Metabolomics insight into the gut microbiome of infants with cow's milk allergy

Zhu, P.

Metabolomics Insight into the Gut Microbiome of Infants with Cow's Milk Allergy

Pingping Zhu

朱萍萍

The publication of the thesis was financially supported by:

Leiden University Libraries

Cover design: **Tingting Zhu & Yan Zhu**

Thesis lay-out: Pingping Zhu

Printing: Ridderprint | www.ridderprint.nl

ISBN: 978-94-6537-050-7

Metabolomics Insight into the Gut Microbiome of Infants with Cow's Milk Allergy

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 13 januari 2026
klokke 10:00 uur

door

Pingping Zhu 朱萍萍

Geboren te Chongqing, China
in 1992

Promotor

Prof. dr. T. Hankemeier

Co-promotor(es)

Dr. A.C. Harms

Dr. ir. A.C. Dubbelman

Promotiecommissie

Prof. dr. M. van Eck

Prof. dr. E.C.M. de Lange

Prof. dr. H.P. Spaink

Dr. R.C. van Wijk

Dr. J. Kirwan

University of Veterinary Medicine Vienna, Austria

Prof. dr. A. Zhernakova

University of Groningen, the Netherlands

The research described in this thesis was performed at Metabolomics and Analytics Center (MAC) of the Leiden Academic Centre for Drug Research (LACDR), Leiden University (Leiden, The Netherlands). The research was financially supported as indicated in each chapter.

Contents

Chapter I

General Introduction and Scope **1**

Chapter II

Development of an untargeted LC-MS metabolomics method with post-column infusion for matrix effect monitoring in plasma and feces **19**

Chapter III

Matrix effects in untargeted LC-MS metabolomics: from creation to compensation with post-column infusion of standards **89**

Chapter IV

Current insights into cow's milk allergy in children: microbiome, metabolome and immune response – a systematic review **135**

Chapter V

Exploring the fecal metabolome in infants with cow's milk allergy: The distinct impacts of cow's milk protein tolerance acquisition and of synbiotic supplementation **175**

Chapter VI

Conclusions and Perspectives **217**

Appendix

Summary **232**

Samenvatting **236**

Curriculum vitae **240**

List of Publications **241**

Acknowledgements **243**

Chapter I

General Introduction and Scope

The prevalence of food allergy has risen over recent decades, with early life recognized as a critical window for its development.[1] However, the mechanisms driving the onset and resolution of food allergy remain incompletely understood. Increasing evidence highlights the gut microbiome, which exerts a dynamic impact on the systematic immune system, as an importance player in regulating these processes during early life.[2] The complex interactions between the gut microbiome and host immunity are gradually being deciphered along with the advances in molecular technologies, such as metagenomics, metatranscriptomics, and metabolomics. Among these tools, metabolomics plays a crucial role by capturing microbial activity from a metabolome perspective, offering valuable insights into microbiome-host interactions. Continued advancement in metabolomics techniques contributes to revealing the cross-talk between the gut microbiome and the host, deepening our understanding of how the gut microbiome influences food allergy in early life.

1. LC-ESI-MS based metabolomics

Along with genomics, transcriptomics, and proteomics, metabolomics is one of the omics strategies applied in systems biology, and the combination of these omics strategies provides a general view of how genotype is linked to phenotype (Figure 1).[3] Metabolites, which are the end products of cellular regulatory processes, reflect the ultimate response of a biological system to genetic or environmental changes.[4] The complete set of metabolites in a biological system is described as "metabolome",[5] which was firstly introduced by Oliver et al. in 1998.[6] Metabolomics is an approach to reveal the metabolome of a studied biological system.[3] This trait makes metabolomics a popular and significant strategy for monitoring ongoing biological processes in an organism.[7] In recent decades, metabolomics has been widely applied in biological studies, especially in the diagnosis, treatment, and prognosis of human diseases.[8–11] The popularity of metabolomics has been greatly enhanced by the emergence of advanced analytical techniques, such as nuclear magnetic resonance (NMR) and mass spectrometry (MS). NMR can quantify organic compounds and provide unbiased metabolite profiles for certain biological samples, but it has rather low sensitivity compared to MS.[12] The high sensitivity of MS is largely due to the breakthroughs in MS

technologies, particularly the development of ionization sources.[13]

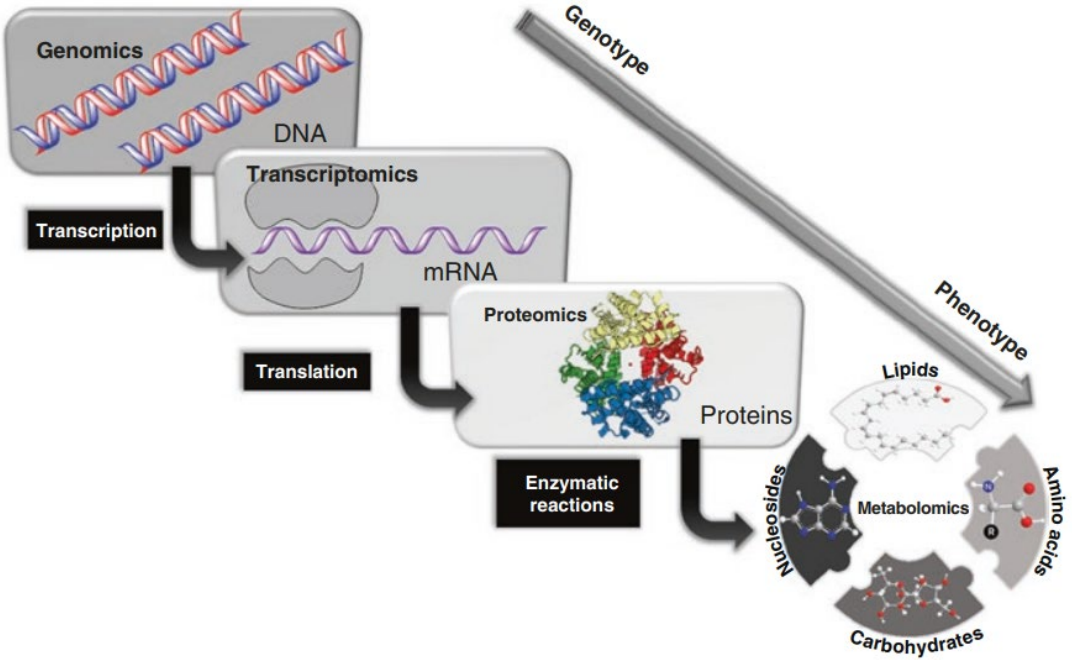


Figure 1. A correlation between the main omics strategies used in systems biology studies. From Klassen *et.al* 2017.[3]

MS is often coupled with various chromatographic separation techniques, such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE). Before the 1980s, coupling GC to MS with an electron ionization (EI) source was the dominant technology for metabolome profiling for decades.[13] However, EI has limitations due to its requirement for high-vacuum and high-temperature conditions, as well as the need for samples to be delivered in gas phase.[14] These constraints restricts its applicability to couple other separation techniques, such as LC and CE, to MS. The exclusive use of EI declined with the development of advanced ionization techniques, such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI).[13] These ionization sources not only allow the detection of intact molecules as "soft" ionization sources, but are also capable of producing stable, gaseous ionized molecules directly from liquid phase, making them perfectly compatible with LC or CE.[13] The ESI source, initially invented by Dole *et al.* in 1968[15] and further developed by Fenn *et al.*,[16–18] is considered a turning point in advancing the application of LC-MS in life science, including metabolomics. The advantages of the ESI source lie in its versatility, sensitivity, high ionization efficiency, and capability of ionizing molecules over a large mass range.[13,14] However, due to its ionization mechanism, the ESI source is more susceptible to matrix effect, particularly ion

suppression.[13,19]

1.1 Matrix effect in the ESI source

The simplified ionization process in an ESI source is as follows: (1) A liquid sample is delivered from the LC to the spray needle, where an intense electric field is generated at the tip, with an electric potential ranging from hundreds to thousands of volts. (2) The strong electric field at the tip of the spray needle forms a Taylor cone, from which a fine spray of charged droplets is emitted. (3) Droplets evaporate under dry gas and heat, causing them to shrink and the charge density increases on their surface until reaching the Rayleigh limit, where the coulombic repulsion counterbalances the surface tension. (4) Ions are ejected from the droplet or released through coulombic explosion when the coulombic repulsion overcomes the surface tension. Through these steps, the gas-phase ions are generated by the ESI process, allowing for MS analysis.[13,20,21] During this process, matrix components that interfere with any of these ionization steps can impact the ion intensity of analytes.[19,22] Figure 2 illustrates the potential mechanisms of matrix effect during the ESI process: (1) In the desolvation process, matrix components can prevent the analyte from accessing the available charge on the surface of the droplets and/or increase the viscosity and surface tension of the droplet, inhibiting further coulombic explosion. (2) During the coulombic explosion, matrix components can compete with the analytes for charge acquisition. (3) After reaching the gas phase, matrix compounds can neutralize or destabilize the charged ions. (4) The analytes can co-precipitate with nonvolatile matrix compounds, reducing the likelihood of their transfer to the gas phase. As a result, the reproducibility, linearity, selectivity, sensitivity, and accuracy of analyte detection can be significantly affected by matrix effect when using LC-ESI-MS-based methods for quantification.[19,23]

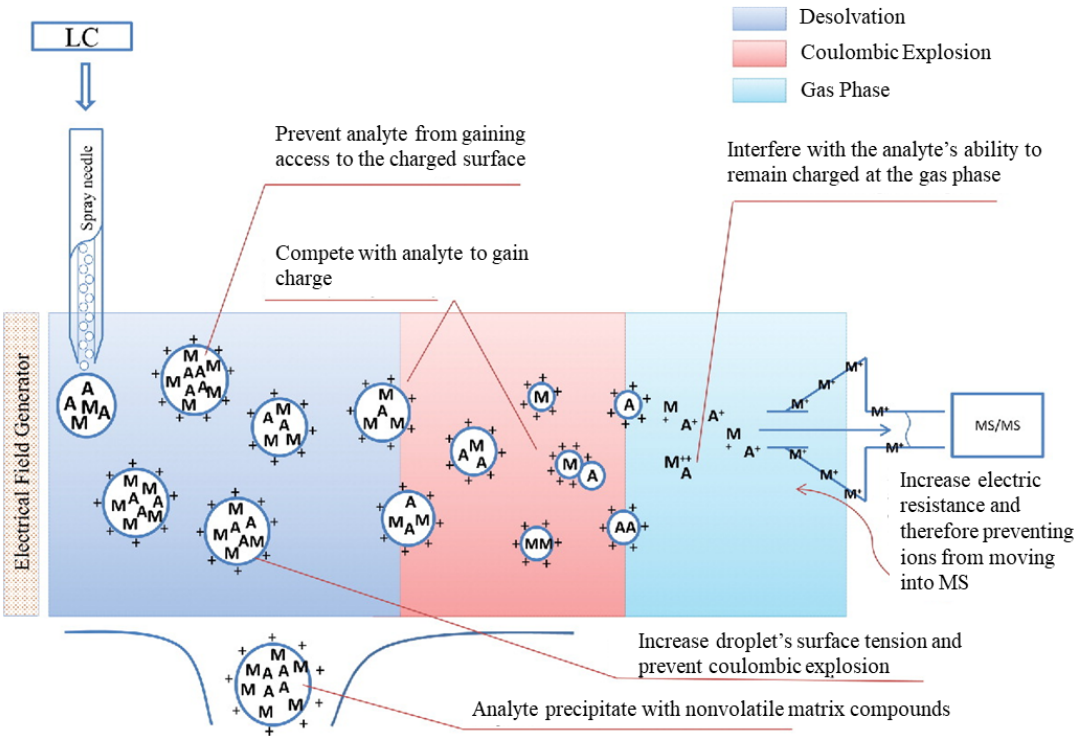


Figure 2. Mechanisms of how matrix components (M) can affect the ionization of analytes (A) in the electrospray ionization (ESI) source. Adapted From Panuwet *et al.* 2016.[19]

1.2 Approaches for addressing matrix effect

Various strategies are employed to minimize and compensate for matrix effect in LC-ESI-MS-based metabolomics studies. In general, matrix effect can be reduced through extensive sample cleanup procedures, tailored LC separation, matrix dilution, and reduced injection volumes.[23,24] Beyond reduction strategies, matrix effect can be assessed using post-extraction spiking (PES) of stable isotopically labeled (SIL) standards and a post-column infusion of standard (PCIS). PES, proposed by Matuszewski *et al.*, is a quantitative approach that evaluates matrix effect by comparing the responses of standards spiked into matrix samples versus matrix-free samples.[25] The other approach, PCIS, introduced by Bonfiglio *et al.* and Choi *et al.* in 1999,[26,27] provides a qualitative assessment of matrix effect by comparing the signals of a post-column infused standard observed with the injections of matrix samples to those of matrix-free samples. As shown in Figure 3, PCIS involves continuously infusing a standard solution via a pump or syringe after separation, then merging it with the LC flow using a T-connector before being injected into the MS. Unlike PES, which assesses

matrix effect at specific retention times,[25] PCIS evaluates it over the entire chromatogram.[26,27] Therefore, PCIS has been recommended as a quality control tool for assessing matrix effect in both targeted and untargeted LC-MS-based metabolomics.[28]

The primary objective of matrix effect evaluation is to identify analytes significantly affected by matrix effect and implement appropriate compensation strategies. In targeted metabolomics, where specific metabolites from known classes are precisely identified and quantified,[3,29] matrix effect compensation is typically achieved by correcting the signal of a target using a surrogate analogue, usually a SIL standard, spiked into the same study sample.[30] Different from targeted metabolomics, untargeted metabolomics aims to profile the metabolome, covering a wide range of metabolites, including unknowns.[3,29] This characteristic makes compensating for matrix effect particularly challenging. Although PCIS, a technique independent of retention time, is a feasible approach for correcting matrix effect in untargeted metabolomics, its application has been rarely reported.[31] Given the importance of untargeted metabolomics in biomarker discovery across diverse fields, such as biomedical research,[32] agriculture,[33] food,[34] and environmental science,[35] addressing the matrix effect in untargeted metabolomics can greatly improve data reliability and expand its applications.

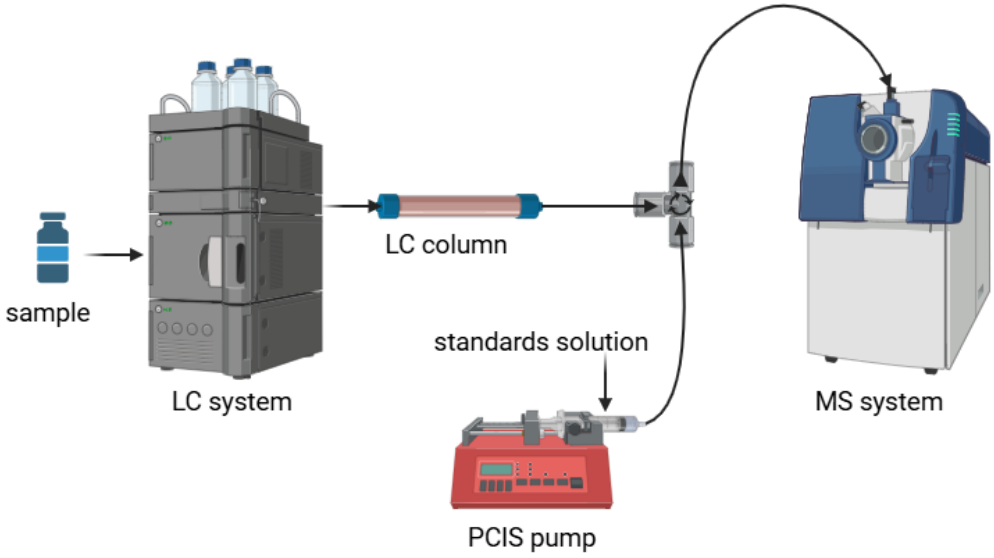


Figure 3. Setup of post-column infusion of standards (PCIS) with LC-MS (Created in https://BioRender.com)

2. Metabolomics and the gut microbiome

The human microbiome was described by Lederberg and McCray in 2001 as "the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space".[36] Our understanding of the human microbiome advanced significantly following the launch of the Human Microbiome Project (HMP) in 2007, an initiative founded by the U.S. National Institutes of Health. The HMP brought together international scientific experts to characterize the human microbial communities and investigate their roles in health and disease.[37] In HMP, biological samples were collected from 15 and 18 body sites in male and female, respectively, across more than 200 donors.[37] Among these sites, the human gut, which harbors the majority of microbes in the body,[38] was found to have an especially diverse microbiota community.[39] This community comprises bacteria, fungi, protists, archaea, and viruses, with bacteria making up around 60% of the dry mass of feces.[40] More than 500 bacterial species colonize the gut of a healthy adult,[38] primarily belonging to the phyla Firmicutes and Bacteroidetes, followed by Proteobacteria, Actinobacteria, Fusobacteria, and Verrucomicrobia.[41]

The co-evolution of the gut microbiome and its human host was initially described as commensal. However, it was later considered more accurate to use the term "mutualistic", which reflects the reciprocal influence and benefits shared between the host and the gut microbiome.[42] The gut microbe is increasingly recognized as a metabolically active "organ" with diverse functions,[43] including fermenting undigested food components, synthesizing essential vitamins, detoxifying harmful compounds, strengthening the intestinal barrier, and regulating the immune system.[44] These functions are tightly interconnected with the host, making gut microbiome a crucial player in human health and disease. Gut microbiota dysbiosis has been observed in many diseases, such as irritable bowel syndrome, inflammatory bowel disease, metabolic dysfunction-associated steatotic liver disease, obesity, diabetes, cardiovascular diseases, colorectal cancer, allergic disease, neurological and psychiatric disorders.[40,45,46] Although the mechanisms underlying the interplay between the gut microbiome and human physiology remain complex, gut microbiome-derived metabolites are believed

to play a critical role in the development and progression of various health conditions.

Figure 4 illustrates examples of well-known gut microbiota-derived metabolites identified over the past decades. The metabolites are primarily generated through three main pathways: (1) digesting dietary compounds (Figure 4a), (2) modifying host-derived metabolites (Figure 4b), and (3) synthesizing them *de novo* (Figure 4c).[47] One major class of gut microbiota-derived metabolites is short chain fatty acids (SCFAs), including formate, acetate, propionate, and butyrate, which are produced via microbial fermentation of undigested carbohydrates in the colon.[48] Another key group involves metabolites derived from an essential amino acid, tryptophan. Microbes in the colon can convert tryptophan into multiple bioactive compounds, including indole, indolepropionic acid (IPA), indole lactic acid (ILA), indoleacetic acid (IAA), indole ethanol (IE), indolealdehyde (IAld), indoleacrylic acid (IA), skatole, and tryptamine.[49]Additionally, gut microbes can metabolize dietary choline, betaine, and L-carnitine to produce trimethylamine (TMA), which can be absorbed in the intestine and subsequently oxidized in the liver to form trimethylamine N-oxide (TMAO).[50] Gut microbiota also play a crucial role in bile acid metabolism. Unconjugated primary bile acids, such as cholic acid (CA) and chenodeoxycholic acid (CDCA), are initially synthesized in the liver from cholesterol and stored in the gallbladder.[51] Upon food intake, they are released into the gut, where certain microbes can convert them into secondary bile acids, primarily deoxycholic acid (DCA) and lithocholic acid (LCA).[51] Apart from metabolizing dietary and host-derived substances, gut microbes are also capable of *de novo* synthesis of important metabolites, such as branched-chain amino acids (BCAAs),[52] polyamines,[53] and vitamins.[54]

Most gut microbiome-derived metabolites play crucial roles in host physiology. For instance, SCFAs are reported to have anti-inflammatory and anti-tumor properties,[48] TMAO has been identified as a predictor of cardiovascular disease pathogenesis,[50,55] and certain secondary bile acids are known as signaling molecules that regulate host endocrine functions.[56] Given the diverse biological functions of these metabolites, integrating metabolomics with other omics approaches, such as metagenomics, metatranscriptomics, and metaproteomics, is essential for gaining deeper insights into

the cross-talk between the gut microbiome and host.

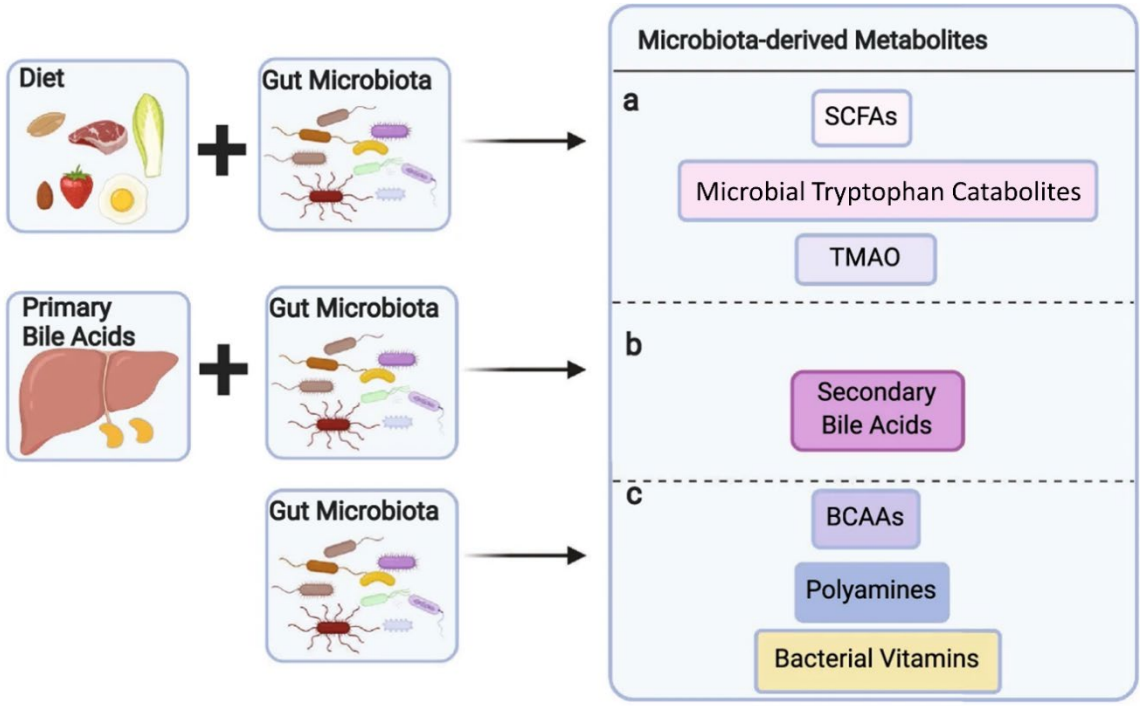


Figure 4. Production of some well-known gut microbiome-derived metabolites. SCFAs: short-chain fatty acids, TMAO: trimethylamine-N-oxide, BCAAs: branched-chain amino acids. From Yang *et al.*[47]

3. The gut microbiome and food allergy in early life

Food allergy is defined as "an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food".[57] It has become a growing global public health concern, particularly in children.[58–61] The estimated prevalence of food allergy is 1-2% in the general population,[62] but rises to 6-8% in children.[63] Most food allergies develop within the first few years of life[1] with major allergens including peanut, tree nuts, fish, shellfish, egg, milk, wheat, soy, fruits, and seeds.[58,61] Among these, cow's milk is the most common food allergen in early childhood.[64,65] It is reported that, in the United States, cow's milk allergy affects approximately 50 % of food-allergic children under one year-old, 40 % of those aged 1-2 years, and 30% of those aged 3-5 years.[59,65]

Numerous factors, including genetics, diet, and environmental influences, contribute to the development and resolution of food allergy.[66,67] Growing evidence also highlights

the gut microbiome as a key player in food allergy of early life. The initial recognition of the role of gut microbiome in allergic diseases dates back to the late 1980s with the proposition of "hygiene hypothesis". It suggests that reduced exposure to infection sources and symbiotic microorganisms may lead to increased rates of allergic diseases.[68] This concept was later extended into the "microflora hypothesis" in 2005, which specifically proposed that the disruptions in the gastrointestinal microbiota during early life impair microbiota-mediated mechanisms of immunological tolerance, thereby increasing the incidence of allergic diseases.[69] Multiple clinical studies have reported altered gut microbiota composition in children with food allergies. For example, Joseph *et al.* observed that children aged 3–5 years with food allergies had significantly lower gut microbiota diversity compared to non-allergic children.[70] Similarly, Japanese children who developed food allergies within their first two years exhibited lower abundances of the bacterial genera *Leuconostoc*, *Weissella*, and *Veillonella* compared to their healthy counterparts.[71] Additionally, reduced levels of *Citrobacter*, *Oscillospira*, *Dorea*, and *Lactococcus* genera in the fecal samples of infants aged 3–6 months have been associated with food allergy development by age three.[72] In contrast, a higher abundance of bacteria from the Firmicutes phyla in infancy (3-6 months) has been linked to the resolution of cow's milk allergy.[73]

As growing evidence supports the role of the gut microbiome in both the development and resolution of food allergy in early life, interest has increased in strategies to modulate its composition and function as a means of preventing and managing food allergy.[74] The gut microbiome can be modified through the administration of probiotics, prebiotics, synbiotics, and fecal microbiome transplantation (FMT).[75] Probiotics, which consist of beneficial live bacteria strains primarily from the *Lactobacillaceae* and *Bifidobacteriaceae* families, aim to directly alter gut microbiota composition and potentially restore microbial balance.[76,77] Prebiotics, on the other hand, are defined as non-digestible food ingredients that can be fermented by gut microbiome, selectively stimulating the growth and/or activity of specific beneficial bacteria.[78] Common prebiotics include fructo-oligosaccharides, galacto-oligosaccharides, and inulin.[74] Synbiotics combine probiotics and prebiotics to enhance the survival and efficacy of probiotic strains.[79] FMT, a more direct approach, is a procedure that transplants fecal

microbiota from a healthy donor to reshape the recipient's gut microbiome.[80] Compared to FMT, probiotics, prebiotics, and synbiotics are more commonly used for food allergy intervention.[76] However, despite their promising potential, clinical evidence supporting the effectiveness of probiotics and/or prebiotics in preventing or treating food allergy remains limited.[76,81,82] Therefore, further research is needed to deepen our understanding of the gut microbiome's role in food allergy and to explore the therapeutic potential of microbiome-targeted interventions.

4. Scope and outline of this thesis

As the intricate relationship between the gut microbiome and food allergy in early life continues to be deciphered, metabolomics offers a powerful tool to explore this cross-communication at the molecular level. Among analytical methods applied in metabolomics, untargeted methods outperform targeted ones in identifying novel metabolites, including those derived from the gut microbiota. One major challenge in untargeted metabolomics is the matrix effect, which can vary between samples, especially those with complex matrices, such as feces. The first hypothesis of this thesis is that the matrix effect in untargeted metabolomics can be monitored and corrected by implementing the PCIS technique with LC-MS methods. The second hypothesis is that the fecal metabolome can provide insights into the cross-talk between the gut microbiome and food allergy in infants with the most prevalent type of food allergy in early life: cow's milk allergy (CMA).

The first hypothesis is examined and verified in **Chapters 2 & 3**. In **chapter 2**, the goal is to develop an untargeted LC-ESI-MS method with PCIS to monitor matrix effect in plasma and fecal samples. To achieve this goal, the first focus is on developing a reverse-phase LC-MS untargeted metabolomics method with PCIS to profile polar to semi-polar metabolites. The development includes injection parameters optimization and validating the method with representative SIL standards. Then, the SIL standards are used to evaluated the capability of PCIS in monitoring the matrix effect of plasma and fecal samples.

As a follow-up, **Chapter 3** aims to investigate the application of PCIS for matrix effect

compensation in untargeted metabolomics. To fulfill this aim, a post-column artificial matrix infusion approach is introduced to the developed LC-PCIS-MS method. This artificial matrix consists of several compounds that are known to disrupt the ionization process of ESI, creating an artificial matrix effect. The artificial matrix effect of a given feature can be determined by comparing its signals obtained with and without artificial matrix infusion. The hypothesis is that the artificial matrix effect can help identify a suitable PCIS for a given feature, and that the selected PCIS can be used to correct the matrix effect of that feature in biological samples. This concept is investigated by comparing the ideal PCISs selected based on compensating artificial and biological matrix effect for diverse SIL standards spiked into plasma, urine, and feces.

In the following two chapters, the focus is on investigating the relationship between the gut microbiome and CMA in early life with the developed untargeted method. To gain insights into this problem, a systematic review is conducted in **Chapter 4**, summarizing existing studies on the microbiome, metabolome, and immune response in CMA children and animal models, including the impacts of interventions with probiotics, prebiotics, and synbiotics. The review highlights a lack of studies on immune responses and metabolomics related to CMA in early life, emphasizing the need for further research in these fields.

The aim of **Chapter 5** is to help address the research gap identified in Chapter 4 concerning CMA in early life. A comprehensive exploration of the fecal metabolome is performed in infants (3-13 months) with CMA by combining the untargeted metabolomics platform developed in Chapter 2 with an additional in-house platform focused on non-polar metabolites. The study includes 39 infants with cow's milk allergy, who were randomized into two intervention groups: one group receiving amino acid-based formula (AFF) and the other group receiving AAF supplemented with synbiotics (inulin, oligofructose, *Bifidobacterium breve* M-16 V) (AAF-S). Fecal samples from all the infants were collected at baseline, as well as six and 12 months after the start of the interventions. By categorizing the infants based on their intervention strategy and cow's milk allergy status after 12-month intervention, the aim is to investigate: (1) the impact of synbiotic supplementation on the fecal metabolome in

infants with cow's milk allergy, and (2) the effect of tolerance acquisition on the fecal metabolome in the infants initially diagnosed with CMA.

This thesis concludes in **Chapter 6** with a general summary and discussion. In this chapter, potential improvements in implementing PCIS to address matrix effect in untargeted metabolomics is discussed, along with recommendations and perspectives on applying metabolomics to investigate the gut microbiome and CMA in early life.

Reference:

1. Hong, X.; Wang, X. Early Life Precursors, Epigenetics, and the Development of Food Allergy. *Semin Immunopathol* **2012**, *34* (5), 655–669. https://doi.org/10.1007/s00281-012-0323-y.
2. Marrs, T.; Jo, J.-H.; Perkin, M. R.; Rivett, D. W.; Witney, A. A.; Bruce, K. D.; Logan, K.; Craven, J.; Radulovic, S.; Versteeg, S. A.; van Ree, R.; McLean, W. H. I.; Strachan, D. P.; Lack, G.; Kong, H. H.; Flohr, C. Gut Microbiota Development during Infancy: Impact of Introducing Allergenic Foods. *Journal of Allergy and Clinical Immunology* **2021**, *147* (2), 613-621.e9. https://doi.org/10.1016/j.jaci.2020.09.042.
3. Klassen, A.; Faccio, A. T.; Canuto, G. A. B.; Da Cruz, P. L. R.; Ribeiro, H. C.; Tavares, M. F. M.; Sussulini, A. Metabolomics: Definitions and Significance in Systems Biology. In *Metabolomics: From Fundamentals to Clinical Applications*; Sussulini, A., Ed.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2017; Vol. 965, pp 3–17. https://doi.org/10.1007/978-3-319-47656-8_1.
4. *Functional Genomics*; Town, C., Ed.; Springer Netherlands: Dordrecht, 2002. https://doi.org/10.1007/978-94-010-0448-0.
5. Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics* **2001**, *2* (3), 155–168. https://doi.org/10.1002/cfg.82.
6. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. Systematic Functional Analysis of the Yeast Genome. *Trends in Biotechnology* **1998**, *16* (9), 373–378. https://doi.org/10.1016/S0167-7799(98)01214-1.
7. Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metabolism* **2007**, *6* (5), 348–351. https://doi.org/10.1016/j.cmet.2007.10.005.
8. Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G. Clinical Applications of Metabolomics in Oncology: A Review. *Clinical Cancer Research* **2009**, *15* (2), 431–440. https://doi.org/10.1158/1078-0432.CCR-08-1059.
9. Tounta, V.; Liu, Y.; Cheyne, A.; Larrouy-Maumus, G. Metabolomics in Infectious Diseases and Drug Discovery. *Mol. Omics* **2021**, *17* (3), 376–393. https://doi.org/10.1039/D1MO00017A.
10. Schmidt, D. R.; Patel, R.; Kirsch, D. G.; Lewis, C. A.; Vander Heiden, M. G.; Locasale, J. W. Metabolomics in Cancer Research and Emerging Applications in Clinical Oncology. *CA: A Cancer Journal for Clinicians* **2021**, *71* (4), 333–358. https://doi.org/10.3322/caac.21670.
11. Alarcon-Barrera, J. C.; Kostidis, S.; Ondo-Mendez, A.; Giera, M. Recent Advances in Metabolomics Analysis for Early Drug Development. *Drug Discovery Today* **2022**, *27* (6), 1763–1773. https://doi.org/10.1016/j.drudis.2022.02.018.
12. Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. Modern Analytical Techniques in Metabolomics Analysis. *Analyst* **2012**, *137* (2), 293–300. https://doi.org/10.1039/C1AN15605E.
13. Siuzdak, G. *Activity Metabolomics and Mass Spectrometry*; MCC Press, 2024. https://doi.org/10.63025/LCUW3037.
14. Famiglini, G.; Palma, P.; Termopoli, V.; Cappiello, A. The History of Electron Ionization in LC-MS, from the Early Days to Modern Technologies: A Review. *Analytica Chimica Acta* **2021**, *1167*, 338350. https://doi.org/10.1016/j.aca.2021.338350.
15. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *The Journal of Chemical Physics* **1968**, *49* (5), 2240–2249. https://doi.org/10.1063/1.1670391.
16. Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.
17. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, Masamichi.; Fenn, J. B. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal. Chem.* **1985**, *57* (3), 675–679. https://doi.org/10.1021/ac00280a023.

18. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71. https://doi.org/10.1126/science.2675315.
19. Panuwet, P.; Hunter Jr., R. E.; D'Souza, P. E.; Chen, X.; Radford, S. A.; Cohen, J. R.; Marder, M. E.; Kartavenka, K.; Ryan, P. B.; Barr, D. B. Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring. *Critical Reviews in Analytical Chemistry* **2016**, *46* (2), 93–105. https://doi.org/10.1080/10408347.2014.980775.
20. Smith, J. N.; Flagan, R. C.; Beauchamp, J. L. Droplet Evaporation and Discharge Dynamics in Electrospray Ionization. *J. Phys. Chem. A* **2002**, *106* (42), 9957–9967. https://doi.org/10.1021/jp025723e.
21. Wilm, M. Principles of Electrospray Ionization. *Molecular & Cellular Proteomics* **2011**, *10* (7). https://doi.org/10.1074/mcp.M111.009407.
22. King, R.; Bonfiglio, R.; Fernandez-Metzler, C.; Miller-Stein, C.; Olah, T. Mechanistic Investigation of Ionization Suppression in Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (11), 942–950. https://doi.org/10.1016/S1044-0305(00)00163-X.
23. Cortese, M.; Gigliobianco, M. R.; Magnoni, F.; Censi, R.; Di Martino, P. Compensate for or Minimize Matrix Effects? Strategies for Overcoming Matrix Effects in Liquid Chromatography-Mass Spectrometry Technique: A Tutorial Review. *Molecules* **2020**, *25* (13), 3047. https://doi.org/10.3390/molecules25133047.
24. Raposo, F.; Barceló, D. Challenges and Strategies of Matrix Effects Using Chromatography-Mass Spectrometry: An Overview from Research versus Regulatory Viewpoints. *TrAC Trends in Analytical Chemistry* **2021**, *134*, 116068. https://doi.org/10.1016/j.trac.2020.116068.
25. Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC−MS/MS. *Anal. Chem.* **2003**, *75* (13), 3019–3030. https://doi.org/10.1021/ac020361s.
26. Bonfiglio, R.; King, R. C.; Olah, T. V.; Merkle, K. The Effects of Sample Preparation Methods on the Variability of the Electrospray Ionization Response for Model Drug Compounds. *Rapid Communications in Mass Spectrometry* **1999**, *13* (12), 1175–1185. https://doi.org/10.1002/(SICI)1097-0231(19990630)13:12<1175::AID-RCM639>3.0.CO;2-0.
27. Choi, B. K.; Gusev, A. I.; Hercules, D. M. Postcolumn Introduction of an Internal Standard for Quantitative LC−MS Analysis. *Anal. Chem.* **1999**, *71* (18), 4107–4110. https://doi.org/10.1021/ac990312o.
28. González, O.; Dubbelman, A.-C.; Hankemeier, T. Postcolumn Infusion as a Quality Control Tool for LC-MS-Based Analysis. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (6), 1077–1080. https://doi.org/10.1021/jasms.2c00022.
29. Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.
30. Thakare, R.; Chhonker, Y. S.; Gautam, N.; Alamoudi, J. A.; Alnouti, Y. Quantitative Analysis of Endogenous Compounds. *Journal of Pharmaceutical and Biomedical Analysis* **2016**, *128*, 426–437. https://doi.org/10.1016/j.jpba.2016.06.017.
31. Tisler, S.; Pattison, D. I.; Christensen, J. H. Correction of Matrix Effects for Reliable Non-Target Screening LC–ESI–MS Analysis of Wastewater. *Anal. Chem.* **2021**, *93* (24), 8432–8441. https://doi.org/10.1021/acs.analchem.1c00357.
32. Pang, H.; Jia, W.; Hu, Z. Emerging Applications of Metabolomics in Clinical Pharmacology. *Clinical Pharmacology & Therapeutics* **2019**, *106* (3), 544–556. https://doi.org/10.1002/cpt.1538.
33. Bedair, M.; Glenn, K. C. Evaluation of the Use of Untargeted Metabolomics in the Safety Assessment of Genetically Modified Crops. *Metabolomics* **2020**, *16* (10), 111. https://doi.org/10.1007/s11306-020-01733-8.
34. Lacalle-Bergeron, L.; Izquierdo-Sandoval, D.; Sancho, J. V.; López, F. J.; Hernández, F.; Portolés, T. Chromatography Hyphenated to High Resolution Mass Spectrometry in Untargeted Metabolomics for Investigation of Food (Bio)Markers. *TrAC Trends in Analytical Chemistry* **2021**, *135*, 116161. https://doi.org/10.1016/j.trac.2020.116161.
35. Wei, S.; Wei, Y.; Gong, Y.; Chen, Y.; Cui, J.; Li, L.; Yan, H.; Yu, Y.; Lin, X.; Li, G.; Yi, L. Metabolomics as a Valid Analytical Technique in Environmental Exposure Research: Application and Progress. *Metabolomics* **2022**, *18* (6), 35. https://doi.org/10.1007/s11306-022-01895-7.
36. Lederberg, J.; McCray, A. T. `Ome Sweet `Omics--A Genealogical Treasury of Words. *The Scientist*. April 2, 2001, p 8.
37. The Human Microbiome Project Consortium. A Framework for Human Microbiome Research. *Nature* **2012**, *486* (7402), 215–221. https://doi.org/10.1038/nature11209.
38. Hooper, L. V.; Midtvedt, T.; Gordon, J. I. HOW HOST-MICROBIAL INTERACTIONS SHAPE THE NUTRIENT ENVIRONMENT OF THE MAMMALIAN INTESTINE. *Annual Review of Nutrition* **2002**, *22* (Volume 22, 2002), 283–307. https://doi.org/10.1146/annurev.nutr.22.011602.092259.

39. The Human Microbiome Project Consortium. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **2012**, *486* (7402), 207–214. https://doi.org/10.1038/nature11234.
40. Gomaa, E. Z. Human Gut Microbiota/Microbiome in Health and Diseases: A Review. *Antonie van Leeuwenhoek* **2020**, *113* (12), 2019–2040. https://doi.org/10.1007/s10482-020-01474-7.
41. Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A. Diversity of the Human Intestinal Microbial Flora. *Science* **2005**, *308* (5728), 1635–1638. https://doi.org/10.1126/science.1110591.
42. Bäckhed, F.; Ley, R. E.; Sonnenburg, J. L.; Peterson, D. A.; Gordon, J. I. Host-Bacterial Mutualism in the Human Intestine. *Science* **2005**, *307* (5717), 1915–1920. https://doi.org/10.1126/science.1104816.
43. O'Hara, A. M.; Shanahan, F. The Gut Flora as a Forgotten Organ. *EMBO reports* **2006**, *7* (7), 688–693. https://doi.org/10.1038/sj.embor.7400731.
44. Heintz-Buschart, A.; Wilmes, P. Human Gut Microbiome: Function Matters. *Trends in Microbiology* **2018**, *26* (7), 563–574. https://doi.org/10.1016/j.tim.2017.11.002.
45. Clemente, J. C.; Ursell, L. K.; Parfrey, L. W.; Knight, R. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* **2012**, *148* (6), 1258–1270. https://doi.org/10.1016/j.cell.2012.01.035.
46. Jethwani, P.; Grover, K. Gut Microbiota in Health and Diseases – A Review. *Int.J.Curr.Microbiol.App.Sci* **2019**, *8* (08), 1586–1599. https://doi.org/10.20546/ijcmas.2019.808.187.
47. Yang, W.; Cong, Y. Gut Microbiota-Derived Metabolites in the Regulation of Host Immune Responses and Immune-Related Inflammatory Diseases. *Cell Mol Immunol* **2021**, *18* (4), 866–877. https://doi.org/10.1038/s41423-021-00661-4.
48. Morrison, D. J.; Preston, T. Formation of Short Chain Fatty Acids by the Gut Microbiota and Their Impact on Human Metabolism. *Gut Microbes* **2016**, *7* (3), 189–200. https://doi.org/10.1080/19490976.2015.1134082.
49. Roager, H. M.; Licht, T. R. Microbial Tryptophan Catabolites in Health and Disease. *Nat Commun* **2018**, *9* (1), 3294. https://doi.org/10.1038/s41467-018-05470-4.
50. Zeisel, S. H.; Warrier, M. Trimethylamine N-Oxide, the Microbiome, and Heart and Kidney Disease. *Annual Review of Nutrition* **2017**, *37* (Volume 37, 2017), 157–181. https://doi.org/10.1146/annurev-nutr-071816-064732.
51. Long, S. L.; Gahan, C. G. M.; Joyce, S. A. Interactions between Gut Bacteria and Bile in Health and Disease. *Molecular Aspects of Medicine* **2017**, *56*, 54–65. https://doi.org/10.1016/j.mam.2017.06.002.
52. Amorim Franco, T. M.; Blanchard, J. S. Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability. *Biochemistry* **2017**, *56* (44), 5849–5865. https://doi.org/10.1021/acs.biochem.7b00849.
53. Di Martino, M. L.; Campilongo, R.; Casalino, M.; Micheli, G.; Colonna, B.; Prosseda, G. Polyamines: Emerging Players in Bacteria–Host Interactions. *International Journal of Medical Microbiology* **2013**, *303* (8), 484–491. https://doi.org/10.1016/j.ijmm.2013.06.008.
54. Magnúsdóttir, S.; Ravcheev, D.; de Crécy-Lagard, V.; Thiele, I. Systematic Genome Assessment of B-Vitamin Biosynthesis Suggests Co-Operation among Gut Microbes. *Front. Genet.* **2015**, *6*. https://doi.org/10.3389/fgene.2015.00148.
55. Wang, Z.; Klipfell, E.; Bennett, B. J.; Koeth, R.; Levison, B. S.; DuGar, B.; Feldstein, A. E.; Britt, E. B.; Fu, X.; Chung, Y.-M.; Wu, Y.; Schauer, P.; Smith, J. D.; Allayee, H.; Tang, W. H. W.; DiDonato, J. A.; Lusis, A. J.; Hazen, S. L. Gut Flora Metabolism of Phosphatidylcholine Promotes Cardiovascular Disease. *Nature* **2011**, *472* (7341), 57–63. https://doi.org/10.1038/nature09922.
56. Thomas, C.; Pellicciari, R.; Pruzanski, M.; Auwerx, J.; Schoonjans, K. Targeting Bile-Acid Signalling for Metabolic Diseases. *Nat Rev Drug Discov* **2008**, *7* (8), 678–693. https://doi.org/10.1038/nrd2619.
57. Guidelines for the Diagnosis and Management of Food Allergy in the United States: Report of the NIAID-Sponsored Expert Panel. *Journal of Allergy and Clinical Immunology* **2010**, *126* (6, Supplement), S1–S58. https://doi.org/10.1016/j.jaci.2010.10.007.
58. Sicherer, S. H.; Sampson, H. A. Food Allergy: A Review and Update on Epidemiology, Pathogenesis, Diagnosis, Prevention, and Management. *Journal of Allergy and Clinical Immunology* **2018**, *141* (1), 41–58. https://doi.org/10.1016/j.jaci.2017.11.003.
59. Gupta, R. S.; Warren, C. M.; Smith, B. M.; Blumenstock, J. A.; Jiang, J.; Davis, M. M.; Nadeau, K. C. The Public Health Impact of Parent-Reported Childhood Food Allergies in the United States. *Pediatrics* **2018**, *142* (6), e20181235. https://doi.org/10.1542/peds.2018-1235.
60. Hong, X.; Wang, X. Epigenetics and Development of Food Allergy (FA) in Early Childhood. *Curr Allergy Asthma Rep* **2014**, *14* (9), 460. https://doi.org/10.1007/s11882-014-0460-6.
61. Sicherer, S. H. Epidemiology of Food Allergy. *Journal of Allergy and Clinical Immunology* **2011**, *127* (3), 594–602. https://doi.org/10.1016/j.jaci.2010.11.044.
62. Young, E. A Population Study of Food Intolerance. *The Lancet* **1994**, *343* (8906), 1127–1130. https://doi.org/10.1016/S0140-6736(94)90234-8.

