Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 2305

Development of advanced chemical biology tools for microbiome metabolism

Chemoselective probes for enhanced metabolomics analysis

WEIFENG LIN

ACTA UNIVERSITATIS
UPSALIENSIS
2023

ISSN 1651-6214
ISBN 978-91-513-1894-3
urn:nbn:se:uu:diva-510431

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in room - A1:107a, BMC, Husargatan 3, Uppsala, Friday, 27 October 2023 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Frank Schroeder (Cornell University, The United States of America).

Abstract
Lin, W. 2023. Development of advanced chemical biology tools for microbiome metabolism. Chemoselective probes for enhanced metabolomics analysis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2305. 75 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1894-3.

The human microbiome has a profound impact on host physiology by generating highly reactive compounds that can contribute to the development of diseases. These microbial metabolites have a substantial potential that can serve as valuable indicators or biomarkers for different health conditions. Nevertheless, elucidating the microbiota composition and function remains challenging due to its remarkable diversity and complex. Furthermore, conducting a comprehensive analysis of the entire metabolome in a single analytical measurement is difficult. Researchers often employ derivatization techniques in analytical chemistry, which involve modifying the chemical structure of molecules to enhance their detectability, ionization properties and stability during analysis. However, derivatization carries the risk of introducing artifacts or chemical alterations that may compromise the accuracy of analytical results. Consequently, more advanced techniques are urgently required to improve the precision of derivatization-based metabolomics.

In response to this challenge, we have developed chemoselective probes immobilized onto magnetic beads to capture metabolites within biological samples. This innovative method improves the mass spectrometric sensitivity by up to a factor of one million, due to the efficient removal of sample matrix background through magnetic separation and improved ionization properties of the metabolites via derivatization. Our approach, termed *quant*-SCHEMA, has demonstrated the qualitative detection of metabolites containing carbonyl and amine groups with exceptional sensitivity and reproducibility. Additionally, we have successfully applied this method with improved probe design to quantitatively analyse carbonyl-containing metabolites, leading to the discovery of four valuable nutritional biomarkers. Furthermore, we have developed a precise quantification method for short-chain fatty acids (SCFAs) based on this chemoselective probe. The successful implementation of our chemoselective probes highlights the importance of chemical biology tools in advancing metabolomics, which we have termed chemical metabolomics,

This comprehensive mass spectrometric analysis expands the horizons of metabolomics-driven biomarker discovery. We envision that our innovative chemical biology tool will find widespread utility in metabolomics analysis, providing valuable insights into microbial interactions with the human host and the development of diseases.

Keywords: Chemical biology, Chemoselective modification, Chemical metabolomics, Bioorganic, Microbiome metabolism

Weifeng Lin, Department of Chemistry - BMC, Analytical Chemistry, Box 599, Uppsala University, SE-75124 Uppsala, Sweden. Science for Life Laboratory, SciLifeLab, Box 256, Uppsala University, SE-75105 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1894-3
URN urn:nbn:se:uu:diva-510431 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-510431)

To my family,

to myself.

吾心向光光自来

獨上高樓，
望盡天涯路。

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Lin, W.**, Conway, L.P., Block, A., Sommi, G., Vujasinovic, M., Löhr, J.-M., Globisch, D. (2020) Sensitive mass spectrometric analysis of carbonyl metabolites in human urine and fecal samples using chemoselective modification. *Analyst*, 145(11):3822-3831.

II. **Lin, W.**, Yang, Z., Kaur, A., Block, A., Vujasinovic, M., Löhr, J.-M., Globisch, D. (2021) Squaric acid as a new chemoselective moiety for mass spectrometry-based metabolomics analysis of amines. *RSC Chemical Biology*, 2(5): 1479-1483.

III. **Lin, W.**, Conway, L.P., Vujasinovic, M., Löhr, J.-M., Globisch, D. (2021) Chemoselective and highly sensitive quantification of gut microbiome and human metabolites. *Angewandte Chemie International Edition*, 133(43): 23420-23428.

IV. **Lin, W.**, Romero García, F., Norin, E.L., Kart, D., Engstrand, L., Du, J., Globisch, D. (2023) Sensitive quantification of short-chain fatty acids combined with global metabolomics in microbiome cultures. *Chemical Communications*, 59 (39): 5843-5846.

V. **Lin, W.**, Mellinghaus, K., Rodriguez-Mateos, A., Globisch, D. (2023) Identification of nutritional biomarkers through highly sensitive and chemoselective metabolomics. *Food Chemistry*, 425, 136481.

List of Additional Papers

VI. Conway, L.P., Garg, N., **Lin, W.**, Vujasinovic, M., Löhr, J.-M., Globisch, D. (2019) Chemoselective probe for detailed analysis of ketones and aldehydes produced by gut microbiota in human samples. *Chemical Communications*, 55(62): 9080-9083.

VII. Correia, M.S.P., **Lin, W.**, Aria, A.J., Jain, A., Globisch, D. (2020) Rapid preparation of a large sulfated metabolite library for structure validation in human samples. *Metabolites*, 10(10): 415.

VIII. Vallianatou, T., **Lin, W.**, Bechet, N.B., Correia, M.S.P., Shanbhag, N.C., Lundgaard, I., Globisch, D. (2021) Differential regulation of oxidative stress, microbiota-derived, and energy metabolites in the mouse brain during sleep. *Journal of Cerebral Blood Flow & Metabolism*, 41(12): 3324-3338.

IX. Kaur, A., **Lin, W.**, Dovhalyuk, V., Driutti, L., Di Martino, M., Vujasinovic, M., Löhr, J.-M., Sellin, M.E., Globisch, D. (2023) Chemoselective bicyclobutane-based mass spectrometric detection of biological thiols uncovers human and bacterial metabolites. *Chemical Science*, 14(20): 5291-5301.

X. Bi, H., Tranell, J., Harper, D., **Lin, W.**, Li, J. Hellström, A.R., Larsson, M., Rubin, C.-J., Wang, C., Sayyab, S., Kerje, S., Bed'hom, B., Gourichon, D., Ito, S., Wakamatsu, K., Tixier-Boichard, M., Marks, M.S., Globisch, D., Andersson, L. (2023) A frame-shift mutation in COMTD1 is associated with impaired pheomelanin pigmentation in chicken. *PLOS Genetics*, 19(4): e1010724.

Contents

Abbreviations

LC-MS	Liquid chromatography mass spectrometry
UPLC-MS	Ultra-performance liquid chromatography mass spectrometry
GC	Gas chromatography
ESI	Electrospray ionization
qTOF	Quadrupole time-of-flight
MS/MS	Tandem mass spectrometry
EIC	Extracted ion chromatogram
NMR	Nuclear magnetic resonance
TLC	Thin Layer Chromatography
HMDB	Human metabolome database
FooDB	Food database
CSF	Cerebrospinal fluid
ROS	Reactive oxygen species
SCFAs	Short-chain fatty acids
TCA cycle	Tricarboxylic acid cycle
QC	Quality control
ROC	Receiver operating characteristic
AUC	Area under the curve
PCA	Principal Component analysis
CV%	Coefficient of variation
LOD	Limit of detection
LOQ	Limit of quantification
Boc	*tert*-Butyloxycarbonyl
Fmoc	Fluorenylmethyloxycarbonyl
Cbz	Benzyl carbamates
DCM	Dichloromethane
DMF	Dimethylformamide
THF	Tetrahydrofuran
TFA	Trifluoroacetic acid
DIPEA	*N,N*-Diisopropylethylamine
TCEP	Tris(2-carboxyethyl)phosphine
NHS	*N*-Hydroxysuccinimide

STP	4-Sulfo-2,3,5,6-Tetrafluorophenol
MBA	Methylbenzylamine
BCB	Bicyclobutane
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
DMAP	4-Dimethylaminopyridine
HOBT	Hydroxybenzotriazole
HBTU	Hexafluorophosphate benzotriazole tetramethyl uronium
TEA	Triethylamine
NOC	p-Nitrocinnamyloxy-carbonyl

Introduction

Trillions of diverse microorganisms inhabit the surfaces and cavities of the human body, with a particularly significant presence in the gastrointestinal tract[1-3]. These microbial communities play an active role in various metabolic processes and engage in the exchange of metabolites, both among themselves and with their human host[4-7]. A major advantage of the co-evolved relationship between these communities and their host lies in their collective defences against pathogenic infections[8-10]. In contrast, abnormal alterations of the composition of the gut microbiota, termed microbiota dysbiosis, have been associated with the development of diseases and a range of other pathological conditions[11-13]. While major progress has been made through metagenomic sequencing, which has revealed alterations in the abundance of specific bacterial species, the molecular mechanisms underlying communication between the host and the microbiota largely remain a mystery[3, 14-17]. It is worth highlighting that the genetic pool of the microbiota contains approximately 400 times more genetic information than the human genome. Within this vast genetic reservoir, numerous genes encode metabolic reactions orthogonal to the human metabolism, giving rise to the metabolic modification of xenobiotics—com-

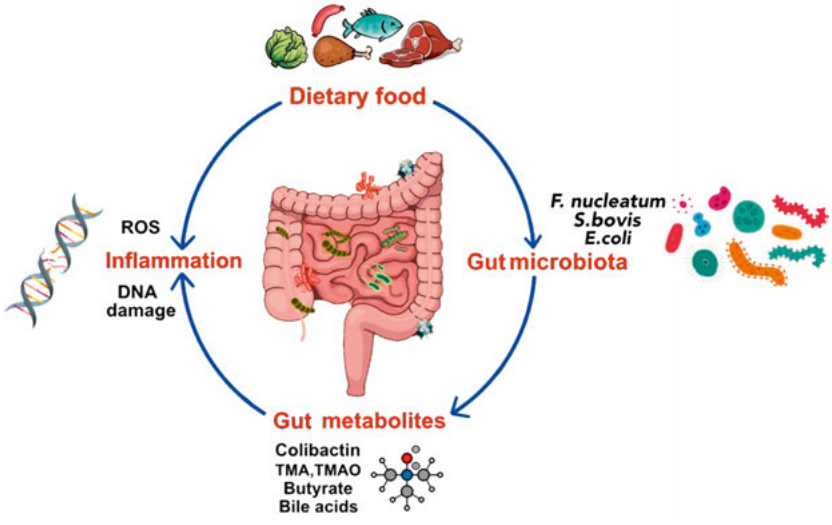


Figure 1 Overview of the connections among xenobiotic, gut microbiota, metabolites, and diseases. Figure adapted from reference 17.

pounds that cannot be produced by the human body[5, 6, 18] (Figure 1). Importantly, many of these microbial-derived compounds possess unknown bioactivity[19-21]. Elucidating the chemical structures of these bioactive metabolites and understanding their roles in human physiology holds the potential to provide insights into the complex interactions between the microbiota and their host, thereby enhancing our comprehension of relevant disease mechanisms. Notably, data from The Human Metabolome Database (HMDB) revealed that more than 80% of predicted metabolites remain not detected and identified[22-25].

To date, a principal goal of metabolomics biomedical research is to explore certain specific biomarkers of multiple diseases, drug toxicity, and efficacy[26-28]. Besides, personalized medicine and precision medicine are the other benefits from biomedical research in metabolomics[29, 30]. For instance, these biomarkers could be applied for the selection of optimal therapeutic intervention and evaluation of response to treatment. Metabolomics-derived biomarkers can have a heavy impact on clinical and translational study because most of the metabolites are species-independent and the evaluation methods of these metabolites are usually non-invasive[26, 31, 32]. Metabolomics research is mainly focused on the analysis of the metabolome – the small molecules (< 1 kDa), which can be found in a specific biological sample or an organism – and is therefore the emerging method of choice for the analysis and discovery of known as well as unknown microbiota-derived metabolites[33]. However, the analysis of these metabolites poses significant challenges due to the complexity of the microbiota, chemical diversity, and the requirement for high detection sensitivity[34-36]. Therefore, the development of advanced techniques is urgently needed to overcome analytical limitation in metabolomics research.

The following part of the introduction will focus on the current development of metabolomics approaches, existing challenges and a detailed discussion on chemical metabolomics strategies.

Current analytical approaches for metabolomics

Metabolomics aims at the comprehensive investigation of all metabolites present within biological samples. These metabolite extracts undergo analysis using either nuclear magnetic resonance (NMR) spectroscopy[37, 38] or mass spectrometry (MS)[39-41]. Each of these techniques has their distinct strengths and limitations in metabolomics analysis. NMR stands out as a non-destructive technique, allowing samples to be utilized for further studies without any loss. Additionally, NMR based metabolomics can provide direct quantitative analysis. However, NMR's drawback often lies in its relatively low sensitivity, which limits its capability to detect a broad spectrum of metabolite signals. In

contrast, MS exhibits heightened sensitivity compared to NMR and it is able to detect metabolites that NMR is not capable of. Nevertheless, the exceptional sensitivity of MS introduces the challenge of background noise, which unavoidably complicates the subsequent data analysis. This often results in metabolites with weak ionisability hidden through the background noise. Moreover, the application of MS frequently requires coupling with high performance separation techniques, such as liquid chromatography (LC) or gas chromatography (GC), which are the relatively popular technique for effectively separating metabolites within analysed samples[42]. Mass spectrometric analysis has emerged as the method of choice for global metabolomics analysis due to its ability to detect and analyse thousands of metabolites with superior sensitivity compared to NMR analysis. Furthermore, MS based metabolomics analysis allows us to perform MS/MS fragmentation to obtain each fingerprints of the each desired molecule[43]. Since molecule has their specific fragmentation patterns, we can validate the metabolite structure via MS/MS spectrum comparison with online database.

Challenging in conventional metabolomics

To comprehensively analyse the complete or near-complete metabolome demands methodologies that exhibit high metabolic coverage, sensitive detection, accurate quantification, and confident metabolite identification. However, it is analytically challenging due to the extreme complexity of the metabolome and various practical considerations[44]. Firstly, the complex diversity in the chemical and physical properties in metabolites often yields inadequate instrument responses simultaneously within a single analysis[41]. Consequently, employing multiple instrumental approaches for a single biological sample becomes necessary to maximize the coverage of the metabolome (Figure 2A). It unfortunately introduced extra time and costs. Secondly, the wide range of concentrations (from pM to mM) in biological molecules renders the impossibility to obtain all metabolic information qualitatively or quantitatively at the same time[45, 46]. Additionally, the detection and quantification of the low abundant metabolites are difficult, particularly when these metabolites generate weak instrumental responses[44]. Thirdly, biological samples usually are very precious, irrespective of where they are collected from. In certain cases, the sample amount can be exceedingly limited. For instance, cerebrospinal fluid (CSF) is one of the best samples to be investigated for metabolic profiling of brain activities. However, the collection of the CSF demands professional operation with care, and the volume of this sample type can be little (i.e., approximately 10 μL from a mouse)[47, 48]. Thus, highly sensitive methods are required to analyse these sample types. Finally, while numerous metabolite databases have emerged in the past two decades, rapid identification of both known and unknown metabolites in metabolomics research remains an

additional challenge. It is mainly because of the limited availability of the authentic standards and complicated sample matrices that can leave the difficulty in the acquisition of high-quality MS and MS/MS spectra.

Derivatization

Derivatization is one of the most promising approaches to conquer these limitations. This strategy involves the integration of chemical modification with conventional metabolomics to improve the mass spectrometric analysis[49-51]. The rationale behind this strategy is that over 95% of the metabolites contain at least one functional group among amine, carboxylic acid, hydroxyl, carbonyl and thiol, offering opportunities for chemical modification[22]. Using chemical modification on metabolites not only allows for the enhancement of the ionization of metabolites during MS analysis but also enable the improvement of chemical property for better separation performance in GC or LC. The advancement of derivatization holds the potential to discover previously undetectable metabolites that are inaccessible using standard metabolomics approaches (Figure 2B). While numerous chemical derivatization methods have been reported over the last decades[52-59], a crucial aspect – the isolation step – has often been overlooked, leading to the generation of inferior sample matrices. The high abundance of derivatization reagents can cause ion suppression and instrument contamination, while the harsh conditions applied may even alter metabolite structures.

Although a few methods have been developed over the past two decades to address these challenges, these approaches frequently suffer from other limitations[60, 61]. Therefore, it is crucial to develop chemical biology techniques that allow for the selective discovery and analysis of these compounds. These

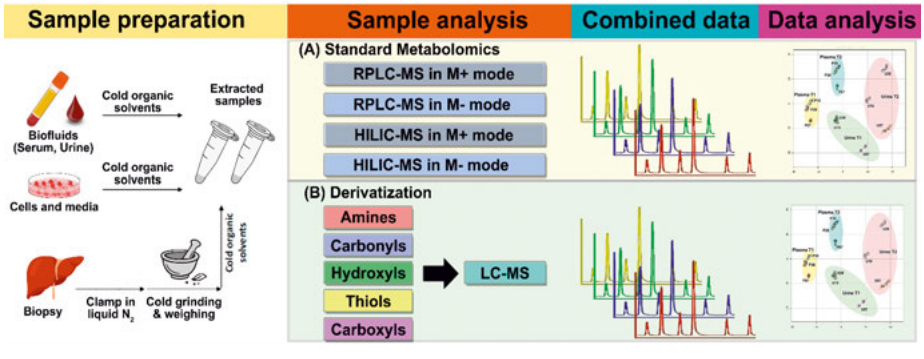


Figure 2 Overview of global metabolomics and the comparison between standard metabolomics process (A) and derivatization-based metabolomics (B). The conventional metabolomics includes sample preparation, sample analysis and data analysis. Figure inspired by literature S. Zhao et al.

methodologies are urgently needed to unveil the bioactivity of these metabolites and their impact on the human host. By investigating the interplay between microbiota and their metabolic by-products, we can gain profound insights into the complex web of interactions that govern human health and pave the way for novel therapeutic approaches.

Chemical derivatization in metabolomics

Chemical derivatization, a well-established practice in analytical chemistry, has been used to tackle many analytical issues. The goal of chemical derivatization lies in transforming analytes into new compounds with altered chemical and physical properties for improved analysis using chemical reagents. The application of chemical derivatization in metabolomics were part of many successes over recent two decades. For example, METPR was firstly developed by Cravatt and his laboratory to tag, capture and profile specific portions of the metabolome[61]. 4-channel chemical isotope labelling (CIL) has been developed to profile metabolomics with covering amine/phenol, hydroxyl, carboxyl, and carbonyl containing metabolites by Li and co-workers[62]. Similarly, MetFish has also been developed using similar labelling reagents to identify and quantify amine, carboxyl, carbonyl and hydroxyl metabolites with an additional step for metabolites isolation using liquid-liquid extraction[60]. There are some other studies included chemical derivatization, but their coverage of the metabolite classes is limited[57]. The simple utilization of derivatization reagents without organic chemistry thoughts behind gives rise to many issues, such as chemoselectivity challenges, side reactions, limited yield and harsh condition to be applied. The ideal chemical reaction for metabolite derivatization should encompass following key factors: 1) highly chemoselectivity; 2) mild reaction condition, ideally in room temperature; 3) green chemistry, no additional reagents needed; 4) high reaction conversion rate. To attain these goals, expertise in organic chemistry is highly recommended in the development of optimal chemoselective modifications for metabolite analysis.

Current chemoselective reaction overview

Many chemical derivatization reagents have been developed over the past decades. The summary of diverse derivatization reagents can be found in Figure 3. Most of the chemoselective modifications in metabolomics have been inspired by chemoproteomics as the development of proteomics is more advanced than metabolomics field[63]. However, it is noted that the reagents require different properties for the investigation of metabolites compared to proteins.

Chemoselective moieties for amine modification

Amine-containing metabolites are crucial in many biological processes, such as cellular functions, signalling pathways, and overall physiological homeostasis[64]. They also constitute the predominant portion of metabolites derived from the gut microbiota. For example, dopamine, serotonin, norepinephrine, and histamine, which function as neurotransmitters and are representative amine-containing metabolites[65]. In addition, amino acids, essential components of proteins, contain amine functional groups. Amino acids metabolism also provide valuable insights into investigating various disease developments[66].

Although metabolites with amine groups are easy to protonate under acidic mobile phase condition and ionize in the ESI source, the high polarity of amine-containing metabolites presents challenges in achieving adequate retention on reverse-phase columns for effective separation. There are several reagents widely used in amine modification including dansyl chloride[62], 4-aminophenyl disulfide (APDS)[67] and NHS-ester[68].

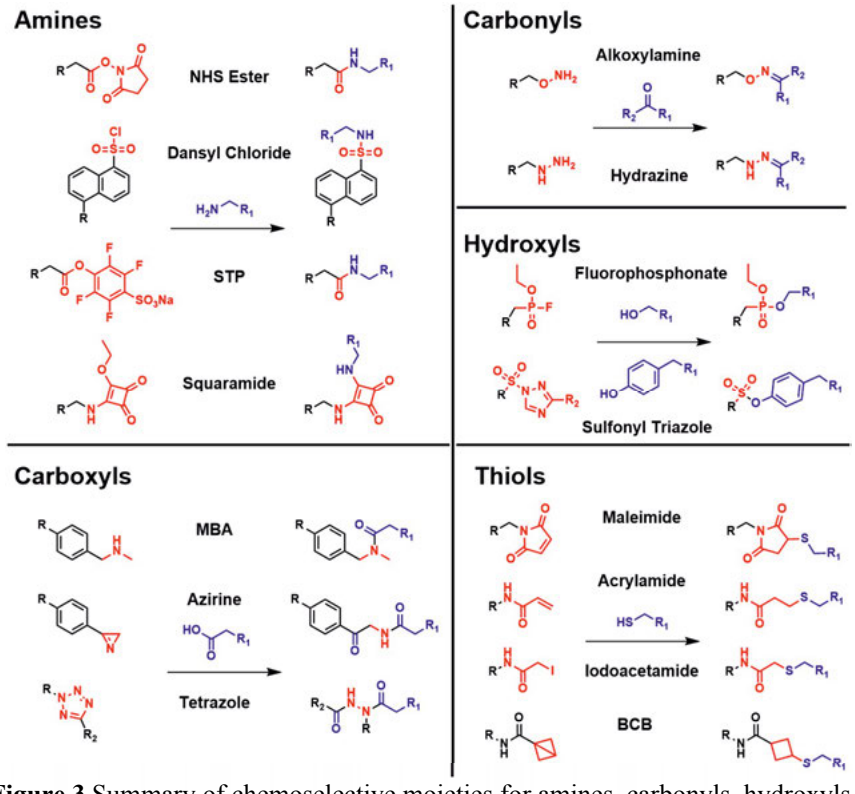


Figure 3 Summary of chemoselective moieties for amines, carbonyls, hydroxyls, carboxyls and thiols. *N*-Hydroxysuccinimide (NHS); 4-Sulfo-2,3,5,6-Tetrafluorophenol (STP); Methylbenzylamine (MBA); Bicyclobutane (BCB).

