolites were curated on a database level over the past 15 years of use at UC Davis[92], but for LC-MS/MS data, no such data repository has been finalized. Necessary improvements to increase automated metabolite annotation accuracy include better metrics to determine MS/MS match quality, consideration of non-unique MS/MS matches (MS/MS false discovery rate), identification of in-source fragment ions, increased size of retention time libraries, improved accuracy of retention time prediction, automated assessment of biological likelihood of a metabolite being present in a sample, and ultimately integration of all of these lines of evidence to result in a meaningful metric for the likelihood of metabolite identification.

1.8 Chapter 1 methods

1.8.1 Minimal Genome Metabolomics analysis methods

Cell pellet samples were fully randomized through sample extraction, analytical analysis, and data processing. 250 µL of 3:3:2 (isopropanol/acetonitrile/water) (v/v/v) was added to the original tube prepared for metabolomics analysis, vortexed vigorously for 30 seconds, sonicated for 1 minute, and all contents were transferred to a new clean 2 mL round-bottomed microcentrifuge tube. An additional 250 µL of 3:3:2 was added to the original tube, vortexed, sonicated, and also transferred to the round-bottomed 2 mL microcentrifuge tube to maximize recovery. Two three mm stainless steel balls were added to the round bottom microcentrifuge tube and processed in a Geno/Grinder tissue

homogenizer (Spex SamplePrep; Metuchen, NJ) at 1500 rpm for 1 minute. The tube was then centrifuged at 14000 *rcf* for 3 minutes, and 450 µL of supernatant was transferred to a clean 1.5 mL microcentrifuge tube that was dried under vacuum overnight. Dried samples were stored at -80 °C until analysis. Eight method blanks were generated by starting with 50 µL of methanol and no cell pellet, and these blanks were extracted at times evenly distributed throughout all experimental sample extractions. A pooled quality control (QC) sample was created by transferring 50 µL of the residual supernatant after centrifugation and sample removal, to a clean 5 mL centrifuge tube. This pooled QC was vortexed thoroughly to mix, and 450 µL of the QC mix was transferred into eight 1.5 mL microcentrifuge tubes, dried and stored in the same manner as samples. The LC-MS grade water, acetonitrile, and methanol used in metabolomics analysis were purchased from Fisher Scientific (Waltham, MA), and isopropanol, formic acid, and ammonium formate were purchased from Sigma-Aldrich (St. Louis, MO).

Prior to analytical analysis, dried samples were resuspended in 100 µL 8:2 (acetonitrile/ water) (v/v) with internal standards (42 isotopically labeled or synthetic compounds), vortexed for 30 seconds, sonicated for 1 minute, centrifuged at 14000 rcf for 2 minutes, and transferred to amber screw top vials with micro-inserts for LC-MS/MS analysis. Analytical analysis was performed using a Vanquish Focused UHPLC coupled to a Q-Exactive HF mass spectrometer (ThermoFisher Scientific) as previously described[93]. In summary the analysis used a Waters Acquity UPLC BEH Amide column (150mm x 2.1 mm id, 1.7 µm particle size) coupled to Acquity UPLC BEH Amide VanGuard precolumn (5 x 2.1 mm; 1.7µm) (Waters, Milford, MA) with mobile phases of LC-MS grade water, and 95:5 (v/v) acetonitrile/water, each at 10mM ammonium formate and 0.125% formic

acid. Mass spectra were collected in data dependent mode with the top four ions from each MS1 scan being selected for MS/MS fragmentation. Electrospray ionization analysis was conducted using a spray voltage of positive 3kV and negative 3 kV as separate analyses. Three microliters of sample were injected for positive mode ionization, and 5 µL injected for negative mode analysis. Blanks and QC samples were analyzed at even intervals spaced throughout all MS analyses. Additionally, a pooled quality control sample was measured with a mass scan range window of 100 m/z sequentially 9 times to cover m/z 60-900 in order to increase data dependent MS/MS spectra coverage for use in compound identification.

Data was processed using open source software MS-DIAL[3] version 3.70. MS-DIAL performed baseline correction, deconvolution, peak detection, alignment, gap filling, adduct identification, m/z-RT library matching, and MS/MS library matching. An in-house m/z-RT library developed from authentic standards was matched to peaks identified by MS-DIAL, and MS/MS spectra were matched to library spectra from the Mass Bank of North America (MoNA; https://massbank.us), and NIST17. Manual inspection of each annotated compound was conducted to confirm m/z-RT library match, and/or MS/MS library match. Peak height was reported as intensity for each annotated metabolite in each sample. Six outliers were removed as determined by deviation of the summed intensity of all internal standards, or the summed intensity of all annotated known compounds by more than two standard deviations in either positive or negative mode ionization analysis which are expected to be failed injections via autosampler error. Calculation of ANOVA was carried out for all features comparing genotypes (n=7, 8 or 9), and multiple comparisons were accounted for using Bonferroni correction.

1.8.2 N^6-formylated lysine in bacterial protein digest analysis methods

Chemicals

N^6-Formyl lysine (>98.0% purity) was purchased from TCI Chemicals (Catalog No. F0136). L-Lysine (>98%) (Product 616214) and L-lysine-3,3,4,4,5,5,6,6-d_8 (lysine-D_8) (Product L5501) (98%) were purchased from Sigma-Aldrich. N^2-Formyl lysine (>98%) was purchased from Iris Biotech GmbH (Product XAA1330). Additional internal standard information is listed in Supplementary Table 1. Formaldehyde (37% *w/w* solution in water) containing 10-15% methanol as preservative was purchased from Fisher Scientific (Catalog No. BP531-500). Water and acetonitrile were LC-MS grade and purchased from Fisher Scientific. All other chemicals were purchased from Sigma-Aldrich.

Sample preparation

LC run solvents used for sample resuspension were prepared using a stock LC solvent solution of 4:1 acetonitrile/ water (*v/v*) with 20 isotopically labeled internal standards including lysine-D_8 combined from individual stock solutions as described in Supplementary Table 1 (Supplementary Table 1). Internal standards were used to monitor analytical variance and retention time stability. This stock run solvent was used to prepare the spiked LC run solvents used for this experiment. The spiked LC solvents had an additional 1 µg/mL lysine, 5 µg/mL lysine, 1 µg/mL lysine-D_8, or 5 µg/mL lysine-D_8 made by diluting authentic standards of lysine or lysine-D_8 in 4:1 acetonitrile/water (*v/v*), drying standard solutions under vacuum (vacuum centrifuge), and dissolving standards with stock run solvent to reach desired concentrations. Calibration curves of lysine and N^6-formyl lysine were prepared through dilutions of authentic standards in 4:1

acetonitrile/water (*v/v*), drying under vacuum, and dissolving in stock run solvent to achieve desired concentrations. Final concentrations of N^6-formyl lysine were 5 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL, and 500 ng/mL, and final concentrations of lysine were 0.5 µg/mL, 1 µg/mL, 5 µg/mL, 10 µg/mL, and 20 µg/mL. All prepared samples were transferred to amber LC-MS vials with glass micro-inserts and split top caps and stored in autosampler at 4 °C until analysis.

To prepare sample matrix, proteins were extracted from *E. coli* as previously described.[25] Briefly, cells of various genotypes were grown in M9 minimal medium or M9 medium containing 1 mM formaldehyde. Cells were pelleted by centrifugation, pellets were suspended in 1 mL PBS buffer, sonicated, and centrifuged to clear. Soluble proteins in the supernatant were precipitated with 80% acetone and pelleted by centrifugation at 5000 *g* for 3 min at 4 °C. Protein pellets were air-dried and stored at -20 °C until subsequent steps. Protein pellets were dissolved in 0.2 mL of 1 M NaOH, combined with 0.4 mL of 100 mM ammonium bicarbonate buffer, pH 8.5, desalted on a PD-25 column, eluted with 100 mM ammonium bicarbonate buffer, pH 8.5, and protein content was measured. *Streptomyces griseus* protease was added at mass ratio of 1:10 (protease/protein) and incubated at 37 °C for 16 h. Protein digests were dried under vacuum, and stored at -20 °C.

Dried protein digests were dissolved in 0.5 mL 4:1 acetonitrile/ water (*v/v*), vortexed, and sonicated for 2 min. From each sample tube, 0.02 mL was combined into one clean microcentrifuge tube to create a pooled sample, vortexed vigorously for 30 sec and aliquots of 0.05 mL were transferred to clean microcentrifuge tubes. These identical pooled samples were dried under vacuum and stored at -20 °C until resuspension for LC-

MS/MS analysis. Pooled samples were dissolved in stock run solvent, and stock solvent with addition of 1 µg/mL lysine, 5 µg/mL lysine, 1 µg/mL lysine-D_8, or 5 µg/mL lysine-D_8 (*n*=3). These microcentrifuge tubes were vortexed vigorously for 30 sec, sonicated for 2 min, centrifuged at 16000 *g* for 2 min, transferred to amber LC-MS vials with glass micro-inserts, and stored in autosampler at 4 °C until analysis. Unless otherwise stated, sample preparation was carried out at room temperature without protection from ambient light.

For fraction collection, the UHPLC eluent within 10 sec of the lysine chromatographic peak was collected from five injections of 10 µg/mL authentic lysine standard. This fraction was dried under vacuum, dissolved in 0.05 mL stock run solvent, vortexed vigorously for 30 sec, sonicated for 2 min, centrifuged at 16000 *g* for 2 min, and transferred to an LC-MS vial with a glass micro-insert and stored at 4 °C until LC-MS/MS analysis.

LC-MS/MS data acquisition

LC-MS/MS analysis was carried out using a Vanquish UHPLC (ThermoFisher Scientific) coupled to a QExactive HF^+ orbitrap mass spectrometer (ThermoFisher Scientific). LC separation used a BEH amide hydrophilic interaction LC column (Waters Acquity UPLC BEH Amide, 150-mm length, 2.1-mm inner diameter and 1.7-µm particle size) with a guard column (Acquity VanGuard BEH Amide pre-column, 5-mm length, 2.1-mm inner diameter and 1.7-µm particle size). The column compartment and mobile phase preheater were held at 45 °C, and mobile phase flow rate was set to 0.4 mL/min. Two mobile phases were used; mobile phase A was water, and mobile phase B was 95:5 acetonitrile/water (*v/v*). Both mobile phases were modified to 10 mM ammonium formate, and 0.1% formic acid. The mobile phase gradient started at 100% mobile phase B from

0 to 2 min; brought to 70% B between 2 and 7.7 min; went from 70% to 40% B between 7.7 and 9.5 min, was held at 40% B until 12.75 min; brought back to 100% B between 12.75 and 12.85 min, and held at 100% B until 17 min. Injection volume was 0.003 mL. The injection needle was washed for 10 sec before and after each injection with 1:1 acetonitrile and water.

Positive mode electrospray ionization used a spray voltage of 3.6 kV, capillary temperature of 320 °C, sheath gas flow rate of 60 units nitrogen, and auxiliary gas flow rate of 25 units nitrogen. Data was collected from 0 to 15 min of the LC gradient using data dependent acquisition DDA with the top four ions from each MS1 scan being selected for MS/MS fragmentation which had a selection window of +/- 1 *m/z* of the precursor ion and an exclusion time of 2 sec for each ion selected for fragmentation. MS^1 spectra were collected with resolving power setting of 60,000, and MS/MS spectra collected at resolving power setting of 15,000. DDA MS/MS was acquired with normalized collision energy of 20%, 30% and 40%. All spectra were stored in centroid, ".raw" format. Automatic gain control was 1E6 for MS^1 and 1.5E5 for MS/MS.

Data analysis

Data was converted from ".raw" format to ".abf" format using the Analysis Base File converter and processed through open source software MS-DIAL (version 4.16)[94,95] with parameters described previously.[96] MSDIAL performed peak picking, peak smoothing, peak alignment, deisotoping, gap filling, and automated matching to MS/MS and retention time libraries. LC-MS features were identified using spectral matching to all positive mode LC-MS/MS spectra from the Massbank of North America (MoNA) and NIST17. Retention time matching used experimental accurate masses and retention times from authentic

chemical standards analyzed under identical LC conditions and can be found as a library on MoNA labeled “Fiehn HILIC”. Isotopically labeled and non-exogenous internal standards from the stock run solvent were used to assess injection quality. Peak height was used as spectral intensity for all data analysis. The coefficient of variation across all isotopically labeled internal standards was less than 4.3% across all samples containing sample matrix. One sample was removed due to a difference in internal standards intensity greater than three standard deviations compared to the rest of samples suggesting an injection failure. A linear model was fit to external calibration curve to quantify N^6-formyl lysine.

1.8.3 Rat lipidomics analysis methods

Male Wistar rats weighing 280–300 g (approximately 9 weeks old) were obtained from Envigo Laboratories (California, USA). Animals were individually caged under controlled temperature (22 ± 2 °C) and lighting (12-h light, 6 AM–6 PM; 12-h dark, 6 PM–6 AM) with free access to water and standard rat chow or experimental diets (see below). All procedures involving animals were approved by the Institutional Animal Care and Use Committee at the University of Southern California.

Animals were fed for 1 week with diets containing different amounts of Na^+ and K^+ (n = 7 for each diet; a total of 14 rats). The diets were prepared from K^+-deficient powdered rat diet (TD.88239.PWD; Envigo Teklad) with or without supplementation with KCl or NaCl[39]. The control ("healthy") diet contained 0.29% Na^+, which is the level in normal rat diets, and 2% K^+, which is higher than the normal 1%. In the HNaLK diet, Na^+ content was increased from 0.29% to 0.79%, a level similar to those in Western diets[97], and K^+ content

was decreased to 0.1%, which was selected to be low, but not low enough to deplete plasma K^+[98]. The diets were prepared by dissolving 30 g agarose in 500 mL of deionized water, and adding 500 g of powdered diet supplemented with KCl and NaCl, and heating to effect gellation[99]. After one-week of feeding, animals were anesthetized with isoflurane at ~7 AM, and blood samples for lipidomics analysis were collected using a heparinized syringe through the abdominal aorta. We chose this time for blood sampling because the interaction of the diet with the gut microbiota can be better observed immediately after the feeding period than in a fasting state. Blood samples were rapidly spun, and plasma was isolated and frozen immediately in liquid N_2. The plasma samples were stored at -80 °C until analysis.

This feeding experiment was also conducted in animals treated with antibiotics to deplete gut bacteria, as previously described[39]. Animals were maintained on drinking water (autoclaved tap water) containing vancomycin, metronidazole, neomycin, and ampicillin (0.5 mg/mL for vancomycin and 1 mg/mL for the others) for a week. During the antibiotic treatment, food intake and weight gain decreased during the initial 2-3 days but returned to normal values thereafter. After a one-week treatment, fecal DNA content decreased to 3% of control (data not shown), indicating that the antibiotic treatment was effective to remove most gut bacteria. After the antibiotic treatment, the animals were fed either the control or the HNaLK diet (n = 7 for each diet; a total of 14 rats) for an additional week, while being continuously treated with antibiotics. After the feeding, blood samples were collected and processed, as described above. These experiments in antibiotic-treated rats were conducted simultaneously with those in untreated rats using the same batch of rats. Thus, the present study employed a 2×2 factorial design (i.e., 4 groups; n = 7 for

each group; a total of 28 rats) to test the effects of the HNaLK diet and antibiotic treatment on circulating lipids.

Plasma samples (*n* = 7 for each group; a total of 28 samples) were extracted as previously described[100]. Briefly, 40 uL of plasma was extracted using a biphasic extraction method with methanol (containing 13 internal standards), water, and methyl *tert*-butyl ether (MTBE), adapted from the study of Matyash *et al.*[101] Three quality-control samples of pooled human-blood plasma (BioIVT; Westbury, NY), and three method blanks (solvents added to empty tubes) were extracted and analyzed alongside the samples. The non-polar MTBE portion of extract was dried under vacuum and reconstituted in 0.11 mL methanol/toluene (9:1, *v/v*) containing the internal standard 12[[(cyclohexylamino)-carbonyl]amino]-dodecanoic acid. Non-polar lipids were analyzed on a Vanquish Focused UHPLC coupled to a Q-Exactive HF mass spectrometer (ThermoFisher Scientific). Liquid chromatography was performed using a Waters Acquity UPLC CSH C18 column (100mm x 2.1mm, 1.7 µm particle size) coupled to an Acquity UPLC CSH C18 VanGuard precolumn (5 x 2.1 mm; 1.7 µm) (Waters, Milford, MA) with mobile phases of 60:40 (*v/v*) acetonitrile/ water, and 90:10 (*v/v*) isopropanol/acetonitrile, both of which had 10 mM ammonium formate and 0.1% formic acid for positive ionization analysis, and 10 mM ammonium acetate and 0.1% acetic acid for negative ionization analysis. Three and five microliters of reconstituted sample were injected onto column for positive and negative ionization analysis respectively. The mobile phase gradient was identical to that of Cajka *et al.*[100] Data were collected from 120 to 1200 m/z in a data dependent manner, with the top four ions from each MS1 scan being selected for MS/MS fragmentation.

Samples were analyzed in a randomized order with a method blank and quality control plasma sample analyzed throughout analyses.

Data were processed using open-source software MS-DIAL version 2.82[3]. Raw data is available on the Metabolomics Workbench (https://metabolomicsworkbench.org/) (Project doi: 10.21228/M8Z090). MS-DIAL performed baseline correction, deconvolution, peak detection, alignment, gap filling, adduct identification, accurate mass/ retention time (m/z-RT) library matching, and MS/MS library matching. MS-DIAL parameters were set to 10 for minimum peak width, 10000 for minimum peak height, MS/MS fragmentation match to library of at least 80%, MS1 tolerance of 0.005 Da, MS2 tolerance of 0.01 Da, smoothing level of 3, and retention time window of 0.15 minutes for retention time matching to an in-house m/z-RT library. Tandem MS spectra were matched to library spectra from the Mass Bank of North America (MoNA), NIST17, and LipidBlast[102]. Features that appeared in method blanks were removed from further investigation. Duplicate peaks, isotopes, and adducts were investigated and removed using Mass Spectral Feature List Optimizer[103]. Manual inspection of each annotated compound was conducted to confirm m/z-RT library match, and/or MS/MS library match. Peak height was used as mass spectral intensity at a specific retention time for each annotated lipid. After data processing, each sample was normalized to the total summed intensity of all known features (excluding internal standards).

Chemical similarity enrichment analysis was conducted using ChemRICH available at www.chemrich.fiehnlab.ucdavis.edu, which groups annotated lipids into non-overlapping classes based on Medial Subject Headings (MeSH) terms and Tanimoto chemical similarity calculations (e.g, unsaturated triglycerides, saturated triglycerides, unsaturated

fatty acids, sphingomyelins)[104]. ChemRICH input included the annotated feature name, chemical identifiers including SMILES code, uncorrected *p*-values of Student t-tests between two groups of interest, and fold change between the two groups of interest for each annotated lipid. ChemRICH then uses the Kolmogorov-Smirnov test to calculate significance level (*p*-value) between two treatment groups for all detected classes of lipids. ChemRICH analysis was calculated to detect differential regulation of chemical classes in antibiotic-treated compared to untreated groups, with or without considering diet treatment.

The lipidomic dataset generated in section 2.3 (excluding quality control samples and internal standards) was investigated using the Lipid Ontology (LION/web) enrichment analysis web application (http://www.lipidontology.com/)[105]. Differences between the dietary treatment groups and the antibiotic treatment groups were investigated using the T-Test option (n = 14) in LION; 328 lipid species were submitted to LION for enrichment statistics.

Total plasma FFA levels were measured using an acyl-CoA oxidase-based colorimetric kit from Wako Chemicals Inc. (Richmond, VA). Total plasma TG levels were analyzed using a Ponte Scientific TG reagent (Thermo Fisher Scientific; Waltham, MA). Plasma levels of Angiopoietin-like Protein 4 (ANGPTL4) was measured using an ELISA kit from Cloud Clone (Katy, TX).

All data are expressed as means ± S.E.M. The significance of differences in the mean value was assessed by one-way ANOVA followed by *ad hoc* analysis using the Bonferonni method for multiple comparisons. In the lipidomic analysis, *P* values were adjusted for multiple comparisons using Benjamini and Hochberg false discovery rate

correction[106]. A *P* value less than 0.05 was considered to be statistically significant. Principal component analysis (PCA) was calculated using all known lipids (excluding internal standards) from a log transformed dataset generated from LC-MS/MS data processing (section 2.3). PLS-DA analysis and variable importance in projection (VIP scoring) was calculated using a lipidomic dataset generated in section 2.3, excluding internal standards and quality control samples. For PCA and PLS-DA analysis, no missing value replacement was performed. PLS-DA and PCA figures were generated using MetaboAnalyst 4.0[107].

1.8.4 Metabolomics analysis of serum and stool collected during dietary flaxseed intervention study

A randomized crossover flaxseed intervention was conducted through the Roswell Cancer Center. The subjects participating in this study was a cohort of 179 healthy postmenopausal women from New York. Health was assessed by a medical history questionnaire prior to study enrollment. Fasting serum was collected from each subject before and after a 6-week intervention period when 10 grams per day of ground flaxseed was consumed (Figure 1.13). One stool sample was collected within 48 hours prior to flaxseed intervention began, and after 6 weeks of flaxseed consumption. Study participants were provided with a 500-gram bag of ground flaxseed and a standard scoop. Unused ground flaxseed was returned after the dietary intervention to monitor compliance. Subjects maintained their usual diet throughout the study except for flaxseed supplementation. Women having consumed antibiotics, hormone replacement therapy, supplements for treatment of menopausal symptoms, or flaxseed supplements were

excluded for this study. Subjects were instructed to avoid additional foods containing flaxseed for the duration of the study.

LC-MS/MS analysis of serum was performed using HILIC chromatography with methods as previously described[15] with modifications as follows. Data processing was performed using MS-DIAL version 3.96 with key parameters of minimum peak height set to 100,000, MS/MS libraries of NIST17 and MassBank of North America, match identification score cutoff of 88, and retention time library of Fiehn HILIC library found at the MassBank of North America. The resulting aligned feature table was further curated. Blank filtering was performed by removing features that had a ratio of the maximum intensity of any sample / average of method blanks less than 5. All annotated metabolites were manually checked for MS/MS matches, accurate *m/z* match, and retention time match. Each MS/MS library match was also checked for other possible metabolites that could be MS/MS matches signifying non-unique MS/MS matches and not fulfilling MS/MS match in context of MSI confidence levels reported elsewhere[108]. Stool samples were analyzed by using an extraction procedure as previously reported[15], except the serum sample was replaced with 5 mg +/- 1mg of stool homogenized in a genogrinder shaker with two 3mm stainless steel balls prior to addition of MTBE. Statistical analyses were performed as paired T-tests using pre and post intervention samples from the same subject as pairs and correction for multiple comparisons performed using methods of Benjamin and Hochberg[106]. GC-MS analysis of serum was performed as previously described[15] for serum and stool samples, with the exception that data was normalized to the sum of all annotated metabolites, instead of random forest signal correction.

1.9 References

1 A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning, Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer, *Anal. Chem.*, 2006, **78**, 2113–2120.

2 J. W. Jorgenson, Capillary liquid chromatography at ultrahigh pressures, *Annu. Rev. Anal. Chem.*, 2010, **3**, 129–150.

3 H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn and M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat. Methods*, 2015, **12**, 523–526.

4 H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z. J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita and M. Arita, A lipidome atlas in MS-DIAL 4, *Nat. Biotechnol.*, 2020, **38**, 1159–1163.