63. Bock, S. A. Prospective Appraisal of Complaints of Adverse Reactions to Foods in Children During the First 3 Years of Life. *Pediatrics* **1987**, *79* (5), 683–688. https://doi.org/10.1542/peds.79.5.683.
64. Fleischer, D. M.; Perry, T. T.; Atkins, D.; Wood, R. A.; Burks, A. W.; Jones, S. M.; Henning, A. K.; Stablein, D.; Sampson, H. A.; Sicherer, S. H. Allergic Reactions to Foods in Preschool-Aged Children in a Prospective Observational Food Allergy Study. *Pediatrics* **2012**, *130* (1), e25–e32. https://doi.org/10.1542/peds.2011-1762.
65. Sicherer, S. H.; Warren, C. M.; Dant, C.; Gupta, R. S.; Nadeau, K. C. Food Allergy from Infancy Through Adulthood. *The Journal of Allergy and Clinical Immunology: In Practice* **2020**, *8* (6), 1854–1864. https://doi.org/10.1016/j.jaip.2020.02.010.
66. Lack, G. Update on Risk Factors for Food Allergy. *Journal of Allergy and Clinical Immunology* **2012**, *129* (5), 1187–1197. https://doi.org/10.1016/j.jaci.2012.02.036.
67. Zhang, Q.; Zhang, C.; Zhang, Y.; Liu, Y.; Wang, J.; Gao, Z.; Sun, J.; Li, Q.; Sun, J.; Cui, X.; Wang, Y.; Fu, L. Early-Life Risk Factors for Food Allergy: Dietary and Environmental Factors Revisited. *Comprehensive Reviews in Food Science and Food Safety* **2023**, *22* (6), 4355–4377. https://doi.org/10.1111/1541-4337.13226.
68. Strachan, D. P. Hay Fever, Hygiene, and Household Size. *BMJ* **1989**, *299* (6710), 1259–1260.
69. Noverr, M. C.; Huffnagle, G. B. The 'Microflora Hypothesis' of Allergic Diseases. *Clinical & Experimental Allergy* **2005**, *35* (12), 1511–1520. https://doi.org/10.1111/j.1365-2222.2005.02379.x.
70. Joseph, C. L.; Sitarik, A. R.; Kim, H.; Huffnagle, G.; Fujimura, K.; Yong, G. J. M.; Levin, A. M.; Zoratti, E.; Lynch, S.; Ownby, D. R.; Lukacs, N. W.; Davidson, B.; Barone, C.; Cole Johnson, C. Infant Gut Bacterial Community Composition and Food-Related Manifestation of Atopy in Early Childhood. *Pediatric Allergy and Immunology* **2022**, *33* (1), e13704. https://doi.org/10.1111/pai.13704.
71. Tanaka, M.; Korenori, Y.; Washio, M.; Kobayashi, T.; Momoda, R.; Kiyohara, C.; Kuroda, A.; Saito, Y.; Sonomoto, K.; Nakayama, J. Signatures in the Gut Microbiota of Japanese Infants Who Developed Food Allergies in Early Childhood. *FEMS Microbiology Ecology* **2017**, *93* (8), fix099. https://doi.org/10.1093/femsec/fix099.
72. Savage, J. H.; Lee-Sarwar, K. A.; Sordillo, J.; Bunyavanich, S.; Zhou, Y.; O'Connor, G.; Sandel, M.; Bacharier, L. B.; Zeiger, R.; Sodergren, E.; Weinstock, G. M.; Gold, D. R.; Weiss, S. T.; Litonjua, A. A. A Prospective Microbiome-Wide Association Study of Food Sensitization and Food Allergy in Early Childhood. *Allergy* **2018**, *73* (1), 145–152. https://doi.org/10.1111/all.13232.
73. Bunyavanich, S.; Shen, N.; Grishin, A.; Wood, R.; Burks, W.; Dawson, P.; Jones, S. M.; Leung, D. Y. M.; Sampson, H.; Sicherer, S.; Clemente, J. C. Early-Life Gut Microbiome Composition and Milk Allergy Resolution. *Journal of Allergy and Clinical Immunology* **2016**, *138* (4), 1122–1130. https://doi.org/10.1016/j.jaci.2016.03.041.
74. Aitoro, R.; Paparo, L.; Amoroso, A.; Di Costanzo, M.; Cosenza, L.; Granata, V.; Di Scala, C.; Nocerino, R.; Trinchese, G.; Montella, M.; Ercolini, D.; Berni Canani, R. Gut Microbiota as a Target for Preventive and Therapeutic Intervention against Food Allergy. *Nutrients* **2017**, *9* (7), 672. https://doi.org/10.3390/nu9070672.
75. *Probiotics, Prebiotics, Synbiotics, and Postbiotics: Human Microbiome and Human Health*; Kothari, V., Kumar, P., Ray, S., Eds.; Springer Nature Singapore: Singapore, 2023. https://doi.org/10.1007/978-981-99-1463-0.
76. Renz, H.; Skevaki, C. Early Life Microbial Exposures and Allergy Risks: Opportunities for Prevention. *Nat Rev Immunol* **2021**, *21* (3), 177–191. https://doi.org/10.1038/s41577-020-00420-y.
77. Hill, C.; Guarner, F.; Reid, G.; Gibson, G. R.; Merenstein, D. J.; Pot, B.; Morelli, L.; Canani, R. B.; Flint, H. J.; Salminen, S.; Calder, P. C.; Sanders, M. E. The International Scientific Association for Probiotics and Prebiotics Consensus Statement on the Scope and Appropriate Use of the Term Probiotic. *Nat Rev Gastroenterol Hepatol* **2014**, *11* (8), 506–514. https://doi.org/10.1038/nrgastro.2014.66.
78. Gibson, G. R.; Probert, H. M.; Loo, J. V.; Rastall, R. A.; Roberfroid, M. B. Dietary Modulation of the Human Colonic Microbiota: Updating the Concept of Prebiotics. *Nutrition Research Reviews* **2004**, *17* (2), 259–275. https://doi.org/10.1079/NRR200479.
79. Pandey, K. R.; Naik, S. R.; Vakil, B. V. Probiotics, Prebiotics and Synbiotics- a Review. *J Food Sci Technol* **2015**, *52* (12), 7577–7587. https://doi.org/10.1007/s13197-015-1921-1.
80. Vindigni, S. M.; Surawicz, C. M. Fecal Microbiota Transplantation. *Gastroenterology Clinics* **2017**, *46* (1), 171–185. https://doi.org/10.1016/j.gtc.2016.09.012.
81. Castellazzi, A. M.; Valsecchi, C.; Caimmi, S.; Licari, A.; Marseglia, A.; Leoni, M. C.; Caimmi, D.; Miraglia del Giudice, M.; Leonardi, S.; La Rosa, M.; Marseglia, G. L. Probiotics and Food Allergy. *Ital J Pediatr* **2013**, *39* (1), 47. https://doi.org/10.1186/1824-7288-39-47.
82. Cuello-Garcia, C.; Fiocchi, A.; Pawankar, R.; Yepes-Nuñez, J. J.; Morgano, G. P.; Zhang, Y.; Agarwal, A.; Gandhi, S.; Terracciano, L.; Schünemann, H. J.; Brozek, J. L. Prebiotics for the Prevention of Allergies: A

Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Clinical & Experimental Allergy* **2017**, *47* (11), 1468–1477. https://doi.org/10.1111/cea.1304

Chapter II

Development of an untargeted LC-MS metabolomics method with post-column infusion for matrix effect monitoring in plasma and feces

Based on:

Pingping Zhu, Anne-Charlotte Dubbelman, Christie Hunter, Michele Genangeli, Naama Karu, Amy Harms, Thomas Hankemeier

Development of an untargeted LC-MS metabolomics method with post-column infusion for matrix effect monitoring in plasma and feces

Journal of the American Society for Mass Spectrometry

DOI: 10.1021/jasms.3c00418

Abstract

Untargeted metabolomics based on RPLC-MS plays a crucial role in biomarker discovery across physiological and disease states. Standardizing the development process of untargeted methods requires attention to critical factors that are under discussed or easily overlooked, such as injection parameters, performance assessment and matrix effect evaluation. In this study, we developed an untargeted metabolomics method for plasma and fecal samples with the optimization and evaluation of these factors. Our results showed that optimizing the reconstitution solvent and sample injection amount were critical for achieving the balance between metabolites coverage and signal linearity. Method validation with representative stable-isotopically labeled standards (SILs) provided insights into the analytical performance evaluation of our method. To tackle the issue of matrix effect, we implemented a post-column infusion (PCI) approach to monitor the overall absolute matrix effect (AME) and relative matrix effect (RME). The monitoring revealed distinct AME and RME profiles in plasma and feces. Comparing RME data obtained for SILs through post-extraction spiking with those monitored using PCI compounds demonstrated the comparability of these two methods for RME assessment. Therefore, we applied the PCI approach to predict the RME of 305 target compounds covered in our in-house library and found that targets detected in the negative polarity were more vulnerable to RME, regardless of the sample matrix. Given the value of this PCI approach in identifying the strengths and weaknesses of our method in terms of matrix effect, we recommend implementing a PCI approach during method development and applying it routinely in untargeted metabolomics.

1. Introduction

Untargeted metabolomics is a powerful approach that has demonstrated great potential in exploring metabolic changes in health and disease conditions[1–3]. Its application has extended beyond biomedical research to fields such as food, agricultural and environmental studies[4–6], thereby making it a highly valuable tool for diverse scientific research. One of the most widely used techniques for untargeted metabolomic analysis is ultra-performance liquid chromatography coupled to mass spectrometry (UPLC-MS)[7]. Among the different types of UPLC-MS, reverse phase (RP) LC-MS is the most popular choice for UPLC-MS due to its versatility, robustness, stability and good retention of semi-polar to non-polar metabolites[8]. As the popularity of untargeted metabolomics has increased, researchers have focused on standardizing the development process of this method, especially when aiming at semi-quantitative analysis beyond qualitative compound discovery and screening. Parameters such as sample extraction, LC-MS system selection and setup, quality management, and analysis batch design have been extensively studied and advised upon[9–11]. However, some critical factors required to develop a reliable untargeted RPLC-MS platform are either easily overlooked or still under discussion for standardization. Those factors include the optimization of injection solvent and sample injection amount. In the sample preparation process of untargeted methods, an evaporation and reconstitution step is typically performed to allow for flexibility in modifying the injection solvent composition and the sample loading amount. This step is important to prevent mismatch between the mobile phase and injection solvent, and to balance the challenge of maximizing the metabolome coverage, minimizing signal saturation and reducing matrix effect[12,13]. Studies have shown that the reconstitution solvent can affect peak shape and metabolite coverage in RPLC for untargeted analysis of small molecules[14,15], which emphasizes the importance of investigating the injection solvent during the development of a RPLC-MS method. Another critical parameter is the injection amount, which was reported to impact the data quality and repeatability in terms of overloading, signal saturation and featuremissingness[16]. Therefore, a systematic investigation of the injection amount is also critical when developing an RPLC-MS method[13].

The investigation of reconstitution solvents typically involves assessing the peak shape and signal intensity of representative metabolites[14,15]. When investigating the injection amount, serially diluted standards or samples are commonly used to evaluate signal linearity[13,16]. In order to maintain signal linearity in high-resolution MS, techniques such as Dynamic Ion Transmission Control (ITC) and Automatic Gain Control (AGC) have been developed to modulate the ion amounts in various regions of the MS system. The ITC technique, implemented in all QTOF systems from SCIEX, modulates the ion current scan by scan to ensure it remains within the dynamic range of the detection system. For trap-based MS instruments from Thermo Fisher, AGC is employed to automatically regulate the ion amount in the ion-trap by adjusting the fill time for every scan. These techniques not only extend the dynamic range of the MS system but also offer insights into the ion transmission status through the MS system. Recently, they have been employed as effective approaches to investigate ion transmission during method development with high-resolution MS[17], making them promising readouts for the optimization of sample injection amount to avoid the risk of signal non-linearity.

Another challenge in untargeted metabolomics is the method performance assessment and validation. Despite the recommendations for addressing quality assurance and quality control challenges[9,18], there is currently no consensus on the performance validation of untargeted methods during the development phase. However, it has been recommended that in addition to monitoring signal drift and repeatability with pooled quality control (QC) samples, an untargeted method can be validated in a targeted way with representative metabolites[19]. This strategy has been widely applied in untargeted metabolomics research to validate the parameters of linearity, precision, recovery, and accuracy with selected endogenous metabolites[9,20–22]. However, in these studies, serially diluted pooled QC samples were commonly used to evaluate the linearity, leading to the dilution of both the targeted analyte and the matrix, which reduces the reliability of this strategy[19]. Matrix effect has been regarded as one of the most significant challenges of LC-MS methods, especially when analyzing complex biological matrices[23,24]. Therefore, matrix effect is another widely discussed factor in untargeted metabolomics because of its impact on reproducibility, linearity, selectivity, accuracy and sensitivity[25].

The phenomenon of matrix effect was first reported in 1993 by Tang and Kebarle, who observed that the signal of an analyte ionized by the electrospray ionization (ESI) source can be strongly affected by the presence of other electrolytes in the solution[26]. Although the exact mechanism of how the matrix effect occurs is still unclear, it is commonly assumed that the co-eluted matrix components can affect the ionization of an analyte by preventing or competing with the analyte to gain charge, increasing the surface tension of the charged droplet, interfering with the stability of charged analytes in the gas phase and/or co-precipitating with the analyte[27]. To overcome matrix effect in LC-MS, two main strategies have been proposed: matrix effect reduction and matrix assessment/correction. Matrix effect reduction can be achieved through extensive sample cleaning procedures, enhanced LC separation efficiency, sample dilution, or adopting alternative MS ionization sources other than ESI[25,28–30].

Matrix effect assessment can mainly be achieved by post-extraction spiking and post-column infusion (PCI) of compounds[24,25]. The post-extraction spike method was proposed by Matuszewski *et al.* to quantitatively assess matrix effect by comparing the response in neat standard solution samples with that in post-extraction spiked samples. They also introduced the terms absolute matrix effect (AME) and relative matrix effect (RME) to describe matrix effect, where AME is the response ratio of an analyte at a given concentration spiked in post-extraction biological samples compared to neat solution samples, and RME is the variability of AME among different lots of biological samples[30]. Following the introduction of the post-extraction spiking method, the term matrix factor (MF) was introduced as a quantitative measure of matrix effect and shares the same concept with AME[31]. The MF was later applied in accordance with the European Medicine Agency (EMA) guideline released in 2011 for the ME assessment in bioanalytical method validation[32]. According to the guideline, the MF variability (RME) should not exceed 15%. In contrast to the post-extraction spiking method which assesses the matrix effect at specific time points, the PCI technique was proposed by Bonfiglio *et al.* as a method for matrix effect assessment across the entire LC chromatogram[33]. In PCI, a compound is constantly infused into the MS after joining the column effluent using a T-connector. This enables the infusion profile of the compound

to be observed across the entire chromatogram with the injection of a matrix sample and a blank sample, allowing for real-time monitoring of the matrix effect. Due to this advantage, PCI has been utilized in targeted analysis for matrix effect evaluation and correction for small molecules and drugs in urine and plasma samples[34–37].

Unfortunately, although multiple strategies have been proposed for reducing and assessing matrix effect, few are applicable to untargeted LC-MS methods. In these methods, simple and unbiased sample preparation is required to broaden the metabolite coverage, and in order to represent a compromise that accommodates most classes, the LC separation is typically not tailored for specific compound classes[12,38]. Therefore, in terms of matrix effect reduction, aside from switching to an ionization source other than ESI, sample dilution is the only applicable approach in untargeted metabolomics. For the matrix effect assessment and correction, the post-extraction spiking method is more suitable for targeted metabolomics due to the requirement of authentic standards. Hence, PCI is recommended as a more appropriate tool for matrix effect evaluation in untargeted metabolomics[29,39], but only few studies about its application have been reported[40,41]. Although stable isotope labelling has also been applied to matrix effect evaluation in untargeted metabolomics, this technique is limited to specific sample types like yeast, cells or plants due to the requirement of globally labeled growth medium[42–44].

In this study, we developed an RPLC-MS untargeted metabolomics method suitable for the measurement of plasma and feces, taking into account both matrix diversity and the growing popularity of fecal metabolome studies. Initially, we optimized the injection solvent and injection amount for both matrices and validated the optimized platform in a targeted manner. To evaluate the matrix effect of plasma and feces alongside other performance parameters (precision, accuracy, recovery) and guarantee the reliability of the linearity, stable-isotopically labeled standards (SILs), instead of endogenous metabolites, were used in the validation. These SILs are well distributed in terms of class, retention time, physicochemical properties, and abundance according to their endogenous analogues. By validating our method with these SILs, we have gained insight into its analytical performance. Additionally, we augmented this untargeted

method with a PCI approach for matrix effect monitoring, which offers the advantage of overall matrix effect evaluation of plasma and fecal samples. This allows us to identify the strengths and weaknesses of our method in terms of matrix effect, ensuring better data reliability in untargeted metabolomics. The successful application of PCI for matrix effect monitoring in this untargeted metabolomics method strongly suggests that this approach can be widely implemented in the development and routine analysis of an LC-MS untargeted method.

2. Methods

2.1 Chemicals and materials

LC-MS-grade acetonitrile (ACN) and Methanol (MeOH) were purchased from Actu-all chemicals (Randmeer, The Netherlands). Methyl tert-butyl ether (MTBE, ≥99.8%) and sodium hydroxide were purchased from Sigma Aldrich (St. Louis, Missouri, United States). Formic acid (FA) was purchased from Biosolve B.V. (Valkenswaard, Netherlands), and hydrochloric acid (37% solution in water) was purchased from Acros organics (Geel, Belgium). Purified water was obtained from a Milli-Q PF Plus system (Merck Millipore, Burlington, Massachusetts, United States). Most chemical standards and stable isotopic-labelled standards (SILs) were purchased from CDN Isotopes (C/D/N Isotopes Inc, Quebec, Canada), Cambridge Isotope Laboratories (Tewksbury, MA, USA) and TRC (Toronto Research Chemicals, Toronto, Canada). Table S1 provides the supplier details of all standards. Pooled EDTA plasma was obtained from Innovative Research (Peary Court Novi, MI, USA), pooled male and female ETDA plasma were purchased from Sanquin (Sanquin, Amsterdam, The Netherlands), and ETDA plasma from individual donors was purchased from BioIVT (Westbury, NY, USA). Fecal samples were collected from four healthy adults, including three female and one male volunteer (age range: 23-35 years old).

2.2 Solution Preparation

2.2.1 Preparation of calibrant solutions

The stock solutions of 28 authentic SILs were prepared at different concentrations in appropriate solvents (Table S1). For certain SILs, ammonium hydroxide or hydrochloric

acid was added to improve solubility. Standard mixture solutions were prepared by mixing 21 (plasma validation) or 16 (feces validation) SILs. Those mixtures were serially diluted with water to obtain working calibration solutions at 9 (plasma) or 11 (feces) concentration levels (see table S2-S3). Stock solutions and standard mixtures were stored at -80 °C until use, and calibration solutions were freshly prepared before experiments.

2.2.2 Preparation of internal standards and reconstitution solution

Fludrocortisone-d_5, glucose-$^{13}C_6$-d_7, caffeine-d_9 and valine-d_8 were added as internal standards (IS) for signal drift monitoring. Detailed information of those IS are shown in Table S1. Four IS were spiked in plasma validation, while three IS (except glucose-$^{13}C_6$-d_7) were spiked for fecal validation. Cortisone-d_8 in water with 0.1% FA was prepared as the reconstitution solution.

2.2.3 Preparation of PCI compounds solutions

Leucine-enkephalin, fludrocortisone, 5-fluoroisatin, caffeine-$^{13}C_3$ and 3-fluoro-DL-valine were selected as the PCI compounds considering the physical properties, ionization behaviors, availability and cost. All the PCI compounds were prepared with water or MeOH or water/MeOH (1:1, v/v) (Table S1). The post-column infusion mixture solution was prepared with water/ACN (1:1, v/v). In positive mode, the PCI comprised Leucine-enkephalin, fludrocortisone, 5-fluoroisatin, and caffeine-$^{13}C_3$, while in negative mode, it included Leucine-enkephalin, fludrocortisone and 3-Fluoro-DL-valine. Table S4 provides the final concentrations of each PCI compound in the mixture solutions.

2.3 Sample preparation

2.3.1 Plasma sample preparation

Protein precipitation was used to prepare plasma samples. Aliquots of 25 µL of plasma were mixed with 10 µL of IS working solution and quenched with 90 µL of ice-cold MeOH. All samples were then vortex mixed (1 min, high speed), incubated on ice (20 min), and centrifuged (15 min, 15600 g, 4°C). Afterwards, 100 µL of supernatant from each sample was transferred to 1.5 mL Eppendorf tubes and evaporated to dryness in a SpeedVac (Labcono, Kansas City, MO, USA). The residuals were reconstituted in 75

µL of water with 0.1% FA, vortex mixed (1 min, high speed) and centrifuged (5 min, 2300 g, 4°C). Finally, 70 µL of the supernatant was transferred to autosampler vials, and 1 µL was injected into the LC-MS.

During method development, extracted plasma samples were reconstituted in 50 µL of 0.1% FA in water with 0%, 10% or 20% of ACN (v/v/v) to optimize reconstitution solvent and in 50, 75 or 150 µL of 0.1% FA in water to evaluate sample dilution factors (DF) of 1:2, 1:3 and 1:6 (v/v), respectively. Of those samples, 1 and 2 µL was injected to compare injection volumes.

For method validation, calibration lines (n=3) were created using pooled EDTA plasma with 10 µL of spiked calibration working solutions. Precision was evaluated at each concentration level from the calibration lines. Pooled EDTA plasma, pooled male EDTA plasma, pooled female EDTA plasma and one individual EDTA plasma were used as four different plasma samples for recovery, accuracy and matrix effect evaluation. For recovery, plasma samples were prepared by spiking 10 µL of calibration working solutions to get concentrations at low (cal4), medium (cal6) and high (cal8) levels before extraction and after drying. The samples spiked before extraction were also used for the evaluation of accuracy. Samples for the matrix effect evaluation were prepared by spiking 10 µL of calibration working solution at three concentration levels in plasma and matrix-free (solvent) samples after drying.

2.3.2 Fecal sample preparation

Final sample preparation procedure

Fecal samples were stored at -20°C immediately after collection. Samples were thawed at ambient temperature and homogenized as proposed by Hosseinkhani *et al.* (involving stirring, sonication for 5 min and vortex mixing for 10 min)[45], with the adjustment that 1 mL water per gram of sample was added at the start to improve homogenization. The homogenized and aliquoted samples (around 2g per tube) were stored at -80 °C for more than 48h before lyophilization. Freeze-drying was conducted overnight (20 h, 4 mBar, -110 °C) with a CHRIST Alpha 3-4 LSCbasic freezer-dryer (Martin Christ, Germany)

and 20 mg (± 0.3 mg) lyophilized sample aliquots were weighed and stored at -80 °C until extraction.

Liquid-liquid extraction (LLE) was performed as recommended by Hosseinkhani *et al*.[45], whereby the starting amount was adapted to 20 mg dried feces, considering the added water and limited sample size of clinical samples. Added volumes for extraction were changed accordingly. Briefly, 108 µL ice-cold MeOH (5.4 µL mg^{-1} dried feces) and 36 µL ice-cold water (1.8 µL mg^{-1} dried feces) were added to 1.5 mL Eppendorf tubes with 20mg freeze-dried feces, followed by vortex mixing (2 min). Then, 60 µL ice-cold MTBE (3 µL mg^{-1} dried feces) was added, followed by vortex mixing (2 min) and centrifugation (15 min, 16000g, 4 °C,). Next, 140 µL of the supernatant was transferred to clean tubes. Phase separation was induced by adding 84 µL of ice-cold MTBE (4.2 µL mg^{-1} dried feces) and 100 µL of ice-cold water (5 µL mg^{-1} dried feces). Then samples were remixed for 2 min and kept at 4°C for 10 min to obtain better protein precipitation. After centrifugation (20 min, 16000g, 4 °C), 90 µL of the aqueous layer was transferred to 1.5 mL Eppendorf tubes and evaporated to dryness. The remainder of the aqueous layer was saved for other analyses. The dried residuals were reconstituted in 50 µL of reconstitution solution, resulting in the ratio of dried feces to reconstitution solvent being around 1:8 (mg/v) (calculation details are provided in table S5). All the samples were vortex mixed (5 min) and centrifuged (5 min, 16000g, 4 °C) before transfer to autosampler vials, and 1 µL was injected into the LC-MS.

Sample preparation for reconstitution solvent, dilution factor, and injection volume comparison

Pooled fecal samples from three individuals were used to optimize the reconstitution solution, dilution factor, and injection volume for feces. The individual samples were homogenized separately, and equal amounts were aliquoted, pooled, mixed and homogenized. Freeze-dried feces (50 mg) from the pooled sample were aliquoted and extracted with MTBE/MeOH/water (3.6/2.7/3.4, v/v/v). After LLE, the aqueous layer was transferred to 1.5 mL Eppendorf tubes and evaporated in the SpeedVac. Dried fecal extracts were reconstituted in 300 µL of 0.1% FA in water with 0%, 10% or 20% of ACN (v/v/v) to evaluate reconstitution solvent, and in 150 or 300 µL of 0.1% FA in

water to evaluate sample DF of 1:3 and 1:6 (mg/v), respectively. Of those samples, 1 and 2 µL were injected to optimize injection volume.

Sample preparation for validation

A pooled sample from four donors was used to build the calibration line and assess precision and recovery. Samples from each individual were used for accuracy and matrix effect evaluation. Calibration lines were constructed by spiking the calibrant solution at each level to the samples after LLE extraction. Samples for recovery evaluation were prepared by spiking calibrant solution in fecal samples to get concentration at low (cal4) and high (cal10) concentration levels before LLE extraction and after drying. The samples spiked after drying were also used for the evaluation of accuracy. Samples for the matrix effect evaluation were prepared by spiking calibrant solutions in fecal and matrix-free (solvent) samples to get concentrations at low (cal4), medium (cal7) and high (cal10) levels after drying. Final sample preparation procedure for feces was followed for the steps of extraction, reconstitution and injection.

2.4 Method validation

Linearity

The linearity of selected SILs in both plasma and feces was evaluated by calibration lines (n = 3). The calibration lines of the SILs applied in plasma and feces were designed based on the concentration levels of their endogenous analogues (Figure S1). The calibration points and ranges of SILs after spiking in plasma and feces are presented in table S2-S3.

Precision, accuracy and recovery

Precision was expressed as the relative standard deviation (RSD) of the peak area for each calibration point in three calibration lines. Accuracy and recovery were evaluated at different concentration levels with four samples. The accuracy was calculated by dividing the calibration line back-calculated concentration by the nominal concentration at each level. The recovery was calculated as the ratio of the SILs peak area obtained in the samples spiked before extraction and after drying at each concentration level.

Matrix effect

Absolute matrix effect (AME) and relative matrix effect (RME) were both evaluated with four different plasma samples. The AME was assessed by calculating the ratio of peak area obtained in the matrix (post-extraction) and matrix-free sample (solvent sample). The RME was expressed as the RSD of the AME.

2.5 LC-MS conditions and post-column infusion setup

Analysis was performed on a reverse phase UPLC-MS untargeted platform. The platform consisted of a Shimadzu Nexera X2 LC system coupled to a TripleTOF 6600 mass spectrometer (SCIEX, Foster City, CA, USA) with an electrospray ionization source (ESI) that operated at both positive and negative ion modes. The ESI source parameters were set as follows: spray voltage ±4.5 kV, capillary temperature 400 ◦C, sheath gas 40, auxiliary gas 40, curtain gas 45. Data were acquired under full scan mode over the *m/z* range of 60-800 Da. The LC separation was carried out using a Waters Acquity UPLC HSS T3 column (1.8 μm, 2.1 mm × 100 mm) with the oven temperature maintained at 40 °C. The mobile phase A was 0.1% FA in water, and the mobile phase B was 0.1% FA in ACN. With a flow rate of 0.4 mL min^{-1}, the gradient started at 100% A and was held for 0.5 min, then B linearly increased to 20% over 2.5 min and continuously increased to 98% from 2.5 to 7.5 min. This condition was maintained for 4.5 min, then returned to 100% A in 0.1 min, at which the column was equilibrated for 3 min, resulting in a 15 min run time per analysis. The autosampler temperature was set at 10 °C. To decelerate the contamination of the MS, the LC flow was diverted to waste at 7 min of the gradient by an external valve (Valco instruments, USA). During the analysis, the PCI compounds were continuously pumped by a binary Agilent 1260 Infinity pump (Agilent Technologies, Santa Clara, USA) at a flow rate of 20 μL min^{-1} and combined to the LC flow with a T-piece (IDEX, PEEK Tee, 0.02 Thru hole, F-300) before entering the ESI source.

2.6 Data processing

The raw data was obtained using Analyst TF software 1.7.1 (SCIEX) and processed using SCIEX OS (version 2.1, SCIEX) and PeakView (version 2.2, SCIEX) software.

Extracted ion chromatograms (EICs) were obtained for each compound, including PCI compounds with an *m/z* window of 0.02 Da. A maximum mass error of 5 ppm was applied for peak integration of all the compounds, and the retention times of endogenous compounds were verified using authentic standards. Count conversion factor plots were viewed in PeakView. This option can be enabled by closing the PeakView software, copying the "Instrument Utilities.dll" file from the "C:\Program Files\AB SCIEX\PeakView 2\bin" folder to the "C:\Program Files\AB SCIEX\PeakView 2\Help" folder, and restarting the software. Then, when opening a datafile and extracting the TOF MS TIC, navigate to the "Help" menu in PeakView software, click on the "Instrument Utilities.dll" and select "Plot Count Conversion Factors". The PCI infusion profiles were generated by smoothing the extracted EIC data using the simple moving average function (SMA, n = 20) in R (version 4.2.1). To generate matrix effect profiles (MEPs), the matrix effect of each time point was calculated as reported in the literature[41]. This calculation involved dividing the EIC response (R) of each PCI compound in the matrix sample by that in the blank sample (Eq 1) and smoothing accordingly.

Equation 1: $\mathrm{ME}\ (\%) = \frac{R_{matrix}}{R_{blank}} * 100$

Equation 2: $\overline{\mathrm{ME}}_{S_i} = \mathrm{Mean}[\mathrm{ME}(C_{S_i}^1) \ldots \mathrm{ME}(C_{S_i}^j)]$

Equation 3: $\mathrm{RME}_S = \frac{\mathrm{SD}(\overline{ME}_{S_1} \ldots \overline{ME}_{S_i})}{\mathrm{Mean}(\overline{ME}_{S_1} \ldots \overline{ME}_{S_i})} * 100$

To evaluate the RME among four individuals, the absolute matrix effect signal of one sample ($\overline{\mathrm{ME}}_{S_i}$) was firstly calculated by averaging the matrix effect of all the PCI compounds (C^1- C^j) in that sample (Eq 2). Then, the RME among four individual samples (RME_S) was calculated as the RSD of $\overline{\mathrm{ME}}_{S_i}$ (Eq 3). The calculated RME profile from certain samples was used to predict RME for targets detected in those samples based on their retention times.

3. Results and discussion

3.1 Analytical performance evaluation of the developed method

Before method validation, we optimized the reconstitution solvent and injection amount for both plasma and fecal samples, as detailed in the supplementary section "reconstitution solvent and injection amount optimization". In summary, 0.1% FA in water outperformed solutions with 10% and 20% ACN and was chosen as the final injection solvent, considering the peak shape for the metabolites of interest. After signal intensity comparison and detector saturation checking through Dynamic Ion Transmission Control (ITC), dilution factors DF3 (1:3, v/v) and DF8 (1:8, mg/v) were selected for plasma and fecal samples, respectively, with an injection volume of 1µL. In the analytical performance evaluation, we validated the untargeted method in both plasma and fecal samples. The dynamic range, precision, accuracy, recovery and matrix effect were evaluated with selected SILs.

3.1.1 Plasma validation

The linearity range and precision are summarized in Table S7. We obtained good linearity ($R^2 > 0.98$) with a wide range for 19 of 21 SIL targets. The inclusion and exclusion criteria for the calibration points were based on the acceptable residual error (< 20%) compared to the nominal concentration. At least five consecutive concentration levels were required to build a calibration curve. DCA-d_4 could not form a calibration curve as only three continuous concentration levels were within acceptable criteria, probably caused by its solubility issue as described in the "reconstitution solvent and injection amount optimization "supplementary section. Good precision (RSD < 15 %) was achieved for most of the acceptable concentration levels. The accuracy, recovery and matrix effect were assessed with three concentration levels (Low, Medium, High). However, only medium and high concentrations were evaluated for n-methyl-d_3-l-histidine, indole-d_5-3-acetic acid and GCA-d_4 because the low-level concentration fell below the detection limit. None of them was evaluated for DCA-d_4 due to unavailability of the calibration curve.

Good recoveries were obtained for 20 SILs (within 80-120%), except for TMAO-d_9, which exhibited a recovery of around 65% at low and medium concentration levels (Figure S5a). The accuracy between the back-calculated and the nominal concentration was within 67-122% for all SIL targets, except for citric acid-d_4, which had an accuracy

close to 200% (Figure S5b). The imprecise accuracy of citric acid-d_4 was caused by the varying levels of citric acid in the different plasma samples. We observed that the citric acid level in the pooled plasma used for creating the calibration curve was much higher than the other plasma we used for accuracy evaluation. Therefore, with an identical spiked concentration of citric acid-d_4, a higher response was observed in the plasma with lower endogenous citric acid due to lower rate of ion suppression. When applying the calibration line built with suppressed signal to the samples that suffered less ion suppression, the back-calculated concentrations will be higher than the spiked ones due to the higher observed response, resulting in the inaccuracy of citric acid-d_4. The impact of ion suppression on accuracy emphasizes the importance of matrix effect evaluation, especially the relative matrix effect among samples.

The results of the matrix effect evaluation are presented in Figure 1. As shown in Figure 1a, for 45% of the SIL targets, the absolute matrix effects (AME) met the criteria acceptable by most bioanalytical laboratories (80%-120%)[46] at all the concentration levels. Severe AME was observed for some early-eluting targets (L-ornithine-d_6, n-methyl-d_3-l-histidine and L-glutamine-d_5) with a value below 20%. TMAO-d_9, L-carnitine-d_3, betaine-d_9 and lactic acid-$^{13}C_3$ had AME lower than 80%. These SIL targets eluted in regions with a high intensity of co-eluting ions, as shown in the total ion chromatogram (TIC) (see Figure S6); therefore ion suppression could be expected for compounds eluting in those regions. The AME of citric acid-d_4 and octanoyl-l-carnitine-d_3 were above 120% at low and medium concentrations, while indole-d_5-3-acetic acid and GCA-d_4 had AME larger than 120% at all the detected concentrations. The precision of AME was determined by the RSD of the AME, which is also called the relative matrix effect (RME). As presented in Figure 1b, L-lactic acid-$^{13}C_3$ and citric acid-d_4 had RME larger than 15%, and the other targets all had RME less than 15%.

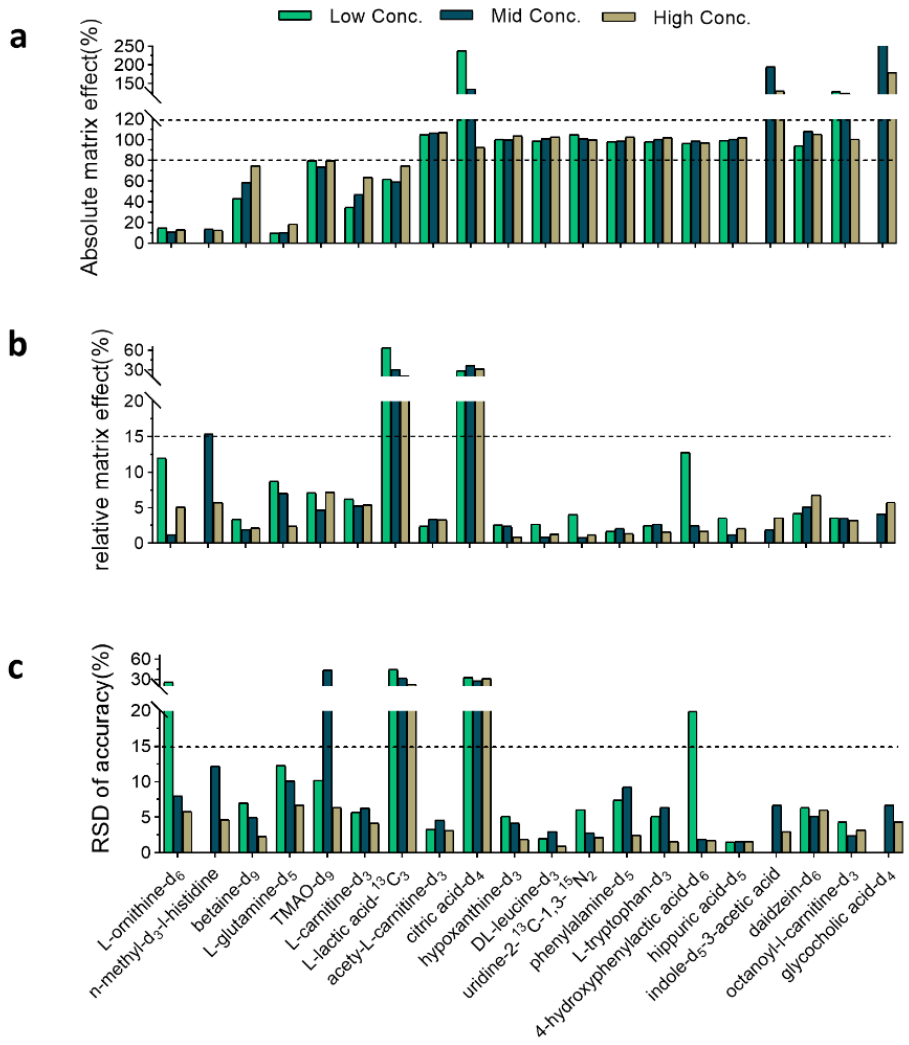


Figure 1. Matrix effect and precision of accuracy for spiked SIL targets in plasma. (a) Absolute matrix effect (AME), the dashed lines point out the range of 80-120%. (b) Relative matrix effect (RME), the dashed line indicates the RME at 15%. (c) The precision (RSD%) of the accuracy among four different donors, the dashed line indicates the RSD% at 15%.

3.1.2 Feces validation

The linearity range and precision are summarized in Table S8. All the targets, except u-15N-guanosine, obtained good linearity (R2 > 0.98) with a wide range, and at least six consecutive calibration points were included for building the calibration curve. Good precision (RSD < 15%) was achieved for most of the calibration points (Table S8). Additionally, good accuracy (80-120%) was obtained for almost all the targets (Figure S7). Nevertheless, slightly lower accuracy was observed at either low or high concentrations for some targets (hippuric acid-d_5, L-tyrosine-$^{13}C_9$-^{15}N, DL-leucine-d_3, phenylalanine-d_5, L-tryptophan-d_3) because they were close to the boundary of the linear range. The accuracies of choline-d_4 and DL-proline-d_7 at the high level are lower than 60% due to exceeding the linear range, and the low levels of some targets were excluded because they were below the lower detection limit.

The recovery for fecal LLE extraction was validated at low and high concentration levels (Figure S8a). The RSD of recovery among four replicates was calculated to show the repeatability of the extraction process (Figure S8b). Overall, although almost all targets had a recovery below 80%, good repeatability (RSD < 10%) was obtained. However, attention needs to be paid to cytidine-$^{15}N_3$, u-^{15}N-guanosine and citric acid-d_4, which have recoveries below 30%.

The matrix effect results for spiked SILs in feces are described in Figure 2. Overall, the AME for most spiked SILs was around 80%, at least for two concentration levels, except cytidine-$^{15}N_3$ and octanoyl-L-carnitine-d3 with AME above 120% for all detectable concentrations (Figure 4a). The overall ion suppression for all the SILs spiked in fecal sample aligns with the intensity variation of TICs for fecal samples, as presented in Figure S6. An RME below 15% was obtained for most of the spiked SILs, with only indole-d_5-3-acetic acid showing larger variability at three concentration levels (Figure 2b).

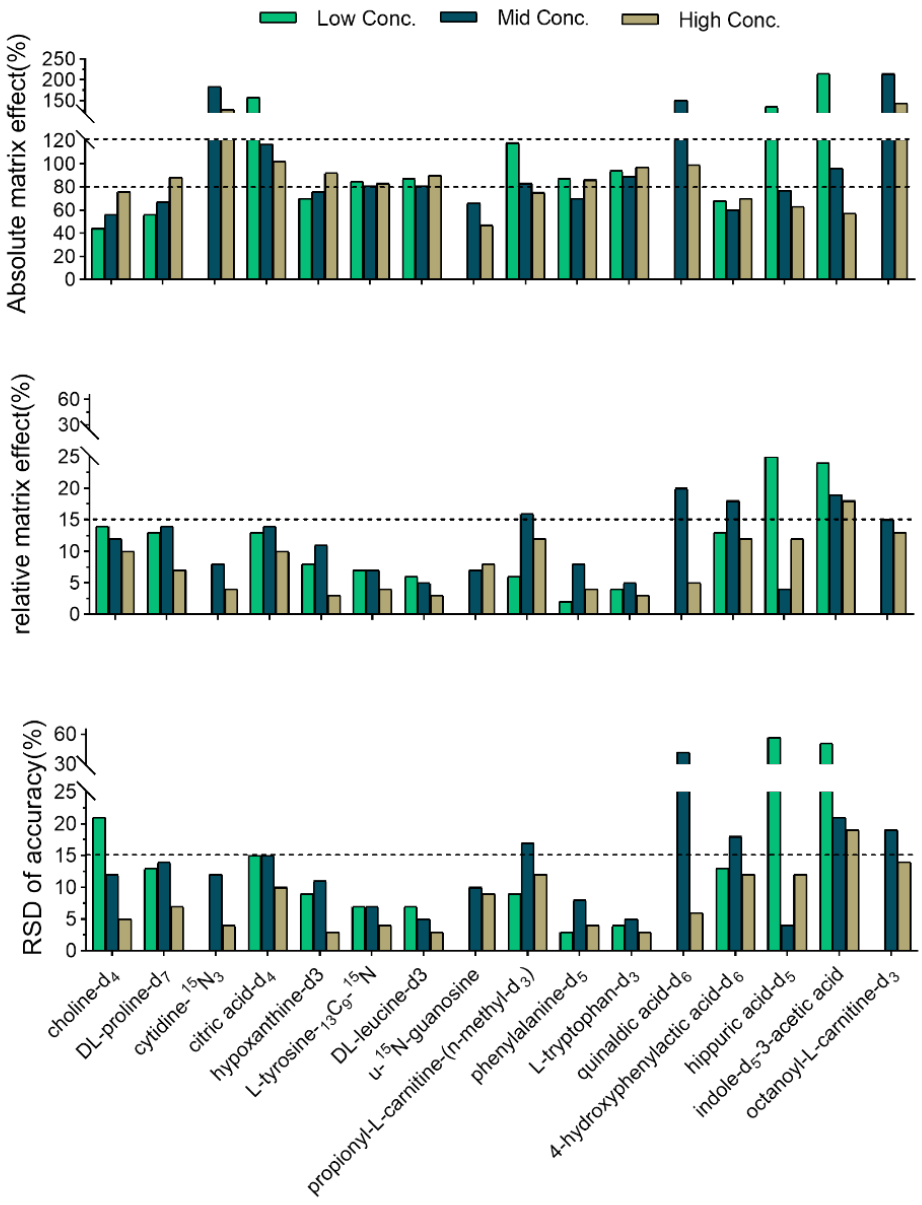


Figure 2. Matrix effect and precision of accuracy for spiked SIL targets in feces. (a) Absolute matrix effect (AME), the dash lines point out the range of 80-120%. (b) Relative matrix effect (RME), the dashed line indicates the RME at 15%. (c) The precision (RSD%) of the accuracy among four different donors, the dashed line indicates the RSD% at 15%.

In conclusion, by validating the method with selected SILs, we explored the linear dynamic range of different classes of compounds measured in plasma and feces, and also demonstrated that our method has good precision, accuracy and acceptable recovery and recovery repeatability. Additionally, the matrix effect of plasma and feces were assessed with selected SILs. In our validation, we used the original terms AME and RME to describe the matrix effect evaluation to avoid confusion. An AME value above 100% indicates ion enhancement, and less than 100% indicates ion suppression[31]. Although most of the bioanalytical laboratories use 80%-120% as the criteria for AME[46], besides the acceptable RME criteria (< 15%), there is no admissible value suggested by the EMA guideline. Therefore, it demonstrates that guaranteeing the reproducibility of AME is more critical for measurable compounds in bioanalytical method validation. Our validation data shows that L-lactic acid-$^{13}C_3$ and citric acid-d4 in plasma, and indole-d_5-3-acetic acid in feces have RME larger than 15%. To elucidate the impact of RME on the reproducibility of quantification, the precision (RSD %) of accuracy for spiked SILs are plotted for plasma and feces in figure 3C and figure 4C, respectively. The RSD accuracy values in both matrices align with the RME trends. The three targets with larger RME have accuracy RSD % above 15%, indicating that a high RME affects the accuracy and reproducibility of measurements among samples.

It is worth noting that, as suggested by the EMA guideline, it is possible to compensate for both AME and RME with internal standards in targeted metabolomics. In untargeted metabolomics, however, this approach is not feasible due to the unknown identity of some features and the lack of appropriate internal standards. To ensure the accuracy and reliability of data detection and interpretation, it is imperative to obtain information on the RME of all detected features in untargeted metabolomics measurements. With validation utilizing a wide diversity of SILs, we have highlighted the problem of the matrix effect variation in plasma and feces, while a comprehensive analysis of matrix effect variation for all detected features is still missing. Hence, how to evaluate or at least monitor the overall matrix effect variability in one or different types of matrices in untargeted metabolomics is a highly relevant problem to be addressed.

3.2 Matrix effect monitoring with PCI compounds

In order to monitor the overall AME and RME for plasma and fecal samples, we have developed a PCI approach using xenobiotic compounds. The infusion profile of each PCI compound was acquired with different plasma samples (n = 4), different fecal samples (n = 4), and blank samples in both positive and negative ion modes. The matrix effect profile (MEP) of each sample assessed with every PCI compound was generated, and distinct MEPs were obtained for different samples with all PCI compounds (Figure S9). Those MEPs were utilized to assess the AME and RME in plasma and feces, as described in the data processing section (2.6).

3.2.1 AME monitoring with PCI compounds

To ensure a fair assessment of AME and RME, a PCI compound-independent MEP was generated for each individual plasma and fecal sample. The averaged MEP intensity ($\overline{\mathrm{ME}_S}$) was calculated for each sample to form the PCI compounds-independent MEP (represented by the solid line in Figure S10-S13). The MEP variation plots with different individuals were created in both polarities accordingly, and the variation range among different individuals is represented by the shaded area in Figure 3a. Additionally, the averaged MEP intensity of the four samples was used to construct a real-time profile of AME (represented by the solid line in Figure 3a).

The AME profile provides a qualitative evaluation of the matrix effect in plasma and feces. Ion enhancement was rare in both matrices, while ion suppression was observed in specific regions of plasma and almost the entire chromatogram of feces. Severe ion suppression occurred before 1 min regardless of matrix and polarity, likely caused by unretained nonvolatile solutes such as highly polar metabolites and ionic species (e.g., inorganic electrolytes, salts)[47,48].

In plasma, the matrix effect dropped below 60% at around 1.6 min in both polarities and at around 4.6 min in positive and 1.2 min in negative polarity. The mass spectrum in those regions was inspected and showed a high signal of citric acid (RT at 1.58 min) and lactic acid (RT at 1.25 min) in at least one of those plasma samples (Figure S14-S15). This suggests that citric acid and lactic acid are most likely the causes of the drastic signal decrease observed at around 1.6 min and 1.2 min, as a high concentration of co-eluting compounds has been considered one of the prime factors to induce ion

suppression[48]. Nevertheless, no other feature with a high signal was recorded around 4.6 min in the plasma samples, presumably due to an undetected compound or compounds outside our targeted mass range (60-800 Da).We suspect that compounds with higher masses could be suppressing the signal of small molecules we detected during this elution time[49]. Furthermore, a high signal of EDTA was detected in plasma samples at approximately 1 min. This suggests that EDTA, a widely used anticoagulant, is a contributing factor to the significant ion suppression observed in plasma, which is consistent with reviewed literature[50,51]. Phospholipids, a recognized source of matrix effect in plasma[52,53], were not observed in our study, since the lipids elute after 7 min, when the LC flow was diverted to waste.

Similar to plasma, lipids are also considered as one of the major sources of matrix effect in feces[54]. However, compared to plasma, the matrix complexity of feces makes it more challenging to investigate the sources of ion suppression. We zoomed in on the mass spectrum where the most severe ion suppression occurred in feces (around 1 and 3 min) (Figure S16-S17), but we only putatively matched the prominent signal observed at around 3 min in positive polarity with phenylalanine according to our in-house target library. Further efforts would be required to identify the co-eluting compounds that induce matrix effect in feces, but this was considered beyond the scope of this study.

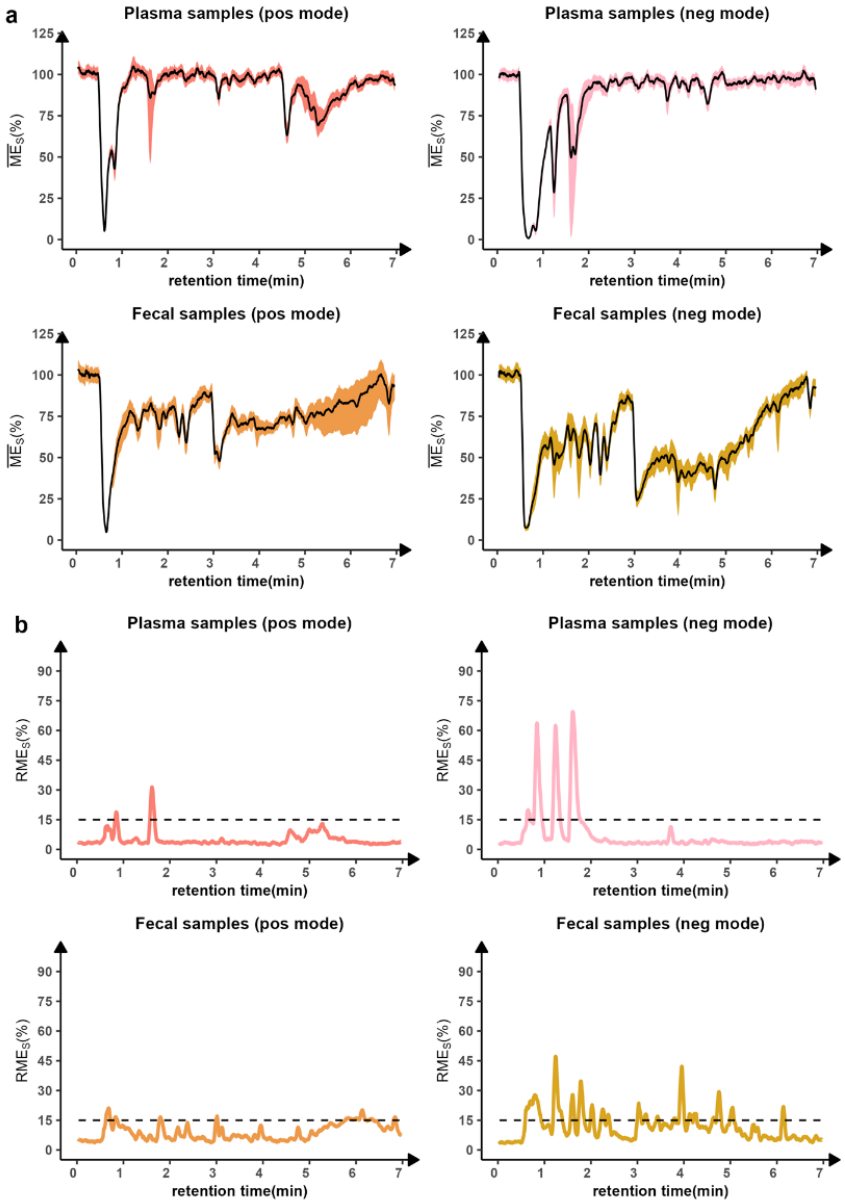


Figure 3. AME and RME profile in plasma and feces. (a) AME monitoring of plasma and feces using samples from four individuals in positive and negative mode. The solid line represents the averaged absolute matrix effect profile (MEP), and the shaded area shows the MEP variations among different individuals. (b) RME monitoring in plasma and feces using samples from four individuals in positive and negative mode.

3.2.2 RME monitoring with PCI compounds

The variation in AME (shaded area in Figure 3A) shows the matrix diversity of plasma and feces between different individuals. Accordingly, the RSD% of the AME indicates the RME of the entire runs (Figure 3B). In positive ion mode, the monitored RME in plasma and feces remains around or below 15% throughout almost the entire chromatogram. However, around 1.6 min in plasma, the RME exceeds over 30%, which is likely due to a large concentration variability of citric acid in those samples. In negative ion mode, there are more regions with high monitored RME in both plasma and feces. Three major spikes in the RME plot, up to 60%, are observed in plasma, and two of them are probably caused by high concentration variability of lactic acid and citric acid. In feces, the RME fluctuates within 45% in most regions. The RME overview demonstrates that it is reasonable to compare the detected signals in plasma or feces from different donors across most regions of the chromatogram, regardless of severe AME. Still, caution should be exercised for certain regions, particularly in negative ion mode.

To validate the accuracy of using PCI compounds to monitor RME, we extracted the monitored RME values at specific time points matching the RT of the spiked SILs, and compared them to the RME assessed with spiked SILs (Figure 4). The results reveal consistency between the RME monitored by PCI compounds and the RME assessed using spiked SILs. In plasma, both evaluation methods demonstrated that L-lactic acid-$^{13}C_3$ and citric acid-d_4 had an RSD% around 30%, while the other SILs had acceptable RSD% (< 15%). In feces, both methods indicated that indole-d_5-3-acetic acid had high variability (RSD > 15%). These results demonstrate that using PCI compounds for RME evaluation is comparable to spiking SILs, making it a compelling approach to evaluate RME for both known targets and unknown features.