Chemoselective moieties for carbonyl modification

Carbonyl-containing metabolites are a highly reactive compound class that can form conjugates with DNA and proteins[69]. They participate in a wide range of biochemical pathways, and their dysregulation has been linked to many pathological conditions[70-72]. Moreover, the International Agency for Research on Cancer has classified six carbonyl-containing metabolites as carcinogens[73]. Within this class, reactive oxygen species (ROS) play a crucial role in cellular signalling across many biochemical processes. However, ROS dysregulation can prompt the reaction of carbonyl species with cellular macromolecules, resulting in detrimental disruptions to normal physiology[74, 75]. Imbalances between ROS production and antioxidant defences have been associated with pathologies in diseases such as cancer, retinopathy, and asthma[76].

Carbonyl, a neutral functional group, poses challenges for ionization in ESI, regardless of whether in positive or negative mode analysis. Additionally, carbonyl-containing metabolites, often important and low in abundance such as hormones. The derivatization can significantly enhance detection sensitivity, and several labelling reagents are widely used[77]. For instance, hydrazine and alkoxyamine are two very common reagents that react with carbonyls to form hydrazone and alkoxime respectively. The main difference between these two reactive moieties is the reversibility. Hydrazones tend to be hydrolysed more easily under neutral conditions compared to alkoxime due to the inductive effect of oxygen in alkoxime compared to nitrogen in hydrazine[78]. In addition, reductive amination is another strategy to react with carbonyls, but limited application has been found due to harsh conditions using reducing reagent.

Chemoselective moieties for carboxylic acid modification

Carboxylic acids are important metabolites due to their central roles in a wide range of biochemical processes in human body. They serve as essential constituents in energy-producing pathways like the Krebs cycle, contributing to the generation of ATP. The carboxylic acids analysis allows for the detailed investigation of entire TCA cycle, but the multi-carboxylate groups in these metabolites present a challenging in the accuracy of the analysis. Furthermore, carboxylic acids serve as foundational building blocks for the biosynthesis of many biomolecules. Beyond their structural functions, carboxylic acids also function as buffering agents in the regulation of biological fluid pH and the maintenance of a stable internal environment. Notably, short-chain fatty acids (SCFAs), a subset of carboxylic acids, hold significant importance and are extensively produced by the gut microbiome[79].

The detection of carboxylic acids using LC-MS is usually conducted in negative mode analysis. However, the sensitivity of carboxylic acids is relatively low, often suffered from matrix effects. One of the most common strategies to

label carboxylates is to use primary or secondary amine to react with carboxylic acid via amide coupling, such as DMED[80] and 4-BNMA[53]. However, these reactions required amide coupling reagents like EDC or DMAP, which can cause undesired side reactions between endogenous amines and targeted carboxylic acids.

Chemoselective moieties for hydroxyl modification

Alcoholic hydroxyl groups are generally neutral moiety with low ionization efficiency. Derivatization of hydroxyl can significantly improve their detection efficiency. Since the hydroxyl is also nucleophile but weaker compared to amines, the highly chemoselective reaction are required to avoid any side reaction with amines in biological samples[81]. To date, there is limited methods available to chemoselectively react with alcohols, but dansyl chloride has been used to label hydroxyl group with different condition compared to amine reaction as a compromised choice[82]. Similarly, phenol also contains hydroxyl group, but it has different chemical properties compared to alcohols. The availability of developed chemoselective reaction is limited.

Chemoselective moieties for thiol modification

Thiols represent an important class of metabolites that is crucial for the regulation of homeostasis[83]. Cysteine (Cys) and glutathione (GSH) stand as essential biological thiols, working to maintain cellular redox balance[84, 85]. Altered levels of these two thiols lead to higher mortality rates attributed to pathological conditions including coronary artery disease, Crohn's disease, and type 2 diabetes[86, 87]. Additionally, the host-microbiota mutualism, these microorganisms also synthesize and maintain specific thiol-containing molecules to several essential cellular functions[88].

The derivatization of thiols requires not only the enhancement of detection efficiency but also the consideration of stability improvement. Biological thiols, unlike other functional groups, exhibit limited stability and often undergo natural oxidation to form disulfides during sample preparation processes[89]. Consequently, in some cases, pretreatment with agents like TCEP or other reducing reagents becomes necessary to accurately detect biological thiols. Commonly used derivatization reagents for reacting with thiols include bromoacetonylquinolinium bromide[90] and iodoacetamide[91]. Recently, the development of bicyclobutane by our laboratory has emerged as a mild-condition method for detecting thiols[92].

Chemical biology tools

Chemical biology is to study and manipulate biological systems at the molecular level, which requires instrumental or molecular tools to facilitate it. These

tools bridge the gap between chemistry and biology, enabling chemical biologists to explore and understand complex biological processes. Researchers in this field often utilize small molecule libraries, chemical probes, fluorescent labelling, bioorthogonal chemistry, optogenetics and protein engineering tools[93]. Cleavable linkers are also an essential component of chemical biology tools, which enables precise control over the release of molecules and providing valuable insights into biological processes and therapeutic interventions[94].

Cleavable linker for sample matrix removal

Cleavable linkers have been broadly utilized in other chemical biological fields, such as proteomics, drug delivery and controlled release systems (Figure 4)[94]. These linkers are designed to be cleaved under specific chemical conditions. Depending on the conditions of cleavage, they can be categorized as photocleavable, acid-labile, redox-sensitive, enzyme-cleavable, and more. Despite their broad application in related fields, the incorporation of cleavable linkers in metabolomics remains relatively limited. There are several major advantages to include a cleavable linker in chemical metabolomics workflow. (1) It affords chemical accessibility, enabling the connection between the reactive site and solid phase support moiety. This facilitates the isolation and enrichment of captured metabolites. (2) The step of sample matrix removal results in reduced ion suppression and improved peak shape due to cleaner chromatography. (3) The metabolite isolation step can reduce the risk of the false positive identification of metabolites. (4) The complex sample matrix removal can help us to unify the LC-MS output across different sample types

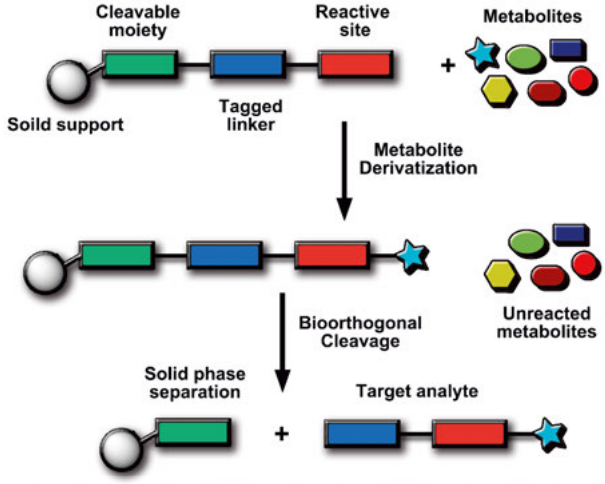


Figure 4 Strategy of using cleavable linker in metabolomics analysis. Solid support can be used for metabolite isolation. Cleavable moiety can help for the release of metabolites after capturing. Tagged linker can be designed for specific diagnostic properties. Reactive site helps for targeting at desired metabolites.

with different salt concentration. A uniform analytical solution matrix can enhance accuracy of comparing different metabolites among different sample types. These mentioned advantages are often absent in the conventional chemical derivatization methods.

Other chemical biology developments in metabolomics

An increasing number of chemical biologists are engaging in metabolomics to explore and elucidate the metabolism of humans and other organisms. For example, the Schroeder laboratory has undertaken extensive research, using the *C. elegans* model organism, to identify previously unknown metabolites that regulate various biological processes[95-98]. Their dedication to validate these metabolites of interest through total synthesis or isolation followed by 2D NMR spectroscopic analysis has unveiled the structures and biological functions of unknown metabolites. Similarly, the Johannsen laboratory has utilized their in-house chemical probes (alkyne-labelled metabolite analogue) to elucidate newly discovered metabolites[99-101]. Meanwhile, the Balskus laboratory has combined enzymatic assays with metabolomics to decipher the human microbiome[102-104]. These chemical biology tools have not only advanced the field of metabolomics but have also contributed significantly to our improved understanding of human physiology and metabolism.

Quantification in metabolomics

Quantification is one of the essential methods in analytical chemistry to determine the amount or concentration of specific analytes in samples. Calibration curves are needed to build up for the quantification of molecules of interest, which is plotted by measuring a series of known concentration of target molecules. Internal standard calibration is usually more preferable due to its higher accuracy and precision. An internal standard is added into unknown samples, and the ratio of the analyte's peak area to the internal standard's peak area is used to determine the concentration of the analyte, compensating for variations in sample preparation and instrument response.

Isotopic labelling is a powerful technique in quantification in mass spectrometry-based analysis, such as proteomics and metabolomics[105]. There are two main types of isotopic labelling techniques commonly used for quantification. The first one is to incorporate the stable isotopes in cleavable linkers allow us to tag the analytes to accurately quantify the abundance of molecules in different samples. For example, Isotope-Coded Affinity Tags (ICAT) in protein quantification[106] and Chemical isotopic labelling (CIL) in metabolomics[62]. Another quantification method using isotopic labelling is called Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC)[107]. Cells are grown in a

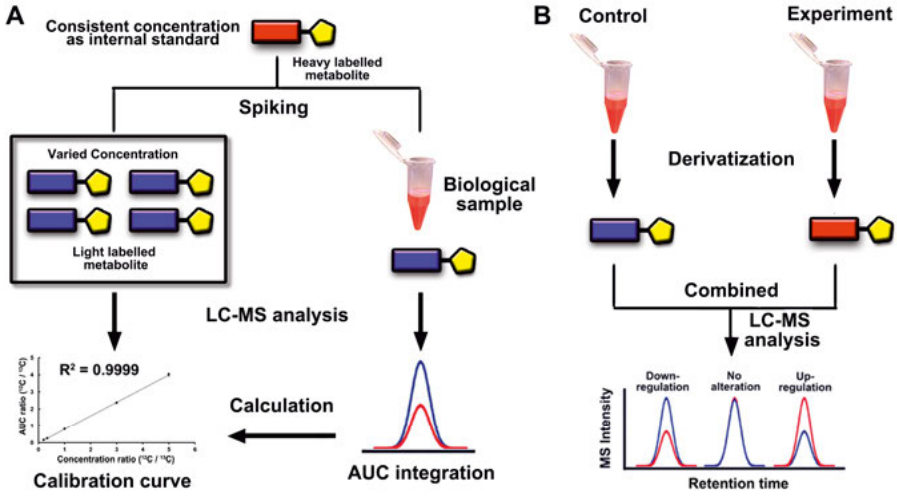


Figure 5 Overview of (A) precise quantification and (B) relative quantification using isotopical labelling.

culture medium containing stable isotopic labelled amino acids to investigate the protein expression levels quantitatively.

Quantification using isotopic labelling have major advantages. Firstly, it allows for precise and accurate quantification of analytes in complex mixtures due to the distinguishing mass difference between labelled and unlabelled samples. The direct comparison between samples in a single MS experiment provide better quality of the data outcomes compared to the conventional semi-quantitative analysis. Secondly, isotopically labelled molecules can also serve as internal standards to eliminate the variation in sample preparation.

Precise quantification using internal standards

Precise quantification using internal standards is a common analytical strategy to enhance the accuracy and reliability of measurement of specific molecules[108]. By using internal standards, we can not only obtain and calculate the concentration of target molecules, but also remove any variations due to sample handling or matrix effects to obtain highly accurate measurements[109]. Furthermore, when we quantify target molecules using internal standards, calibration curves can be recorded for different sample types to ensure reliable quantification[108]. Additionally, the ideal internal standards are isotopic labelled compounds with at least 3 Da difference to avoid any interference from natural isotopic pattern[110]. However, these isotopic labelled internal standards are generally expensive or even not commercially available, which leads to the difficulty to precise quantify molecules of interest. In derivatization strategies, we can have isotopic labelled in cleavable linker, which allows us to

readily synthesize any molecules of interest within limited organic chemistry steps.

Relative quantification for biomarker discovery

Metabolomics is one of the powerful methods for biomarker discovery, as it enables the identification of specific metabolites can serve as indicators of biological process, disease states and response to treatments[111-113]. Untargeted metabolomics can rapidly quantify thousands of metabolites simultaneously to obtain alteration of their abundance between different sample sets. However, the conventional metabolomics studies usually are semi-quantitative meaning that they required to compare the abundance of interesting metabolites in two different injections in LC-MS. The comparison between two different LC-MS injections can introduce many instrumental errors since the LCMS responses can be varied in different injections. Although many computational methods have been reported to use internal standards or quality control (QC) samples for data normalization[114-118], it is not likely to remove instrumental errors entirely. Derivatization can solve this problem properly, as it will only need a single LC-MS analysis by combining the control sample (light-tagged) and comparative sample (heavy-tagged). The major advantage of this method is that we can significantly remove the entire instrumental error, as the natural metabolites and isotopic labelled metabolites are analysed simultaneously due to their nearly identical chemical and physical properties. The isotopic labelling tags will allow us to differentiate where the metabolite comes from. Additionally, chemical metabolomics does not require normalization step after obtaining the MS data, which can be more accurate and more time efficient.

Aims

In this PhD thesis, we aimed at developing new chemical biology methods that can advance the analysis of metabolites qualitatively and quantitatively. These methods were focused on using chemical tools to chemoselectively investigate metabolites and ultimately to identify new biomarkers. The detailed aims were:

1. Development of the new reactive sites for multifunctional metabolomics analysis targeting amines, carbonyls (ketones/aldehydes), and carboxylic acids.

2. Design and development of chemoselective probes.

3. Identification of amine- and carbonyl-containing metabolites from human urine, plasma, fecal samples.

4. Application of chemoselective probes for quantitative analysis of carbonyl-containing metabolites in human samples.

5. Application of chemoselective probes for precise quantification of SCFAs in bacterial cultures.

6. Nutritional biomarker discovery using chemical metabolomics in human urine samples.

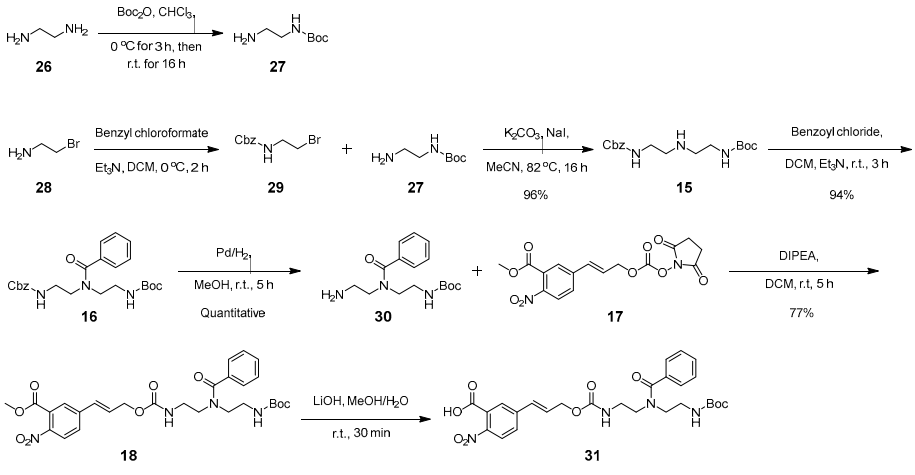
Methods

General

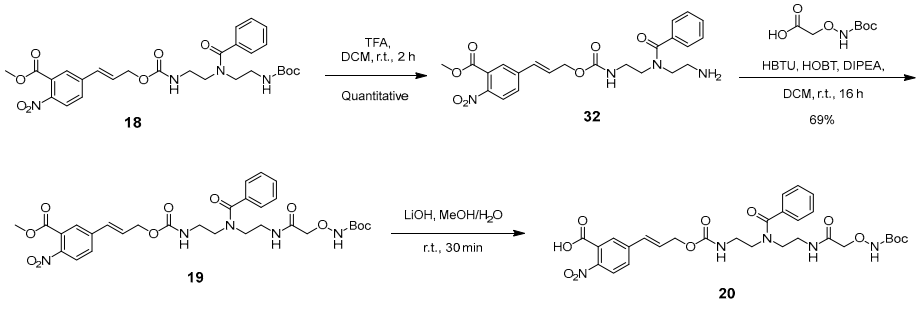
All non-aqueous reactions were performed using flame- or oven dried glassware under an atmosphere of dry nitrogen. All reagents and solvents were purchased from Sigma-Aldrich or Fischer Scientific and were used without further purification. The in-house built metabolite library was obtained from MetaSci. Mass spectrometry grade solvents were used for UHPLC-ESI-MS analysis. Solutions were concentrated in vacuo on a Heidolph or an IKA rotary evaporator. Thin Layer Chromatography (TLC) was performed on silica gel 60 F-254 plates. Visualization of the developed chromatogram was performed using fluorescence quenching or staining with CAM (cerium ammonium molybdate), ninhydrin. Chromatographic purification of products was accomplished using flash column chromatography on Merck silica gel 60 (40−63 μm. All synthesized compounds were $\geqslant$95% pure as determined by NMR. NMR spectra were recorded on Agilent 400 MHz spectrometer (1H NMR: 399.97 MHz, 13C NMR: 100.58 MHz). Chemical shifts are reported in parts per million (ppm) on the δ scale from an internal standard. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Glass vials used for handling magnetic beads were microwave vials from Biotage (0.2-0.5 mL or 0.5-2.0 mL). All chemical synthesis protocols and characterization data are available in the Supplementary Information.

Synthesis Schemes

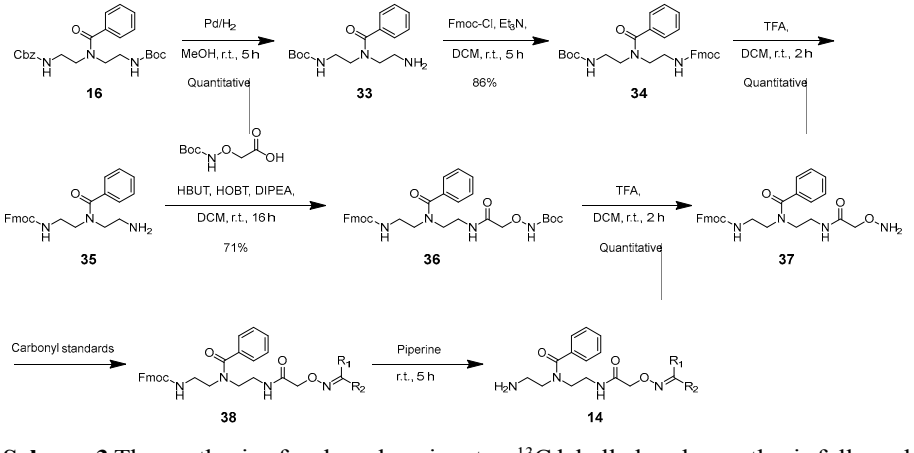
Synthesis Scheme of each chemical probes are included here. The details of each chemical synthesis step can be found in Supplementary Information. Notably, phenyl-$^{13}C_6$ labelled compounds have been synthesized following the same scheme.



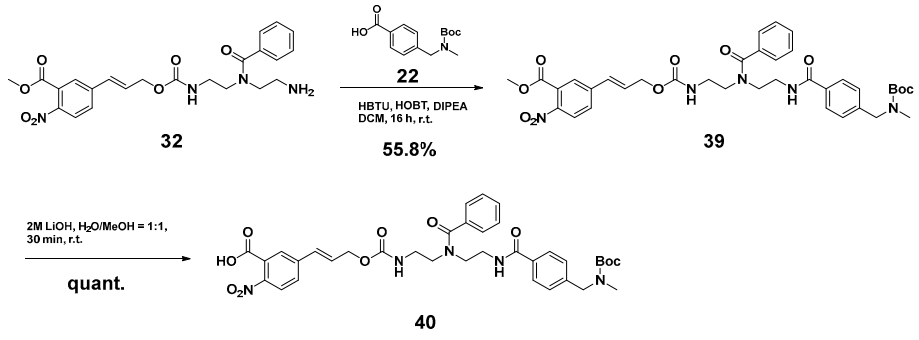
Scheme 1 The synthesis of general probe. ^{13}C labelled probe synthesis followed the same procedure.



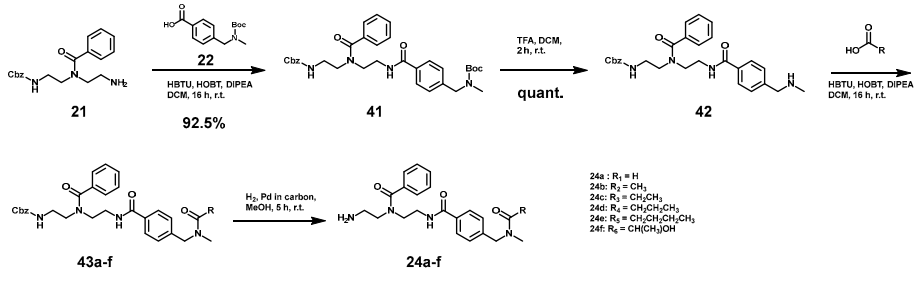
Scheme 2 The synthesis of carbonyl specific probe. ^{13}C labelled probe synthesis followed the same procedure.



Scheme 3 The synthesis of carbonyl conjugates. ^{13}C labelled probe synthesis followed the same procedure. The library construction procedure starts from compound **35**.