5 O. D. Myers, S. J. Sumner, S. Li, S. Barnes and X. Du, Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data, *Anal. Chem.*, 2017, **89**, 8689–8695.

6 R. Krajmalnik-Brown, Z. E. Ilhan, D. W. Kang and J. K. DiBaise, Effects of gut microbes on nutrient absorption and energy regulation, *Nutr. Clin. Pract.*, 2012, **27**, 201–214.

7 J. G. LeBlanc, C. Milani, G. S. de Giori, F. Sesma, D. van Sinderen and M. Ventura, Bacteria as vitamin suppliers to their host: A gut microbiota perspective, *Curr. Opin. Biotechnol.*, 2013, **24**, 160–168.

8 J. M. Ridlon, D. J. Kang, P. B. Hylemon and J. S. Bajaj, Bile acids and the gut microbiome, *Curr. Opin. Gastroenterol.*, 2014, **30**, 332–338.

9 M. R. Viant, I. J. Kurland, M. R. Jones and W. B. Dunn, How close are we to complete annotation of metabolomes?, *Curr. Opin. Chem. Biol.*, 2017, **36**, 64–69.

10 A. D. Hanson, C. S. Henry, O. Fiehn and V. de Crécy-Lagard, Metabolite Damage and Metabolite Damage Control in Plants., *Annu. Rev. Plant Biol.*, 2016, **67**, 131–52.

11 T. Kind, M. Scholz and O. Fiehn, How Large Is the Metabolome? A Critical Analysis of Data Exchange Practices in Chemistry, *PLoS One*, , DOI:10.1371/journal.pone.0005440.

12 R. R. da Silva, P. C. Dorrestein and R. A. Quinn, Illuminating the dark matter in metabolomics, *Proc. Natl. Acad. Sci.*, 2015, **112**, 12549–12550.

13 A. D. Southam, R. J. M. Weber, J. Engel, M. R. Jones and M. R. Viant, A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics, *Nat. Protoc.*, 2017, **12**, 255–273.

14 R. González-Domínguez, T. García-Barrera and J. L. Gómez-Ariza, Metabolomic study of lipids in serum for biomarker discovery in Alzheimer's disease using direct

infusion mass spectrometry, *J. Pharm. Biomed. Anal.*, 2014, **98**, 321–326.

15 D. K. Barupal, Y. Zhang, T. Shen, S. Fan, B. S. Roberts, P. Fitzgerald, B. Wancewicz, L. Valdiviez, G. Wohlgemuth, G. Byram, Y. Y. Choy, B. Haffner, M. R. Showalter, A. Vaniya, C. S. Bloszies, J. S. Folz, T. Kind, A. M. Flenniken, C. McKerlie, L. M. J. Nutter, K. C. Lloyd and O. Fiehn, A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international mouse phenotyping consortium, *Metabolites*, 2019, **9**, 9050101.

16 C. A. Hutchison, R. Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z. Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson and J. C. Venter, Design and synthesis of a minimal bacterial genome, *Science (80-.).*, 2016, **351**, 1–11.

17 Z. Yang and S. K. W. Tsui, Functional Annotation of Proteins Encoded by the Minimal Bacterial Genome Based on Secondary Structure Element Alignment, *J. Proteome Res.*, 2018, **17**, 2511–2520.

18 I. Ali, R. J. Conrad, E. Verdin and M. Ott, Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics, *Chem. Rev.*, 2018, **118**, 1216–1252.

19 J. R. Wiśniewski, A. Zougman and M. Mann, Nε-Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function, *Nucleic Acids Res.*, 2008, **36**, 570–577.

20 J. Lenčo, M. A. Khalikova and F. Švec, Dissolving Peptides in 0.1% Formic Acid

Brings Risk of Artificial Formylation, *J. Proteome Res.*, 2020, **19**, 993–999.

21 B. Edrissi, K. Taghizadeh, B. C. Moeller, D. Kracko, M. Doyle-Eisele, J. A. Swenberg and P. C. Dedon, Dosimetry of N6-formyllysine adducts following [13C2H2]-formaldehyde exposures in rats, *Chem. Res. Toxicol.*, 2013, **26**, 1421–1423.

22 Y. Zhao, C. K. Chan, K. K. Jason Chan and W. Chan, Quantitation of N6-Formyl-lysine Adduct following Aristolochic Acid Exposure in Cells and Rat Tissues by Liquid Chromatography-Tandem Mass Spectrometry Coupled with Stable Isotope-Dilution Method, *Chem. Res. Toxicol.*, 2019, **32**, 2086–2094.

23 T. Baldensperger, S. Di Sanzo, A. Ori and M. A. Glomb, Quantitation of Reactive Acyl-CoA Species Mediated Protein Acylation by HPLC-MS/MS, *Anal. Chem.*, 2019, **91**, 12336–12343.

24 T. Jiang, X. Zhou, K. Taghizadeh, M. Dong and P. C. Dedon, N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 60–65.

25 J. A. Patterson, H. He, J. S. Folz, Q. Li, M. A. Wilson and O. Fiehn, Thioproline formation as a driver of formaldehyde toxicity in Escherichia coli, *Biochem. J.*, 2020, **April 17**, 1745–1757.

26 M. Smuda, M. Voigt and M. A. Glomb, Degradation of 1-Deoxy-D-erythro-hexo-2, 3-diulose in the Presence of Lysine Leads to Formation of Carboxylic Acid Amides, *Agric. Biol. Chem.*, 2010, **58**, 6458–6464.

27 L. Trézl, I. Rusznák, E. Tyihák, T. Szarvas and B. Szende, Spontaneous N epsilon-methylation and N epsilon-formylation reactions between L-lysine and formaldehyde inhibited by L-ascorbic acid., *Biochem. J.*, 1983, **214**, 289–292.

28 C. E. Dempsey, Selective Formylation of a- or 6-Amino-groups of Peptides, *J. Chem. Soc. Perkin Trans. 1*, 1982, 2625–2629.

29 S. Rakete, A. Klaus and M. A. Glomb, Investigations on the maillard reaction of dextrins during aging of pilsner type beer, *J. Agric. Food Chem.*, 2014, **62**, 9876–9884.

30 S. M. Stamenova and R. Hoffmann, Diversity of advanced glycation end products in the bovine milk proteome, *Amino Acids*, 2019, **51**, 891–901.

31 F. H. Westheimer, The magnitude of the primary kinetic isotope effect for compounds of hydrogen and deuterium, *Chem. Rev.*, 1961, **61**, 265–273.

32 G. R. Wagner and R. M. Payne, Widespread and enzyme-independent Nε-acetylation and Nε-succinylation of proteins in the chemical conditions of the mitochondrial matrix, *J. Biol. Chem.*, 2013, **288**, 29036–29045.

33 I. Lagkouvardos, K. Kläring, S. S. Heinzmann, S. Platz, B. Scholz, K. H. Engel, P. Schmitt-Kopplin, D. Haller, S. Rohn, T. Skurk and T. Clavel, Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men, *Mol. Nutr. Food Res.*, 2015, **59**, 1614–1628.

34 M. Rådjursöga, H. M. L. Id, A. Pedersen, D. Malmodin, L. Ellegård and A. W. Id, Nutritional Metabolomics: Postprandial Response of Meals Relating to Vegan,

Lacto-Ovo Vegetarian, and, *Nutrients*, 2018, **10**, 1–13.

35 A. A. Mcdonough, L. C. Veiras, C. A. Guevara and D. L. Ralph, Cardiovascular benefits associated with higher dietary K+ vs . lower dietary Na+ : evidence from population and mechanistic studies, *Am. J. Physiol. Endocrinol. Metab.*, 2017, **312**, 348–356.

36 R. Baudrand, C. Campino, C. A. Carvajal, O. Olivieri, G. Guidi, G. Faccini, P. A. Vohringer, J. Cerda, G. Owen, A. . Kalergis and C. . Fardella, High sodium intake is associated with increased glucocorticoid production , insulin resistance and metabolic syndrome, *Clin. Endocrinol. (Oxf).*, 2014, **80**, 677–684.

37 R. Burcelin, Regulation of Metabolism : A Cross Talk Between Gut Microbiota and Its Human Host, *Physiology*, 2012, **27**, 300–307.

38 Z. Wang, E. Klipfell, B. J. Bennett, R. Koeth, B. S. Levison, B. Dugar, A. E. Feldstein, E. B. Britt, X. Fu, Y. Chung, Y. Wu, P. Schauer, J. D. Smith, H. Allayee, W. H. W. Tang, J. A. Didonato, A. J. Lusis and S. L. Hazen, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease, *Nature*, 2011, **472**, 57–63.

39 I. Blazenovic, Y. T. Oh, F. Li, J. Ji, A. Nguyen, B. Wancewicz, J. M. Bender, O. Fiehn and J. H. Youn, Effects of Gut Bacteria Depletion and High-Na + and Low-K + Intake on Circulating Levels of Biogenic Amines, *Mol. Nutr. Food Res.*, 2019, **63**, e1801184.

40 P. Stahel, C. Xiao, R. A. Hegele and G. F. Lewis, The Atherogenic Dyslipidemia Complex and Novel Approaches to Cardiovascular Disease Prevention in Diabetes,

Can. J. Cardiol., 2018, **34**, 595–604.

41 V. R. Velagapudi, R. Hezaveh, C. S. Reigstad, P. Gopalacharyulu, L. Yetukuri, S. Islam, J. Felin, R. Perkins, J. Borén, M. Orešič and F. Bäckhed, The gut microbiota modulates host energy and lipid metabolism in mice, *J. Lipid Res.*, 2010, **51**, 1101–1112.

42 E. P. Rhee, S. Cheng, M. G. Larson, G. A. Walford, G. D. Lewis, E. McCabe, E. Yang, L. Farrell, C. S. Fox, C. J. O'Donnell, S. A. Carr, R. S. Vasan, J. C. Florez, C. B. Clish, T. J. Wang and R. E. Gerszten, Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans, *J. Clin. Invest.*, 2011, **121**, 1402–1411.

43 T. Arora, V. Velagapudi, D. J. Pournaras, R. Welbourn, C. W. le Roux, M. Oresic and F. Backhed, Roux-en-Y Gastric Bypass Surgery Induces Early Plasma Metabolomic and Lipidomic Alterations in Humans Associated with Diabetes Remission, *PLoS One*, 2015, **10**, e0126401.

44 U. Schwab, T. Seppanen-Laakso, L. Yetukuri, A. Jyrki, M. Kolehmainen, D. E. Laaksonen, A. Ruskeepaa, H. Gylling, M. Uusitupa, M. Oresic and G. Study Group, Triacylglycerol Fatty Acid Composition in Diet-Induced Weight Loss in Subjects with Abnormal Glucose Metabolism – the GENOBIN Study, *PLoS One*, 2008, **3**, e2630.

45 K. Vanschoonbeek, moniek P. M. de Maat and J. W. M. Heemskerk, Fish Oil Consumption and Reduction of Arterial Disease, *J. Nutr.*, 2003, **133**, 657–660.

46 P. M. Kris-etherton, D. S. Taylor, S. Yu-poth, P. Huth, K. Moriarty, V. Fishell, R. L. Hargrove, G. Zhao and T. D. Etherton, Polyunsaturated fatty acids in the food chain

in the United States, *Am. J. Clin. Nutr.*, 2000, **71**, 179S-188S.

47 L. M. Arterburn, E. B. Hall and H. Oken, Distribution , interconversion , and dose response of n-3 fatty acids, *Am. J. Clin. Nutr.*, 2006, **83**, 1467–1476.

48 B. Brayner, G. Kaur, M. A. Keske and K. M. Livingstone, FADS Polymorphism, Omega-3 Fatty Acids and Diabetes Risk: A Systematic Review, *Nutrients*, 2018, **10**, E758.

49 M. J. Takkunen, U. S. Schwab, V. D. F. De Mello, J. G. Eriksson, J. Lindstrom, J. Tuomilehto, D. S. Group, M. I. J. Uusitupa and S. G. DPS, Longitudinal associations of serum fatty acid composition with type 2 diabetes risk and markers of insulin secretion and sensitivity in the Finnish Diabetes Prevention Study, *Eur. J. Nutr.*, 2016, **55**, 967–979.

50 M. M. Sethom, S. Fares, M. Feki, M. Elasmi, S. Omar, H. Sanhaji, R. Jemaa and N. Kaabachi, Plasma fatty acids profile and estimated elongase and desaturases activities in Tunisian patients with the metabolic syndrome, *Prostaglandins Leukot. Essent. Fat. Acids*, 2011, **85**, 137–141.

51 A. Kindt, G. Liebisch, T. Clavel, D. Haller, G. Hörmannsperger, H. Yoon, D. Kolmeder, A. Sigruener, S. Krautbauer, C. Seeliger, A. Ganzha, S. Schweizer, R. Morisset, T. Strowig, H. Daniel, D. Helm, B. Küster, J. Krumsiek and J. Ecker, The gut microbiota promotes hepatic fatty acid desaturation and elongation in mice, *Nat. Commun.*, 2018, **9**, 3760.

52 R. Kübeck, C. Bonet-ripoll, C. Hoffmann, A. Walker, V. M. Müller, V. L. Schüppel, I. Lagkouvardos, B. Scholz, K. Engel, H. Daniel, P. Schmitt-kopplin, D. Haller, T.

Clavel and M. Klingenspor, Dietary fat and gut microbiota interactions determine diet-induced obesity in mice, *Mol. Metab.*, 2016, **5**, 1162–1174.

53 A. Olsen, J. Christensen, N. F. Johnsen, K. Overvad, K. E. B. Knudsen and A. Tjønneland, Prediagnostic plasma enterolactone levels and mortality among women with breast cancer, *Breast Cancer Res. Treat.*, 2011, **128**, 883–889.

54 H. Adlercreutz, T. Fotsis, C. Bannwart, K. Wahala, T. Makela, G. Brunow and T. Hase, Determination of urinary lignans and phytoestrogen metabolites, potential antiestrogens and anticarcinogens, in urine of women on various habitual diets, *J. Steroid Biochem.*, 1986, **25**, 791–797.

55 Q. Sun, N. Wedick, A. Pan, M. Townsend, A. Cassidy, E. Franke, Adrian, Rimm, F. Hu and R. Dam, Gut Microbiota Metabolites of Dietary Lignans and Risk of Type 2 Diabetes : A Prospective Investigation in Two Cohorts of, *Diabetes Care*, 2014, **37**, 1287–1295.

56 J. Hallund, I. Tetens, S. Bügel, T. Tholstrup and J. M. Bruun, The effect of a lignan complex isolated from flaxseed on inflammation markers in healthy postmenopausal women, *Nutr. Metab. Cardiovasc. Dis.*, 2008, **18**, 497–502.

57 A. Patade, B. P. Daggy, K. Korlagunta, E. A. Lucas, L. Devareddy and B. H. Arjmandi, Flaxseed Reduces Total and LDL Cholesterol Concentrations in Native American Postmenopausal Women, *J. Women's Heal.*, 2008, **17**, 355–366.

58 B. H. Arjmandi, D. A. Khan, S. Juma, M. L. Drum, S. Venkatesh, E. Sohn, L. Wei and R. Derman, Whole flaxseed consumption lowers serum LDL-cholesterol and lipoprotein(a) concentrations in postmenopausal women, *Nutr. Res.*, 1998, **18**,

1203–1214.

59 M. C. Colli, C. G. Bôer, C. G. M. de Souza, R. M. Peralta, A. A. Soares, A. L. de Oliveira and A. Bracht, Evaluation of the Efficacy of Flaxseed Meal and Flaxseed Extract in Reducing Menopausal Symptoms, *J. Med. Food*, 2012, **15**, 840–845.

60 A. Lemay, N. Kadri, S. Dodin, J.-C. Forest and H. Jacques, Flaxseed Dietary Supplement Versus Hormone Replacement Therapy in Hypercholesterolemic Menopausal Women, *Obstet. Gynecol.*, 2016, **100**, 495–504.

61 M. Abu-Taha, J.-M. Cerda-Nicolas, E. J. Morcillo, P. J. Jose, I. Noguera, J. Cortijo, C. Rius, A. C. Issekutz, C. Hermenegildo and M.-J. Sanz, Menopause and Ovariectomy Cause a Low Grade of Systemic Inflammation that May Be Prevented by Chronic Treatment with Low Doses of Estrogen or Losartan, *J. Immunol.*, 2009, **183**, 1393–1402.

62 J. Hallund, S. Bügel, G. Ravn-Haren, T. Tholstrup and I. Tetens, A Lignan Complex Isolated from Flaxseed Does Not Affect Plasma Lipid Concentrations or Antioxidant Capacity in Healthy Postmenopausal Women, *Nutr. Dis.*, 2006, **136**, 112–116.

63 S. Dodin, S. C. Cunnane, B. Mâsse, A. Lemay, H. Jacques, G. Asselin, J. Tremblay-Mercier, I. Marc, B. Lamarche, F. Légaré and J. C. Forest, Flaxseed on cardiovascular disease markers in healthy menopausal women: a randomized, double-blind, placebo-controlled trial, *Nutrition*, 2008, **24**, 23–30.

64 M. Kristensen, M. G. Jensen, J. Aarestrup, K. E. N. Petersen, L. Søndergaard, M. S. Mikkelsen and A. Astrup, Flaxseed dietary fibers lower cholesterol and increase fecal fat excretion, but magnitude of effect depend on food type, *Nutr. Metab.*, 2012,

9, 1–8.

65 D. Rodriguez-Leyva, C. M. C. Bassett, R. McCullough and G. N. Pierce, The cardiovascular effects of flaxseed and its omega-3 fatty acid, alpha-linolenic acid, *Can. J. Cardiol.*, 2010, **26**, 489–496.

66 W. Zhang, X. Wang, Y. Liu, H. Tian, B. Flickinger, M. W. Empie and S. Z. Sun, Dietary flaxseed lignan extract lowers plasma cholesterol and glucose concentrations in hypercholesterolaemic subjects, *Br. J. Nutr.*, 2008, **99**, 1301–1309.

67 J. M. Landete, Plant and mammalian lignans: A review of source, intake, metabolism, intestinal bacteria and health, *Food Res. Int.*, 2012, **46**, 410–424.

68 M. Axelson and K. D. R. Setchell, The excretion of lignans in rats - evidence for an intestinal bacterial source for this new group of compounds, *FEBS Lett.*, 1981, **123**, 337–342.

69 S. Possemiers, S. Bolca, E. Eeckhaut, H. Depypere and W. Verstraete, Metabolismof isofavones, lignans and prenylflavonoidsby intestinal bacteria: producer phenotyping and relation with intestinal community, *FEMS Microbiol. Ecol.*, 2007, **61**, 372–383.

70 E. Eeckhaut, K. Struijs, S. Possemiers, J. P. Vincken, D. De Keukeleire and W. Verstraete, Metabolism of the lignan macromolecule into enterolignans in the gastrointestinal lumen as determined in the simulator of the human intestinal microbial ecosystem, *J. Agric. Food Chem.*, 2008, **56**, 4806–4812.

71 E. Hålldin, A. K. Eriksen, C. Brunius, A. B. da Silva, M. Bronze, K. Hanhineva, A.-M. Aura and R. Landberg, Factors Explaining Inter-Personal Variation in Plasma Enterolactone Concentrations in Humans, *Mol. Nutr. Food Res.*, 2019, **1801159**, 1801159.

72 S. O. Mueller, S. Simon, K. Chae, M. Metzler and K. S. Korach, Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor α (ERα) and ERβ in human cells, *Toxicol. Sci.*, 2004, **80**, 14–25.

73 D. D. Kitts, Y. V Yuan, A. N. Wijewickreme and L. U. Thompson, Antioxidant activity of the flaxseed lignan secoisolariciresinol diglycoside and its mammalian lignan metabolites enterodiol and enterolactone, *Mol. Cell. Biochem.*, 1999, **202**, 91–100.

74 E. Corsini, M. Dell'Agli, A. Facchi, E. De Fabiani, L. Lucchi, M. S. Boraso, M. Marinovich and C. L. Galli, Enterodiol and enterolactone modulate the immune response by acting on nuclear factor-κB (NF-κB) signaling, *J. Agric. Food Chem.*, 2010, **58**, 6678–6684.

75 F. L. Miles, S. L. Navarro, Y. Schwarz, H. Gu, D. Djukovic, T. W. Randolph, A. Shojaie, M. Kratz, M. A. J. Hullar, P. D. Lampe, M. L. Neuhouser, D. Raftery and J. W. Lampe, Plasma metabolite abundances are associated with urinary enterolactone excretion in healthy participants on controlled diets, *Food Funct.*, 2017, **8**, 3209–3218.

76 S. P. Ndou, H. M. Tun, E. Kiarie, M. C. Walsh, E. Khafipour and C. M. Nyachoti, Dietary supplementation with flaxseed meal and oat hulls modulates intestinal histomorphometric characteristics, digesta- and mucosa-associated microbiota in

pigs, *Sci. Rep.*, 2018, **8**, 1–15.

77 J. Martens-Lobenhoffer, R. N. Rodionov, A. Drust and S. M. Bode-Böger, Detection and quantification of α-keto-δ-(N G,N G-dimethylguanidino)valeric acid: A metabolite of asymmetric dimethylarginine, *Anal. Biochem.*, 2011, **419**, 234–240.

78 H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn and M. Arita, Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software, *Anal. Chem.*, 2016, **88**, 7946–7958.

79 J. F. O'Sullivan, J. E. Morningstar, Q. Yang, B. Zheng, Y. Gao, S. Jeanfavre, J. Scott, C. Fernandez, H. Zheng, S. O'Connor, P. Cohen, R. S. Vasan, M. T. Long, J. G. Wilson, O. Melander, T. J. Wang, C. Fox, R. T. Peterson, C. B. Clish, K. E. Corey and R. E. Gerszten, Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes, *J. Clin. Invest.*, 2017, **127**, 4394–4402.

80 P. Vitaglione, S. Sforza, G. Galaverna, C. Ghidini, N. Caporaso, P. P. Vescovi, V. Fogliano and R. Marchelli, Bioavailability of trans -resveratrol from red wine in humans, *Mol. Nutr. Food Res.*, 2005, **49**, 495–504.

81 G. Borges, J. I. Ottaviani, J. J. J. van der Hooft, H. Schroeter and A. Crozier, Absorption, metabolism, distribution and excretion of (−)-epicatechin: A review of recent findings, *Mol. Aspects Med.*, 2018, **61**, 18–30.

82 J. W. Lampe, Biomarkers of nutritional exposure and nutritional status: An overview, *J. Nutr.*, 2003, **133**, 956–964.

83 A. Toda, M. Shimizu, S. Uehara, T. Sasaki, T. Miura, M. Mogi, M. Utoh, H. Suemizu and H. Yamazaki, Plasma and hepatic concentrations of acetaminophen and its primary conjugates after oral administrations determined in experimental animals and humans and extrapolated by pharmacokinetic modeling, *Xenobiotica*, 2021, **51**, 316–323.

84 B. D. Oomah, E. O. Kenaschuk and G. Mazza, Tocopherols in Flaxseed, *J. Agric. Food Chem.*, 1997, **45**, 2076–2080.

85 C. Müller, P. Schäfer, M. Störtzel, S. Vogt and W. Weinmann, Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2002, **773**, 47–52.

86 C. A. Chamberlain, V. Y. Rubio and T. J. Garrett, Impact of matrix effects and ionization efficiency in non-quantitative untargeted metabolomics, *Metabolomics*, 2019, **15**, 1–9.