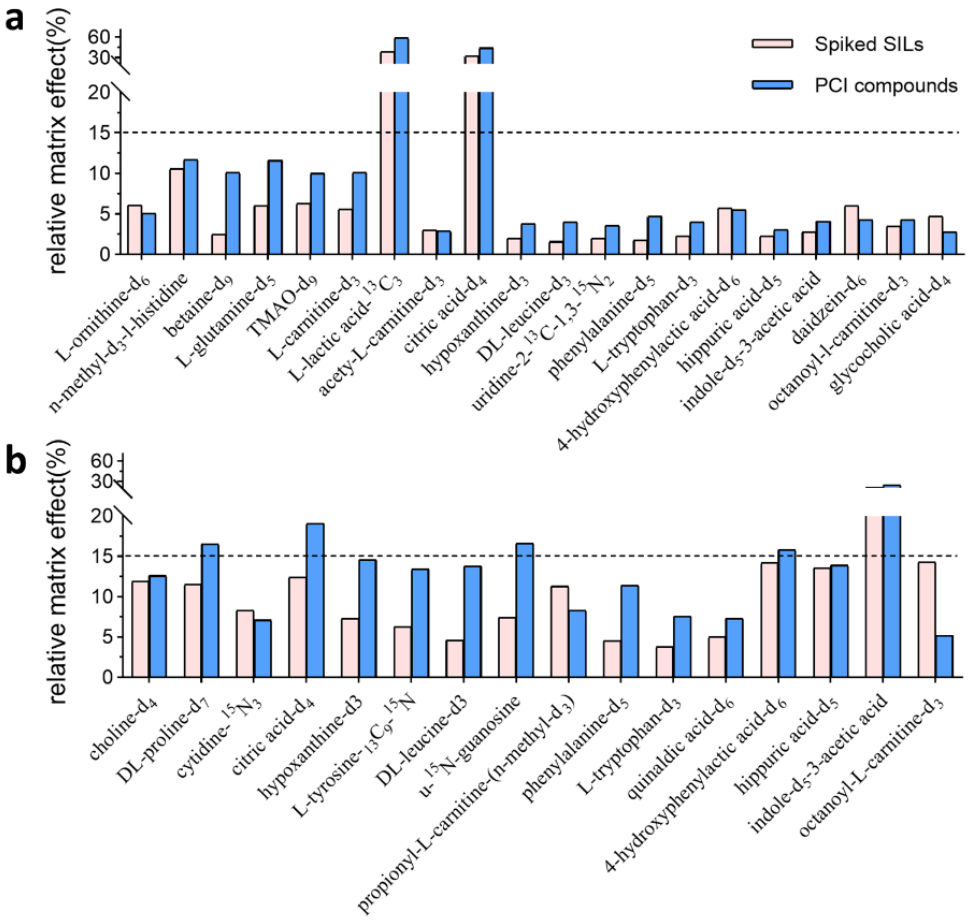


Figure 4. Comparison of RME evaluated with spiked SILs and PCI compounds in plasma (a) and feces (b). The averaged RME data from different concentrations of the spiked SILs were used. For the SILs that are detectable in both polarities, the selected polarity is consistent between the two methods.

3.3 RME monitoring application to targets included in an in-house LC-MS library

Together with the LC-MS untargeted method, an in-house targeted library containing retention time and accurate mass information was established by measuring commercially available authentic standards. The library included 305 targets that eluted before 6 minutes, and those targets were distributed across various classes, including amines, benzenoids, organic acids, indoles, nucleosides and nucleotides, and bile acids. In light of the effectiveness of PCI compounds on RME monitoring, we predicted the RME of the 305 targets based on their RT and the acquired RME profiles in plasma and fecal samples, respectively. Figure 5A provides an overview of the predicted RME for

55 targets that are only measurable in positive ion mode and 25 targets that are only detectable in negative ion mode (refer to table S9 for more information about the targets and predicted RME values). As expected, there were more targets within a caution zone (15% < RME <=30%) in feces than in plasma. A higher proportion of targets in negative ion mode were predicted to be affected by sample diversity compared to positive ion mode. In plasma, only one target (glycolic acid, with an RT of 0.80 min) detected in negative ion mode shows RSD > 30%. Figure 5B presents the predicted RME of the 225 targets that are detectable in both positive and negative modes (refer to table S10 for more information about the targets and predicted RME values). In general, we observed that more targets are susceptible to the matrix effect diversity in negative ion mode than in positive ion mode, regardless of matrix type, and that there are more targets predicted with a RME > 15% in feces compared to plasma. For the targets that are detectable in both ionization modes, the predicted RME needs to be taken into account when selecting the appropriate polarity for quantitation, along with other parameters such as signal-to-noise ratio.

Although the predicted RME in our study is only based on four individual plasma and fecal samples and only the predicted value at the apex of the peak was used (without considering the peak width), our results demonstrate the potential of the PCI approach in identifying the regions of caution regarding to RME and predicting RME for both known and unknown features based on their retention times. Some high-resolution MS instruments have the option to continuously infuse a compound after LC separation for calibration purposes, which also can be utilized for ME monitoring. However, including multiple PCI compounds enhances the possibility of capturing various ME profiles compared to using just one, as demonstrated in our study, especially for fecal samples in negative mode (Figure S6D). Moreover, ideal PCI compounds should have exogenous *m/z* values that do not interfere with the targets of interest and should not induce significant additional ME. Overall we strongly recommend applying a PCI approach both during the method development and routine studies. Its application in method development aids in identifying cautionary areas in the chromatography that

suffer from matrix effect. This information is crucial in guiding the optimization of specific LC parameters, such as gradient and injection amount, to minimize matrix effect. Additionally, the routine application of PCI is crucial in improving the reliability of data interpretation in studies that apply untargeted methods, particularly for cohorts with an anticipated range of abnormal or unusually high compound concentrations. For instance, plasma samples from individuals with kidney disease may exhibit wider zones of ion suppression due to the specific nature of the health condition, which involves the accumulation of various compounds in the blood. Likewise, when comparing fecal samples from individuals consuming a ketogenic diet with those from vegetarians, it is important to examine ion suppression due to the high variation in fat content.

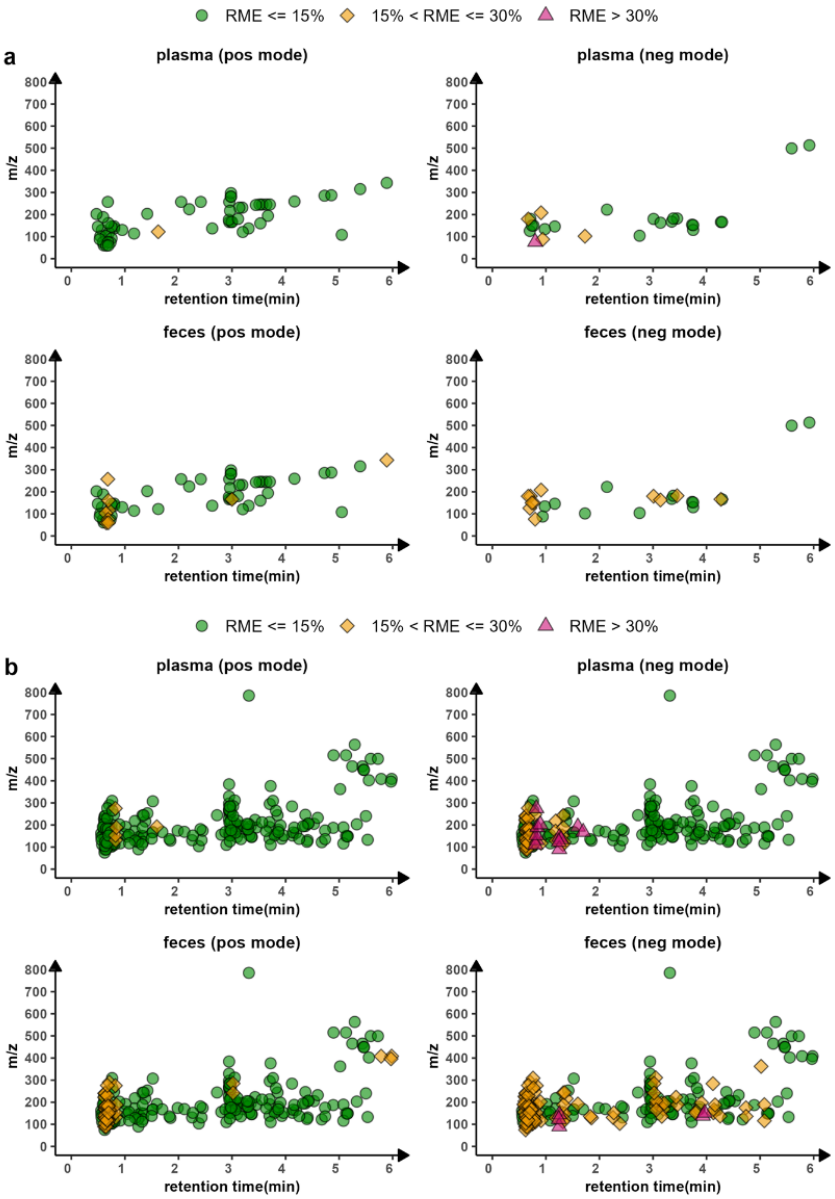


Figure 5. RME assessment of targets included in the in-house library. (a) Predicted RME by PCI compounds for targets that are only detectable in one polarity mode. Positive (55 targets) or negative (25 targets). (b) Predicted RME by PCI compounds for targets that are detectable in both positive and negative modes (225 targets).

4. Conclusion

In this study, we propose a comprehensive framework for the development of untargeted metabolomics methods with a PCI approach for matrix effect monitoring. To the best of our knowledge, our research is the first study offering practical strategies that combine

the optimization of sample injection amount and reconstitution solvent, performance validation and matrix effect evaluation in the development of an untargeted metabolomics method.

Our study demonstrates that optimization of sample injection amount, utilizing ion transmission monitoring techniques such as ITC in the TripleTOF system, is critical for balancing metabolite coverage and signal linearity. Additionally, considering specific LC gradients and metabolite classes of interest, it is crucial to optimize reconstitution solvents to avoid potential issues of peak shape distortion and poor solubility in untargeted methods. Furthermore, validating an untargeted metabolomics method in a targeted manner provides valuable insights into the analytical performance of the method, including the linear dynamic range, precision, accuracy, recovery and matrix effect.

To address the challenge of matrix effect, we highly recommend implementing a PCI approach during the development phase of an untargeted metabolomics method and suggest also applying it in routine studies. Our results demonstrate that the PCI approach effectively monitors the matrix effect for plasma and fecal samples, allowing the identification of regions with high matrix effect variation in the untargeted metabolomics method that should be interpreted with caution. More impressively, the PCI approach yields comparable RME data when compared to the traditional post-extraction spiking method, making it a compelling technique for assessing RME for both known targets and unknown features detected in untargeted metabolomics. This approach shows great promise for generating reliable data from an untargeted method and advancing quantitative analysis in untargeted metabolomics.

Acknowledgements

Pingping Zhu Would like to acknowledge the China Scholarship Council (CSC, No. 201906240049), and the EARLYFIT project funded by NWO project 16490. This research was part of the Netherlands X-omics Initiative and partially founded by NWO project 184.034.019. This publication is part of the 'Building the infrastructure for Exposome research: Exposome-Scan' project (with project number 175.2019.032) of the

program ‘Investment Grant NWO Large', which is funded by the Dutch Research Council (NWO). Additionally, the Authors acknowledge Mariyana Savova from Leiden University for her invaluable advice and assistance in fecal samples preparation.

Reference:

1. Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. Identification of bioactive metabolites using activity metabolomics. Nat Rev Mol Cell Biol. 2019;20(6):353-367. doi:10.1038/s41580-019-0108-4
2. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Published online 2016. doi:10.1038/nrd.2016.32
3. Pang H, Jia W, Hu Z. Emerging Applications of Metabolomics in Clinical Pharmacology. Clin Pharmacol Ther |. 2019;106. doi:10.1002/cpt.1538
4. Lacalle-Bergeron L, Izquierdo-Sandoval D, Sancho J V., López FJ, Hernández F, Portolés T. Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers. TrAC - Trends Anal Chem. 2021;135. doi:10.1016/j.trac.2020.116161
5. Wei S, Wei Y, Gong Y, et al. Metabolomics as a valid analytical technique in environmental exposure research: application and progress. Metabolomics. 123AD;1:35. doi:10.1007/s11306-022-01895-7
6. Bedair M, Glenn KC. Evaluation of the use of untargeted metabolomics in the safety assessment of genetically modified crops. Metabolomics. 123AD;1:111. doi:10.1007/s11306-020-01733-8
7. Miggiels P, Wouters B, van Westen GJP, Dubbelman AC, Hankemeier T. Novel technologies for metabolomics: More for less. TrAC - Trends Anal Chem. 2019;120. doi:10.1016/J.TRAC.2018.11.021
8. Pezzatti J, González-Ruiz V, Codesido S, et al. A scoring approach for multi-platform acquisition in metabolomics. J Chromatogr A. 2019;1592:47-54. doi:10.1016/j.chroma.2019.01.023
9. Broadhurst D, Goodacre R, Reinke SN, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics. 2018;14(6):1-17. doi:10.1007/s11306-018-1367-3
10. Dudzik D, Barbas-Bernardos C, García A, Barbas C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. J Pharm Biomed Anal. 2018;147:149-173. doi:10.1016/j.jpba.2017.07.044
11. Pezzatti J, Boccard J, Codesido S, et al. Implementation of liquid chromatography–high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: A tutorial. Anal Chim Acta. 2020;1105:28-44. doi:10.1016/J.ACA.2019.12.062
12. Vuckovic D, Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal Bioanal Chem. 2012;403:1523-1548. doi:10.1007/s00216-012-6039-y
13. Wu ZE, Kruger MC, Cooper GJS, Poppitt SD, Fraser K. Tissue-specific sample dilution: An important parameter to optimise prior to untargeted lc-ms metabolomics. Metabolites. 2019;9(7). doi:10.3390/metabo9070124
14. Manier SK, Meyer MR. Impact of the used solvent on the reconstitution efficiency of evaporated biosamples for untargeted metabolomics studies. Metabolomics. 2020;16(3):1-6. doi:10.1007/s11306-019-1631-1
15. Lindahl A, Sääf S, Lehtiö J, Nordström A. Tuning Metabolome Coverage in Reversed Phase LC-MS Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition. Anal Chem. 2017;89(14):7356-7364. doi:10.1021/acs.analchem.7b00475
16. Murray K. J K., Boyd R. K., Eberlin M. N., Langley G. J., Li L. NY. Guide to Achieving Reliable Quantitative LC-MS Measurements. Vol 68.; 2013.
17. Schulze B, Bader T, Seitz W, Winzenbacher R. Column bleed in the analysis of highly polar substances: an overlooked aspect in HRMS. Anal Bioanal Chem. 2020;412(20):4837-4847. doi:10.1007/s00216-020-02387-0
18. Kirwan JA, Gika H, Beger RD, et al. Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management on behalf of the metabolomics Quality Assurance and Quality Control Consortium (mQACC). 2022;18:70. doi:10.1007/s11306-022-01926-3
19. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. J Chromatogr A. 2014;1353:99-105. doi:10.1016/j.chroma.2014.04.071
20. Neveu V, Moussy A, Rouaix HH, et al. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. Nucleic Acids Res. 2016;45. doi:10.1093/nar/gkw980

21. Hosseinkhani F, Huang L, Dubbelman AC, Guled F, Harms AC, Hankemeier T. Systematic Evaluation of HILIC Stationary Phases for Global Metabolomics of Human Plasma. Metab 2022, Vol 12, Page 165. 2022;12(2):165. doi:10.3390/METABO12020165
22. Leogrande P, Jardines D, Dayamin Martinez-Brito |, et al. Metabolomics workflow as a driven tool for rapid detection of metabolites in doping analysis. Development and validation. Published online 2021. doi:10.1002/rcm.9217
23. Bussche J Vanden, Marzorati M, Laukens D, Vanhaecke L. Validated High Resolution Mass Spectrometry-Based Approach for Metabolomic Fingerprinting of the Human Gut Phenotype. Anal Chem. 2015;87(21):10927-10934. doi:10.1021/ACS.ANALCHEM.5B02688
24. González O, Blanco ME, Iriarte G, Bartolomé L, Maguregui MI, Alonso RM. Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect. J Chromatogr A. 2014;1353:10-27. doi:10.1016/J.CHROMA.2014.03.077
25. Cortese M, Gigliobianco MR, Magnoni F, Censi R, Di Martino P. Compensate for or minimize matrix effects? strategies for overcoming matrix effects in liquid chromatography-mass spectrometry technique: A tutorial review. Molecules. 2020;25(13). doi:10.3390/molecules25133047
26. Kebarle P, Tang L. From Ions in Solution To Ions in the Gas Phase. Anal Chem. 1993;65(22):972A-986A. doi:10.1021/ac00070a715
27. Panuwet P, Hunter RE, D'Souza PE, et al. Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring. Crit Rev Anal Chem. 2016;46(2):93-105. doi:10.1080/10408347.2014.980775
28. Cappiello A, Famiglini G, Palma P, Pierini E, Termopoli V, Trufelli H. Overcoming matrix effects in liquid chromatography-mass spectrometry. Anal Chem. 2008;80(23):9343-9348. doi:10.1021/ac8018312
29. Tulipani S, Mora-Cubillos X, Jaúregui O, et al. New and Vintage Solutions To Enhance the Plasma Metabolome Coverage by LC-ESI-MS Untargeted Metabolomics: The Not-So-Simple Process of Method Performance Evaluation. Anal Chem. 2015;87. doi:10.1021/ac503031d
30. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem. 2003;75(13):3019-3030. doi:10.1021/ac020361s
31. Viswanathan CT, Bansal S, Booth B, et al. Commentary Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays. Published online 2007. doi:10.1007/s11095-007-9291-7
32. European Medicines Agency. Guideline on bioanalytical method validation. 2011;44(July 2011):1-23.
33. Bonfiglio R, King RC, Olah T V, Merkle K. The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. Rapid Commun Mass Spectrom. 1999;13(12):1175-1185. doi:10.1002/(SICI)1097-0231(19990630)13:12<1175::AID-RCM639>3.0.CO;2-0
34. González O, Van Vliet M, Damen CWN, Van Der Kloet FM, Vreeken RJ, Hankemeier T. Matrix Effect Compensation in Small-Molecule Profiling for an LC-TOF Platform Using Multicomponent Postcolumn Infusion. Anal Chem. 2015;87(12):5921-5929. doi:10.1021/ac504268y
35. Liao HW, Chen GY, Wu MS, Liao WC, Lin CH, Kuo CH. Development of a Postcolumn Infused-Internal Standard Liquid Chromatography Mass Spectrometry Method for Quantitative Metabolomics Studies. J Proteome Res. 2017;16(2):1097-1104. doi:10.1021/acs.jproteome.6b01011
36. Rossmann J, Renner LD, Oertel R, El-Armouche A. Post-column infusion of internal standard quantification for liquid chromatography-electrospray ionization-tandem mass spectrometry analysis – Pharmaceuticals in urine as example approach. J Chromatogr A. 2018;1535:80-87. doi:10.1016/j.chroma.2018.01.001
37. Huang CS, Kuo CH, Lo C, et al. Investigating the association of the biogenic amine profile in urine with therapeutic response to neoadjuvant chemotherapy in breast cancer patients. J Proteome Res. 2020;19(10):4061-4070. doi:10.1021/acs.jproteome.0c00362
38. Patti GJ. Separation strategies for untargeted metabolomics. J Sep Sci. 2011;34(24):3460-3469. doi:10.1002/JSSC.201100532
39. González O, Dubbelman A-C, Hankemeier T. Postcolumn Infusion as a Quality Control Tool for LC-MS-Based Analysis. Cite This J Am Soc Mass Spectrom. 2022;2022. doi:10.1021/jasms.2c00022
40. Chepyala D, Kuo HC, Su KY, et al. Improved Dried Blood Spot-Based Metabolomics Analysis by a Postcolumn Infused-Internal Standard Assisted Liquid Chromatography-Electrospray Ionization Mass Spectrometry Method. Anal Chem. 2019;91(16):10702-10712. doi:10.1021/acs.analchem.9b02050
41. Tisler S, Pattison DI, Christensen JH. Correction of Matrix Effects for Reliable Non-target Screening LC–ESI–MS Analysis of Wastewater. Anal Chem. Published online June 7, 2021:acs.analchem.1c00357. doi:10.1021/acs.analchem.1c00357

42. Bueschl C, Kluger B, Lemmens M, et al. A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research. Metabolomics. 2014;10:754-769. doi:10.1007/s11306-013-0611-0
43. Ćeranić A, Bueschl C, Doppler M, et al. Enhanced Metabolome Coverage and Evaluation of Matrix Effects by the Use of Experimental-Condition-Matched 13C-Labeled Biological Samples in Isotope-Assisted LC-HRMS Metabolomics. Metab 2020, Vol 10, Page 434. 2020;10(11):434. doi:10.3390/METABO10110434
44. Mizuno H, Ueda K, Kobayashi Y, et al. The great importance of normalization of LC-MS data for highly-accurate non-targeted metabolomics. Published online 2016. doi:10.1002/bmc.3864
45. Hosseinkhani F, Dubbelman AC, Karu N, Harms AC, Hankemeier T. Towards Standards for Human Fecal Sample Preparation in Targeted and Untargeted LC-HRMS Studies. Metab 2021, Vol 11, Page 364. 2021;11(6):364. doi:10.3390/METABO11060364
46. Kollipara S, Bende G, Agarwal N, Varshney B, Paliwal J. International Guidelines for Bioanalytical Method Validation: A Comparison and Discussion on Current Scenario. doi:10.1007/s10337-010-1869-2
47. King R, Bonfiglio R, Fernandez-Metzler C, Miller-Stein C, Olah T. Mechanistic investigation of ionization suppression in electrospray ionization. J Am Soc Mass Spectrom. 2000;11(11):942-950. doi:10.1016/S1044-0305(00)00163-X
48. Antignac JP, De Wasch K, Monteau F, De Brabander H, Andre F, Le Bizec B. The ion suppression phenomenon in liquid chromatography-mass spectrometry and its consequences in the field of residue analysis. Anal Chim Acta. 2005;529(1-2 SPEC. ISS.):129-136. doi:10.1016/J.ACA.2004.08.055
49. Sterner JL, Johnston M V, Nicol GR, Ridge DP. Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples. J MASS Spectrom J Mass Spectrom. 2000;35:385-391. doi:10.1002/(SICI)1096-9888(200003)35:3<385::AID-JMS947>3.0.CO;2-O
50. Chin C, Zhang ZP, Karnes HT. A study of matrix effects on an LC/MS/MS assay for olanzapine and desmethyl olanzapine. J Pharm Biomed Anal. 2004;35(5):1149-1167. doi:10.1016/J.JPBA.2004.01.005
51. Ghosh C, Shinde CP, Chakraborty BS. Influence of ionization source design on matrix effects during LC-ESI-MS/MS analysis. J Chromatogr B Anal Technol Biomed Life Sci. 2012;893-894:193-200. doi:10.1016/J.JCHROMB.2012.03.012
52. Ismaiel OA, Zhang T, Jenkins RG, Karnes HT. Investigation of endogenous blood plasma phospholipids, cholesterol and glycerides that contribute to matrix effects in bioanalysis by liquid chromatography/mass spectrometry. J Chromatogr B Anal Technol Biomed Life Sci. 2010;878(31):3303-3316. doi:10.1016/J.JCHROMB.2010.10.012
53. Ismaiel O, Halquist M, Elmamly M, Shalaby A, Thomaskaren H. Monitoring phospholipids for assessment of ion enhancement and ion suppression in ESI and APCI LC/MS/MS for chlorpheniramine in human plasma and the importance of multiple source matrix effect evaluations. J Chromatogr B. 2008;875(2):333-343. doi:10.1016/j.jchromb.2008.08.032
54. Jeffery J, Vincent ZJ, Ayling RM, Lewis SJ. Development and validation of a liquid chromatography tandem mass spectrometry assay for the measurement of faecal metronidazole. Clin Biochem. 2017;50(6):323-330. doi:10.1016/J.CLINBIOCHEM.2016.11.031

Supplementary Material

Method optimization: reconstitution solvent and injection amount

In developing the RPLC-MS untargeted method, we first optimized the reconstitution solvent for feces and plasma by reconstituting dried extracted samples in 100% water, water/ACN (9:1, v/v), and water/ACN (8:2, v/v). The chromatographic peak shape and height of representative metabolites were evaluated for reconstitution solvent selection.

Figure S2a presents the chromatography of the early eluting endogenous metabolite guanine and the late eluter deoxycholic acid (DCA) in both plasma and fecal samples with three reconstitution solvents. Peak shape deterioration of guanine was observed with the increased ACN in the injection solvent. It has been reported that peak distortion of polar metabolites in RPLC is caused by viscosity and elution strength mismatch between the injection solvent and the mobile phase, and that a potent injection solvent causes peak splitting and fronting due to the migration time differences between the analyte, injected solvent and mobile phase[1–3]. In our gradient, the proportion of mobile phase B reaches 20% around 2.5 min, at which the strength and viscosity of the mobile phase is equal to the injection solvent. This explains why peak distortion only occurred for metabolites eluting before 2.5 min, like guanine, when injected with 20% ACN in our method, and the peak shape of metabolites eluting after 2.5 min, like DCA, were retained.

Interestingly, although the peak shape of DCA was not distorted, the peak height increased along with the increment of ACN in both feces and plasma. We also observed that the signal boost of DCA caused by ACN is much higher in feces than in plasma. This phenomenon suggests that adding ACN to the reconstitution solvent facilitates the solubility of DCA regardless of sample type. However, the signal improvement is sample or concentration related, as the concentration of DCA in feces is higher than in plasma.

Lindahl *et al.* reported that when comparing reconstitution with different proportions of MeOH in water, using 100% water as a reconstitution solvent increases the response of metabolites with logP < 5 [4]. However, the data we obtained about DCA (logP = 3.5)[5] disagrees with their conclusion and restricts the range to lower logP. To investigate the correlation of metabolite polarity and the effect of reconstitution solvent on metabolite response in our method, we compared the peak areas of 26 endogenous metabolites in feces with three injection solvents (Figure S2b). These 26 metabolites were widely distributed in retention time (RT), representing a wide range of polarity. As shown in Figure 1b, the peak areas were comparable in three reconstitution solvents for the metabolites eluting before 6 min, while after that, the peak areas increased (area

percentage higher than 33%) with raised ACN ratio in the reconstitution solvents. Together with other studies[2,4], it proves that the effect of the injection solvents on the peak shape of polar metabolites and on the solubility of less polar metabolites is ubiquitous in the RPLC method, and our result suggests that the affected regions are dependent on the injection solvent and the LC gradient. Thus, it is necessary to consider the peak shape and solubility of metabolites of interest when selecting the injection solvent for RPLC untargeted methods. We decided to use 100% water with 0.1% FA as the final injection solvent as we are more interested in polar and semi-polar metabolites. For metabolites eluting after 6 min, more caution is needed in interpreting the results in clinical applications, given their potentially incomplete solubility. This aspect was not further examined in plasma, as the peak distortion and solubility are more dependent on the LC method and metabolites rather than the sample itself.

Next, we investigated the effect of the dilution factor (DF) and injection volume on metabolome coverage and signal saturation for plasma and fecal samples. In total, we compared three DFs (1:6, 1:3, 1:2) in plasma (v/v) and two DFs (1:6, 1:3) in feces (mg/v), along with two injection volumes (1µL and 2µL) in both matrices. The combination of DF and injection volume is presented as DF_µL.

To compare the metabolome coverage, we integrated 47 identified metabolites (details provided in Table S6) with diverse endogenous abundance, for all the DF and injection volume combinations. Figure S3a and Figure S3b present the metabolite distributions of different combinations in plasma and feces, respectively. The 47 metabolites are detectable in all the combinations, and as expected, the overall signal increased with a higher injected concentration in both matrices. Potential signal saturation was evaluated by visualizing the count conversion factors (ccf) plots for each combination (Figure 3C-F). The %ccf value reflects the degree to which the dynamic ion transmission control (ITC) has modulated the MS ion current in TOF MS spectrum for that point in the chromatogram. This dynamic ITC is a feature of the TripleTOF 6600 system which functions to reduce the risk of detector saturation when high amounts of MS ion current are present in a particular sample. As ion current reaches a predefined upper TargetTIC (total ion chromatogram) signal, the ion current is modulated down by adjusting a lens

voltage in the front end of the instrument. This adjustment is done scan-by-scan based on feedback from the detection system, the higher the ion current goes above the TargetTIC, the higher the modulation applied. The %ccf value shows the % of ion beam modulation that was used in each spectrum. ITC considers both the total ion current and the ion current of a dominant ion. Although the peak area written to the datafile is corrected back to the value it would have been at 100% ion current, ideally, to ensure quantitative linearity, an ion load that is causing maximal ion modulation or exceeding the ion modulation limit (2%) should be avoided.

The %ccf plots of injections in plasma show that, in general, the ion current is modulated to a greater extent with a higher injected concentration (Figure S3c). It reaches the lowest value around 0.85 min, where except DF6_1µL, the other combinations either surpass (DF2_2µL, DF2_1µL, DF3_2µL) or stay near (DF3_1µL, DF6_2µL) the limit of the ion modulation (Figure S3d). In fecal injections, the %ccf plots show that the ion modulation reaches the limit around 4.7 min (Figure S3f). DF6_2µL and DF3_2µL reach the limit of ion modulation, and DF3_1µL stays close to the limit (Figure S3e).

Additionally, to assess the relationship between injection amount and matrix effect, post-column infusion (PCI) compounds were used to monitor the matrix effect for all the injection combinations in plasma. As expected, a higher-injected sample amount resulted in greater ion suppression in the area that suffers from matrix effect. DF2_2 µL caused the most pronounced ion suppression and DF6_1 µL experienced the least ion suppression (Figure S4a-b). The % ccf plots exhibited similar fluctuation trends to the matrix effect profiles in those samples (Figure S4c-d), indicating that both reflect the amount of ion current injected into the MS system. When more ions are injected, there is increased ionization competition in the source, and the ITC undergoes greater modulation to reduce the risk of detector saturation.

According to the selected 47 metabolites, there is no metabolite coverage difference among all the combinations. Hence, DF6_1µL is the optimum injection condition for both plasma and feces to avoid potential signal saturation and reduce matrix effect throughout the chromatogram. However, in plasma there are more metabolites with low abundance, and DF6_1µL results in higher number of peaks with area below 1000

(Figure S3a), which may cause repeatability and quantification issues for these metabolites. Therefore, to achieve a compromise between the signal sensitivity, detector saturation and matrix effect, DF3_1μL and DF6_1μL should be selected as the final injection condition for plasma and feces, respectively. Nevertheless, considering the limited fecal sample availability in some clinical studies, DF8_1μL was utilized for further method validation in feces. The assessment of metabolome coverage was not extended with DF8_1μL, as we do not anticipate significant signal reduction with the injection volume fold changes less than 1.5 times.

Reference:

1. Castells CB, Castells RC. Peak distortion in reversed-phase liquid chromatography as a consequence of viscosity differences between sample solvent and mobile phase. *J Chromatogr A*. 1998;805:55-61. doi:10.1016/s0021-9673(98)00042-9
2. Keunchkarian S, Reta M, Romero L, Castells C. Effect of sample solvent on the chromatographic peak shape of analytes eluted under reversed-phase liquid chromatogaphic conditions. *J Chromatogr A*. 2006;1119(1-2):20-28. doi:10.1016/J.CHROMA.2006.02.006
3. Vanmiddlesworth BJ, Dorsey JG. Quantifying injection solvent effects in reversed-phase liquid chromatography. *J Chromatogr A*. 2012;1236:77-89. doi:10.1016/j.chroma.2012.02.075
4. Lindahl A, Sääf S, Lehtiö J, Nordström A. Tuning Metabolome Coverage in Reversed Phase LC-MS Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition. *Anal Chem*. 2017;89(14):7356-7364. doi:10.1021/acs.analchem.7b00475
5. Roda A, Minutello A, Angellotti MA, Fini A. Bile acid structure-activity relationship: Evaluation of bile acid lipophilicity using 1-octanol/water partition coefficient and reverse phase HPLC. *J Lipid Res*. 1990;31(8):1433-1443. doi:10.1016/s0022-2275(20)42614-8

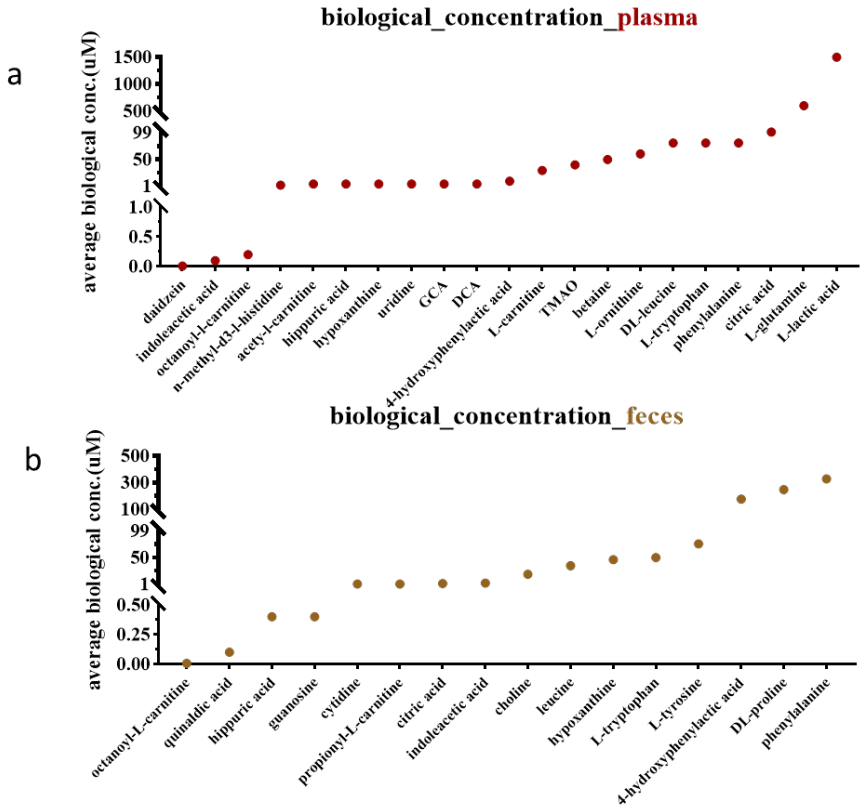


Figure S1. Distribution of endogenous compound abundance for selected metabolites. (a) metabolites selected for plasma. Abundance levels were determined by referring to the reported concentration in HMDB (b) metabolites selected for feces. Abundance levels were determined by standard addition to the collected fecal samples due to the inconsistency of the reported concentration.

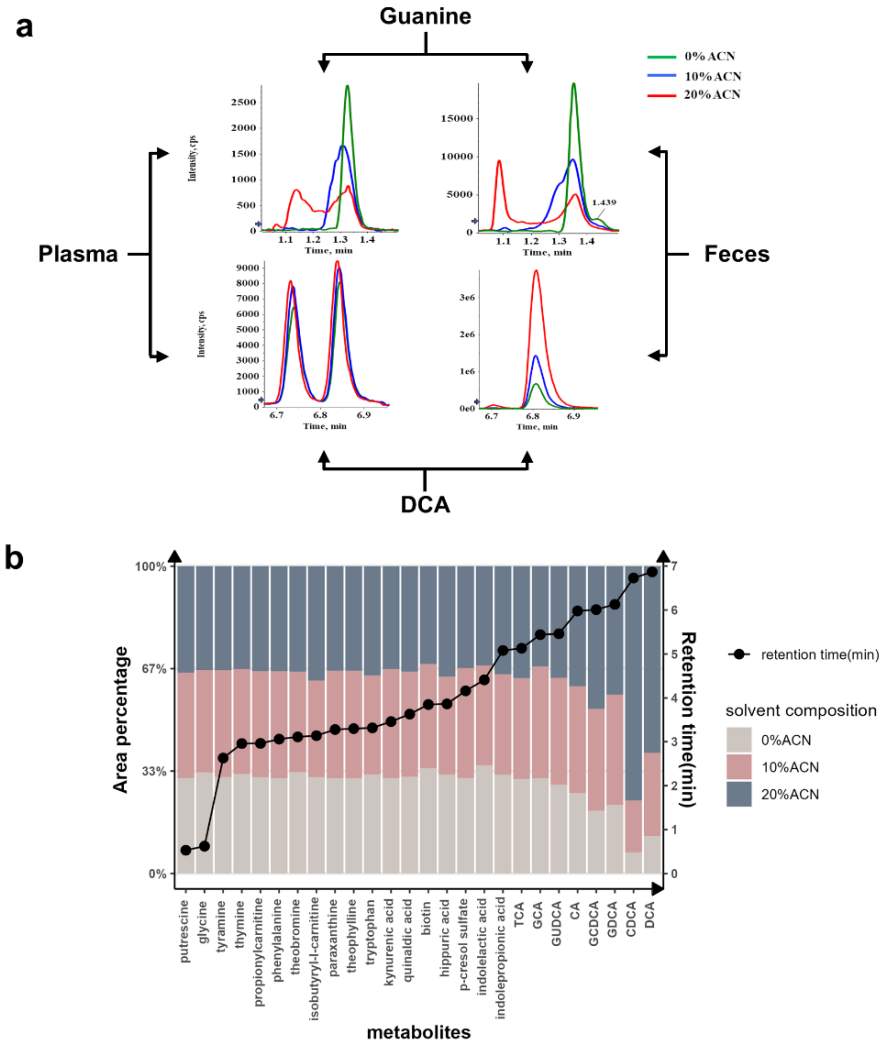


Figure S2. Reconstitution solvent comparison in plasma and feces. Plasma (dilution factor (DF) 2) and feces (DF6) were injected with 1µL in three reconstitution solvents. (a) Chromatogram of guanine and DCA. Guanine as an example of how reconstitution solvent affects the peak shape for polar metabolites; DCA (RT at 6.83min, the other peak at around 6.73 is chenodeoxycholic acid) as an example of how reconstitution solvent affects the signal for less polar metabolites. (b) Peak area comparison of 26 metabolites (bile acids were analyzed in negative mode) in feces injected with three reconstitution solvents. The peak area percentage for each metabolite was calculated by dividing the peak area in a specific reconstitution solvent by the sum of the peak area in three reconstitution solvents. If there is around 33% area percentage for each condition, no obvious solubility issue was observed for that metabolite. To avoid the impact of peak shape distortion on integration, only metabolites with acceptable peak shape in 20% ACN were involved in the comparison (metabolites which eluted around void volume or after 2.5 min). Complete names for the abbreviation: taurocholic acid (TCA),

glycocholic acid (GCA), cholic acid (CA), glycochenodeoxycholic acid (GCDCA), glycodeoxycholic acid (GDCA), chenodeoxycholic acid (CDCA).

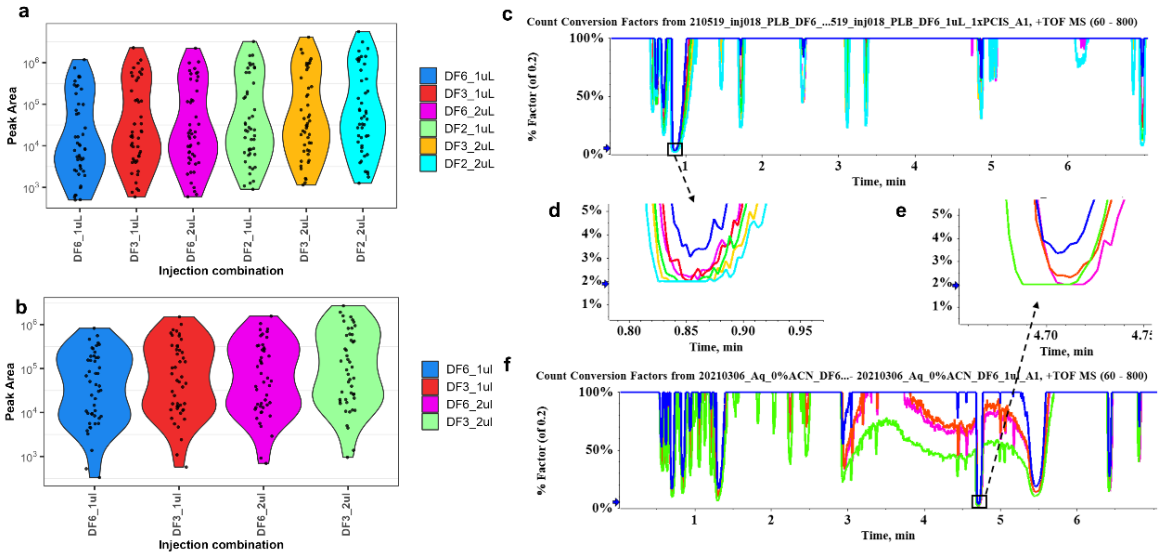


Figure S3. Injection concentration optimization for plasma and feces. Plasma and feces were reconstituted in 0.1% FA in water. In the metabolite coverage comparison, 47 metabolites (measured in positive ion mode), all eluting before 6 min to avoid the solubility issue of late-eluting compounds, were included. (a) Metabolite coverage comparison of different injected concentrations in plasma. (b) Metabolite coverage comparison of different injected concentrations in feces. (c) %Ccf plots for different injected concentrations in plasma. (d) Zoom in on the %ccf plots in the region of highest ion current modulation in plasma. (e) Zoom in on the %ccf plots in the region of highest ion current modulation in feces. (f) %Ccf plots for different injected concentrations in feces.

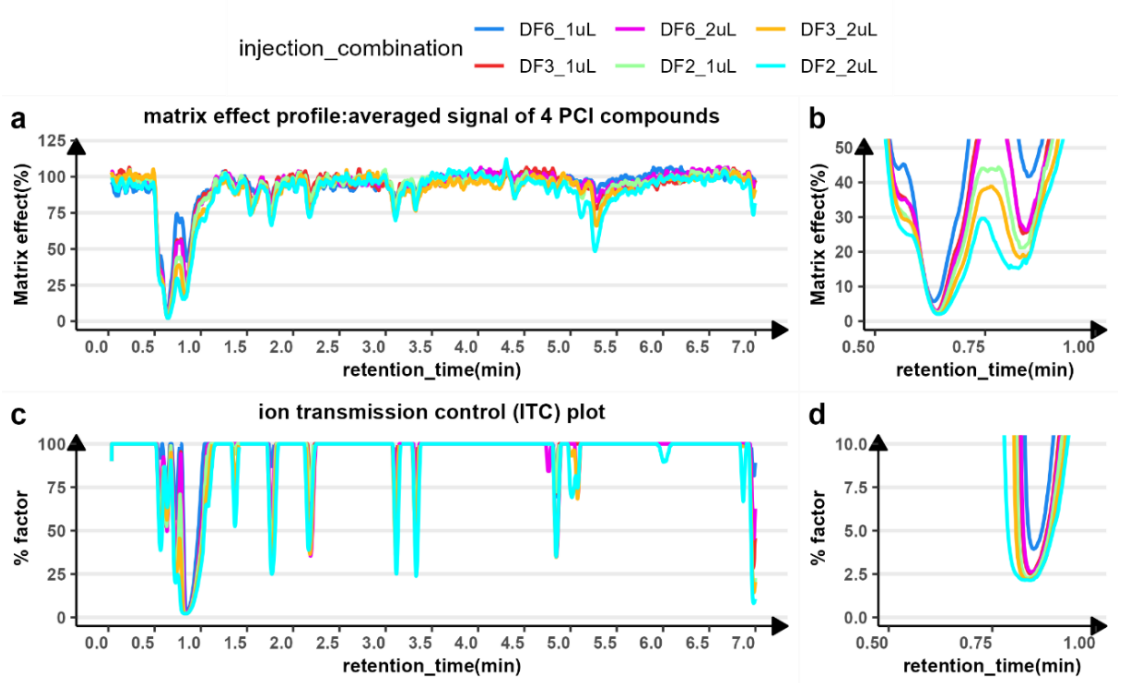


Figure S4. Matrix effect profile and ion transmission control (ITC) plot for all the injection combinations of plasma in positive mode.(a) matrix effect profile presented with the averaged signal of 4 PCI; (b) Zoom in on the plot of (a) in the region of highest ion suppression ; (c) ITC plots of all the injection combination;(d) Zoom in on the plot of (c) in the region of highest ion current modulation.

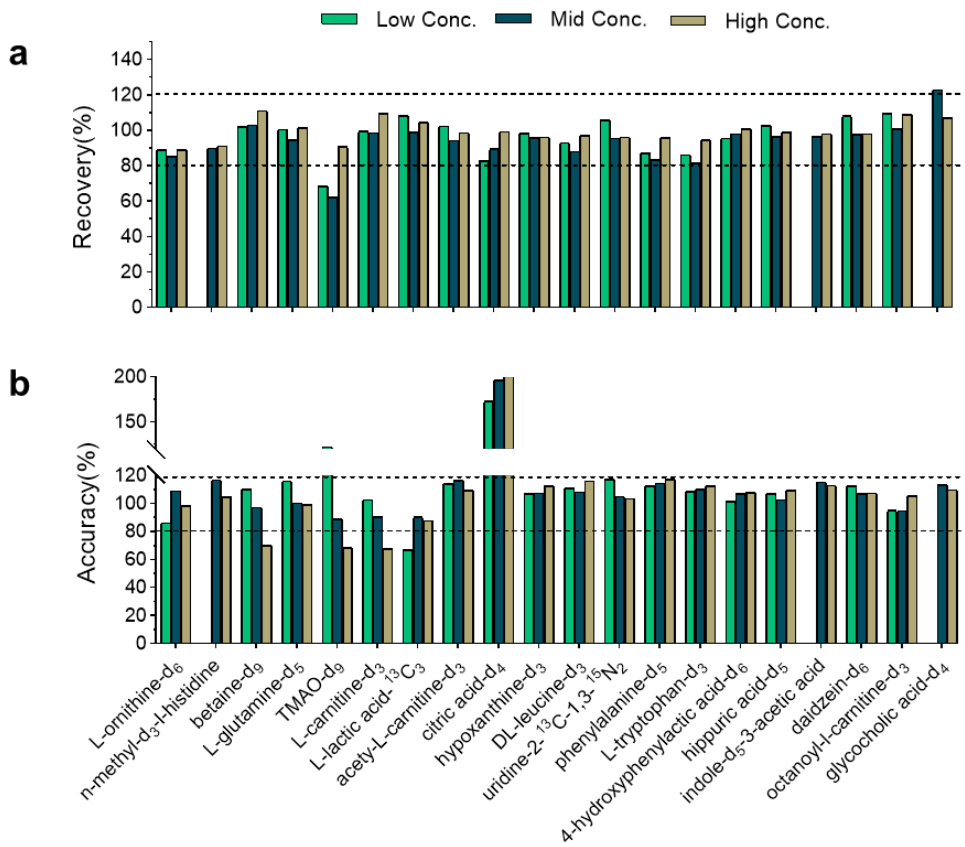


Figure S5. Recovery (a) and accuracy (b) of the spiked SILs in plasma

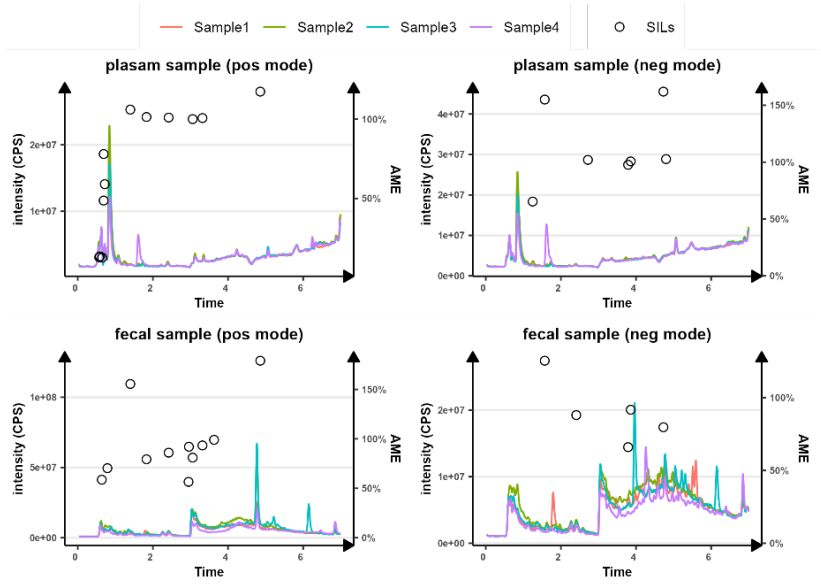


Figure S6. Averaged AME of spiked SILs and TIC intensity in plasma and feces. The averaged AME was calculated from different concentrations of spiked SILs.