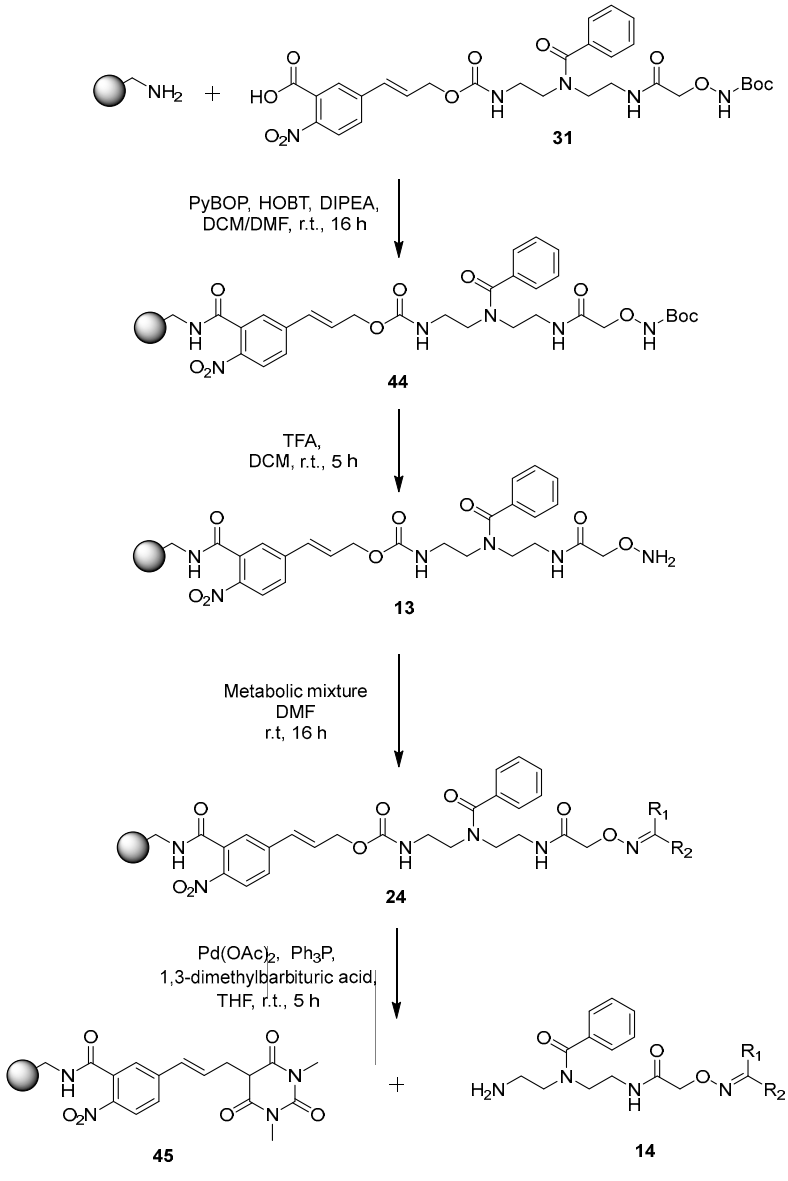


Scheme 4 The synthesis of SCFAs specific probe.



Scheme 5 The synthesis of SCFAs conjugates. ^{13}C labelled probe synthesis followed the same procedure.

Magnetic beads general handling scheme



UPLC-MS analysis

The UPLC-MS analysis was performed in a Synapt G2 Q-TOF mass spectrometer or a Maxis II ETD Q-TOF mass spectrometer using an electrospray ionization (ESI) source with an Acquity UPLC system or an Elite UHPLC system and equipped with an Acquity UPLC® BEH C18 column (1.7 μm, 100×2.1 mm) or a Waters ACQUITY UPLCVR HSS T3 column (1.8 mm, 100 2.1 mm). Milli-Q water with 0.1% formic acid was used as mobile phase A and LCMS grade methanol with 0.1% formic acid was used as mobile phase B. The column temperature was kept at 40 °C, and the autosampler was kept at 4 °C. The flow rate was set to 0.22 mL/min. The gradient used was as follows: 0–2 min, 0% B; 2–15 min, 0–100% B; 15–16 min, 100% B; 16–17 min, 100-0% B; 17–23 min, 0% B. The system was controlled using the Compass HyStar software package from Bruker. High-resolution mass spectra were acquired in positive mode at a mass range of m/z 50–1200. The samples were introduced into the q-TOF using positive electrospray ionization. A solution of sodium formate (0.5 mM in 2-propanol: water, 90:10, v/v) was used to calibrate the instrument. Data acquisition was performed in MSE mode. The samples were injected to the UPLC-MS system in a randomized order with QC samples injected in the beginning and end of the sample list in both ionization modes, as well as after every eight samples (7 QCs in each ionization mode in total).

Magnetic beads bound chemoselective probe preparation procedure

Preparation of carbonyl-specific probes

MagnaBind Amine Derivatized Beads slurry (50 μL, Thermo ScientificTM) was transferred into a 1.5 mL Eppendorf tube. Original solution from supplier was taken out by magnetic separation. The beads were washed with THF (2×150 μL) followed by phosphate buffer (2×150 μL, 25 nM, pH 7.5). DMF (150 μL) was added to the Eppendorf followed by 5 μL DIPEA and then vortexed for at least 30 s to yield the unprotonated amine. The beads were washed with DMF (150 μL) followed by DCM (150 μL). An amide coupling solution (4.5 mM PyBop, 3.3 mM HOBT, 1% DIPEA v/v in DCM) and probe solution (3 mM $^{12}C_6$ or $^{13}C_6$ probe in DMF) were freshly prepared in separate. The probe solution (100 μL) and amide coupling solution (100 μL) were combined into the Eppendorf tube containing magnetic beads. The mixture was shaken and incubated using a Thermomixer (1,500 rpm, 25 ℃, overnight.). The solution was removed and the beads consecutively washed with 2×150 μL THF and 2×150 μL DCM. After removal of all the solution, DCM (190 μL) and

TFA (10 μL) were added in sequence to the Eppendorf for Boc deprotection. The mixture was shaken and incubated with a Thermomixer (1,500 rpm, 25 ℃, 5 h). The reaction mixture was removed and followed by washing with THF (2×150 μL). DCM (150 μL) and DIPEA (10 μL) were added in sequence to the Eppendorf for amine deprotonation and TFA neutralization. The beads were washed with DMF (2×150 μL). The beads were suspended in the DMF (300 μL), ready to be used for sample treatment. **Note:** for carbonyl-containing metabolite detection only, ^{13}C probes then are not needed to be prepared.

Preparation of amine-specific probes

MagnaBind Amine Derivatized Beads slurry (50 μL, 320 nmol, Thermo ScientificTM) was transferred into a 1.5 mL Eppendorf tube. Original solution from supplier was taken out by magnetic separation. The beads were washed with THF (2×150 μL) followed by phosphate buffer (2×150 μL, 25 nM, pH 7.5). DMF (150 μL) was added to the Eppendorf followed by 5 μL DIPEA and then vortexed for at least 30 s to yield the unprotonated amine. The beads were washed with DMF (150 μL) followed by DCM (150 μL). An amide coupling solution (4.5 mM PyBop, 3.3 mM HOBT, 1% DIPEA v/v in DCM) and probe solution (3 mM probe in DMF) were freshly prepared in separate. The probe solution (100 μL) and amide coupling solution (100 μL) were combined into the Eppendorf tube containing magnetic beads. The mixture was shaken and incubated using a Thermomixer (1,600 rpm, 25 ℃, overnight.). The solution was removed and the beads consecutively washed with 2×150 μL THF and 2×150 μL DCM. After removal of all the solution, DCM (190 μL) and TFA (10 μL) were added in sequence to the Eppendorf for Boc deprotection. The mixture was shaken and incubated with a Thermomixer (1,500 rpm, 25 ℃, 5 h). The reaction mixture was removed and followed by washing with THF (2×150 μL). DCM (150 μL) and DIPEA (10 μL) were added in sequence to the Eppendorf for amine deprotonation and TFA neutralization. The beads were washed with DMF (2×150 μL) and EtOH (2×150 μL). The beads, suspended in the EtOH (300 μL), were added with 3,4-Diethoxy-3-cyclobutene-1,2-dione (5 μL) and trimethylamine (3 μL). The suspension was agitated at 25 °C in a ThermoMixer (1,600 rpm) for 16 h. After the reaction was complete, the supernatant was removed and the beads were washed with THF (3 × 200 μL) followed by EtOH (2× 200 μL). The beads were suspended in the EtOH (300 μL), ready to be used for sample treatment.

Preparation of SCFAs-specific probes

MagnaBind Amine Derivatized Beads slurry (50 μL, Thermo ScientificTM) was transferred into a 1.5 mL Eppendorf tube. Original solution from supplier was taken out by magnetic separation. The beads were washed with THF (2×150 μL) followed by phosphate buffer (2×150 μL, 25 nM, pH 7.5). DMF (150 μL) was added to the Eppendorf followed by 5 μL DIPEA and then vortexed for at least 30 s to yield the unprotonated amine. The beads were washed with DMF (150 μL) followed by DCM (150 μL). An amide coupling solution (4.5 mM PyBop, 3.3 mM HOBT, 1% DIPEA v/v in DCM) and probe solution (3 mM in DMF) were freshly prepared in separate. The probe solution (100 μL) and amide coupling solution (100 μL) were combined into the Eppendorf tube containing magnetic beads. The mixture was shaken and incubated using a Thermomixer (1,500 rpm, 25 ℃, overnight.). The solution was removed and the beads consecutively washed with 2×150 μL THF and 2×150 μL DCM. After removal of all the solution, DCM (190 μL) and TFA (10 μL) were added in sequence to the Eppendorf for Boc deprotection. The mixture was shaken and incubated with a Thermomixer (1,500 rpm, 25 ℃, 5 h). The reaction mixture was removed and followed by washing with THF (2 ×150 μL). DCM (150 μL) and DIPEA (10 μL) were added in sequence to the Eppendorf for amine deprotonation and TFA neutralization. The beads were washed with DMF (2×150 μL). The beads were suspended in the DMF (100 μL), ready to be used for sample treatment.

Sample preparation for chemical metabolomics

Preparation of fecal metabolite extracts

A scalpel was used to collect approximately 30 mg of the frozen fecal sample from ten different patients (stored at 80 ℃) in specialized tube D (MP Biomedicals). Ultrapure water (50 μL) and LCMS grade methanol (200 μL) were added into each tube. The mixture was vortexed and subsequently homogenized by a FastPrep 24 homogenizer (3 cycles, 6 m/s, 40 s, MP Biomedicals). The mixture was taken out from tube D into Eppendorf tubes and stored at -20 ℃ for at least 1 h for protein precipitation. The supernatant was collected after centrifugation (18,620 g, 5 min, 4 ℃). The extracts from ten patients were combined and direct used in the bead treatment.

Preparation of urine metabolite extracts

Ice cold methanol (200 μL) was added to urine sample aliquots (50 μL) for protein precipitation. Each sample was vigorously shaken for 30 s and then

cooled at 4 ℃ for 30 min. The supernatant was taken out and transferred into empty Eppendorf tubes and the organic solvents were removed from the supernatant through vacuum centrifugation. The residue was directly used in bead treatment.

Preparation of plasma metabolite extracts

Ice cold methanol (200 µL) was added to plasma sample aliquots (50 µL) for protein precipitation. Each sample was vigorously shaken for 30 s and then cooled at 4 ℃ for 30 min. The supernatant was taken out and transferred into empty Eppendorf tubes and the organic solvents were removed from the supernatant through vacuum centrifugation. The residue was directly used in beads treatment.

Preparation of bacterial metabolite extracts

Anaerobic Cultivated Human Intestinal Microflora system (ACHIM) culture was originally obtained from the fresh feces of a healthy Scandinavian donor on an ordinary Western diet. The culture has been re-cultivated every week in anaerobic conditions as described previously (Patent Number: WO 2013/053836A). This ACHIM culture is referred to as "healthy gut microbiota" in this thesis and the 1:10 dilution of the ACHIM culture is represented here as "weak gut microbiota". Salmonella typhimurium SL1344 at 107 or 106 colony forming units (CFU) were co-cultured with the "Healthy" and "Weak" gut microbiota in vitro, respectively. Further antibiotic treatment with 50 µg/mL of Spectinomycin was added to both conditions to evaluate the influence of the antibiotic treatment on the interaction between the gut microbiota and Salmonella. "Healthy" and "Weak" gut microbiota without Salmonella nor antibiotic were used as negative control for the analysis. Salmonella culture without gut microbiota culture is considered the positive control for the CFU comparison.

All tubes were grown anaerobically in a 9 mL culture system. On day 1 and day 5 after co-culture, Salmonella and gut microbial numbers were counted by series dilution and plated on CHROMagar Salmonella agar plates (CHROMagar, Paris France) and modified YCFA plates (recommended from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) supplemented with 2 g/L maltose and 2 g/L cellobiose, all reagents from Sigma-Aldrich, Germany), and grown aerobically and anaerobically, respectively as published before.2

In addition, on day 1 and day 5, 2 mL supernatant without bacteria from all the conditions was collected after filtering the culture mix through the 0.2 µm

PES-membrane filter (Whatman Uniflo, Cytva UK). Afterwards, the supernatant was mixed with 8 ml ice-cold methanol (Sigma-Aldrich, Germany) and kept at -80°C until further *quant*-SCHEMA analysis. 10 µL bacterial extract (2 µL bacterial sample + 8 µL MeOH) was direct used in the bead treatment.

Ethical approval

Patient fecal, plasma and urine samples were obtained in accordance with the World Medical Association Declaration of Helsinki and all patients gave written informed consent. Approval for the study was obtained from the ethical committee at Karolinska University Hospital (Ethical approval number: Dnr 2017/290-31). Fecal, plasma and urine samples were collected using routine clinical collection protocols and all patient codes have been removed in this publication. All samples were stored at −80 °C.

The dietary intervention study was conducted in accordance with the guidelines stated in the current revision of the Declaration of Helsinki, and informed consent was obtained for all subjects. All procedures involving human subjects were approved by King's College London Research Ethics Committee (HR-17/18–5353) and registered at the National Institutes of Health clinicaltrials.gov as NCT03573414.

Chemical metabolomics procedure

Treatment condition for carbonyl-specific probes

A suspension of activated beads in DMF (300 µL) was used to mix with the biological sample extract. One drop of 1M HCl were added into the mixture for protonation of carbonyl group before it was shaken for 16 h at 1,500 rpm and 25 ℃. The biological sample extract solution was removed from the beads via magnetic separation and the beads were washed with THF (2 x 200 µL) before being resuspended in THF (300 µL), ready to be used for bioorthogonal cleavage treatment. Note: for the separate treatment of ^{12}C and ^{13}C probes, different time points of biological samples can be treated correspondingly.

Treatment condition for amine-specific probes

A suspension of activated beads in DMF (300 µL) was used to mix with the biological sample extract in a solution of 1% v/v trimethylamine in ethanol. The mixture was shaken for 16 h at 1500 rpm and 55 °C. The biological sample extract solution was removed from the beads via magnetic separation

and the beads were washed with THF (2 x 200 µL) before being resuspended in THF (300 µL), ready to be used for bioorthogonal cleavage treatment.

Treatment condition for SCFAs-specific probes

A suspension of activated beads in DMF (100 µL) was used to mix with the biological sample extract. A capture solution (4.8 mM HBTU, 3.6 mM HOBT, 1% DIPEA v/v in DCM) was freshly prepared in separate and 100 µL of this solution was added into Eppendorf for SCFAs capture before it was shaken for 16 h at 1,500 rpm and 25 ℃. The remaining biological sample extract solution was removed from the beads via magnetic separation and the beads were washed with THF (2 x 200 µL) before being resuspended in THF (300 µL), ready to be used for bioorthogonal cleavage treatment.

Bioorthogonal cleavage treatment

The suspension of beads was transferred to a glass vial. Triphenylphosphine (97.0 µL, 12.9 mM in THF, 1.25 µmol) and dimethylbarbituric acid (90.0 µL, 30.7 mM in THF, 2.76 µmol) solutions were added to the vial, followed by palladium (II) acetate solution (84.0 µL, 6.53 mM in THF, 549 nmol). The vial was quickly sealed and a stream of nitrogen was passed through until approximately half the volume of the suspension remained. The vial was agitated at intervals on a vortexer and the reaction was allowed to continue 16 h at 25 ℃. In parallel, a sample of unmodified beads was treated with the same cleavage conditions as the activated beads treated with human sample extract and used as control sample. The supernatant was removed from the beads using magnetic separation and the solvent removed using a vacuum centrifuge. The residues were redissolved in MeOH (30 µL each) and triphenylphosphine and triphenylphosphine oxide were precipitated through the addition of water (120 µL each). The suspension was centrifuged (benchtop centrifuge, 12,000 g, 5 min), the supernatant removed, and the solvent was again removed with the vacuum centrifuge. The residues were redissolved in water/acetonitrile solution (95:5 v/v) and submitted for LC-MS analysis.

Preparation of probe-conjugated standards

The probe-conjugated standards can be prepared by either magnetic beads-bound probe treatment or simplified probes (Fmoc-probes) treatment. The magnetic beads bound probe treatment can follow the procedure above by replacing biological sample extracts into standard solution. The Fmoc probe treatment procedure can be found below.

A solution of Fmoc-protected probe (50 µL, 1.0 mM in MeOH) was evaporated under reduced pressure. The residue was combined with DCM (50 µl) and TFA (100 µL). The solution was shaken at 1,500 rpm for 2 h at 25 ℃, before the solvents were removed under reduced pressure. The residue was then combined with a solution of four standards (0.5 equiv. each in 400 µL DMF). The resulting solution was then shaken under the corresponding reaction conditions. The solvents were then removed under reduced pressure, and the residues were treated with piperidine (80 µL) and shaken at 1,500 rpm for 4 h at 25 ℃. The piperidine was then removed under reduced pressure, and the residue was redissolved in MeOH (100 µL) followed by water (400 µL). The solution was diluted as necessary in a solution of water and acetonitrile (95:5 v/v) before being submitted for UPLC-MS analysis.

Data analysis

Data files from the LC-MS analysis were converted into either NetCDF file format using MassLynx 4.1 (Waters) or the MZML file format using MS convert (ProteoWizard). The chromatograms and mass spectra were processed using the XCMS R package for peak alignment[119]. The feature list was reduced by eliminating those features with an m/z value less than 293.1609 (the m/z value corresponding to the monoprotonated probe with the smallest carbonyl compound – formaldehyde). Features with intensities > 30,000 ion count, retention time > 1.5 min and %CV of the QCs < 30 were selected for further statistical analysis as considerably higher than noise. Mass values of each feature with 263.1503 subtracted (corresponding to the mass of the tagged probe) were compared to the human metabolome database in order to find plausible candidates for the parent metabolites with carbonyl functional group. Then the filtered dataset was normalized based on the QCs using SERRF to eliminate the technical error. (Fan et al. 2019) Commercial or synthetic standards were then used to confirm the identity of the metabolites and identification of the correct regioisomers. An overview of the data was provided by principal component analysis (PCA) and heatmap, prior to which the data was auto-scaled using the metabolomics platform www.metaboanalyst.ca. The normality of the test statistics, p values and ROC analysis were evaluated using the same platform, and the data were distributed normally. The combinatory variables used for the calculation of AUC in ROC analysis were obtained using binary logistic regression in SPSS.

Method development

LOD and LOQ measurement.

Synthetic conjugated standards were prepared in a solution of water and acetonitrile (95:5 v/v) at a range of concentrations (1 μM, 100 nM, 50 nM, 25 nM, 10 nM, 5 nM, 1 nM, 0.5 nM, 0.3 nM, 0.1 nM) before being submitted for UPLC-MS analysis. More specific concentrations can be prepared if it is needed. The measurement was based on the conventional concepts that signal to noise ratios at least higher than 3 and 10 corresponding to Limit of Detection (LOD) and Limit of Quantification (LOQ), respectively. The signal to noise ratios were calculated according to European Pharmacopoeia guidelines by using MassLynx 4.1(Waters) or using DataAnalysis 5.0 (Bruker).

Bead loading capacity experiment

In five independent experiments, 50 μL beads was utilized to conjugate the general probes **18** (1.67 mg, 3.0 μmol, 10 equiv.) following the same procedure as above. Then synthetic $^{13}C_6$ labelling cleavage product **$^{13}C_6$-30** (5.0 μL in 30 mM) were spiked into these reaction mixtures after these bead-bound simplified probes were independently cleavage under the same bioorthogonal cleavage condition. These five independent cleaved samples were prepared into 50 μL of water and acetonitrile (95:5 v/v) before being submitted for UPLC-MS analysis.

Chemical metabolomics procedure optimization

To optimize the chemoselective probe procedure for large-scale analysis, its cleavage behavior in different vessel types was compared to the previously described standard procedure. The cleavage reagents were added to the probe solution (30 μM in MeOH) in the different vessel types (Standard vials, Eppendorf tubes, Eppendorf tubes with screw cap). 0.5 μL of a ^{13}C-labelled internal standard was added into each vessel before the metabolite release treatment as described above. In addition, cleavage was performed in standard 1.5 mL Eppendorf tubes in ambient air as a negative reference. The experiment was conducted in triplicates for each vessel type. After UHPLC-MS analysis, the data was analyzed for the peak areas of the cleaved probe product to perform the comparison.

Results and discussion

This thesis primarily focuses on the development of advanced chemical metabolomics tools for qualitative and quantitative analysis of metabolites in human and bacterial samples. The ultimate goal is to discover potential biomarkers that reflect various biological states. Additionally, the thesis also explores the development of novel chemical modification reactions within the field of metabolomics. We began with the qualitative detection of carbonyl and amine-containing metabolites for the discovery of previously undetectable metabolites[64, 120, 121]. This was achieved through either the adaption of more advanced moieties or the development of novel chemical modification moieties for improving reactivity and stability. Our approach involved the construction of conjugated libraries, which enabled the rapid validation of the chemical structures of detected carbonyl-containing metabolites (as discussed in Paper I). Notably, utilization of our method validated three to four times more metabolites compared to other published techniques for carbonyl detection. Furthermore, we introduced a novel chemical modification moiety known as squaramides, designed to selectively conjugate with amine-containing metabolites. The advantageous chemical properties of squaramides, including stability and reactivity, significantly enhanced the detection of amine-containing metabolites, as detailed in Paper II. Moreover, we have expanded the application of chemoselective probe from qualitative detection to quantitative analysis. We developed a chemoselective quantitative analysis, termed *quant*-SCHEMA, utilizing isotopically labelling technology (as presented in Paper III)[69]. The application of *quant*-SCHEMA led to the discovery of a list of previously undetectable metabolites. The detection of the pairs of ^{12}C and ^{13}C with specific 6.0201 Da difference in mass spectrometric analysis enhances the confidence level of metabolite validation. Expanding the utility of *quant*-SCHEMA, we designed a chemical reactive moiety tailored for short-chain fatty acids, as outlined in Paper IV[122]. This method exhibited remarkable sensitivity, requiring only 2% of the sample volume for quantifying SCFAs, leaving the majority of the sample volume available for conventional metabolomics methods without any loss. In the end of the thesis, our novel methodology was successfully applied to the discovery of dietary biomarkers in an upscaled analysis (as detailed in Paper V)[123]. Four carbonyl-containing metabolites demonstrated high sensitivity and specificity

as biomarkers in both training and validation datasets. Further discussion on these findings will follow in subsequent sections.