87 S. Fan, T. Kind, T. Cajka, S. L. Hazen, W. H. W. Tang, R. Kaddurah-Daouk, M. R. Irvin, D. K. Arnett, D. K. Barupal and O. Fiehn, Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data, *Anal. Chem.*, 2019, **91**, 3590–3596.

88 J. E. Schollée, E. L. Schymanski, M. A. Stravs, R. Gulde, N. S. Thomaidis and J. Hollender, Similarity of High-Resolution Tandem Mass Spectrometry Spectra of Structurally Related Micropollutants and Transformation Products, *J. Am. Soc.*

Mass Spectrom., 2017, **28**, 2692–2704.

89 P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal and O. Fiehn, Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics, *Anal. Chem.*, 2020, **92**, 7515–7522.

90 Y. F. Xu, W. Lu and J. D. Rabinowitz, Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics, *Anal. Chem.*, 2015, **87**, 2273–2281.

91 D. K. Barupal and O. Fiehn, Generating the Blood Exposome Database Using a Comprehensive Text Mining and, *Environ. Health Perspect.*, 2019, **127**, 1–10.

92 T. Kind, G. Wohlgemuth, D. Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz and O. Fiehn, FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry, *Anal. Chem.*, 2009, **81**, 10038–10048.

93 T. D. Niehaus, J. A. Patterson, D. C. Alexander, J. S. Folz, M. Pyc, B. S. Mactavish, S. D. Bruner, R. T. Mullen, O. Fiehn and A. D. Hanson, The metabolite repair enzyme Nit1 is a dual-targeted amidase that disposes of damaged glutathione in Arabidopsis, *Biochem. J.*, , DOI:10.1042/BCJ20180931.

94 H. Tsugawa, R. Nakabayashi, T. Mori, Y. Yamada, M. Takahashi, A. Rai, R. Sugiyama, H. Yamamoto, T. Nakaya, M. Yamazaki, R. Kooke, J. A. Bac-Molenaar, N. Oztolan-Erol, J. J. B. Keurentjes, M. Arita and K. Saito, A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms, *Nat. Methods*, 2019, **16**, 295–298.

95 H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. Vandergheynst, O. Fiehn and M. Arita, MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat. Methods*, 2015, **12**, 523–526.

96 J. Folz, Y. T. Oh, I. Blaženović, J. Richey, O. Fiehn and J. H. Youn, Interaction of Gut Microbiota and High-Sodium, Low-Potassium Diet in Altering Plasma Triglyceride Profiles Revealed by Lipidomics Analysis, *Mol. Nutr. Food Res.*, 2019, **63**, 1–9.

97 L. Tobian, Dietary sodium pathophysiology chloride and potassium have effects on the of hypertension in humans and animals, *Am. J. Clin. Nutr.*, 1997, **65**, 606–611.

98 Z. Wang, P. Sun, W. Xing, C. Pan, D. Lin and W. Wang, Decrease in dietary K intake stimulates the generation of superoxide anions in the kidney and inhibits K secretory channels in the CCD, *Am. J. Physiol. Ren. Physiol.*, 2010, **298**, F1515-1522.

99 Y. T. Oh, J. Kim and J. H. Youn, Role of pituitary in K+ homeostasis : impaired renal responses to altered K+ intake in hypophysectomized rats, *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, 2013, **304**, R1166-1174.

100 T. Cajka, J. T. Smilowitz and O. Fiehn, Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography − High-Resolution Mass Spectrometry Platforms, *Anal. Chem.*, 2017, **89**, 12360–12368.

101 V. Matyash, G. Liebisch, T. V. Kurzchalia, A. Shevchenko and D. Schwudke, Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics, *J. Lipid Res.*,

2008, **49**, 1137–1146.

102 T. Kind, K. H. Liu, D. Y. Lee, B. Defelice, J. K. Meissen and O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, *Nat. Methods*, 2013, **10**, 755–758.

103 B. C. DeFelice, S. S. Mehta, S. Samra, T. Čajka, B. Wancewicz, J. F. Fahrmann and O. Fiehn, Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing, *Anal. Chem.*, 2017, **89**, 3250–3255.

104 D. K. Barupal and O. Fiehn, Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets, *Sci. Rep.*, 2017, **7**, 14567.

105 M. R. Molenaar, A. Jeucken, T. A. Wassenaar, C. H. A. Van De Lest, J. F. Brouwers and J. B. Helms, LION/web: A web-based ontology enrichment tool for lipidomic data analysis, *Gigascience*, 2019, **8**, 1–10.

106 Y. Benjamini and Y. Hochberg, Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B*, 1995, **57**, 289–300.

107 J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart and J. Xia, MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis, *Nucleic Acids Res.*, 2018, **46**, W486–W494.

108 E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer and J.

Hollender, Identifying small molecules via high resolution mass spectrometry: Communicating confidence, *Environ. Sci. Technol.*, 2014, **48**, 2097–2098.

Chapter 2: Metabolomics Analysis of Time-Series Human Small Intestine Lumen Samples Collected in vivo

Adapted from "Metabolomics Analysis of Time-Series Human Small Intestine Lumen Samples Collected *in vivo*". 2021. Jacob S. Folz, Dari Shalon, Oliver Fiehn. *Food and Function.* 10.1039/D1FO01574E

2.1 Abstract

The human small intestine remains an elusive organ to study due to the difficulty of retrieving samples in a non-invasive manner. Stool samples as a surrogate do not reflect events in the upper gut intestinal tract. As proof of concept, this study investigates time-series samples collected from the upper gastrointestinal tract of a single healthy subject. Samples were retrieved using a small diameter tube that collected samples in the stomach and duodenum as the tube progressed to the jejunum, and then remained positioned in the jejunum during the final 8.5 hours of the testing period. Lipidomics and metabolomics liquid chromatography tandem mass spectrometry (LC-MS/MS) assays were employed to annotate 828 unique metabolites using accurate mass with retention time and/or tandem MS library matches. Annotated metabolites were clustered based on correlation to reveal sets of biologically related metabolites. Typical clusters included bile metabolites, food metabolites, protein breakdown products, and endogenous lipids. Acylcarnitines and phospholipids were clustered with known human bile components supporting their presence in human bile, in addition to novel human bile compounds 4-hydroxyhippuric acid, N-acetylglucosaminoasparagine and 3-methoxy-4-hydroxyphenylglycol sulfate. Food metabolites were observed passing through the small

intestine after meals. Acetaminophen and its phase II metabolism products appeared for hours after the initial drug treatment, due to enterohepatic recirculation. This exploratory study revealed novel trends in timing and chemical composition of the human jejunum under standard living conditions.

2.2 Introduction

The human gastrointestinal tract (GI tract) performs essential functions for life including absorption of energy from food and vitamin transformations. This makes the GI tract an important area of research to better understand how humans interact with food. The upper human GI tract has been investigated using techniques such as endoscope imaging[1] and sampling[2], external imaging[3], ileostomy studies[4,5], *in vitro* small intestine models[6], and humanized animal models[7]. These techniques have been primarily used for single timepoint sample collections. However, the human small intestine is a highly complex and dynamic system requiring time-series profiles to understand the *in vivo* functionality of the upper GI tract[8]. This study investigates a unique set of samples collected from the upper GI tract of a single healthy subject at approximately 30-minute intervals for 8 hours.

One important class of biofluids for human GI tract function is bile. Bile is made in the liver, stored in the gallbladder, and excreted into the duodenum. Bile consists of roughly 95% water with 5% organic and inorganic components including bile acids, bile pigments, phospholipids, cholesterol, electrolytes, and proteins[9]. Bile is excreted in response to meals[10] and has important functions including acting as a lipid emulsion stabilizer[11,12] to aid in lipid absorption, as excretory route for cholesterol and other exposome metabolites[13], and to modify the composition of gut microbiota[14]. To better understand

how the GI tract interacts with diet and lifestyle, we investigated bile related metabolites and bile excretion patterns over the course of 8 hours in a single test subject.

In addition to endogenous metabolites, exogenous compounds from food and drugs and their metabolites also pass through the small intestine. These compounds represent extremely broad classes of chemicals encompassing the metabolome of plants, animals, microbes, and the environment. While many exposome compounds are listed in databases such as FooDB and DrugBase, nontargeted metabolomics can detect additional compounds to add knowledge on transformations in the human GI tract. Such food biomarkers may classify dietary patterns and be used in human cohort studies[15–17]. We show here that some foods are more readily predicted from GI tract fluid analyses than others. We also demonstrate the utility of a non-targeted metabolomics analysis leading to novel observations of endogenous and exogenous metabolite dynamics associated with the human GI tract *in vivo*.

2.3 Methods

Ethics statement

The study was approved by and complied with the guidelines approved by the Association for the Accreditation of Human Research Protection Programs (AAHRPP) certified review board WCG IRB (Study Number 1186513, IRB Tracking Number: 20181298) as a Non-Significant-Risk study not requiring an Investigational Device Exemption (IDE) review by the FDA. The single participant provided written informed consent for study participation conducted at Silicon Valley Gastroenterology, 2490 Hospital Drive, Suite 211, Mountain

View, California 94040. For the chemical analyses of samples delivered to UC Davis, the UC Davis IRB Administration approved the study under IRB ID 1307967-1.

Sample collection

A single human volunteer, a male between the age of 45-65 years without any known morbidities and a BMI <30 swallowed a 1.2 mm outer diameter silicon tube with one end attached to a capsule shaped sinker element 6 mm in diameter and 15 mm long. A silicone tube ran through the center of this sinker element with an opening covered with a 150-micron mesh filter. The capsule was swallowed with water and the tube was stopped from advancing when the end of the tube was 200 cm past the mouth, which equates to 125 cm past the pylorus, placing the sampling end of the tube in the jejunum. Samples were aspirated by pulling out one milliliter aliquots of GI tract fluid using a syringe attached to the proximal end of the sampling tube every 30 minutes. The first two samples were collected while the tube was still in the stomach and the rest were taken from the upper small intestine with the final 14 samples taken from 125 cm past the pylorus (Supplementary Table 1). During the sample collection period, food, water, and acetaminophen were consumed *ad libitum* and time of ingestion recorded (Figure 2.1). Specific timepoints of sample collection, food consumption, and sample details can be found in Supplementary Table 1. Samples were transferred to microcentrifuge tubes and frozen at -20 °C until sample extraction (less than 1 week). After sampling, the silicon tube was cut and evacuated along with the capsule sinker element through normal bowel movement.

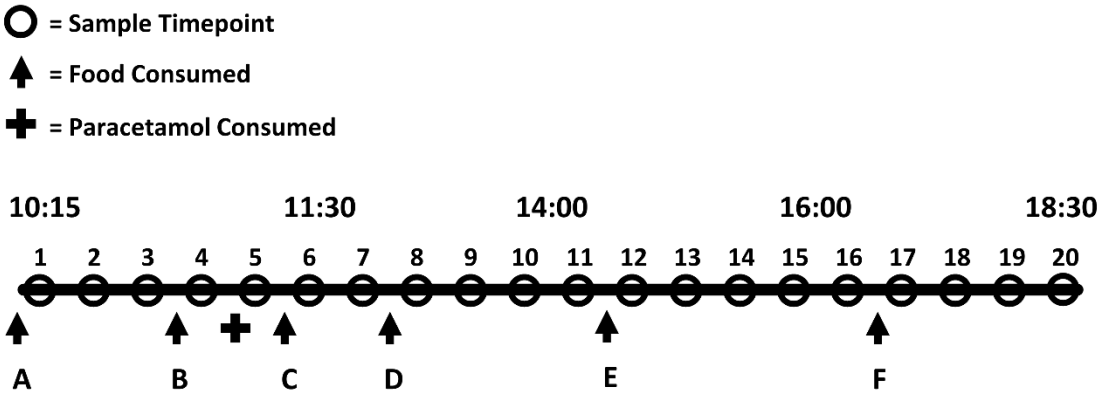


Figure 2.1. Method summary of GI tract sampling. Timeline of sample collections that occurred at approximately 30-minute intervals throughout one day. *Cross*: 500 mg acetaminophen. *Black arrows*: food or beverage consumption. (A) apple juice, espresso (B) 60 g puree (pear, peach, pumpkin, apple, cinnamon) (C) 10 g puree as before. (D) 20 g olive oil, 20 g white bread, espresso (E) 25 g whole wheat cereal, water, espresso (F) 210 g fermented milk drink. Samples 1-2 were collected from the stomach and 3-20 were collected from the small intestine.

Sample Preparation

Samples were separated into hydrophobic and hydrophilic portions using a modified liquid bilayer extraction[18]. Frozen samples were thawed on ice and 20 μL of intestinal liquid was transferred to a clean microcentrifuge tube. Blank samples were created using 20 μL of LC-MS grade water instead of intestinal fluid. 225 μL of ice-cold methanol (containing internal standards[19]) was added to each microcentrifuge tube. Tubes were vortexed vigorously for 10 seconds and then 750 μL of methyl-*tert* butyl ether (containing internal standard cholesterol ester 22:1) was added to each tube. All tubes were vortexed vigorously for 10 seconds and shaken on an orbital shaker at 4 °C for 6 minutes. 190 μL

of ice-cold water was added to each tube followed by 10 seconds of mixing by vortex. Microcentrifuge tubes were centrifuged at 14,000 RCF for two minutes. Two aliquots of 350 μL of the upper MTBE layer were aliquoted into two clean microcentrifuge tubes for lipidomic analysis, and two aliquots of 125 μL of the lower layer was transferred into two clean microcentrifuge tubes for analysis of hydrophilic metabolites. A portion from the remaining upper and lower layers from all samples were combined with portions of other GI tract samples external to this study to generate pooled quality control samples to assess technical variation in analytical measurement. All tubes were dried to completion in a rotary vacuum dryer and stored at -20 °C for less than two weeks until LC-MS/MS analysis.

LC-MS/MS analysis

Reverse phased liquid chromatography tandem mass spectrometry (RPLC-MS/MS) was used to perform lipidomic analysis and began by adding 100 μL run solvent (9:1 methanol/toluene (*v/v*)) to microcentrifuge tubes from the dried upper layer of extraction. Tubes were vortexed for 10 seconds, sonicated for 2 minutes, vortexed for 10 seconds, centrifuged at 14,000 RCF for 2 minutes and the supernatant was transferred to amber 2 mL LC-MS vials with 200 μL glass insert. Chromatography was performed using a Vanquish Focus UHPLC (ThermoFisher Scientific) and mass spectra collected with a QExactive HF^+ mass spectrometer. An Acquity UPLC CSH C18 (100 mm × 2.1 mm, 1.7 μm particle size) column (Waters, Milford MA) with an Acquity UPLC CSH C18 (5 mm × 1.2 mm, 1.7 μm particle size) pre-column (Waters, Milford MA) was used with mobile phase A (6:4 acetonitrile/ water (*v/v*)) and mobile phase B (9:1 isopropanol acetonitrile (*v/v*)). Mobile phases A and B were modified with 10mM ammonium formate and 0.1%

formic acid for positive mode ionization, and 10mM ammonium acetate for negative mode ionization. The LC gradient started at 15% B, increased to 30% B from 0-2 minutes, increased to 48% B from 2-2.5 minutes, increased to 82% B from 2.5-11 minutes, increased to 99% B from 11-11.5 minutes, held at 99% B from 11.5-12 minutes, returned to 15% B from 12-12.1 minutes and held at 15% B from 12.1-15 minutes. The autosampler was held at 4 °C and needle wash was performed before and after sample injections for 10 seconds with isopropanol. Injection volumes were 4 μL for both positive and negative mode ionization analyses. Additional MS parameters were used as previously reported[19]. Spectral data was collected with scan range of 120-1700 *m/z*. MS/MS fragmentation used data dependent acquisition (DDA) and was collected for the top 4 most abundant ions from each MS scan.

Hydrophilic interaction liquid chromatography tandem mass spectrometry (HILIC-MS/MS) was used to measure hydrophilic metabolites and began by adding 0.1 mL run solvent (8:2 acetonitrile/water (*v/v*) with 20 deuterated or synthetic internal standards as previously reported[19] to dried microcentrifuge tubes from the bottom aqueous phase of extraction. Tubes were vortexed, sonicated, vortexed, centrifuged, and transferred as in the lipidomic analysis. The same instruments and parameters were used as lipidomic analysis with the following exceptions. A BEH Amide (150 mm × 2.1 mm, 1.7 μm particle size) column (Waters, Milford MA) with BEH Amide (5 mm × 1.2 mm, 1.7 μm particle size) pre-column Waters, Milford MA) was used with mobile phases A (water) and B (95:5 acetonitrile/ water (*v/v*)) both modified with 10mM ammonium formate, and 0.1% formic acid. The gradient started at 100% B, was held at 100% B from 0-2 minutes, decreased to 70% B between 2-7.7 minutes, decreased to 40% between 7.7 and 9.5 minutes, was

held at 40% B from 9.5-12.75 minutes, returned to 100% B between 12.75-12.85 minutes, and was held at 100% B from 12.85-17 minutes. Needle wash solution was 1:1 acetonitrile/ water (*v/v*). Injection volumes were 3 μL for positive mode electrospray analyses and 5 μL for negative mode electrospray analyses. The scan range was 90-900 *m/z* with MS/MS acquired using DDA. Four quality control samples were analyzed evenly spaced throughout samples for all analytical platforms and used to assess injection reproducibility and instrument stability.

Data Analysis

LC-MS/MS data was processed using open source software MS-DIAL[20] (version 4.24) which performed peak picking, deisotoping, automated peak annotation, alignment and gap filling. Data processing parameters can be found in Supplemental Table 2. Blank subtraction was performed by removing features that had a maximum sample intensity / average blank intensity ratio of less than 5 and also any features that had a maximum sample intensity of less than 30k. Adduct and duplicate features were flagged using Mass Spectral Feature List Optimizer (MS-FLO)[21]. Data from each of the four analytical platforms (RPLC-MS/MS ESI+/-, HILIC-MS/MS ESI+/-) were processed separately and combined after data curation. No data normalization was performed because no trend in data intensities was observed from the internal standards during data acquisition. Peak height was used for all quantitation. Raw data may be found on the Metabolomics Workbench (ST001794). Metabolite annotations were made using defined confidence levels[22] based on accurate mass, MS/MS library matching to experimental data, and retention time from authentic standards run on the same instrument (Table 2.1). Tandem MS/MS libraries of the MassBank of North America (MassBank.us) and NIST17 (NIST,

Gaithersberg, MD) were used for spectral matching. Manual curation of datasets was performed to reduce in-source fragment annotations identified by very similar RT and high correlation between features. Predicted retention times calculated using Retip[23] were used to help identify in-source fragments and eliminate low confidence annotations. Manual review of MS/MS matches was performed to remove poor spectral matches since false positive annotations can occur when automatically matching MS/MS from complex biological samples to large MS/MS spectral libraries[24].

Table 2.1. Classes of annotated metabolites. Subclass level ontology determined by ClassyFire software. Subclasses < 10 metabolites omitted.

chemical subclass	count
Amino acids, peptides, and analogues	325
Glycerophosphocholines	57
Carbohydrates and carbohydrate conjugates	45
Fatty acids and conjugates	45
Triacylglycerols	31
Fatty acid esters	23
Glycosphingolipids	20
Phosphosphingolipids	16
Ceramides	13
Diacylglycerols	12
Amines	10
Bile acids, alcohols and derivatives	10

Correlation based clustering was performed using the “hclust” function from r package “stats v3.6.2” with method “ward.D2”[25] and tree cutting was conducted using “cutree” function from “stats v3.6.2” with input of k=18-26. Spearman rank correlations were used for all correlation analyses in this study. The optimal number of clusters calculated through the elbow, and silhouette[26] methods using Nbclust R package[27] suggested 2 and 3 as the optimal number of clusters respectively for the 828 annotated metabolites. This few of clusters was not useful to find biologically relevant trends due to large and variable clusters. Using a previous timeseries metabolomics study[28] as a guide, a value of 26

clusters was chosen as a starting point, and then the number of clusters was reduced by 1 until biologically distinct metabolite clusters with differing metabolite profiles began to be combined which resulted in 20 being chosen as the optimal number of clusters (Supplemental Figure 1). Helpful R source scripts for the clustering analysis was adapted from others[29]. Other figures, tables, and correlation analyses were created using custom R scripts.

2.4 Results and Discussion

2.4.1 Analyses of GI tract metabolites

A total of 828 unique metabolites were annotated using non-targeted UHPLC-MS/MS analyses of samples retrieved from the upper GI tract. The most abundant chemical subclasses encompassed in this dataset are summarized in Table 2.1. The largest chemical subclass is amino acids, peptides and analogues which contains 325 metabolites including 267 dipeptides and tripeptides. The internal standard 12-(cyclohexylcarbamoylamino)dodecanoic acid (CUDA) was used to determine injection reproducibility with raw average relative standard deviations in the quality control between 4-10% for all LC-MS/MS assays (Supplementary Table 4 (see online)). Metabolites were annotated based on a combination of accurate mass, tandem mass spectral library matching, and retention time matching leading to level 1 through 3 identification confidence levels as previously defined[22]. In total, 6902 chromatographic features with associated MS/MS spectra were detected after blank subtraction, and 12% of these features were annotated with a chemical structure. Unknown features were omitted from the current report, but data are publicly available to enable discovery of additional

metabolites in the future (see methods). Metabolites that share biological regulation or origin will strongly correlate in intensity across the testing period. Therefore, we used correlation-based clustering to identify these groups of functionally related metabolites (Figure 2.2). Multiple clusters were found following consumption of different foods. For example, after consumption of a puree of fruit and vegetables, a group of sugars and arbutin (a biomarker of pear consumption[30–32]) were clustered (Food cluster 5 in Figure 2.2) and show maximum intensities approximately 30 minutes after puree consumption (Figure 2.3 A). In total 20 clusters were generated and manually categorized based on the biological function of metabolites within each cluster. Clusters were manually classified as containing a high proportion of food metabolites, bile metabolites, di- and tripeptides and amino acids, and one cluster of di- and tripeptides with many ceramide lipids. Clusters were labeled as mixed/unknown if a cluster contained metabolites split between multiple or unknown functions. The cluster assignment for all metabolites can be found in Supplementary Table 4(see online).