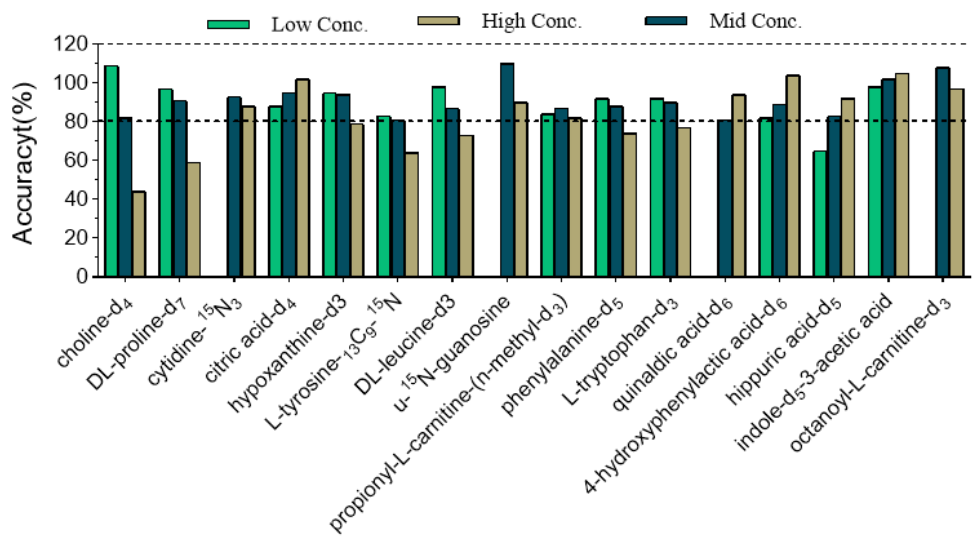


Figure S7. Accuracy of spiked SILs in feces

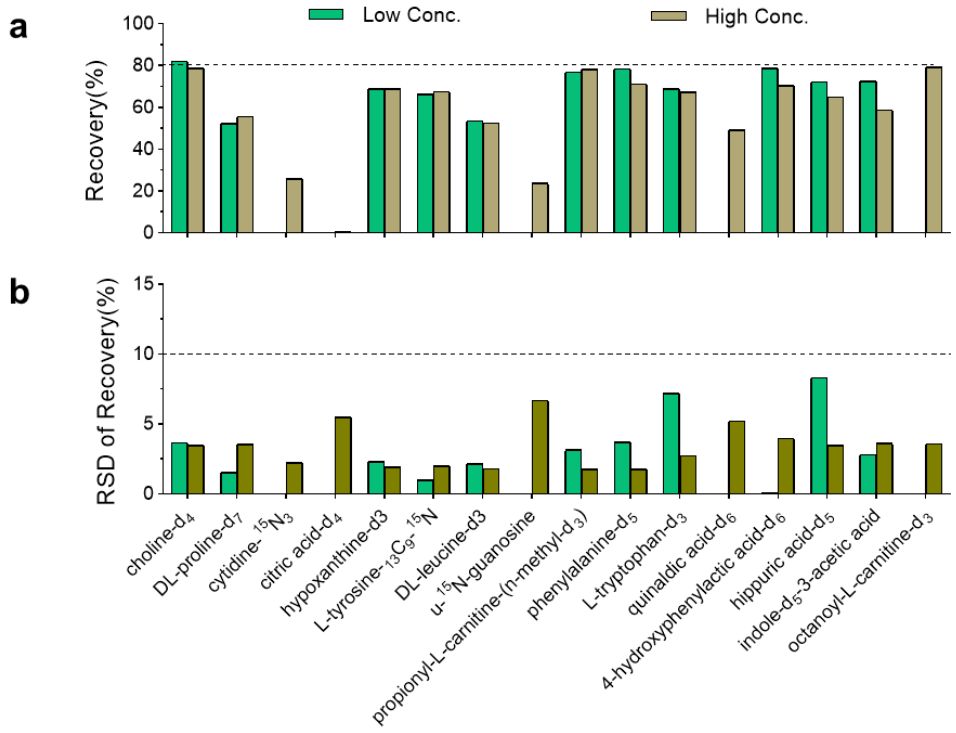


Figure S8. Recovery (a) and RSD of recovery (b) of the spiked SILs in feces

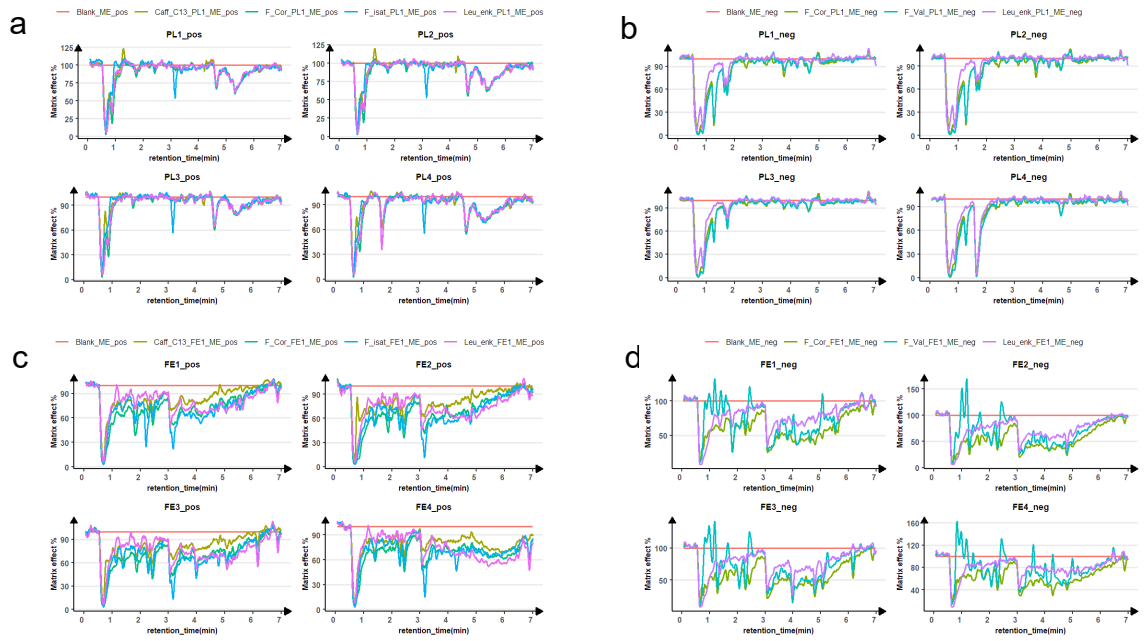


Figure S9. Matrix effect profile (MEP) of each sample with all the PCI compounds. (a) plasma in positive mode. (b) plasma in negative mode. (c) feces in positive mode. (d) feces in negative mode.

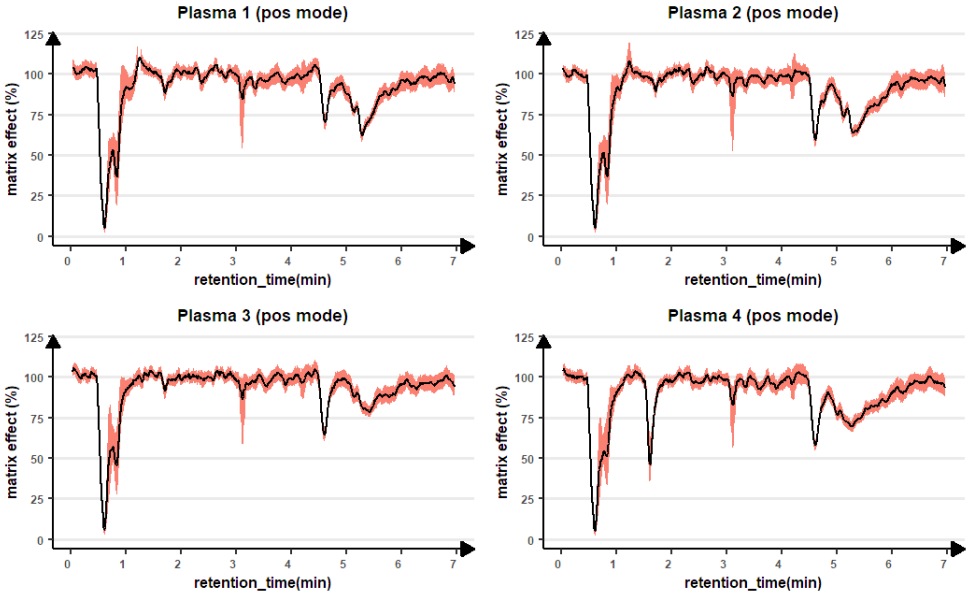


Figure S10. Overlapped MEPs of four PCI compounds for each plasma sample in positive mode.

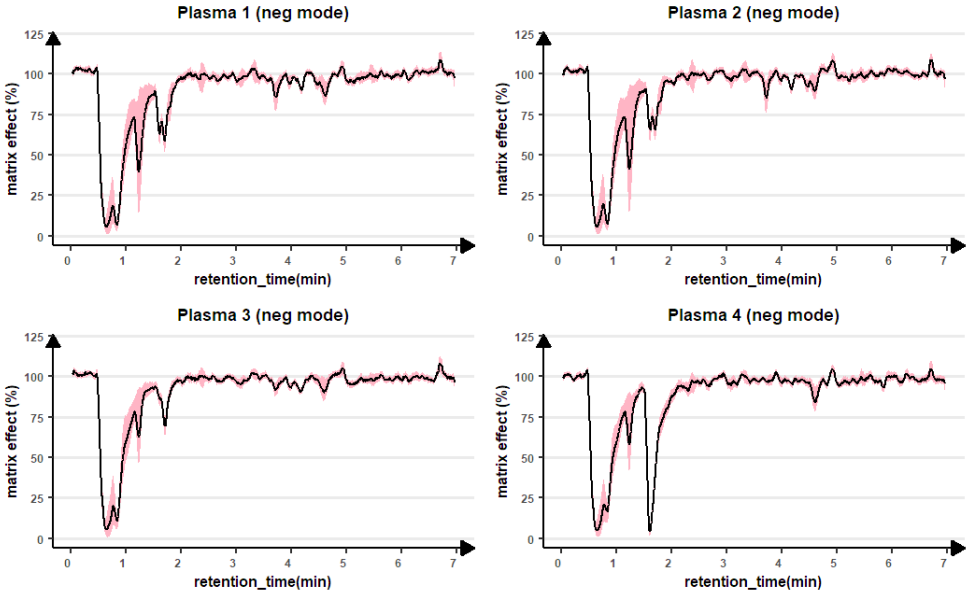


Figure S11. Overlapped MEPs of four PCI compounds for each plasma sample in negative mode.

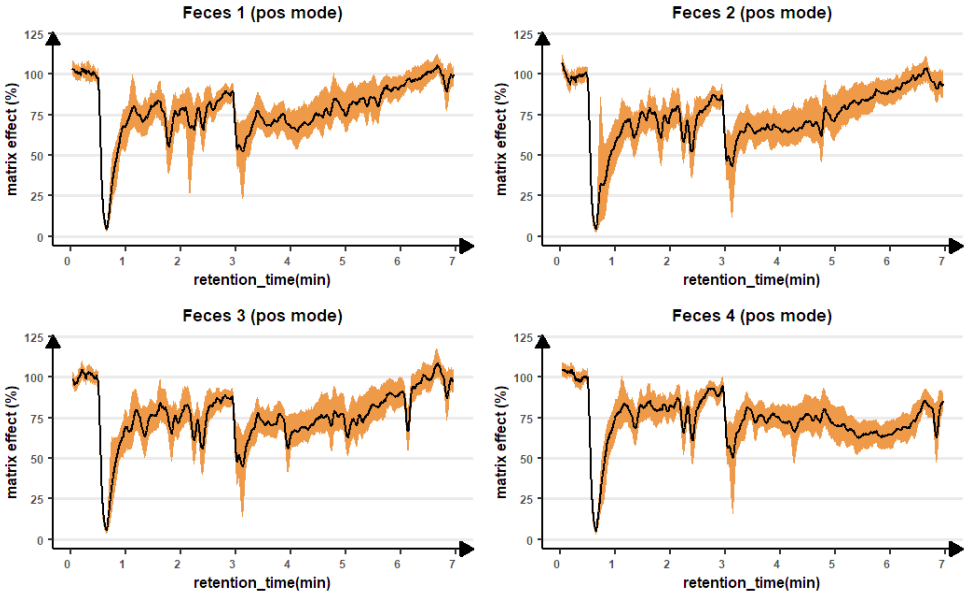


Figure S12. Overlapped MEPs of four PCI compounds for each fecal sample in positive mode.

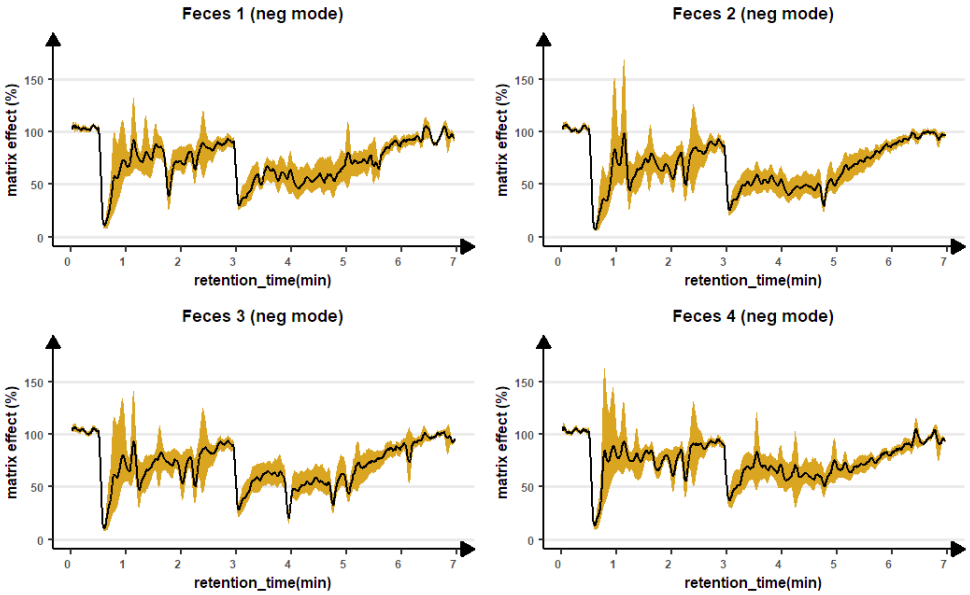


Figure S13. Overlapped MEPs of four PCI compounds for each fecal sample in negative mode.

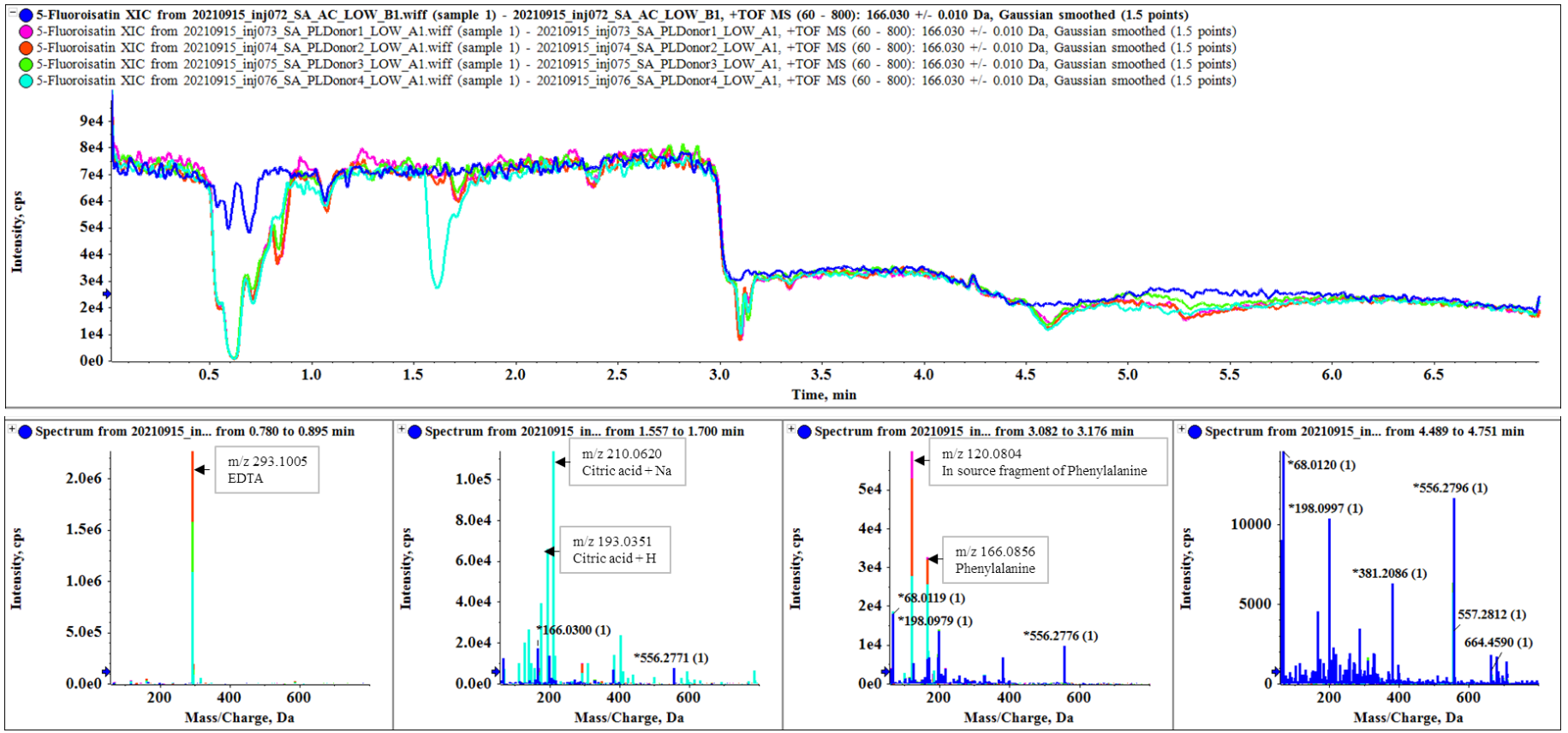


Figure S14. Mass spectrum inspection of the suppressed areas in plasma samples with 5-fluoroisatin in positive mode

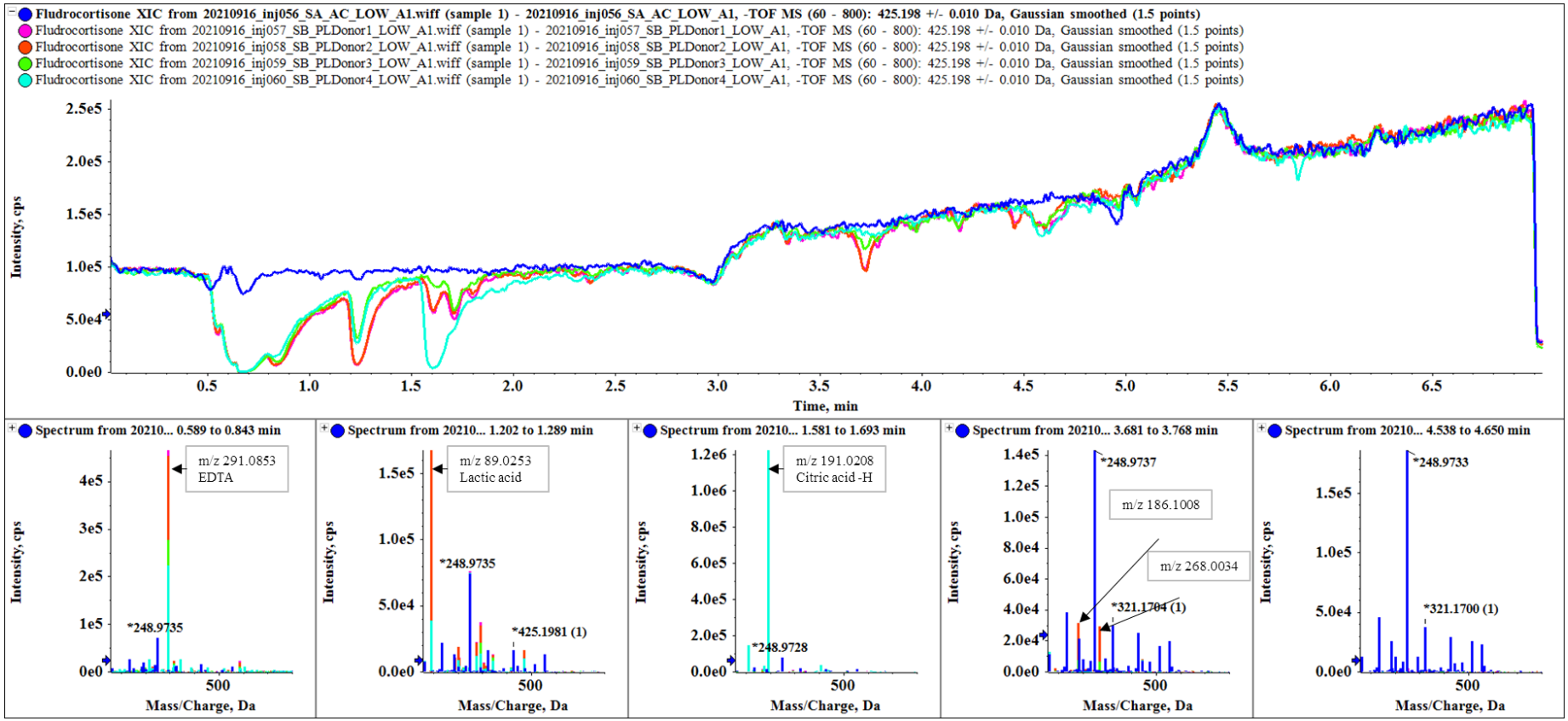


Figure S15. Mass spectrum inspection of the suppressed areas in plasma samples with fludrocortisone in negative mod

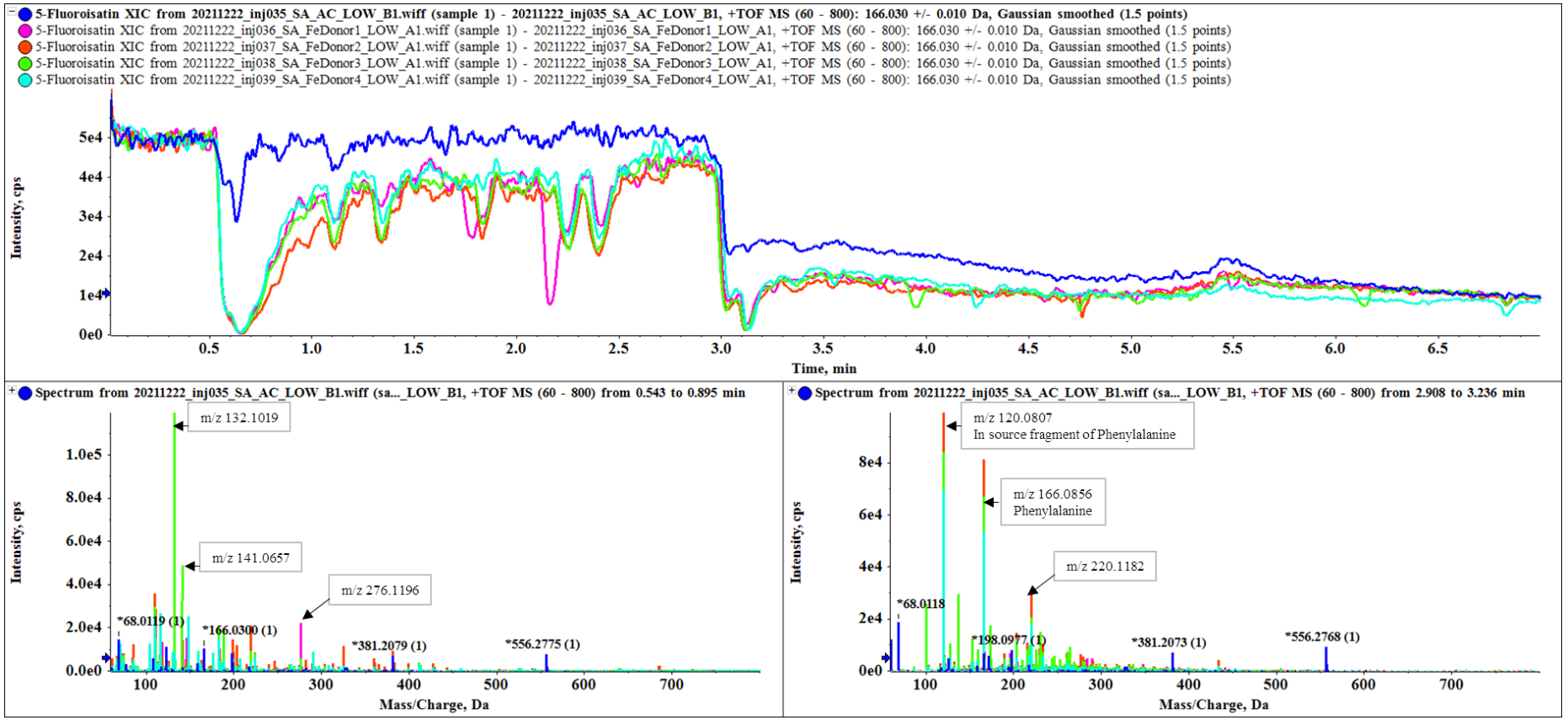


Figure S16. Mass spectrum inspection of the suppressed areas in fecal samples with 5-fluoroisatin in positive mode

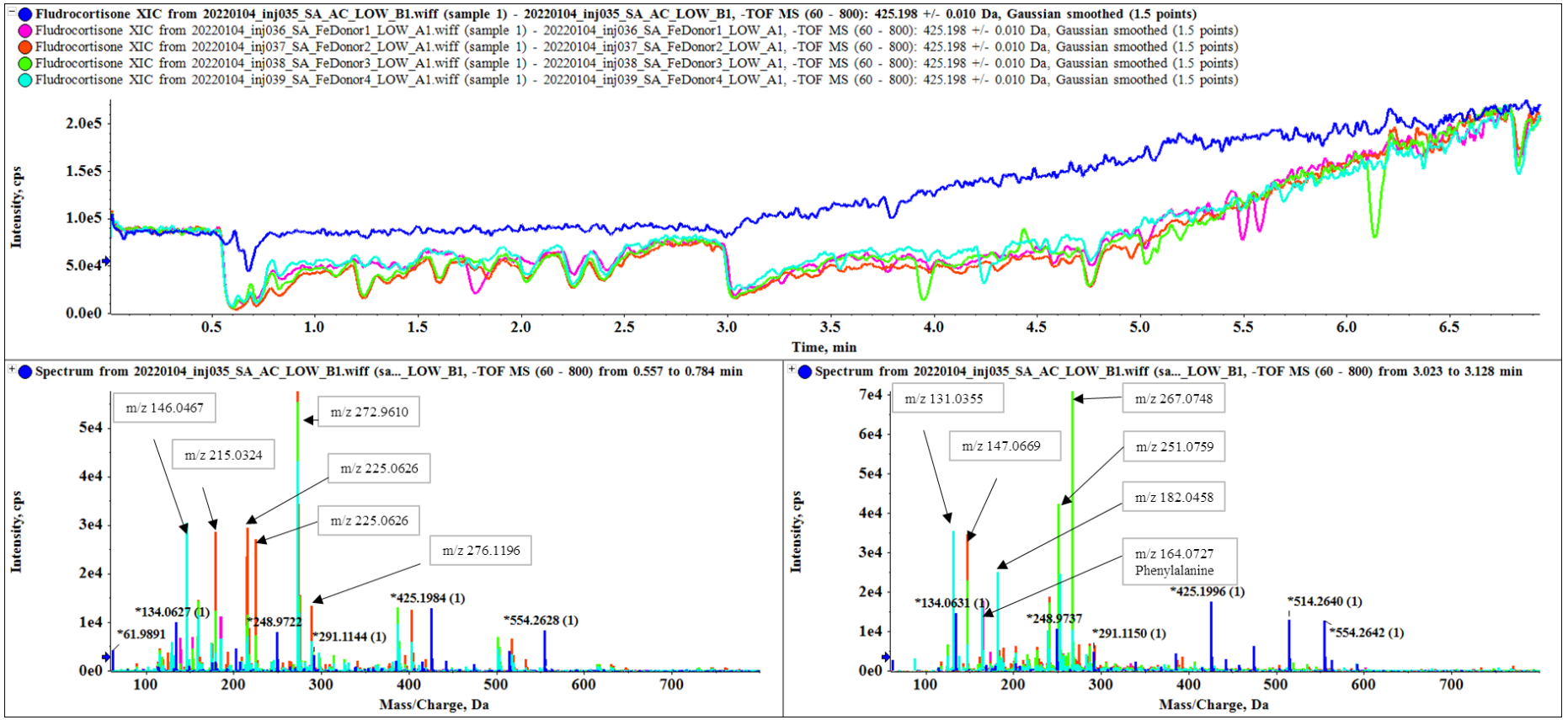


Figure S17. Mass spectrum inspection of the suppressed areas in fecal samples with fludrocortisone in negative mod

Table S1: general information and stock solution preparation for all the authentic standards

Compound Name	Compound Formula	Da	sub_class	CAS number	supplier	product number	stock mM	Solvent	RT/min	usage
L-ornithine-d_6	C3D9NO	84.1249	Amino acids, peptides, and analogues	347841-40-1	CDN	D-3759	250.0	H2O	0.57	plasma validation
n-methyl-d_3-l-histidine	C7D3H9N3O2	172.1040	Amino acids, peptides, and analogues	91037-48-8	CDN	D-2493	25.0	H2O	0.59	plasma validation
Betaine-d_9	C5H2D9NO2	126.1355	Amino acids, peptides, and analogues	285979-85-3	CDN	D-3352	250.0	H2O	0.62	plasma validation
L-glutamine-d_5	C5H5D5N2O3	151.1005	Amino acids, peptides, and analogues	14341-78-7	cambridge Isotope laboratories	DLM-1826-0.1	250.0	H2O (1% NH3.H2O)	0.65	plasma validation
TMAO-d_9	C3D9NO	84.1249	Aminoxides	1161070-49-0	cambridge Isotope laboratories	DLM-4779-1	500.0	H2O	0.69	plasma validation
L-carnitine-d_3	C7H12D3NO3	164.1240	carnitine and Acetylcarnitines	350818-62-1	CDN	D-5069	125.0	H2O	0.69	plasma validation
L-lactic acid-$^{13}C_3$	13C3H6O3	93.0418	organic acids and derivatives	201595-71-3	TRC	L113507	170.0	H2O	1.25	plasma validation
acety-L-carnitine-d_3	C9H14D3NO4	206.1346	carnitine and Acetylcarnitines	1334532-17-0	CDN	D-6534	50.0	H2O	1.40	plasma validation
citric acid-d_4	C6H4D4O7	196.0521	organic acids and derivatives	147664-83-3	cambridge Isotope laboratories	DLM-3487-0.5	1250.0	H2O	1.57	plasma and feces validation
hypoxanthine-d_3	C5D3HN4O	140.0636	Purines and purine derivatives	NA	cambridge Isotope laboratories	DLM-2923-0.1	62.5	10% MeOH (0.2M HCL)	1.83	plasma and feces validation
DL-leucine-d3	C6H10D3NO2	134.1135	Amino acids, peptides, and analogues	87828-86-2	CDN	D-2400	62.5	10% MeOH (1% NH3.H2O)	2.42	plasma and feces validation
uridine-2-^{13}C-1,3-$^{15}N_2$	C8[13C]1H12[15N]2O6	247.0670	Pyrimidine nucleosides	369656-75-7	TRC	U829907	31.3	H2O	2.72	plasma validation
phenylalanine-d_5	C9H6D5NO2	170.1104	Amino acids, peptides, and analogues	28466-89-7	CDN	D-1597	50.0	15% MeOH	3.06	plasma and feces validation
L-tryptophan-d_3	C11H9D3N2O2	207.1087	Indolyl carboxylic acids and derivatives	133519-78-5	CDN	D-7419	50.0	H2O (0.5% NH3.H2O)	3.32	plasma and feces validation
4-hydroxy-phenylactic acid-d_6	C8H2D6O3	158.0850	Benzoic acids and derivatives	100287-06-7	TRC	H949062	125.0	H2O (1.5% NH3.H2O)	3.79	plasma and feces validation
hippuric acid-d_5	C9H4D5NO3	184.0896	Benzoic acids and derivatives	53518-98-2	chem Cruz	sc-490158	12.5	H2O	3.86	plasma and feces validation
indole-d_5-3-acetic acid	C10H4D5NO2	180.0947	Indolyl carboxylic acids and derivatives	76937-78-5	TRC	I577344	1.3	MeOH	4.73	plasma and feces validation
daidzein-d_6	C15H4D6O4	260.0956	Isoflavonoids	291759-05-2	TRC	D103502	0.1	MeOH	4.80	plasma validation
octanoyl-l-carnitine-d_3	C15H26D3NO4	290.2285	carnitine and Acetylcarnitines	1334532-24-9	CDN	D-6651	2.5	H2O	4.86	plasma and feces validation
glycocholic acid-d_4	C26H39D4NO6	469.3341	Bile acids_conjugated	1201918-15-1	CDN	D-3878	62.5	MeOH	5.46	plasma validation
deoxycholic acid-d_4	C24H36D4O4	396.3178	Bile acids_unconjugated	112076-61-6	CDN	D-2941	62.5	MeOH	6.87	plasma validation
choline-d_4	C5H9D4NO	107.1248	Quaternary ammonium salts	285979-70-6	CDN	D-2464	475.0	H2O	0.64	feces validation

DL-proline-d_7	C5H2D7NO2	122.1073	Amino acids, peptides, and analogues	65807-21-8	cambridge Isotope laboratories	DLM-2657-0	310.0	H2O	0.79	feces validation
cytidine-$^{15}N_3$	C9H13[15N]3O5	246.0766	Pyrimidine nucleosides	NA	cambridge Isotope laboratories	NLM-3797-50	100.0	H2O	1.40	feces validation
L-tyrosine-$^{13}C_9$-^{15}N	[13C]9H11[15N]O3	191.1011	Amino acids, peptides, and analogues	129077-96-9	cambridge Isotope laboratories	CNLM-439-H-0.1	47.7	H2O (0.2M HCL)	2.41	feces validation
u-^{15}N-guanosine	C10H13[15N]5O5	288.0768	Purine nucleosides	NA	Silantes	125303603	10.0	H2O (0.1M HCL)	2.95	feces validation
propionyl-L-carnitine-(n-methyl-d_3)	C10H16D3NO4	220.1502	Fatty Acyls	1334532-24-9	CDN	D-6651	50.0	H2O	2.96	feces validation
quinaldic acid-d_6	C10HD6NO2	179.0853	Quinoline carboxylic acids	1219802-18-2	CDN	D-6514	50.0	H2O	3.63	feces validation
leucine-enkephalin	C28H37N5O7	555.2693	Amino acids, peptides, and analogues	81678-16-2	Sigma-Aldrich	L9133	1.8	H2O	/	PCI compound
fludrocortisone	C21H29FO5	380.1999	Hydroxysteroids	127-31-1	TRC	F428100	1.3	MeOH	/	PCI compound
5-fluoroisatin	C8H4FNO2	165.0226	Indolines	443-69-6	Sigma-Aldrich	366978	6.1	50% MeOH	/	PCI compound
caffeine-$^{13}C_3$	C5[13C]3H10N4O2	197.0904	Purines and purine derivatives	78072-66-9	TRC	C080101	2.5	50% MeOH	/	PCI compound
3-fluoro-DL-valine	C5H10FNO2	135.0696	Amino acids, peptides, and analogues	43163-94-6	Sigma-Aldrich	47581	14.8	50% MeOH	/	PCI compound
fludrocortisone-d_5	C21H24D5FO5	385.2313	Hydroxysteroids	NA	TRC	F428102	2.6	MeOH	5.04	internal standard
D-glucose-$^{13}C_6$, d_7	[13C]6H5D7O6	193.1275	Carbohydrates and carbohydrate conjugates	201417-01-8	TRC	G595001	15.5	50% MeOH	0.69	internal standard
5-chloroisatin	C8H4ClNO2	180.9931	Indolines	17630-76-1	chem Cruz	sc-254819	5.5	MeoH	4.95	internal standard
caffeine-d_9	C8HD9N4O2	203.1369	Purines and purine derivatives	722358-85-8	TRC	C080102	4.9	50% MeOH	3.64	internal standard
valine-d_8	C5H3D8NO2	125.1292	Amino acids, peptides, and analogues	203784-63-8	cambridge Isotope laboratories	DLM-488	24.0	MeOH	1.08	internal standard
cortisone-d_8	C21H20D8O5	368.2439	Hydroxysteroids	NA	TRC	C696502	6.8	MeOH	5.06	external standard

Table S2: calibration line of the SILs in plasma validation

Compound	**concentration of each calibration point (µM)**								
	Cal1	**Cal2**	**Cal3**	**Cal4**	**Cal5**	**Cal6**	**Cal7**	**Cal8**	**Cal9**
L-ornithine-d_6	0.600	1.200	3.000	6.000	30.00	60	300	600	1500
n-methyl-d_3-l-histidine	0.030	0.060	0.150	0.300	1.50	3	15	30	75
betaine-d_9	0.500	1.000	2.500	5.000	25.00	50	250	500	1250
L-glutamine-d_5	6.000	12.000	30.000	60.000	300.00	600	3000	6000	15000
TMAO-d_9	0.400	0.800	2.000	4.000	20.00	40	200	400	1000
L-carnitine-d_3	0.300	0.600	1.500	3.000	15.00	30	150	300	750
L-lactic acid-$^{13}C_3$	0.050	0.100	0.250	0.500	2.50	5	25	50	125
acety-L-carnitine-d_3	1.500	3.000	7.500	15.000	75.00	150	750	1500	3750
citric acid-d_4	1.000	2.000	5.000	10.000	50.00	100	500	1000	2500

hypoxanthine-d_3	0.050	0.100	0.250	0.500	2.50	5	25	50	125
DL-leucine-d_3	0.800	1.600	4.000	8.000	40.00	80	400	800	2000
uridine-2-^{13}C-1,3-$^{15}N_2$	0.050	0.100	0.250	0.500	2.50	5	25	50	125
phenylalanine-d_5	0.800	1.600	4.000	8.000	40.00	80	400	800	2000
L-tryptophan-d_3	0.800	1.600	4.000	8.000	40.00	80	400	800	2000
4-hydroxyphenylactic acid-d_6	0.010	0.020	0.050	0.100	0.50	1	5	10	25
hippuric acid-d_5	0.050	0.100	0.250	0.500	2.50	5	25	50	125
indole-d_5-3-acetic acid	0.010	0.020	0.050	0.100	0.50	1	5	10	25
daidzein-d_6	0.005	0.010	0.025	0.050	0.250	0.5	2.5	5	12.5
octanoyl-l-carnitine-d_3	0.002	0.004	0.010	0.020	0.10	0.2	1	2	5
glycocholic acid-d_4	0.050	0.100	0.250	0.500	2.50	5	25	50	125
deoxycholic acid-d_4	0.050	0.100	0.250	0.500	2.50	5	25	50	125

Table S3: calibration line of the SILs in feces validation

Compound	**concentration of each calibration point (µM)**										
	cal1	**cal2**	**cal3**	**cal4**	**cal5**	**cal6**	**cal7**	**cal8**	**cal9**	**cal10**	**cal11**
choline-d_4	0.190	0.380	0.950	1.900	3.800	9.500	19.000	38.000	95.000	190.000	475.000
DL-proline-d_7	2.480	4.960	12.400	24.800	49.600	124.000	248.000	496.000	1240.000	2480.000	6200.000
cytidine-$^{15}N_3$	0.010	0.020	0.050	0.100	0.200	0.500	1.000	2.000	5.000	10.000	25.000
citric acid-d_4	0.020	0.040	0.100	0.200	0.400	1.000	2.000	4.000	10.000	20.000	50.000
hypoxanthine-d3	0.460	0.920	2.300	4.600	9.200	23.000	46.000	92.000	230.000	460.000	1150.000
L-tyrosine-$_{13}C_9$-^{15}N	0.750	1.500	3.750	7.500	15.000	37.500	75.000	150.000	375.000	750.000	1875.000
DL-leucine-d3	0.350	0.700	1.750	3.500	7.000	17.500	35.000	70.000	175.000	350.000	875.000
u-^{15}N-guanosine	0.004	0.008	0.020	0.040	0.080	0.200	0.400	0.800	2.000	4.000	10.000
propionyl-L-carnitine-(n-methyl-d_3)	0.010	0.020	0.050	0.100	0.200	0.500	1.000	2.000	5.000	10.000	25.000
phenylalanine-d_5	0.652	1.304	3.260	6.520	13.040	32.600	65.200	130.400	326.000	652.000	1630.000
L-tryptophan-d_3	0.500	1.000	2.500	5.000	10.000	25.000	50.000	100.000	250.000	500.000	1250.000
quinaldic acid-d_6	0.001	0.002	0.005	0.010	0.020	0.050	0.100	0.200	0.500	1.000	2.500
4-hydroxyphenylactic acid-d_6	1.760	3.520	8.800	17.600	35.200	88.000	176.000	352.000	880.000	1760.000	4400.000
hippuric acid-d_5	0.004	0.008	0.020	0.040	0.080	0.200	0.400	0.800	2.000	4.000	10.000
indole-d_5-3-acetic acid	0.030	0.060	0.150	0.300	0.600	1.500	3.000	6.000	15.000	30.000	75.000
octanoyl-L-carnitine-d_3	0.000	0.000	0.000	0.001	0.001	0.004	0.007	0.014	0.035	0.070	0.175

Table S4: final concentration of each PCIS compound in the infused mixture solutions

PCI compound	**concentration (ng/mL)**	**applied polarity**
included Leucine-enkephalin	212.4	positive mode
fludrocortisone	154.0	positive mode
5-fluoroisatin	1069.2	positive mode
caffeine-13C3	219.2	positive mode
Leucine-enkephalin	344.9	negative mode

fludrocortisone	371.0	negative mode
3-Fluoro-DL-valine	4264.3	negative mode

Table S5: dilution ratio calculation of feces

$V_{reconstitution}$	50	mL
$m_{dry\ feces}$	20	mg
LLE extraction steps		
Step1		
$V_{MeOH\ (\mu L)}$	108	
$V_{MilliQ(\mu L)}$	36	
$V_{MTBE(\mu L)}$	60	
$V_{total\ (\mu L)}$	204	no layer seperation between aqueous(aq) and organic layer
$V_{taken(\mu L)}$	140	all the liquid can be taken from the sample
% of taken(µL)	68.6%	assuming MeOH is 100% miscible with MilliQ water
V_aq_taken(µL)	98.8	
mg/feces_taken	13.7	
Step2		
$V_{MTBE\ (\mu L)}$	84	
$V_{MilliQ(\mu L)}$	100	
Final		
$V_{aq(\mu L)}$	198.8	
$V_{taken(\mu L)}$	90	
Ratio calculation		
Feces_conc. in aq (mg/µL)	0.069	
feces(in 90µL of Aq)	6.21	
V_reconstitute(µL)	50	
ratio (mg:µL)	"6.12": 50	~ 1:8

Table S6: general information of metabolites applied for injection volume and dilution factor (DF) optimization

Index	HMDB_ID	Metabolite Name	Compound Formula	Monoisotopic Mass/Da	Retention time/min
1	HMDB0000214	Ornithine	C5H12N2O2	132.0899	0.57
2	HMDB0000123	Glycine	C2H5NO2	75.0320	0.62
3	HMDB0000251	Taurine	C2H7NO3S	125.0147	0.64
4	HMDB0000097	Choline	C5H13NO	103.0997	0.64
5	HMDB0000001	1-Methylhistidine	C7H11N3O2	169.0851	0.65
6	HMDB0000641	Glutamine	C5H10N2O3	146.0691	0.65
7	HMDB0000906	Trimethylamine	C3H9N	59.0735	0.65
8	HMDB0000925	TMAO	C3H9NO	75.0684	0.69
9	HMDB0000062	L-Carnitine	C7H15NO3	161.1052	0.69

10	HMDB0000043	Betaine	C5H11NO2	117.0790	0.70
11	HMDB0000562	Creatinine	C4H7N3O	113.0589	0.74
12	HMDB0000162	Proline	C5H9NO2	115.0633	0.79
13	HMDB0013222	3-Guanidinopropanoate	C4H9N3O2	131.0695	0.83
14	HMDB0000883	Valine	C5H11NO2	117.0790	1.08
15	HMDB0000190	Lactic acid	C3H6O3	90.0317	1.25
16	HMDB0000767	Pseudouridine	C9H12N2O6	244.0695	1.35
17	HMDB0000696	Methionine	C5H11NO2S	149.0510	1.35
18	HMDB0000301	Urocanic acid	C6H6N2O2	138.0429	1.50
19	HMDB0000094	Citric acid	C6H8O7	192.0270	1.57
20	HMDB0000289	Uric acid	C5H4N4O3	168.0283	1.69
21	HMDB0000157	Hypoxanthine	C5H4N4O	136.0385	1.83
22	HMDB0000172	Isoleucine	C6H13NO2	131.0946	2.25
23	HMDB0000292	Xanthine	C5H4N4O2	152.0334	2.27
24	HMDB0000158	Tyrosine	C9H11NO3	181.0739	2.41
25	HMDB0000687	Leucine	C6H13NO2	131.0946	2.42
26	HMDB0000296	Uridine	C9H12N2O6	244.0695	2.72
27	HMDB0000050	Adenosine	C10H13N5O4	267.0968	2.94
28	HMDB0000262	Thymine	C5H6N2O2	126.0429	2.96
29	HMDB0001886	3-Methylxanthine	C6H6N4O2	166.0491	3.00
30	HMDB0000684	Kynurenine	C10H12N2O3	208.0848	3.06
31	HMDB0000159	Phenylalanine	C9H11NO2	165.0790	3.06
32	HMDB0002825	Theobromine	C7H8N4O2	180.0647	3.11
33	HMDB0002013	Isobutyryl-L-Carnitine	C11H21NO4	231.1471	3.14
34	HMDB0000929	Tryptophan	C11H12N2O2	204.0899	3.32
35	HMDB0000422	2-Methylglutaricacid	C6H10O4	146.0579	3.37
36	HMDB0000715	Kynurenic acid	C10H7NO3	189.0426	3.46
37	HMDB0000378	2-Methylbutyroylcarnitine	C12H23NO4	245.1627	3.57
38	HMDB0000842	Quinaldic acid	C10H7NO2	173.0477	3.63
39	HMDB0001847	Caffeine	C8H10N4O2	194.0804	3.67
40	HMDB0000714	Hippuric acid	C9H9NO3	179.0582	3.86
41	HMDB0000671	Indolelactic acid	C11H11NO3	205.0739	4.42
42	HMDB0011621	Cinnamoylglycine	C11H11NO3	205.0739	4.53
43	HMDB0000197	Indoleacetic acid	C10H9NO2	175.0633	4.73
44	HMDB0002302	Indolepropionic acid	C11H11NO2	189.0790	5.08
45	HMDB0000138	Glycocholic acid	C26H43NO6	465.3090	5.44
46	HMDB0000951	Taurochenodeoxycholic acid	C26H45NO6S	499.2968	5.59
47	HMDB0000619	Cholic aicd	C24H40O5	408.2876	5.98

Table S7: validation parameters in plasma: linearity range and precision

Compound	polarity	Weighting factor	Linear range (uM)	cal points	r^2	Precision								
						Cal01	Cal02	Cal03	Cal04	Cal05	Cal06	Cal07	Cal08	Cal09
L-ornithine-d_6	POS	1/X^2	30-1500	6	0.984	ND	ND	ND	13.9%	2.3%	4.3%	12.1%	11.0%	2.1%
n-methyl-d_3-l-histidine	POS	1/X^2	0.3-75	5	0.985	ND	ND	ND	ND	2.8%	11.5%	7.4%	8.6%	7.6%
Betaine-d_9	POS	1/X^2	0.5-50	6	0.989	15.3%	5.3%	8.9%	8.3%	2.0%	3.6%	10.2%	12.5%	1.4%
L-glutamine-d_5	POS	1/X^2	12-15000	8	0.981	ND	7.0%	17.1%	7.1%	4.8%	3.7%	11.6%	9.3%	2.6%
TMAO-d_9	POS	1/X^2	0.4-200	7	0.977	6.3%	14.5%	3.7%	10.7%	14.1%	6.1%	1.4%	3.3%	4.2%
L-carnitine-d_3	POS	1/X^2	0.3-30	6	0.991	8.9%	1.8%	9.8%	7.6%	6.7%	2.1%	15.2%	15.7%	5.2%
L-lactic acid-$^{13}C_3$	NEG	1/X^2	15-3750	6	0.982	ND	ND	ND	8.9%	6.6%	1.3%	8.6%	10.3%	2.5%
acety-L-carnitine-d_3	POS	1/X^2	0.05-125	9	0.991	23.2%	13.9%	7.4%	10.9%	3.1%	2.3%	12.7%	11.4%	0.9%
citric acid-d_4	NEG	1/X^2	1-2500	9	0.987	10.2%	6.7%	14.5%	12.3%	2.7%	7.1%	10.5%	17.4%	1.8%
hypoxanthine-d_3	POS	1/X^2	0.1-125	8	0.990	ND	9.1%	4.1%	6.8%	2.9%	1.5%	13.0%	11.7%	4.2%
DL-leucine-d_3	POS	1/X^2	0.8-2000	9	0.991	21.0%	3.0%	8.9%	9.5%	3.1%	3.0%	14.4%	11.8%	3.1%
uridine-2-^{13}C-1,3-$^{15}N_2$	NEG	1/X^2	0.25-125	7	0.991	ND	ND	14.9%	9.0%	0.9%	1.3%	9.7%	11.0%	1.6%
phenylalanine-d_5	POS	1/X^2	0.8-2000	9	0.992	10.8%	2.4%	6.1%	6.6%	3.2%	2.7%	11.5%	12.1%	1.0%
L-tryptophan-d_3	POS	1/X^2	0.8-2000	9	0.992	7.8%	3.9%	10.5%	11.0%	3.0%	2.2%	11.6%	10.3%	4.4%
4-hydroxyphenylactic acid-d_6	NEG	1/X^2	0.1-25	6	0.986	ND	ND	ND	17.9%	8.2%	0.4%	11.3%	12.4%	2.4%
hippuric acid-d_5	NEG	1/X^2	0.05-125	9	0.987	14.8%	4.3%	8.7%	9.2%	2.4%	1.0%	12.1%	11.3%	2.3%
indole-d_5-3-acetic acid	NEG	1/X^2	0.5-25	5	0.986	ND	ND	ND	ND	4.7%	2.1%	15.5%	13.6%	1.9%
daidzein-d_6	NEG	1/X^2	0.01-12.5	8	0.991	ND	13.3%	2.9%	6.3%	4.3%	5.8%	8.2%	10.7%	2.8%
octanoyl-l-carnitine-d_3	POS	1/X^2	0.004-5	7	0.987	ND	ND	19.6%	16.3%	8.0%	13.2%	9.0%	13.5%	5.9%
glycocholic acid-d_4	NEG	1/X^2	2.5-125	5	0.990	ND	ND	0.2%	2.8%	5.3%	4.2%	8.8%	10.8%	4.9%
deoxycholic acid-d_4	NEG	1/X^2	25-125	3	0.978	ND	ND	ND	ND	0.3%	0.4%	6.4%	14.4%	2.3%

* ND: calibration points below the limit of the detection; the number highlighted in red indicates calibration points out of the linear range (residual error >20%), which are excluded in the calibration line.