Chemical metabolomics for qualitative analysis

Metabolic ketones and aldehydes

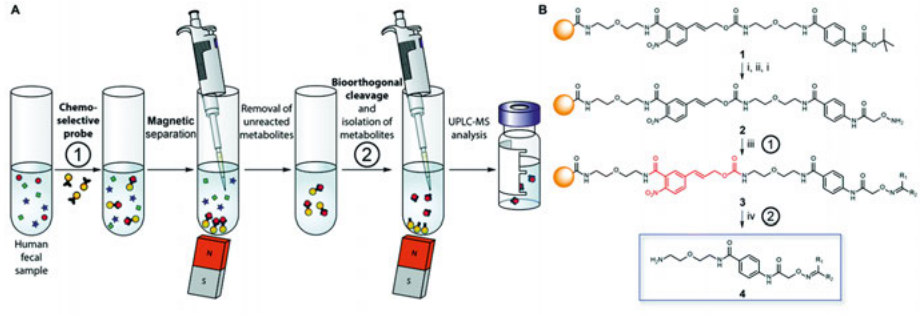


Figure 6 Chemoselective probe for carbonyl-containing metabolite analysis. (A) Schematic overview of our methodology using chemoselective probes immobilized to magnetic beads for metabolite conjugation. (B) Chemoselective probe chemical workflow.

In my first leading publication during my PhD studies, we have further utilized the chemoselective reactive site alkoxyamine to capture carbonyl-containing metabolites present in fecal and urine samples from the pancreatic cancer patients. This research is based on a previously reported research communication[120] led by Dr. Louis Conway, which has developed the first application of alkoxyamine as a reactive site for capturing carbonyl-containing metabolites in fecal sample. This advanced probe is immobilized onto magnetic beads and enables facile probe activation for simple sample treatment and subsequent metabolite extraction. The key moiety of the probe is the bioorthogonal cleavage site *p*-nitrocinnamyloxy-carbonyl (Noc)[124], which can be efficiently cleaved under mild bioorthogonal conditions using palladium (0) catalysis.

Carbonyl-containing compounds are reactive metabolites that can be produced by a range of bacteria commonly found in the human gut. However, due to their poor ionizability, it is difficult to analyse the importance of these compounds in the human gut using mass spectrometric-based analysis. Our study design is based on the magnetic bead-immobilized probe **1**. After Boc-deprotection, the straightforward activation yields the immobilized and Boc-protected alkoxyamine **2** that is used to treat human samples to conjugate metabolic ketones and aldehydes **3**. Under mild bioorthogonal cleavage

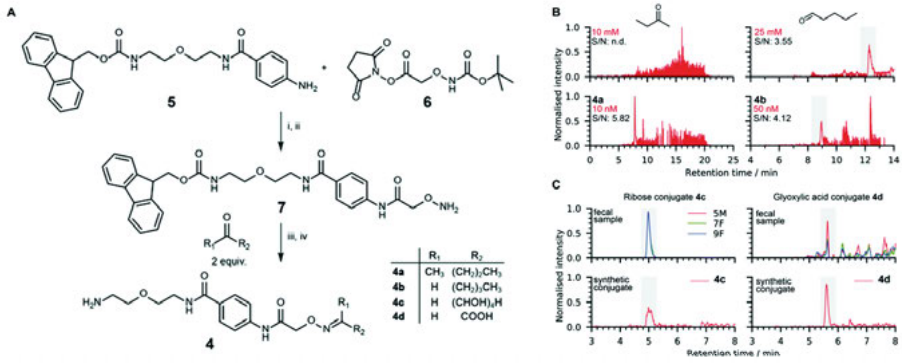


Figure 7 (A) Reference conjugate synthesis for metabolite validation. (B) LOD measurement. (C) Metabolite validation.

conditions, the captured carbonyl-containing metabolites were released as the general structure **4**, ready for analysis using UPLC-MS (Figure 6).

To validate our developed method, we synthesized a simplified probe **7** synthesized from aniline **5** and NHS-activated carboxylic acid **6** under peptide coupling conditions. We optimized probe activation conditions and confirmed the stability of the conjugate under treatment conditions. The synthetic metabolite conjugates **4** were further used in metabolite validation through co-injection experiments and to measure the ionizability of these metabolite class. The limit of detection (LOD) for the butanone conjugate 4a was determined to be less than 10 nM in positive ionization mode, which corresponds to 50 fmol. However, no clear peak was observed for unconjugated butanone even at 10 mM demonstrating signal enhancement by a factor of at least one million (Figure 7).

One major challenge in metabolomics analysis is the validation of the exact chemical structures of regioisomers. Metabolites catalogued in databases such as HMDB[22], Sirius[125], MZmine[126], and Metlin[127] can be identified through MS/MS fragmentation comparisons. Another method for compound validation with higher confidence level is the co-injection of a synthetic or commercially available reference compound with the natural metabolite in the sample to compare chromatographic and mass spectrometric properties and unambiguously distinguish regioisomers. Consequently, we have developed a straightforward synthetic approach to prepare reference compounds for the validation of the chemical structures of metabolites, allowing us to synthesize these references within 24 hours (Figure 8). We prepared a library of 94 carbonyl-containing metabolite conjugates for structure validation in a pooled fecal sample for maximizing metabolite coverage. This metabolite library is suitable for large-scale carbonyl screening in any type of biological sample with enhanced mass spectrometric sensitivity. In this study, after we have captured carbonyl-containing metabolites using our chemoselective probe, we

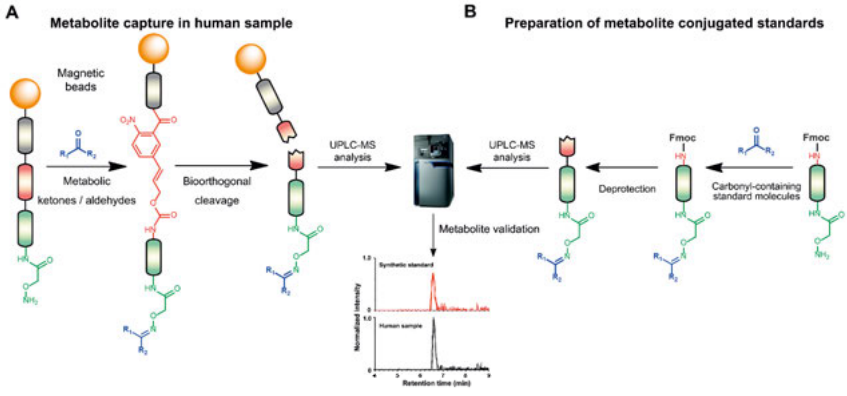
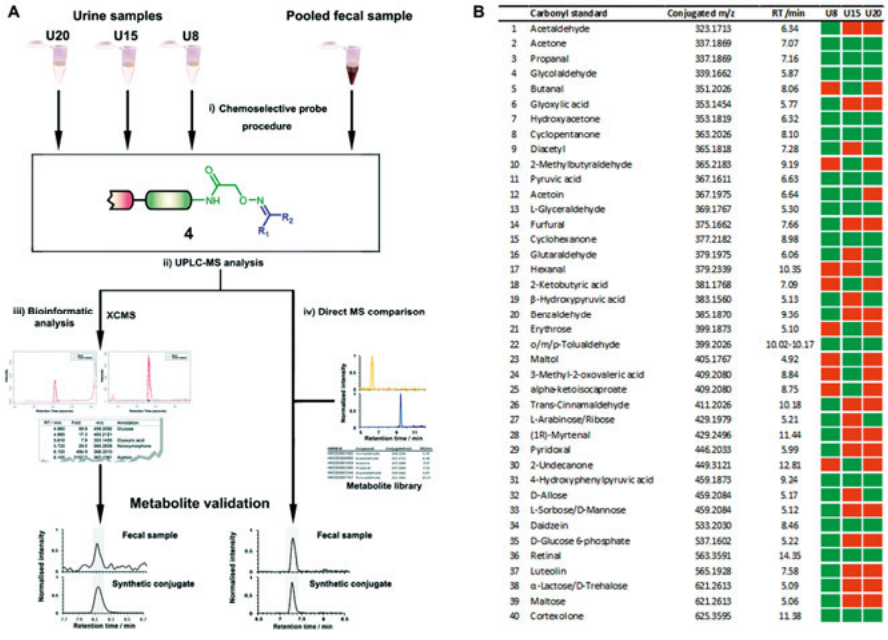


Figure 8 Chemoselective probe procedure and metabolite validation. (A) Procedure for the capture and extraction of carbonyl-containing metabolites from human samples. (B) Preparation of metabolite-conjugate standards (3) Validation of captured metabolites using

have firstly performed the standard procedure using XCMS to annotate tentative metabolite structures and secondly directly compared the captured metabolites to our newly built in-house carbonyl conjugates reference library to rapidly obtain the carbonyl-containing metabolite profile (Figure 9A).

Urinary carbonyl-containing metabolites have not yet been analysed in the context of pancreatic cancer. As carbonyl-containing compounds are highly reactive and can interact with other biomolecules, leading to the degradation of regulatory proteins, enzymes and DNA modifications, we applied our method to the analysis of patient urine samples. Following the identical procedure for sample treatment and bioinformatics processing, we identified a total of 99 ketones and aldehydes in the urine samples. By comparing the carbonyl-containing metabolite UPLC-MS information with our newly constructed in-house reference, we validated 40 metabolites present in at least one of these three samples through co-injection experiments with these synthetic conjugate-standards. Among these compounds, 35 had been reported previously in human samples according to HMDB. The high sensitivity of our methods led to the discovery and confirmation of five metabolites glycolaldehyde, cyclopentanone, glutaraldehyde, hexanal, and myrtenal, which have not yet been reported in this human sample type demonstrating the versatility of our method (Figure 9B).

The versatility of this metabolite library was demonstrated by investigating a fecal sample pooled from eight different patients to discover metabolic ketones and aldehydes derived from microbiota metabolism. After LC-MS based analysis of the captured and released metabolite conjugates, we compared the obtained features with the recorded conjugate library data. The comparison of retention times and *m/z* values between the metabolite library



	Carbonyl standard	Conjugated m/z	RT /min	U8	U15	U20
1	Acetaldehyde	323.1713	6.34			
2	Acetone	337.1869	7.07			
3	Propanal	337.1869	7.16			
4	Glycolaldehyde	339.1662	5.87			
5	Butanal	351.2026	8.06			
6	Glyoxylic acid	353.1454	5.77			
7	Hydroxyacetone	353.1819	6.32			
8	Cyclopentanone	363.2026	8.10			
9	Diacetyl	365.1818	7.28			
10	2-Methylbutyraldehyde	365.2183	9.19			
11	Pyruvic acid	367.1611	6.63			
12	Acetoin	367.1975	6.64			
13	L-Glyceraldehyde	369.1767	5.30			
14	Furfural	375.1662	7.66			
15	Cyclohexanone	377.2182	8.98			
16	Glutaraldehyde	379.1975	6.06			
17	Hexanal	379.2339	10.35			
18	2-Ketobutyric acid	381.1768	7.09			
19	β-Hydroxypyruvic acid	383.1560	5.13			
20	Benzaldehyde	385.1870	9.36			
21	Erythrose	399.1873	5.10			
22	o/m/p-Tolualdehyde	399.2026	10.02-10.17			
23	Maltol	405.1767	4.92			
24	3-Methyl-2-oxovaleric acid	409.2080	8.84			
25	alpha-ketoisocaproate	409.2080	8.75			
26	Trans-Cinnamaldehyde	411.2026	10.18			
27	L-Arabinose/Ribose	429.1979	5.21			
28	(1R)-Myrtenal	429.2496	11.44			
29	Pyridoxal	446.2033	5.99			
30	2-Undecanone	449.3121	12.81			
31	4-Hydroxyphenylpyruvic acid	459.1873	9.24			
32	D-Allose	459.2084	5.17			
33	L-Sorbose/D-Mannose	459.2084	5.12			
34	Daidzein	533.2030	8.46			
35	D-Glucose 6-phosphate	537.1602	5.22			
36	Retinal	563.3591	14.35			
37	Luteolin	565.1928	7.58			
38	α-Lactose/D-Trehalose	621.2613	5.09			
39	Maltose	621.2613	5.06			
40	Cortexolone	625.3595	11.38			

Figure 9 (A) Overview of the analysis of human samples using the chemoselective probe methodology using either a bioinformatic workflow (left) or comparison with a metabolite library. A pooled fecal sample constituted from eight samples and three individual urines. (B) Validated carbonyl-containing metabolites in human urine samples sorted by molecular weight.

and the dataset led to the validation of 33 carbonyl-containing metabolites, which is approximately twice as many as validated in individual samples. Our procedure also identified five metabolites, which until now have not been detected in human fecal samples (glycolaldehyde, lactaldehyde, glutaraldehyde, myrtenal, and cyclopentanone).

In summary, our advanced chemical metabolomics method was successfully applied for the first time to the analysis of carbonyl compounds in urine samples leading to the discovery of five previously undetected metabolites. We also report the construction of a metabolite conjugation library of 94 ketone- or aldehyde-containing metabolites for unambiguous structure validation and demonstrate its utility by applying it to the analysis of a pooled fecal sample. In this sample, we discovered five metabolites previously unknown to be analysed in feces. These results highlight that our bipartite approach using our chemoselective probe combined with either sophisticated bioinformatic analysis or a unique conjugation library allows for the discovery of unknown as well as the validation of known metabolites in complex human samples. This study lays the foundation for selective metabolomics analysis of carbonyl-containing compounds at low concentrations in any sample type and the discovery of biomarkers of disease.

Metabolic amines

In Paper II, we introduced a novel chemoselective moiety, squaric acid, for the first analysis of metabolic amines. Amines play crucial roles in biological processes, with this class of metabolites containing the largest percentage of gut microbiota-derived compounds. Specifically, amino acids, short peptides, and neurotransmitters are vital for human physiology[128]. However, it is estimated that many metabolites in human fecal samples remain undiscovered, along with their chemical structures. Identifying this metabolite class in fecal samples is essential for gaining deeper insights into gut microbiome metabolism and its association with human health[128]. Fecal samples are rich in gut microbiota-derived metabolites. Chemical modification is a useful technique for enriching metabolites and enhancing their detection. *N*-hydroxysuccinimide (NHS) and sulfo-*N*-hydroxysuccinimide (sulfo-NHS) are commonly used moieties for labelling amines. However, the stability of these moieties poses a critical challenge, limiting their use in amine-specific chemoselective reactions[129, 130].

We sought to activate our chemoselective probe with a new and more stable moiety for the analysis of amines. 3,4-Diamino analogues of squaric acid, known as squaramides, are utilized for selective amine conjugation in various research fields, including materials science, medicinal chemistry, and

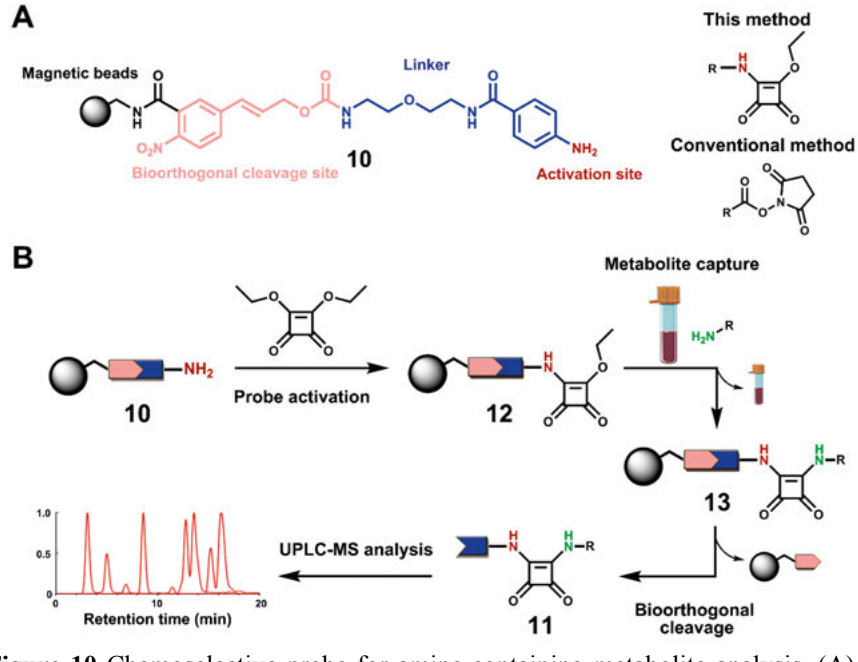


Figure 10 Chemoselective probe for amine-containing metabolite analysis. (A) Probe design: magnetic beads (black), bioorthogonal cleavage site (pink), linker (blue), and activation site (red). (B) Chemoselective probe treatment workflow that includes a magnetic separation step from the sample matrix after incubation.

sensors[131]. Squaric acids are also used for bioconjugation of macromolecules, specific cell labelling, and bioisosteric replacement due to their high selectivity for reacting with amines under slightly basic conditions. However, this chemoselective functional group has not been widely reported in metabolomics-based investigations. Due to the favourable properties of the squaric acid moiety, we activated our recently developed chemoselective probe with this moiety for the analysis of amine-containing metabolites in human samples (Figure 10). Before applying this new methodology to human samples, we first tested the properties of probe activation, conjugation chemistry, and the stability of the released metabolite conjugates using a synthesized simplified probe **8**. We optimized the conditions for the synthesis of **8** from intermediate **9** with squaric acid diethyl ester to activate

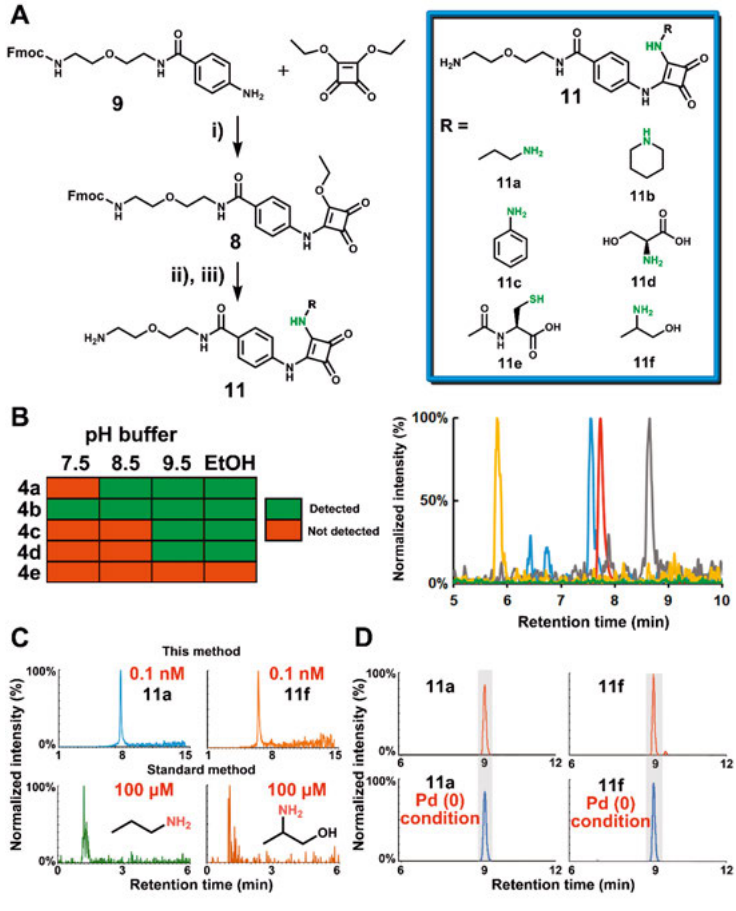


Figure 11 Method validation. (A) Synthetic scheme for conjugated metabolites and six representative examples. (B) Reactivity condition analysis for five metabolites. Colour-code green: detected; red: not detected. (C) Extracted ion chromatograms (EICs) for LOD comparison of **11a** and 1-aminopropane as well as **11f** and 2-amino-1-propanol. (D) Stability of **11a** and **11f** for treatment with $Pd(OAc)_2$, PPh_3, dimethylbarbituric acid.

immobilized probe **10**. Monosquaramide **8** was then tested for reactivity and chemoselectivity with five representative metabolites that cover diverse amine functionalities commonly found in metabolites. The reactions of **8** with 1-aminopropane (a primary amine), piperidine (a secondary amine), aniline (an aromatic amine), *L*-serine (an amino acid), and *N*-acetyl-1-cysteine (a thiol) were monitored under different reaction conditions, including pH values of 7.5, 8.5, and 9.5, as well as 1% trimethylamine in ethanol. The results were analysed via UPLC–MS analysis and confirmed that basic conditions or the addition of an organic base increased reactivity. Importantly, this moiety only reacted with amine-containing molecules and not with the thiol nucleophile *N*-acetyl-*L*-cysteine, demonstrating its chemoselectivity for amines under the tested conditions (Figure 11). Based on these successful results, we synthesized metabolite conjugates **11a–e** to determine the limit of detection (LOD). The lowest LOD value was determined to be 80 pM for the conjugate of 1-aminopropane **11a** in positive MS ionization mode, which corresponds to 400 amol. The LOD for unconjugated 1-aminopropane was determined to be 100 μM, highlighting a mass spectrometric sensitivity increase by a factor of 1.25 million for **11a**. The LOD for the other tested metabolites ranged from 100 pM to 1.0 μM. Next, we determined the stability of squaramides of the general structure **11** under the cleavage conditions involving the bioorthogonal moiety Noc. The stability under cleavage conditions, using $Pd(OAc)_2$, PPh_3, and dimethylbarbituric acid, was analysed to eliminate the possibility of undesired decomposition of the captured metabolites in human samples. All conjugates remained unaltered even after 16 hours of treatment, confirming the stability of 4 throughout our entire sample preparation procedure (Figure 11C).