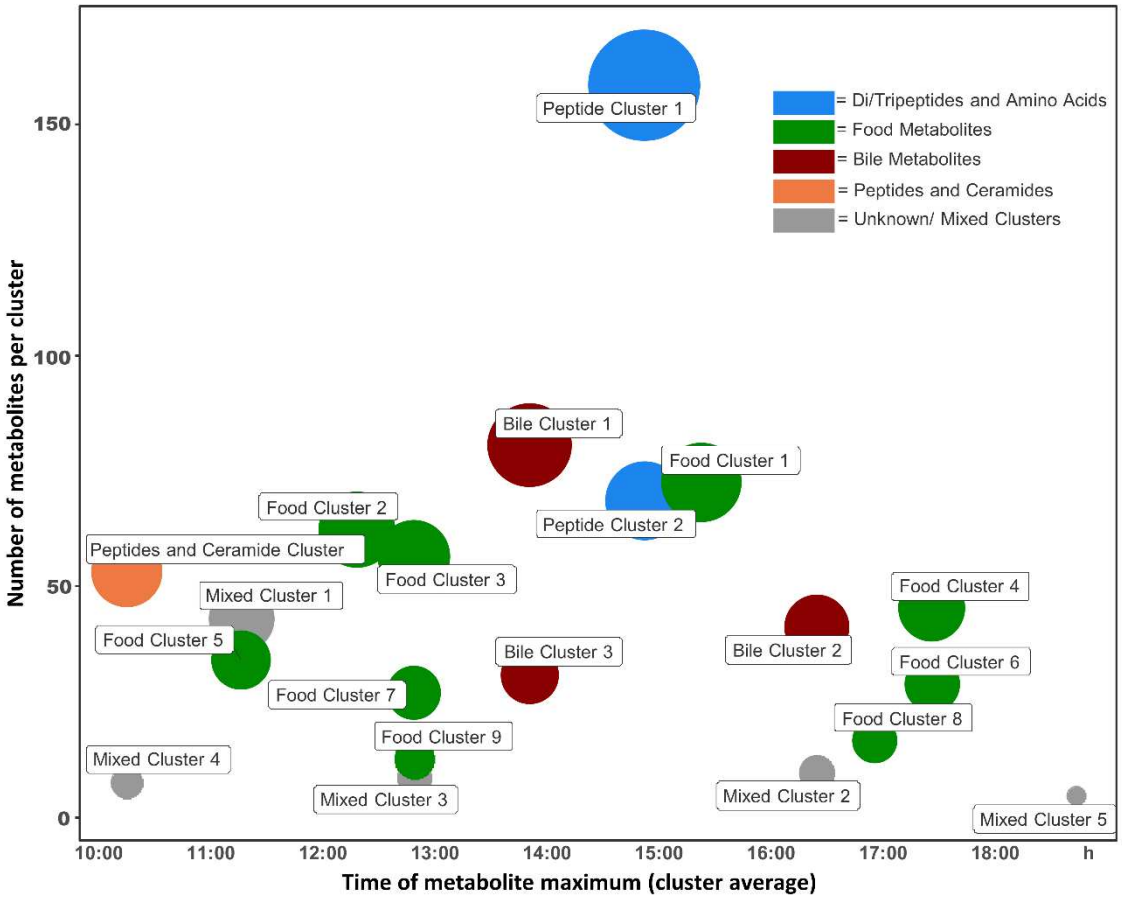


Figure 2.2. Metabolite groups using correlation-based clustering. The number of metabolites per cluster is represented by node size and position on the y-axis. The position on the x axis gives the time of day at which the clustered metabolites showed an average maximum. Cluster names are given manually by key metabolites in each cluster.

2.4.2 Food related metabolites

Several meals and beverages were consumed during the sample collection period between 10:15 – 18:30 h (Figure 2.1, Supplementary Table 1(see online)). Some meals led to food metabolite clusters unique to a single meal like arbutin and sugars linked to a

fruit puree (Figure 2.3 A). These food metabolites showed a clear maximum after food consumption, and then decreased to baseline levels. Similarly, some metabolites were relatively unique to milk (oligosaccharide sialyllactose[33,34], N-acetyl-lactosamine[34], butyrylcarnitine[35]) or fermented milk (cadaverine[36]) and showed distinct maximum levels after consumption of a fermented milk product at 16:35 h (Figure 2.3B). Alternatively, some food related clusters did not clearly indicate which meal the metabolites came from. One example of a non-specific food metabolite cluster is food cluster 1 which did not contain metabolites unique to a specific meal and had multiple spikes in intensity throughout the day (Figure 2.3 C). Similarly, food cluster 9 also showed multiple spikes throughout the day with high variability. Cluster 9 consisted of coffee biomarkers (trigonelline[37,38], caffeine[37,38] and catechol[39]) which generally follow coffee consumption timepoints (Figure 2.3 D) linking this cluster to drinking coffee products. These metabolites did not show well-defined peaks after coffee consumption. A previous study reported that coffee metabolites dwell for a long time after coffee consumption by quantifying trigonelline in saliva samples 16 hours after coffee consumption even after rinsing of the mouth[40]. Our study here confirms that some metabolites from food and beverage can remain in the upper GI tract for long periods after consumption. Some metabolites correlated with coffee biomarkers might be caused by an endogenous response to coffee intake, which warrants an interesting aspect for future investigations

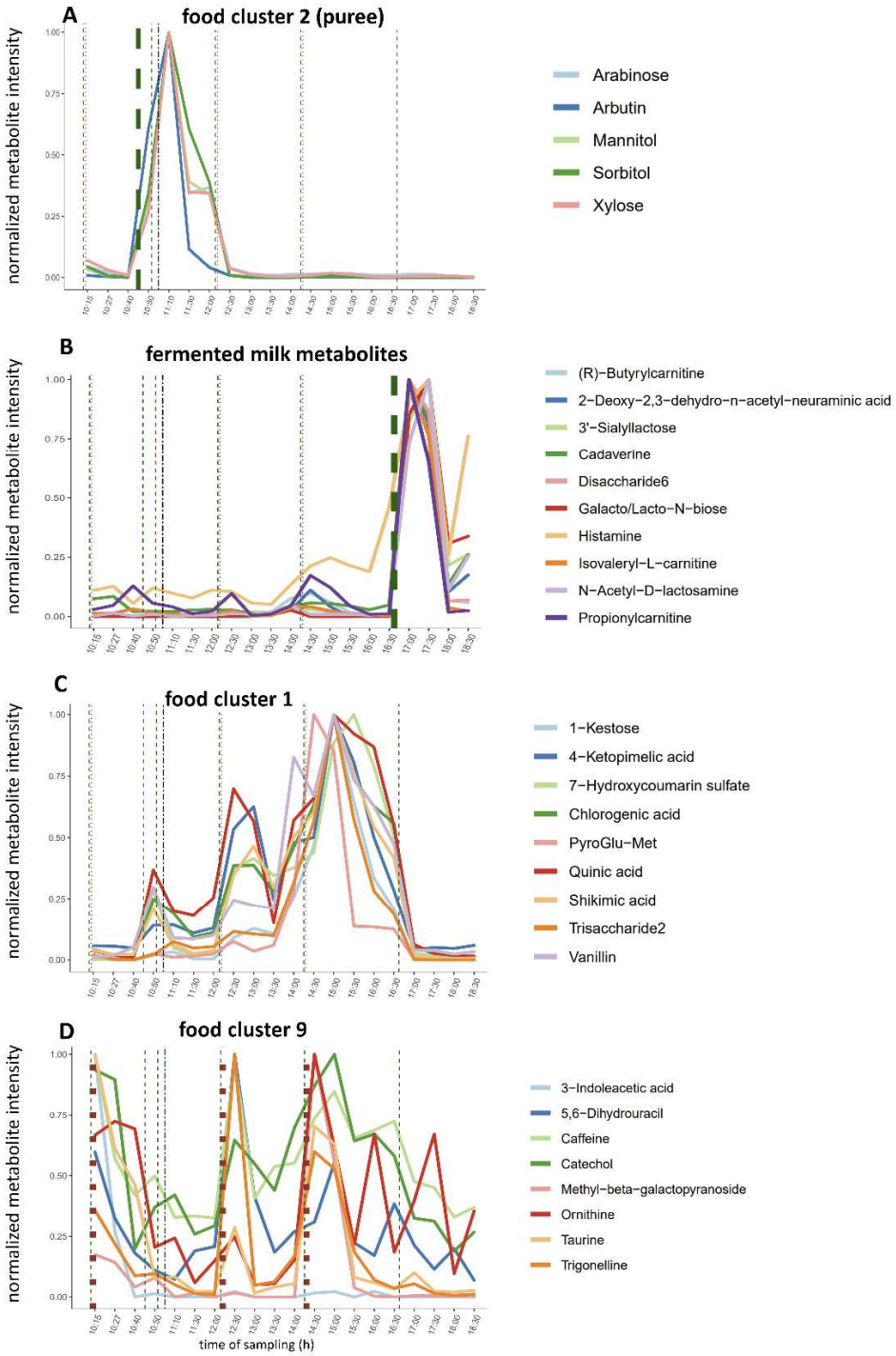


Figure 2.3. Intensity profiles of food related metabolites across the sampling period. Vertical lines represent meals (green dashed lines), coffee consumption (brown dotted lines) and acetaminophen consumption (black dashed line). Panel

A: bold green dashed line for fruit/vegetable puree. Panel B: bold green dashed line for fermented milk beverage. Panel C: metabolites grouped in food cluster 1 without direct relationship to timing of puree consumption. Panel D: bold brown dashed lines for coffee consumption.

Well-defined maxima in intensity of dietary diacylglycerides (DGs) and triacylglycerides (TGs) were observed immediately following consumption of 20 grams of olive oil with a slice of bread at 12:15 h (Figure 2.4 A). TGs are the main component of olive oil. The most common acyl chain lengths in olive oil are 18:1, 18:2, 16:0, and 18:0[41,42]. As expected, these same acyl chain lengths were highly represented in the TGs measured after olive oil consumption (Figure 2.4 A). The profile of free fatty acids showed less defined peaks throughout the sampling period such as oleic acid (fatty acid 18:1) which became more abundant after olive oil consumption but fluctuated in intensity for the rest of the afternoon (Figure 2.4 C). This trend may be explained by TGs of olive oil being hydrolyzed to free fatty acids by lipases present in the upper GI tract, followed by gradual absorption of the fatty acids over the following hours[43,44]. Interestingly, we observed a second spike of the same TGs three hours after olive oil consumption (Figure 2.4 A). One explanation for this second peak is the highly variable gastric retention time of humans which can be between 0 and 4 hours depending on food type[45,46] . This finding implies that some olive oil proceeded directly to the small intestine after consumption and the rest remained in the stomach with the solid bread of the meal until gastric emptying occurred hours later. The physical appearance of samples supports these findings as there was a lipid layer that formed on the top of the samples with high lipid abundance (samples collected from 12:30-13:00 h and 15:30-16:00 h), and not in samples between the two

abundant dietary TG spikes (Supplementary Table 1(see online)). Oleic acid (fatty acid 18:1) and arachidic acid (fatty acid 20:0) had higher abundances after olive oil consumption and showed maximum intensities at the same timepoints as the dietary TGs previously mentioned. Two fatty acyl esters of hydroxy fatty acids (FAHFAs) showed similar intensity profiles to oleic and arachidic acid (Figure 2.4 C). FAHFAs are a recently discovered class of lipids that have been shown to decrease insulin resistance and inflammation[47]. FAHFAs are produced endogenously[47] and also found in a variety of plants.[48,49] These FAHFAs are likely derived from olive oil given their strong correlation with other olive oil metabolites and their related fatty acyl constituents (Figure 2.4 C). We here report for the first time the presence of oleic acid-hydroxy oleic acid and oleic acid-hydroxy stearic acid FAHFAs in association with olive oil, and for the first time the finding of FAHFAs in the human GI tract. Finding these FAHFA constituents closely correlated with free fatty acids associated with olive oil suggests that FAHFAs are likely hydrolyzed and absorbed at similar rates compared to free fatty acids.

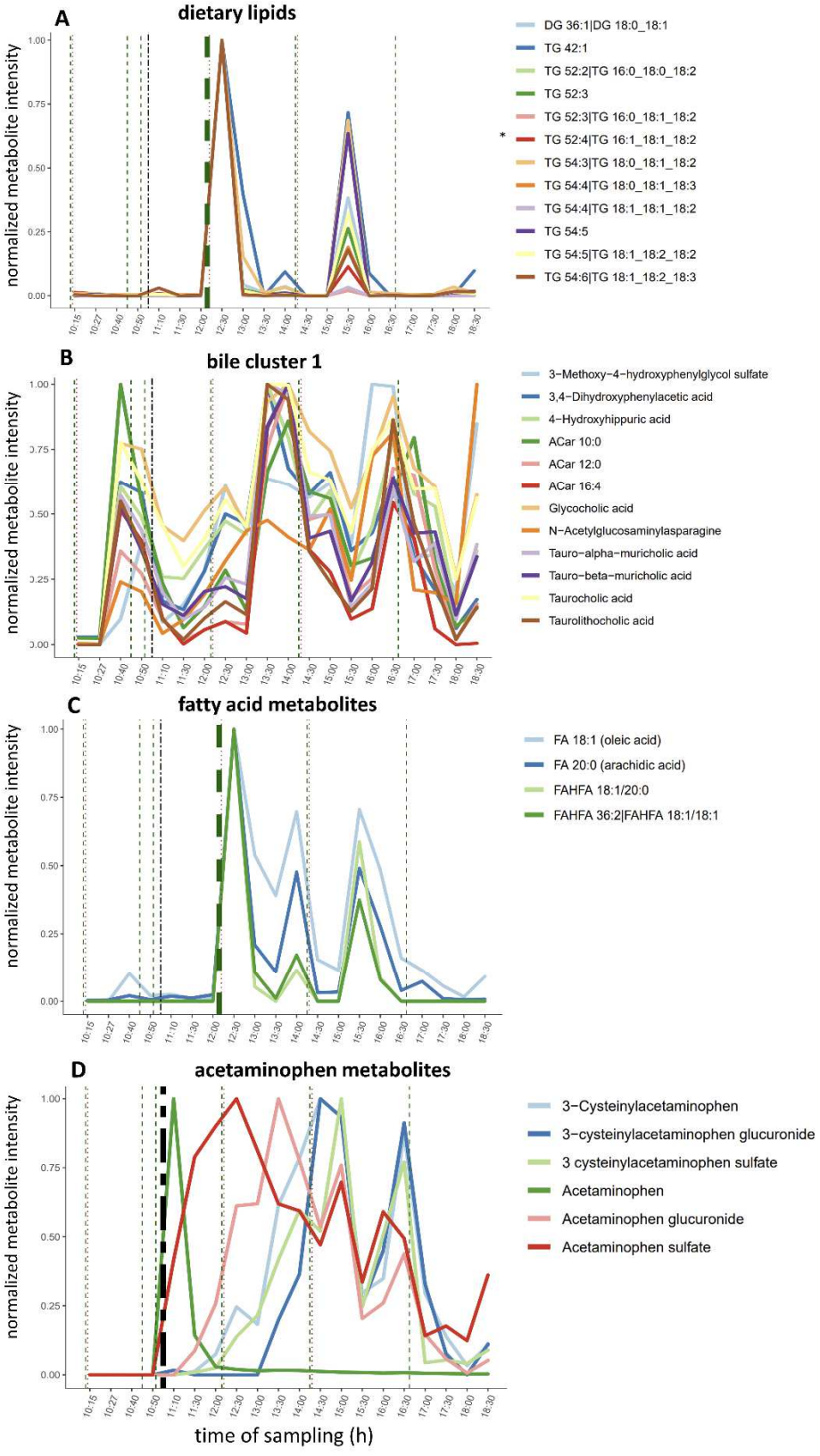

A
dietary lipids
normalized metabolite intensity
DG 36:1|DG 18:0_18:1
TG 42:1
TG 52:2|TG 16:0_18:0_18:2
TG 52:3
TG 52:3|TG 16:0_18:1_18:2
TG 52:4|TG 16:1_18:1_18:2
TG 54:3|TG 18:0_18:1_18:2
TG 54:4|TG 18:0_18:1_18:3
TG 54:4|TG 18:1_18:1_18:2
TG 54:5
TG 54:5|TG 18:1_18:2_18:2
TG 54:6|TG 18:1_18:2_18:3
*
B
bile cluster 1
normalized metabolite intensity
3−Methoxy−4−hydroxyphenylglycol sulfate
3,4−Dihydroxyphenylacetic acid
4−Hydroxyhippuric acid
ACar 10:0
ACar 12:0
ACar 16:4
Glycocholic acid
N−Acetylglucosaminylasparagine
Tauro−alpha−muricholic acid
Tauro−beta−muricholic acid
Taurocholic acid
Taurolithocholic acid
C
fatty acid metabolites
normalized metabolite intensity
FA 18:1 (oleic acid)
FA 20:0 (arachidic acid)
FAHFA 18:1/20:0
FAHFA 36:2|FAHFA 18:1/18:1
D
acetaminophen metabolites
normalized metabolite intensity
3−Cysteinylacetaminophen
3−cysteinylacetaminophen glucuronide
3 cysteinylacetaminophen sulfate
Acetaminophen
Acetaminophen glucuronide
Acetaminophen sulfate
time of sampling (h)

Figure 2.4. Intensity profiles of metabolite clusters across the sampling period. Vertical lines represent meals (green dashed lines) and acetaminophen consumption (black dashed line). Panel A: bold green dashed line for olive oil / bread with profiles of dietary lipids. Panel B: intensity profiles of 12 bile-associated metabolites. Panel C: bold green dashed line for olive oil / bread with profiles of fatty acids and FAHFAs. Panel D: bold black dashed line for acetaminophen intake, followed by profile of acetaminophen metabolite profiles. *marked annotation encompasses TG 18:0_18:1_18:2, and TG 18:1_18:1_18:1 due to coelution.

2.4.3 Bile related metabolites

Three separate metabolite clusters of bile-related metabolites were found with average maxima at 14:00 h and 16:30 h (Figure 2.2). These data suggest that over the course of 8 hours of sampling there were two primary bile excretion events, accompanied with lesser bile acid excretion events occurring throughout the sampling period such as a smaller peak at 12:30 h (Figure 2.4 B). These excretion events are characterized by the relative maxima of known bile related metabolites including all twelve detected bile acids, three steroid hormones, cholesterol, phase II exposome metabolites, phospholipids, and acylcarnitines[50,51] (Figure 2.2, Supplementary Table 4 (see online), Supplementary Figure 2 b,d). Apart from the key bile acids, bile cluster 3 mostly contained phosphatidylcholines (PCs) and cluster 1 had many lysophosphatidylcholines (LPCs), which are both known components of human bile. These clusters have maxima at similar timepoints compared to bile acids but show more defined spikes and then drop to baseline

levels for the remainder of the samples, likely due to rapid lipase action on PCs and LPCs (Supplementary Figure 1). Bile acids are the most focused on of bile components due to their importance in food digestion[12], cell injury and protection[52], and function as a signaling axis between the gut microbiota and host[53].

Bile acids are difficult to study in humans *in vivo* due to difficulty in sampling the human small intestine making this a valuable and unique view into bile excretion under normal conditions. The time dependent excretion of bile appeared to be strongly related to passage of dietary lipids through the upper small intestine. Lipids linked to dietary sources showed two sharp spikes in abundance at 12:30 h and 15:30 h (Figure 2.4 A). One hour after these spikes in lipid concentrations, major bile excretion events were recorded at 13:30 h and 16:30 h (Figure 2.4 B). This observation is in alignment with the known progression of lipid rich stomach chyme stimulating cells of the proximal small intestine to release hormone cholecystokinin into circulation leading to contraction of the gallbladder and excretion of bile into the intestinal tract[54]. The postprandial response of bile components also aligns with the finding that bile acids in circulation are governed by meal intake, as opposed to hepatic bile acid synthesis, which has a defined circadian rhythm[55]. Previous studies of bile excretion report “time after test meal” to predict gallbladder contraction[3,10,56,57]. In our study, we are better able to give a specific metric of dietary lipids in the small intestine to better measure the excretion of bile compared to time after consuming meals. Our study gives a unique perspective on the time dependency of known bile metabolite excretion relating to dietary lipids.

In addition, we found a range of both known and novel metabolites to be correlated with bile. Specifically, the metabolite 4-hydroxyhippuric acid is clustered with bile metabolites

and has a Spearman rank correlation coefficient of $r_{xy} > 0.9$ to six different bile acids and several other bile related metabolites (Supplementary Table 5 (see online)). This metabolite is a product of microbial transformation of dietary polyphenols followed by liver transformation and glycination[58] and has been reported as a urinary biomarker for fruit and vegetable intake in humans[59]. Metabolite 4-hydroxyhippuric acid has been measured in mouse bile[60], however this is the first report of this metabolite detected in relation to human bile. Another metabolite, N-acetylglucosaminylasparagine, is a glycoprotein breakdown product reported to be disposed of through urine[61] and is here associated with bile metabolites suggesting bile as an additional disposal route for this metabolite. Another metabolite clustered with bile metabolites is a product of norepinephrine metabolism, 3-methoxy-4-hydroxyphenylglycol sulfate. This metabolite has been reported as a urinary metabolite, but not as a bile related metabolite. It has been previously proposed that bile is a plausible disposal route for 3-methoxy-4-hydroxyphenylglycol sulfate in addition to urine[62]; however, we here present the first direct measurements to validate this link. Urine and bile are routes of excretion for metabolic end products and exogenous metabolites[13] which makes the link of these known urinary metabolites in bile for the first time a plausible biological outcome.

2.4.4 Acetaminophen and related metabolites

A dose of 500 mg of the general analgesic acetaminophen was taken orally mid-morning on the day of this experiment (Figure 1.1). Acetaminophen abundance in the upper GI tract lumen immediately rose to its highest level 15 minutes later. After 35 minutes the abundance of acetaminophen fell to about 15% of its maximal abundance and continued to fall to baseline level over the following hours. After absorption, acetaminophen is

transformed in the liver primarily to acetaminophen sulfate, acetaminophen glucuronide, and cysteinylacetaminophen[63,64]. These three conjugates were detected, and the intensity profiles support the route of liver transformation of acetaminophen, biliary excretion, and collection from the upper small intestine (Figure 2.4 D). Interestingly, each conjugate showed a unique profile. Acetaminophen sulfate appeared immediately after acetaminophen ingestion and had a maximal abundance after 1.5 hours, in contrast to acetaminophen glucuronide that did not appear until 30 minutes after acetaminophen ingestion and had maximal abundance after 2.5 hours. The phase III conjugate 3-cysteinylacetaminophen slowly increased in abundance and did not reach maximal intensity until 3.5 hours after acetaminophen ingestion (Figure 2.4 D). Biliary excretion of these conjugated metabolites is further supported as the source of these metabolites because we detected a local maximum for these conjugates at 16:30 h, the same timepoint as found for one of the major bile excretion events (Figure 2.4 B). The unique profile for each of these acetaminophen related metabolites is in agreement with the known pharmacokinetic properties of acetaminophen. After oral ingestion of acetaminophen, plasma acetaminophen has maximum concentration within less than 1 hour, followed by the maximum of the sulfate conjugate between 1 and 1.5 hours, and the glucuronide conjugate maximum at 2 to 2.5 hours[64]. The immediate appearance of acetaminophen sulfate in this experiment, and no immediate appearance in acetaminophen glucuronide, is expected because intestinal epithelium cells have the capacity to convert acetaminophen to acetaminophen sulfate *in vitro*, while showing little to no capacity to convert acetaminophen to acetaminophen glucuronide[65].

Next, we wondered if additional downstream metabolites of acetaminophen might be detectable in the GI tract. To make this end, we used the BioTransformer software[66] that predicts possible metabolic products of drugs using known phase II metabolism reactions from humans. BioTransformer predicted the presence of acetaminophen sulfate and acetaminophen glucuronide from acetaminophen, however 3-cysteinylacetaminophen was not predicted, likely because it requires multiple reactions to form (conjugation to glutathione then degradation to 3-cysteinylacetaminophen[67]). Importantly, BioTransformer predicted additional acetaminophen metabolites for which no public MS/MS spectra are available, neither in MassBank.us nor NIST20. Instead, we queried unknown LC-MS/MS features using accurate mass matches to the candidate list provided through BioTransformer. Two BioTransformer candidate accurate mass matches were 3-cysteinylacetaminophen glucuronide and 3-cysteinylcysteine acetaminophen sulfate, both with less than 1.8 ppm mass difference to the theoretical elemental compositions. The glucuronide conjugate was too low abundant to trigger a data-dependent MS/MS event, but the MS/MS spectrum for the 3-cysteinylcysteine acetaminophen sulfate included a typical M-80 representing neutral loss of SO_3, confirming the likely identification of this compound (Supplementary Table 4 (see online)). While these two metabolites have been reported before in urine[68], and other authors predicted these acetaminophen metabolites to occur in mammals[69], no MS/MS spectra were published to date. Both metabolites showed similar time profile as 3-cysteinylacetaminophen (Supplementary Figure 2), providing additional biological support for structural annotation of these acetaminophen metabolites. Measurement of these acetaminophen metabolites

in relation to bile highlights the use of bile clusters in the GI tract for understanding drug metabolism in humans.