Table S8: validation parameters in feces: linearity range and precision

Compound	polarity	Weighting factor	Linear range (uM)	Cal points	r^2	Precision										
						Cal01	Cal02	Cal03	Cal04	Cal05	Cal06	Cal07	Cal08	Cal09	Cal10	Cal11
choline-d_4	POS	1/X^2	0.190-19.0	7	0.989	1.7%	3.3%	2.4%	1.1%	1.0%	2.3%	1.8%	1.5%	4.8%	3.4%	2.2%
DL-proline-d_7	POS	1/X^2	2.480-496.0	8	0.996	1.4%	2.8%	0.8%	0.7%	1.5%	1.5%	0.6%	1.3%	0.8%	1.7%	1.9%
cytidine-$^{15}N_3$	POS	1/X^2	0.200-25.0	7	0.978	ND	ND	ND	ND	2.8%	13.2%	7.3%	0.7%	1.0%	2.3%	4.6%
citric acid-d_4	NEG	1/X^2	0.100-50.0	9	0.987	ND	ND	5.4%	10.7%	12.1%	2.8%	2.7%	5.0%	2.3%	3.4%	2.7%
hypoxanthine-d3	POS	1/X^2	0.460-460.0	10	0.993	2.7%	5.6%	3.2%	2.7%	0.8%	1.7%	1.4%	0.6%	1.5%	0.9%	2.3%
L-tyrosine-$_{13}C_9$-^{15}N	NEG	1/X^2	0.750-750.0	10	0.988	0.2%	1.4%	2.9%	0.5%	1.4%	2.2%	3.9%	0.8%	1.3%	0.6%	3.2%
DL-leucine-d3	POS	1/X^2	0.350-350.0	10	0.991	1.8%	2.4%	4.1%	2.3%	2.1%	1.1%	0.6%	0.5%	0.5%	0.9%	0.8%
u-^{15}N-guanosine	POS	1/X^2	0.200-10.0	6	0.958	ND	ND	ND	ND	ND	8.1%	15.4%	5.2%	3.0%	6.4%	5.6%
propionyl-L-carnitine-(n-methyl-d_3)	POS	1/X^2	0.050-25.0	9	0.984	ND	ND	2.3%	1.8%	5.1%	1.7%	2.6%	2.5%	1.3%	1.6%	0.7%
phenylalanine-d_5	POS	1/X^2	0.652-652.0	10	0.985	2.0%	3.1%	1.4%	0.7%	4.7%	0.9%	1.1%	3.2%	2.2%	1.2%	2.8%
L-tryptophan-d_3	POS	1/X^2	0.500-500.0	10	0.993	1.2%	1.6%	2.1%	2.5%	2.1%	2.1%	1.5%	1.6%	0.8%	0.8%	0.1%
quinaldic acid-d_6	POS	1/X^2	0.050-2.5	6	0.983	ND	ND	ND	ND	ND	1.9%	20.8%	4.6%	0.8%	3.8%	5.6%
4-hydroxyphenylactic acid-d_6	NEG	1/X^2	1.760-4400.0	11	0.991	2.6%	4.9%	2.5%	0.8%	0.4%	1.5%	1.9%	1.6%	1.1%	0.2%	1.5%
hippuric acid-d_5	NEG	1/X^2	0.040-4.0	7	0.983	ND	ND	ND	6.1%	9.4%	7.5%	0.7%	0.9%	4.3%	1.2%	1.5%
indole-d_5-3-acetic acid	NEG	1/X^2	0.350-75.0	8	0.985	ND	ND	ND	7.0%	16.2%	8.8%	6.9%	0.6%	0.7%	1.5%	3.4%
octanoyl-L-carnitine-d_3	POS	1/X	0.001-0.2	7	0.998	ND	ND	ND	ND	4.6%	4.8%	13.1%	6.8%	6.3%	2.0%	0.8%

* ND: calibration points below the limit of the detection; the number highlighted in red indicates calibration points out of the linear range (residual error >20%), which are excluded in the calibration line.

Table S9: predicted RME for targets measurable in positive or negative mode

index	HMDB_ID	Compound name	Compound Formula	Monoisotopic Mass/Da	detection_polarity	Retention time/min	Predicted_RME_plasma	PL_RME_criteria	Predicted_RME__feces	FE_RME_criteria
1	HMDB0001256	Spermine	C10H26N4	202.2157	POS	0.47	3.40	RME <= 15%	4.57	RME <= 15%
2	HMDB0001257	Spermidine	C7H19N3	145.1579	POS	0.49	2.97	RME <= 15%	4.45	RME <= 15%
3	HMDB0001414	Putrescine	C4H12N2	88.1000	POS	0.53	4.02	RME <= 15%	4.86	RME <= 15%
4	HMDB0002322	cadaverine	C5H14N2	102.1157	POS	0.54	4.03	RME <= 15%	5.33	RME <= 15%
5	HMDB0001432	agmatine	C5H14N4	130.1218	POS	0.57	5.12	RME <= 15%	8.57	RME <= 15%
6	HMDB0000237	propionic acid	C3H6O2	74.0368	POS	0.59	5.83	RME <= 15%	10.83	RME <= 15%

7	HMDB0001325	N6,N6,N6-Trimethyl-L-lysine	C9H20N2O2	188.1525	POS	0.59	6.21	RME <= 15%	10.93	RME <= 15%
8	HMDB0000149	Ethanolamine	C2H7NO	61.0527	POS	0.60	7.78	RME <= 15%	11.91	RME <= 15%
9	HMDB0000097	Choline	C5H13NO	103.0997	POS	0.64	11.73	RME <= 15%	18.83	15% < RME <= 30%
10	HMDB0000906	trimethylamine (TMA)	C3H9N	59.0735	POS	0.65	11.57	RME <= 15%	19.15	15% < RME <= 30%
11	HMDB0000143	Galactose	C6H12O6	180.0634	NEG	0.67	18.61	15% < RME <= 30%	21.95	15% < RME <= 30%
12	HMDB0000086	glycerophosphorylcholine (GPC)	C8H20NO6P	257.1028	POS	0.68	9.68	RME <= 15%	21.02	15% < RME <= 30%
13	HMDB0000925	TMAO	C3H9NO	75.0684	POS	0.69	9.55	RME <= 15%	20.65	15% < RME <= 30%
14	HMDB0000062	L-Carnitine	C7H15NO3	161.1052	POS	0.69	9.65	RME <= 15%	20.88	15% < RME <= 30%
15	HMDB0000294	ureaÂ	CH4N2O	60.0324	POS	0.69	9.65	RME <= 15%	20.88	15% < RME <= 30%
16	HMDB0000043	Betaine	C5H11NO2	117.0790	POS	0.70	9.75	RME <= 15%	19.34	15% < RME <= 30%
17	HMDB0003903	Isethionic acid	C2H6O4S	125.9987	NEG	0.70	14.70	RME <= 15%	24.48	15% < RME <= 30%
18	HMDB0000660	Fructose	C6H12O6	180.0634	NEG	0.71	14.12	RME <= 15%	24.73	15% < RME <= 30%
19	HMDB0001522	methylguanidine	C2H7N3	73.0640	POS	0.73	8.89	RME <= 15%	13.25	RME <= 15%
20	HMDB0001644	xylulose	C5H10O5	150.0528	NEG	0.74	14.06	RME <= 15%	22.87	15% < RME <= 30%
21	HMDB0001161	Deoxycarnitine	C7H15NO2	145.1103	POS	0.75	6.20	RME <= 15%	11.77	RME <= 15%
22	HMDB0000131	Glycerol	C3H8O3	92.0473	POS	0.75	6.20	RME <= 15%	11.77	RME <= 15%
23	HMDB0029878	tartaric acid	C4H6O6	150.0164	NEG	0.77	13.91	RME <= 15%	26.41	15% < RME <= 30%
24	HMDB0000115	Glycolic acid	C2H4O3	76.0160	NEG	0.80	37.10	RME > 30%	27.77	15% < RME <= 30%
25	HMDB0000895	Acetylcholine'	C7H16NO2	146.1181	POS	0.80	10.05	RME <= 15%	14.05	RME <= 15%
26	HMDB0000699	1-methylnicotinamide	C7H9N2O	137.0715	POS	0.80	10.05	RME <= 15%	14.05	RME <= 15%
27	HMDB0031159	Hydroxycitric acid	C6H8O8	208.0219	NEG	0.91	29.54	15% < RME <= 30%	17.67	15% < RME <= 30%
28	HMDB0000243	pyruvate	C3H4O3	88.0160	NEG	0.94	18.83	15% < RME <= 30%	14.02	RME <= 15%
29	HMDB0002064	N-acetylputrescine	C6H14N2O	130.1106	POS	0.95	3.86	RME <= 15%	11.80	RME <= 15%
30	HMDB0000156	L-Malic acid	C4H6O5	134.0215	NEG	0.98	10.32	RME <= 15%	11.36	RME <= 15%
31	HMDB0000208	ketoglutaric acid	C5H6O5	146.0215	NEG	1.17	7.53	RME <= 15%	8.61	RME <= 15%

32	HMDB0000076	5,6-dihydrouracil	C4H6N2O2	114.0429	POS	1.17	3.59	RME <= 15%	8.88	RME <= 15%
33	HMDB0000201	L-Acetylcarnitine	C9H17NO4	203.1158	POS	1.42	2.97	RME <= 15%	6.12	RME <= 15%
34	HMDB0001406	Nicotinamideï¼ˆform of vitamin Bï¼‰	C6H6N2O	122.0480	POS	1.62	30.80	RME > 30%	5.18	RME <= 15%
35	HMDB0000005	2-Ketobutyric acid'	C4H6O3	102.0317	NEG	1.73	19.66	15% < RME <= 30%	14.83	RME <= 15%
36	HMDB0000982	5-Methylcytidine	C10H15N3O5	257.1012	POS	2.05	3.76	RME <= 15%	4.76	RME <= 15%
37	HMDB0010325	Ethyl glucuronide	C8H14O7	222.0740	NEG	2.14	4.60	RME <= 15%	6.31	RME <= 15%
38	HMDB0000732	3-Hydroxykynurenine	C10H12N2O4	224.0797	POS	2.20	3.36	RME <= 15%	9.34	RME <= 15%
39	HMDB0240577	3-Methylcytidine (*methosulfate)	C10H15N3O5	257.1012	POS	2.42	3.77	RME <= 15%	9.52	RME <= 15%
40	HMDB0000306	tyramineÂ	C8H11NO	137.0841	POS	2.63	3.43	RME <= 15%	5.83	RME <= 15%
41	HMDB0000008	2-hydroxybutyric acidÂ	C4H8O3	104.0473	NEG	2.75	2.69	RME <= 15%	5.87	RME <= 15%
42	HMDB0001046	Cotinine	C10H12N2O	176.0950	POS	2.94	3.78	RME <= 15%	4.48	RME <= 15%
43	HMDB0242132	2-O-Methylcytidine	C10H15N3O5	257.1012	POS	2.95	3.57	RME <= 15%	4.61	RME <= 15%
44	HMDB0010738	1-methylxanthine	C6H6N4O2	166.0491	POS	2.95	3.49	RME <= 15%	5.02	RME <= 15%
45	HMDB0000824	Propionylcarnitine	C10H19NO4	217.1314	POS	2.96	3.42	RME <= 15%	8.68	RME <= 15%
46	HMDB0003331	1-Methyladenosine	C11H15N5O4	281.1124	POS	2.98	3.08	RME <= 15%	13.27	RME <= 15%
47	HMDB0001107	7-Methylguanosine	C11H15N5O5	297.1073	POS	2.98	3.08	RME <= 15%	13.27	RME <= 15%
48	HMDB0004044	N6-methyladenosine	C11H15N5O4	281.1124	POS	2.98	3.08	RME <= 15%	13.27	RME <= 15%
49	HMDB0000022	3-Methoxytyramine	C9H13NO2	167.0946	POS	2.99	3.21	RME <= 15%	14.69	RME <= 15%
50	HMDB0001886	3-methylxanthine	C6H6N4O2	166.0491	POS	3.00	3.21	RME <= 15%	15.85	15% < RME <= 30%
51	HMDB0000707	4-Hydroxyphenylpyruvic acid	C9H8O4	180.0423	NEG	3.01	3.67	RME <= 15%	18.29	15% < RME <= 30%
52	HMDB0002825	theobromineÂ	C7H8N4O2	180.0647	POS	3.11	5.49	RME <= 15%	11.20	RME <= 15%
53	HMDB0002013	Isobutyryl-L-carnitine	C11H21NO4	231.1471	POS	3.14	5.13	RME <= 15%	10.41	RME <= 15%
54	HMDB0001890	N-acetyl cysteine	C5H9NO3S	163.0303	NEG	3.14	3.27	RME <= 15%	15.44	15% < RME <= 30%
55	HMDB0000736	Butyrylcarnitine	C11H21NO4	231.1471	POS	3.19	3.42	RME <= 15%	5.28	RME <= 15%

56	HMDB0012275	Phenylethylamine (PEA)	C8H11N	121.0891	POS	3.20	3.40	RME <= 15%	5.23	RME <= 15%
57	HMDB0001392	4-Aminobenzoic acid (PABA)	C7H7NO2	137.0477	POS	3.31	3.69	RME <= 15%	5.99	RME <= 15%
58	HMDB0001336	3,4-dihydroxyphenylacetic acid (DOPAC)	C8H8O4	168.0423	NEG	3.35	3.25	RME <= 15%	11.55	RME <= 15%
59	HMDB0000755	Hydroxyphenyllactic acid	C9H10O4	182.0579	NEG	3.38	3.22	RME <= 15%	12.87	RME <= 15%
60	HMDB0002366	Tiglylcarnitine	C12H21NO4	243.1471	POS	3.45	2.85	RME <= 15%	7.08	RME <= 15%
61	HMDB0000423	3,4-Dihydroxyhydrocinnamic acid (Dihydro-caffeic acid)	C9H10O4	182.0579	NEG	3.45	3.65	RME <= 15%	15.92	15% < RME <= 30%
62	HMDB0041993	Pivaloylcarnitine	C12H23NO4	245.1627	POS	3.51	3.65	RME <= 15%	6.76	RME <= 15%
63	HMDB0000303	tryptamine	C10H12N2	160.1000	POS	3.53	3.52	RME <= 15%	6.25	RME <= 15%
64	HMDB0000378	2-Methylbutyroylcarnitine	C12H23NO4	245.1627	POS	3.57	3.47	RME <= 15%	6.35	RME <= 15%
65	HMDB0000688	Isovalerylcarnitine	C12H23NO4	245.1627	POS	3.64	3.62	RME <= 15%	7.26	RME <= 15%
66	HMDB0001847	caffeine	C8H10N4O2	194.0804	POS	3.67	3.66	RME <= 15%	6.40	RME <= 15%
67	HMDB0013128	Valerylcarnitine	C12H23NO4	245.1627	POS	3.71	3.27	RME <= 15%	5.53	RME <= 15%
68	HMDB0001856	3,4-dihydroxybenzoic acid	C7H6O4	154.0266	NEG	3.73	11.05	RME <= 15%	11.18	RME <= 15%
69	HMDB0000491	3-methyl-2-oxovalerate	C6H10O3	130.0630	NEG	3.75	8.49	RME <= 15%	13.10	RME <= 15%
70	HMDB0000703	Mandelic acid	C8H8O3	152.0473	NEG	3.75	8.49	RME <= 15%	13.10	RME <= 15%
71	HMDB0000705	Hexanoylcarnitine	C13H25NO4	259.1784	POS	4.17	2.63	RME <= 15%	5.71	RME <= 15%
72	HMDB0000375	3-(3-hydroxyphenyl)propionic acid (hMPP)	C9H10O3	166.0630	NEG	4.27	3.17	RME <= 15%	17.85	15% < RME <= 30%
73	HMDB0000779	3-Phenyllactic acid	C9H10O3	166.0630	NEG	4.29	3.51	RME <= 15%	13.22	RME <= 15%
74	HMDB0013324	2-Octenoylcarnitine	C15H27NO4	285.1940	POS	4.73	5.62	RME <= 15%	7.81	RME <= 15%
75	HMDB0000791	L-Octanoylcarnitine	C15H29NO4	287.2097	POS	4.86	4.03	RME <= 15%	5.55	RME <= 15%
76	HMDB0001858	p-cresol	C7H8O	108.0575	POS	5.05	8.80	RME <= 15%	9.70	RME <= 15%
77	HMDB0000651	Decanoylcarnitine	C17H33NO4	315.2410	POS	5.39	7.87	RME <= 15%	11.65	RME <= 15%
78	HMDB0000874	Tauroursodeoxycholic acid	C26H45NO6S	499.2968	NEG	5.59	2.76	RME <= 15%	11.32	RME <= 15%
79	HMDB0002250	Lauroylcarnitine	C19H37NO4	343.2723	POS	5.89	3.83	RME <= 15%	15.84	15% < RME <= 30%
80	HMDB0002639	Sulfolithocholylglycine	C26H43NO7S	513.2760	NEG	5.92	3.33	RME <= 15%	5.65	RME <= 15%

Table S10: predicted RME for tagets measurable in positive and negative mode

Index	Compound name	HMDB_ID	Compound Formula	Monoisotopic Mass/Da	detection_polarity	Retention time/min	Predicted_RME_pos_plasma (PL)	PL_pos_RME_criteria	Predicted_RME_neg_plasma (PL)	PL_neg_RME_criteria	Predicted_RME_pos_feces (FE)	FE_pos_RME_criteria	Predicted_RME_neg_feces (FE)	FE_neg_RME_criteria
1	5-Hydroxylysine	HMDB0000450	C6H14N2O3	162.1004	POS_NEG	0.56	4.60	RME <= 15%	9.39	RME <= 15%	6.26	RME <= 15%	7.36	RME <= 15%
2	Lysine	HMDB0000182	C6H14N2O2	146.1055	POS_NEG	0.56	4.73	RME <= 15%	9.23	RME <= 15%	6.61	RME <= 15%	7.74	RME <= 15%
3	Ornithine	HMDB0000214	C5H12N2O2	132.0899	POS_NEG	0.56	5.10	RME <= 15%	9.22	RME <= 15%	7.03	RME <= 15%	8.41	RME <= 15%
4	Histamine	HMDB0000870	C5H9N3	111.0796	POS_NEG	0.58	5.09	RME <= 15%	10.03	RME <= 15%	9.34	RME <= 15%	10.13	RME <= 15%
5	N(omega)-Hydroxyarginine	HMDB0004224	C6H14N4O3	190.1066	POS_NEG	0.59	5.83	RME <= 15%	10.63	RME <= 15%	10.83	RME <= 15%	15.73	15% < RME <= 30%
6	Histidine	HMDB0000177	C6H9N3O2	155.0695	POS_NEG	0.60	6.52	RME <= 15%	11.03	RME <= 15%	11.16	RME <= 15%	18.16	15% < RME <= 30%
7	Diaminopimelic acid	HMDB0001370	C7H14N2O4	190.0954	POS_NEG	0.61	8.89	RME <= 15%	11.71	RME <= 15%	12.38	RME <= 15%	19.91	15% < RME <= 30%
8	Carnosine	HMDB0000033	C9H14N4O3	226.1066	POS_NEG	0.61	10.49	RME <= 15%	11.91	RME <= 15%	12.56	RME <= 15%	19.82	15% < RME <= 30%
9	N(6)-Carboxymethyllysine (CML)	HMDB0240347	C8H16N2O4	204.1110	POS_NEG	0.61	10.49	RME <= 15%	11.91	RME <= 15%	12.56	RME <= 15%	19.82	15% < RME <= 30%
10	Arginine	HMDB0000517	C6H14N4O2	174.1117	POS_NEG	0.61	10.49	RME <= 15%	11.91	RME <= 15%	12.56	RME <= 15%	19.82	15% < RME <= 30%
11	Glycine	HMDB0000123	C2H5NO2	75.0320	POS_NEG	0.62	11.14	RME <= 15%	14.29	RME <= 15%	14.95	RME <= 15%	21.04	15% < RME <= 30%
12	Beta-Alanine	HMDB0000056	C3H7NO2	89.0477	POS_NEG	0.62	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
13	Glycylglycine	HMDB0011733	C4H8N2O3	132.0535	POS_NEG	0.62	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
14	Cystathionine	HMDB0000099	C7H14N2O4S	222.0674	POS_NEG	0.62	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
15	Anserine	HMDB0000194	C10H16N4O3	240.1222	POS_NEG	0.62	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%

16	1-Methylhisti-dine	HMDB0000001	C7H11N3O2	169.0851	POS_NEG	0.63	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
17	Serine	HMDB0000187	C3H7NO3	105.0426	POS_NEG	0.63	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
18	O-Phosphoeth-anolamine	HMDB0000224	C2H8NO4P	141.0191	POS_NEG	0.63	11.27	RME <= 15%	16.09	15% < RME <= 30%	15.84	15% < RME <= 30%	21.59	15% < RME <= 30%
19	Asparagine	HMDB0000168	C4H8N2O3	132.0535	POS_NEG	0.63	11.50	RME <= 15%	17.16	15% < RME <= 30%	16.67	15% < RME <= 30%	21.55	15% < RME <= 30%
20	Homocarnosine	HMDB0000745	C10H16N4O3	240.1222	POS_NEG	0.63	11.64	RME <= 15%	18.06	15% < RME <= 30%	17.52	15% < RME <= 30%	21.38	15% < RME <= 30%
21	Taurine	HMDB0000251	C2H7NO3S	125.0147	POS_NEG	0.64	11.61	RME <= 15%	19.40	15% < RME <= 30%	18.31	15% < RME <= 30%	21.45	15% < RME <= 30%
22	3-methyl histi-dine	HMDB0000479	C7H11N3O2	169.0851	POS_NEG	0.64	11.66	RME <= 15%	18.89	15% < RME <= 30%	18.64	15% < RME <= 30%	21.75	15% < RME <= 30%
23	myo-inositol	HMDB0000211	C6H12O6	180.0634	POS_NEG	0.64	11.66	RME <= 15%	18.89	15% < RME <= 30%	18.64	15% < RME <= 30%	21.75	15% < RME <= 30%
24	Alanine	HMDB0000161	C3H7NO2	89.0477	POS_NEG	0.64	11.73	RME <= 15%	18.79	15% < RME <= 30%	18.83	15% < RME <= 30%	21.90	15% < RME <= 30%
25	Aspartic acid	HMDB0000191	C4H7NO4	133.0375	POS_NEG	0.65	11.44	RME <= 15%	19.41	15% < RME <= 30%	18.93	15% < RME <= 30%	21.42	15% < RME <= 30%
26	Sarcosine	HMDB0000271	C3H7NO2	89.0477	POS_NEG	0.65	11.44	RME <= 15%	19.41	15% < RME <= 30%	18.93	15% < RME <= 30%	21.42	15% < RME <= 30%
27	Phosphoserine	HMDB0000272	C3H8NO6P	185.0089	POS_NEG	0.65	11.23	RME <= 15%	19.96	15% < RME <= 30%	18.96	15% < RME <= 30%	20.94	15% < RME <= 30%
28	Glutamine	HMDB0000641	C5H10N2O3	146.0691	POS_NEG	0.66	11.82	RME <= 15%	19.38	15% < RME <= 30%	19.71	15% < RME <= 30%	21.50	15% < RME <= 30%
29	Homo-L-argi-nine	HMDB0000670	C7H16N4O2	188.1273	POS_NEG	0.66	11.82	RME <= 15%	19.38	15% < RME <= 30%	19.71	15% < RME <= 30%	21.50	15% < RME <= 30%
30	Homoserine	HMDB0000719	C4H9NO3	119.0582	POS_NEG	0.66	11.72	RME <= 15%	18.41	15% < RME <= 30%	19.71	15% < RME <= 30%	21.53	15% < RME <= 30%
31	Threonine	HMDB0000167	C4H9NO3	119.0582	POS_NEG	0.67	11.60	RME <= 15%	18.62	15% < RME <= 30%	20.34	15% < RME <= 30%	21.85	15% < RME <= 30%
32	ribose-5-P	HMDB0001548	C5H11O8P	230.0192	POS_NEG	0.67	11.60	RME <= 15%	18.62	15% < RME <= 30%	20.34	15% < RME <= 30%	21.85	15% < RME <= 30%

33	4-Hydroxypro-line	HMDB0000725	C5H9NO3	131.0582	POS_NEG	0.67	11.33	RME <= 15%	18.61	15% < RME <= 30%	20.48	15% < RME <= 30%	21.95	15% < RME <= 30%
34	Saccharopine	HMDB0000279	C11H20N2O6	276.1321	POS_NEG	0.67	11.33	RME <= 15%	18.61	15% < RME <= 30%	20.48	15% < RME <= 30%	21.95	15% < RME <= 30%
35	Gamma-Amino-butyric acid	HMDB0000112	C4H9NO2	103.0633	POS_NEG	0.68	10.83	RME <= 15%	18.42	15% < RME <= 30%	20.80	15% < RME <= 30%	22.97	15% < RME <= 30%
36	Galactitol	HMDB0000107	C6H14O6	182.0790	POS_NEG	0.68	9.68	RME <= 15%	18.29	15% < RME <= 30%	21.02	15% < RME <= 30%	23.27	15% < RME <= 30%
37	O-Acetylserine	HMDB0003011	C5H9NO4	147.0532	POS_NEG	0.68	9.68	RME <= 15%	18.29	15% < RME <= 30%	21.02	15% < RME <= 30%	23.27	15% < RME <= 30%
38	Dimethylglycine	HMDB0000092	C4H9NO2	103.0633	POS_NEG	0.68	9.68	RME <= 15%	18.29	15% < RME <= 30%	21.02	15% < RME <= 30%	23.27	15% < RME <= 30%
39	Glycerol 3-phos-phate	HMDB0000126	C3H9O6P	172.0137	POS_NEG	0.68	9.68	RME <= 15%	18.29	15% < RME <= 30%	21.02	15% < RME <= 30%	23.27	15% < RME <= 30%
40	Glutamic acid	HMDB0000148	C5H9NO4	147.0532	POS_NEG	0.68	9.68	RME <= 15%	18.29	15% < RME <= 30%	21.02	15% < RME <= 30%	23.27	15% < RME <= 30%
41	N-monomethylarg-inine	HMDB0029416	C7H16N4O2	188.1273	POS_NEG	0.68	9.49	RME <= 15%	17.74	15% < RME <= 30%	21.11	15% < RME <= 30%	23.74	15% < RME <= 30%
42	5-Ami-nolevulinic acid	HMDB0001149	C5H9NO3	131.0582	POS_NEG	0.68	9.49	RME <= 15%	17.74	15% < RME <= 30%	21.11	15% < RME <= 30%	23.74	15% < RME <= 30%
43	Mannitol	HMDB0000765	C6H14O6	182.0790	POS_NEG	0.69	9.55	RME <= 15%	16.79	15% < RME <= 30%	20.65	15% < RME <= 30%	24.23	15% < RME <= 30%
44	Sorbitol	HMDB0000247	C6H14O6	182.0790	POS_NEG	0.69	9.55	RME <= 15%	16.79	15% < RME <= 30%	20.65	15% < RME <= 30%	24.23	15% < RME <= 30%
45	Guanidineacetic acid	HMDB0000128	C3H7N3O2	117.0538	POS_NEG	0.69	9.65	RME <= 15%	16.92	15% < RME <= 30%	20.88	15% < RME <= 30%	24.17	15% < RME <= 30%
46	xylitol	HMDB0002917	C5H12O5	152.0685	POS_NEG	0.70	9.72	RME <= 15%	15.34	15% < RME <= 30%	20.15	15% < RME <= 30%	24.13	15% < RME <= 30%
47	Citrulline	HMDB0000904	C6H13N3O3	175.0957	POS_NEG	0.70	9.75	RME <= 15%	14.70	RME <= 15%	19.34	15% < RME <= 30%	24.48	15% < RME <= 30%
48	D-Gluconate	HMDB0000625	C6H12O7	196.0583	POS_NEG	0.70	9.75	RME <= 15%	14.70	RME <= 15%	19.34	15% < RME <= 30%	24.48	15% < RME <= 30%

49	Cysteine	HMDB0000574	C3H7NO2S	121.0198	POS_NEG	0.70	9.75	RME <= 15%	14.70	RME <= 15%	19.34	15% < RME <= 30%	24.48	15% < RME <= 30%
50	Methionine sulfoxide	HMDB0002005	C5H11NO3S	165.0460	POS_NEG	0.71	9.78	RME <= 15%	13.98	RME <= 15%	18.38	15% < RME <= 30%	24.80	15% < RME <= 30%
51	Argininosuccinic acid	HMDB0000052	C10H18N4O6	290.1226	POS_NEG	0.71	9.96	RME <= 15%	14.12	RME <= 15%	17.73	15% < RME <= 30%	24.73	15% < RME <= 30%
52	Guanidinosuccinic acid (GSA)	HMDB0003157	C5H9N3O4	175.0593	POS_NEG	0.73	8.89	RME <= 15%	14.15	RME <= 15%	13.25	RME <= 15%	22.89	15% < RME <= 30%
53	Allantoin	HMDB0000462	C4H6N4O3	158.0440	POS_NEG	0.73	8.89	RME <= 15%	14.15	RME <= 15%	13.25	RME <= 15%	22.89	15% < RME <= 30%
54	taurocyamine	HMDB0003584	C3H9N3O3S	167.0365	POS_NEG	0.73	8.89	RME <= 15%	14.15	RME <= 15%	13.25	RME <= 15%	22.89	15% < RME <= 30%
55	Norepinephrine	HMDB0000216	C8H11NO3	169.0739	POS_NEG	0.73	8.36	RME <= 15%	14.38	RME <= 15%	12.83	RME <= 15%	23.01	15% < RME <= 30%
56	creatinine	HMDB0000562	C4H7N3O	113.0589	POS_NEG	0.74	7.93	RME <= 15%	14.06	RME <= 15%	12.61	RME <= 15%	22.87	15% < RME <= 30%
57	Asymmetric dimethylarginine	HMDB0001539	C8H18N4O2	202.1430	POS_NEG	0.75	6.62	RME <= 15%	13.57	RME <= 15%	12.05	RME <= 15%	23.16	15% < RME <= 30%
58	Epinephrine	HMDB0000068	C9H13NO3	183.0895	POS_NEG	0.75	6.20	RME <= 15%	13.76	RME <= 15%	11.77	RME <= 15%	23.51	15% < RME <= 30%
59	3-Aminoisobutanoic acid	HMDB0003911	C4H9NO2	103.0633	POS_NEG	0.75	6.20	RME <= 15%	13.76	RME <= 15%	11.77	RME <= 15%	23.51	15% < RME <= 30%
60	Aminoadipic acid	HMDB0000510	C6H11NO4	161.0688	POS_NEG	0.75	6.20	RME <= 15%	13.76	RME <= 15%	11.77	RME <= 15%	23.51	15% < RME <= 30%
61	Homocysteine	HMDB0000742	C4H9NO2S	135.0354	POS_NEG	0.75	6.20	RME <= 15%	13.76	RME <= 15%	11.77	RME <= 15%	23.51	15% < RME <= 30%
62	Methylcysteine	HMDB0002108	C4H9NO2S	135.0354	POS_NEG	0.75	6.20	RME <= 15%	13.76	RME <= 15%	11.77	RME <= 15%	23.51	15% < RME <= 30%
63	Alpha-aminobutyric acid	HMDB0000452	C4H9NO2	103.0633	POS_NEG	0.75	5.89	RME <= 15%	13.36	RME <= 15%	11.36	RME <= 15%	23.36	15% < RME <= 30%
64	N-Acetylneuraminic acid	HMDB0000230	C11H19NO9	309.1060	POS_NEG	0.76	5.09	RME <= 15%	13.16	RME <= 15%	10.50	RME <= 15%	24.51	15% < RME <= 30%
65	Thiamine (vitamin B1)	HMDB0000235	C12H17N4OS	265.1123	POS_NEG	0.77	4.97	RME <= 15%	12.93	RME <= 15%	10.16	RME <= 15%	25.56	15% < RME <= 30%

66	L_Pipecolic acid	HMDB0000716	C6H11NO2	129.0790	POS_NEG	0.77	5.07	RME <= 15%	13.91	RME <= 15%	10.24	RME <= 15%	26.41	15% < RME <= 30%
67	creatine	HMDB0000064	C4H9N3O2	131.0695	POS_NEG	0.78	5.69	RME <= 15%	16.19	15% < RME <= 30%	10.22	RME <= 15%	26.81	15% < RME <= 30%
68	Proline	HMDB0000162	C5H9NO2	115.0633	POS_NEG	0.78	6.01	RME <= 15%	22.00	15% < RME <= 30%	10.97	RME <= 15%	27.21	15% < RME <= 30%
69	Malonylcarnitine	HMDB0002095	C10H17NO6	247.1056	POS_NEG	0.80	8.05	RME <= 15%	29.87	15% < RME <= 30%	12.88	RME <= 15%	27.55	15% < RME <= 30%
70	Cytosine	HMDB0000630	C4H5N3O	111.0433	POS_NEG	0.81	11.12	RME <= 15%	38.97	RME > 30%	14.60	RME <= 15%	27.37	15% < RME <= 30%
71	N2-gamma-Glutamylglutamine	HMDB0011738	C10H17N3O6	275.1117	POS_NEG	0.82	15.28	15% < RME <= 30%	53.67	RME > 30%	16.51	15% < RME <= 30%	27.27	15% < RME <= 30%
72	Dopamine	HMDB0000073	C8H11NO2	153.0790	POS_NEG	0.82	15.28	15% < RME <= 30%	53.67	RME > 30%	16.51	15% < RME <= 30%	27.27	15% < RME <= 30%
73	3-Guanidinopropanoate	HMDB0013222	C4H9N3O2	131.0695	POS_NEG	0.83	15.93	15% < RME <= 30%	56.68	RME > 30%	16.58	15% < RME <= 30%	26.73	15% < RME <= 30%
74	Homocitrulline	HMDB0000679	C7H15N3O3	189.1113	POS_NEG	0.84	18.58	15% < RME <= 30%	63.45	RME > 30%	16.52	15% < RME <= 30%	25.73	15% < RME <= 30%
75	Symmetric dimethylarginine	HMDB0003334	C8H18N4O2	202.1430	POS_NEG	0.90	9.46	RME <= 15%	33.47	RME > 30%	12.39	RME <= 15%	19.27	15% < RME <= 30%
76	5-Aminovaleric acid	HMDB0003355	C5H11NO2	117.0790	POS_NEG	0.91	7.13	RME <= 15%	29.54	15% < RME <= 30%	11.84	RME <= 15%	17.67	15% < RME <= 30%
77	Dihydroorotic acid	HMDB0000528	C5H6N2O4	158.0328	POS_NEG	1.00	3.38	RME <= 15%	6.70	RME <= 15%	11.97	RME <= 15%	11.13	RME <= 15%
78	isocitric acid	HMDB0000193	C6H8O7	192.0270	POS_NEG	1.01	2.90	RME <= 15%	5.49	RME <= 15%	11.92	RME <= 15%	11.25	RME <= 15%
79	N-Acetylserine	HMDB0002931	C5H9NO4	147.0532	POS_NEG	1.05	3.70	RME <= 15%	4.74	RME <= 15%	10.47	RME <= 15%	11.70	RME <= 15%
80	Orotic Acid	HMDB0000226	C5H4N2O4	156.0171	POS_NEG	1.08	3.42	RME <= 15%	5.37	RME <= 15%	10.97	RME <= 15%	12.56	RME <= 15%
81	Valine	HMDB0000883	C5H11NO2	117.0790	POS_NEG	1.08	3.50	RME <= 15%	5.58	RME <= 15%	11.00	RME <= 15%	12.68	RME <= 15%
82	Glycylproline	HMDB0000721	C7H12N2O3	172.0848	POS_NEG	1.18	3.66	RME <= 15%	17.41	15% < RME <= 30%	8.37	RME <= 15%	13.11	RME <= 15%
83	gamma-Glutamylalanine	HMDB0006248	C8H14N2O5	218.0903	POS_NEG	1.19	3.73	RME <= 15%	20.46	15% < RME <= 30%	8.17	RME <= 15%	14.92	RME <= 15%

84	Pipecolic acid	HMDB0000070	C6H11NO2	129.0790	POS_NEG	1.19	4.15	RME <= 15%	27.12	15% < RME <= 30%	7.96	RME <= 15%	19.42	15% < RME <= 30%
85	Picolinic acid	HMDB0002243	C6H5NO2	123.0320	POS_NEG	1.24	5.15	RME <= 15%	62.59	RME > 30%	6.60	RME <= 15%	47.13	RME > 30%
86	lactic acid	HMDB0000190	C3H6O3	90.0317	POS_NEG	1.25	5.24	RME <= 15%	60.52	RME > 30%	6.19	RME <= 15%	46.11	RME > 30%
87	4-guanidinobu-tyric acid	HMDB0003464	C5H11N3O2	145.0851	POS_NEG	1.25	5.24	RME <= 15%	60.52	RME > 30%	6.19	RME <= 15%	46.11	RME > 30%
88	Adenine	HMDB0000034	C5H5N5	135.0545	POS_NEG	1.29	5.53	RME <= 15%	41.83	RME > 30%	6.02	RME <= 15%	29.55	15% < RME <= 30%
89	guanine	HMDB0000132	C5H5N5O	151.0494	POS_NEG	1.31	5.07	RME <= 15%	29.50	15% < RME <= 30%	7.51	RME <= 15%	22.28	15% < RME <= 30%
90	Uracil	HMDB0000300	C4H4N2O2	112.0273	POS_NEG	1.32	4.64	RME <= 15%	23.12	15% < RME <= 30%	8.47	RME <= 15%	20.47	15% < RME <= 30%
91	cytidine	HMDB0000089	C9H13N3O5	243.0855	POS_NEG	1.33	4.29	RME <= 15%	21.63	15% < RME <= 30%	8.67	RME <= 15%	20.37	15% < RME <= 30%
92	N-Acetylgluta-mine	HMDB0006029	C7H12N2O4	188.0797	POS_NEG	1.33	3.97	RME <= 15%	18.93	15% < RME <= 30%	9.20	RME <= 15%	19.70	15% < RME <= 30%
93	Nicotinic acid	HMDB0001488	C6H5NO2	123.0320	POS_NEG	1.34	3.35	RME <= 15%	15.91	15% < RME <= 30%	9.66	RME <= 15%	19.89	15% < RME <= 30%
94	Pseudouridine	HMDB0000767	C9H12N2O6	244.0695	POS_NEG	1.35	3.21	RME <= 15%	13.23	RME <= 15%	9.25	RME <= 15%	19.64	15% < RME <= 30%
95	Methionine	HMDB0000696	C5H11NO2S	149.0510	POS_NEG	1.35	3.21	RME <= 15%	13.23	RME <= 15%	9.25	RME <= 15%	19.64	15% < RME <= 30%
96	alpha-N-Acety-larginine	HMDB0004620	C8H16N4O3	216.1222	POS_NEG	1.38	2.97	RME <= 15%	9.03	RME <= 15%	8.39	RME <= 15%	16.75	15% < RME <= 30%
97	neopterin	HMDB0000845	C9H11N5O4	253.0811	POS_NEG	1.42	3.02	RME <= 15%	5.63	RME <= 15%	6.28	RME <= 15%	10.98	RME <= 15%
98	urocanic acid	HMDB0000301	C6H6N2O2	138.0429	POS_NEG	1.50	3.68	RME <= 15%	4.76	RME <= 15%	4.62	RME <= 15%	7.59	RME <= 15%
99	Quinolinic acid	HMDB0000232	C7H5NO4	167.0219	POS_NEG	1.52	3.12	RME <= 15%	4.60	RME <= 15%	4.69	RME <= 15%	8.13	RME <= 15%
100	Pyridoxal	HMDB0001545	C8H9NO3	167.0582	POS_NEG	1.52	3.12	RME <= 15%	4.60	RME <= 15%	4.69	RME <= 15%	8.13	RME <= 15%
101	Glutathione	HMDB0000125	C10H17N3O6S	307.0838	POS_NEG	1.52	3.12	RME <= 15%	4.60	RME <= 15%	4.69	RME <= 15%	8.13	RME <= 15%
102	Citric acid	HMDB0000094	C6H8O7	192.0270	POS_NEG	1.60	27.28	15% < RME <= 30%	66.09	RME > 30%	6.93	RME <= 15%	26.57	15% < RME <= 30%

103	uric acid	HMDB0000289	C5H4N4O3	168.0283	POS_NEG	1.70	8.37	RME <= 15%	33.43	RME > 30%	4.47	RME <= 15%	7.90	RME <= 15%
104	hypoxanthine	HMDB0000157	C5H4N4O	136.0385	POS_NEG	1.83	3.77	RME <= 15%	13.53	RME <= 15%	14.60	RME <= 15%	23.11	15% < RME <= 30%
105	5-oxoproline(pyroglutamic acid)	HMDB0000267	C5H7NO3	129.0426	POS_NEG	1.85	3.52	RME <= 15%	12.97	RME <= 15%	12.46	RME <= 15%	16.82	15% < RME <= 30%
106	cis-Aconitic acid	HMDB0000072	C6H6O6	174.0164	POS_NEG	1.99	3.19	RME <= 15%	7.81	RME <= 15%	6.15	RME <= 15%	13.61	RME <= 15%
107	pyridoxine (vitamin B6)	HMDB0000239	C8H11NO3	169.0739	POS_NEG	2.11	3.29	RME <= 15%	4.66	RME <= 15%	5.66	RME <= 15%	6.85	RME <= 15%
108	Isoleucine	HMDB0000172	C6H13NO2	131.0946	POS_NEG	2.26	3.04	RME <= 15%	4.38	RME <= 15%	6.91	RME <= 15%	20.57	15% < RME <= 30%
109	xanthine	HMDB0000292	C5H4N4O2	152.0334	POS_NEG	2.27	3.46	RME <= 15%	4.62	RME <= 15%	6.80	RME <= 15%	20.25	15% < RME <= 30%
110	3-hydroxybutyric acid	HMDB0000357	C4H8O3	104.0473	POS_NEG	2.38	3.95	RME <= 15%	3.97	RME <= 15%	13.82	RME <= 15%	17.64	15% < RME <= 30%
111	Tyrosine	HMDB0000158	C9H11NO3	181.0739	POS_NEG	2.41	3.88	RME <= 15%	3.57	RME <= 15%	9.85	RME <= 15%	13.43	RME <= 15%
112	Leucine	HMDB0000687	C6H13NO2	131.0946	POS_NEG	2.43	3.74	RME <= 15%	3.29	RME <= 15%	7.99	RME <= 15%	9.42	RME <= 15%
113	trans-Aconitic acid	HMDB0000958	C6H6O6	174.0164	POS_NEG	2.45	3.56	RME <= 15%	3.39	RME <= 15%	6.66	RME <= 15%	7.43	RME <= 15%
114	uridine	HMDB0000296	C9H12N2O6	244.0695	POS_NEG	2.72	3.71	RME <= 15%	2.69	RME <= 15%	6.47	RME <= 15%	5.77	RME <= 15%
115	2,5-Furandicarboxylic acid	HMDB0004812	C6H4O5	156.0059	POS_NEG	2.80	3.24	RME <= 15%	3.50	RME <= 15%	4.37	RME <= 15%	5.41	RME <= 15%
116	p-hydroxymandelate	HMDB0000822	C8H8O4	168.0423	POS_NEG	2.88	3.27	RME <= 15%	3.03	RME <= 15%	4.13	RME <= 15%	5.00	RME <= 15%
117	hydroquinone	HMDB0002434	C6H6O2	110.0368	POS_NEG	2.90	3.66	RME <= 15%	2.99	RME <= 15%	3.85	RME <= 15%	5.23	RME <= 15%
118	2-Deoxyuridine	HMDB0000012	C9H12N2O5	228.0746	POS_NEG	2.94	3.78	RME <= 15%	2.98	RME <= 15%	4.48	RME <= 15%	5.37	RME <= 15%
119	adenosine	HMDB0000050	C10H13N5O4	267.0968	POS_NEG	2.95	3.57	RME <= 15%	2.95	RME <= 15%	4.61	RME <= 15%	5.40	RME <= 15%
120	2-Methylguanosine	HMDB0005862	C11H15N5O5	297.1073	POS_NEG	2.95	3.57	RME <= 15%	2.95	RME <= 15%	4.61	RME <= 15%	5.40	RME <= 15%
121	cAMP (Adenosine 3,5-cyclic monophosphate)	HMDB0000058	C10H12N5O6P	329.0525	POS_NEG	2.95	3.57	RME <= 15%	2.95	RME <= 15%	4.61	RME <= 15%	5.40	RME <= 15%
122	S-Adenosylhomocysteine (SAH)	HMDB0000939	C14H20N6O5S	384.1216	POS_NEG	2.95	3.57	RME <= 15%	2.95	RME <= 15%	4.61	RME <= 15%	5.40	RME <= 15%
123	2-deoxyguanosine	HMDB0000085	C10H13N5O4	267.0968	POS_NEG	2.95	3.57	RME <= 15%	2.95	RME <= 15%	4.61	RME <= 15%	5.40	RME <= 15%

124	ketoadipic acid	HMDB0000225	C6H8O5	160.0372	POS_NEG	2.95	3.49	RME <= 15%	2.91	RME <= 15%	5.02	RME <= 15%	5.41	RME <= 15%
125	Guanosine	HMDB0000133	C10H13N5O5	283.0917	POS_NEG	2.95	3.49	RME <= 15%	2.91	RME <= 15%	5.02	RME <= 15%	5.41	RME <= 15%
126	inosine	HMDB0000195	C10H12N4O5	268.0808	POS_NEG	2.95	3.49	RME <= 15%	2.91	RME <= 15%	5.02	RME <= 15%	5.41	RME <= 15%
127	2-deoxyinosine	HMDB0000071	C10H12N4O4	252.0859	POS_NEG	2.95	3.49	RME <= 15%	2.91	RME <= 15%	5.02	RME <= 15%	5.41	RME <= 15%
128	Methyldopa	HMDB0011754	C10H13NO4	211.0845	POS_NEG	2.95	3.49	RME <= 15%	2.91	RME <= 15%	5.02	RME <= 15%	5.41	RME <= 15%
129	thymine	HMDB0000262	C5H6N2O2	126.0429	POS_NEG	2.96	3.47	RME <= 15%	3.16	RME <= 15%	7.46	RME <= 15%	5.97	RME <= 15%
130	2-Thiocytidine dihydrate	CHEBI:19780; PUBCHEM CID:3011746	C9H13N3O4S	259.0627	POS_NEG	2.96	3.47	RME <= 15%	3.16	RME <= 15%	7.46	RME <= 15%	5.97	RME <= 15%
131	N1-Methyl-2-pyridone-5-carboxamide	HMDB0004193	C7H8N2O2	152.0586	POS_NEG	2.96	3.47	RME <= 15%	3.16	RME <= 15%	7.46	RME <= 15%	5.97	RME <= 15%
132	Xanthosine	HMDB0000299	C10H12N4O6	284.0757	POS_NEG	2.96	3.47	RME <= 15%	3.16	RME <= 15%	7.46	RME <= 15%	5.97	RME <= 15%
133	N1-Methyl-4-pyridone-3-carboxamide	HMDB0004194	C7H8N2O2	152.0586	POS_NEG	2.97	3.42	RME <= 15%	3.22	RME <= 15%	8.68	RME <= 15%	6.67	RME <= 15%
134	4-Pyridoxic acid	HMDB0000017	C8H9NO4	183.0532	POS_NEG	2.97	3.42	RME <= 15%	3.22	RME <= 15%	8.68	RME <= 15%	6.67	RME <= 15%
135	3-Methoxytyrosine	HMDB0001434	C10H13NO4	211.0845	POS_NEG	2.97	3.09	RME <= 15%	3.32	RME <= 15%	11.03	RME <= 15%	9.10	RME <= 15%
136	8-Hydroxy-2-deoxyguanosine	HMDB0003333	C10H13N5O5	283.0917	POS_NEG	2.97	3.09	RME <= 15%	3.32	RME <= 15%	11.03	RME <= 15%	9.10	RME <= 15%
137	1-methylurate	HMDB0003099	C6H6N4O3	182.0440	POS_NEG	2.97	3.09	RME <= 15%	3.32	RME <= 15%	11.03	RME <= 15%	9.10	RME <= 15%
138	2-O-Methyladenosine	HMDB0004326	C11H15N5O4	281.1124	POS_NEG	2.97	3.09	RME <= 15%	3.32	RME <= 15%	11.03	RME <= 15%	9.10	RME <= 15%
139	7-methylxanthine	HMDB0001991	C6H6N4O2	166.0491	POS_NEG	2.98	3.18	RME <= 15%	3.32	RME <= 15%	12.32	RME <= 15%	9.92	RME <= 15%
140	5-Methyluridine	HMDB0000884	C10H14N2O6	258.0852	POS_NEG	2.98	3.08	RME <= 15%	3.50	RME <= 15%	13.27	RME <= 15%	11.25	RME <= 15%
141	5-Hydroxy-L-tryptophan	HMDB0000472	C11H12N2O3	220.0848	POS_NEG	2.98	3.08	RME <= 15%	3.50	RME <= 15%	13.27	RME <= 15%	11.25	RME <= 15%
142	Serotonin	HMDB0000259	C10H12N2O	176.0950	POS_NEG	2.98	3.21	RME <= 15%	3.57	RME <= 15%	14.15	RME <= 15%	12.23	RME <= 15%
143	Thymidine	HMDB0000273	C10H14N2O5	242.0903	POS_NEG	3.02	3.39	RME <= 15%	3.72	RME <= 15%	16.62	15% < RME <= 30%	21.56	15% < RME <= 30%
144	N4-Acetylcytidine	HMDB0005923	C11H15N3O6	285.0961	POS_NEG	3.02	3.39	RME <= 15%	3.72	RME <= 15%	16.62	15% < RME <= 30%	21.56	15% < RME <= 30%
145	N2,N2-Dimethylguanosine	HMDB0004824	C12H17N5O5	311.1230	POS_NEG	3.05	3.30	RME <= 15%	3.33	RME <= 15%	8.29	RME <= 15%	20.43	15% < RME <= 30%