Upon method optimization, we treated human fecal samples collected from three different pancreatic cancer patients with probe **12** (Scheme 1). The treatment of fecal samples with **12** was performed according to our previously reported procedures with minor modification. The outcome from the bioinformatic analysis identified more than 3000 significantly altered features after removal of features following these criteria: (i) m/z values smaller than the ammonium conjugate (<319.1401 Da), (ii) less abundant in the metabolite extract compared to the control sample, (iii) not significantly altered features (p-values > 0.05), (iv) features that eluted before 1.5 min. The filtering process was followed by the comparison of the m/z values from the candidate features with HMDB to determine their tentative chemical structures.

This study represents the first application of squaric acid for chemoselective analysis of amine metabolites in metabolomics-based studies. The high stability of the formed conjugates of squaramide with metabolites and the enhanced ionization properties of these conjugates led to the successful capture and analysis of amine-containing metabolites in human fecal samples.

The highly improved LOD at pM concentrations resulted in the detection of three previously undetected metabolites in this sample type, as well as a series of microbiota-derived metabolites. This analysis resulted in 165 metabolites annotated with chemical structures with a 10-ppm accuracy. Using our constructed in-house conjugated metabolite library, we confirmed 39 metabolites at the highest confidence level through UPLC–MS validation experiments. Furthermore, we detected 12 microbiota-derived metabolites in these fecal samples, providing direct insights into gut microbiota metabolism and its interaction with the human host. This method can now be utilized for monitoring amine metabolites in any biological sample.

Chemical metabolomics for qualitative analysis

Relative quantification (*quant*-SCHEMA)

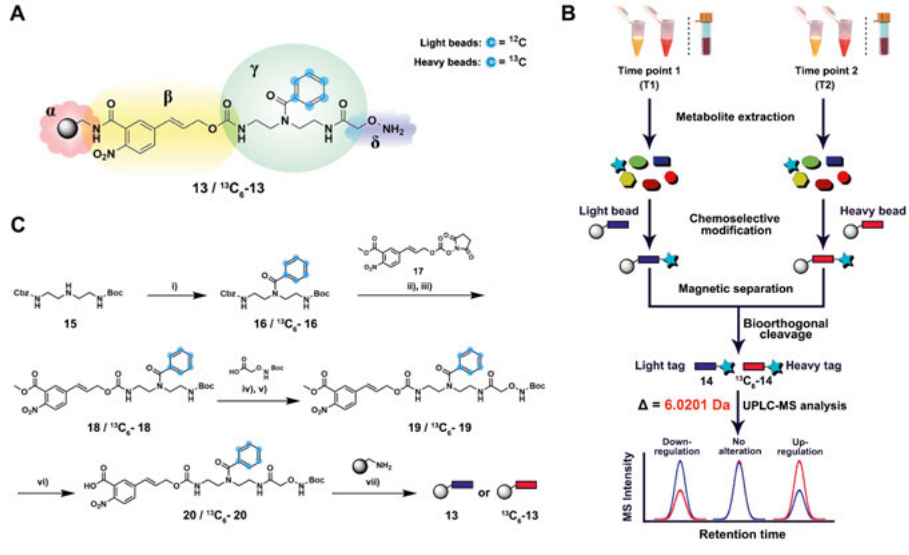


Figure 12 Chemoselective probe for quantitative analysis of metabolites in human samples. (A) Chemical probe design. (B) Workflow for *quant*-SCHEMA. (C) Simplified synthetic scheme.

In paper III, we have described an innovative methodology to quantify gut microbiome and human metabolites with high sensitivity. As all our previous publication were focused on the qualitative analysis to identify and discover metabolites, this second-generation chemical probe method allows for direct quantitative analysis of carbonyl-metabolites. The high sensitivity for detailed analysis of the importance of those microbial metabolites and to understand how they involved in the disease development.

The novel quantitative chemical probes have been redesigned to enable a new chemical biology methodology for advanced mass spectrometric analysis that we have termed Quantitative Sensitive CHEmoselective MetAbolomics (*quant*-SCHEMA). The synthesis of $^{12}C/^{13}C$ isotopically labelled analogues of the probe and separate sample treatment allows for comparative and quantitative analysis of endogenous and microbiota-derived metabolites in human samples at low concentrations (Scheme 2). While isotope-labelling is a well-established part of the (chemo)proteomic workflow, it is rarely used in metabolomics due to the higher structural diversity of metabolites and concentration differences. The few reports are based on simple chemistry without separation from the sample matrix, require harsh conditions for coupling, are limited in the validation of metabolites, or have only been applied to a single sample type. This second generation of chemoselective probes described in this study now facilitates quantitative analysis and overcome the limitations that other method has to encounter. The redesign of the probe scaffold that only contains the bioorthogonal cleavage site from the first-generation probe and now provides several advantages. i) quantitative and comparative analysis between two sample sets by use of a light and a heavy labelled probe; ii) a new coupling chemistry to amino-activated magnetic beads facilitates a shorter synthetic route; iii) the reactive site is introduced to a primary amine for more efficient activation; iv) precise quantification of selected metabolites at low fmol concentrations without the need of commercially available isotope labelled compounds; v) treatment of three different human sample types with the identical procedure; and vi) biomarker discovery using a combination of isotope-labelled chemoselective probe methodology and bioinformatic metabolomics analysis (Figure 12A).

The treatment of this new generation probe is different from the first-generation probe, even though the magnetic separation and bioorthogonal cleavage remain. In brief, this new quantitative methodology is based on chemoselective probe **13** that is immobilized to magnetic beads and a corresponding stable isotopically labelled analogue **$^{13}C_6$-13**. After separate incubation (light probe for sample collected from time point 1 and heavy probe for sample collected from time point 2) for 16 h to capture carbonyl-containing metabolites, the beads were isolated from each sample using a magnet, washed, and combined for bioorthogonal cleavage. The resulting tagged metabolites of the general structure **14/$^{13}C_6$-14** were analysed using UPLC-MS. The obtained data was analysed using the bioinformatic software package XCMS in R. Both conjugated metabolites **14** and **$^{13}C_6$-14** can be distinguished by their mass difference of 6.0201 Da without interference from the natural isotope distribution, while both elute at the same retention time due to the same physicochemical properties. These advantages allow us to easily identify the desired metabolites when the pairs of the signals with 6.0201 Da difference are observed at the same retention time (Figure 12B).

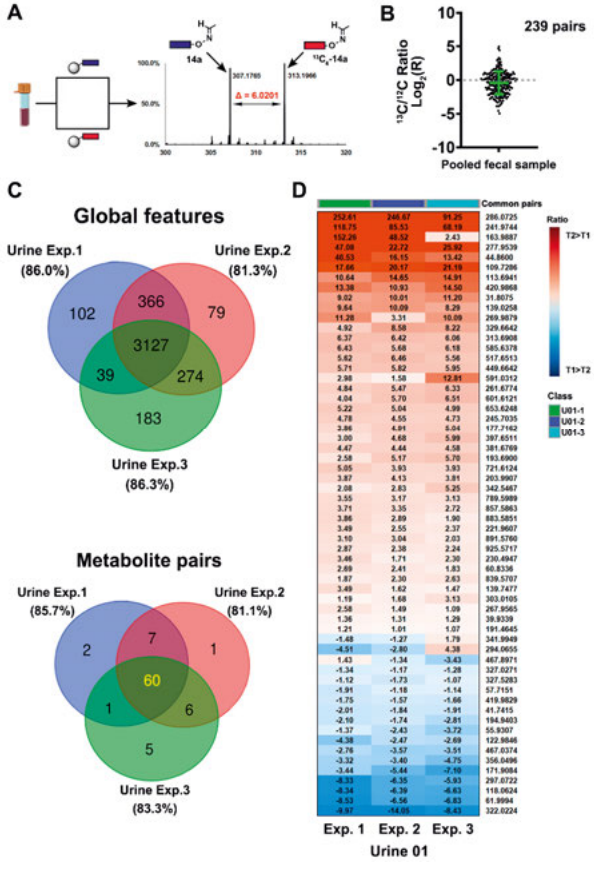


Figure 13 Method validation. The same biological sample has been treated 3 times to determine the reproducibility of this method.

Before utilization of this method in human samples, we have conducted a series of experiments to validate the applicability of this new method. We initially determined the bead load capacity for the coupling of the chemoselective probe that was applied for all following experiments at 0.8±0.1 µmol*g^{-1}. An additional analysis of equimolar solutions for each metabolite as well as 1:3 and 3:1 ratios demonstrates the quality of this methodology. We have also determined the recovery rate for three selected carbonyl compounds that are representative for different relativities of carbonyls in human samples: octanal as an alkylaldehyde, 2,3-pentanedione as an alkylketone and phenylacetaldehyde as an aromatic aldehyde that is sterically hindered. Comparison of an aqueous solution of these compounds (c=100 µM each) to human urine samples with metabolites spiked in at the same concentration resulted in a recovery rate of 49–67 % with an average RSD of 8.2 % demonstrating the highly reproducibility of this methods. Our method was designed for straightforward identification of the captured carbonyl-containing metabolites via bioinformatic analysis using XCMS through the difference of Δ=6.0201 Da in the MS chromatogram. The adapted XCMS can help us to collect the pairs with these two criteria: i) difference of 6.0201 Da; ii) within retention time <0.05 min. The efficiency and reproducibility of the novel method were tested using a pooled fecal sample from ten patients to maximize the number of carbonyl-containing metabolites. The sample was split in two equal parts and each was treated either with the light (**13**) or with the heavy ^{13}C-labeled chemoselective probe

(**$^{13}C_6$-13**), the conjugates were cleaved from the magnetic beads after magnetic separation and analysed using UPLC-MS analysis. This experiment led to the identification of 239 metabolite pairs representing the identification of 239 carbonyl-containing metabolites with higher level of the confidence than the conventional *m/z* value-based metabolite assignments. The ratios of these detected 239 pairs were mainly clustered in the expected $Log_2 = 0$ demonstrating the precision of this method as it accounts for differences between both synthetic analogues, coupling efficiency to the magnetic beads, metabolite conjugation, bioorthogonal cleavage, and UPLC-MS analysis (Figure 13).

This analysis was followed by an additional reproducibility test in urine samples from one patient but collected 6 months apart. Samples T1 and T2 were split in three parts each and each aliquot of T1 was treated with probe **13**, while each aliquot of T2 was treated with probe **$^{13}C_6$-13**. In this reproducibility experiment, 60 out of 82 pairs of labelled and unlabelled tagged metabolites were commonly detected in all independently replicas, which were identified out of 3,127 common features. As we expected, these 60 metabolite pairs were found to be up- and down-regulated in the same direction in repeated analyses (Figure 13). The LOD and LOQ of this compound class were also determined. The acetone-conjugate **14b** can be detected at a concentration of 100 pM, which corresponds to an LOD of 500 amol. Furthermore, the LOQ for conjugates of acetone (**14b**) and butanone (**14c**) using this new method is c=0.5-5nM (Scheme 3).

For this study, we collected urine and plasma samples from the same high-risk patients for pancreatic cancer at two different time points T1 and T2. We have applied our developed *quant*-SCHEMA method for direct comparative analysis of both time points for each of the three individuals. The total number of detected carbonyls in these samples, confirmed by the pairs of the light and heavy tags, are 180 metabolites for plasma and 203 for urine samples. This reliable extraction and isolation of large numbers of carbonyl metabolites enables the large-scale analysis of this reactive compound class in human samples for biomarker discovery and is enhanced to previous studies due to the improved MS sensitivity and metabolite identification procedure. The treated urine and plasma samples were analysed in parallel with the same control sample in a randomized sequence allowing for comparison of the samples through principal components analysis. This unsupervised multivariate analysis clearly separated plasma and urine samples, but also clearly grouped samples according to collection time, while replicates remained tightly clustered. Importantly, our method permits identification of metabolic variations in individuals over time that can in the future be utilized to identify disease onset. This has so far been unavailable for broad coverage of carbonyl-metabolites in clinical human samples.

As our *quant*-SCHEMA analysis is based on isotope-labelled metabolite tags, it allows not only for the identification of metabolites but also their precise quantification in complex biological matrices. Additionally, our method provides a new straightforward possibility to synthesize $^{13}C_6$-labelled standards for the precise quantification of a metabolite of interest. This is advantageous as not all metabolites are commercially available as stable isotope labelled analogues. We synthesized light and heavy labelled standards of the acetone conjugate **14 b** and **$^{13}C_6$-14** b to measure a calibration curve enabling quantification of acetone in urine and fecal samples (Figure 14).

Our sensitive mass spectrometric analysis also led to the identification of previously undetected metabolites in all three investigated sample types. Due to removal of the captured metabolites from the complex sample matrix using magnetic separation, analysis in parallel of metabolites present in all three human sample types is possible. The retention time and peak shape for our analysis is identical for every sample type. This is an important advantage as the magnetic separation removes the common matrix effect that other

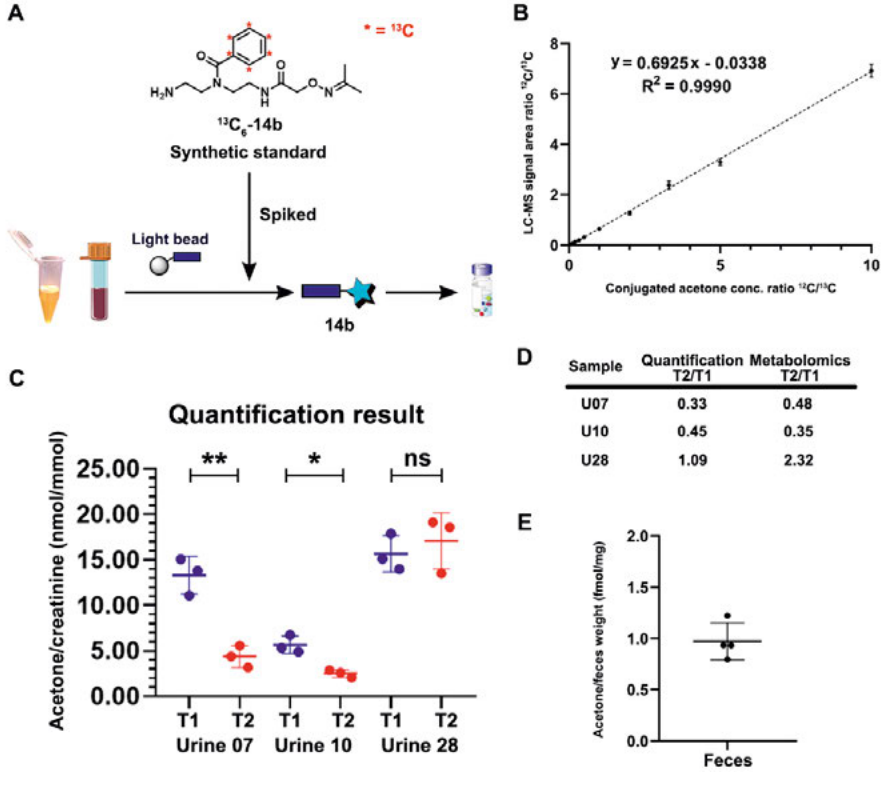


Figure 14 Quantification of acetone. (A) Workflow. (B) Calibration curve. (C) Quantification results in urine sample. (D) Comparison between metabolomics output and quantification output. (E) Quantification results in feces.

derivatization methods and standard metabolomics studies need to consider through additional normalization procedures.

In summary, the developed new method represents a reproducible chemical biology tool to enhance metabolomics and metabolite analysis in complex human samples. This new methodology is based on a chemoselective probe immobilized to magnetic beads with a light and heavy isotope tag that allows for advanced analysis to compare with other derivatization methods for metabolite analysis, which lack separation from the sample matrix and must overcome matrix interferences. Due to the combination of natural and isotope-labelled conjugates before LC-MS analysis, this new method does not require any normalization procedure including internal standards or normalization of different sample types. This is necessary in general metabolomics analysis to reduce technical errors. This comprehensive analysis allows for enhanced mass spectrometric approaches to biomarker discovery and the possibility to monitor patients and their disease progression at different timepoints is a new tool towards personalized medicine and diagnostics.

Precise quantification – Short chain fatty acids (SCFAs)

In Paper IV, we continued to develop *quant*-SCHEMA to expand its applications. We noticed the high sensitivity, reproducibility of this method, and the importance of short-chain fatty acids (SCFAs)[79, 132]. This led us to the idea of developing a precise quantification method for SCFAs with minimal sample cost. The enhanced sensitivity not only improves the detection limit but also allows for a reduction in sample volume usage. What if we could accurately measure SCFAs using only 2% of the sample volume, while the remaining sample could be used for conventional metabolomics processes? To achieve this, we conducted a series of experiments to further refine our *quant*-SCHEMA (Figure 15, Scheme 4). Baed on the research from Paper III, we incorporated our chemoselective probes with *N*-methylbenzylamine as a new carboxylic acid chemoselective reactive moiety.

The SCFA conjugates and their $^{13}C_6$-labelled SCFAs conjugates for the calibration curve were synthesized from a simplified probe. We coupled synthetic intermediate **21** with 4-(Boc-aminomethyl)benzoic acid **22** using common amide coupling reagents HBTU, HOBT, and DIPEA (Scheme 5). The simplified probe was then treated under acidic conditions to remove the Boc protecting group for activation. The activated probe **23** was coupled with formic acid (FA), acetic acid (AA), propanoic acid (PA), butyric acid (BA), valeric acid (VA), and lactic acid (LA) using amide coupling conditions to obtain the corresponding SCFAs conjugates **24a-f** and their respective $^{13}C_6$-

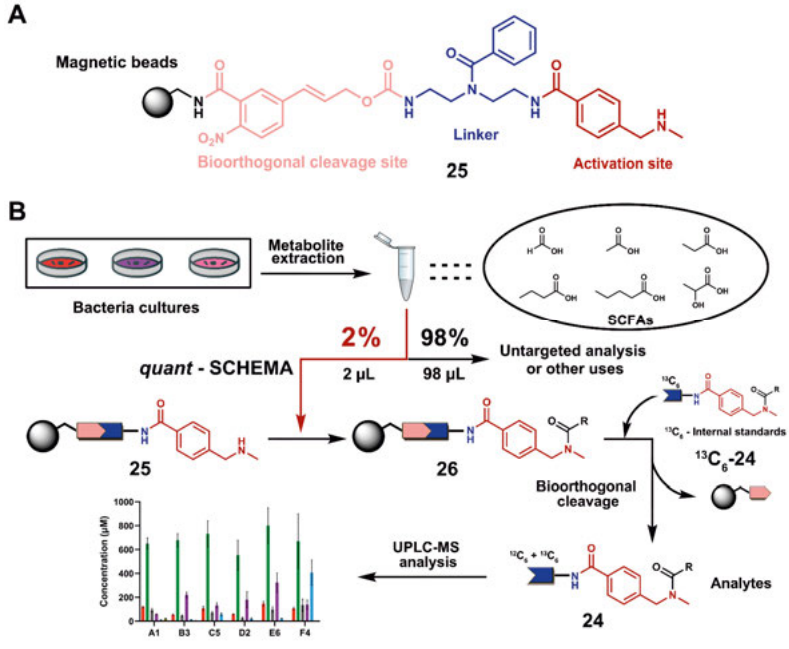


Figure 15 General overview of this methodology. (A) Chemoselective probe's chemical structure. (B) General workflow in this chemical probe utilization for SCFAs quantification.

labelled conjugates **24a-f*** (see Figure 16A). These optimized conditions were also applied for the immobilized probe experiments in biological samples. After synthesizing these SCFA conjugates, we prepared a series of concentration solutions to determine the six LOD/LOQ values and the six calibration curves. Serial dilutions were prepared for each SCFA conjugate to calculate the signal-to-noise ratios (see Figure 16B). The rough LOQ for these analytes was determined to be either 10 nM (50 femtomoles for FA, AA, and LA) or 1 nM (5 femtomoles for PA, BA, and VA). The calibration curves were prepared through three independent experimental procedures, demonstrating near-perfect linearity. The details of LOD/LOQ were also determined (R^2 > 0.9994, see Figs. 16C). Additionally, a linear concentration range was tested from 1 nM to 1000 nM. Using these calibration curves, we validated the method by quantifying an identical bacterial sample six times to assess its reproducibility. The relative standard deviations (RSD) ranged between 7.1% and 17.7%. These highly reproducible results (average RSD of 12.40%) demonstrate the robustness of this new method (see Figure 17A). We also conducted an experiment to determine the matrix effect by quantifying the same sample four times, with sample matrix and without sample matrix. The matrix effect led to a reduction in the detection signal from 68.3% to 83.9%.

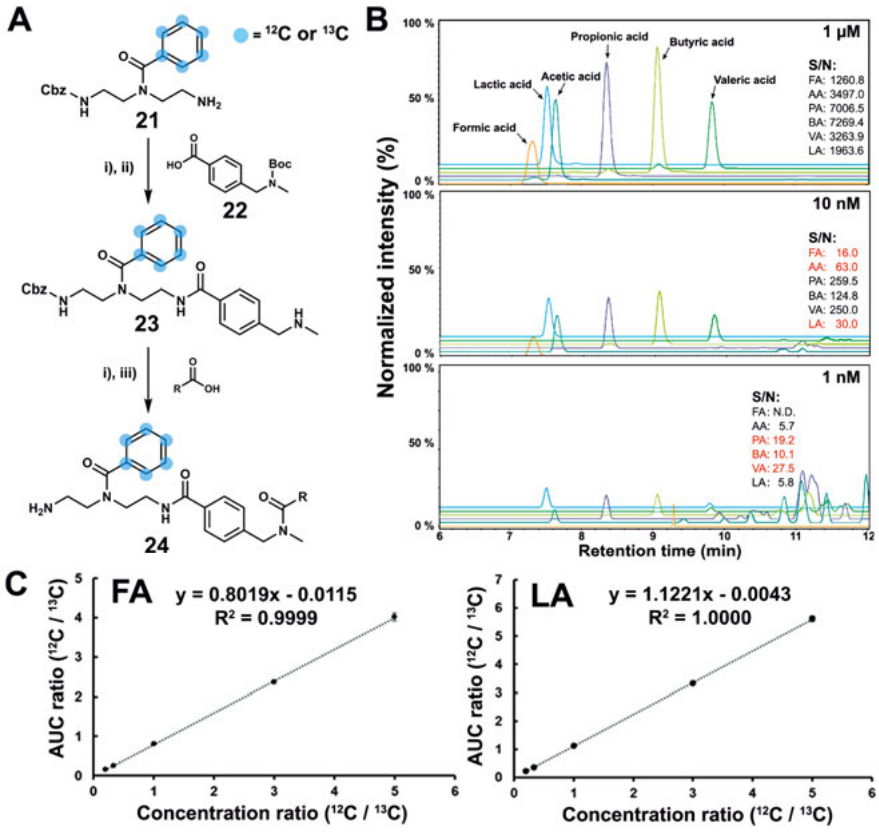


Figure 16 Method validation. (A) Synthetic scheme for SCFA conjugates; i) HBTU (1.6 eqv.), HOBT (1.2 eqv.), DIPEA (3 eqv.), DCM, 16 h, r.t.; ii) TFA/DCM = 1:1, 2 h, r.t.; iii) Piperidine in DMF, 5 h, r.t.(B) LOQ experiment in 1 µM, 10 nM and 1 nM. The values highlighted in red represent the LOQ of the corresponding SCFA. (C) Calibration curves for SCFAs with formic acid (**FA**) and lactic acid (**LA**) as examples.