2.4.5 Ceramide and Peptide Clusters

One cluster of metabolites contained many ceramides with high degree of intercorrelations (Supplementary Figure 2A). Ceramide abundance showed five very pronounced maxima throughout the eight-hour sampling period that did not correlate to other physiological or environmental stimuli such as acetaminophen, dietary ingestions, bile excretions or dietary lipids. Ceramides are important as structural lipids and signaling molecules particularly within the GI tract[70]. These lipids are produced in many tissues throughout the body including GI tract epithelium and are also present in many foods[70]. Our findings presented here support the notion that GI tract ceramides are produced endogenously and are regulated tightly in frequent intervals.

The cluster of GI tract ceramides had maxima that trended with dozens of di- and tripeptides and free amino acids maxima (Supplementary Figure 2C). Such short-chain peptides in these clusters might originate from incomplete protein degradation and found in the GI tract either through gastric or pancreatic excretions. Both the stomach and pancreas secrete proteases that could be responsible for digesting proteins down to short peptides and amino acids. The abundance profile of these clusters show 2-hourly maxima at 10:30h, 12:30h, 14:30h and 16:30h with an additional minor maximum at 13:30h. Similar to ceramides, these maxima did not directly coincide with major bile excretions or dietary lipids. Pancreatic juice is a relatively little investigated biofluid[71]. However, multiple studies have reported a selection of free amino acids as important components of

pancreatic juice[71] including phenylalanine, tyrosine, tryptophan, valine, leucine, isoleucine and alanine. We here find all of these amino acids associated with pancreatic juice in conjunction with di- and tripeptide clusters (Supplementary Figure 2B). Di- and tripeptides are regularly reported in metabolomics studies[72,73] but are absent from most classic nutritional studies. Our data support the notion that the observed dipeptide clusters may be derived from pancreatic juice; yet this evidence is not conclusive because amino acids are certainly not unique to pancreatic juice.

2.5 Conclusion

We here present a technique to enable a unique perspective into the human upper GI tract *in vivo* over the course of eight hours. Correlation based clustering connected metabolites of similar biological function to be investigated in temporal profiles and to be associated with important physiological and dietary events. Interestingly, even classic nutritional compounds like amino acids showed clear temporal profiles that were not exclusively related to dietary input. Bile metabolites spiked roughly 1.5 hours after measurement of dietary lipids passing the upper small intestine. This observation presents a unique measurement because in humans, bile excretion has been measured almost exclusively under fasted conditions and not during a normal day of *ad libitum* meal consumption. We here report specific metabolites (4-hydroxyhippuric acid, N-acetylglucosaminylasparagine, 3-methoxy-4-hydroxyphenylglycol sulfate) to be linked to human bile. This experiment also gave a unique insight into acetaminophen metabolism and excretion and presents the utility of bile associated metabolites to find endogenously modified drug metabolites. The human GI tract is an extraordinarily complex and dynamic system and metabolomics experiments offer a valuable approach to discover how the gut

interacts with food. This single subject study presents trends that may not be representative of the general population. Future experiments are needed to determine which trends can be expected from the population as a whole. Although this is a single subject study the dataset is rich in information and provides unique findings into the metabolome of the upper human GI tract *in vivo*.

2.6 References

1 R. Kiesslich, M. Goetz, E. M. Angus, Q. Hu, Y. Guan, C. Potten, T. Allen, M. F. Neurath, N. F. Shroyer, M. H. Montrose and A. J. M. Watson, Identification of Epithelial Gaps in Human Small and Large Intestine by Confocal Endomicroscopy, *Gastroenterology*, 2007, **133**, 1769–1778.

2 C. L. Wright and R. H. Riddell, Histology of the Stomach and Duodenum in Crohn's Disease, *Am. J. Surg. Pathol.*, 1998, **22**, 383–390.

3 P. Portincasa, A. Moschetta, A. Colecchia, D. Festi and G. Palasciano, Measurements of gallbladder motor function by ultrasonography: Towards standardization, *Dig. Liver Dis.*, 2003, **35**, 56–61.

4 L. Ellegård and H. Andersson, Oat bran rapidly increases bile acid excretion and bile acid synthesis: An ileostomy study, *Eur. J. Clin. Nutr.*, 2007, **61**, 938–945.

5 L. Ohlsson, E. Hertervig, B. A. G. Jönsson, R. D. Duan, L. Nyberg, R. Svernlöv and Å. Nilsson, Sphingolipids in human ileostomy content after meals containing milk sphingomyelin, *Am. J. Clin. Nutr.*, 2010, **91**, 672–678.

6 I. Mainville, Y. Arcand and E. R. Farnworth, A dynamic model that simulates the human upper gastrointestinal tract for the study of probiotics, *Int. J. Food Microbiol.*,

2005, **99**, 287–296.

7 Q. Zhang, G. Widmer and S. Tzipori, A pig model of the human gastrointestinal tract, *Gut Microbes*, 2013, **4**, 193–200.

8 E. G. Zoetendal, J. Raes, B. Van Den Bogert, M. Arumugam, C. C. Booijink, F. J. Troost, P. Bork, M. Wels, W. M. De Vos and M. Kleerebezem, The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates, *ISME J.*, 2012, **6**, 1415–1426.

9 J. L. Boyer, Bile formation and secretion, *Compr. Physiol.*, 2013, **3**, 1035–1078.

10 P. J. Howard, G. M. Murphy and R. H. Dowling, Gall bladder emptying patterns in response to a normal meal in healthy subjects and patients with gall stones: Ultrasound study, *Gut*, 1991, **32**, 1406–1411.

11 A. J. Fillery-Travis, L. H. Foster and M. M. Robins, Stability of emulsions stabilised by two physiological surfactants: l-α-phosphatidylcholine and sodium taurocholate, *Biophys. Chem.*, 1995, **54**, 253–260.

12 M. Armand, P. Borel, B. Pasquier, C. Dubois, M. Senft, M. Andre, J. Peyrot, J. Salducci and D. Lairon, Physicochemical characteristics of emulsions during fat digestion in human stomach and duodenum, *Am. J. Physiol. - Gastrointest. Liver Physiol.*, 1996, **271**, G172-183.

13 M. J. Zamek-Gliszczynski, K. A. Hoffmaster, K. I. Nezasa, M. N. Tallman and K. L. R. Brouwer, Integration of hepatic drug transporters and phase II metabolizing enzymes: Mechanisms of hepatic excretion of sulfate, glucuronide, and glutathione

metabolites, *Eur. J. Pharm. Sci.*, 2006, **27**, 447–486.

14 V. Urdaneta and J. Casadesús, Interactions between bacteria and bile salts in the gastrointestinal and hepatobiliary tracts, *Front. Med.*, 2017, **4**, 1–13.

15 J. W. Lampe, Biomarkers of nutritional exposure and nutritional status: An overview, *J. Nutr.*, 2003, **133**, 956–964.

16 M. Garcia-Aloy, M. Ulaszewska, P. Franceschi, S. Estruel-Amades, C. H. Weinert, A. Tor-Roca, M. Urpi-Sarda, F. Mattivi and C. Andres-Lacueva, Discovery of Intake Biomarkers of Lentils, Chickpeas, and White Beans by Untargeted LC–MS Metabolomics in Serum and Urine, *Mol. Nutr. Food Res.*, 2020, **64**, 1901137.

17 F. Madrid-Gambin, C. Brunius, M. Garcia-Aloy, S. Estruel-Amades, R. Landberg and C. Andres-Lacueva, Untargeted 1H NMR-Based Metabolomics Analysis of Urine and Serum Profiles after Consumption of Lentils, Chickpeas, and Beans: An Extended Meal Study to Discover Dietary Biomarkers of Pulses, *J. Agric. Food Chem.*, 2018, **66**, 6997–7005.

18 V. Matyash, G. Liebisch, T. V. Kurzchalia, A. Shevchenko and D. Schwudke, Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics, *J. Lipid Res.*, 2008, **49**, 1137–1146.

19 D. K. Barupal, Y. Zhang, T. Shen, S. Fan, B. S. Roberts, P. Fitzgerald, B. Wancewicz, L. Valdiviez, G. Wohlgemuth, G. Byram, Y. Y. Choy, B. Haffner, M. R. Showalter, A. Vaniya, C. S. Bloszies, J. S. Folz, T. Kind, A. M. Flenniken, C. McKerlie, L. M. J. Nutter, K. C. Lloyd and O. Fiehn, A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international

mouse phenotyping consortium, *Metabolites*, 2019, **9**, 9050101.

20 H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z. J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita and M. Arita, A lipidome atlas in MS-DIAL 4, *Nat. Biotechnol.*, 2020, **38**, 1159–1163.

21 B. C. DeFelice, S. S. Mehta, S. Samra, T. Čajka, B. Wancewicz, J. F. Fahrmann and O. Fiehn, Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing, *Anal. Chem.*, 2017, **89**, 3250–3255.

22 E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ru, H. P. Singer and J. Hollender, Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence, *Environ. Sci. Technol.*, 2014, **48**, 2097–2098.

23 P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal and O. Fiehn, Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics, *Anal. Chem.*, 2020, **92**, 7515–7522.

24 J. E. Schollée, E. L. Schymanski, M. A. Stravs, R. Gulde, N. S. Thomaidis and J. Hollender, Similarity of High-Resolution Tandem Mass Spectrometry Spectra of Structurally Related Micropollutants and Transformation Products, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 2692–2704.

25 J. H. Ward, Hierarchical Grouping to Optimize an Objective Function, *J. Am. Stat. Assoc.*, 1963, **58**, 236–244.

26 P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 1987, **20**, 53–65.

27 M. Charrad, N. Ghazzali, V. Boiteau and A. Niknafs, Nbclust: An R package for determining the relevant number of clusters in a data set, *J. Stat. Softw.*, 2014, **61**, 1–36.

28 M. J. Rusilowicz, M. Dickinson, A. J. Charlton, S. O'Keefe and J. Wilson, MetaboClust: Using interactive time-series cluster analysis to relate metabolomic data with perturbed pathways, *PLoS One*, 2018, **13**, e0205968.

29 D. K. Barupal and O. Fiehn, Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets, *Sci. Rep.*, 2017, **7**, 14567.

30 D. C. Nieman, N. D. Gillitt, W. Sha, M. P. Meaney, C. John, K. L. Pappan and J. M. Kinchen, Metabolomics-Based Analysis of Banana and Pear Ingestion on Exercise Performance and Recovery, *J. Proteome Res.*, 2015, **14**, 5367–5377.

31 M. Ulaszewska, N. Vázquez-Manjarrez, M. Garcia-Aloy, R. Llorach, F. Mattivi, L. O. Dragsted, G. Praticò and C. Manach, Food intake biomarkers for apple, pear, and stone fruit Lars Dragsted, *Genes Nutr.*, 2018, **13**, 1–16.

32 A. Escarpa and M. C. González, Evaluation of high-performance liquid chromatography for determination of phenolic compounds in pear horticultural cultivars, *Chromatographia*, 2000, **51**, 37–43.

33 V. Kelly, S. Davis, S. Berry, J. Melis, R. Spelman, R. Snell, K. Lehnert and D.

Palmer, Rapid , quantitative analysis of 3 c - and 6 c -sialyllactose in milk by flow-injection analysis – mass spectrometry : Screening of milks for naturally elevated sialyllactose concentration, *J. Dairy Sci.*, 2013, **96**, 7684–7691.

34 K. A. Al Busadah, S. D. Carrington, S. Albrecht, J. A. Lane, K. Marin, R. M. Hickey and P. M. Rudd, A comparative study of free oligosaccharides in the milk of domestic animals, *Br. J. Nutr.*, 2014, **111**, 1313–1328.

35 K. J. Boudonck, M. W. Mitchell, J. Wulff and J. A. Ryals, Characterization of the biochemical variability of bovine milk using metabolomics, *Metabolomics*, 2009, **5**, 375–386.

36 O. Ozdestan and A. Uren, Biogenic amine content of kefir : a fermented dairy product, *Eur. Food Res. Technol.*, 2010, **231**, 101–107.

37 K. A. Guertin, E. Loftfield, S. M. Boca, J. N. Sampson, S. C. Moore, Q. Xiao, W. Y. Huang, X. Xiong, N. D. Freedman, A. J. Cross and R. Sinha, Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer, *Am. J. Clin. Nutr.*, 2015, **101**, 1000–1011.

38 J. A. Rothwell, P. Keski-Rahkonen, N. Robinot, N. Assi, C. Casagrande, M. Jenab, P. Ferrari, M. C. Boutron-Ruault, Y. Mahamat-Saleh, F. R. Mancini, H. Boeing, V. Katzke, T. Kühn, K. Niforou, A. Trichopoulou, E. Valanou, V. Krogh, A. Mattiello, D. Palli, C. Sacerdote, R. Tumino and A. Scalbert, A Metabolomic Study of Biomarkers of Habitual Coffee Intake in Four European Countries, *Mol. Nutr. Food Res.*, 2019, **63**, 1900659.

39 R. Lang, C. Mueller and T. Hofmann, Development of a stable isotope dilution analysis with liquid chromatography-tandem mass spectrometry detection for the quantitative analysis of di- and trihydroxybenzenes in foods and model systems, *J. Agric. Food Chem.*, 2006, **54**, 5755–5762.

40 R. Lang, A. Wahl, T. Stark and T. Hofmann, in *ACS Symposium Series*, 2012, vol. 1098, pp. 13–25.

41 A. Ranalli, L. Pollastri, S. Contento, G. Di Loreto, E. Iannucci, L. Lucera and F. Russi, Acylglycerol and fatty acid components of pulp, seed, and whole olive fruit oils. Their use to characterize fruit variety by chemometrics, *J. Agric. Food Chem.*, 2002, **50**, 3775–3779.

42 E. Alves, T. Melo, M. P. Barros, M. M. R. Domingues and P. Domingues, Lipidomic Profiling of the Olive (Olea europaea L.) Fruit towards Its Valorisation as a Functional Food: In-depth identification of triacylglycerols and polar lipids in Portuguese olives, *Molecules*, 2019, **24**, 24142555.

43 M. Armand, Lipases and lipolysis in the human digestive tract: Where do we stand?, *Curr. Opin. Clin. Nutr. Metab. Care*, 2007, **10**, 156–164.

44 Z. Ye, C. Cao, R. Li, P. Cao, Q. Li and Y. Liu, Lipid composition modulates the intestine digestion rate and serum lipid status of different edible oils: A combination of in vitro and in vivo studies, *Food Funct.*, 2019, **10**, 1490–1503.

45 R. K. Goyal, Y. Guo and H. Mashimo, Advances in the physiology of gastric emptying, *Neurogastroenterol. Motil.*, 2019, **31**, e13546.

46 S. Hellmig, F. Von Schöning, C. Gadow, S. Katsoulis, J. Hedderich, U. R. Fölsch and E. Stüber, Gastric emptying time of fluids and solids in healthy subjects determined by 13C breath tests: Influence of age, sex and body mass index, *J. Gastroenterol. Hepatol.*, 2006, **21**, 1832–1838.

47 M. M. Yore, I. Syed, P. M. Moraes-Vieira, T. Zhang, M. A. Herman, E. A. Homan, R. T. Patel, J. Lee, S. Chen, O. D. Peroni, A. S. Dhaneshwar, A. Hammarstedt, U. Smith, T. E. McGraw, A. Saghatelian and B. B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects, *Cell*, 2014, **159**, 318–332.

48 M. J. Kolar, S. Konduri, T. Chang, H. Wang, C. McNerlin, L. Ohlsson, M. Härröd, D. Siegel and A. Saghatelian, Linoleic acid esters of hydroxy linoleic acids are anti-inflammatory lipids found in plants and mammals, *J. Biol. Chem.*, 2019, **294**, 10698–10707.

49 A. M. Liberati-Čizmek, M. Biluš, A. L. Brkić, I. C. Barić, M. Bakula, A. Hozić and M. Cindrić, Analysis of Fatty Acid Esters of Hydroxyl Fatty Acid in Selected Plant Food, *Plant Foods Hum. Nutr.*, 2019, **74**, 235–240.

50 I. M. A. Brüggenwirth, R. J. Porte and P. N. Martins, Bile Composition as a Diagnostic and Prognostic Tool in Liver Transplantation, *Liver Transplant.*, 2020, **26**, 1177–1187.

51 B. L. Shneider, P. Rinaldo, S. Emre, J. Bucuvalas, R. Squires, M. Narkewicz, G. Gondolesi, M. Magid, R. Morotti and L. S. Hynan, Abnormal concentrations of esterified carnitine in bile: A feature of pediatric acute liver failure with poor

prognosis, *Hepatology*, 2005, **41**, 717–721.

52 M. J. Perez and O. Britz, Bile-acid-induced cell injury and protection, *World J. Gastroenterol.*, 2009, **15**, 1677–1689.

53 P. B. Hylemon, W. M. Pandak, G. Gil, H. Zhou, S. Ren and P. Dent, Bile acids as regulatory molecules, *J. Lipid Res.*, 2009, **50**, 1509–1520.

54 J. R. Grider, Role of Cholecystokinin in the Regulation of Gastrointestinal Motility, *J. Nutr.*, 1994, **124**, 1334S–1339S.

55 C. Gälman, B. Angelin and M. Rudling, Bile acid synthesis in humans has a rapid diurnal variation that is asynchronous with cholesterol synthesis, *Gastroenterology*, 2005, **129**, 1445–1453.

56 R. P. Jazrawi, P. Pazzi, M. Letizia Petroni, N. Prandini, C. Paul, J. A. Adam, S. Gullini and T. C. Northfield, Postprandial gallbladder motor function: Refilling and turnover of bile in health and in cholelithiasis, *Gastroenterology*, 1995, **109**, 582–591.

57 M. Acalovschi, D. L. Dumitraçcu and I. Csakany, Gastric and gall bladder emptying of a mixed meal are not coordinated in liver cirrhosis - A simultaneous sonographic study, *Gut*, 1997, **40**, 412–417.

58 A. R. Rechner, M. A. Smith, G. Kuhnle, G. R. Gibson, E. S. Debnam, S. K. S. Srai, K. P. Moore and C. A. Rice-Evans, Colonic metabolism of dietary polyphenols: Influence of structure on microbial fermentation products, *Free Radic. Biol. Med.*, 2004, **36**, 212–225.

59 M. Beckmann, T. Wilson, H. Zubair, A. J. Lloyd, L. Lyons, H. Phillips, K. Tailliart, N. Gregory, R. Thatcher, I. Garcia-Perez, G. Frost, J. M. Mathers and J. Draper, A Standardized Strategy for Simultaneous Quantification of Urine Metabolites to Validate Development of a Biomarker Panel Allowing Comprehensive Assessment of Dietary Exposure, *Mol. Nutr. Food Res.*, 2020, **64**, 1–15.

60 N. Orrego-Lagarón, M. Martínez-Huélamo, A. Vallverdú-Queralt, R. M. Lamuela-Raventos and E. Escribano-Ferrer, High gastrointestinal permeability and local metabolism of naringenin: Influence of antibiotic treatment on absorption and metabolism, *Br. J. Nutr.*, 2015, **114**, 169–180.

61 P. Aula, K. O. Raivio and P. Maury, Variation of urinary excretion of aspartylglucosamine and associated clinical findings in aspartylglucosaminuria, *J. Inherit. Metab. Dis.*, 1980, **3**, 159–163.

62 G. Eisenhofer, A. Åneman, D. Hooper, B. Rundqvist and P. Friberg, Mesenteric organ production, hepatic metabolism, and renal elimination of norepinephrine and its metabolites in humans, *J. Neurochem.*, 1996, **66**, 1565–1573.

63 A. Toda, M. Shimizu, S. Uehara, T. Sasaki, T. Miura, M. Mogi, M. Utoh, H. Suemizu and H. Yamazaki, Plasma and hepatic concentrations of acetaminophen and its primary conjugates after oral administrations determined in experimental animals and humans and extrapolated by pharmacokinetic modeling, *Xenobiotica*, 2021, **51**, 316–323.

64 L. Prescott, Kinetics and metabolism of paracetamol and phenacetin., *Br. J. Clin. Pharmacol.*, 1980, **10**, 291S-298S.

65 S. Siissalo, L. Laine, A. Tolonen, A. M. Kaukonen, M. Finel and J. Hirvonen, Caco-2 cell monolayers as a tool to study simultaneous phase II metabolism and metabolite efflux of indomethacin, paracetamol and 1-naphthol, *Int. J. Pharm.*, 2010, **383**, 24–29.

66 Y. Djoumbou-Feunang, J. Fiamoncini, A. Gil-de-la-Fuente, R. Greiner, C. Manach and D. S. Wishart, BioTransformer: A comprehensive computational tool for small molecule metabolism prediction and metabolite identification, *J. Cheminform.*, 2019, **11**, 1–25.

67 M. R. McGill and H. Jaeschke, Metabolism and disposition of acetaminophen: Recent advances in relation to hepatotoxicity and diagnosis, *Pharm. Res.*, 2013, **30**, 2174–2187.

68 E. Pujos-Guillot, G. Pickering, B. Lyan, G. Ducheix, M. Brandolini-Bunlon, F. Glomot, D. Dardevet, C. Dubray and I. Papet, Therapeutic paracetamol treatment in older persons induces dietary and metabolic modifications related to sulfur amino acids, *Age (Omaha).*, 2012, **34**, 181–193.

69 S. J. Hart, I. C. Calder and J. D. Tange, The Metabolism and Toxicity of Paracetamol in Sprague-Dawley and Wistar Rats, *Eur. J. Drug Metab. Pharmacokinet.*, 1982, **7**, 203–222.

70 K. Kurek, B. Łukaszuk, D. M. Piotrowska, P. Wiesiołek, A. M. Chabowska and M. Zendzian-Piotrowska, Metabolism, physiological role, and clinical implications of sphingolipids in gastrointestinal tract, *Biomed Res. Int.*, 2013, **2013**, 908907.

71 N. Cortese, G. Capretti, M. Barbagallo, A. Rigamonti, P. G. Takis, G. F. Castino, D.

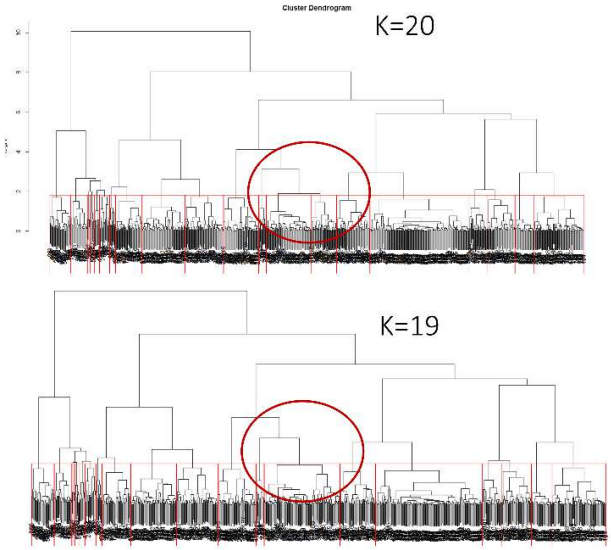
Vignali, G. Maggi, F. Gavazzi, C. Ridolfi, G. Nappo, G. Donisi, M. Erreni, R. Avigni, D. Rahal, P. Spaggiari, M. Roncalli, P. Cappello, F. Novelli, P. Monti, A. Zerbi, P. Allavena, A. Mantovani and F. Marchesi, Metabolome of pancreatic juice delineates distinct clinical profiles of pancreatic cancer and reveals a link between glucose metabolism and PD-1+Cells, *Cancer Immunol. Res.*, 2020, **8**, 493–505.