146	(3-(3-Hydroxy-phenyl)-3-hy-droxypropanoic acid) HPHPA	HMDB0002643	C9H10O4	182.0579	POS_NEG	3.05	3.30	RME <= 15%	3.33	RME <= 15%	8.29	RME <= 15%	20.43	15% < RME <= 30%
147	Kynurenine	HMDB0000684	C10H12N2O3	208.0848	POS_NEG	3.05	3.28	RME <= 15%	3.24	RME <= 15%	7.60	RME <= 15%	19.68	15% < RME <= 30%
148	Phenylalanine	HMDB0000159	C9H11NO2	165.0790	POS_NEG	3.06	3.72	RME <= 15%	3.26	RME <= 15%	7.10	RME <= 15%	19.04	15% < RME <= 30%
149	1,3-dimethylu-rate	HMDB0001857	C7H8N4O3	196.0596	POS_NEG	3.07	3.86	RME <= 15%	3.10	RME <= 15%	7.47	RME <= 15%	18.48	15% < RME <= 30%
150	Isobutyrylgly-cine	HMDB0000730	C6H11NO3	145.0739	POS_NEG	3.09	4.68	RME <= 15%	3.12	RME <= 15%	9.73	RME <= 15%	14.60	RME <= 15%
151	Pantothenic acid ï¼ˆvitamine B5ï¼‰	HMDB0000210	C9H17NO5	219.1107	POS_NEG	3.10	5.14	RME <= 15%	2.93	RME <= 15%	10.76	RME <= 15%	13.22	RME <= 15%
152	1,7-dimethylu-rate	HMDB0011103	C7H8N4O3	196.0596	POS_NEG	3.16	3.86	RME <= 15%	3.23	RME <= 15%	6.50	RME <= 15%	16.81	15% < RME <= 30%
153	n-methylnicotin-amide	HMDB0003152	C7H8N2O	136.0637	POS_NEG	3.17	3.70	RME <= 15%	3.28	RME <= 15%	6.16	RME <= 15%	17.41	15% < RME <= 30%
154	2-Furoylglycine	HMDB0000439	C7H7NO4	169.0375	POS_NEG	3.20	3.43	RME <= 15%	3.18	RME <= 15%	5.30	RME <= 15%	17.59	15% < RME <= 30%
155	4-hydroxyhippu-rate	HMDB0013678	C9H9NO4	195.0532	POS_NEG	3.21	3.17	RME <= 15%	2.80	RME <= 15%	5.67	RME <= 15%	15.21	15% < RME <= 30%
156	paraxanthine	HMDB0001860	C7H8N4O2	180.0647	POS_NEG	3.28	3.03	RME <= 15%	3.17	RME <= 15%	5.20	RME <= 15%	11.14	RME <= 15%
157	Xanthurenic acid	HMDB0000881	C10H7NO4	205.0375	POS_NEG	3.28	3.03	RME <= 15%	3.17	RME <= 15%	5.20	RME <= 15%	11.14	RME <= 15%
158	theophylline	HMDB0001889	C7H8N4O2	180.0647	POS_NEG	3.30	3.41	RME <= 15%	3.30	RME <= 15%	5.35	RME <= 15%	10.93	RME <= 15%
159	Tryptophan	HMDB0000929	C11H12N2O2	204.0899	POS_NEG	3.31	3.86	RME <= 15%	3.39	RME <= 15%	5.94	RME <= 15%	10.03	RME <= 15%
160	FAD	HMDB0001248	C27H33N9O15P2	785.1571	POS_NEG	3.32	3.84	RME <= 15%	3.55	RME <= 15%	6.07	RME <= 15%	9.96	RME <= 15%
161	3-Hydroxyan-thranilic Acid	HMDB0001476	C7H7NO3	153.0426	POS_NEG	3.33	3.94	RME <= 15%	3.40	RME <= 15%	6.32	RME <= 15%	10.16	RME <= 15%
162	3-hydroxyhippu-rate	HMDB0006116	C9H9NO4	195.0532	POS_NEG	3.35	3.94	RME <= 15%	3.25	RME <= 15%	6.98	RME <= 15%	11.55	RME <= 15%
163	3-Methylglutaric Acid	HMDB0000752	C6H10O4	146.0579	POS_NEG	3.36	4.13	RME <= 15%	3.27	RME <= 15%	7.70	RME <= 15%	11.97	RME <= 15%
164	2-Methylglutar-icacid	HMDB0000422	C6H10O4	146.0579	POS_NEG	3.37	4.05	RME <= 15%	3.34	RME <= 15%	7.60	RME <= 15%	12.42	RME <= 15%

165	N-Acetyl-L-tyrosine	HMDB0000866	C11H13NO4	223.0845	POS_NEG	3.43	3.14	RME <= 15%	3.77	RME <= 15%	6.78	RME <= 15%	16.71	15% < RME <= 30%
166	kynurenic acid	HMDB0000715	C10H7NO3	189.0426	POS_NEG	3.46	3.09	RME <= 15%	3.57	RME <= 15%	7.43	RME <= 15%	15.53	15% < RME <= 30%
167	Vanillactic acid	HMDB0000913	C10H12O5	212.0685	POS_NEG	3.56	3.48	RME <= 15%	3.94	RME <= 15%	6.63	RME <= 15%	12.43	RME <= 15%
168	quinaldic acid	HMDB0000842	C10H7NO2	173.0477	POS_NEG	3.64	3.41	RME <= 15%	3.09	RME <= 15%	7.19	RME <= 15%	12.56	RME <= 15%
169	5-hydroxyindoleacetic acid	HMDB0000763	C10H9NO3	191.0582	POS_NEG	3.70	3.43	RME <= 15%	10.20	RME <= 15%	5.80	RME <= 15%	11.48	RME <= 15%
170	indoxyl sulfuric acid	HMDB0000682	C8H7NO4S	213.0096	POS_NEG	3.70	3.43	RME <= 15%	10.20	RME <= 15%	5.80	RME <= 15%	11.48	RME <= 15%
171	4-Hydroxybenzoic acid	HMDB0000500	C7H6O3	138.0317	POS_NEG	3.70	3.27	RME <= 15%	10.60	RME <= 15%	5.83	RME <= 15%	11.33	RME <= 15%
172	indoxyl glucoside (Indican)	HMDB0061755	C14H17NO6	295.1056	POS_NEG	3.71	3.27	RME <= 15%	10.74	RME <= 15%	5.53	RME <= 15%	11.08	RME <= 15%
173	riboflavin (vitamin B2)	HMDB0000244	C17H20N4O6	376.1383	POS_NEG	3.72	3.07	RME <= 15%	11.15	RME <= 15%	4.83	RME <= 15%	10.82	RME <= 15%
174	Pyrocatechol	HMDB0000957	C6H6O2	110.0368	POS_NEG	3.72	2.98	RME <= 15%	11.08	RME <= 15%	4.96	RME <= 15%	10.82	RME <= 15%
175	Indoxyl glucuronide	HMDB0010319	C14H15NO7	309.0849	POS_NEG	3.77	2.88	RME <= 15%	6.25	RME <= 15%	4.66	RME <= 15%	14.55	RME <= 15%
176	4-hydroxyphenylacetic acid	HMDB0000020	C8H8O3	152.0473	POS_NEG	3.80	3.59	RME <= 15%	3.49	RME <= 15%	6.50	RME <= 15%	16.00	15% < RME <= 30%
177	2-AMINOCAPRYLIC ACID	HMDB0000991	C8H17NO2	159.1259	POS_NEG	3.81	3.69	RME <= 15%	3.46	RME <= 15%	7.39	RME <= 15%	15.89	15% < RME <= 30%
178	Phenylacetylglutamine	HMDB0006344	C13H16N2O4	264.1110	POS_NEG	3.82	3.64	RME <= 15%	3.38	RME <= 15%	7.57	RME <= 15%	14.72	RME <= 15%
179	biotin	HMDB0000030	C10H16N2O3S	244.0882	POS_NEG	3.85	3.01	RME <= 15%	3.06	RME <= 15%	7.38	RME <= 15%	13.63	RME <= 15%
180	Vanillic acid	HMDB0000484	C8H8O4	168.0423	POS_NEG	3.87	2.83	RME <= 15%	3.07	RME <= 15%	6.15	RME <= 15%	13.61	RME <= 15%
181	hippuric acid	HMDB0000714	C9H9NO3	179.0582	POS_NEG	3.87	2.83	RME <= 15%	3.07	RME <= 15%	6.15	RME <= 15%	13.61	RME <= 15%
182	Syringic acid	HMDB0002085	C9H10O5	198.0528	POS_NEG	3.91	2.58	RME <= 15%	3.08	RME <= 15%	6.83	RME <= 15%	20.68	15% < RME <= 30%
183	3-hydroxybenzoate	HMDB0002466	C7H6O3	138.0317	POS_NEG	3.93	2.95	RME <= 15%	3.50	RME <= 15%	10.23	RME <= 15%	33.94	RME > 30%
184	3-hydroxyphenylacetic acid	HMDB0000440	C8H8O3	152.0473	POS_NEG	3.97	3.08	RME <= 15%	3.97	RME <= 15%	12.28	RME <= 15%	41.02	RME > 30%
185	p-Cresol glucuronide	HMDB0011686	C13H16O7	284.0896	POS_NEG	4.12	2.53	RME <= 15%	3.67	RME <= 15%	5.50	RME <= 15%	18.06	15% < RME <= 30%

186	2-Hydroxy-phenylacetic acid	HMDB0000669	C8H8O3	152.0473	POS_NEG	4.13	2.31	RME <= 15%	3.87	RME <= 15%	5.74	RME <= 15%	18.16	15% < RME <= 30%
187	4-hydroxy-phenylpropionic acid	HMDB0002199	C9H10O3	166.0630	POS_NEG	4.14	2.13	RME <= 15%	3.81	RME <= 15%	5.90	RME <= 15%	17.71	15% < RME <= 30%
188	p-cresol sulfate	HMDB0011635	C7H8O4S	188.0143	POS_NEG	4.16	2.47	RME <= 15%	4.51	RME <= 15%	6.26	RME <= 15%	16.40	15% < RME <= 30%
189	4-Hy-droxycinnamic acid	HMDB0002035	C9H8O3	164.0473	POS_NEG	4.20	3.33	RME <= 15%	4.17	RME <= 15%	5.06	RME <= 15%	13.80	RME <= 15%
190	Anthranilic acid (2-Aminoben-zoic acid)	HMDB0001123	C7H7NO2	137.0477	POS_NEG	4.26	3.27	RME <= 15%	3.37	RME <= 15%	4.40	RME <= 15%	17.76	15% < RME <= 30%
191	trans-Ferulic acid	HMDB0000954	C10H10O4	194.0579	POS_NEG	4.33	2.61	RME <= 15%	3.17	RME <= 15%	4.51	RME <= 15%	12.23	RME <= 15%
192	N-acetyltrypto-phan	HMDB0013713	C13H14N2O3	246.1004	POS_NEG	4.37	2.34	RME <= 15%	2.99	RME <= 15%	5.14	RME <= 15%	11.65	RME <= 15%
193	2-hydroxyhippu-ric acid (Salicyluric acid)	HMDB0000840	C9H9NO4	195.0532	POS_NEG	4.37	2.34	RME <= 15%	2.99	RME <= 15%	5.14	RME <= 15%	11.65	RME <= 15%
194	4-methyl cate-chol sulfate	HMDB0000873	C7H8O2	124.0524	POS_NEG	4.40	2.86	RME <= 15%	3.02	RME <= 15%	5.03	RME <= 15%	11.54	RME <= 15%
195	3-methylindole (skatol)	HMDB0000466	C9H9N	131.0735	POS_NEG	4.41	3.23	RME <= 15%	2.95	RME <= 15%	4.82	RME <= 15%	11.79	RME <= 15%
196	Indolelactic acid	HMDB0000671	C11H11NO3	205.0739	POS_NEG	4.42	3.37	RME <= 15%	3.33	RME <= 15%	5.02	RME <= 15%	11.70	RME <= 15%
197	Phenylpropio-nylglycine	HMDB0000860	C11H13NO3	207.0895	POS_NEG	4.43	3.35	RME <= 15%	3.65	RME <= 15%	4.74	RME <= 15%	11.52	RME <= 15%
198	Cinnamoylgly-cine	HMDB0011621	C11H11NO3	205.0739	POS_NEG	4.53	6.01	RME <= 15%	2.88	RME <= 15%	4.43	RME <= 15%	9.13	RME <= 15%
199	Melatonin	HMDB0001389	C13H16N2O2	232.1212	POS_NEG	4.60	9.60	RME <= 15%	4.99	RME <= 15%	5.85	RME <= 15%	13.09	RME <= 15%
200	Benzoic acid	HMDB0001870	C7H6O2	122.0368	POS_NEG	4.68	7.48	RME <= 15%	4.43	RME <= 15%	5.04	RME <= 15%	14.86	RME <= 15%
201	Indoleacetic acid	HMDB0000197	C10H9NO2	175.0633	POS_NEG	4.74	5.88	RME <= 15%	3.94	RME <= 15%	9.51	RME <= 15%	26.41	15% < RME <= 30%
202	phenylacetic acid	HMDB0000209	C8H8O2	136.0524	POS_NEG	4.74	5.88	RME <= 15%	3.94	RME <= 15%	9.51	RME <= 15%	26.41	15% < RME <= 30%
203	3-Indoleacrylic acid	HMDB0000734	C11H9NO2	187.0633	POS_NEG	4.89	5.26	RME <= 15%	3.70	RME <= 15%	4.80	RME <= 15%	9.16	RME <= 15%
204	Taurohyocholic acid	PubChem CID:11954195	C26H45NO7S	515.2917	POS_NEG	4.90	5.40	RME <= 15%	3.86	RME <= 15%	4.72	RME <= 15%	9.19	RME <= 15%
205	cortisol	HMDB0000063	C21H30O5	362.2093	POS_NEG	5.02	9.65	RME <= 15%	3.26	RME <= 15%	7.84	RME <= 15%	19.36	15% < RME <= 30%

206	Indolepropionic acid	HMDB0002302	C11H11NO2	189.0790	POS_NEG	5.08	9.24	RME <= 15%	3.51	RME <= 15%	8.30	RME <= 15%	15.35	15% < RME <= 30%
207	4-Methylvaleric acid	HMDB0000689	C6H12O2	116.0837	POS_NEG	5.08	9.24	RME <= 15%	3.51	RME <= 15%	8.30	RME <= 15%	15.35	15% < RME <= 30%
208	Taurocholic acid	HMDB0000036	C26H45NO7S	515.2917	POS_NEG	5.13	9.13	RME <= 15%	3.02	RME <= 15%	7.60	RME <= 15%	10.21	RME <= 15%
209	3-phenylpropa-noic acid	HMDB0000764	C9H10O2	150.0681	POS_NEG	5.13	9.14	RME <= 15%	3.03	RME <= 15%	7.68	RME <= 15%	10.01	RME <= 15%
210	trans-Cinnamic acid	HMDB0000930	C9H8O2	148.0524	POS_NEG	5.16	8.81	RME <= 15%	3.20	RME <= 15%	7.59	RME <= 15%	9.41	RME <= 15%
211	2-phenylpropi-onic acid	HMDB0011743	C9H10O2	150.0681	POS_NEG	5.16	8.77	RME <= 15%	3.20	RME <= 15%	8.09	RME <= 15%	9.86	RME <= 15%
212	Glycohyocholic acid	PubChem CID:71361462	C26H43NO6	465.3090	POS_NEG	5.25	11.77	RME <= 15%	3.52	RME <= 15%	9.57	RME <= 15%	7.79	RME <= 15%
213	Taurolithocholic acid 3-sulfate	HMDB0002580	C26H45NO8S2	563.2587	POS_NEG	5.29	12.63	RME <= 15%	2.73	RME <= 15%	10.50	RME <= 15%	7.49	RME <= 15%
214	Indolebutyric acid	HMDB0002096	C12H13NO2	203.0946	POS_NEG	5.33	10.28	RME <= 15%	2.76	RME <= 15%	11.36	RME <= 15%	7.04	RME <= 15%
215	Glycocholic acid	HMDB0000138	C26H43NO6	465.3090	POS_NEG	5.44	5.63	RME <= 15%	2.68	RME <= 15%	12.89	RME <= 15%	6.09	RME <= 15%
216	Glycohyodeoxy-cholic acid	PubChem CID:114611	C26H43NO5	449.3140	POS_NEG	5.46	5.85	RME <= 15%	2.65	RME <= 15%	13.48	RME <= 15%	7.93	RME <= 15%
217	Glycoursodeox-ycholic acid	HMDB0000708	C26H43NO5	449.3141	POS_NEG	5.47	5.62	RME <= 15%	2.75	RME <= 15%	13.66	RME <= 15%	10.26	RME <= 15%
218	CMPF	HMDB0061112	C12H16O5	240.0998	POS_NEG	5.50	5.86	RME <= 15%	2.91	RME <= 15%	13.08	RME <= 15%	12.84	RME <= 15%
219	indoxyl	HMDB0004094	C8H7NO	133.0528	POS_NEG	5.53	5.58	RME <= 15%	2.50	RME <= 15%	12.52	RME <= 15%	10.07	RME <= 15%
220	dehydrocholic acid (3,7,12-Tri-oxo-5Î²-cholanic acid)	HMDBnone1	C24H34O5	402.2406	POS_NEG	5.56	5.84	RME <= 15%	2.78	RME <= 15%	11.92	RME <= 15%	9.81	RME <= 15%
221	Tau-rochenodeoxy-cholic acid	HMDB0000951	C26H45NO6S	499.2968	POS_NEG	5.59	5.71	RME <= 15%	2.76	RME <= 15%	11.78	RME <= 15%	11.32	RME <= 15%
222	Taurodeoxy-cholic acid	HMDB0000896	C26H45NO6S	499.2968	POS_NEG	5.73	5.80	RME <= 15%	2.96	RME <= 15%	14.06	RME <= 15%	7.43	RME <= 15%
223	hyocholic acid	HMDB0000760	C24H40O5	408.2876	POS_NEG	5.79	3.78	RME <= 15%	3.21	RME <= 15%	15.65	15% < RME <= 30%	6.69	RME <= 15%
224	Pregnenolone Sulfate	HMDB0000774	C21H32O5S	396.1970	POS_NEG	5.97	4.07	RME <= 15%	3.51	RME <= 15%	15.56	15% < RME <= 30%	4.70	RME <= 15%
225	cholic acid	HMDB0000619	C24H40O5	408.2876	POS_NEG	5.98	3.85	RME <= 15%	3.47	RME <= 15%	15.52	15% < RME <= 30%	4.94	RME <= 15%

Chapter III

Matrix effects in untargeted LC-MS metabolomics: from creation to compensation with post-column infusion of standards

Based on:

Pingping Zhu, Amy Harms, Pascal Maas, Manisha Bakas, Julia Josette Whien, Anne-Charlotte Dubbelman, Thomas Hankemeier

Matrix Effects in Untargeted LC-MS Metabolomics: From Creation to Compensation with Post-Column Infusion of Standards

Journal of Chromatography A

DOI: 10.1016/j.chroma.2025.466508

Abstract

Matrix effect is a well-known issue affecting accuracy and repeatability in metabolomics studies using liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS). Post-column infusion of standards (PCIS) is a promising strategy to monitor and correct matrix effect but has been rarely reported in untargeted metabolomics. The major challenges lie in selecting appropriate PCISs and identifying the most suitable PCIS to correct the matrix effect experienced by each feature. In this study, we aim to present a method for selecting suitable PCISs for matrix effect compensation based on the artificial matrix effect (ME_{art}) created by post-column infusion of compounds that disrupt the ESI process. Our hypothesis is that the suitable PCIS for a given analyte can be identified by comparing the PCISs' ability in ME_{art} compensation. We evaluated this approach using 19 stable-isotopically labeled (SIL) standards spiked in plasma, urine, and feces. PCISs selected based on ME_{art} were compared to those selected by biological matrix effect (ME_{bio}), with 17 out of 19 SIL standards (89%) showing consistent PCIS selection, demonstrating the effectiveness of ME_{art} in identifying suitable PCISs. Applying ME_{art}-selected PCISs to correct for the ME_{bio} resulted in improved ME_{bio} for most of the SILs affected by matrix effect and maintained ME_{bio} for those experiencing no matrix effect. We demonstrated the efficacy of ME_{art} in selecting suitable PCISs for ME_{bio} correction within an LC-PCIS-MS method. Importantly, since ME_{art} can be assessed for any detected feature, its application holds great potential for identifying suitable PCISs for matrix effect correction in untargeted metabolomics.

1. Introduction

Matrix effect (ME), primarily caused by coeluting matrix components, poses a significant challenge in liquid chromatography-electrospray ionization-mass spectrometry (LC−ESI−MS). It can alter analytes' ionization efficiency through ion suppression or enhancement, affecting the accuracy and reliability of their quantification.[1,2] ME is categorized as absolute matrix effect (AME) and relative matrix effect (RME). AME describes the response differences of an analyte spiked in a biological sample *vs.* a matrix-free sample. RME is defined as the relative standard deviation (RSD%) of the AME among biological samples from different sources, indicating the sample to sample variation.[3] To mitigate ME, strategies such as extensive sample preparation, sample dilution, and tailored LC separation have been employed.[4] The most common method for evaluating ME is post-extraction spiking (PES) of an analyte or its analogue into biological and matrix-free sample and comparing their responses, which is widely applied in targeted metabolomics.[2] For ME correction, an efficient approach is spiking surrogate analytes or internal standards, typically stable isotopically labeled (SIL) standards, into a study sample, then correct the signal of an analyte by that of a surrogate or SIL standard.[5] Although these approaches are effective for ME evaluation and compensation, their application can be limited by high cost and limited commercial availability of analyte analogues and SIL standards.[4,5] Besides, even deuterium-labeled standards can exhibit retention time shifts compared to the analytes due to altered physicochemical properties, which reduces the efficiency of ME correction.[6,7]

The disadvantages of PES and SIL standards spiking can be mitigated by another technique used for addressing ME in LC-MS-based metabolomics: post-column infusion of standard (PCIS). Unlike PES and SIL standards spiking, which assess and correct ME at specific retention times, PCIS allows for ME evaluation and compensation across the entire chromatographic profile by constantly infusing one or several standards into the LC-MS post-column.[1,2] In 1999, PCIS was introduced to monitor ME in plasma samples,[8] and to correct ME in environmental samples.[9] In PCIS, ME can be evaluated or monitored by comparing the signals of an infused standard

between the injections of matrix and solvent samples.[8] Meanwhile, correction can be achieved by normalizing the analyte signal to the signal of a PCIS in a matrix sample.[9] PCIS has proven effective for monitoring or correcting ME in various targeted LC-MS based studies. Applications include quantifying pharmaceuticals in waste water,[10,11] chicken meat,[12] human urine,[11,13–15] human plasma,[11,16] and dry blood spots samples;[17,18] targeting pesticides in food extracts;[19] analyzing steroids,[20] amino acids,[21–23] phospholipids,[24] and other endogenous metabolites. In most of these studies, a structural analogue of the analyte or a single SIL standard is used as the infused standard, which significantly reduces costs compared to using multiple SIL standards. Importantly, different from PES and SIL standards spiking, which are restricted to targeted metabolomics, PCIS is also applicable in untargeted metabolomics due to its independence from retention time.

Although PCIS has been recommended as a quality control tool for ME evaluation in untargeted analysis,[26] its actual use remains limited. Tisler *et al.* demonstrated that PCIS is a suitable approach for correcting the RME of waste water in untargeted profiling.[27] Our recent study showed that PCIS can efficiently monitor the ME in human plasma and fecal samples in untargeted metabolomics.[28] One of the primary obstacle limiting the implementation of PCIS in untargeted metabolomics probably lies in selecting suitable PCIS candidates for diverse metabolome features. The similarity of hydrophobicity and ionization ability between the analytes and PCIS are important factors for efficient ME correction.[13] However, pre-selecting PCIS candidates for all the detected features in untargeted metabolomics according to the physical-chemical properties is impractical, particularly for the unknown ones. This highlights the necessity of physical-chemical diversity in PCIS candidates applied in untargeted metabolomics. Tisler *et al.* evaluated the diversity of six PCISs by examining the variation of their monitored ME. They concluded that the ME consists of retention-time dependent ME and structural-specific ME.[27] Retention-time dependent matrix were compensated using the median value of the ME obtained with all PCISs, while structure-specific ME were addressed with a quantitative structure-property relationships (QSPR) model.[27] Nevertheless, the QSPR model is target dependent, as it requires the physical-chemical property of a compound to predict the structure-specific ME.[27] Thus, an ideal

approach that considers both co-eluting matrix compounds and structure diversities is still lacking for applying PCIS to correct ME in untargeted metabolomics. Instead of using the median ME obtained from several PCISs, matching each feature to its suitable PCIS could help to address the issue of structure diversity. However, this raises another challenge for implementing PCIS to compensate for ME in untargeted metabolomics: how to match a detected feature to its appropriate PCIS?

In this study, we aim to develop a novel methodology for PCIS matching in an LC-PCIS-MS-based untargeted metabolomics method. To achieve this, we first discussed key factors, including concentration optimization and diversity evaluation, for selecting PCIS candidates. Then, a post-column artificial matrix infusion approach was introduced to the developed LC-PCIS-MS method for PCIS matching. The artificial matrices consist of compounds that disrupt the ionization process in the ESI source. Therefore, by comparing the signals of an analyte with and without artificial matrix infusion, its artificial ME (ME_{art}) can be determined. Our hypothesis is that ME_{art} could be used to identify the suitable PCIS for the analyte by comparing the PCISs' ability for its correction. We demonstrate the utility of this approach in a proof-of-concept study, where 19 diverse SIL standards were spiked into plasma, urine, and feces. Their most suitable PCISs, selected based on compensation for biological ME (ME_{bio}) and ME_{art}, were then compared. Afterward, the efficiency of the ME_{art}-selected PCISs in correcting ME_{bio} was examined for the 19 SIL standards.

2. Material and method

2.1 Chemicals and materials

LC-MS-grade acetonitrile (ACN) and methanol (MeOH) were purchased from Actu-all chemicals (Randmeer, The Netherlands). Methyl tert-butyl ether (MTBE, ≥99.8%) was purchased from Sigma Aldrich (St. Louis, Missouri, United States). Formic acid (FA) was purchased from Biosolve B.V. (Valkenswaard, Netherlands), and hydrochloric acid (37% solution in water) was purchased from Acros organics (Geel, Belgium). Purified water was obtained from a Milli-Q PF Plus system (Merck Millipore, Burlington, Massachusetts, United States). Table S1 provides the supplier details of all standards, including the PCIS candidates, artificial matrix compounds and stable isotopically labeled (SIL) standards. EDTA plasma was obtained from Sanquin (Sanquin, Amsterdam, The Netherlands) and BioIVT (Westbury, NY, USA). Urine and fecal samples were collected from four healthy volunteers (age range: 23-35 years).

2.2 Solution preparation for PCISs, artificial matrix compounds and SIL standards.

Stock solutions of leucine-enkephalin (Leu-enk), fludrocortisone (F-Cor), 5-fluoroisatin (F-Isat), caffeine-$^{13}C_3$ (Caff-$^{13}C_3$), 3-fluoro-DL-valine (F-Val), D-glucose-d_7 (Glu-d_7) were prepared as described in Table S1. The stock solutions of all PCISs were diluted with 50% ACN in water to 50 µg/mL, 5 µg/mL, and 1 µg/mL for concentration optimization. L-homoarginine hydrochloride (hArgHC), sodium dodecyl sulphate (SDS), and tridodecylmethylammonium chloride (TDMAC) were dissolved in 50% ACN in water, while sodium acetate (NaOAc) was prepared in 20% ACN (Table S1). Those standards were used as artificial matrix compounds, and their stock solutions were diluted with 50% ACN in water for concentration optimization. The stock solution preparation of the 19 SIL standards and their concentrations after spiking in plasma, urine, and feces are described in Table S1 and Table S2, respectively.

2.3 Sample preparation

Fecal and plasma samples were prepared as previously reported.[28] Briefly, 20mg freeze-dried fecal samples were extracted by liquid-liquid extraction with the mixture of

water/MeOH/MTBE (v/v/v), then 90 µL of aqueous layer was dried and reconstituted in 50 µL water containing 0.1% FA. Plasma samples were prepared with protein precipitation: 100 µL of ice-cold MeOH was added to 25µL of plasma sample, followed by drying of the supernatant and reconstitution in 75 µL of water containing 0.1% FA. Urine samples were prepared identically to plasma samples. For ME evaluation for SIL standards, the mixture of SILs was spiked into biological and matrix-free samples after extraction.

2.4 LC-MS setup with post-column infusion

Sample measurements were performed using either a Shimadzu Nexera X2 LC system coupled to a TripleTOF 6600 mass spectrometer (SCIEX, Foster City, CA, USA) or a Waters Acquity UPLC Class II LC system coupled to TripleTOF 5600 mass spectrometry (SCIEX, Foster City, CA, USA). For both systems, an ESI source was used, and the same LC-MS conditions and the post column setup were applied, as detailed in our previous study.[28] In short, data were acquired under full scan mode over the *m/z* range of 60-800 Da in both positive and negative modes. The LC separation was achieved by using a Waters Acquity UPLC HSS T3 column (1.8 µm, 2.1 mm × 100 mm) over a 15 min gradient with 0.1% FA in water and 0.1% FA in ACN as mobile phases. The LC flow was diverted to waste at 7 min to decelerate contamination of the MS. A binary Agilent 1260 Infinity pump (Agilent Technologies, Santa Clara, USA) was used for post-column infusion at a flow rate of 20 µL min^{-1}. The post-column flow was combined with the LC eluent using a T connector (IDEX, PEEK Tee, 0.02 Thru hole, F-300) before injecting to the MS.

2.5 Data processing

Raw data were acquired by Analyst TF software 1.7.1 (SCIEX) and processed using SCIEX OS (version 2.1, SCIEX) and PeakView (version 2.2, SCIEX). Extracted ion chromatograms (EICs) for all PCIS candidates were obtained with an *m/z* window of 0.02 Da. A maximum mass error of 5 ppm was applied for peak integration of endogenous compounds and SIL standards. The infusion profiles of the PCIS candidates were generated by smoothing the extracted EIC data using the simple moving average

(SMA, n = 15) function in R (version 4.3.2). The ME profile (MEP) for each PCIS was generated as reported previously.[28]

Different types of ME were calculated for the SIL standards in plasma, urine, and feces, separately. The biological absolute matrix effect (AME_{bio_i}) and relative matrix effect (RME_{bio}) in each type of biological matrix (plasma, urine, feces) were calculated as shown in Equation 1 and 2. The calculation uses the integrated peak area (A) in a biological sample (*i*) from each biological matrix type (bio) (A_{bio_i}) and that in a matrix-free (solvent) sample (A_{sol}). The artificial absolute matrix effect (AME_{art_j}) created by artificial matrix was calculated as Equation 3. The artificial matrix includes individual artificial matrix compounds as well as their mixture, making different artificial matrix combinations (*j*). For each biological matrix, the integrated peak areas of the SIL standards in a pooled biological matrix type ($\overline{bio}$) with ($A_{art_j+\overline{bio}}$) and without ($A_{\overline{bio}}$) artificial matrix infusion were used for AME_{art_j}calculation. The relative artificial matrix effect (RME_{art}) was calculated as the relative standard deviation (RSD %) among the AME_{art_j} obtained from different artificial matrix combinations (Equation 4).

The PCIS-corrected response of each SIL was generated through integrating the ratio obtained from dividing the signal of a SIL standard by that of an individual PCIS at each time point with an in-house software. In each sample, the retention time and peak width of individual SIL standards before PCIS correction were used to identify the regions for ratio integration, and the integration of the PCIS-corrected signal was manually examined. The AME_{bio_i}, RME_{bio}, AME_{art_j}, and RME_{art} after PCIS correction were calculated as described in Equation 1-4, but with the replacement of peak area by PCIS-corrected area. To evaluate the overall ME caused by the biological matrix (ME_{bio}) or the artificial matrix (ME_{art}) for each matrix type, a scoring system combining the absolute matrix effect (AME) and relative matrix effect (RME) was applied, as shown in Table 1. For ME_{bio} score, the averaged AME_{bio_i} score from different individuals ($\overline{AME}_{bio_{1-i}}$) was used for the calculation, while the AME_{art} obtained from the artificial matrix compounds mixture ($AME_{art_{mix}}$) was used for ME_{art} scoring.

Equation 1: $\mathrm{AME}_{bio_i}(\%) = \frac{A_{bio_i}}{A_{sol}} * 100$

Equation 2: $\mathrm{RME}_{bio}(\%) = \frac{\mathrm{SD}(AME_{bio_1} \ldots AME_{bio_i})}{\mathrm{Mean}(AME_{bio_1} \ldots AME_{bio_i})} * 100$

Equation 3: $\mathrm{AME}_{art_j}(\%) = \frac{A_{art_j + \overline{bio}}}{A_{\overline{bio}}} * 100$

Equation 4: $\mathrm{RME}_{art}(\%) = \frac{\mathrm{SD}(AME_{art_1} \ldots AME_{art_j})}{\mathrm{Mean}(AME_{art_1} \ldots AME_{art_j})} * 100$

Table 1. The scoring system for absolute matrix effect (AME), relative matrix effect (RME), biological matrix effect (ME_{bio}), and artificial matrix effect (ME_{art})

Conditions*	**Scoring Formula**
AME <= 100	AME score = 100 * (AME/100)
AME > 100	AME score = 100 / (AME/100)
RME	RME score = 100 - RME
ME_{bio}	ME_{bio} score = ($\overline{AME}_{bio_{1-i}}$ score + RME_{bio} score) /2
ME_{art}	ME_{art} score = ($AME_{art_{mix}}$ score + RME_{art} score) /2

*AME (%) = 100 indicates no matrix effect; AME (%) < 100 indicates ion suppression; AME (%) >100 indicates ion enhancement

3. Results and discussion

3.1 PCIS method development

To develop a suitable PCIS approach for our untargeted metabolomics method, PCIS candidates with diverse structures were examined. Important factors such as adduct formation, infusion profile diversity, infusion concentration, room temperature stability, and matrix effect profile (MEP) diversity were evaluated. Plasma, urine, and feces were used in the selection process, ensuring selected PCIS could be effectively applied across diverse biological matrices.

3.1.1 PCIS selection and infusion concentration optimization

Ideally, a PCIS should be commercially affordable, measurable with specific signal, detectable mainly with protonated $[M+H]^+$ and deprotonated $[M-H]^-$ ions, and stable during analysis.[29] With this in mind, six xenobiotic compounds with different physicochemical properties were evaluated as PCIS candidates in our study. All six standards were examined in positive ionization mode, and five were assessed in negative ionization mode. (Table S3). First, we examined the adduct formation of all the PCIS candidates: $[M+H]^+$ and $[M-H]^-$ were the ions with highest response for most of the candidates, except for Glu-d_7 in positive mode and F-cor in negative mode. The former showed a higher signal with sodium ($[M+Na]^+$) and ammonium ($[M+NH_4]^+$) adducts than with $[M+H]^+$, while the latter had a higher signal as the formic acid adduct ($[M+FA-H]^-$) compared to $[M-H]^-$. In addition to $[M-H]^-$, Glu-d_7 also showed good intensity with $[M+FA-H]^-$. Considering that the infusion profiles of a PCIS may vary with different adducts,[30] for Glu-d_7, we monitored both $[M+Na]^+$ and $[M+NH_4]^+$ in positive mode, as well as both $[M-H]^-$ and $[M+FA-H]^-$ in negative mode. However, only $[M+FA-H]^-$ was monitored for F-Cor in negative mode, as the signal of $[M-H]^-$ was too low to generate a stable infusion profile.

Subsequently, Pearson correlation was applied to evaluate the diversity of the infusion profiles among PCIS candidates. The EICs of all PCIS candidates were extracted and correlated with each other after the injection of plasma samples. This procedure was repeated with four different plasma samples to include sample diversity. As shown in Table S4-5, Leu-enk showed a near identical infusion profile to F-Val in positive mode ($r > 0.95$ in three examined plasma samples), and to F-Isat in negative mode ($r > 0.99$ in all examined plasma samples). Therefore, five PCIS (Leu-enk, F-Cor, F-Isat, Caff-$^{13}C_3$, Glu-d_7) were selected for positive mode, and four (Leu-enk, F-Cor, F-Val, Glu-d_7) for negative mode.

Table 2. Monitored ions and optimized infusion concentrations for selected PCISs

PCIS		Detected ions		Infusion concentration (ng/mL)	
full name	abbreviation	positive	negative	positive	negative
Leucine-enkephalin	Leu-enk	$[M+H]^+$	$[M-H]^-$	212.4	344.9
Fludrocortisone	F-Cor	$[M+H]^+$	$[M+FA-H]^-$	154.0	371.0
D-glucose-d_7	Glu-d_7	$[M+Na]^+$, $[M+NH_4]^+$	$[M-H]^-$, $[M+FA-H]^-$	5734.0	6465.7
5-Fluoroisatin	F-Isat	$[M+H]^+$	/	1069.2	/
Caffeine-$^{13}C3$	Caff-$^{13}C_3$	$[M+H]^+$	/	219.2	/
3-Fluoro-DL-valine	F-Val	/	$[M-H]^-$	/	4264.3

After selection, the infused concentrations of the PCISs were optimized to balance the trade-off between signal intensity and PCIS-induced ME. Concentration optimization is widely discussed in studies applying PCIS,[12,17,22,29] and the ubiquitous goal is to achieve stable infusion signal without inducing additional ME. Figure 1A presents infusion profiles of the PCISs extracted from one plasma sample at the optimized PCIS infusion concentrations (Table 2). In both ionization modes, the initial intensities of the main monitored ions were above 20,000 cps for all PCISs, which was high enough for clear and stable signal monitoring. Stable infusion signals were monitored for all PCISs over plasma injections (Figure S1A-B). Although there were regions (0.5-0.8 min, 1.5-1.8 min) with severe signal suppression, the lowest signals of the PCISs remained above 100 cps. The exceptions were $[Glu\text{-}d_7+NH_4]^+$ in positive mode, and $[Glu\text{-}d_7+FA\text{-}H]^-$ in negative mode at 0.5-0.8 min, as shown in the zoomed-in sections at the top left of Figure 1A.

To assess whether the selected PCISs were also applicable in other biological matrices at the optimized concentrations, the infusion profiles of the PCISs were inspected in three different urine (Figure S1C-D) and fecal samples (Figure S1E-F). The infusion profiles of each PCIS were constructed in urine and feces by averaging signals from three individuals, as presented in Figure 1B and 1C. Similar to plasma, the initial infusion signals were above 20,000 cps and the lowest infusion signals of the PCISs were above 100 cps in both urine and fecal samples. This indicated that abundant and

stable infusion profiles were also achieved for the selected PCISs at the optimized concentrations in plasma and feces.

Then, we evaluated the impact of PCISs on analyte signals by comparing the peak areas of several known metabolites in the same plasma sample, with and without PCIS infusion. In total, 60 targets were used for the signal comparison in positive mode, and 36 in negative mode (Table S6). As shown in Figure 1D, no significant differences in peak areas were found for the examined metabolites with and without PCIS infusion in both ionization modes. Compared with infusion, the signal changes of most examined metabolites with PCIS are within ±30% (Table S6). Additionally, the room temperature stability of the selected PCIS was examined over seven days by injecting the PCIS mixture solution in positive mode. From day 1 to day 7, the signal variations of all PCISs were within 10% compared to the freshly prepared solution on day 0. (Figure S2).

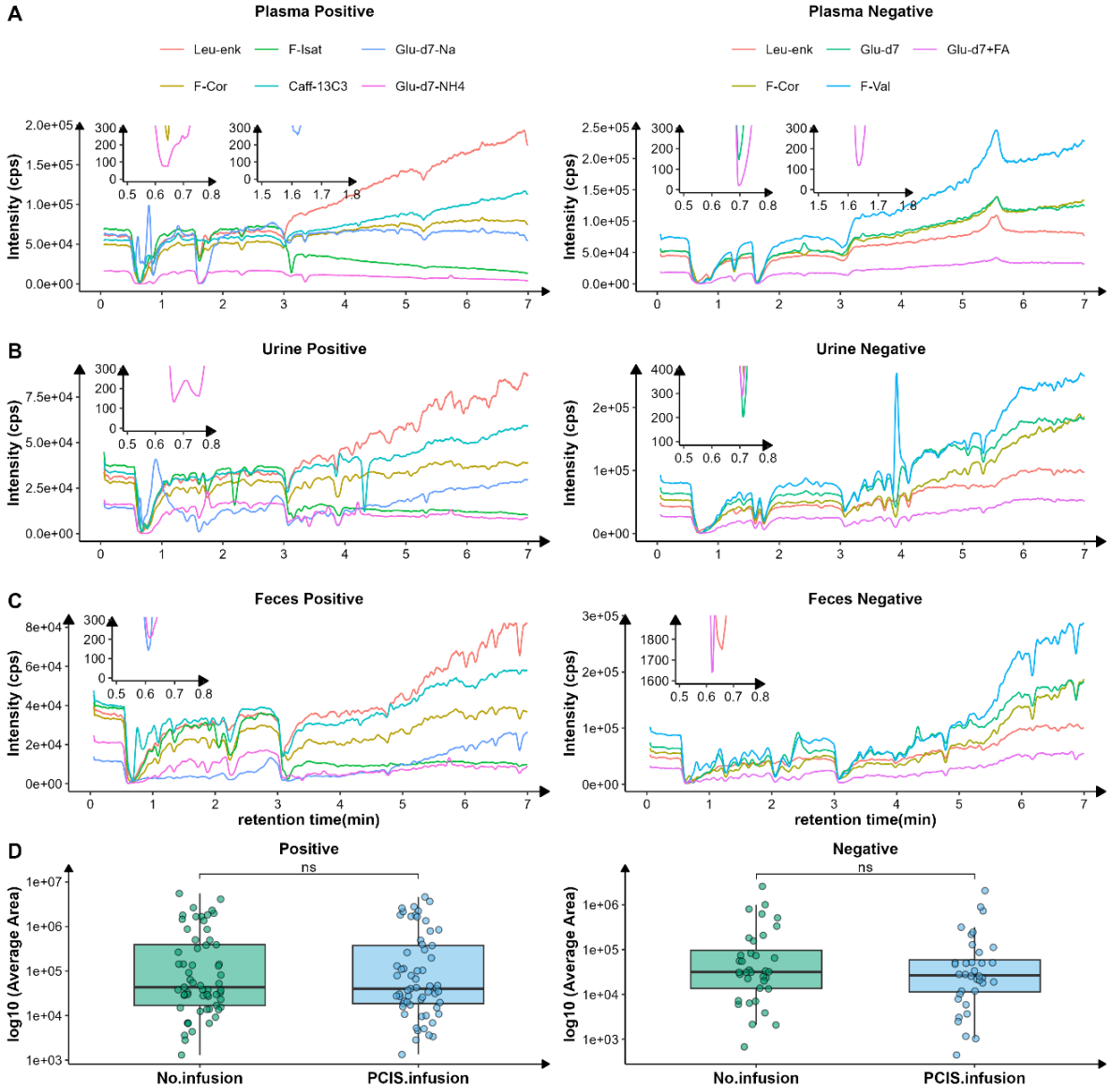


Figure 1. Infusion profiles over 0-7 mins and zoomed-in inspections of regions with severe suppression for the selected PCISs in plasma (A), urine (B) and feces (C), as well as the peak area comparison between plasma injections with and without PCIS infusion in positive and negative ionization modes (D). The intensities plotted in (A) and the peak area for each examined metabolites used in (D) were the mean values from duplicated injections of the same plasma sample; the intensities plotted in (B) and (C) are the mean values of three different individuals; A two-side unpaired t test was applied for statistical assessment in (D).

3.1.2 PCIS diversity evaluation with matrix effect profiles

To clearly identify ion suppression and enhancement over the entire chromatography, the infusion profiles of the PCISs in each biological matrix were normalized against those in the solvent samples, creating the MEP.[19] The MEPs of each PCIS, generated with three different individuals from plasma, urine, and feces, are presented in Figure S3. The averaged MEP for each PCIS ($\overline{\text{MEP}}$) was calculated from the MEPs of different individuals in each biological matrix, and the $\overline{\text{MEP}}$ variation plots were created by overlaying the $\overline{\text{MEP}}$ of all PCISs. As shown in Figure 2A, for each matrix and ionization mode, the solid line represents the overall AME monitored with all PCISs, while the shaded area shows the variations of the $\overline{\text{MEP}}$ among all PCISs.

To directly display the $\overline{\text{MEP}}$ variation among the PCISs, the RSD% of the $\overline{\text{MEP}}$ was calculated per timepoint and plotted for each biological matrix, as presented in Figure 2B. The RSD% of the $\overline{\text{MEP}}$ monitored with the same set of PCISs varied among plasma, urine and feces. In plasma, high diversity was mainly observed in the early elution region (RT < 2 min), with RSD > 15% in both ionization modes. In urine, apart from the early elution regions, diverse MEPs were also noted within 2-4 mins, particularly in positive mode. In feces, the RSD % of the $\overline{\text{MEP}}$ was above or close to 15 % almost throughout the entire chromatogram in both ionization modes. Considering the matrix complexity, it is expected that the $\overline{\text{MEP}}$ variation is larger in feces than in urine and plasma. This observation is consistent with the study by Stahnke *et al*., who reported that a more complex matrix can induce larger variations among infused pesticides.[19] Tisler *et al.* also observed that, compared to diluted waste water, the waste water with more concentrated matrix varied more in MEP compared to the diluted one.[31] In our study, the diversity of the three biological matrices is successfully reflected by the $\overline{\text{MEP}}$ variation of selected PCISs, making it feasible to apply these PCISs to assess the ME from less to more complex biological matrices.

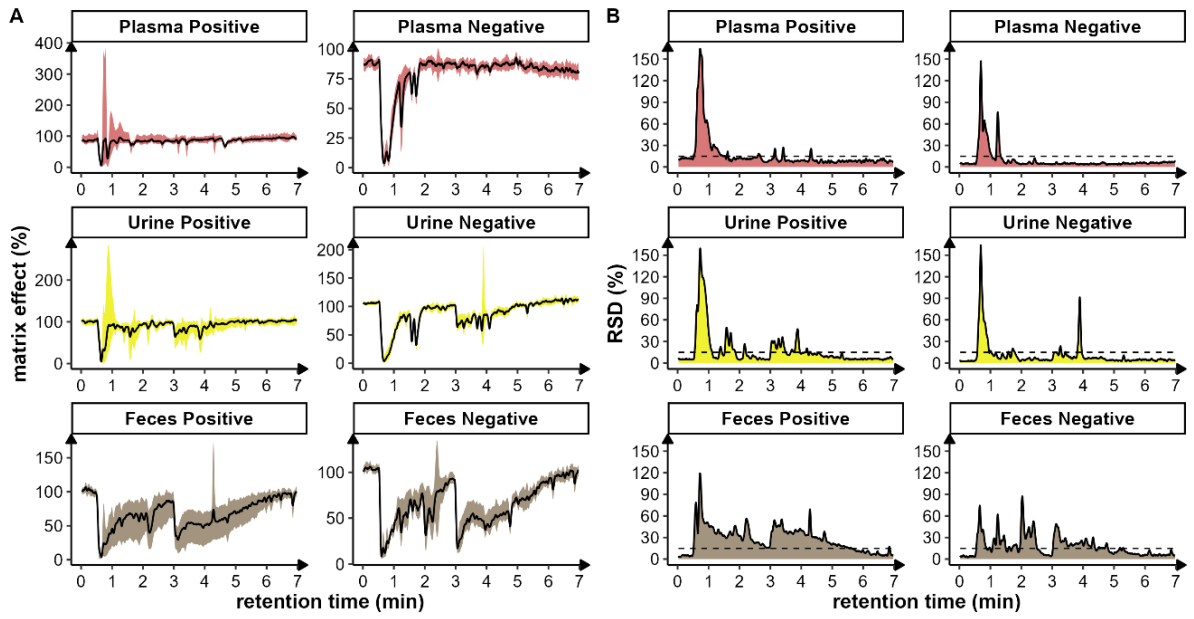


Figure 2. MEP variation plots with overlaid averaged MEP ($\overline{MEP}$) among all the PCIS (A) and RSD of the $\overline{MEP}$ (B) for all PCISs in plasma, urine, and feces for both ionization modes. The dashed line in (B) indicates where RSD is 15%.

3.2 PCIS matching using post-column artificial matrix effect creation

With multiple PCISs available, it's crucial to select one that resembles the analyte in its susceptibility to ion suppression or enhancement to effectively correct for its ME. Therefore, we introduced an approach, post-column artificial ME creation, to the developed LC-PCIS-MS method to match analytes to their suitable PCISs. In this approach, along with the PCIS, the artificial matrix, consisting of a set of compounds that create ME, was continuously infused to the ESI source after the LC column, inducing the ME_{art}. The ME_{art} for a given analyte can be determined by comparing its response with and without artificial matrix infusion. Then, the best-match PCIS for that feature can be selected based on its ability to compensate for the observed ME_{art}. Our hypothesis is that the best-match PCIS selected based on ME_{art} correction should also be effective in compensating for ME_{bio}. This hypothesis depends greatly on how well the infused artificial matrix can mimic the biological matrix to induce ME in the ESI source. Therefore, we selected several compounds according to known ME mechanisms in the ESI source and optimized their concentrations to induce certain ME_{art} for the metabolites examined. This hypothesis was evaluated by comparing the best-match

PCISs selected based on ME_{art} correction with those chosen for ME_{bio} compensation using 19 SIL standards (Table S2).