Meanwhile, the matrix removal step improved the reproducibility of this quantification method by reducing the RSD from 23.6% to 8.4%.

One of the major advantages of our method is its high sensitivity, requiring only a minimal amount of the investigated biological sample. This combines well with global metabolomics analysis, as our method utilizes just 2% of the sample extract for precise SCFAs quantification. The remaining sample volume can then be used for conventional global metabolomics analysis. In this study, *quant*-SCHEMA was applied to 2 µL of the bacterial sample to investigate alterations in SCFAs between the healthy microbiome (H), a co-culture of the healthy microbiome and *Salmonella* (H+S), and the co-culture of the healthy microbiome with *Salmonella* in the presence of the antibiotic Spectinomycin (H+S+Ab) at two different time points (day 1 and day 5). We selected an antibiotic-resistant *Salmonella* strain as a realistic example to

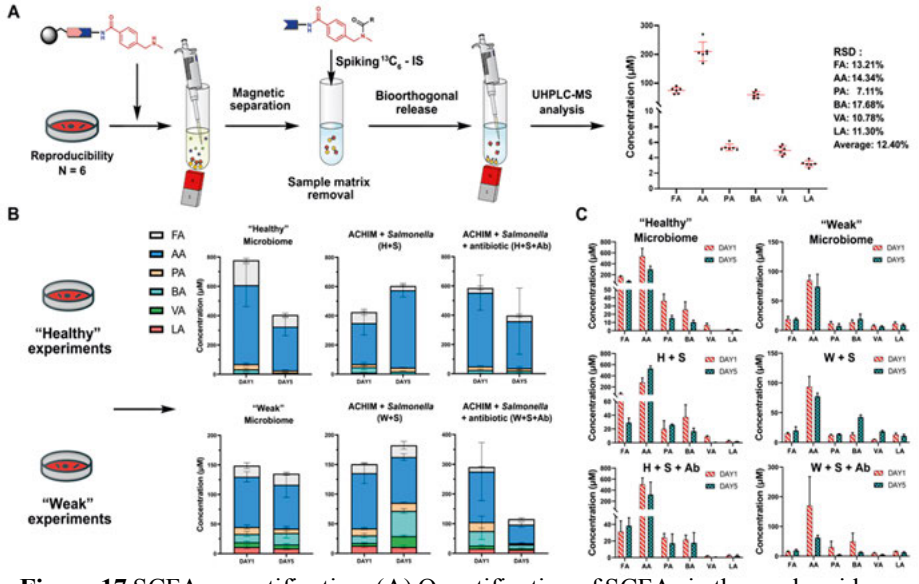


Figure 17 SCFAs quantification. (A) Quantification of SCFAs in the random identical sample 6 times to validate the method reproducibility (See also Table S4). Bioorthogonal cleavage condition: triphenylphosphine, dimethyl-barbituric acid, and $Pd(OAc)_2$ in THF for 5 hours. (B) Total SCFA quantification in biological samples and (C) individual SCFA values. Each condition was analysed as triplicates. Error bars are standard deviation.

represent the disruption of the healthy microbiome after antibiotic treatment, which has been shown to affect SCFA concentrations. Additionally, a diluted ACHIM culture, referred to as a 'weak' gut microbiome with lower bacterial density and SCFA concentration, was tested to evaluate our method's performance in such conditions (see Figure 17B). The antibiotic-treated samples (H+S+Ab) showed a significant decrease in total SCFA levels on day 5 compared to day 1, regardless of the health of the gut microbiome (Figure 17B). This reduction in total SCFA production is consistent with literature reports on antibiotic treatment, confirming the validity of our method. Our results also reveal that acetate, propionate, and butyrate are the SCFAs primarily affected in cultures treated with Spectinomycin. Moreover, the total SCFA concentrations in the healthy microbiome (H) slightly decreased after 5 days of culture compared to an increase observed in the co-cultures with Salmonella. This observation aligns with reports suggesting that the healthy microbiome's SCFA production may be linked to pathogen starvation through increased sugar consumption.

Based on the successful SCFA experiments conducted with a small volume of sample extract, we utilized the remaining sample for global mass spectrometry-based metabolomics analysis. We identified the main metabolites altered in three groups: the healthy microbiome, the *Salmonella* co-culture (H+S), and the antibiotic-administered culture (H+S+Ab). The chemical structures of 44

metabolites from the Top100 features with the highest statistical significance were determined through metabolite annotation using HMDB. A total of 13 metabolites were validated with the highest confidence level through comparison with reference standard compounds. Both *N*-acetylated amino acids were found to be reduced, while the amino acids valine and tyrosine showed significant increases after Streptomycin treatment. The reduction of *N*-acetylated amino acids after antibiotic treatment correlates with the decrease in acetic acid, which is required for the biosynthesis of the acetyl-donor Acetyl-CoA[133]. These findings are novel and indicate that our method holds significant promise in discovering compounds with activity against pathogens, especially antibiotic-resistant bacteria. This demonstrates the powerful application of our combined quantification methodology for volatile metabolites and the metabolites in standard global metabolomics analysis.

In summary, we have developed a novel method for the precise quantification of short-chain fatty acids with high sensitivity. This method only requires 2% of the global metabolomics sample volume, conserving precious biological samples while obtaining SCFA concentrations accurately. To illustrate this significant advantage, we also conducted conventional global metabolomics analysis using the remaining sample extract. Our quantitative methodology underwent validation through a series of analytical experiments, including linearity, limit of quantification, reproducibility, and a positive control experiment in the presence of an antibiotic. As a proof-of-concept study for our methodology, we successfully and precisely quantified the concentrations of SCFAs in 36 bacterial samples and conducted metabolomics analysis in parallel to identify metabolic alterations in these bacterial co-cultures. We envision that these quantitative chemoselective probes will find application in various biomedical studies for SCFA quantification, as they require only small sample volumes. This chemical metabolomics strategy can be combined with other analyses for low-quantity samples.

Chemical metabolomics for biomarker discovery

In Paper V, we applied *quant*-SCHEMA to a dietary intervention study aimed at discovering potential nutritional biomarkers. This was the first application of *quant*-SCHEMA for large-scale sample size analysis (78 vs 78), as in previous publication, we initially developed and tested the method with a small sample number only. As a result, we conducted a series of optimization experiments to enhance the performance of *quant*-SCHEMA in large-scale analysis.

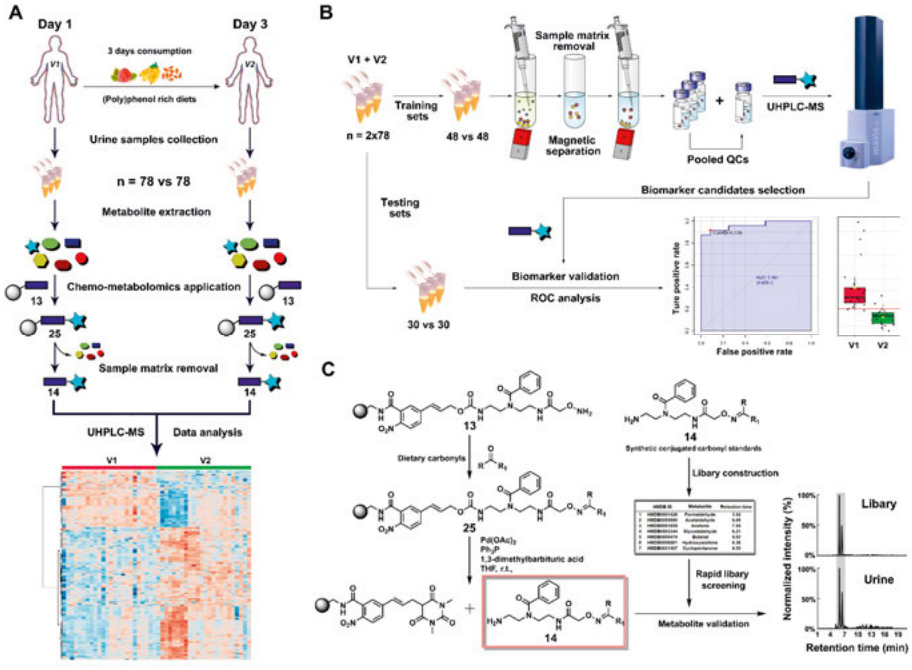


Figure 18 (A) General workflow for the dietary intervention study including sample collection, metabolite extraction, chemical metabolomics treatment, UHPLC-MS analysis, and data interpretation. (B) Workflow for dietary biomarker discovery. The study includes 78 volunteers with the collection of urine samples at two timepoints V1 andV2. The samples were split in 48 urine pairs for two training and 30 urine pairs as validation sets (N = 156 urine samples). (C) Overview of the chemoselective probe treatment procedure and metabolite structure validation of the tagged metabolites with authentic standards.

Polyphenols offer health benefits but have low bioavailability[134-137]. Few studies have investigated the relationship between polyphenols and gut microbiota using metabolomics combined with metagenomics[138, 139]. Our objective was to conduct a detailed analysis of carbonyl-containing metabolites to discover unknown bioactive compounds in a dietary intervention study, elucidating the effects of nutritional metabolism. We successfully employed our recently developed chemical metabolomics method, *quant*-SCHEMA, for semi-quantitative analysis of urine samples collected before and after the consumption of a (poly)phenol-rich breakfast comprising flaxseeds, raspberries, and soy milk, to obtain the dietary carbonyl profile. While individual carbonyl metabolites have been linked to dietary changes, no comprehensive investigation of this compound class has been reported. The results of our extensive analysis of carbonyl metabolites were used to identify metabolic differences resulting from the dietary intervention, potentially revealing candidate biomarkers. To assess the quality of these carbonyl-containing biomarkers, we employed a common metabolomics-

based biomarker discovery strategy. We divided the sample cohort into two training sample sets and one validation set. We selected 48 urine samples (24 vs 24) as Batch 1 Training Set (B1TS), another 48 urine samples (24 vs 24) as Batch 2 Training Set (B2TS), and 60 urine samples (30 vs 30) for the biomarker validation set (VS). Sample volumes were adjusted based on creatinine concentration to normalize metabolite concentrations. Our chemical metabolomics approach was then applied identically to these three sample sets but at different time points. Metabolites identified as statistically significant in both training sample sets were ranked as tentative biomarkers and tested in the validation set to confirm their status as dietary biomarkers (Figure 18).

We initially aimed to investigate global alterations in the carbonyl-containing metabolome during the dietary intervention. To gain a comprehensive

Table 1 Validated carbonyl-containing metabolites using our in-house library. Highlighted acetaldehyde was statistically significantly increased in V2 compared to V1. Highlighted 2,3-Pentanedione was statistically significantly reduced in V2 compared to V1.

	Carbonyl conjugates	Monoisotopic Mass	Conjugated m/z	Library RT (min)	Urine RT (min)	Source
1	Formaldehyde	30.0106	293.1609	10.2	10.17	Seeds
2	Acetaldehyde	44.0262	307.1765	10.89	10.9	Berries, seeds, milk and soy
3	Acetone	58.0419	321.1922	11.01	10.57	Decarboxylation of ketone bodies
4	Propanal	58.0419	321.1922	11.53	10.83	Foods, berries
5	Glycolaldehyde	60.0211	323.1714	9.98/10.27	9.82	-
6	Butanone	72.0575	335.2078	11.93/12.30	12.27/12.46	-
7	Butanal			12.30/12.48		Red raspberry, seeds, soy and milk
8	Hydroxyacetone	74.0368	337.1871	9.92/10.52	9.77/10.54	Foods
9	Lactaldehyde	74.0368	337.1871	9.94/10.56	10.54	Animal foods
10	Diacetyl	86.0368	349.1871	11.08	11.29	Flavoring agent
11	2-Pentanone	86.0732	349.2235	12.66/12.87	12.83/13.00	-
12	Isovaleraldehyde		349.2235	12.83/13.05		Blueberries
13	2-Methylbutyraldehyde		349.2235	12.84/13.02		-
14	Valeraldehyde		349.2235	12.93/13.14		Berries, soy and milk product
15	1-Hydroxy-2-butanone	88.0524	351.2027	13.9	13.9	Plants
16	Cyclohexanone	98.0731	361.2234	12.65/12.93	12.89	Berries and cereal products
17	2,3-Pentanedione	100.0524	363.2027	12.15/12.50	12.00/12.28	Berries and teas
18	Acetylacetone	100.0524	363.2027	11.16	11.2	Berries
19	3-Hexanone	100.0888	363.2391	13.48	13.31/13.47	Raspberry, flaxseed, soy milk
20	2-Methyl-3-pentanone		363.2391	13.51		Corns
21	2-Methylpentanal		363.2391	13.59		Onion
22	2-Hexanone		363.2391	13.33/13.51		-
23	Hexanal		363.2391	13.59/13.73		Raspberry, flaxseed, soy milk
24	Benzaldehyde	106.0419	369.1922	13.56	13.25	Natural plant foods
25	2,5-Dimethyl-3(2H)-furanone	112.0524	375.2027	9.74	9.92	-
26	Levulinic acid	116.0473	379.1976	10.90/11.46	11.94	Cellulose
27	Ketoisocaproic acid	130.0630	393.2133	11.53/12.65	11.65	Animal foods
28	Safranal	150.1045	413.2548	11.01/12.23	11.69	Teas and fruits

overview of carbonyl alterations, we combined the analysis of B1TS, B2TS, and VS (78 vs 78). We selected the top 50 common features among these three distinct datasets based on p-values from paired Student's t-tests. The heatmap illustrates that most carbonyl-containing metabolites decreased as a result of the dietary intervention. This could be attributed to two possible reasons: i) In V1, the volunteers were in a fasting state and generated energy by consuming fats to produce more ketone bodies. The switch to the (poly)phenol-rich breakfast led to a shift in energy generation toward glucose. ii) Another possibility is the antioxidant properties of flaxseeds, raspberries, and soy milk. The consumption of this (poly)phenol-rich diet may facilitate the clearance of carbonyl-containing metabolites from the human body.

Based on the successful global carbonyl profiling, we aimed to obtain structural information for the top hits by applying our in-house library. The highest level of validation for detected metabolite structures was achieved

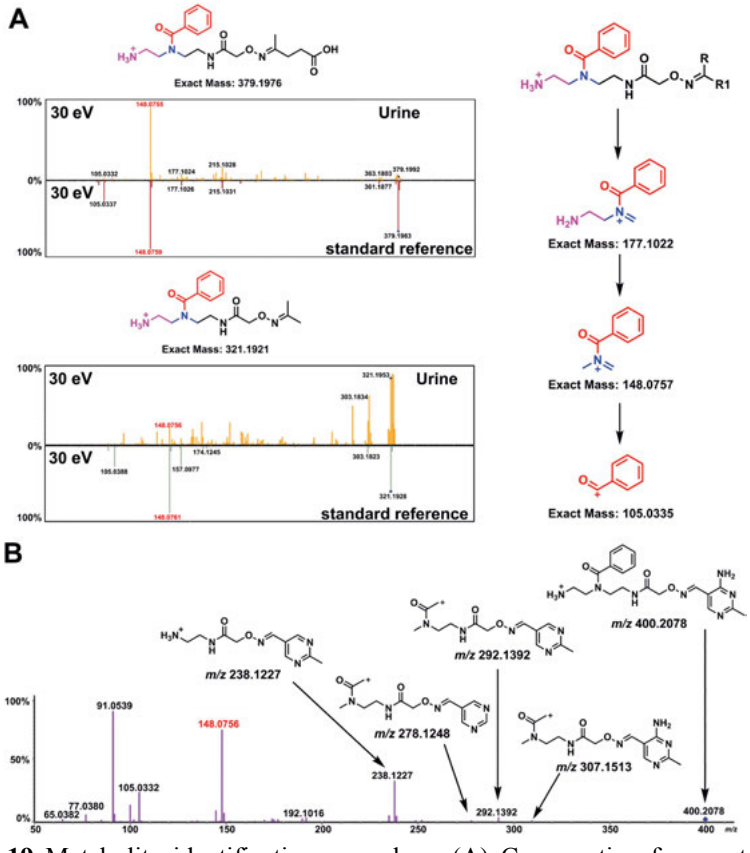


Figure 19 Metabolite identification procedure. (A) Comparative fragmentation analysis of carbonyl-conjugated metabolites with in-house prepared reference standards. (B) MS/MS fragmentation analysis of one of the tentative biomarkers 4-amino-2-methylpyrimidine-5-carbaldehyde with annotated MS fragmentation ions.

through co-injection analysis with synthetic or commercially available reference compounds that exhibited identical chromatographic properties and MS/MS fragmentation patterns as the metabolite in the biological sample (Confidence level 1). Metabolites annotated in databases could be identified through MS/MS fragmentation comparisons (Confidence level 2). The in-house metabolite conjugate library described in Paper I facilitates rapid structure validation by directly comparing the retention time and MS/MS fragmentation pattern between the natural metabolite and the constructed library compound (Figure 18C). This straightforward method provided rapid insights into the chemical structure of carbonyl-containing metabolites. We successfully validated 30 carbonyl-containing metabolites by directly comparing them to the conjugated library. As expected, most of these validated carbonyls were previously associated with specific food sources based on comparisons with HMDB and the Food Database (FooDB). For instance, acetaldehyde, butanal, valeraldehyde, 3-hexanone, and hexanal were directly linked to the consumption of various berries, seeds, and soy products. Propanal, cyclohexanone, 2,3-pentanedione, acetylacetone, benzaldehyde, and 2-undecanone primarily originated from fruits, particularly berries. Kynurenine, a product of milk consumption, could be linked to soy products (Table 1). To facilitate the discovery of other carbonyl-containing metabolites, we also identified compounds through MS/MS fragmentation analysis. To achieve this, we first generated MS/MS fragmentation spectra for standards from our conjugated library. This analysis revealed the presence of the fragment 148.0757 in all conjugated standards, resulting from the partial loss of the probe linker. This finding confirmed the bioinformatic data analysis output and provided an additional diagnostic fragment to enhance confidence in the identification of carbonyl-containing metabolites (Figure 19A).

The first training set (B1TS) yielded 132 features with a 1.5-fold alteration and p-values < 0.05. Consistent with the global analysis, more carbonyl-containing metabolites were found at reduced levels after the dietary intervention. A similar analysis was conducted for B2TS, resulting in over 275 significantly altered features with a fold change of over 1.5. Upon comparison of the feature lists from these two sets, we identified 24 features commonly altered in both training data sets, providing a higher level of confidence in the validity of these metabolites as dietary biomarkers. We successfully annotated 5 of these 24 biomarker candidates by comparing the detected high-resolution mass without the probe tag with HMDB. The presence of the diagnostic MS/MS tag fragment ($m/z = 148.0754$) for all five metabolites confirmed the presence of a carbonyl group in their chemical structure. The MS/MS fragmentation analysis for 4-amino-2-methylpyrimidine-5-carbaldehyde serves as a representative example of structural elucidation (see Figure 19B). The loss of the methyl group and amino moiety from the dietary compound was observed (m/z =307.1513, 292.1404, 278.1248). These three fragment

ions are specific and were not observed in the MS/MS analysis of our constructed library of over 100 compounds. For instance, 1-(methyldithio)acetone (**a** / p-value = 0.0129 in B1TS, 0.0005 in B2TS), glucose ketone (**c** / p-value < 0.0001 in B1TS, 0.0261 in B2TS), and thiacremonone (**e** / p-value = 0.0019 in B1TS, 0.0016 in B2TS) increased after the consumption of flaxseeds, raspberries, and soy milk, whereas 4-amino-2-methylpyrimidine-5-carbaldehyde (**b** / p-value = 0.0021 in B1TS, 0.0082 in B2TS) and 6'-methoxypolygoacetophenoside (**d** / p-value < 0.0001 in both B1TS and B2TS) were found at reduced levels.