72 S. Ichikawa, M. Morifuji, H. Ohara, H. Matsumoto, Y. Takeuchi and K. Sato, Hydroxyproline-containing dipeptides and tripeptides quantified at high concentration in human blood after oral administration of gelatin hydrolysate, *Int. J. Food Sci. Nutr.*, 2010, **61**, 52–60.

73 C. Ibáñez, C. Simó, V. García-Cañas, Á. Gómez-Martínez, J. A. Ferragut and A. Cifuentes, CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation, *Electrophoresis*, 2012, **33**, 2328–2336.

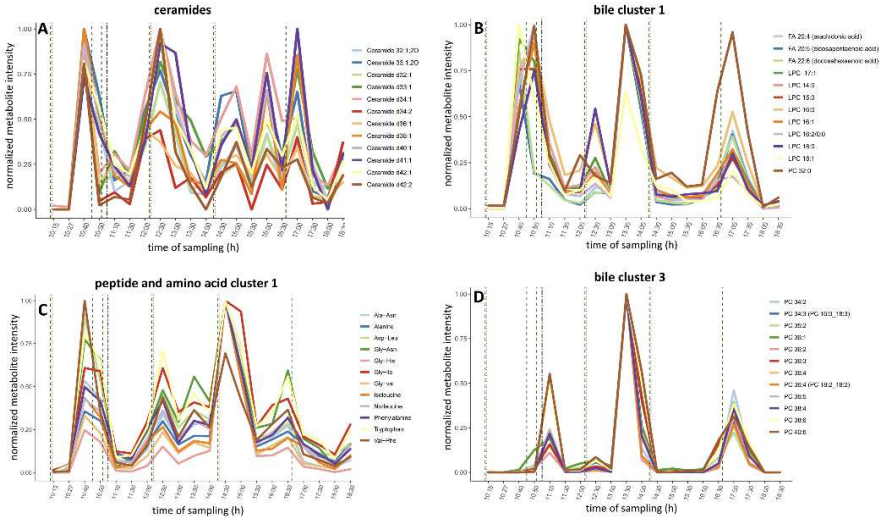
2.7 Supplemental Figures

Supplemental Figure 1
for **Metabolomics Analysis of Time-Series Human Small Intestine Lumen Samples Collected *in vivo***
Jacob S. Folz[a], Dari Shalon[b], Oliver Fiehn*[a]
[a] West Coast Metabolomics Center and Department of Food Science and Technology, University of California Davis, Davis, CA, USA
[b] EnVivo Bio, Inc., San Carlos, CA 94070, USA



Supplemental Figure 1. The smallest optimal number of clusters was determined by reducing the number of clusters (k) by one from 26 original clusters. Dendograms are generated from correlation of intensities from metabolites annotated in this study. These dendograms show two groups (Food cluster 1 and bile cluster 2) being combined between k=20 and k=19 (cluster highlighted with red circle).

Supplemental Figure 2
for **Metabolomics Analysis of Time-Series Human Small Intestine Lumen Samples Collected *in vivo***
Jacob S. Folz[a], Dari Shalon[b], Oliver Fiehn*[a]
[a] West Coast Metabolomics Center and Department of Food Science and Technology, University of California Davis, Davis, CA, USA
[b] EnVivo Bio, Inc., San Carlos, CA 94070, USA



Supplemental Figure 2. Four groups of metabolites. Each graph shows the intensity profile of metabolites across the testing period. Vertical lines represent meals (green dashed lines), coffee consumption (brown dotted lines) and acetaminophen consumption (black dashed line). Intensities are normalized to the highest intensity for each metabolite. Panel A contains ceremide lipids. Panel B contains bile related metabolites showing mostly lysophosphatidylcholines (LPCs). Panel C includes metabolites clustered in the di- and tripeptide and amino acid cluster 1. Panel D contains metabolites of bile cluster 3 showing exclusively phosphatidylcholines (PC).

Chapter 3: Bile Acids in Humans with Emphasis on the GI Tract

3.1 Abstract

Bile acids have been known to be important in the digestion of lipids for decades. More recently, bile acids have also been implicated as systemic signaling molecules[1] and are associated with many diseases. Bile acids are regulated by human liver enzymes in interplay with gut microbiota and gut intestinal enzymes. Bile acids are at highest concentration in the intestinal tract but are present and detectable at lower concentrations throughout nearly all tissues of the body. Accurate measurement of bile acids in different sample types requires careful consideration of analytical techniques to cover the large variation in concentrations between tissues, and between different bile acid species. This chapter first provides background on the biological importance of bile acids and LC-MS/MS measurement of bile acids. Experimental data in this chapter further highlights the use of bile acid quantification in human serum, stool, and samples acquired from the lumen of the intestinal tract of 15 subjects. A flaxseed dietary intervention did not cause significant shifts in bile acids of serum nor stool. In contrast, analysis of the bile acid profiles of the intestinal tract of 15 healthy human volunteers revealed trends that have never been observed in live humans before.

3.2 Biological importance of bile acids in humans

Bile acids are important in digestion and absorptions of lipids in the gastrointestinal tract. Bile acids act as emulsion stabilizers in the beginning of the small intestine[2,3], but play other important biological roles throughout the body. Bile acids act as signaling molecules by binding to bile acid receptor proteins present across most human tissues[4,5].

Bile acids have been implicated in diseases including Alzheimers[6], liver diseases[7], cancers[8], inflammatory bowel disease[9], and metabolic disorders[9]. Bile acids are synthesized from cholesterol in the liver and transformed extensively by microbes in the intestinal tract. The primary bile acids cholic acid and chenodeoxycholic acid are the only bile acids synthesized by the liver[10]. Further modifications of hydroxyl groups located on the steroid backbone of bile acids are performed in the digestive tract during conversion of primary to secondary bile acids. The liver also performs conjugation of both primary and secondary bile acids to form glycine or taurine conjugates. Bile acids are excreted into the duodenum. In the ileum, approximately 95% of the bile acids are actively transported through the intestinal epithelium into the portal vein which flows directly to the liver[10]. These re-absorbed bile acids are then conjugated to amino acids and recycled back to the gallbladder where they are again excreted into the duodenum. The cyclic flow of bile acids through the small intestine, portal vein and liver is referred to as enterohepatic circulation. Other metabolites including drugs[11], toxins[12], and plant metabolites[13] also cycle in enterohepatic circulation. During enterohepatic circulation, gut microbes transform bile acids into a complex mixture of bile acid species.

Bile acids are transformed by gut microbes through reactions including deconjugation, dehydroxylation, and epimerization reactions[14]. More than 60 bile acid species have been shown to exist in human circulation[15] with many more proposed to exist in stool[14]. A selection of the most abundant and biologically relevant bile acids will be considered here. These bile acids fall into four general classes based on their biological progression from primary bile acids (synthesized in the liver), to secondary bile acids (bile acids microbially modified on the cholesterol backbone), and also conjugated (bile acids conjugated to

either glycine or taurine by the liver), or unconjugated (bile acids that have been deconjugated by microbial enzymes)[16]. The bile acids classes to be focused on in this work include primary conjugated bile acids, primary unconjugated bile acids, secondary conjugated bile acids, and secondary unconjugated bile acids (Table 3.1).

Table 3.1. Bile acid names, abbreviations, and their classification as primary or secondary, and conjugated or unconjugated.

Bile acid name	**Abbreviation**	**Class**
Taurocholic acid	TCA	Primary Conjugated
Glycocholic acid	GCA	Primary Conjugated
Taurochenodeoxycholic acid	TCDCA	Primary Conjugated
Glycochenodeoxycholic acid	GCDCA	Primary Conjugated
Cholic acid	CA	Primary Unconjugated
Chenodeoxycholic acid	CDCA	Primary Unconjugated
Tauro-alpha-Muricholic acid	T-a-MCA	Secondary Conjugated
Tauroursodeoxycholic acid	TUDCA	Secondary Conjugated
Glycoursodeoxycholic acid	GUDCA	Secondary Conjugated
Glycohyodeoxycholic acid	GHDCA	Secondary Conjugated
Taurodeoxycholic acid	TDCA	Secondary Conjugated
Glycodeoxycholic acid	GDCA	Secondary Conjugated
Taurolithocholic acid	TLCA	Secondary Conjugated
Glycolithocholic acid	GLCA	Secondary Conjugated
Ursodeoxycholic acid	UDCA	Secondary Unconjugated
Deoxycholic acid	DCA	Secondary Unconjugated
Lithocholic acid	LCA	Secondary Unconjugated
Tyrosocholic acid	TyroCA	Microbially Conjugated
Phenylalanocholic acid	PhenylCA	Microbially Conjugated
Leucholic acid	LeuCA	Microbially Conjugated

During enterohepatic circulation bile acids are excreted in the proximal small intestine, absorbed in the distal small intestine and colon, transferred to the portal vein, and then sequestered by the liver. Some bile acids are not sequestered by the liver and flow into systemic circulation. Either spillover or regulated transport of bile acids leads to bile acids flowing throughout the body. The bile acid composition in circulation is important since bile acids act on receptors located in many organs of the human body[4,5]. Several human proteins bind bile acids including the bile acid membrane receptor (TGR5/GPBAR1), nuclear receptor farnesoid X receptor (FXR), pregnane X receptor (PXR), and vitamin D receptor (VDR)[17]. FXR and TGR5 are the most considered proteins with respect to bile acids as signaling molecules. FXR is important in bile acid homeostasis and is highly expressed in intestinal and liver cells. When FXR binds bile acids, a signaling cascade is stimulated leading to inhibition of CYP7A1, the enzyme responsible for the rate limiting step in bile acid synthesis from cholesterol[18]. FXR is also expressed in other organs throughout the body including heart, blood vessels, kidney, adrenal glands, and brain[4]. TGR5 is a cell membrane G-protein coupled protein receptor expressed in a similarly diverse number of human tissues and when bound to bile acids causes a cAMP dependent signaling cascade[5]. These receptors have varying affinity for different bile acid species. TGR5 is most strongly stimulated by LCA, followed by DCA[5] . FXR is most strongly activated by CDCA, followed by DCA and LCA[17]. In summary, bile acids cause systemic effects through these signaling pathways, which are dependent on the different proportions of bile acids present. These signaling effects are important for consideration of how diet and the gut microbiota influence the human body.

Another important biological role of bile acids is regulating the gut microbiota. Bile acids are toxic at high concentrations and disrupt cell membranes causing cell death. Different bile acids have different levels of toxicity which is associated with the bile acid hydrophobicity[19]. Hydrophobic bile acids are more prone to perturb cell membranes compared to hydrophilic bile acids. Gram positive bacteria are generally less susceptible to bile acid toxicity compared to gram negative bacteria[20]. The composition of the bile acid pool regulates the gut microbiota, but the gut microbiota also regulates the bile acid pool. Germ free mice have a much larger and less complex bile acid pool compared to conventionally raised mice[21]. The ability of the gut microbiota to regulate bile acids and for bile acids to regulate microbes makes up an axis of communication between the gut microbiota and host. This communication is of particular importance because bile acids are important to many aspects of human health[22]. This bile acid axis of communication is complex and difficult to study because bile acids and gut microbes interact in the human intestinal tract which has until now been difficult to investigate due to retrieving samples in a non-invasive manner. This chapter discusses data from samples that were collected from human blood plasma, human stool, and from within the intestinal tract of humans *in vivo*.

3.3 Overview of targeted bile acid analysis of human serum, stool, and GI tract samples

Targeted approaches in metabolomics studies are useful when there is a defined set of chemicals of interest for an experiment. One of the most important benefits of targeted LC-MS/MS analysis is improved sensitivity and selectivity for the targeted metabolites. This improved performance is achieved by considering only specific fragment ions.

Signal/noise ratios improve because background (noise) signals are filtered out in these precursor/fragment ion selections. Selectivity is even improved if more than one fragment ion is monitored, which is limited for some bile acids that show little fragmentation even under high collision energies. Either triple quadrupole or quadrupole-ion trap instruments are used to perform multiple reaction monitoring where two mass analyzers are used as sequential filters before ions reach the mass detector. This type of analysis reduces noise from co-eluting metabolites, and thus lower concentrations of metabolites can be detected. Another benefit of this type of analysis is that accurate concentration values can be achieved for the chemicals of interest. Concentrations are calculated based on analysis of known concentrations of standard chemicals for each target. Another benefit of targeted analysis is the ability to account for ion suppression through use of internal standards. Internal standards are frequently isotopically labeled forms of the targets of interest and are analyzed alongside all samples and standard chemicals used to calculate concentration. The ratio of the peak area of the metabolite of interest and the internal standard is then used to calculate concentration. This process of quantification accounts for ion suppression since both the internal standard and metabolite of interest will experience ion suppression at the same proportion.

Bile acids are therefore commonly analyzed using targeted analysis methods. The benefits of increased sensitivity, and accurate quantification are important in measuring bile because some bile acid species exist at very low concentrations, especially when considering experimental conditions under which bile acid concentrations radically differ between studies. Another reason for using a targeted approach to measure bile acids is that there are many isomers within the chemical class of bile acids. Isomers usually do

not have unique fragment ions to allow MS/MS ion transitions to distinguish between the different bile acids. This leads to the need for separation of these chemically similar isomers by retention time during liquid chromatography. The LC conditions required to separate these chemicals therefore must be highly specific and robust against retention time shifts. Thus, an optimized method specifically for bile acids is usually used for separation and quantification of these metabolites. Since there is not a plethora of other metabolites to measure during this bile acid optimized LC run, a targeted run that focuses only on a select number of ions is necessary; 20 mass transition scans were used in the experimental data of this thesis. By only measuring a select number of masses, less expensive instrumentation is needed in comparison to high-resolution accurate mass non-targeted analysis. Targeted instruments, however, cannot collect comprehensive non-targeted metabolomics data at comparable scan rates compared to mass spectrometers equipped with time of flight or orbital ion trap mass analyzers. Therefore, targeted studies necessarily follow classic hypothesis-driven experimental designs rather than exploratory or hypothesis-generating designs in non-targeted metabolomics. This shows the drawback of targeted analyses, which is that only ions of interest that are input to the analytical method will be measured, and no further information will be collected from any other metabolites. For the samples analyzed in this thesis, only targeted bile acid data are discussed here, in addition to non-classical microbially conjugated bile acids that were discovered in non-targeted analyses. Further analysis of non-targeted data of the 15-human subject study was beyond the scope of this thesis.

An additional challenge in bile acid quantification is the large range of concentrations between different bile acids. Some bile acid species are present at trace levels, while

others are present at very high concentrations (*e.g.* > 5mg bile acid /mL human bile)[23]. The concentration of bile acids in biological samples including blood plasma and stool span up to six orders of magnitude[24]. MS instruments are limited to quantify metabolites across up to five orders of magnitude due to the linear dynamic range of response signal to analyte concentration[25]. To detect all bile acids within the linear dynamic range for biological samples, the sample can be diluted to low concentrations. Dilution of samples should proceed to the point where the highest concentration of bile acids falls within the linear dynamic range. By analyzing samples at only a low concentration there is a disregard for bile acids that may be present below the limit of detection. To increase the range of concentrations able to be accurately quantified, samples can be analyzed at multiple concentrations. Samples can be analyzed at high concentration to accurately quantify low abundant bile acids, and then reanalyzed at low concentration to accurately quantify high abundant bile acids. Analyzing each sample twice requires twice as much time to collect data, adds additional data processing time, and an additional step of determining which analysis (high or low concentration analysis) should be reported as the most accurate concentration. For this thesis, both approaches were tested and used. The approach for the flaxseed dietary study ofserum and stool measured all samples at one specific dilution factor, whereas for the GI tract samples from the 15-volunteer study, bile acid concentrations were measured using up to three dilutions due to the very large differences in bile acid concentration between samples and between GI tract locations.

3.4 Bile acids in serum and stool in response to flaxseed dietary intervention

During the flaxseed dietary intervention study discussed in chapter one, we hypothesized that bile acid profiles change in serum and/or stool samples of subjects who consumed

flaxseed for six weeks. The biological rational for expecting a change in bile acids is that flaxseed contains a high content of dietary fiber[26] that is known to cause changes in the composition and activity of the gut microbiota, and additionally interact with bile acids on the molecular level to reduce bioavailability of the bile acids[27]. Hence, flaxseed consumption was expected to change the microbial composition of the intestinal tract and thereby cause changes in the bile acid pool. Flaxseed was also expected to provide fiber that would increase the excretion of bile acids to cause an increase in stool total bile acid concentration, and a decrease in serum total bile concentration. To test for these hypothesized changes, serum and stool bile acids were quantified from samples of obtained from 178 postmenopausal women taken before and after 6 weeks of flaxseed consumption. The hypothesized change in bile acids was tested using paired Student t-tests comparing the pre- and post-intervention samples for each subject. However, there were not statistically significantly different bile acids found between pre and post intervention samples in either serum or stool (FDR corrected $p < 0.05$). The failure to detect a change in bile acids may be due to a lack of dietary control during this study, or possibly too little dietary fiber (~2.5 grams per day)[28] to have an impact on microbial metabolism or bile acid metabolism. Subjects were not limited in diet throughout the study except for the intervention of consuming 10 g/day of ground flaxseed. Diet plays an important role in shaping both the gut microbiota and the bile acid pool[20]. The different diets consumed by subjects may have introduced enough variation to obscure any effect that the flaxseed may have had on bile acid metabolism. Future studies may be better prepared to elucidate the effect that flaxseed consumption has on the bile acid pool by using more strict dietary regiments during the testing period. Obviously, another possible

approach would be to increase the amount of flaxseed consumed per day to increase the amount of dietary fiber consumed within realistic doses.

The total concentration of bile acids in serum from the 178 women tested showed an average of 4.67 μg/mL, which is much lower than the total concentration of bile acids measured in stool from a subset of the same subjects (254.3 μg/g stool (wet weight)). Stool has approximately 50-fold higher concentration of bile acids compared to serum, and also shows dramatically different proportions of bile acids (Figure 3. A and C). Stool samples have the highest proportion of lithocholic, taurolithocholic, ursodeoxycholic, and deoxycholic acids, while serum samples have a more even distribution across a variety of bile acids. Serum bile acid proportions contained considerable proportions of primary and secondary glycine and taurine conjugated bile acids, a considerable proportion of unconjugated primary bile acids, and deoxycholic acid. In contrast to stool samples from these same subjects, little lithocholic acid was present in the serum. This large difference in the proportion of bile acids present in serum compared to stool shows that stool bile acids do not represent the proportions of bile acids present in circulation.

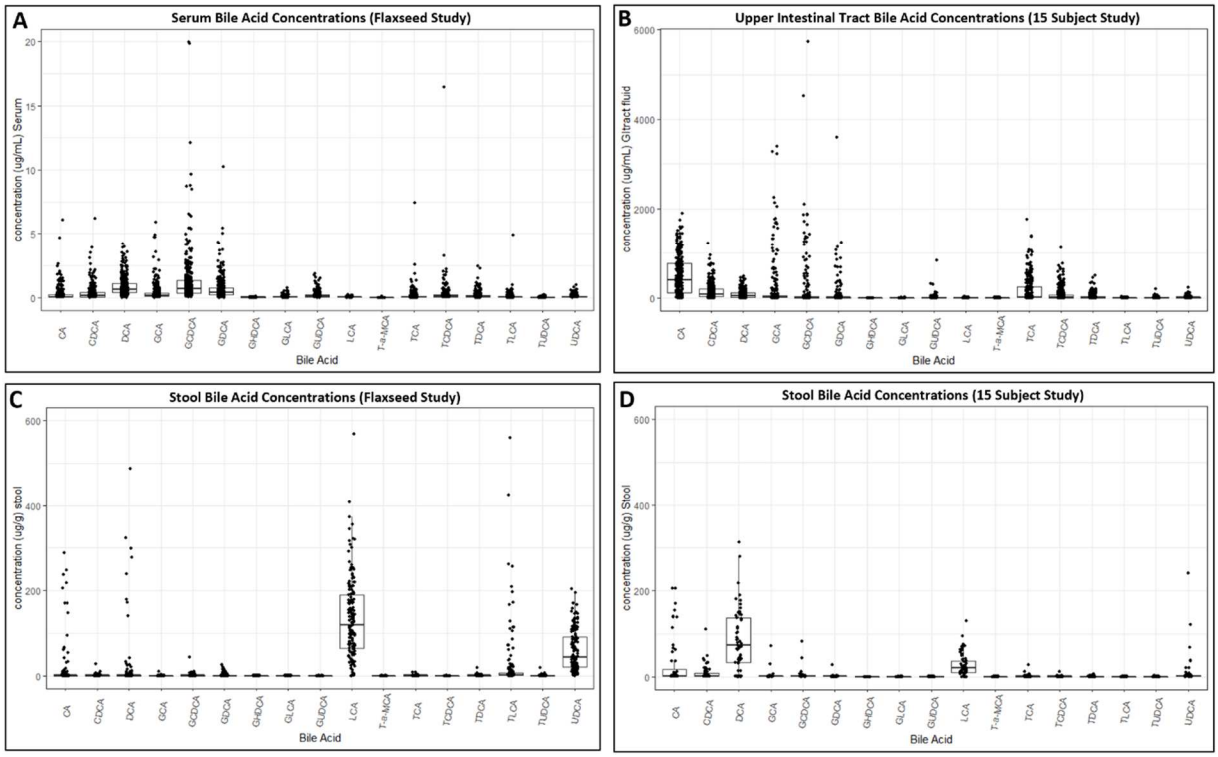


Figure 3.1. Bile acid concentrations in mg/mL for liquid samples and mg/g for stool samples across multiple studies. Plot A shows serum bile acid concentrations from a study of 178 women sampled before and after a flaxseed dietary intervention. Plot B is bile acid concentrations in GI tract luminal fluid from samples collected in the intestinal tract. Plot C is bile acid concentrations from stool samples collected from a subset of the 178 women participating in the flaxseed dietary intervention study from panel A. Plot D is bile acid concentrations from subjects that had samples taken from intestinal tract in panel B.

3.5 Bile acids along the human GI tract

Bile acids are excreted into the duodenum and flow through the small intestine to the ileum where a majority of them are reabsorbed. Non-absorbed bile acids then progress to the colon where they can be passively reabsorbed into intestinal epithelium or excreted in feces. Bile acids in the small intestine of humans have been difficult to study as it is difficult to sample this region. Through use of an ingestible sampling device that can sample this region of the GI tract, samples were retrieved from multiple parts of the small intestine of 15 healthy human subjects. During sample collection, four sampling devices were swallowed simultaneously. As the devices progressed through the digestive tract, the devices dissolved their outer layers in a time dependent manner releasing a hollow elastic bladder that automatically expanded to pull in samples of the lumen of the intestinal tract in a staggered fashion. Four different pH sensitive coatings were applied to the outer layer of the capsule to cause location dependent release of the elastic bladder throughout the intestinal tract. Each of the four capsule types were taken simultaneously leading to four samples being collected between the duodenum and ascending colon for each sampling time point. Trends in the general classes of bile acids were targeted for analysis in attempt to measure the spatial representation of bile acids in the human GI tract and determine whether the trends agree with hypothesized biological activities occurring throughout the human intestinal tract[29].