3.2.1 Selection and concentration optimization of post-column infused artificial matrix compounds

In the ESI source, matrix compounds can disturb the ionization by competing with the analytes for charge in liquid phase and affecting the analytes' ability to remain charged in gas phase.[2] Considering this, four compounds, l-homoarginine hydrochloride (hArgHC), sodium dodecyl sulphate (SDS), sodium acetate (NaOAc), and tridodecylmethylammonium chloride (TDMAC), were chosen as artificial matrix compounds to interrupt the ionization process in the ESI source. These compounds contain salts and/or ionic compounds which can easily form charged ions, competing with analytes for ionization. Additionally, hArg has a high proton affinity in gas-phase;[32] SDS and TDMAC can prevent the coulombic explosion by increasing the droplet's surface tension as surfactants.[33] Based on their ionization properties, hArgHC, NaOAc, and TDMAC were infused in positive mode, while SDS and NaOAc in negative mode.

The concentrations of the artificial matrix compounds were optimized by infusing them individually as well as in a mixture with the injection of pooled plasma, urine and fecal samples. This allowed us to calculate both AME_{art} and RME_{art} with a pooled biological sample, as shown in Equation 3-4. To balance the trade-off between ME_{art} and signal intensity of endogenous metabolites, the optimization aimed to get around 70% AME_{art} and more than 15% RME_{art}. During the optimization process, 19 and 24 endogenous metabolites were evaluated in positive and negative ionization modes, respectively. The average AME_{art} of all the evaluated metabolites at the optimized concentrations (Table 3) are plotted in Figure 3A-B, while the individual AME_{art} are shown in Figure S4. Compared to infusion without artificial matrix (PCIS only), infusing the mixtures successfully induced AME_{art} to 60-70% in plasma, urine and feces for both ionization modes. For individual artificial matrix compounds, hArgHC and SDS at 1µM barely caused signal suppression in positive and negative modes, respectively; NaOAc showed a pronounced ion suppression effect, bringing AME_{art} to 70-75% for both ionization

modes, except for urine in negative mode; TDMAC suppressed the signal of the examined metabolites in plasma and urine, resulting in around 75% AME_{art} in positive mode. Given that diverse ion suppression effects were observed with different artificial matrix combinations, RME_{art} was calculated with all combinations for each ionization mode. As illustrated in Figure 3C-D, the RME_{art} of most metabolites were above 15 % in both ionization modes for all three biological matrices, with several of them exceeding 30%.

Meanwhile, with artificial matrix infusion, the PCISs should experience similar signal suppression to ensure their ability to correct ME_{art}, hence, their infusion profiles were inspected to determine whether their signals were correspondingly suppressed. The PCIS profiles in feces, with and without the mixture, are presented in Figure S5-6 as examples. The artificial matrix successfully suppressed the signal of all PCISs in positive mode, except for $[Glu\text{-}d_7 + Na]^+$. Its signal was largely enhanced by the artificial matrix infusion, which was likely due to the high sodium content in NaOAc (Figure S5A). This resulted in a much lower signal for $[Glu\text{-}d7 + NH_4]^+$, especially in regions with severe ion suppression (Figure S5B). Due to the distorted adduct distributions caused by artificial matrix, $Glu\text{-}d_7$ was not considered as a suitable PCIS for ME correction in positive mode in our study. In negative mode, the artificial matrix also suppressed the signal of all PCISs, except for $[F\text{-}Cor+FA\text{-}H]^-$ and $[Glu\text{-}d7 +FA\text{-}H]^-$, which had comparable intensity with and without artificial matrix infusion over 1-5 mins (Figure S6). These results proved that, at the optimized concentration, infusing the mixture of artificial matrix compounds could successfully induce ion suppression for both metabolites and PCISs.

Table 3. Information and optimized infusion concentrations of artificial matrix compounds

Artificial matrix compound		**Infusion concentration (µM)**	
full name	abbreviation	positive	negative
L-homoarginine hydrochloride	hArgHC	1.0	/
sodium dodecyl sulphate	SDS	/	1.0

sodium acetate	NaOAc	500	375.0
tridodecylmethylammonium chloride	TDMAC	37.5	/

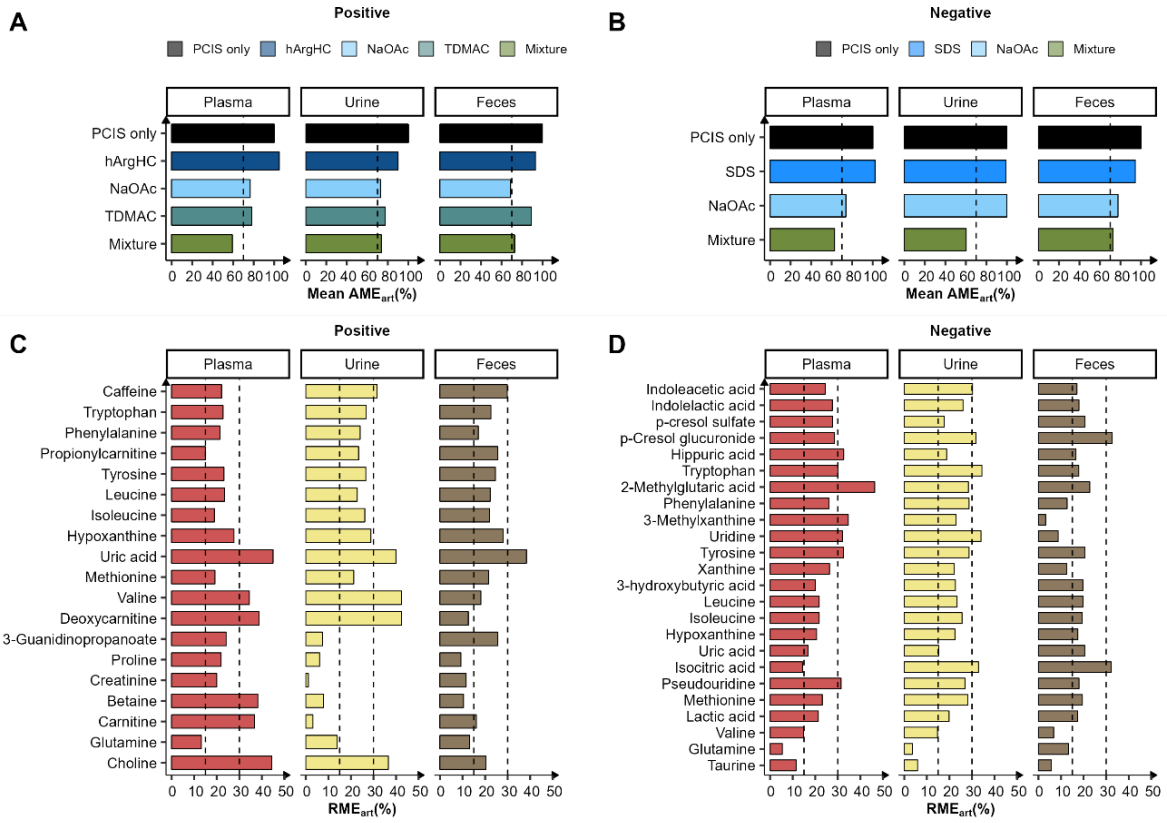


Figure 3. The artificial matrix induced AME_{art} (A, B) and RME_{art} (C, D) for metabolites examined in plasma, urine, and feces. The mean AME_{art} was calculated by averaging the AME_{art} for the examined metabolites in each ionization mode with duplicates, and PCIS only was used as reference with AME_{art} = 100%. The dashed lines in (A) and (B) indicate 70% AME_{art}. The RME_{art} for each metabolite is presented as the RSD % of the AME_{art} among all infusion combinations in positive (C) and negative (D) modes, including PCIS only. The dashed lines in (C) and (D) indicate 15% and 30 % RME_{art}.

3.2.2 PCIS matching: ME_{art} correction vs. ME_{bio} correction

With the optimized artificial matrix concentration, we compared the best-match PCISs selected based on their ability to correct the ME_{art} or the ME_{bio} for 19 SIL standards. These standards were widely distributed in class and physical properties, representing

diverse endogenous metabolites. The AME and RME of the SIL standards were calculated after being spiked into plasma, urine, and fecal samples, as described in 2.5.

To combine both AME and RME for PCIS selection, they were scored as described in Table 1. The final ME scores were calculated by average the AME and RME scores. AME_{bio} and RME_{bio} were calculated at two concentration levels (Table S2), and the averaged areas of two levels were used for ME_{bio} scoring. AME_{art} and RME_{art} were calculated at one concentration level (Table S2). An AME within 80-120% and a RME ≤ 30% are commonly accepted in untargeted analysis,[34] which results in an AME score ≥ 80 and a RME score ≥ 70. Therefore, a PCIS is considered suitable for correcting ME_{bio} or ME_{art} for a SIL standard if it returns a ME score ≥ 75 after correction.

Figure S7 shows the ME_{bio} and ME_{art} scores for 19 SIL standards spiked in plasma, urine, and feces before and after PCIS correction. To identify the suitable PCISs for ME correction, the PCISs were filtered with an ME score ≥ 75. Figure 4 presents the ME scores before and after correction with the filtered PCISs. Before PCIS correction, early-eluting (retention time < 1 min) SIL standards (L-ornithine-d_6, L-glutamine-d_5, TMAO-d_9, L-carnitine-d_3, N-methy-d_3-L-histidine, and betaine-d_9) in three biological matrices, hippuric acid-d_5 in urine, and daidzein-d_6 in feces suffered from more severe ME_{bio}, with scores < 75 (Figure 4A). In contrast to ME_{bio}, most of the early eluting SIL standards had ME_{art} above 75, while more later eluters got ME_{art} scores < 75 before PCIS correction (Figure 4B). It is likely that with artificial matrix infusion, the biological matrix remained as the major source of ionization competition in the early elution region, where the ME_{art} was masked by severe ME_{bio}. Some known endogenous ion suppressors, such as inorganic electrolytes, salts, and highly polar compounds, are poorly retained on the RP column, leading to pronounced ion suppression that overwhelms the influence of artificial matrix in the early elution region. This was reversed in the late elution regions where the artificial matrix had a greater impact on ionization than the biological matrix. The filtered PCISs improved or maintained the ME scores for the SIL standards with initial scores ≥ 75 and successfully compensated for the ME for most of the SIL standards with ME_{bio} or ME_{art} scores ≤ 75 before PCIS correction (Figure 4). For those SIL standards had no PCISs to improve their ME scores

to 75, mainly for the ME_{bio} correction of early-eluting ones in urine and feces, most of them still obtained improved scores after PCIS correction (Figure S7A).

Considering that more than one PCISs managed to correct either ME_{bio} or ME_{art} for most SIL standards (Figure 4), to obtain an overview of which PCIS was appropriate for compensating ME regardless of matrix types, we summed the ME_{bio} or ME_{art} scores of the filtered PCISs for individual SIL standards across plasma, urine, and feces. With 75 as the acceptable ME score, a PCIS score sum between 75 and 100 indicated its capability to correct ME in one biological matrix; between 150 and 225 indicated effective correction in two biological matrices; a score sum above 225 indicated correction in all three biological matrices. Then, the matrix-independent PCIS can be identified by selecting the one with the highest score sum for each SIL standard.

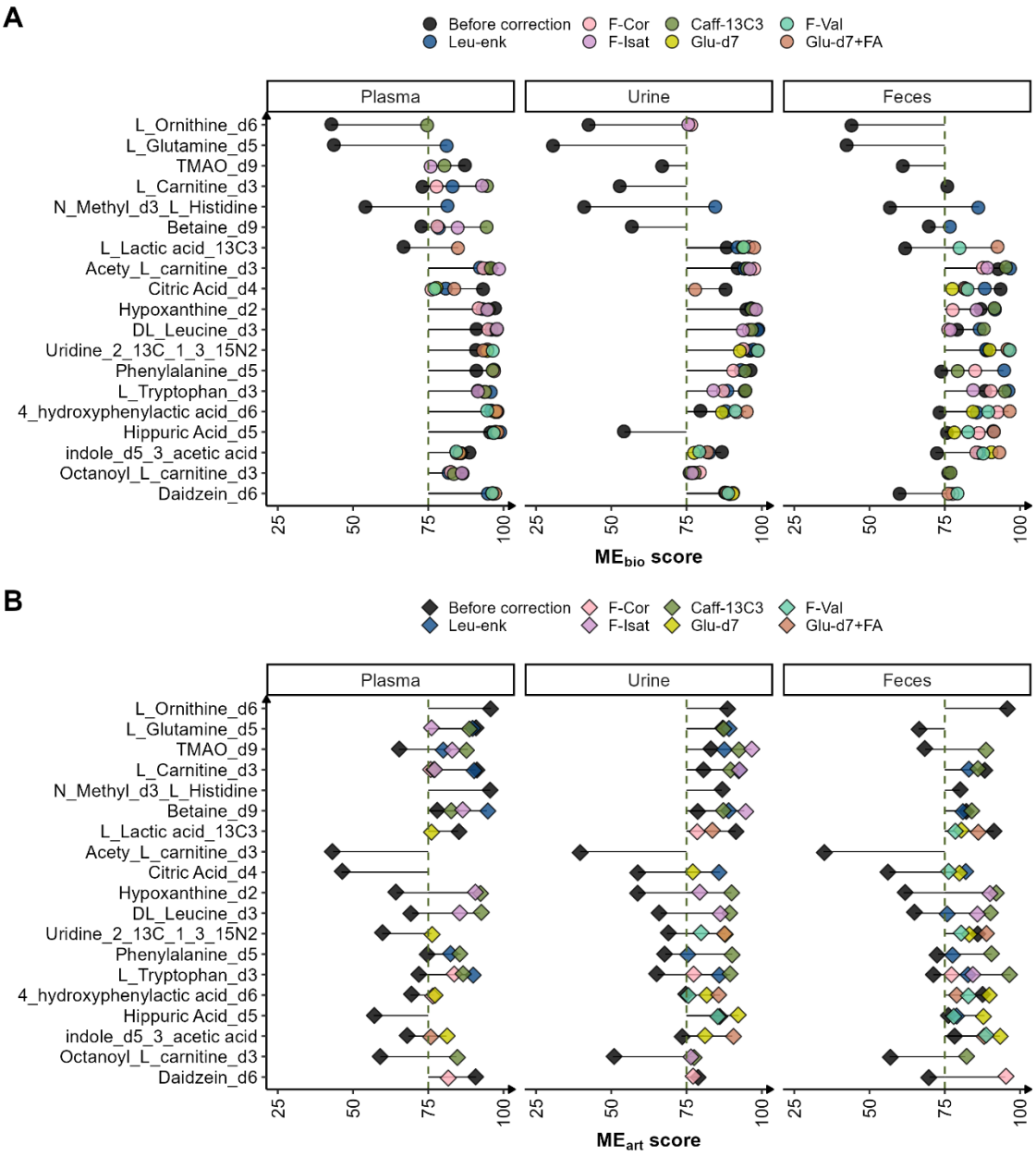


Figure 4. ME_{bio} (A) and ME_{art} (B) scores of the SIL standards before correction (dots and diamonds in black) and after correction with suitable PCISs (dots and diamonds with colors) in plasma, urine, and feces. The SIL standards are plotted in increasing order of retention times from top to bottom and the dashed lines indicate a score of 75, and triplicates were used for ME_{bio} score and ME_{art} score calculation.

The score sums of the filtered PCISs for the SIL standards are presented in Figure 5A. In total, 13 and 12 out of 19 SIL standards had at least one PCIS with an ME_{bio} score sum and ME_{art} score sum ≥ 225, respectively. More PCISs returned an ME_{art} score sum

≥ 150 compared to ME_{bio} score sum for the early-eluting SIL standards. In contrast, for the SIL standards eluting after one minute, all the filtered PCISs returned ME_{bio} score sums ≥ 150, except for lactic acid-$^{13}C_3$. More PCISs achieved an ME_{bio} score sum ≥ 225 for those SIL standards compared to ME_{art} score sum. Given that the early-eluting SIL standards experienced more ME_{bio} than ME_{art}, while the later-eluting ones were more affected by ME_{art} than ME_{bio} (Figure 4), these results suggest that correcting a severe ME is likely to facilitate the selection of the matrix-independent PCIS for a SIL standard. This is also evidenced by comparing the filtered PCIS for ME_{bio} correction in plasma, urine, and feces (Figure 4A). For instance, lactic acid-$^{13}C_3$ experienced more ME_{bio} in plasma and feces. Compared to five PCIS suitable for ME_{bio} correction in urine, only one and two were suitable for the correction in plasma and feces, respectively. Similarly, hippuric acid-d_5 experienced severe ME_{bio} only in urine, with no PCIS suitable for correction, whereas multiple PCISs corrected its ME_{bio} in plasma and feces. Additionally, since daidzein-d_6 had an ME_{bio} score < 75 only in feces before correction, three PCISs were ideal for correcting its ME_{bio} in feces, while all PCISs were suitable for the correction in urine and plasma.

Therefore, we assumed that for the SIL standards experiencing more severe ME_{art} than ME_{bio}, the PCIS selected based on ME_{art} compensation would also be effective in correcting their ME_{bio}. To evaluate this assumption, we compared the best-match PCIS identified by the highest ME_{bio} score sum to those selected by the highest ME_{art} score sum. Three SIL standards (L-ornithine-d_6, N-methyl-d_3-L-histidine, acety-L-carnitine-d_3) did not have a suitable PCIS with an ME_{art} score ≥ 75 in any of the examined biological matrices. Their ME_{art}-based best-match PCISs were still selected based on the highest ME_{art} score sum to include them in the comparison. Figure 5B presents the selected best-match PCISs for all SIL standards according to the highest ME_{bio} score sum or the highest ME_{art} score sum. Ten SIL standards (connected by solid line) obtained identical best-match PCISs based on the selection of ME_{bio} and ME_{art} score sums. Seven standards (connected by dashed line) had different best-match PCISs. However, their PCISs selected based on the highest ME_{art} score sums returned comparable ME_{bio} score sums to those chosen with the highest ME_{bio} score sum, making them equally suitable for ME_{bio} correction. Two standards (unconnected), L-ornithine-

d6 and citric acid-d4, exhibited different best-match PCISs. Consequently, 17 out of 19 SIL standards (89%) showed consistency in PCIS selection based on ME_{bio} and ME_{art} score sums, including some SIL standards experiencing severe ME_{bio} than ME_{art}, such as most of the early eluters and L-lactic acid-$^{13}C_3$. This suggests that, although ME_{art} may be less effective in identifying the matrix-independent PCISs for the SIL standards experiencing a more severe ME_{bio} than ME_{art}, utilizing the ME_{art} score sum across diverse matrices may enhance the likelihood of making suitable selections.

These results demonstrate that ME_{art} compensation obtained comparable PCIS selections to the ME_{bio} correction for the examined SIL standards across diverse matrices. The ME_{bio} correction, namely matching PCISs to analytes by assessing their ability to correct for ME quantified with spiked SIL standards, has been commonly applied in targeted metabolomics studies[13,14,23] However, this approach is impractical for untargeted metabolomics due to the reliance on SIL standards. Another strategy for PCIS selection is to evaluate the improvement in linearity and precision across matrix dilution series before and after PCIS correction.[15,18,29] Although this method is applicable for untargeted analysis as it does not require authentic standards spiking, it can be problematic for metabolites with rather high or low endogenous abundance due to potential solubility and detection limit issues.[29] Therefore, the reliability of ME_{art} compensation in PCIS selection is supported not only by its consistency with the ME_{bio} correction, but also by mitigating the risk of the analyte signals falling beyond their limits of detection/quantification. More importantly, since the ME_{art} can be determined for any detected feature, the ME_{art} compensation represents an ideal approach for PCIS matching in untargeted metabolomics.

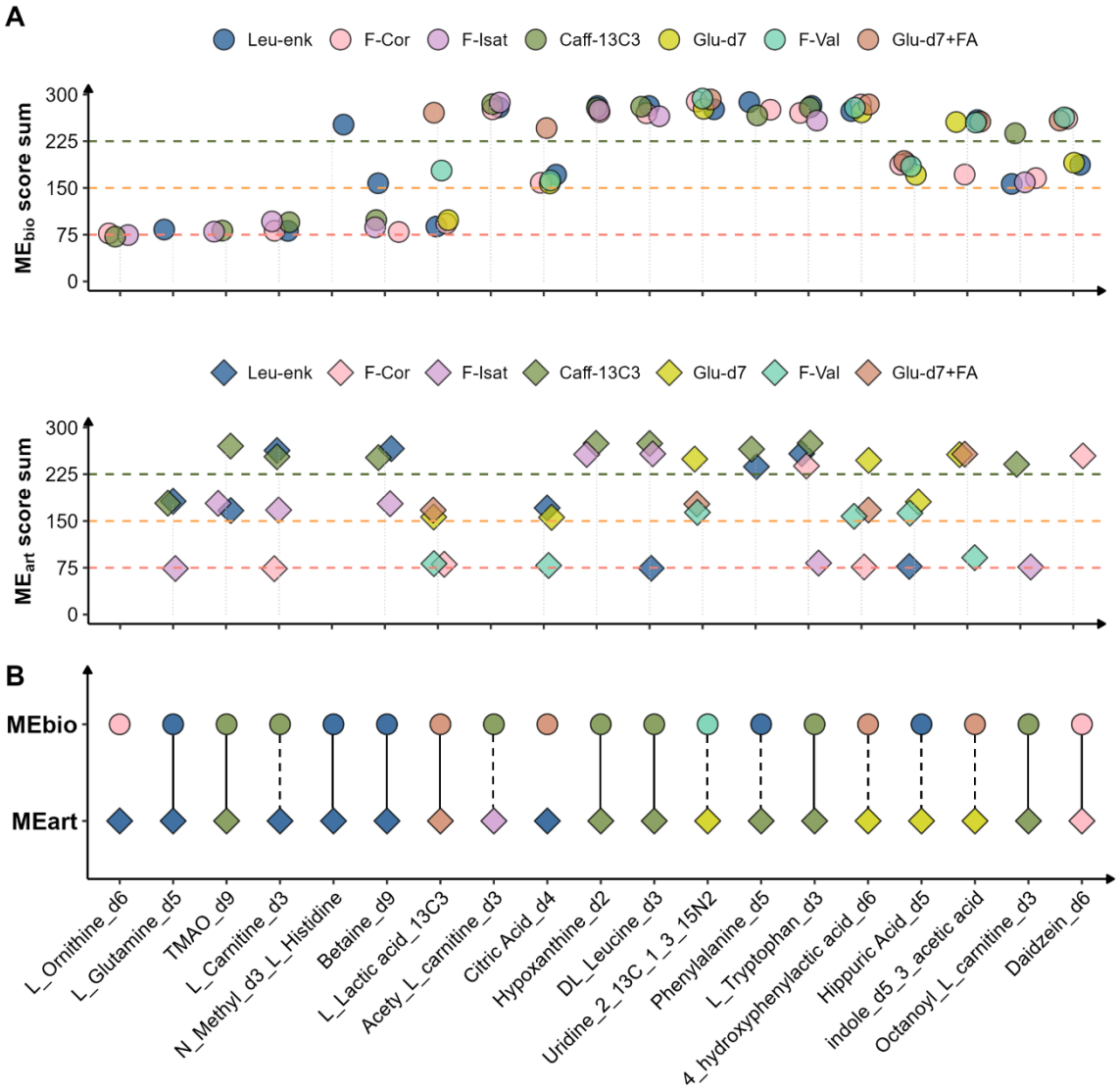


Figure 5. The ME_{bio} and ME_{art} score sum of all the SIL standards across plasma, urine, and feces for the PCISs that returned scores ≥ 75 in at least one biological matrix (A). The selected PCISs according to the highest ME_{bio} and ME_{art} score sums across plasma, urine, and feces for the SIL standards (B), where the solid line connection indicates identical PCIS selection, while the dashed line connection indicates PCIS selection with comparable score sums, and no connection indicates different PCIS selection. The SIL standards are plotted in increasing order of retention times from left to right.

3.3 ME_{bio} correction with PCIS selected by ME_{art}

To assess the effectiveness of ME_{art}-selected PCIS in ME_{bio} correction, we applied the PCIS selected with the highest ME_{art} score sum for ME_{bio} correction of the 19 SIL

standards spiked in plasma, urine, and feces. The ME_{bio} scores of all the SIL standards before and after PCIS correction is plotted in Figure S8. The selected PCIS improved or maintained the ME_{bio} score for 19 (100%), 16 (84%), and 18 (95%) SIL standards spiked in plasma, urine, and feces, respectively.

To illustrate the improvement in ME after PCIS correction, the AME_{bio} and RME_{bio} values of the 19 SIL standards were compared before and after correction in each biological matrix. As presented in Figure 6, the dots represent the AME_{bio} value (left y axis), whereas the bars indicate the RME_{bio} value (right y axis). In plasma (Figure 6A), seven SIL standards, namely the six early eluters and lactic acid-$^{13}C_3$, experienced ion suppression with AME_{bio} < 80% before correction. The PCIS improved the AME_{bio} towards 80 -120% for five of them, bringing the AME_{bio} of N-methyl-d_3-L-histidine and lactic acid-$^{13}C_3$ within the acceptable range. The RME_{bio} of all SIL standards were below 30% after PCIS correction, with significant improvements for L-ornithine-d_6, L-glutamine-d_5, and lactic acid-$^{13}C_3$. In urine (Figure 6B), nine SIL standards, including the early eluters, 4-hydroxyphenylactic acid-d_6, hippuric acid-d_5, and octanoyl-L-carnitine-d_3, had AME_{bio} outside 80 -120%. After PCIS correction, five standards showed improved AME_{bio}, with N-methyl-d_3-L-histindine and 4-hydroxyphenylactic acid-d_6 reaching the range of 80-120%. The RME_{bio} of all SIL standards were within 30% after PCIS correction in urine, except for hippuric acid-d_5. Significant improvements were noted for L-glutamine-d_5, L-carnitine-d_3, and betanine-d_3, which had RME_{bio} greater than 30% before correction. In feces (Figure 6C), 13 SIL standards suffered from ion suppression (AME_{bio} < 80%) or enhancement (AME_{bio} > 120%) before PCIS correction. After correction, 10 of them showed improved AME_{bio} approaching the 80-120% range, with five within the range. Four SIL standards exhibited RME_{bio} close to or exceeding 30% before correction; the PCIS successfully reduced the RME_{bio} of lactic acid-$^{13}C_3$ and daidzein-d_6 to below 15%.

Over-corrected AME_{bio} or increased RME_{bio} were observed for TMAO-d_9, L-carnitine-d_3, and betaine-d_9, in all biological matrices. This seems to be caused by the permanent positive charge at the quaternary ammonium group in these early eluters, which makes them less susceptible to ion suppression than the PCISs in the region with severe

suppression. Considering that their PCIS selections were consistent based on the ME_{bio} and ME_{art} score sums (Figure 5B), another PCIS candidate with permanent charge may need to be included for ME correction of those standards. In addition, the selected PCIS failed to maintain or improve the AME_{bio} or/and RME_{bio} for citric acid-d_4, hippuric acid-d_5, and indole-d_5-acetic acid in urine. These SIL standards showed maintained or improved ME in the other two biological matrices, except for citric acid-d_4 in plasma. This inefficient correction is likely due to specific co-eluting matrix compounds present in urine, suggesting that a more acidic PCIS may be needed to mimic the ionization of those SIL standards for effective ME correction in urine. Overall, compared with no correction, PCISs selected by ME_{art} improved the ME_{bio} for most SILs affected by ME in the examined matrices, and maintained the ME_{bio} for those with acceptable ME prior to correction, demonstrating the reliability of ME_{art}-selected PCIS for ME_{bio} compensation.

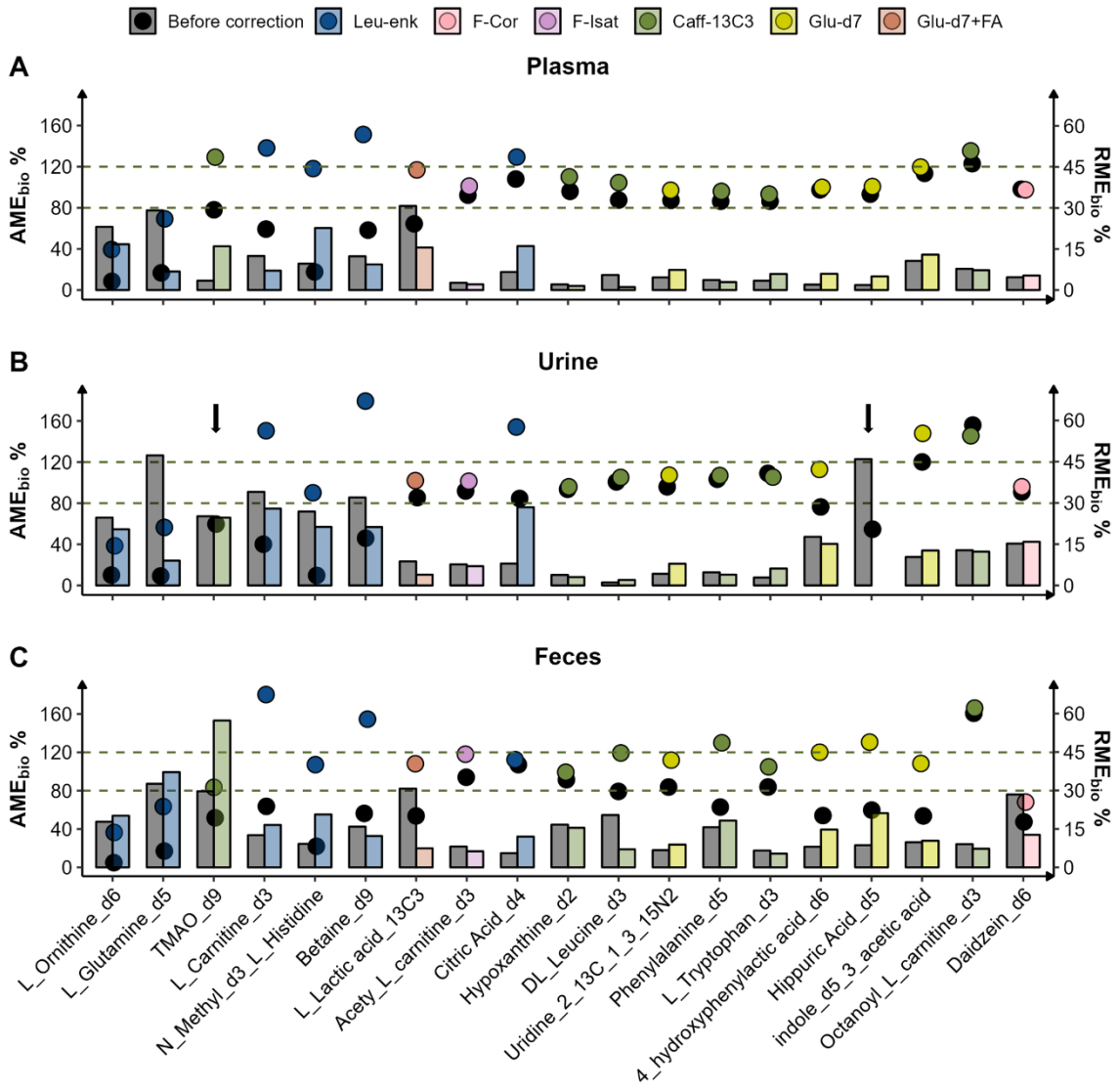


Figure 6. AME_{bio} (dots, left y axis) and RME_{bio} (bars, right y axis) of the 19 SIL standards spiked in plasma, urine, and feces before and after correction with the PCISs selected by the highest ME_{art} score sum. The SIL standards are plotted in increasing order of retention time from left to right. The dashed lines indicated 80-120% of AME_{bio}. The black arrows in (B) indicate the AME_{bio} higher than 160% and/or RME_{bio} larger than 60%.

4. Conclusion

In this study, we presented a strategy in an LC-PCIS-MS method for selecting suitable PCISs to compensate for ME in untargeted metabolomics. This is achieved by comparing the PCISs' ability to correct for the ME_{art} created through post-column

infusion of compounds that affect the ionization in the ESI source. A ME score system was introduced to incorporate AME and RME into the selection. To give equal importance to AME and RME, their average score was used as the final ME score in our study. Different weights can be assigned to AME and RME if one is considered more important than the other in a particular study.

The feasibility of ME_{art} compensation in identifying suitable PCIS was evaluated using 19 SIL standards spiked in plasma, urine, and feces. This evaluation was conducted by comparing the PCISs selected based on the ME_{art} and ME_{bio} compensation across the three matrices. As a result, 89% of the SIL standards showed consistent PCIS selection between ME_{bio} and ME_{art}, demonstrating the effectiveness of ME_{art} compensation for PCIS selection. Subsequently, we applied the ME_{art}-selected PCISs to correct for the ME_{bio} for the SIL standards, resulting in improved ME_{bio} for most of the SIL standards experiencing ME and maintaining ME_{bio} for those with acceptable ME before correction.

In conclusion, we demonstrate the concept of applying ME_{art} creation and compensation for PCIS matching in an LC-PCIS-MS method to correct for ME across diverse biological matrices. Importantly, this strategy is independent of retention time and standards spiking, making it universally applicable for any detectable feature in untargeted metabolomics. Ideally, based on the ME_{art} -selected PCISs, a feature-PCIS-matched library could be developed. Depending on the purpose of the study, such a library could be constructed with diverse or specific matrices and applied for ME correction in future studies. Overall, our study has proposed a novel approach to compensate for the ME in untargeted metabolomics with PCIS, which contributes to improving data reliability and comparability for untargeted metabolomic studies across varied matrices.

Acknowledgements

Pingping Zhu Would like to acknowledge the China Scholarship Council (CSC, No. 201906240049). This research was part of the Netherlands X-omics Initiative and partially founded by NWO project 184.034.019. This publication is part of the 'Building the infrastructure for Exposome research: Exposome-Scan' project (with project number 175.2019.032) of the program ‘Investment Grant NWO Large', which is funded by the Dutch Research Council (NWO).

Reference:

(1) González, O.; Blanco, M. E.; Iriarte, G.; Bartolomé, L.; Maguregui, M. I.; Alonso, R. M. Bioanalytical Chromatographic Method Validation According to Current Regulations, with a Special Focus on the Non-Well Defined Parameters Limit of Quantification, Robustness and Matrix Effect. *Journal of Chromatography A* **2014**, *1353*, 10–27. https://doi.org/10.1016/j.chroma.2014.03.077.

(2) Panuwet, P.; Hunter Jr., R. E.; D’Souza, P. E.; Chen, X.; Radford, S. A.; Cohen, J. R.; Marder, M. E.; Kartavenka, K.; Ryan, P. B.; Barr, D. B. Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring. *Critical Reviews in Analytical Chemistry* **2016**, *46* (2), 93–105. https://doi.org/10.1080/10408347.2014.980775.

(3) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC−MS/MS. *Anal. Chem.* **2003**, *75* (13), 3019–3030. https://doi.org/10.1021/ac020361s.

(4) Cortese, M.; Gigliobianco, M. R.; Magnoni, F.; Censi, R.; Di Martino, P. Compensate for or Minimize Matrix Effects? Strategies for Overcoming Matrix Effects in Liquid Chromatography-Mass Spectrometry Technique: A Tutorial Review. *Molecules* **2020**, *25* (13), 3047. https://doi.org/10.3390/molecules25133047.

(5) Thakare, R.; Chhonker, Y. S.; Gautam, N.; Alamoudi, J. A.; Alnouti, Y. Quantitative Analysis of Endogenous Compounds. *Journal of Pharmaceutical and Biomedical Analysis* **2016**, *128*, 426–437. https://doi.org/10.1016/j.jpba.2016.06.017.

(6) Wang, S.; Cyronak, M.; Yang, E. Does a Stable Isotopically Labeled Internal Standard Always Correct Analyte Response?: A Matrix Effect Study on a LC/MS/MS Method for the Determination of Carvedilol Enantiomers in Human Plasma. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43* (2), 701–707. https://doi.org/10.1016/j.jpba.2006.08.010.

(7) Berg, T.; Strand, D. H. 13C Labelled Internal Standards—A Solution to Minimize Ion Suppression Effects in Liquid Chromatography–Tandem Mass Spectrometry Analyses of Drugs in Biological Samples? *Journal of Chromatography A* **2011**, *1218* (52), 9366–9374. https://doi.org/10.1016/j.chroma.2011.10.081.

(8) Bonfiglio, R.; King, R. C.; Olah, T. V.; Merkle, K. The Effects of Sample Preparation Methods on the Variability of the Electrospray Ionization Response for Model Drug Compounds. *Rapid Communications in Mass Spectrometry* **1999**, *13* (12), 1175–1185. https://doi.org/10.1002/(SICI)1097-0231(19990630)13:12%253C1175::AID-RCM639%253E3.0.CO;2-0.

(9) Choi, B. K.; Gusev, A. I.; Hercules, D. M. Postcolumn Introduction of an Internal Standard for Quantitative LC−MS Analysis. *Anal. Chem.* **1999**, *71* (18), 4107–4110. https://doi.org/10.1021/ac990312o.

(10) Zhao, X.; Metcalfe, C. D. Characterizing and Compensating for Matrix Effects Using Atmospheric Pressure Chemical Ionization Liquid Chromatography−Tandem Mass Spectrometry: Analysis of Neutral Pharmaceuticals in Municipal Wastewater. *Anal. Chem.* **2008**, *80* (6), 2010–2017. https://doi.org/10.1021/ac701633m.

(11) Rossmann, J.; Gurke, R.; Renner, L. D.; Oertel, R.; Kirch, W. Evaluation of the Matrix Effect of Different Sample Matrices for 33 Pharmaceuticals by Post-Column Infusion. *Journal of Chromatography B* **2015**, *1000*, 84–94. https://doi.org/10.1016/j.jchromb.2015.06.019.

(12) Chang, K. C.; Su, J. J.; Cheng, C. Development of Online Sampling and Matrix Reduction Technique Coupled Liquid Chromatography/Ion Trap Mass Spectrometry for Determination Maduramicin in Chicken Meat. *Food Chemistry* **2013**, *141* (2), 1522–1529. https://doi.org/10.1016/j.foodchem.2013.04.016.

(13) Liao, H.-W.; Chen, G.-Y.; Tsai, I.-L.; Kuo, C.-H. Using a Postcolumn-Infused Internal Standard for Correcting the Matrix Effects of Urine Specimens in Liquid Chromatography–Electrospray Ionization Mass

Spectrometry. *Journal of Chromatography A* **2014**, *1327*, 97–104. https://doi.org/10.1016/j.chroma.2013.12.066.
(14) González, O.; Van Vliet, M.; Damen, C. W. N.; Van Der Kloet, F. M.; Vreeken, R. J.; Hankemeier, T. Matrix Effect Compensation in Small-Molecule Profiling for an LC–TOF Platform Using Multicomponent Postcolumn Infusion. *Anal. Chem.* **2015**, *87* (12), 5921–5929. https://doi.org/10.1021/ac504268y.
(15) Rossmann, J.; Renner, L. D.; Oertel, R.; El-Armouche, A. Post-Column Infusion of Internal Standard Quantification for Liquid Chromatography-Electrospray Ionization-Tandem Mass Spectrometry Analysis – Pharmaceuticals in Urine as Example Approach. *Journal of Chromatography A* **2018**, *1535*, 80–87. https://doi.org/10.1016/j.chroma.2018.01.001.
(16) Chiu, H.-H.; Liao, H.-W.; Shao, Y.-Y.; Lu, Y.-S.; Lin, C.-H.; Tsai, I.-L.; Kuo, C.-H. Development of a General Method for Quantifying IgG-Based Therapeutic Monoclonal Antibodies in Human Plasma Using Protein G Purification Coupled with a Two Internal Standard Calibration Strategy Using LC-MS/MS. *Analytica Chimica Acta* **2018**, *1019*, 93–102. https://doi.org/10.1016/j.aca.2018.02.040.
(17) Liao, H.-W.; Lin, S.-W.; Chen, G.-Y.; Kuo, C.-H. Estimation and Correction of the Blood Volume Variations of Dried Blood Spots Using a Postcolumn Infused-Internal Standard Strategy with LC-Electrospray Ionization-MS. *Anal. Chem.* **2016**, *88* (12), 6457–6464. https://doi.org/10.1021/acs.analchem.6b01145.
(18) Jhang, R.-S.; Lin, S.-Y.; Peng, Y.-F.; Chao, H.-C.; Tsai, I.-L.; Lin, Y.-T.; Liao, H.-W.; Tang, S.-C.; Kuo, C.-H.; Jeng, J.-S. Using the PCI-IS Method to Simultaneously Estimate Blood Volume and Quantify Nonvitamin K Antagonist Oral Anticoagulant Concentrations in Dried Blood Spots. *Anal. Chem.* **2020**, *92* (3), 2511–2518. https://doi.org/10.1021/acs.analchem.9b04063.
(19) Stahnke, H.; Reemtsma, T.; Alder, L. Compensation of Matrix Effects by Postcolumn Infusion of a Monitor Substance in Multiresidue Analysis with LC−MS/MS. *Anal. Chem.* **2009**, *81* (6), 2185–2192. https://doi.org/10.1021/ac802362s.
(20) Liao, H.-W.; Chen, G.-Y.; Wu, M.-S.; Liao, W.-C.; Tsai, I.-L.; Kuo, C.-H. Quantification of Endogenous Metabolites by the Postcolumn Infused-Internal Standard Method Combined with Matrix Normalization Factor in Liquid Chromatography–Electrospray Ionization Tandem Mass Spectrometry. *Journal of Chromatography A* **2015**, *1375*, 62–68. https://doi.org/10.1016/j.chroma.2014.11.073.
(21) Liao, H.-W.; Chen, G.-Y.; Wu, M.-S.; Liao, W.-C.; Lin, C.-H.; Kuo, C.-H. Development of a Postcolumn Infused-Internal Standard Liquid Chromatography Mass Spectrometry Method for Quantitative Metabolomics Studies. *J. Proteome Res.* **2017**, *16* (2), 1097–1104. https://doi.org/10.1021/acs.jproteome.6b01011.
(22) Huang, M.; Li, H.-Y.; Liao, H.-W.; Lin, C.-H.; Wang, C.-Y.; Kuo, W.-H.; Kuo, C.-H. Using Post-Column Infused Internal Standard Assisted Quantitative Metabolomics for Establishing Prediction Models for Breast Cancer Detection. *Rapid Communications in Mass Spectrometry* **2020**, *34* (S1), e8581. https://doi.org/10.1002/rcm.8581.
(23) Lo, C.; Hsu, Y.-L.; Cheng, C.-N.; Lin, C.-H.; Kuo, H.-C.; Huang, C.-S.; Kuo, C.-H. Investigating the Association of the Biogenic Amine Profile in Urine with Therapeutic Response to Neoadjuvant Chemotherapy in Breast Cancer Patients. *J. Proteome Res.* **2020**, *19* (10), 4061–4070. https://doi.org/10.1021/acs.jproteome.0c00362.
(24) Liao, H.-W.; Kuo, C.-H.; Chao, H.-C.; Chen, G.-Y. Post-Column Infused Internal Standard Assisted Lipidomics Profiling Strategy and Its Application on Phosphatidylcholine Research. *Journal of Pharmaceutical and Biomedical Analysis* **2020**, *178*, 112956. https://doi.org/10.1016/j.jpba.2019.112956.
(25) Chepyala, D.; Kuo, H.-C.; Su, K.-Y.; Liao, H.-W.; Wang, S.-Y.; Chepyala, S. R.; Chang, L.-C.; Kuo, C.-H. Improved Dried Blood Spot-Based Metabolomics Analysis by a Postcolumn Infused-Internal Standard Assisted Liquid Chromatography-Electrospray Ionization Mass Spectrometry Method. *Anal. Chem.* **2019**, *91* (16), 10702–10712. https://doi.org/10.1021/acs.analchem.9b02050.
(26) González, O.; Dubbelman, A.-C.; Hankemeier, T. Postcolumn Infusion as a Quality Control Tool for LC-MS-Based Analysis. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (6), 1077–1080. https://doi.org/10.1021/jasms.2c00022.
(27) Tisler, S.; Pattison, D. I.; Christensen, J. H. Correction of Matrix Effects for Reliable Non-Target Screening LC–ESI–MS Analysis of Wastewater. *Anal. Chem.* **2021**, *93* (24), 8432–8441. https://doi.org/10.1021/acs.analchem.1c00357.
(28) Zhu, P.; Dubbelman, A.-C.; Hunter, C.; Genangeli, M.; Karu, N.; Harms, A.; Hankemeier, T. Development of an Untargeted LC-MS Metabolomics Method with Postcolumn Infusion for Matrix Effect Monitoring in Plasma and Feces. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (3), 590–602. https://doi.org/10.1021/jasms.3c00418.
(29) Dubbelman, A.-C.; van Wieringen, B.; Roman Arias, L.; van Vliet, M.; Vermeulen, R.; Harms, A. C.; Hankemeier, T. Strategies for Using Postcolumn Infusion of Standards to Correct for Matrix Effect in LC-

MS-Based Quantitative Metabolomics. *J. Am. Soc. Mass Spectrom.* **2024**. https://doi.org/10.1021/jasms.4c00408.

(30) González, O.; van Vliet, M.; Damen, C. W. N.; van der Kloet, F. M.; Vreeken, R. J.; Hankemeier, T. Matrix Effect Compensation in Small-Molecule Profiling for an LC–TOF Platform Using Multicomponent Postcolumn Infusion. *Anal. Chem.* **2015**, *87* (12), 5921–5929. https://doi.org/10.1021/ac504268y.

(31) Tisler, S.; Pattison, D. I.; Christensen, J. H. Correction of Matrix Effects for Reliable Non-Target Screening LC–ESI–MS Analysis of Wastewater. *Analytical Chemistry* **2021**, acs.analchem.1c00357. https://doi.org/10.1021/acs.analchem.1c00357.

(32) Rožman, M. Proton Affinity of Several Basic Non-Standard Amino Acids. *Chemical Physics Letters* **2012**, *543*, 50–54. https://doi.org/10.1016/j.cplett.2012.06.048.

(33) Rundlett, K. L.; Armstrong, D. W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis−Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1996**, *68* (19), 3493–3497. https://doi.org/10.1021/ac960472p.

(34) Evans, A. M.; O'Donovan, C.; Playdon, M.; Beecher, C.; Beger, R. D.; Bowden, J. A.; Broadhurst, D.; Clish, C. B.; Dasari, S.; Dunn, W. B.; Griffin, J. L.; Hartung, T.; Hsu, P.-C.; Huan, T.; Jans, J.; Jones, C. M.; Kachman, M.; Kleensang, A.; Lewis, M. R.; Monge, M. E.; Mosley, J. D.; Taylor, E.; Tayyari, F.; Theodoridis, G.; Torta, F.; Ubhi, B. K.; Vuckovic, D.; on behalf of the Metabolomics Quality Assurance, Q. C. C. (mQACC). Dissemination and Analysis of the Quality Assurance (QA) and Quality Control (QC) Practices of LC–MS Based Untargeted Metabolomics Practitioners. *Metabolomics* **2020**, *16* (10), 113. https://doi.org/10.1007/s11306-020-01728-5.

Supplementary Material

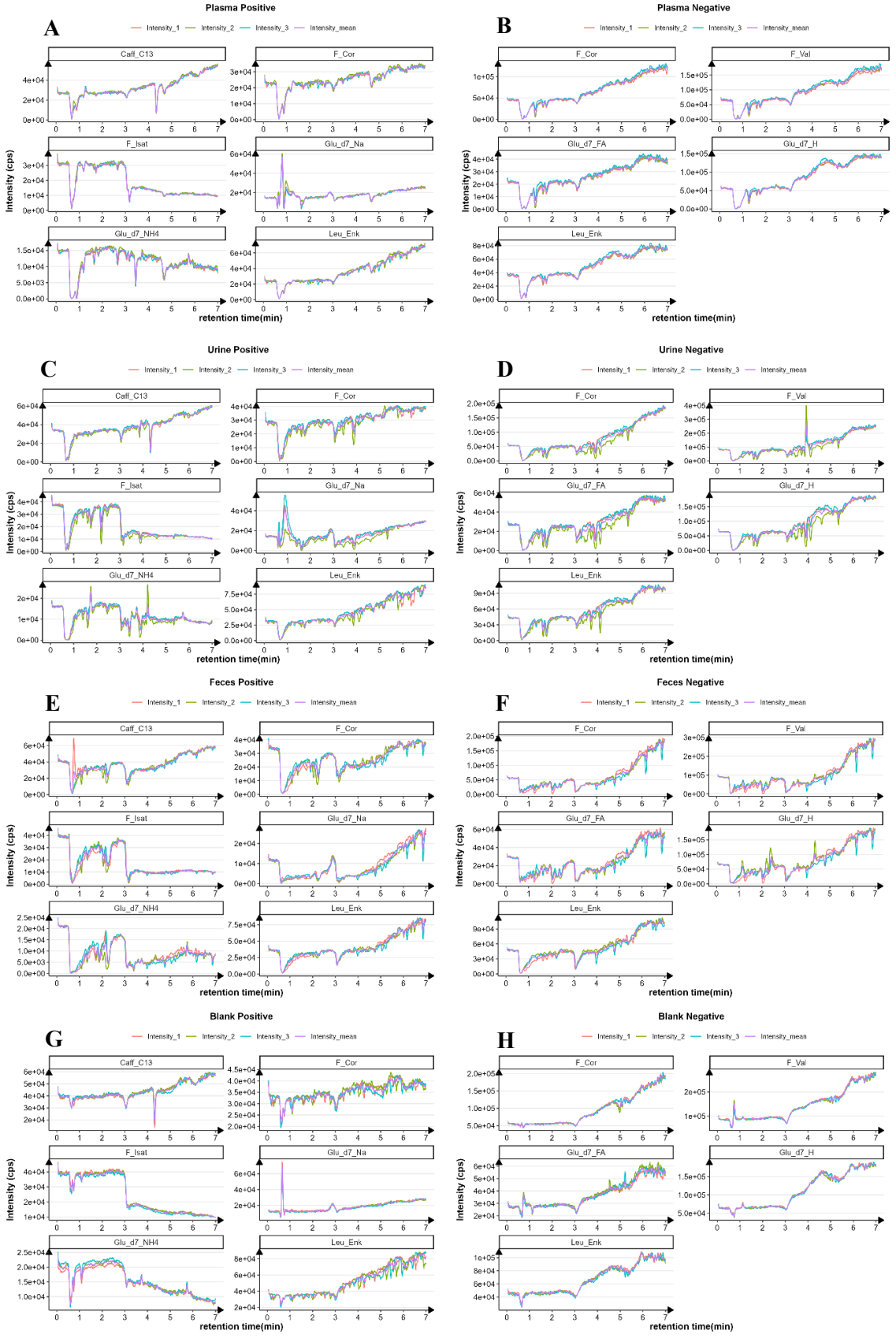


Figure S1. Infusion profiles for 0-7 minutes all PCISs with injections of plasma (A, B), urine (C, D), feces (E,F) and solvent blanks (G, H) in positive and negative modes. Three replicates are plotted for solvent blanks; samples from three different individuals are plotted for plasma, urine, and feces

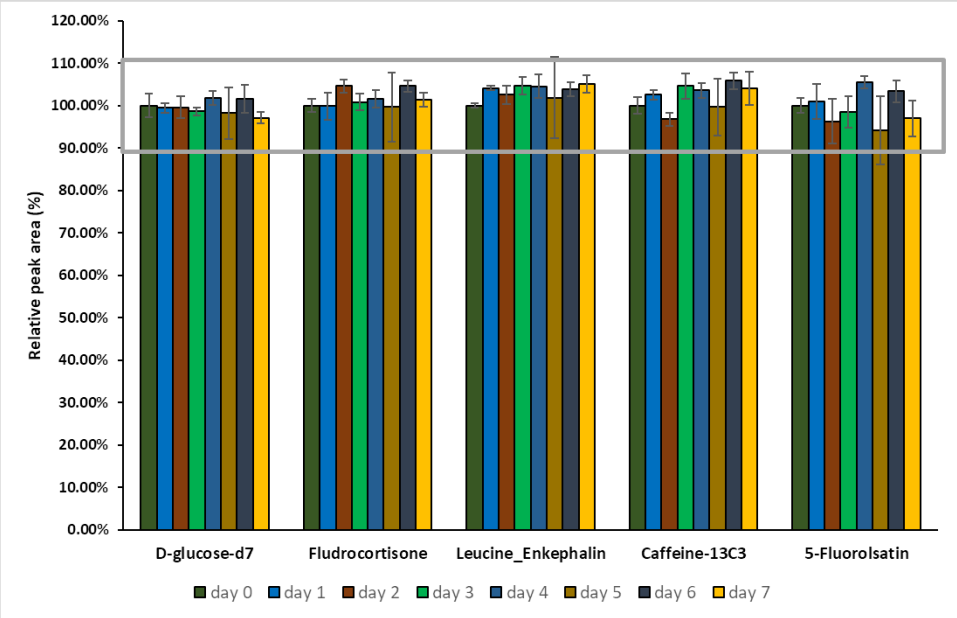


Figure S2. Seven-days room temperature stability test for all the selected PCISs, except for 3-fluoro-DL-valine.