In the validation data set (VS / 30 V1 and 30 V2 samples), we performed a targeted analysis specifically to investigate the five biomarker candidates and determine the alterations within this independent sample cohort. Metabolite e was excluded as a false positive hit in the validation set due to its low statistical significance (p = 0.1598). Since compound a was significantly increased in

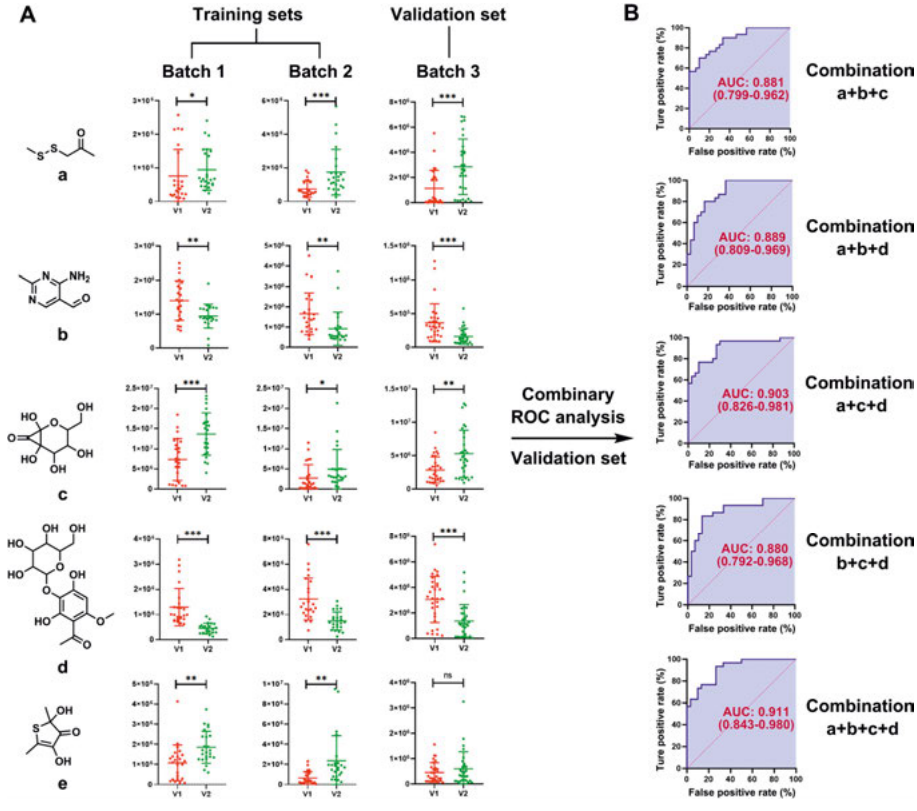


Figure 20 Discovery and evaluation of dietary biomarkers. (A) Tentative structures for the identified biomarkers and their chemical formulas: a: C4H8OS2, b: C6H7N3O, c: C7H10O7, d: C15H20O10, e: C6H8O3S. Mass spectrometric intensities and statistical analysis in the three different data sets (paired Student's t-test, p-values: * p < 0.05; ** p < 0.01 and *** p < 0.001); (B) ROC analysis in the validation set for different combinations of the metabolites.

this third sample set, we further evaluated it through receiver operating characteristic (ROC) analysis to assess its specificity and sensitivity. The ROC analysis validated this metabolite as a promising nutritional biomarker based on AUCs of 0.735, 0.805, and 0.790 for B1TS, B2TS, and VS, respectively (see Figure S7-S9). Additionally, metabolite c was consistently identified as significantly increased after the consumption of the (poly)phenol-enriched breakfast in the three datasets B1TS, B2TS, and VS. Similarly, carbonyls d and b consistently decreased after dietary intervention in B1TS, B2TS, and VS (see Figure 20A). The AUC values for these four candidates ranged from 0.735 to 0.840, confirming a low risk of false positive assessment.

As even the most suitable single metabolite marker carries a high risk of false positives, we combined the top four metabolites and conducted a ROC analysis in various combinations with three or all four metabolites (see Figure 20B). This analysis yielded high AUC values ranging from 0.880 to 0.911, underscoring the high sensitivity and selectivity of these four metabolites as biomarkers for this (poly)phenol-rich diet.

In summary, we applied a new chemical metabolomics approach for the first time in a dietary context, with a specific focus on carbonyl-containing metabolites. This compound class is important due to its relevance to microbiome metabolism and the quenching properties of polyphenols on reactive oxygen species. In this study, we utilized the recently developed *quant*-SCHEMA method, which we optimized for large-scale analysis. Our analysis of a sample cohort comprising 78 individuals who consumed a (poly)phenol-rich breakfast consisting of raspberry powder, flaxseeds, and soy milk for three days led to the discovery of novel dietary biomarkers. In total, we identified a set of four carbonyl-containing metabolites that serve as markers for this diet. This study underscores the advantages of advanced chemical biology methods in nutritional research.

Conclusion

Metabolomics is an evolving scientific research field that demands a diverse array of tools to facilitate its expansion and practical application. To date, metabolomics methodologies have led to significant achievements in biomarker discovery and the exploration of disease development. Metabolites as small molecules with modifiable functional groups such as amines, carbonyls, carboxylic acids, and hydroxyl groups, offer abundant opportunities for chemoselective modifications. This success of dissertation represents the importance of the integration of organic synthesis and chemical biology into metabolomics research. Consequently, there's an increased demand for interdisciplinary skills to advance metabolomics forward. While analytical chemistry remains critical, expertise in chemical biology and organic chemistry is essential for pioneering new discoveries.

This dissertation marks the successful development of a series of novel chemical biology tools for advanced metabolomics analysis including the elucidation of metabolite structure and quantitative analysis for biomarker discovery. We have established and introduced chemical metabolomics step by step from qualitative to quantitative analysis. We have designed a chemical probe consisting of a solid support (magnetic beads), a bioorthogonal cleavage site (Noc), a tagging linker and a chemoselective reactive site. This innovative design enables the chemoselective enrichment of desired metabolites, significantly enhancing the reliability of mass spectrometry detection and minimizing false positives. Magnetic beads expedite metabolite isolation, and the cleavable linker allows us to release target metabolites. The tagging process substantially improves the chemical and physical characteristics of metabolites, similar to other derivatization techniques.

In the first paper, we introduced a strategy for constructing metabolite conjugate libraries to swiftly validate metabolite structures, yielding twice as many validated metabolites as other reported methods. We also identified five previously undetectable metabolites. This demonstrates that our bipartite approach, involving our chemoselective probe combined with sophisticated bioinformatic analysis or a constructed conjugates library, allows for the discovery and validation of known and unknown metabolites in complex human samples. This study lays the foundation for selective metabolomics

analysis of carbonyl-containing compounds at low concentrations and the identification of disease biomarkers.

In the second paper, we described the development of a novel chemoselective reaction moiety for amine conjugation in metabolomics, using squaric acid for the first time. We determined its chemoselectivity, reactivity, stability, and ionizability, successfully capturing as well as analysing amine-containing metabolites in human fecal samples. Due to its superior properties, we have detected three previously undetected metabolites in this sample type and a range of microbiota-derived metabolites. Furthermore, we have designed a more reliable and straightforward method for synthesizing a library of 93 amine-containing metabolites for efficient validation of metabolites with coverage of different metabolite classes.

The third paper transformed our chemical probe from qualitative detection to quantitative analysis by redesigning it to incorporate isotopically labelled linkers. In addition, we have also shortened the synthesis for better utilization in the future. This methodology, based on a chemoselective probe immobilized to magnetic beads with light and heavy isotope tags, outperforms other derivatization methods in metabolite analysis, which lack separation from the sample matrix and must overcome matrix interferences. Due to the combination of natural and isotope-labelled conjugates before LC-MS analysis, it does not require for normalization procedures, internal standards, or adjustments for different sample types, facilitating the detection and quantification of 21 previously undetected metabolites. This comprehensive method enhances mass spectrometric analysis to biomarker discovery and enables personalized medicine and diagnostic monitoring.

In the fourth paper, due to high importance of short-chain fatty acids and biological sample volume, we have designed a specific reactive site facilitating the precise quantification of SCFAs. This method only requires 2% of the global metabolomics sample volume to save precious biological samples and to additionally obtain the SCFA concentrations. Our quantitative methodology has been validated by a series of analytical experiments including linearity, limit of quantification, reproducibility, and a positive control experiment in the presence of an antibiotic. The designed synthetic route of $^{13}C_6$-labelled conjugates allows for the synthesis of the corresponding internal standard for any desired carboxylic acid metabolite as our method does not require a commercially available isotopically labelled standard. We anticipate widespread application of these quantitative chemoselective probes in biomedical studies due to their minimal sample volume requirement.

In the last paper, we have reached an important step, biomarker discovery, of the development of new chemical biology tools for metabolomics. We have

applied a new chemical metabolomics approach for the first time in a dietary context with a focus on carbonyl-metabolites. These metabolites are significant due to their connection with microbiome metabolism and their role in quenching reactive oxygen species, especially in the presence of polyphenols. We optimized our *quant*-SCHEMA method for large-scale analysis in a dietary study, identifying novel dietary biomarkers that represent a (poly)phenol-rich diet. This study demonstrates the value of advanced chemical biology methods in nutritional research.

In conclusion, this dissertation describes the journey of developing chemical probes for metabolomics analysis and lays the groundwork for chemical metabolomics. Initially, our focus was on qualitatively detecting carbonyl and amine-containing metabolites, but our journey evolved to encompass quantitative analysis of carbonyls and SCFAs to better understand their roles in biological processes. Furthermore, this dissertation bridges the gap between analytical chemistry and chemical biology, highlighting the essential role of organic chemistry in this scientific field.

Populärvetenskaplig sammanfattning

Tänk om många sjukdomar kunde diagnostiseras i ett tidigt skede under en vanlig hälsoundersökning?

Den mänskliga mikrobiomen, ett komplext ekosystem av mikroorganismer som bor i våra kroppar, spelar en avgörande roll för vår allmänna hälsa. Ny forskning har visat att dessa mikroskopiska invånare har en djupgående påverkan på värdens fysiologi genom att producera högt aktiva föreningar, d.v.s. mikrobiella metaboliter. Dessa metaboliter kan påverka utvecklingen av olika sjukdomar på ett betydligt sätt. Upptäckten av mikrobiella metaboliter som potentiella indikatorer eller biomarkörer för hälsotillstånd erbjuder en spännande möjlighet. Hur kan vi göra det till verklighet?

"Metabolomik" är studien av molekyler som produceras från ämnesomsättningen. Upptäckten av metaboliter är ett avgörande steg för biomarkörsupptäckt. Men förståelsen av mikrobiotans sammansättning och funktion är fortfarande oklart på grund av dess enastående komplexitet. Dessutom är det mycket svårt att genomföra en omfattande analys av hela metabolomen. Som svar på dessa utmaningar vänder forskare sig ofta till derivatiseringstekniker.

Derivatisering innebär att ändra metaboliters struktur för att förbättra deras detekterbarhet och stabilitet under analysen. Även om derivatisering har varit ett värdefullt verktyg inom metabolomik, men derivatiseringen riskerar att introducera artefakter, som i sin tur kan leda till försämrade noggrannhet i analytiska resultat. Därför finns ett brådskande behov av mer avancerade tekniker för att förbättra precisionen i derivatiseringsbaserad metabolomik.

I denna doktorsavhandling har vi utvecklat en banbrytande metod som kallas kemisk metabolomik. Denna innovativa metod använder kemiska sonder som immobiliserats på magnetiska pärlor för att fånga metaboliter i biologiska prover. De framgångsrika resultaten med kemoselektiva sonder inom kemisk metabolomik understryker betydelsen av kemiska biologiverktyg för att främja metabolomik. Denna omfattande avhandling representerar ett betydande framsteg i vår förmåga att upptäcka biomarkörer och få insikter om

mikrobiella interaktioner med den mänskliga värden och utvecklingen av sjukdomar.

När vi avslöjar mikrobiomens hemligheter, fältet för kemisk metabolomik genomgår en revolutionerande biomarkörsupptäckt inom metabolomik. Denna innovativa metod erbjuder potentialen att ge värdefulla insikter om relationerna mellan mikrobiella samhällen inom våra kroppar och olika hälsotillstånd. Med sin förbättrade precision och känslighet, kemisk metabolomik lovar att bli ett värdefullt verktyg för att främja vår förståelse av den mänskliga mikrobiomens påverkan på mänsklig hälsa. Vi är entusiastiska över våra lovande resultat, eftersom det representerar ett betydande steg framåt i förståelsen av mikrobiotans ämnesomsättning.

科普摘要

如果许多疾病能够在定期的体检中早期被诊断，那会是怎样一种局面呢？

人类人体微生物组，是由居住在我们体内的微小生物构成的复杂生态系统，在维护我们整体健康中发挥着至关重要的作用。最新的研究揭示，这些微不足道的生物通过生成被称为微生物代谢产物的高活性化合物，对宿主的生理产生了深远的影响。这些代谢产物能够显著影响各种疾病的发展。如果将微生物代谢产物视为潜在的健康状况指标或生物标志物，这会为我们带来了令人振奋的机遇。那么，我们该如何将这一机遇变为现实呢？

“代谢组学”是研究由新陈代谢过程所产生的分子的学科。代谢物的检测是生物标志物发现的至关重要的一步。然而，由于微生物组的异常复杂性，了解其组成和功能仍然具有极大的挑战性。此外，进行全面的微生物组分析也非常困难。为应对这些挑战，研究人员通常转向衍生化技术。

衍生化涉及改变代谢物的结构，以提高它们在分析过程中的可检测性和稳定性。尽管衍生化在代谢组学中一直是一个重要的工具，但它也伴随着引入可能损害分析结果准确性的人为因素的风险。因此，我们急需更加先进的技术来提高基于衍生化的代谢组学的精确性。

在这篇博士论文中，我们创新性地提出了一种被称为化学代谢组学的方法。这一创新性方法利用化学探针固定在磁珠上，从生物样本中捕获代谢物。化学代谢组学中化学选择性探针的成功应用，凸显了化学生物学工具在推动代谢组学方面的重要性。这篇全面的论文代表了我们在发现生物标志物、深入了解微生物与人体宿主相互作用以及疾病发展方面的能力取得的巨大进步。

随着我们揭示微生物组的秘密，化学代谢组学领域被认为将彻底改变代谢组学驱动的生物标志物发现。这种创新方法有望为我们提供有关人体内微生物社群与各种健康状况之间关系的宝贵见解。凭借其提高的精确性和敏感性，化学代谢组学将成为深化我们对人体微生物组及其对人类健康的影响理解的宝贵工具。我们为自己的发现感到兴奋，因为这代表了在提升对微生物代谢的理解方面的重大飞跃。

Acknowledgements

As I reflect on my journey through my PhD, I am overwhelmed by the incredible support and assistance I've received. I feel truly fortunate to have met and interacted with the remarkable individuals and colleagues who have become an indelible part of my PhD experience. This journey holds countless unforgettable memories for me, all of which centre around my doctoral studies. I want to extend my heartfelt appreciation to every person who has offered guidance and support to me. Pursuing my PhD in Sweden has filled in me a sense of strength and developed my ability to think independently and critically.

Foremost, my gratitude goes to my supervisor, **Daniel**, who not only granted me the opportunity to pursue a PhD but also provided solid guidance throughout my doctoral studies. Your smart advice for my experiments during the early stages of my scientific career were invaluable. Your mentorship has transformed me into a more self-reliant and capable researcher. I enjoy a lot when we had our scientific discussions, exchange of new ideas, and conversations about life. I am also deeply thankful for your support in my funding application for conferences and other awards. It is beyond doubt that you have played a critical role in shaping my personality and scientific mindset. I am grateful to have you as my PhD supervisor, enabling me to receive sufficient attention to learn the art of research and stay on the right track in my career development.

I would also like to express my appreciation to my co-supervisor, **Mikael**. I always enjoy the time for our conversation. Your encouragement and kindness have contributed significantly to my growth as a scientist. I am happy that I can have you to fill the gaps in my knowledge of microbiology. Your support for my career is highly appreciated.

I am particularly grateful for being the second PhD student in the Globisch lab, which meant there was a predecessor to help overcome any obstacles I might encounter during my doctoral journey. **Mario**, thank you for your patience in guiding me to integrate into the lab. I have learned a lot from you, not only about science but also about understanding Western culture. I am

delighted that our periods of overlap promoted my PhD experience. My life will be unbelievable difficult with your involvement in my PhD.

I extend my thanks to **Louis** as well. Having an English native speaker as my initial direct supervisor in the laboratory was very lucky. Because no matter how bad my English was, you always understand and answer my question. It was an incredible experience. My gratitude also goes out to other first-generation members of the Globisch lab, **Neeraj** and **Caroline**. The warm welcome from you made me feel at home in the lab. I must also acknowledge **Abhishek** for sharing his knowledge of metabolomics and data analysis. I have gained a lot of relevant knowledge when I have scientific discussions with you. To **Theodosia**, who shared an office with me for the longest time, I am thankful for your behaving like my mom and the advice you offered during my down time, although I'm still waiting for an invitation to your housewarming which I know it will not be going to happen. To the second generation of Globisch lab members, **Amanpreet**, **Sydney**, **Vladyslav**, **Carina**, **Fan**, our shared experiences made our days at the office unbelievably better. I appreciate the contributions of the students in the lab, particularly **Lars**, **Amelie**, **Anaïs**, **Kiana**, **Léna**, **Weiming**, **Spyke**, **Zhen** and **Marko**. I enjoyed the time when we discussed about science and provided my supervision in the lab.

A special thank you to **Ioanna**, whose remarkable support I cherish. Spending time with you in the office has been a true pleasure, and I appreciate the unique relationship we share. It's unbelievable that I was fortunate to be a PhD student between Mario and you. I believe my lab life would be a disaster if I had myself as a senior or junior PhD student. You have added immeasurable value to my life in the lab, and I cannot express enough gratitude for your support with my poor vocabulary. Thank you again for your help and time you spent with me.

I would like to extend my thanks to all my former and current colleagues at the Department of Medicinal Chemistry and the Department of Chemistry (BMC). Your kindness, shared conversations, and moments of relaxation and fun enriched my days. Special mention to **Leonidas**, **Johan**, and **Catia,** you guys are always funny**. Mark, Lucie, Varun,** and **Sandy**, thanks for your kindness and helps.

My days in Sweden were made brighter by the company of **Carlos**, **Kristina**, **David**, **Sara**, **Jens** and **Sakis**. You all have been wonderful friends, and our lives outside of work were more enjoyable because of the time we spent together. I also cherish the memories created with **Alex**, **Iselin**, **Claire**, **Manuel**, **Simon**, and **Yaelle**. I'm grateful for the reunions we've made and our continued connection. I still remember our regular Zoom meeting during the

pandemic, as well as the fun card games and travels we shared. To **Hongyi** and **Ruiying**, thank you for your support and companionship during my PhD journey, and I'm delighted to have found someone from my home region. To **Huasi and your husband**, I enjoyed a lot when we visited your home and time we shared with your dogs.

I would like to express my sincere gratitude to whom provided the encouragement and inspiration that motivated me to pursue my PhD when I was still innocent. I also appreciate whom cheered me up during my darkest moments in my PhD journey. I am grateful to all the other people I met in Sweden for your kindness and friendliness, which made me feel truly welcome.

I cannot forget the support I received from my hometown, Shenzhen. I deeply appreciate my friends in Shenzhen, who always hosted me very well during my holidays. Special thanks to my family, I am very happy to have my parents and my siblings who are very open-minded. Thank you for giving me the freedom to choose what I want to do. I could not have completed my PhD without your support.

Lastly, **Mengqi**, I cannot believe all that we have been through. Although you are such a troublemaker in my daily life, I still determinedly choose you. Until now, I am convinced that the challenges you brought into my life have ultimately made me stronger. As the saying goes, 'What doesn't kill me makes me stronger.' I wholeheartedly believe it now. Our time together has undoubtedly led to personal growth, and I enjoy a lot when I make fun of you. If I could go back to the moment we met and choose my life path again, I would still choose to donate my blood as your experimental volunteer without any hesitation. Because you are worthy.

謹以此文，獻於昔年方十七之己，余不禁感慨良多。雖遺憾之傷，或時常有之，然感激之情，仍深沉如海。若非往昔之己，常能於逆境之中，重拾信心，堅韌不拔，今日之工亦难以臻於斯境。且盼來日之途，黃金榜上，莫再失龍頭望。

吾深憾於太白之神工，故常詠其詩句以自勉矣。

歌曰：

“長風破浪會有時，直掛雲帆濟滄海。”

References

1. T. S. B. Schmidt, J. Raes and P. Bork, *Cell*, 2018, **172**, 1198-1215.
2. H. M. P. Consortium, *Nature*, 2012, **486**, 207-214.
3. S. R. Gill, M. Pop, R. T. Deboy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett and K. E. Nelson, *Science*, 2006, **312**, 1355-1359.
4. A. Visconti, C. I. Le Roy, F. Rosa, N. Rossi, T. C. Martin, R. P. Mohney, W. Li, E. de Rinaldis, J. T. Bell, J. C. Venter, K. E. Nelson, T. D. Spector and M. Falchi, *Nat Commun*, 2019, **10**, 4505.
5. M. S. Donia and M. A. Fischbach, *Science*, 2015, **349**, 1254766.
6. J. K. Nicholson, E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia and S. Pettersson, *Science*, 2012, **336**, 1262-1267.
7. C. Whidbey, N. C. Sadler, R. N. Nair, R. F. Volk, A. J. DeLeon, L. M. Bramer, S. J. Fansler, J. R. Hansen, A. K. Shukla, J. K. Jansson, B. D. Thrall and A. T. Wright, *J Am Chem Soc*, 2019, **141**, 42-47.
8. Y. K. Lee and S. K. Mazmanian, *Science*, 2010, **330**, 1768-1773.
9. Y. Belkaid and T. W. Hand, *Cell*, 2014, **157**, 121-141.
10. J. M. Pickard, M. Y. Zeng, R. Caruso and G. Núñez, *Immunol Rev*, 2017, **279**, 70-89.
11. E. Vogtmann and J. J. Goedert, *Br J Cancer*, 2016, **114**, 237-242.
12. A. K. DeGruttola, D. Low, A. Mizoguchi and E. Mizoguchi, *Inflamm Bowel Dis*, 2016, **22**, 1137-1150.
13. S. Carding, K. Verbeke, D. T. Vipond, B. M. Corfe and L. J. Owen, *Microb Ecol Health Dis*, 2015, **26**, 26191.
14. W. L. Wang, S. Y. Xu, Z. G. Ren, L. Tao, J. W. Jiang and S. S. Zheng, *World J Gastroenterol*, 2015, **21**, 803-814.
15. E. P. Culligan, J. R. Marchesi, C. Hill and R. D. Sleator, *Gut Microbes*, 2012, **3**, 394-397.
16. F. Bäckhed, C. M. Fraser, Y. Ringel, M. E. Sanders, R. B. Sartor, P. M. Sherman, J. Versalovic, V. Young and B. B. Finlay, *Cell Host Microbe*, 2012, **12**, 611-622.
17. N. Dalal, R. Jalandra, N. Bayal, A. K. Yadav, Harshulika, M. Sharma, G. K. Makharia, P. Kumar, R. Singh, P. R. Solanki and A. Kumar, *J Cancer Res Clin Oncol*, 2021, **147**, 3141-3155.
18. N. Koppel, V. Maini Rekdal and E. P. Balskus, *Science*, 2017, **356**.
19. K. B. Martinez, V. Leone and E. B. Chang, *J Biol Chem*, 2017, **292**, 8553-8559.
20. A. Milshteyn, D. A. Colosimo and S. F. Brady, *Cell Host Microbe*, 2018, **23**, 725-736.