Four sampling locations of the intestinal tract were determined by the type of capsule ingested (capsule type 1,2,3 or 4). These capsule types were designed to sample across the length of the small intestine. The capsule types sampled between the proximal and distal small intestine with the general regions sampled being the duodenum (capsule type

1), duodenum/ jejunum (capsule type 2), jejunum/ ileum (capsule type 3), and ileum/ ascending colon (capsule type 4). A gradient was observed in the proportion of different bile acid classes across the four sample collection points (Figure 3. A). Conjugated primary bile acids accounted for 59% of bile acids in the duodenum region and decreased to 17% of the bile acids in the ileum region of the small intestine. The inverse trend was observed for unconjugated primary bile acids which increased from 27% of the bile acids in the duodenum (capsule type 1) to 69% of bile acids in the ileum (capsule type 4). These two classes of bile acids expectedly showed inverse trends because bile acids are deconjugated during transit of the small intestine causing conjugated primary bile acids to become unconjugated primary bile acids[30]. This deconjugation reaction is performed by bile salt hydrolase enzymes (EC 3.5.1.24) which are present in approximately 25% of bacterial strains identified in the human gut[31]. Similarly, conjugated secondary bile acids decreased from 9.9% to 3.2% of bile acids between the proximal and distal small intestine, while unconjugated secondary bile acids increased from 3.5% to 10.4% of bile acids between the proximal and distal small intestine (Figure 3.). These trends in secondary bile acids are also due to the activity of bile salt hydrolases enzymes deconjugating secondary bile acids[31].

The proportion of total primary and total secondary bile acids between the proximal and distal small intestine remains remarkably constant (Figure 3. A). To convert primary to secondary bile acids a dehydroxylation reaction is required[16]. Enzymes capable of this reaction are less prevalent across the gut microbiota than bile salt hydrolases and found primarily in gut bacteria *Clostridum scindnes, C. hylemonae,* and the food-borne pathogen *C. perfringens*[32]. Conversion of primary to secondary bile acids occurs primarily

in the lower intestinal tract[32–34]. In agreement with this these findings we observed little conversion of total primary to secondary bile acids during transit of the small intestine, but observed a high proportion of unconjugated secondary bile acids in stool (67.2% of stool bile acids) (Figure 3. A). Secondary bile acids dominated stool bile acids, likely for two reasons. The first reason is that microbes containing enzymes that can dehydroxylate bile acids have more time to act on bile acid species while the bile acids progress through the lower intestinal tract. Secondly, conjugated bile acids and primary bile acids are more strongly re-absorbed in the ileum as compared to unconjugated bile acids and secondary bile acids.

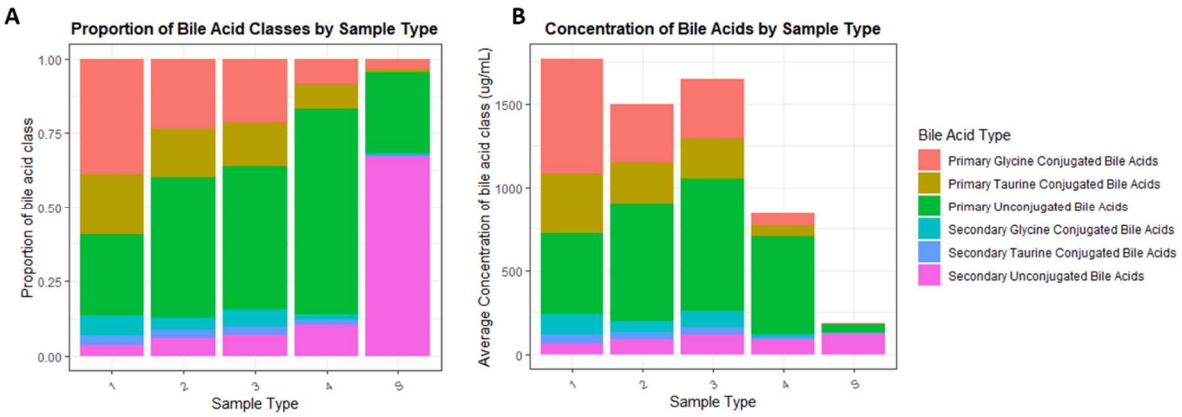


Figure 3.2. Bile acid profiles from samples collected by capsules in the intestinal tract with sample type 1 corresponding to the most proximal small intestine sampling location and sample type four corresponding to the most distal intestinal tract sampling location at approximately the ileum. Sample type S are stool samples. Bile acid classes contain the following bile acids: Primary glycine conjugated(GCA, GCDCA), Primary taurine conjugated(TCA, TCDCA), Primary unconjugated(CA, CDCA), Secondary glycine conjugated(GDCA, GUDCA, GLCA, GHDCA), Secondary taurine conjugated(TDCA, TUDCA, TLCA, THDCA), Secondary unconjugated(DCA, UDCA, LCA, HDCA). Panel A shows the proportion of different bile acid classes for the different sampling types, and panel B shows the same data presented by average concentration values in μ/mL (or μ/g for stool samples). Sample size for each sample type was n=75, n=69, n=69, n=62, n=57 for capsule types 1, 2, 3, 4, and Stool respectively.

The highest average concentration of total bile acids (1769.7 μg/mL with standard deviation of 1869.3 μg/mL) was observed at the first sampling location which was closest to the duodenum where bile acids are excreted into the digestive tract. The second and

third sampling points had an average of about 1500 μg/mL of bile acids whereas the most distal sampling point showed the lowest average concentration of bile acids (839.5 μg/mL with standard deviation of 831 μg/mL). Stool samples had an average total bile acid concentration of 186.7 μg/g with a standard deviation of 161.0 μg/mL. A statistically significant difference in total bile acid concentration between capsule types was found based on Kruskal-Wallis test by rank (p-value < 0.05). Further analysis using two-sample Wilcoxon tests revealed that the first three capsule types were not significantly different than each other (FDR corrected p-value > 0.05) but capsule type 4 and stool samples were statistically significantly different than each other, and different than the first three capsule types (FDR corrected p-value < 0.05). The decrease in concentration of bile acids in the fourth capsule type compared to the first three capsule types was likely due to samples being collected in the ileum during or after active uptake of luminal bile acids through intestinal epithelium. The much lower concentration of bile acids in stool compared to the intestinal samples is explained by a vast majority of bile acids being reabsorbed in the ileum. In additional to ileum uptake, bile acids are also passively absorbed through epithelium of the colon to further reduce the total amount of bile acids in feces. This thesis presents for the first time measurements of bile acids from samples acquired *in-vivo* throughout the human intestinal tract using a non-invasive approach.

3.6 Stool bile acids do not represent circulating bile acids

Bile acids in the small intestine are at a higher concentration than feces (Figure 3. B). Furthermore, bile acids in the small intestine have a different proportional profile compared to stool. To investigate whether stool or the intestinal tract was more similar to

bile acids circulating in blood, serum measurements collected during the study of flaxseed consumption was compared to samples collected from the intestinal tract and feces. Interestingly, the proportions of bile acids in circulation are much more similar to the proportion of bile acids in the small intestine compared to feces (Figure 3. A and B are more similar than A and D). This fact can be explained in the following way: bile acids are present at a much higher concentration in the small intestine compared to feces, and additionally, most of the bile acids in the small intestine are reabsorbed into the portal vein and flow to the liver where a portion of them continue into systemic circulation. The dissimilarity of bile acid pools in stool compared to the bloodstream raises the question of whether the stool bile acid pool can be expected to reveal biologically significant differences in bile acid metabolism in biomedical studies. The bile acids in circulation are likely responsible for many of the biological effects associated with bile acids. Despite this fact, stool bile acids are still commonly measured to investigate the effect of treatments on bile acid metabolism[35–40]. Stool is easily sampled and contains the spent remains after digestion, including bile acids, which makes it a popular specimen for investigation of the gastrointestinal tract. Samples collected during this study of 15 human subjects suggests that differences in bile acid metabolism in the upper intestinal tract is not well represented by the profile of bile acids observed in stool samples.

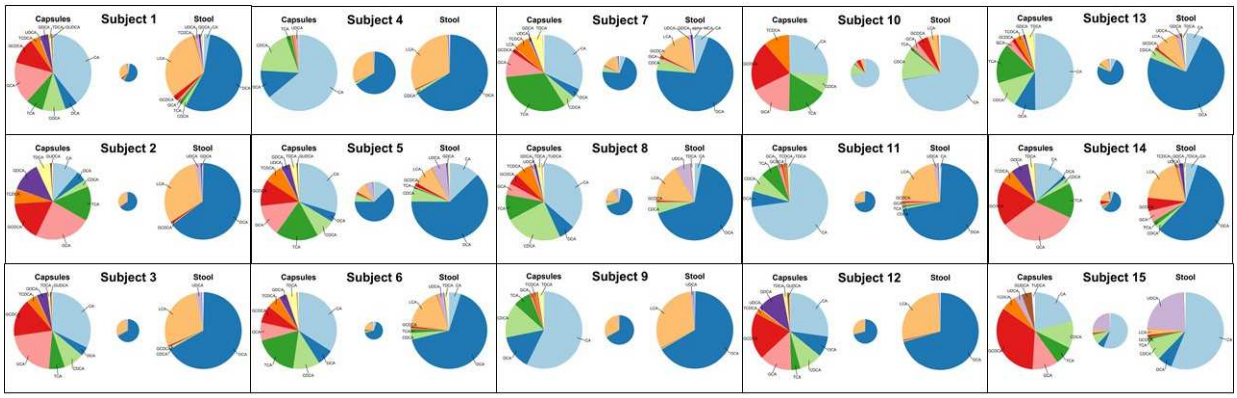


Figure 3.3. Pie charts separated by subject and also separated by capsule samples and stool samples. The small circle between capsule and stool pie charts are the identical stool pie chart except scaled by area to the total average concentration of capsule samples. The most abundant 12 bile acids across this study were plotted to reduce labeling complexity. The color scheme is as follows, light blue:CA, dark blue:DCA, light green:CDCA, dark green:TCA, pink:GCA, red:GCDCA, peach:LCA, orange:TCDCA, light purple:UDCA, purple:GDCA, yellow:TDCA, light brown:GUDCA. White space left in the charts represents bile acids omitted from these plots.

Bile acids were measured in samples collected from the small intestine of humans alongside stool samples collected from the same subjects. Subjects showed variable bile acid profiles in the small intestine samples (Figure 3.3 left pie charts for each subject). For example GCDCA (Figure 3.3, red colored portions) was nearly absent in small intestine samples of some subjects (subjects 4, 9, and 11), and was more than 20% of all small intestine bile acids in other subjects (subjects 10, 12, and 15). Other bile acids showed similar variation across subjects, yet this variation largely disappeared in stool

samples, which converged to mostly DCA and LCA (Figure 3.3 right pie charts for each subject) except for outlier subjects 10 and 15. Interestingly, subjects 10 and 15 are the only two subjects to consume antibiotics within 4 months of sample collection, which may explain the large difference in the bile acid profiles of these subjects compared to other subjects. Together these observations show that the small intestine and stool have different proportions and concentrations of bile acids. Furthermore, the bile acid profiles in the small intestine have greater variation than the profile of bile acids in stool. The source of this variation in bile acids is likely due to the time of day the sample was acquired, age, sex, or other lifestyle variables. Future studies with larger sample size, controlled diet, and lifestyle meta-data will provide the resolution necessary to find how other variables impact bile acids of the intestinal tract.

3.7 Phenylalacholic, tyrosocholic, and leucocholic acid across the GI tract

Recently three bile acids conjugated to amino acids other than glycine and taurine were discovered in the gastrointestinal tract of mice and humans[41]. The formation of glycine and taurine conjugated bile acids is performed by known enzymes in the liver; however, formation of these newly discovered bile acid conjugates have been reported to be synthesized by microbes in the intestinal tract[41]. These are expected to be biologically important bile acids as they were found to differ significantly between diseased and healthy states in multiple studies that were re-analyzed to consider the identification of these new bile acid conjugates[41]. Additionally these bile acids were shown to be agonists of human FXR[41].

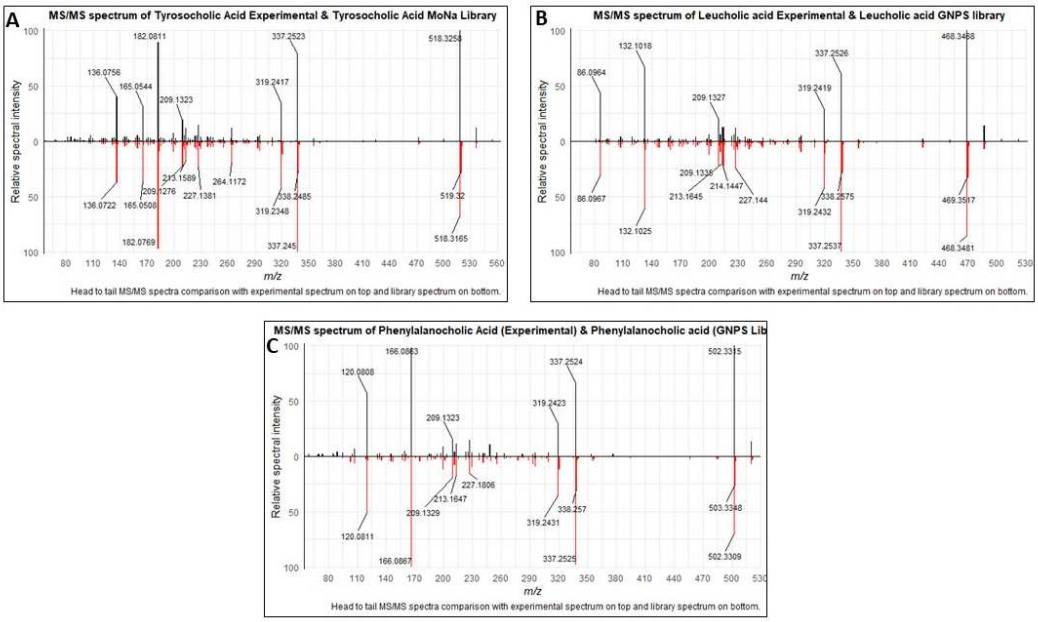


Figure 3.4. MS/MS spectral matches between experimental spectra (top spectrum in all panels) compared to library spectra (bottom spectrum in all panels). Panel A shows match from experimental to tyrosocholic acid, Panel B shows spectral match to leucholic acid, and panel C shows spectral match to phenylalanocholic acid.

These newly discovered bile acids were not observed in the targeted bile acid measurements because they were previously not known to exist when my studies began. Targeted methods only measure the specific *m/z* transitions that are monitored across the chromatogram, and the *m/z* of these novel bile acid species were not included in the list of bile acids monitored. Fortunately, bile acids are also detected during non-targeted HILIC LC-MS/MS metabolomics analysis. The MS/MS fragmentation spectra of these novel bile acid conjugates were matched to MS/MS spectra collected during non-targeted metabolomics analysis of the small intestine and stool samples of this experiment (Figure

3.4). These three bile acids were detected in all subjects but showed a subject-dependent trend (Figure 3.5). Subjects 3, 8, and 13 had the highest concentration of these microbially conjugated bile acids, which could be due to differences in the abundance or activity of specific microbes in the intestinal tract. The genes that create these conjugates have not yet been elucidated. Further investigation into the production of these bile acids may reveal their possible role as a component in the axis of communication between the gut microbiota and human host.

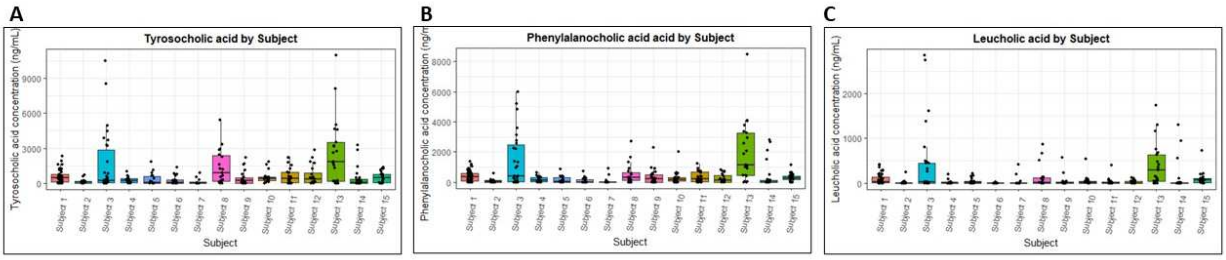


Figure 3.5. Newly discovered microbially conjugated bile acid concentrations (approximate ng/mL) measured in samples collected from the intestinal tract of different subjects. Panel A shows tyrosocholic acid concentrations across subjects, panel B shows pheylalanocholic acid concentration across subjects, and panel C shows concentration of leucholic acid across subjects. Boxplots of concentration values separated by subject (n=17). Boxes represent distance between quartiles 1 and 3 and error bars extend up to 1.5 inter quartile ranges (IQR) from the upper (third quartile) or lower (first quartile) portion of boxes. Samples greater than 1.5 IQR from the first or third quartile are plotted as individual points outside of error bars. Concentration was calculated using GCA as a surrogate standard in non-targeted HILIC analysis and plotted against the known concentration of GCA from targeted quantification; the resulting equation was applied to quantify these three bile acids.

3.8 Methods

3.8.1 Targeted bile acid analysis of serum and stool from flaxseed dietary intervention

Samples of serum and stool were collected before and after a 6-week flaxseed dietary intervention as previously reported in section 1.8.4. Serum samples were prepared as previously described[42]. In summary 50 µL of serum was transferred to 96 well polypropylene plates. To each well, 25 µL of deuterium labeled internal standards in acetonitrile with internal standards at 250 ng/mL was added. Addition of 25 µL of 1000 nM internal standards (CUDA and PHAU) in methanol to each well was done as well. Then 25 µL of antioxidant solution and 125 µL of 1:1 acetonitrile/methanol were added to bring the total volume of each well to 250 µL. Blank samples were made by using 50 µL of LC-MS grade water instead of serum. The plate was sealed, vortexed for 30 seconds, centrifuged for 5 minutes at 15,000 *rcf* at 6 °C, and supernatants were transferred to a clean 96 well plate, sealed, and stored at -20 °C until analysis. The extraction was carried out on ice except for during vortex and centrifugation. Stool samples were extracted by weighing 2 mg (+/- 0.2 mg) into 2 mL microcentrifuge tubes, and adding 500 µL of ice-cold methanol, 25 µL of deuterium labeled bile acid standards at 500 nM, 25 µL of CUDA/PHAU solution at 1000 nM, and two 3mm stainless steel grinding beads. Tubes were homogenized in a genogrinder shaker for 1 minute at 1500 rpm. Tubes were then subjected to centrifugation at 14,000 *rcf*, and the supernatant was transferred to a clean microcentrifuge tube and dried in rotary vacuum dryer until dry. Samples were then stored at -20 °C until resuspension. At the time of resuspension, 250 µL of methanol was added to each tube, vortexed for 20 seconds, sonicated for 5 minutes at room temperature, and

then centrifuged for 2 minutes at 14000 *rcf*. 30 µL of sample from each tube was then transferred to a well of a 96 well plate and stored in the autosampler at 4 °C until analysis.

LC-MS/MS was carried out using an AB SCIEX 6500+ QTRAP mass spectrometer coupled to a ThermoFisher Vanquish LC system. Multiple reaction monitoring scans were collected and processed as previously reported[42]. Briefly, a Waters Aquity BEH C18 column (1.7µm, 2.1mm x 100mm) with guard column Waters Acquity BEH C18 (1.7µm, 2.1mm x 5mm) was used for separation with mobile phases of A: LC-MS grade water with 0.1% formic acid, and B: LC-MS grade acetonitrile with 0.1% formic acid with a flow rate of 400 µL/min. The mobile phase gradient was identical to that previously reported[42]. Concentration curves were run during sample analysis (3 standard curve points between every 20 samples) with 9 concentration points ranging from 0.2 ng/mL to 1500 ng/mL. Multiquant version 3.0.2 was used to perform smoothing, peak integration, and concentration calculations as previously reported[42]. The ratio of bile acid analyte to deuterium internal standards was used to calculate absolute concentration of bile acids in samples.

3.8.2 Targeted bile acid analysis of intestinal tract samples

Samples of GI tract lumen supernatant (n=275), and corresponding stool samples (n=57) were collected using ingestible sampling capsules. Four different enteric coatings were applied as the exterior of capsules that were 23 mm long and 6.5 mm wide to target specific regions of the intestinal tract. Once the coating degraded, an elastic bladder expanded and pulled in approximately 300 µL of luminal liquid through a one-way valve. The coating of capsule type 1 was designed to quickly dissolve at pH 6 to sample

immediately after passing the pyloric sphincter where pH shifts from approximately pH 2 in the stomach to 6 in the duodenum[43–46]. Capsule type 2 was designed to open at the same pH; however, had a more robust coating to delay sampling in time compared to capsule 1. The coating of capsule type 3 was designed to open at pH 7.5, to target the middle of the small intestine, as the pH rises from approximately pH 6 in the duodenum to 8 in the ileum[43–46]. Capsule type four had the same pH sensitive coating as capsule type 3 but had a more robust coating to delay capsule opening in time compared to capsule type 3.

This study was approved by WIRB-Copernicus Group IRB and informed consent was acquired. Subjects were healthy with no known prior GI tract complications, between the ages of 18 and 70, not pregnant, fluent in English, and highly willing to comply with study requirements. Fifteen subjects each swallowed a total of 17 sampling capsules over the course of three days. During this period all subject stool samples were immediately frozen after collection, and capsules were retrieved by staff after thawing. Upon retrieval, elastic bladders were rinsed with isopropyl alcohol, and liquid sample retrieved using a syringe from the elastic bladder. Liquid sample was transferred to a microcentrifuge tube and centrifuged at 10,000 *rcf* for 3 minutes. The supernatant was transferred to a new microcentrifuge tube and immediately frozen at -80 °C. Samples with mostly gas in the elastic bladders were excluded from analysis. One capsule was consumed on the first day of the three-day testing period. On the second and third day, four capsules were consumed after two evening meals. Diet, time of meals and time of bowel movements were all recorded throughout the three-day testing period.

Samples of intestinal liquid were extracted using a modified 96 well plate biphasic extraction[47]. Samples in microcentrifuge tubes were thawed on ice and 10 µL of intestinal lumen supernatant was added to wells of a 2 mL polypropylene 96 well plate in a pre-determined randomized order. A quality control (QC) sample was prepared by mixing intestinal tract samples acquired from previous pilot studies, and was used in this study to assess analytical variation. QC sample matrix (10 µL) and blanks (10 µL of LC-MS grade water) were included in every row of the 96 well plate corresponding to every 10th sample. A volume of 170 µL of Methanol containing internal standards (UltimateSPLASH Avanti Polar Lipids, Alabaster, Alabama) was added to each well, followed by 490 µL of methyl-tert-butyl-ether (MTBE) containing internal standard CE 22:1. Plates were then sealed, vortexed vigorously for 30 seconds, and shaken on orbital shaking plate for 5 minutes at 4 °C. The plate was unsealed and 150 µL of cold water was added to each well. Plates were then re-sealed and vortexed vigorously for 30 seconds, and centrifuged for 12 minutes at 4,000 *rcf* and 4 °C. From the top phase of the extraction, two aliquots of 180 µL were transferred to clean 96 well plates, and two aliquots of 70 µL from the bottom phase were transferred to two other clean 96 well plates for a total of four 96 well plates. Plates were dried in rotary vacuum until dry, sealed, and stored at -20 °C until LC-MS/MS analysis which occurred within one week. One of the 96 well plates containing the aqueous phase of extract was dissolved in 35 µL of HILIC run solvent as described in chapter 2. A volume of 5 µL was analyzed using non-targeted HILIC analysis. Immediately after HILIC analysis, the 96 well plates were dried in rotary vacuum until dry, sealed, and stored at -20 °C until targeted bile acid analysis.