21. A. R. Healy and S. B. Herzon, *J Am Chem Soc*, 2017, **139**, 14817-14824.
22. D. S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B. L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V. W. Lui, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H. B. Schiöth, R. Greiner and V. Gautam, *Nucleic Acids Res*, 2022, **50**, D622-D631.
23. D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel and I. Forsythe, *Nucleic Acids Res*, 2009, **37**, D603-610.
24. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res*, 2013, **41**, D801-807.
25. D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach and A. Scalbert, *Nucleic Acids Res*, 2018, **46**, D608-D617.
26. J. L. Spratlin, N. J. Serkova and S. G. Eckhardt, *Clin Cancer Res*, 2009, **15**, 431-440.
27. M. Mamas, W. B. Dunn, L. Neyses and R. Goodacre, *Arch Toxicol*, 2011, **85**, 5-17.
28. Q. Yang, A. H. Zhang, J. H. Miao, H. Sun, Y. Han, G. L. Yan, F. F. Wu and X. J. Wang, *RSC Adv*, 2019, **9**, 37245-37257.
29. M. Jacob, A. L. Lopata, M. Dasouki and A. M. Abdel Rahman, *Mass Spectrom Rev*, 2019, **38**, 221-238.
30. D. S. Wishart, *Nat Rev Drug Discov*, 2016, **15**, 473-484.
31. G. A. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah and D. Raftery, *Expert Rev Mol Diagn*, 2008, **8**, 617-633.
32. G. D. Lewis, A. Asnani and R. E. Gerszten, *J Am Coll Cardiol*, 2008, **52**, 117-123.
33. A. Bauermeister, H. Mannochio-Russo, L. V. Costa-Lotufo, A. K. Jarmusch and P. C. Dorrestein, *Nat Rev Microbiol*, 2022, **20**, 143-160.
34. E. A. Eloe-Fadrosh and D. A. Rasko, *Annu Rev Med*, 2013, **64**, 145-163.
35. G. Berg, D. Rybakova, D. Fischer, T. Cernava, M. C. Vergès, T. Charles, X. Chen, L. Cocolin, K. Eversole, G. H. Corral, M. Kazou, L. Kinkel, L. Lange, N. Lima, A. Loy, J. A. Macklin, E. Maguin, T. Mauchline, R. McClure, B. Mitter, M. Ryan, I. Sarand, H. Smidt, B. Schelkle, H. Roume, G. S. Kiran, J.

Selvin, R. S. C. Souza, L. van Overbeek, B. K. Singh, M. Wagner, A. Walsh, A. Sessitsch and M. Schloter, *Microbiome*, 2020, **8**, 103.
36. J. F. Cryan, K. J. O'Riordan, C. S. M. Cowan, K. V. Sandhu, T. F. S. Bastiaanssen, M. Boehme, M. G. Codagnone, S. Cussotto, C. Fulling, A. V. Golubeva, K. E. Guzzetta, M. Jaggar, C. M. Long-Smith, J. M. Lyte, J. A. Martin, A. Molinero-Perez, G. Moloney, E. Morelli, E. Morillas, R. O'Connor, J. S. Cruz-Pereira, V. L. Peterson, K. Rea, N. L. Ritz, E. Sherwin, S. Spichak, E. M. Teichman, M. van de Wouw, A. P. Ventura-Silva, S. E. Wallace-Fitzsimons, N. Hyland, G. Clarke and T. G. Dinan, *Physiol Rev*, 2019, **99**, 1877-2013.
37. J. L. Markley, R. Brüschweiler, A. S. Edison, H. R. Eghbalnia, R. Powers, D. Raftery and D. S. Wishart, *Curr Opin Biotechnol*, 2017, **43**, 34-40.
38. P. Giraudeau, *Analyst*, 2020, **145**, 2457-2472.
39. A. H. Emwas, *Methods Mol Biol*, 2015, **1277**, 161-193.
40. S. Alseekh, A. Aharoni, Y. Brotman, K. Contrepois, J. D'Auria, J. Ewald, J. C Ewald, P. D. Fraser, P. Giavalisco, R. D. Hall, M. Heinemann, H. Link, J. Luo, S. Neumann, J. Nielsen, L. Perez de Souza, K. Saito, U. Sauer, F. C. Schroeder, S. Schuster, G. Siuzdak, A. Skirycz, L. W. Sumner, M. P. Snyder, H. Tang, T. Tohge, Y. Wang, W. Wen, S. Wu, G. Xu, N. Zamboni and A. R. Fernie, *Nat Methods*, 2021, **18**, 747-756.
41. K. Dettmer, P. A. Aronov and B. D. Hammock, *Mass Spectrom Rev*, 2007, **26**, 51-78.
42. G. A. Gowda and D. Djukovic, *Methods Mol Biol*, 2014, **1198**, 3-12.
43. A. Vaniya and O. Fiehn, *Trends Analyt Chem*, 2015, **69**, 52-61.
44. J. F. Xiao, B. Zhou and H. W. Ressom, *Trends Analyt Chem*, 2012, **32**, 1-14.
45. N. Bar, T. Korem, O. Weissbrod, D. Zeevi, D. Rothschild, S. Leviatan, N. Kosower, M. Lotan-Pompan, A. Weinberger, C. I. Le Roy, C. Menni, A. Visconti, M. Falchi, T. D. Spector, J. Adamski, P. W. Franks, O. Pedersen, E. Segal and I. D. consortium, *Nature*, 2020, **588**, 135-140.
46. N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T. L. Pedersen, S. R. Smith, F. Bamforth, R. Greiner, B. McManus, J. W. Newman, T. Goodfriend and D. S. Wishart, *PLoS One*, 2011, **6**, e16957.
47. E. Holmes, T. M. Tsang, J. T. Huang, F. M. Leweke, D. Koethe, C. W. Gerth, B. M. Nolden, S. Gross, D. Schreiber, J. K. Nicholson and S. Bahn, *PLoS Med*, 2006, **3**, e327.
48. T. Vallianatou, W. Lin, N. B. Bèchet, M. S. Correia, N. C. Shanbhag, I. Lundgaard and D. Globisch, *J Cereb Blood Flow Metab*, 2021, **41**, 3324-3338.
49. E. Rampler, Y. E. Abiead, H. Schoeny, M. Rusz, F. Hildebrand, V. Fitz and G. Koellensperger, *Anal Chem*, 2021, **93**, 519-545.
50. L. Xu, C. Basheer and H. K. Lee, *J Chromatogr A*, 2009, **1216**, 701-707.
51. W. Liu, L. Zhang, Z. Wei, S. Chen and G. Chen, *J Chromatogr A*, 2009, **1216**, 5340-5346.

52. Y. Iwasaki, Y. Nakano, K. Mochizuki, M. Nomoto, Y. Takahashi, R. Ito, K. Saito and H. Nakazawa, *J Chromatogr B Analyt Technol Biomed Life Sci*, 2011, **879**, 1159-1165.
53. B. J. Marquis, H. P. Louks, C. Bose, R. R. Wolfe and S. P. Singh, *Chromatographia*, 2017, **80**, 1723-1732.
54. M. Karlsen, H. Liu, J. E. Johansen and B. H. Hoff, *Molecules*, 2015, **20**, 5329-5345.
55. Y. Z. Wang, Y. Y. Chen, X. Z. Wu, P. R. Bai, N. An, X. L. Liu, Q. F. Zhu and Y. Q. Feng, *Anal Chem*, 2023, **95**, 11550-11557.
56. M. J. Bañuls, F. García-Piñón, R. Puchades and A. Maquieira, *Bioconjug Chem*, 2008, **19**, 665-672.
57. B. A. Garcia, S. Mollah, B. M. Ueberheide, S. A. Busby, T. L. Muratore, J. Shabanowitz and D. F. Hunt, *Nat Protoc*, 2007, **2**, 933-938.
58. Z. Wei, X. Zhang, J. Wang, S. Zhang and R. G. Cooks, *Chem Sci*, 2018, **9**, 7779-7786.
59. N. An, Q. F. Zhu, Y. Z. Wang, C. F. Xiong, Y. N. Hu and Y. Q. Feng, *Anal Chem*, 2021, **93**, 11321-11328.
60. C. Xu, S. P. Couvillion, R. L. Sontag, N. G. Isern, Y. Maezato, S. R. Lindemann, T. Roy Chowdhury, R. Zhao, B. R. Morton, R. K. Chu, R. J. Moore, J. K. Jansson, V. L. Bailey, P. J. Mouser, M. F. Romine, J. F. Fredrickson and T. O. Metz, *mSystems*, 2021, **6**, e0105820.
61. E. E. Carlson and B. F. Cravatt, *Nat Methods*, 2007, **4**, 429-435.
62. S. Zhao, H. Li, W. Han, W. Chan and L. Li, *Anal Chem*, 2019, **91**, 12108-12115.
63. J. N. Spradlin, E. Zhang and D. K. Nomura, *Acc Chem Res*, 2021, **54**, 1801-1813.
64. W. Lin, Z. Yang, A. Kaur, A. Block, M. Vujasinovic, J. M. Löhr and D. Globisch, *RSC Chem Biol*, 2021, **2**, 1479-1483.
65. G. Caspani, S. Kennedy, J. A. Foster and J. Swann, *Microb Cell*, 2019, **6**, 454-481.
66. J. Zeier, *Plant Cell Environ*, 2013, **36**, 2085-2103.
67. T. Santa, *Biomed Chromatogr*, 2011, **25**, 1-10.
68. W. C. Yang, H. Mirzaei, X. Liu and F. E. Regnier, *Anal Chem*, 2006, **78**, 4702-4708.
69. W. Lin, L. P. Conway, M. Vujasinovic, J. M. Löhr and D. Globisch, *Angew Chem Int Ed Engl*, 2021, **60**, 23232-23240.
70. M. Singh, A. Kapoor and A. Bhatnagar, *Chem Biol Interact*, 2015, **234**, 261-273.
71. C. C. Benz and C. Yau, *Nat Rev Cancer*, 2008, **8**, 875-879.
72. S. D. Yan, X. Chen, A. M. Schmidt, J. Brett, G. Godman, Y. S. Zou, C. W. Scott, C. Caputo, T. Frappier and M. A. Smith, *Proc Natl Acad Sci U S A*, 1994, **91**, 7787-7791.
73. P. Pinsky, L. Rabeneck and B. Lauby-Secretan, *N Engl J Med*, 2018, **379**, 301-302.
74. B. C. Dickinson and C. J. Chang, *Nat Chem Biol*, 2011, **7**, 504-511.

75. E. F. Holmquist, U. B Keiding, R. Kold-Christensen, T. Salomón, K. A. Jørgensen, P. Kristensen, T. B. Poulsen and M. Johannsen, *Anal Chem*, 2017, **89**, 5066-5071.
76. R. L. Auten and J. M. Davis, *Pediatr Res*, 2009, **66**, 121-127.
77. C. L. Hawkins and M. J. Davies, *J Biol Chem*, 2019, **294**, 19683-19708.
78. R. Nguyen and I. Huc, *Chem Commun (Camb)*, 2003, 942-943.
79. B. Dalile, L. Van Oudenhove, B. Vervliet and K. Verbeke, *Nat Rev Gastroenterol Hepatol*, 2019, **16**, 461-478.
80. J. Ding, T. Kind, Q. F. Zhu, Y. Wang, J. W. Yan, O. Fiehn and Y. Q. Feng, *Anal Chem*, 2020, **92**, 5960-5968.
81. D. J. Trader and E. E. Carlson, *Mol Biosyst*, 2012, **8**, 2484-2493.
82. K. Guo and L. Li, *Anal Chem*, 2009, **81**, 3919-3932.
83. B. D'Autréaux and M. B. Toledano, *Nat Rev Mol Cell Biol*, 2007, **8**, 813-824.
84. G. Wu, Y. Z. Fang, S. Yang, J. R. Lupton and N. D. Turner, *J Nutr*, 2004, **134**, 489-492.
85. L. Zheng, S. Cardaci, L. Jerby, E. D. MacKenzie, M. Sciacovelli, T. I. Johnson, E. Gaude, A. King, J. D. Leach, R. Edrada-Ebel, A. Hedley, N. A. Morrice, G. Kalna, K. Blyth, E. Ruppin, C. Frezza and E. Gottlieb, *Nat Commun*, 2015, **6**, 6001.
86. R. N. Carter and N. M. Morton, *J Pathol*, 2016, **238**, 321-332.
87. R. S. Patel, N. Ghasemzadeh, D. J. Eapen, S. Sher, S. Arshad, Y. A. Ko, E. Veledar, H. Samady, A. M. Zafari, L. Sperling, V. Vaccarino, D. P. Jones and A. A. Quyyumi, *Circulation*, 2016, **133**, 361-369.
88. C. E. Hand and J. F. Honek, *J Nat Prod*, 2005, **68**, 293-308.
89. H. Peng, W. Chen, Y. Cheng, L. Hakuna, R. Strongin and B. Wang, *Sensors (Basel)*, 2012, **12**, 15907-15946.
90. V. Liem-Nguyen, K. Huynh, C. Gallampois and E. Björn, *Anal Chim Acta*, 2019, **1067**, 71-78.
91. B. G. Hill, C. Reily, J. Y. Oh, M. S. Johnson and A. Landar, *Free Radic Biol Med*, 2009, **47**, 675-683.
92. A. Kaur, W. Lin, V. Dovhalyuk, L. Driutti, M. L. Di Martino, M. Vujasinovic, J. M. Löhr, M. E. Sellin and D. Globisch, *Chem Sci*, 2023, **14**, 5291-5301.
93. A. C. Conibear, *Nat Rev Chem*, 2020, **4**, 674-695.
94. J. D. Bargh, A. Isidro-Llobet, J. S. Parker and D. R. Spring, *Chem Soc Rev*, 2019, **48**, 4361-4374.
95. F. C. Schroeder, *Chem Biol*, 2015, **22**, 7-16.
96. M. J. Helf, B. W. Fox, A. B. Artyukhin, Y. K. Zhang and F. C. Schroeder, *Nat Commun*, 2022, **13**, 782.
97. J. S. Hoki, H. H. Le, K. E. Mellott, Y. K. Zhang, B. W. Fox, P. R. Rodrigues, Y. Yu, M. J. Helf, J. A. Baccile and F. C. Schroeder, *J Am Chem Soc*, 2020, **142**, 18449-18459.
98. J. Yu, M. C. Vogt, B. W. Fox, C. J. J. Wrobel, D. Fajardo Palomino, B. J. Curtis, B. Zhang, H. H. Le, A. Tauffenberger, O. Hobert and F. C. Schroeder, *Nat Chem Biol*, 2023, **19**, 141-150.

99. T. Salomón, C. Sibbersen, J. Hansen, D. Britz, M. V. Svart, T. S. Voss, N. Møller, N. Gregersen, K. A. Jørgensen, J. Palmfeldt, T. B. Poulsen and M. Johannsen, *Cell Chem Biol*, 2017, **24**, 935-943.e937.
100. R. Kold-Christensen and M. Johannsen, *Trends Endocrinol Metab*, 2020, **31**, 81-92.
101. C. Sibbersen, A. M. Schou Oxvig, S. Bisgaard Olesen, C. B. Nielsen, J. J. Galligan, K. A. Jørgensen, J. Palmfeldt and M. Johannsen, *ACS Chem Biol*, 2018, **13**, 3294-3305.
102. C. C. Le, M. Bae, S. Kiamehr and E. P. Balskus, *Annu Rev Biochem*, 2022, **91**, 475-504.
103. A. Y. M. Woo, M. A. Aguilar Ramos, R. Narayan, K. C. Richards-Corke, M. L. Wang, W. J. Sandoval-Espinola and E. P. Balskus, *Nat Rev Chem*, 2023, **7**, 319-339.
104. D. J. Kenny and E. P. Balskus, *Chem Soc Rev*, 2018, **47**, 1705-1729.
105. A. Chokkathukalam, D. H. Kim, M. P. Barrett, R. Breitling and D. J. Creek, *Bioanalysis*, 2014, **6**, 511-524.
106. S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb and R. Aebersold, *Nat Biotechnol*, 1999, **17**, 994-999.
107. S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann, *Mol Cell Proteomics*, 2002, **1**, 376-386.
108. M. Wang, C. Wang and X. Han, *Mass Spectrom Rev*, 2017, **36**, 693-714.
109. P. Panuwet, R. E. Hunter, P. E. D'Souza, X. Chen, S. A. Radford, J. R. Cohen, M. E. Marder, K. Kartavenka, P. B. Ryan and D. B. Barr, *Crit Rev Anal Chem*, 2016, **46**, 93-105.
110. C. Jang, L. Chen and J. D. Rabinowitz, *Cell*, 2018, **173**, 822-837.
111. A. Zhang, H. Sun, G. Yan, P. Wang and X. Wang, *Biomed Res Int*, 2015, **2015**, 354671.
112. C. B. Clish, *Cold Spring Harb Mol Case Stud*, 2015, **1**, a000588.
113. R. D. Beger, W. Dunn, M. A. Schmidt, S. S. Gross, J. A. Kirwan, M. Cascante, L. Brennan, D. S. Wishart, M. Oresic, T. Hankemeier, D. I. Broadhurst, A. N. Lane, K. Suhre, G. Kastenmüller, S. J. Sumner, I. Thiele, O. Fiehn, R. Kaddurah-Daouk and f. P. M. a. P. T. G.-M. S. Initiative, *Metabolomics*, 2016, **12**, 149.
114. 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, *Anal Chem*, 2019, **91**, 3590-3596.
115. B. Li, J. Tang, Q. Yang, S. Li, X. Cui, Y. Li, Y. Chen, W. Xue, X. Li and F. Zhu, *Nucleic Acids Res*, 2017, **45**, W162-W170.
116. B. Li, J. Tang, Q. Yang, X. Cui, S. Li, S. Chen, Q. Cao, W. Xue, N. Chen and F. Zhu, *Sci Rep*, 2016, **6**, 38881.
117. A. K. Boysen, K. R. Heal, L. T. Carlson and A. E. Ingalls, *Anal Chem*, 2018, **90**, 1363-1369.
118. M. Sysi-Aho, M. Katajamaa, L. Yetukuri and M. Oresic, *BMC Bioinformatics*, 2007, **8**, 93.
119. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, *Anal Chem*, 2006, **78**, 779-787.

120. L. P. Conway, N. Garg, W. Lin, M. Vujasinovic, J. M. Löhr and D. Globisch, *Chem Commun (Camb)*, 2019, **55**, 9080-9083.
121. W. Lin, L. P. Conway, A. Block, G. Sommi, M. Vujasinovic, J. M. Löhr and D. Globisch, *Analyst*, 2020, **145**, 3822-3831.
122. W. Lin, F. R. García, E. L. Norin, D. Kart, L. Engstrand, J. Du and D. Globisch, *Chem Commun (Camb)*, 2023, **59**, 5843-5846.
123. W. Lin, K. Mellinghaus, A. Rodriguez-Mateos and D. Globisch, *Food Chem*, 2023, **425**, 136481.
124. N. Garg, L. P. Conway, C. Ballet, M. S. P. Correia, F. K. S. Olsson, M. Vujasinovic, J. M. Löhr and D. Globisch, *Angew Chem Int Ed Engl*, 2018, **57**, 13805-13809.
125. K. Dührkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu and S. Böcker, *Nat Methods*, 2019, **16**, 299-302.
126. R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T. S. Dyrlund, R. Bushuiev, K. J. Murray, N. Hoffmann, M. Lu, A. Sarvepalli, Z. Zhang, M. Fleischauer, K. Dührkop, M. Wesner, S. J. Hoogstra, E. Rudt, O. Mokshyna, C. Brungs, K. Ponomarov, L. Mutabdžija, T. Damiani, C. J. Pudney, M. Earll, P. O. Helmer, T. R. Fallon, T. Schulze, A. Rivas-Ubach, A. Bilbao, H. Richter, L. F. Nothias, M. Wang, M. Orešič, J. K. Weng, S. Böcker, A. Jeibmann, H. Hayen, U. Karst, P. C. Dorrestein, D. Petras, X. Du and T. Pluskal, *Nat Biotechnol*, 2023, **41**, 447-449.
127. C. Guijas, J. R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A. E. Aisporna, D. W. Wolan, M. E. Spilker, H. P. Benton and G. Siuzdak, *Anal Chem*, 2018, **90**, 3156-3164.
128. A. Parker, S. Fonseca and S. R. Carding, *Gut Microbes*, 2020, **11**, 135-157.
129. C. Y. Lim, N. A. Owens, R. D. Wampler, Y. Ying, J. H. Granger, M. D. Porter, M. Takahashi and K. Shimazu, *Langmuir*, 2014, **30**, 12868-12878.
130. O. Koniev and A. Wagner, *Chem Soc Rev*, 2015, **44**, 5495-5551.
131. F. R. Wurm and H. A. Klok, *Chem Soc Rev*, 2013, **42**, 8220-8236.
132. D. Parada Venegas, M. K. De la Fuente, G. Landskron, M. J. González, R. Quera, G. Dijkstra, H. J. M. Harmsen, K. N. Faber and M. A. Hermoso, *Front Immunol*, 2019, **10**, 277.
133. A. J. Wolfe, *Microbiol Mol Biol Rev*, 2005, **69**, 12-50.
134. C. Di Lorenzo, F. Colombo, S. Biella, C. Stockley and P. Restani, *Nutrients*, 2021, **13**.
135. P. C. Hollman and M. B. Katan, *Free Radic Res*, 1999, **31 Suppl**, S75-80.
136. P. C. Hollman and M. B. Katan, *Food Chem Toxicol*, 1999, **37**, 937-942.
137. J. Oracz, E. Nebesny, D. Zyzelewicz, G. Budryn and B. Luzak, *Crit Rev Food Sci Nutr*, 2020, **60**, 1947-1985.
138. T. Ozdal, D. A. Sela, J. Xiao, D. Boyacioglu, F. Chen and E. Capanoglu, *Nutrients*, 2016, **8**, 78.
139. D. Plamada and D. C. Vodnar, *Nutrients*, 2021, **14**.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2305

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-510431

ACTA UNIVERSITATIS
UPSALIENSIS
2023