Multiple dilutions were prepared for bile acid analysis as follows. The dried sample was dissolved in 60 µL of bile acid run solvent (1:1 acetonitrile/ methanol containing 5 deuterated internal standards) through 30 seconds of vortex and were shaken on an orbital shaker for 5 minutes. From this plate, 5 µL was transferred to a new 96 well plate and combined with 145 µL of bile acid run solvent. Both dilutions were analyzed for all samples, and samples that still presented bile acids above the highest concentration of the standard curve (1500 µg/mL) were diluted once more at the ratio of 5:145 and re-analyzed. A 9-point standard curve was used for bile acids and ranged from 0.2 ng/mL to 1500 ng/mL. The standard curve solutions were created by drying the appropriate amount of bile acid mixture to achieve desired concentrations and dissolving in bile acid run solvent prior to analysis. Three standard curve points were analyzed between every 20 samples. A blank sample was also analyzed between every 20 samples.

Stool samples were prepared by weighing 4 mg (+/- 1 mg wet stool) into 2 mL microcentrifuge tubes. Quality control samples were made by using 20 µL of the QC mix described above. Blanks began with 20 µL of LC-MS grade water. To each tube 225 µL of ice cold methanol containing internal standards (as above) was added to each tube, followed by 750 µL ice cold MTBE with CE 22:1. Two 3mm stainless steel grinding beads were added to each tube and tubes were processed in a genogrinder at 1500 rpm for 1 minute. A volume of 188 µL of cold water was then added to each tube. Tubes were vortexed vigorously and centrifuged at 14000 *rcf* for 2 minutes. Two aliquots of 180 µL of the MTBE layer and two aliquots of 50 µL of the lower layer were transferred to four separate 96 well plates, dried in rotary vacuum, sealed, and stored at -20 °C until analysis alongside intestinal samples.

Samples were analyzed using a Thermofisher Vanquish UHPLC system coupled to a Thermofisher TSQ Altis triple quadrupole mass spectrometer. An Aquity BEH C18 column (1.7µm, 2.1mm x 100mm) with guard column Acquity BEH C18 (1.7µm, 2.1mm x 5mm) was used for chromatographic separation with mobile phases of A: LC-MS grade water with 0.1% formic acid, and B: LC-MS grade acetonitrile with 0.1% formic acid with a flow rate of 400 µL/minute and column temperature of 50 °C. The gradient began at 20% B for 1 minute, then shifted to 45% B between 1 and 11 minutes, then to 95%B between 11 and 14 minutes, then to 99% B between 14 and 14.5 minutes, 99% B was maintained until 15.5 minutes, then transitioned form 99% B to 20% B between 15.5 and 16.5 minutes and maintained at 20% B until 18 minutes. Injection volume was 5 µL and MRM scans were collected as reported (Table 3.2).

Table 3.2. MRM scan list. Reactions monitored through targeted bile acid analysis. Multiple bile acids have the same transitions with different retention times and included in the same row. Bile acid names with “d” followed by a number indicate the number of deuterium atoms attached to the isotopically labeled internal standards.

Compound(s)	Precursor (m/z)	Product (m/z)	Collision Energy (V)	Min Dwell Time (ms)	RF Lens (V)	Source Fragmentation
PHAU (internal standard)	248.95	130.1	14.18	197.628	55	0
TDHCA	508.3	79.94	53.44	72.825	232	100
T-ω,α,β-MCA & TCA	514.475	80.042	55	72.825	152	0
TUDCA & TCDCA & TDCA	498.475	80.042	55	57.72	138	0
ω,α,β MCA & CA	407.25	407.25	0	57.662	95	77.6
GUDCA & GHDCA	448.288	74.054	33.18	57.72	93	69.4
GCA	464.425	74.054	35.92	57.72	143	0
GCA-d4 (internal standard)	468.288	74.143	36.17	57.72	121	100
TCDCA-d4 (internal standard)	502.8	80.042	55	57.72	138	0
CA-d4 (internal standard)	411.388	411.388	0	57.662	141	100
UDCA & CDCA & DCA	391.338	391.338	0	57.662	140	100
GCDCA-d4 (internal standard)	452.338	73.988	38.11	57.662	98	71.4
GCDCA & GDCA	448.25	74.089	34.4	57.662	90	65.3
TLCA	482.338	80.03	52.05	57.662	249	100
CUDA (internal standard)	339.15	214.271	21.68	57.662	83	0
CDCA-d4 & DCA-d4 (internal standards)	395.338	395.338	0	57.662	126	89.8
DCA	391.2	391.2	0	57.662	95	63.3
GLCA	432.4	74.071	32.63	57.662	112	2
LCA	375.162	375.162	0	83.426	119	100
LCA-d5 (internal standard)	380.362	380.362	0	83.426	119	100

All MRM scans were imported to Skyline[48] open source software. Skyline performed peak integration for all peaks detected. Precursor and product ion masses were provided to Skyline, in addition to the retention time of each metabolite, and a RT window of less than 0.15 minutes. The chromatogram for each bile acid and internal standard (31 total

chemicals monitored) was individually checked for correct peak integration. Peak area was exported for all monitored chemicals for all dilution values. Bile acid data was removed if there was not a true peak observed in at least one sample (for example ω-MCA, a mouse specific bile acid, was not observed in any of these samples). The ratio of analyte to internal standards within samples was calculated for all bile acids. A linear model was fit to standard curve points for each bile acid (R^2 >0.995 for all bile acids) and the model was applied to all samples and blanks to calculate concentration. The average value of concentration in blank measurements was subtracted from sample concentrations. Because multiple dilutions were analyzed for each sample, the measurement from sample dilution with the concentration value closest to the center of the standard curve (750 ng/mL) was kept. Zero values were imputed with a concentration value between 0.001 and 0.1 ng/mL. Dilution factors were applied to each measured concentration value dependent upon which dilution was kept in order to determine the concentration of bile acids in original sample. Statistical analysis and data processing steps used the R programming language. FDR correction of Benjamini and Hochberg[49] was used.

3.9 References

1 M. Makishima, A. Y. Okamoto, J. J. Repa, H. Tu, R. M. Learned, A. Luk, M. V. Hull, K. D. Lustig, D. J. Mangelsdorf and B. Shan, Identification of a nuclear receptor for bite acids, *Science (80-.).*, 1999, **284**, 1362–1365.

2 A. J. Fillery-Travis, L. H. Foster and M. M. Robins, Stability of emulsions stabilised by two physiological surfactants: l-α-phosphatidylcholine and sodium taurocholate,

Biophys. Chem., 1995, **54**, 253–260.

3 M. Armand, P. Borel, B. Pasquier, C. Dubois, M. Senft, M. Andre, J. Peyrot, J. Salducci and D. Lairon, Physicochemical characteristics of emulsions during fat digestion in human stomach and duodenum, *Am. J. Physiol. - Gastrointest. Liver Physiol.*, , DOI:10.1152/ajpgi.1996.271.1.g172.

4 C. Huang, J. Wang, W. Hu, C. Wang, X. Lu, L. Tong, F. Wu and W. Zhang, Identification of functional farnesoid X receptors in brain neurons, *FEBS Lett.*, 2016, **590**, 3233–3242.

5 V. Stepanov, K. Stankov and M. Mikov, The bile acid membrane receptor TGR5: A novel pharmacological target in metabolic, inflammatory and neoplastic disorders, *J. Recept. Signal Transduct.*, 2013, **33**, 213–223.

6 S. MahmoudianDehkordi, M. Arnold, K. Nho, S. Ahmad, W. Jia, G. Xie, G. Louie, A. Kueider-Paisley, M. A. Moseley, J. W. Thompson, L. St John Williams, J. D. Tenenbaum, C. Blach, R. Baillie, X. Han, S. Bhattacharyya, J. B. Toledo, S. Schafferer, S. Klein, T. Koal, S. L. Risacher, M. A. Kling, A. Motsinger-Reif, D. M. Rotroff, J. Jack, T. Hankemeier, D. A. Bennett, P. L. De Jager, J. Q. Trojanowski, L. M. Shaw, M. W. Weiner, P. M. Doraiswamy, C. M. van Duijn, A. J. Saykin, G. Kastenmüller and R. Kaddurah-Daouk, Altered bile acid profile associates with cognitive impairment in Alzheimer's disease—An emerging role for gut microbiome, *Alzheimer's Dement.*, 2019, **15**, 76–92.

7 A. Asgharpour, D. Kumar and A. Sanyal, Bile acids: emerging role in management of liver diseases, *Hepatol. Int.*, 2015, **9**, 527–533.

8 M. Baptissart, A. Vega, S. Maqdasy, F. Caira, S. Baron, J. M. A. Lobaccaro and D. H. Volle, Bile acids: From digestion to cancers, *Biochimie*, 2013, **95**, 504–517.

9 A. Baars, A. Oosting, J. Knol, J. Garssen and J. van Bergenhenegouwen, The Gut Microbiota as a Therapeutic Target in IBD and Metabolic Disease: A Role for the Bile Acid Receptors FXR and TGR5, *Microorganisms*, 2015, **3**, 641–666.

10 J. Y. L. Chiang, Bile acid metabolism and signaling, *Compr. Physiol.*, 2013, **3**, 1191–1212.

11 M. R. McGill and H. Jaeschke, Metabolism and disposition of acetaminophen: Recent advances in relation to hepatotoxicity and diagnosis, *Pharm. Res.*, 2013, **30**, 2174–2187.

12 C. Thiel, K. Thiel, W. Klingert, A. Diewold, K. Scheuermann, E. Hawerkamp, J. Lauber, J. Scheppach, M. H. Morgalla, A. Königsrainer and M. Schenk, The enterohepatic circulation of amanitin : Kinetics and therapeutical implications, *Toxicol. Lett.*, 2011, **203**, 142–146.

13 M. Zeng, R. Sun, S. Basu, Y. Ma, S. Ge, T. Yin, S. Gao, J. Zhang and M. Hu, Disposition of flavonoids via recycling: Direct biliary excretion of enterically or extrahepatically derived flavonoid glucuronides, *Mol. Nutr. Food Res.*, 2016, **60**, 1006–1019.

14 D. V. Guzior and R. A. Quinn, Review: microbial transformations of human bile acids, *Microbiome*, 2021, **9**, 1–13.

15 M. H. Sarafian, M. R. Lewis, A. Pechlivanis, S. Ralphs, M. J. W. McPhail, V. C.

Patel, M. E. Dumas, E. Holmes and J. K. Nicholson, Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry, *Anal. Chem.*, 2015, **87**, 9662–9670.

16 J. M. Ridlon, D.-J. Kang and P. B. Hylemon, Bile salt biotransformations by human intestinal bacteria, *J. Lipid Res.*, 2006, **47**, 241–259.

17 C. Mazuy, A. Helleboid, B. Staels and P. Lefebvre, Nuclear bile acid signaling through the farnesoid X receptor, *Cell. Mol. Life Sci.*, 2015, **72**, 1631–1650.

18 Y. Wang, S. Gunewardena, F. Li, D. J. Matye, C. Chen, X. Chao, T. Jung, Y. Zhang, M. Czerwiński, H. M. Ni, W. X. Ding and T. Li, An FGF15/19-TFEB regulatory loop controls hepatic cholesterol and bile acid homeostasis, *Nat. Commun.*, , DOI:10.1038/s41467-020-17363-6.

19 M. Watanabe, S. Fukiya and A. Yokota, Comprehensive evaluation of the bactericidal activities of free bile acids in the large intestine of humans and rodents, *J. Lipid Res.*, 2017, **58**, 1143–1152.

20 J. M. Ridlon, D. J. Kang, P. B. Hylemon and J. S. Bajaj, Bile acids and the gut microbiome, *Curr. Opin. Gastroenterol.*, 2014, **30**, 332–338.

21 S. I. Sayin, A. Wahlström, J. Felin, S. Jäntti, H. U. Marschall, K. Bamberg, B. Angelin, T. Hyötyläinen, M. Orešič and F. Bäckhed, Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist, *Cell Metab.*, 2013, **17**, 225–235.

22 A. Molinaro, A. Wahlström and H. U. Marschall, Role of Bile Acids in Metabolic

Control, *Trends Endocrinol. Metab.*, 2018, **29**, 31–41.

23 J. L. Boyer, Bile formation and secretion, *Compr. Physiol.*, 2013, **3**, 1035–1078.

24 M. Dutta, J. Cai, W. Gui and A. D. Patterson, A review of analytical platforms for accurate bile acid measurement, *Anal. Bioanal. Chem.*, 2019, **411**, 4541–4549.

25 H. Liu, L. Lam and P. K. Dasgupta, Expanding the linear dynamic range for multiple reaction monitoring in quantitative liquid chromatography-tandem mass spectrometry utilizing natural isotopologue transitions, *Talanta*, 2011, **87**, 307–310.

26 M. Kristensen, M. G. Jensen, J. Aarestrup, K. E. N. Petersen, L. Søndergaard, M. S. Mikkelsen and A. Astrup, Flaxseed dietary fibers lower cholesterol and increase fecal fat excretion, but magnitude of effect depend on food type, *Nutr. Metab.*, 2012, **9**, 1–8.

27 J. Singh, R. Metrani, S. R. Shivanagoudra, G. K. Jayaprakasha and B. S. Patil, Review on Bile Acids: Effects of the Gut Microbiome, Interactions with Dietary Fiber, and Alterations in the Bioaccessibility of Bioactive Compounds, *J. Agric. Food Chem.*, 2019, **67**, 9124–9138.

28 P. Kajla, A. Sharma and D. R. Sood, Flaxseed — a potential functional food source, 2015, **52**, 1857–1871.

29 J. M. Ridlon, S. C. Harris, S. Bhowmik, D. J. Kang and P. B. Hylemon, Consequences of bile salt biotransformations by intestinal bacteria, *Gut Microbes*, 2016, **7**, 22–39.

30 T. C. Northfield and I. Mccoll, Postprandial concentrations of free and conjugated

bile acids down the length of the normal human small intestine, *Gut*, 1973, **1970**, 513–518.

31 Z. Song, Y. Cai, X. Lao, X. Wang, X. Lin and Y. Cui, Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome, *Microbiome*, 2019, **7**, 1–16.

32 H. Doden, L. A. Sallam, S. Devendran, L. Ly, G. Doden, S. L. Daniel and J. M. Ridlon, Metabolism of Oxo-Bile Acids and Characterization of Recombinant 12-alpha-Hydroxysteroid Dehydrogenases from Bile Acid 7-alpha-Dehydroxylating Human Gut Bacteria, *Appl. Environ. Microbiol.*, 2018, **84**, e00235-18.

33 V. Aries, J. S. Crowther, B. S. Drasar and M. J. Hill, Degradation of bile salts by human intestinal bacteria, *Gut*, 1969, **10**, 575–576.

34 F. BERR, G. KULLAK–UBLICK, G. PAUMGARTNER, W. MUNZING and H. B. PHILLIP, 7a-Dehydroxylating Bacteria Enhance Deoxycholic Acid Input and Cholesterol Saturation of Bile in Patients With Gallstones, *Gastroenterology111*, 1996, **111**, 1611–1620.

35 G. Kakiyama, W. M. Pandak, P. M. Gillevet, P. B. Hylemon, D. M. Heuman, K. Daita, H. Takei, A. Muto, H. Nittono, J. M. Ridlon, M. B. White, N. A. Noble, P. Monteith, M. Fuchs, L. R. Thacker, M. Sikaroodi and J. S. Bajaj, Modulation of the fecal bile acid profile by gut microbiota in cirrhosis q, *J. Hepatol.*, 2013, **58**, 949–955.

36 B. P. Vaughn, T. Kaiser, C. Staley, M. J. Hamilton, J. Reich, C. Graiziger, S. Singroy, A. J. Kabage, M. J. Sadowsky and A. Khoruts, A pilot study of fecal bile

acid and microbiota profiles in inflammatory bowel disease and primary sclerosing cholangitis, 9–19.

37 M. Camilleri, P. Carlson, A. Acosta and I. Busciglio, Colonic mucosal gene expression and genotype in irritable bowel syndrome patients with normal or elevated fecal bile acid excretion, 2021, 10–20.

38 P. Vijayvargiya, D. Gonzalez, I. Gerardo, C. Sarah, T. Sarah, H. Saifuddin, P. Duggan, V. Melo, T. Thomas, M. Heeney, A. Beyde, J. Miller, J. Kenneth, V. Kafayat, O. Joseph, F. B. Jessica, A. Leslie and M. Camilleri, Increased Fecal Bile Acid Excretion in a Significant Subset of Patients with Other Inflammatory Diarrheal Diseases, *Dig. Dis. Sci.*, , DOI:10.1007/s10620-021-06993-5.

39 A. R. Weingarden, C. Chen, A. Bobr, D. Yao, Y. Lu, V. M. Nelson, M. J. Sadowsky, A. Khoruts, W. Ar, C. Chen, A. Bobr, D. Yao, Y. Lu and N. Vm, Microbiota transplantation restores normal fecal bile acid composition in recurrent Clostridium difficile infection, 2021, 310–319.

40 M. Arrieta, L. T. Stiemsma, P. A. Dimitriu, L. Thorson, S. Russell, S. Yurist-doutsch, B. Kuzeljevic, M. J. Gold, H. M. Britton, D. L. Lefebvre, P. Subbarao, P. Mandhane, A. Becker, K. M. Mcnagny, M. R. Sears, T. Kollmann, S. Investigators, W. W. Mohn, S. E. Turvey and B. B. Finlay, Early infancy microbial and metabolic alterations affect risk of childhood asthma.

41 R. A. Quinn, A. V. Melnik, A. Vrbanac, T. Fu, K. A. Patras, M. Christy, Z. Bodai, P. Belda-Ferre, A. Tripathi, L. K. Chung, M. Downes, R. D. Welch, M. Quinn, G. Humphrey, M. Panitchpakdi, K. Weldon, A. Aksenov, R. da Silva, J. Avila-Pacheco,

C. Clish, S. Bae, H. Mallick, E. A. Franzosa, J. Lloyd-Price, R. Bussell, T. Thron, A. T. Nelson, M. Wang, E. Leszczynski, F. Vargas, J. M. Gauglitz, M. J. Meehan, E. Gentry, T. D. Arthur, A. C. Komor, O. Poulsen, B. S. Boland, J. T. Chang, W. J. Sandborn, M. Lim, N. Garg, J. C. Lumeng, R. J. Xavier, B. I. Kazmierczak, R. Jain, M. Egan, K. E. Rhee, D. Ferguson, M. Raffatellu, H. Vlamakis, G. G. Haddad, D. Siegel, C. Huttenhower, S. K. Mazmanian, R. M. Evans, V. Nizet, R. Knight and P. C. Dorrestein, Global Chemical Impact of the Microbiome Includes Novel Bile Acid Conjugations Summary Paragraph HHS Public Access, *Nature*, 2020, **579**, 123–129.

42 D. K. Barupal, Y. Zhang, T. Shen, S. Fan, B. S. Roberts, P. Fitzgerald, B. Wancewicz, L. Valdiviez, G. Wohlgemuth, G. Byram, Y. Y. Choy, B. Haffner, M. R. Showalter, A. Vaniya, C. S. Bloszies, J. S. Folz, T. Kind, A. M. Flenniken, C. McKerlie, L. M. J. Nutter, K. C. Lloyd and O. Fiehn, A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international mouse phenotyping consortium, *Metabolites*, 2019, **9**, 9050101.

43 D. F. Evans, G. Pye, R. Bramley, A. G. Clark, T. J. Dyson and J. D. Hardcastle, Measurement of gastrointestinal pH profiles in normal ambulant human subjects, *Gut*, 1988, **29**, 1035–1041.

44 M. Koziolek, M. Grimm, D. Becker, V. Iordanov, H. Zou, J. Shimizu, C. Wanke, G. Garbacz and W. Weitschies, Investigation of pH and Temperature Profiles in the GI Tract of Fasted Human Subjects Using the Intellicap R System, *J. Pharm. Sci.*, 2015, **104**, 2855–2863.

45 J. B. Dressman, R. R. Berardi, L. C. Dermentzoglou, T. L. Russle, S. P. Schmaltz, J. L. Barnett and K. M. Jarvenpaa, Upper Gastrointestinal (GI) pH in Young, Healthy Men and Women, *Pharm. Rsearch*, 1990, **7**, 756–761.

46 F. Xu, G. Yan, Z. Wang and P. Jiang, Continuous accurate pH measurements of human GI tract using a digital pH-ISFET sensor inside a wireless capsule, *Measurement*, 2015, **64**, 49–56.

47 V. Matyash, G. Liebisch, T. V. Kurzchalia, A. Shevchenko and D. Schwudke, Lipid extraction by methyl- *tert* -butyl ether for high-throughput lipidomics, *J. Lipid Res.*, 2008, **49**, 1137–1146.

48 Richards et al., Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics, *Physiol. Behav.*, 2018, **176**, 139–148.

49 Y. Benjamini and Y. Hochberg, Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B*, 1995, **57**, 289–300.

Dissertation Conclusion

This work provides valuable insight into the composition of the human intestinal tract metabolome over time and by intestinal tract location. Considerations for analytical analysis was thoroughly addressed to present the complexity of chemicals within the human intestinal tract and provide a technically sound approach for measuring a large portion of the intestinal tract metabolome. Metabolite identification is a key step in metabolomics analysis and is essential for generating accurate metabolomics datasets. A dietary intervention study with ground flaxseed reported significantly changed metabolites in serum and stool in response to consuming flaxseed. This study presents an example of how non-targeted metabolomics reveals changes due to diet, while targeted hypothesis driven analysis of bile acids in these same samples did not reveal changes due to flaxseed consumption. The overall diet of subjects was not controlled during this study which was presented as a confounding variable and limitation for studies investigating the metabolic impact of diet. A study of aspirates taken from the upper small intestine presented an in-depth analysis of one single subject from one area of the intestinal tract. These data revealed highly variable and complex chemical composition of the human small intestine. Many metabolites that showed similar trends were not able to be linked to known biological functions. These facts show that there is tremendous opportunity to discover novel functionality of the digestive tract and how it interacts with food. Investigation of samples collected from 15 healthy volunteers provided the first ever measurement of bile acid profiles across the length of the human intestinal tract *in vivo* using a non-invasive sampling approach. This study revealed trends in bile acids that are in agreement with the trends expected based on known bile acid metabolism. The bile

acids in intestinal tract samples were found to differ profoundly in proportion and concentration compared to stool samples. There is high inter-individual variation in bile acid profiles, specifically with respect to microbially conjugated bile acids. These intestinal tract samples will be further analyzed using multiple analytical platforms to maximize the coverage of metabolite measurement from the intestinal tract. The human intestinal tract is an extraordinarily complex system. By investigating the metabolome of the human intestinal tract, unknown relations between food, humans and microbes will be discovered.