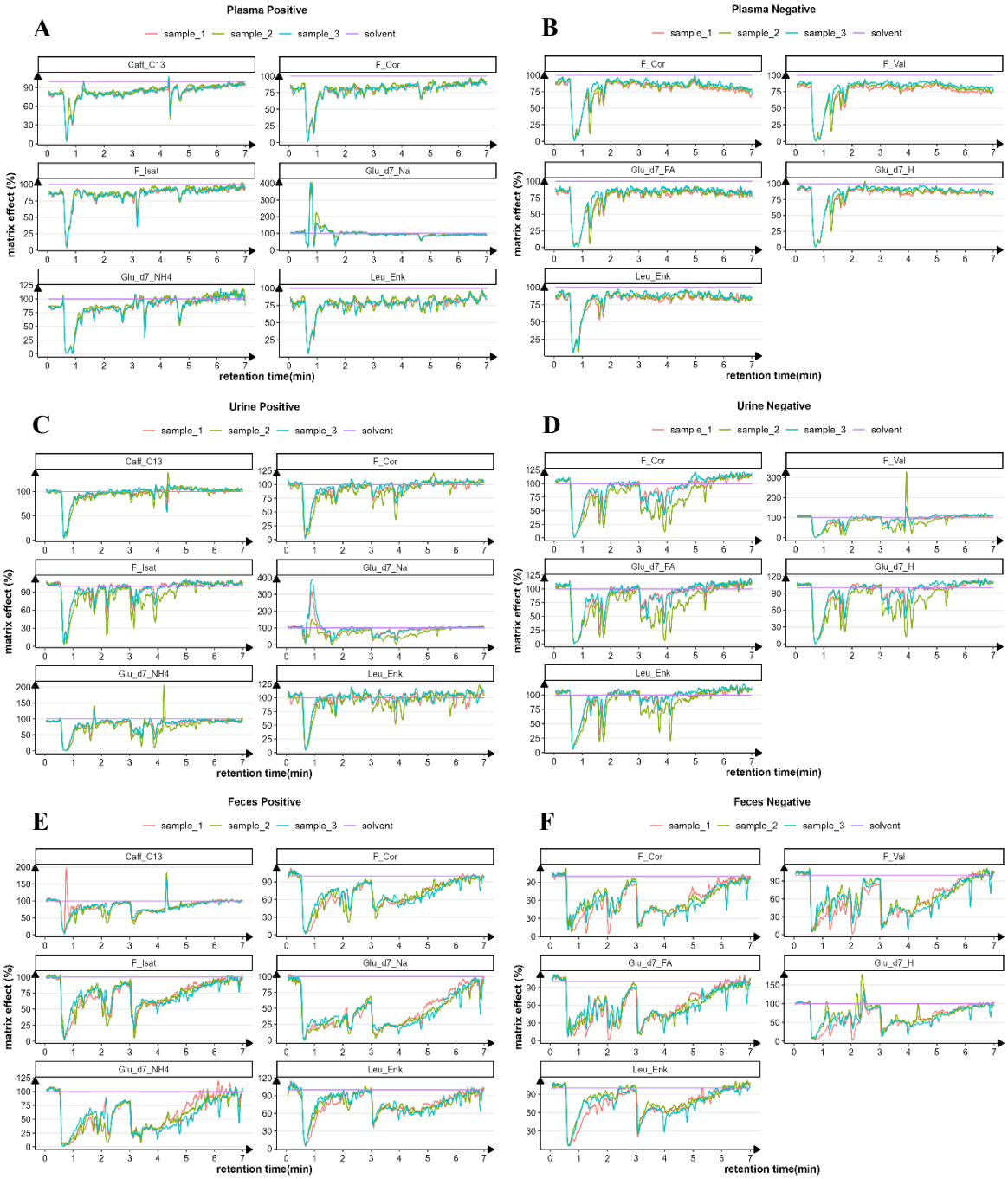


Figure S3. Matrix effect profiles from 0-7 minutes for all PCISs monitored in plasma (A, B), urine (C, D), and fecal samples (E, F) in positive and negative modes. For each ionization mode and PCIS, the averaged intensity of the infusion profile from three solvent samples was used as the reference (100% matrix effect).

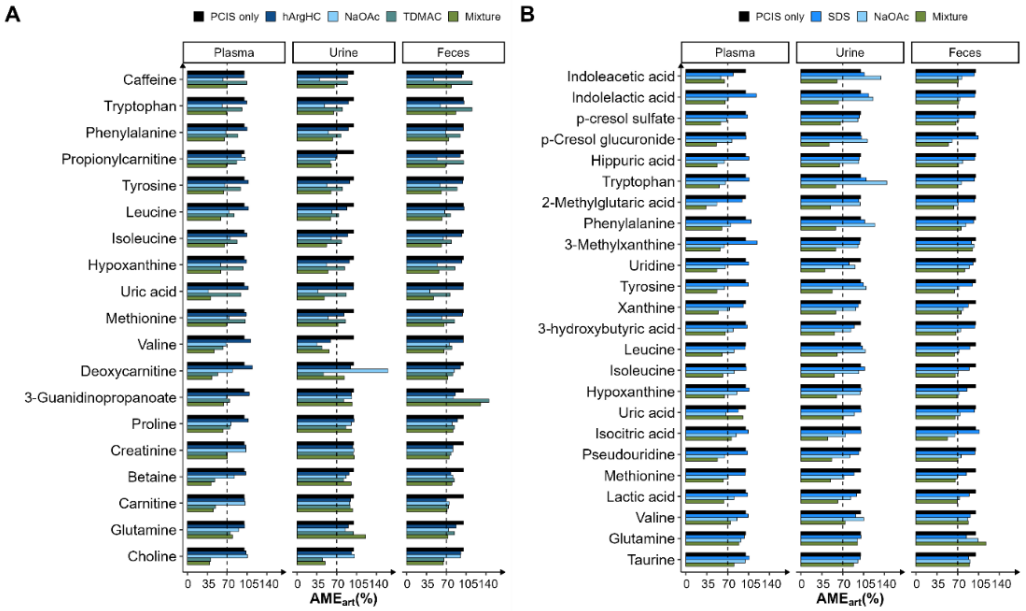


Figure S4. MEart induced by individual artificial matrix compounds and the mixture of all artificial matrix compounds (Mixture) in positive (A) and negative (B) modes for plasma, urine, and feces, across all examined metabolites. "PCIS only" is used as a reference with no induced ME_{art} (MEart = 100%), and the dashed line indicates 70% of ME_{art}.

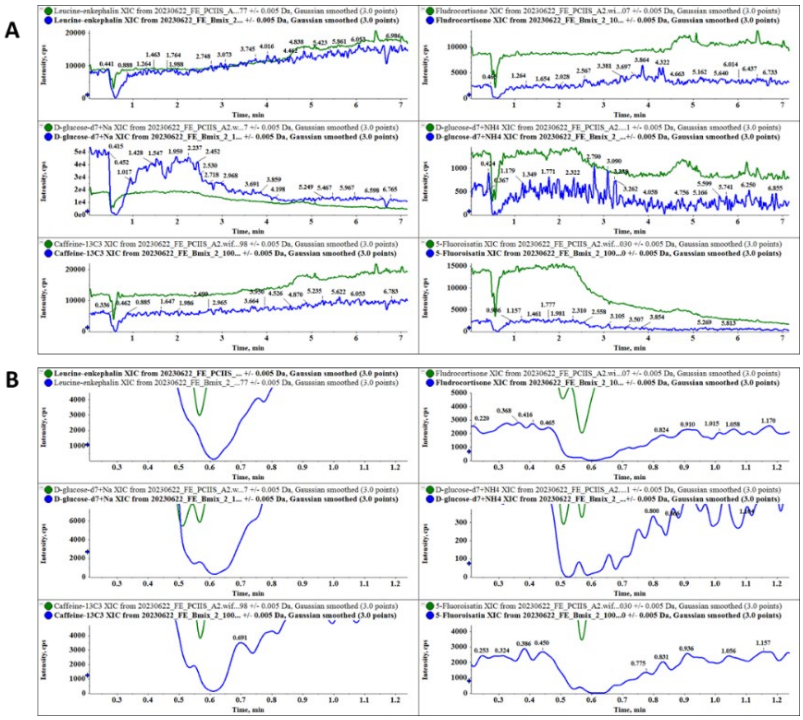


Figure S5. Infusion profiles of PCIS (A) and zoomed-in plots of the region with severe suppression (B) for a pooled fecal sample with infusion of PCIS (green line) and PCIS plus the mixture of artificial matrix compounds (blue line) in positive mode.

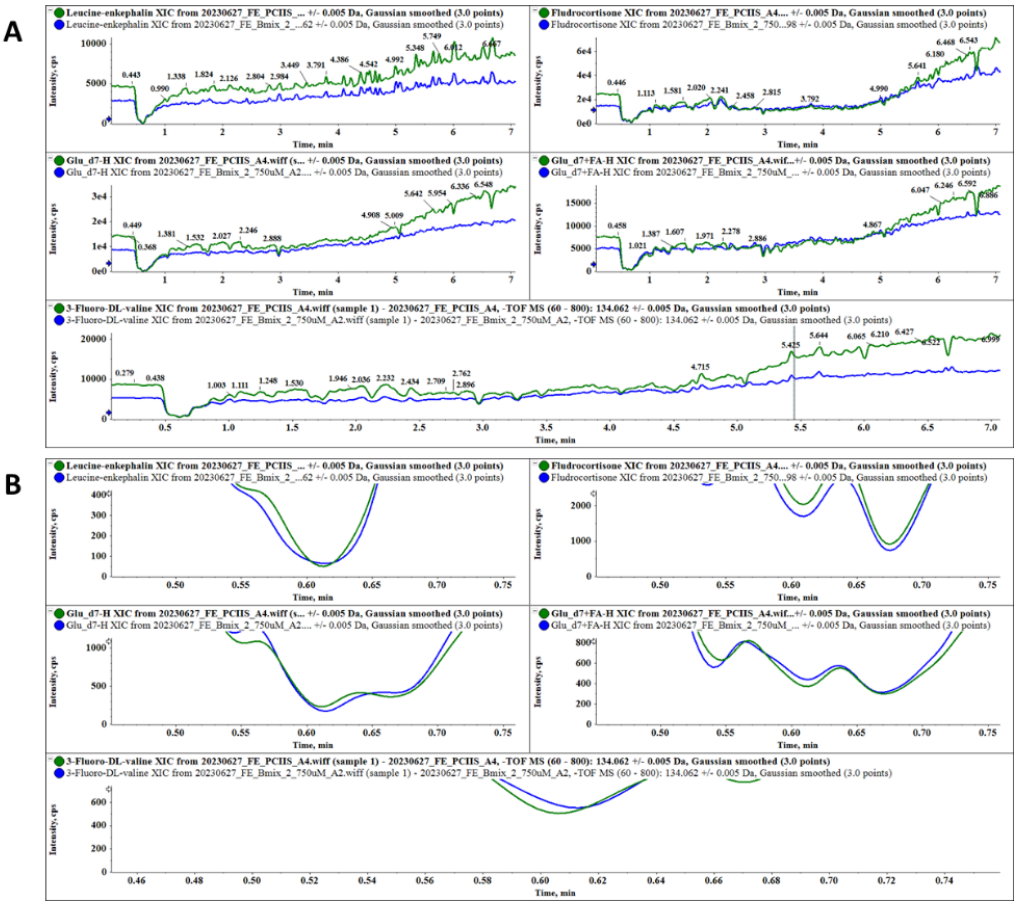


Figure S6. Infusion profiles of PCISs (A) and zoomed-in plots of the region with severe suppression (B) for a pooled fecal sample with infusion of PCIS (green line) and PCIS plus the mixture of artificial matrix compounds (blue line) in negative mode.

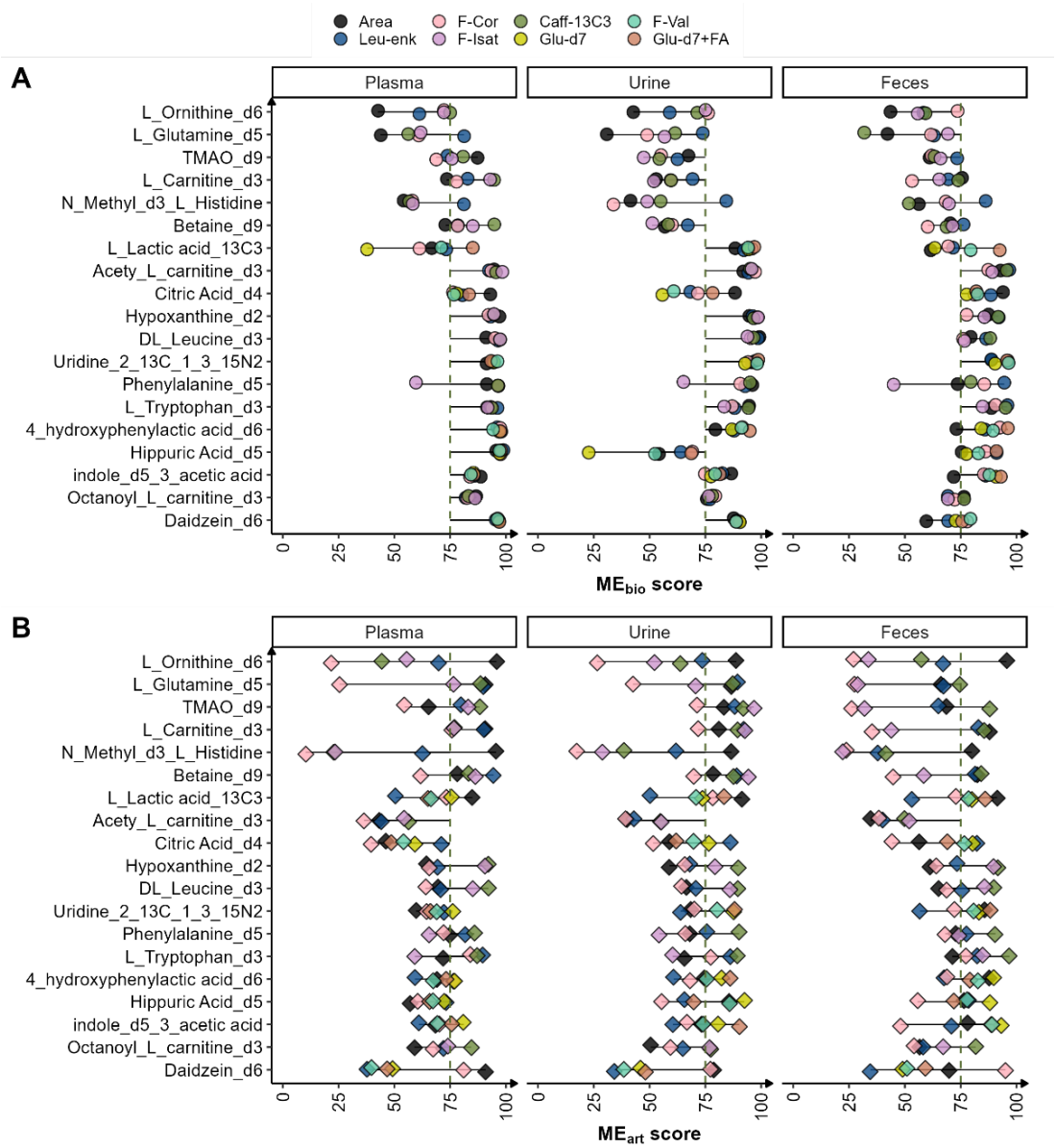


Figure S7. The ME_{bio} (A) and ME_{art} (B) scores of 19 SIL standards spiked in plasma, urine and feces before and after PCIS correction.

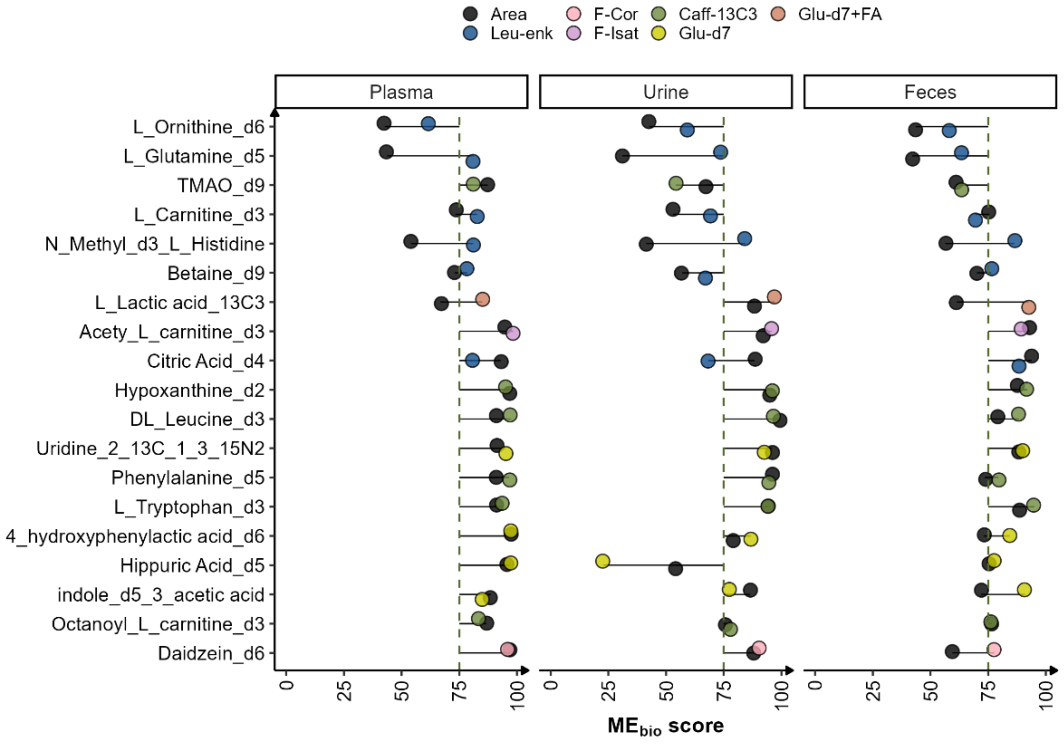


Figure S8. The ME_{bio} score of the 19 SIL standards spiked in plasma, urine and feces before and after correction with the PCIS selected with artificial matrix infusion. The dashed line indicates an ME_{bio} score of 75.

Table S1: General information and stock solution preparation for all the authentic standards

Compound Name	Compound Formula	Monoisotopic Mass/Da	CAS number	supplier	stock /mM	Solvent	usage
L-ornithine-d_6	C3D9NO	84.1249	347841-40-1	CDN	250.0	H2O	SIL
L-glutamine-d_5	C5H5D5N2O3	151.1005	14341-78-7	cambridge Isotope laboratories	250.0	H2O (1% NH3.H2O)	SIL
TMAO-d_9	C3D9NO	84.1249	1161070-49-0	cambridge Isotope laboratories	500.0	H2O	SIL
L-carnitine-d_3	C7H12D3NO3	164.1240	350818-62-1	CDN	125.0	H2O	SIL
n-methyl-d_3-l-histidine	C7D3H9N3O2	172.1040	91037-48-8	CDN	25.0	H2O	SIL
Betaine-d_9	C5H2D9NO2	126.1355	285979-85-3	CDN	250.0	H2O	SIL
L-lactic acid-$^{13}C_3$	13C3H6O3	93.0418	201595-71-3	TRC	170.0	H2O	SIL
acety-L-carnitine-d_3	C9H14D3NO4	206.1346	1334532-17-0	CDN	50.0	H2O	SIL
citric acid-d_4	C6H4D4O7	196.0521	147664-83-3	cambridge Isotope laboratories	1250.0	H2O	SIL
hypoxanthine-d_2	C5D3HN4O	140.0636	NA	cambridge Isotope laboratories	62.5	10% MeOH (0.2M HCL)	SIL
DL-leucine-d3	C6H10D3NO2	134.1135	87828-86-2	CDN	62.5	10% MeOH (1% NH3.H2O)	SIL
uridine-2-^{13}C-1,3-$^{15}N_2$	C8[13C]1H12[15N]2O6	247.0670	369656-75-7	TRC	31.3	H2O	SIL

phenylalanine-d_5	C9H6D5NO2	170.1104	28466-89-7	CDN	50.0	15% MeOH	SIL
L-tryptophan-d_3	C11H9D3N2O2	207.1087	133519-78-5	CDN	50.0	H2O (0.5% NH3.H2O)	SIL
4-hydroxyphenylactic acid-d_6	C8H2D6O3	158.0850	100287-06-7	TRC	125.0	H2O (1.5% NH3.H2O)	SIL
hippuric acid-d_5	C9H4D5NO3	184.0896	53518-98-2	chem Cruz	12.5	H2O	SIL
indole-d_5-3-acetic acid	C10H4D5NO2	180.0947	76937-78-5	TRC	1.3	MeOH	SIL
daidzein-d_6	C15H4D6O4	260.0956	291759-05-2	TRC	0.1	MeOH	SIL
octanoyl-l-carnitine-d_3	C15H26D3NO4	290.2285	1334532-24-9	CDN	2.5	H2O	SIL
leucine-enkephalin	C28H37N5O7	555.2693	81678-16-2	Sigma-Aldrich	1.8	H2O	PCIS
fludrocortisone	C21H29FO5	380.1999	127-31-1	TRC	1.3	MeOH	PCIS
5-fluoroisatin	C8H4FNO2	165.0226	443-69-6	Sigma-Aldrich	6.1	50% MeOH	PCIS
caffeine-$^{13}C_3$	C5[13C]3H10N4O2	197.0904	78072-66-9	TRC	2.5	50% MeOH	PCIS
3-fluoro-DL-valine	C5H10FNO2	135.0696	43163-94-6	Sigma-Aldrich	14.8	50% MeOH	PCIS
L-homoarginine hydrochloride	C7H17ClN4O2	224.1040	1483-01-8	sigma	3.0	50% ACN	artificial matrix compound
sodium dodecyl sulphate	NaSO4C12H25	288.1371	151-21-3	J.T. Baker	1.5	50% ACN	artificial matrix compound
sodium acetate	C2H3NaO2	82.0031	127-09-3	Alfa Aesar	100.0	20% ACN	artificial matrix compound
tridodecylmethylammonium chloride	C37H78ClN	571.5823	7173-54-8	Fluka	1.5	50% ACN	artificial matrix compound

Table S2: Retention time, detection polarity, and final concentrations in the biological samples after post-extration spiking of all the SIL standards

SIL standards	ME_{bio} calculation		ME_{art} calculation (uM)	Retention time/min in Triple TOF 6600 system	Retention time/min in Triple TOF 5600 system	Polarity for detection
	low (uM)	high (uM)				
L-ornithine-d_6	60.0	600.0	300	0.57	0.56	Positive
L-glutamine-d_5	600.0	6000.0	300	0.66	0.66	Positive
TMAO-d_9	40.0	400.0	40	0.69	0.69	Positive
L-carnitine-d_3	30.0	300.0	10	0.69	0.69	Positive
n-methyl-d_3-l-histidine	3.0	30.0	75	0.70	0.70	Positive
Betaine-d_9	50.0	500.0	10	0.71	0.71	Positive
L-lactic acid-$^{13}C_3$	150.0	1500.0	75	1.25	1.20	Negative
acety-L-carnitine-d_3	5.0	50.0	2.5	1.40	1.27	Positive
citric acid-d_4	100.0	1000.0	50	1.57	1.57	Negative
hypoxanthine-d_2	5.0	50.0	25	1.83	1.75	Positive
DL-leucine-d3	80.0	800.0	40	2.42	2.19	Positive
uridine-2-^{13}C-1,3-$^{15}N_2$	5.0	50.0	25	2.72	2.44	Negative
phenylalanine-d_5	80.0	800.0	16	3.06	3.03	Positive
L-tryptophan-d_3	80.0	800.0	40	3.32	3.55	Positive
4-hydroxyphenylactic acid-d_6	1.0	10.0	10	3.79	4.07	Negative
hippuric acid-d_5	5.0	50.0	10	3.86	4.13	Negative
indole-d_5-3-acetic acid	1.0	10.0	50	4.73	5.14	Negative
daidzein-d_6	0.5	5.0	2.5	4.80	5.28	Negative
octanoyl-l-carnitine-d_3	0.2	2.0	1	4.86	5.22	Positive

Table S3: structures overview of PCIS candidates

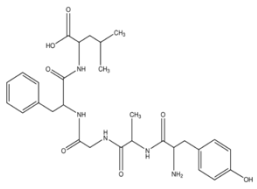
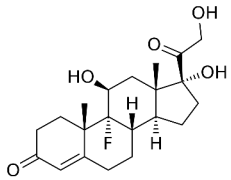
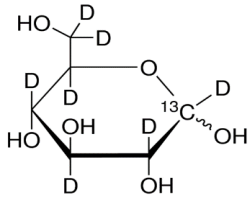
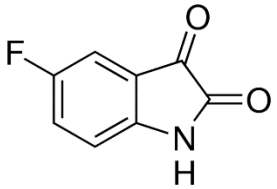
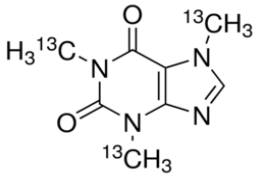
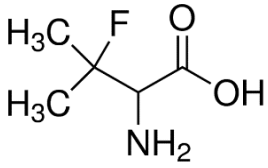
Standards name	Molecular structure	Polarity assessed
Leucine-enkephalin		positive and negative
Fludrocortisone		positive and negative
D-glucose-d_7		positive and negative
5-Fluoroisatin		positive and negative
Caffeine-$^{13}C3$		positive
3-Fluoro-DL-valine		positive and negative

Table S4: Pearson correlation coeffecient (r) of the infusion profiles between PCIS candidates in diverse plasma samples (positive ionization mode)

Sanquine Plasma	[D-glucose-d_7 + Na]+	[Fludrocortisone +H]+	[Leucine-enkepahlin +H]+	[Caffeine$^{13}C_3$ +H]+	[3-Fluoro-DL-valine+H]+	[5-Fluorosatin+H]+
[D-glucose-d_7 + Na]+	1					
[Fludrocortisone +H]+	0.770804431	1				
[Leucine-enkepahlin +H]+	0.772241398	0.824544	1			
[Caffeine$^{13}C_3$ +H]+	0.807158021	0.880798384	0.905081916	1		
[3-Fluoro-DL-valine+H]+	0.72848998	0.735984463	0.968348889	0.894836281	1	
[5-Fluorosatin+H]+	-0.200570801	-0.01464634	-0.532458303	-0.246552806	-0.575077475	1

Divbiosc Plasma	[D-glucose-d_7 + Na]+	[Fludrocortisone +H]+	[Leucine-enkepahlin +H]+	[Caffeine$^{13}C_3$ +H]+	[3-Fluoro-DL-valine+H]+	[5-Fluorosatin+H]+
[D-glucose-d_7 + Na]+	1					
[Fludrocortisone +H]+	0.630635012	1				
[Leucine-enkepahlin +H]+	0.694528207	0.834120488	1			
[Caffeine$^{13}C_3$ +H]+	0.683401118	0.868356752	0.901571591	1		
[3-Fluoro-DL-valine+H]+	0.538261395	0.705531452	0.921834288	0.877180876	1	
[5-Fluorosatin+H]+	-0.190154901	-0.031721979	-0.526404679	-0.283939985	-0.606670633	1

Fasting Plasma	[D-glucose-d_7 + Na]+	[Fludrocortisone +H]+	[Leucine-enkepahlin +H]+	[Caffeine$^{13}C_3$ +H]+	[3-Fluoro-DL-valine+H]+	[5-Fluorosatin+H]+
[D-glucose-d_7 + Na]+	1					
[Fludrocortisone +H]+	0.683430579	1				
[Leucine-enkepahlin +H]+	0.703838424	0.828812768	1			
[Caffeine$^{13}C_3$ +H]+	0.740523259	0.867595416	0.899915573	1		
[3-Fluoro-DL-valine+H]+	0.663754647	0.73810718	0.968591541	0.897303108	1	
[5-Fluorosatin+H]+	-0.256899697	-0.06528373	-0.572740147	-0.304372961	-0.616763643	1

Non-fasting Plasma	[D-glucose-d_7 + Na]+	[Fludrocortisone +H]+	[Leucine-enkepahlin +H]+	[Caffeine$^{13}C_3$ +H]+	[3-Fluoro-DL-valine+H]+	[5-Fluorosatin+H]+
[D-glucose-d_7 + Na]+	1					
[Fludrocortisone +H]+	0.647767404	1				
[Leucine-enkepahlin +H]+	0.653235124	0.792877554	1			
[Caffeine$^{13}C_3$ +H]+	0.70200815	0.813079254	0.890842094	1		
[3-Fluoro-DL-valine+H]+	0.587276227	0.681877501	0.965422476	0.882096272	1	
[5-Fluorosatin+H]+	-0.161839449	-0.019331194	-0.564482804	-0.310651998	-0.626161036	1

Table S5: Pearson correlation coeffecient (r) of the infusion profiles between PCIS candidates in diverse plasma samples (negative ionization mode)

Sanquine Plasma	[D-glucose-d_7 -H]-	[Fludrocortisone+FA-H]-	[Leucine-enkepahlin -H]-	[3-Fluoro-DL-valine-H]-	[5-Fluoroisatin-H]-
[D-glucose-d_7 -H]-	1				
[Fludrocortisone+FA-H]-	0.971665897	1			
[Leucine-enkepahlin -H]-	0.913084057	0.877045789	1		
[3-Fluoro-DL-valine-H]-	0.985792414	0.97978939	0.904547957	1	
[5-Fluoroisatin-H]-	0.931715540	0.894316261	0.992686476	0.917705787	1

Divbiosc Plasma	[D-glucose-d_7 -H]-	[Fludrocortisone+FA-H]-	[Leucine-enkepahlin -H]-	[3-Fluoro-DL-valine-H]-	[5-Fluoroisatin-H]-
[D-glucose-d_7 -H]-	1				
[Fludrocortisone+FA-H]-	0.981822481	1			
[Leucine-enkepahlin -H]-	0.962978629	0.952940878	1		
[3-Fluoro-DL-valine-H]-	0.9891261	0.98361744	0.959161467	1	
[5-Fluoroisatin-H]-	0.964913125	0.956276829	0.994226765	0.953515605	1

Fasting Plasma	[D-glucose-d_7 -H]-	[Fludrocortisone+FA-H]-	[Leucine-enkepahlin -H]-	[3-Fluoro-DL-valine-H]-	[5-Fluoroisatin-H]-
[D-glucose-d_7 -H]-	1				
[Fludrocortisone+FA-H]-	0.981660938	1			
[Leucine-enkepahlin -H]-	0.947698008	0.921857285	1		
[3-Fluoro-DL-valine-H]-	0.984958237	0.982635835	0.933553917	1	
[5-Fluoroisatin-H]-	0.960629360	0.940410637	0.99140311	0.941998665	1

Non-fasting Plasma	[D-glucose-d_7 -H]-	[Fludrocortisone+FA-H]-	[Leucine-enkepahlin -H]-	[3-Fluoro-DL-valine-H]-	[5-Fluoroisatin-H]-
[D-glucose-d_7 -H]-	1				
[Fludrocortisone+FA-H]-	0.984870223	1			
[Leucine-enkepahlin -H]-	0.94116293	0.918152598	1		
[3-Fluoro-DL-valine-H]-	0.98722257	0.983336006	0.928348359	1	
[5-Fluoroisatin-H]-	0.957162007	0.93747786	0.991424244	0.937538212	1

Table S6: Endogenous metabolites used for peak area comparison in plasam injections with and without PCIS infusion

Index	HMDB_ID	Metabolite_Name	Formula	MS.ready.monoiso-topic.mass	RT	Polarity	change in per-centage (%)*
1	HMDB0000378	2-Methyl-butyroylcarnitine	C12H23NO4	245.1627	3.565	Positive	-31.2
2	HMDB0013324	2-Octenoylcarnitine	C15H27NO4	285.194	4.727	Positive	-27.2
3	HMDB0000043	Betaine	C5H11NO2	117.079	0.699	Positive	-1.0
4	HMDB0000097	Choline	C5H13NO	103.0997	0.643	Positive	-0.7
5	HMDB0001161	Deoxycarnitine	C7H15NO2	145.1103	0.748	Positive	23.6
6	HMDB0000705	Hexanoylcarnitine	C13H25NO4	259.1784	4.166	Positive	-26.0
7	HMDB0002013	Isobutyrylcarnitine	C11H21NO4	231.1471	3.135	Positive	-18.4
8	HMDB0000201	Acetylcarnitine	C9H17NO4	203.1158	1.416	Positive	-7.7
9	HMDB0002250	Lauroylcarnitine	C19H37NO4	343.2723	5.89	Positive	25.4
10	HMDB0000062	Carnitine	C7H15NO3	161.1052	0.689	Positive	-9.6
11	HMDB0000791	Octanoylcarnitine	C15H29NO4	287.2097	4.855	Positive	-29.4
12	HMDB0005066	Myristoylcarnitine	C21H41NO4	371.3036	6.371	Positive	11.0
13	HMDB0000824	Propionylcarnitine	C10H19NO4	217.1314	2.963	Positive	-22.8
14	HMDB0002366	Tiglylcarnitine	C12H21NO4	243.1471	3.446	Positive	-24.0
15	HMDB0000925	TMAO	C3H9NO	75.06841	0.685	Positive	-22.9
16	HMDB0000641	Glutamine	C5H10N2O3	146.0691	0.654	Positive	-20.5
17	HMDB0000172	Isoleucine	C6H13NO2	131.0946	2.25	Positive	12.6
18	HMDB0000684	Kynurenine	C10H12N2O3	208.0848	3.059	Positive	-14.0
19	HMDB0000687	Leucine	C6H13NO2	131.0946	2.42	Positive	20.3
20	HMDB0000696	Methionine	C5H11NO2S	149.051	1.353	Positive	1.8
21	HMDB0000214	Ornithine	C5H12N2O2	132.0899	0.568	Positive	3.9
22	HMDB0000159	Phenylalanine	C9H11NO2	165.079	3.061	Positive	-10.7
23	HMDB0000162	Proline	C5H9NO2	115.0633	0.787	Positive	24.1
24	HMDB0000929	Tryptophan	C11H12N2O2	204.0899	3.317	Positive	-2.0
25	HMDB0000158	Tyrosine	C9H11NO3	181.0739	2.41	Positive	29.5
26	HMDB0000883	Valine	C5H11NO2	117.079	1.082	Positive	9.4
27	HMDB0002825	Theobromine	C7H8N4O2	180.0647	3.112	Positive	-25.7
28	HMDB0000357	3-hydroxybutyric acid	C4H8O3	104.0473	2.38	Positive	1.3
29	HMDB0000619	CA	C24H40O5	408.2876	5.98	Positive	20.8
30	HMDB0000562	Creatinine	C4H7N3O	113.0589	0.737	Positive	-8.3
31	HMDB0000626	DCA	C24H40O4	392.2927	6.867	Positive	21.9
32	HMDB0000714	Hippuric acid	C9H9NO3	179.0582	3.863	Positive	2.3
33	HMDB0000197	Indoleacetic acid	C10H9NO2	175.0633	4.733	Positive	-2.1
34	HMDB0000715	Kynurenic acid	C10H7NO3	189.0426	3.457	Positive	-16.9

35	HMDB0000036	TCA	C26H45NO7S	515.2917	5.126	Positive	-32.8
36	HMDB0000289	Uric acid	C5H4N4O3	168.0283	1.692	Positive	-16.4
37	HMDB0000292	Xanthine	C5H4N4O2	152.0334	2.268	Positive	-25.8
38	HMDB0000422	2-MethylglutaricÂ acid	C6H10O4	146.0579	3.371	Positive	-27.8
39	HMDB0000157	Hypoxanthine	C5H4N4O	136.0385	1.83	Positive	-31.7
40	HMDB0002302	Indolepropionic acid	C11H11NO2	189.079	5.08	Positive	-1.0
41	HMDB0000262	Thymine	C5H6N2O2	126.0429	2.96	Positive	-22.0
42	HMDB0000301	Urocanic acid	C6H6N2O2	138.0429	1.5	Positive	-14.2
43	HMDB0000906	Trimethylamine (TMA)	C3H9N	59.0735	0.654	Positive	3.0
44	HMDB0001886	3-methylxanthine	C6H6N4O2	166.0491	3	Positive	-11.8
45	HMDB0013222	3-Guanidinopropano-ate	C4H9N3O2	131.0695	0.825	Positive	-3.2
46	HMDB0004824	N2.N2-dimethylguano-sine	C12H17N5O5	311.123	3.05	Positive	27.9
47	HMDB0000767	Pseudouridine	C9H12N2O6	244.0695	1.35	Positive	-9.8
48	HMDB0001847	Caffeine	C8H10N4O2	194.0804	3.67	Positive	-25.5
49	HMDB0000671	Indolelactic acid	C11H11NO3	205.0739	4.42	Positive	-7.2
50	HMDB0000842	Quinaldic acid	C10H7NO2	173.0477	3.63	Positive	2.2
51	HMDB0005862	2-Methylguanosine	C11H15N5O5	297.1073	2.94	Positive	1.4
52	HMDB0011621	Cinnamoylglycine	C11H11NO3	205.0739	4.53	Positive	-7.2
53	HMDB0000631	GDCA	C26H43NO5	449.3141	6.12	Positive	-3.8
54	HMDB0000698	GLCA	C26H43NO4	433.3192	6.85	Positive	-6.5
55	HMDB0000896	TDCA	C26H45NO6S	499.2968	5.72	Positive	-22.1
56	HMDB0000637	GCDCA	C26H43NO5	449.3141	6.01	Positive	-6.5
57	HMDB0000138	GCA	C26H43NO6	465.309	5.44	Positive	-21.0
58	HMDB0000951	TCDCA	C26H45NO6S	499.2968	5.59	Positive	-28.2
59	HMDB0000708	GUDCA	C26H43NO5	449.3141	5.47	Positive	-19.1
60	HMDB0000518	CDCA	C24H40O4	392.2927	6.73	Positive	23.2
61	HMDB0000641	Glutamine	C5H10N2O3	146.0691	0.66	Negative	-36.1
62	HMDB0000172	Isoleucine	C6H13NO2	131.0946	2.26	Negative	-38.4
63	HMDB0000684	Kynurenine	C10H12N2O3	208.0848	3.05	Negative	-7.9
64	HMDB0000687	Leucine	C6H13NO2	131.0946	2.44	Negative	-34.8
65	HMDB0000696	Methionine	C5H11NO2S	149.051	1.35	Negative	-10.3
66	HMDB0000159	Phenylalanine	C9H11NO2	165.079	3.06	Negative	-52.1
67	HMDB0000162	Proline	C5H9NO2	115.0633	0.78	Negative	-47.2
68	HMDB0000251	Taurine	C2H7NO3S	125.0147	0.64	Negative	-17.2
69	HMDB0000158	Tyrosine	C9H11NO3	181.0739	2.41	Negative	-49.4
70	HMDB0000883	Valine	C5H11NO2	117.079	1.08	Negative	-44.6

71	HMDB0002825	Theobromine	C7H8N4O2	180.0647	3.15	Negative	-26.8
72	HMDB0000357	3-hydroxybutyric acid	C4H8O3	104.0473	2.38	Negative	-16.1
73	HMDB0000714	Hippuric acid	C9H9NO3	179.0582	3.87	Negative	-11.5
74	HMDB0000197	Indoleacetic acid	C10H9NO2	175.0633	4.74	Negative	-47.5
75	HMDB0000715	Kynurenic acid	C10H7NO3	189.0426	3.46	Negative	-34.0
76	HMDB0000289	Uric acid	C5H4N4O3	168.0283	1.7	Negative	-19.7
77	HMDB0000292	Xanthine	C5H4N4O2	152.0334	2.27	Negative	-5.1
78	HMDB0000422	2-MethylglutaricÂ acid	C6H10O4	146.0579	3.371	Negative	4.5
79	HMDB0000157	Hypoxanthine	C5H4N4O	136.0385	1.83	Negative	-22.6
80	HMDB0000190	Lactic acid	C3H6O3	90.0317	1.25	Negative	-11.0
81	HMDB0001886	3-methylxanthine	C6H6N4O2	166.0491	3.00	Negative	9.0
82	HMDB0013222	3-Guanidinopropano-ate	C4H9N3O2	131.0695	0.83	Negative	-48.8
83	HMDB0004824	N2.N2-dimethylguano-sine	C12H17N5O5	311.123	3.05	Negative	-45.9
84	HMDB0011635	p-cresol sulfate	C7H8O4S	188.0143	4.16	Negative	-9.0
85	HMDB0000767	Pseudouridine	C9H12N2O6	244.0695	1.35	Negative	-8.9
86	HMDB0000296	Uridine	C9H12N2O6	244.0695	2.72	Negative	-30.9
87	HMDB0000671	Indolelactic acid	C11H11NO3	205.0739	4.41	Negative	-18.4
88	HMDB0011686	p-Cresol glucuronide	C13H16O7	284.0896	4.12	Negative	-36.1
89	HMDB0005862	2-Methylguanosine	C11H15N5O5	297.1073	2.95	Negative	-50.5
90	HMDB0011621	Cinnamoylglycine	C11H11NO3	205.0739	4.53	Negative	-18.4
91	HMDB0000631	GDCA	C26H43NO5	449.3141	6.12	Negative	-20.1
92	HMDB0000896	TDCA	C26H45NO6S	499.2968	5.73	Negative	-10.9
93	HMDB0000637	GCDCA	C26H43NO5	449.3141	6.01	Negative	-38.1
94	HMDB0000138	GCA	C26H43NO6	465.309	5.43	Negative	-38.1
95	HMDB0000951	TCDCA	C26H45NO6S	499.2968	5.59	Negative	-18.4
96	HMDB0000708	GUDCA	C26H43NO5	449.3141	5.46	Negative	-38.3

* change in percentage is calculated as ($A_{\text{with PCIS}}$ - $A_{\text{without PCIS}}$)/$A_{\text{without PCIS}}$ *100; A stands for peak area

Chapter IV

Current insights into cow's milk allergy in children: microbiome, metabolome and immune response – a systematic review

Based on:

Mariyana V Savova*, Pingping Zhu*, Amy C Harms, Renate G van der Molen, Clara Belzer, Diana M Hendrickx

Current insights into cow's milk allergy in children: microbiome, metabolome and immune response – a systematic review

Pediatric Allergy and Immunology

DOI: 10.1111/pai.14084

*Authors contributed equally

Abstract

The increasing prevalence of IgE-mediated cow's milk allergy (CMA) in childhood is a worldwide health concern. There is a growing awareness that the gut microbiome (GM) might play an important role in CMA development. Therefore, treatment with probiotics and prebiotics has gained popularity. This systematic review provides an overview on the alterations of the GM, metabolome and immune response in CMA-children and animal models, including post-treatment modifications. MEDLINE, PubMed, Scopus and Web of Science were searched for studies on the GM in CMA-diagnosed children, published before March 1, 2023. A total of 21 articles (13 on children, 8 on animal models) were included. The studies suggest that the GM, characterized by an enrichment of the Clostridia class and reductions in the Lactobacillales order and *Bifidobacterium* genus, is associated with CMA in early life. Additionally, reduced levels of short chain fatty acids (SCFAs) and altered amino acid metabolism were reported in CMA-children. Commonly used probiotic strains belong to the *Bifidobacterium* and *Lactobacillus* genera. However, only *Bifidobacterium* levels were consistently upregulated after intervention, while alterations of other bacteria taxa remain inconclusive. These interventions appear to contribute to the restoration of SCFAs and amino acid metabolism balance. Mouse models indicate that these interventions tend to restore the T_h2/T_h1 balance, increase the T_{reg} response, and/or silence the overall pro- and anti-inflammatory cytokine response. Overall, this systematic review highlights the need for multi-omics related research in CMA-children to gain a mechanistic understanding of this disease and to develop effective treatments and preventive strategies.

Introduction

One of the most common food allergies in early childhood is cow's milk allergy (CMA).[1,2] Allergic reactions can be IgE-mediated, non-IgE-mediated, or a mix of both.[3] Multiple studies have shown that among the children diagnosed with CMA those with IgE-mediated reactions to CM tend to have persistent symptoms and acquire tolerance slower than those with non–IgE-mediated reactions.[4–7] At present, infants diagnosed with CMA are placed on an elimination diet consisting of an extensively hydrolyzed formula (EHF) or, if symptoms persist, an amino-acid formula (AAF).[8] Because of the increasing evidence linking food allergies with alterations in gut microbial composition,[9,10] modifying the gut microbiome (GM) with probiotics, prebiotics or synbiotics has emerged as a promising way to prevent and treat allergies.[11] However, there is still little mechanistic understanding on how the GM influences host immune health, leading to allergies, including CMA.[12] Recent technological innovations in the field of microbiome, proteomics and metabolomics have opened new doors for research and provided opportunities to address the gap in understanding the role of GM in CMA. The objective of this systematic review is to further the understanding of the relationship between the GM and CMA, by reviewing existing studies examining microbiome, metabolome, proteome, and immune response data on IgE-mediated CMA in children and animal models.

1. Methods

This systematic review is registered in PROSPERO (CRD42021290177).

2.1 Search strategy

A search in MEDLINE, PubMed, Scopus and Web of Science was performed using the queries in Table S1. The search was limited to research articles published in English before March 1, 2023.

2.2 Inclusion and exclusion criteria

Human case, case-control, and intervention studies were included only if they examined children with IgE-mediated CMA aged 0-12 years. The allergy had to be medically

diagnosed by either a skin prick test (SPT) or an IgE-specific test combined with a cow's milk food challenge. In studies with fecal transplantation (FT), the IgE-mediated CMA status of the donor must be confirmed by the diagnosis criteria used for human studies. For studies reporting data on groups of subjects diagnosed with different types of CMA, only the group with IgE-mediated CMA was reviewed. For animal studies, only case-control and intervention studies on models that included both sensitization and challenge steps were included. The studies were included only if they contained analytical data that examined the GM or metabolome and were excluded when they failed to meet the inclusion criteria, had unclear diagnosis, or involved antibiotic treatment.

2.3 Study selection

Titles, abstracts, and methods were screened independently by two of the authors MVS, PZ, DMH, and by a third author in case of disagreement. Subsequently, the full text of the studies marked as potentially eligible was retrieved and independently checked for eligibility by at least two of the authors MVS, PZ, DMH, and by a third author in case of disagreement or doubts.

2.4 Data extraction

For human studies, the extracted data included general study details (author, year), participant information (age, sample size), CMA diagnosis, analytical data types, data acquisition techniques, measured analytical parameters and significant results. For intervention studies, the intervention details were also extracted. If available, the age range for each group in the study was reported. When only the mean and standard deviation (SD) were available, the age was reported as mean ± sd. The results were split in two: increased and decreased variables between the compared groups. For animal intervention studies, the extracted data included general study details, model information, challenge information, intervention details, data acquisition techniques, measured analytical parameters and significant results.

3 Results

3.1 Search strategy

Our search yielded 733, 479, 512, 897 articles in respectively Scopus, PubMed, MEDLINE and Web of Science. Forty-nine studies were eligible for inclusion. Figure 1 shows the PRISMA[13] flow diagram. Of the 49 papers, 28 were excluded after careful consideration (Table S2).

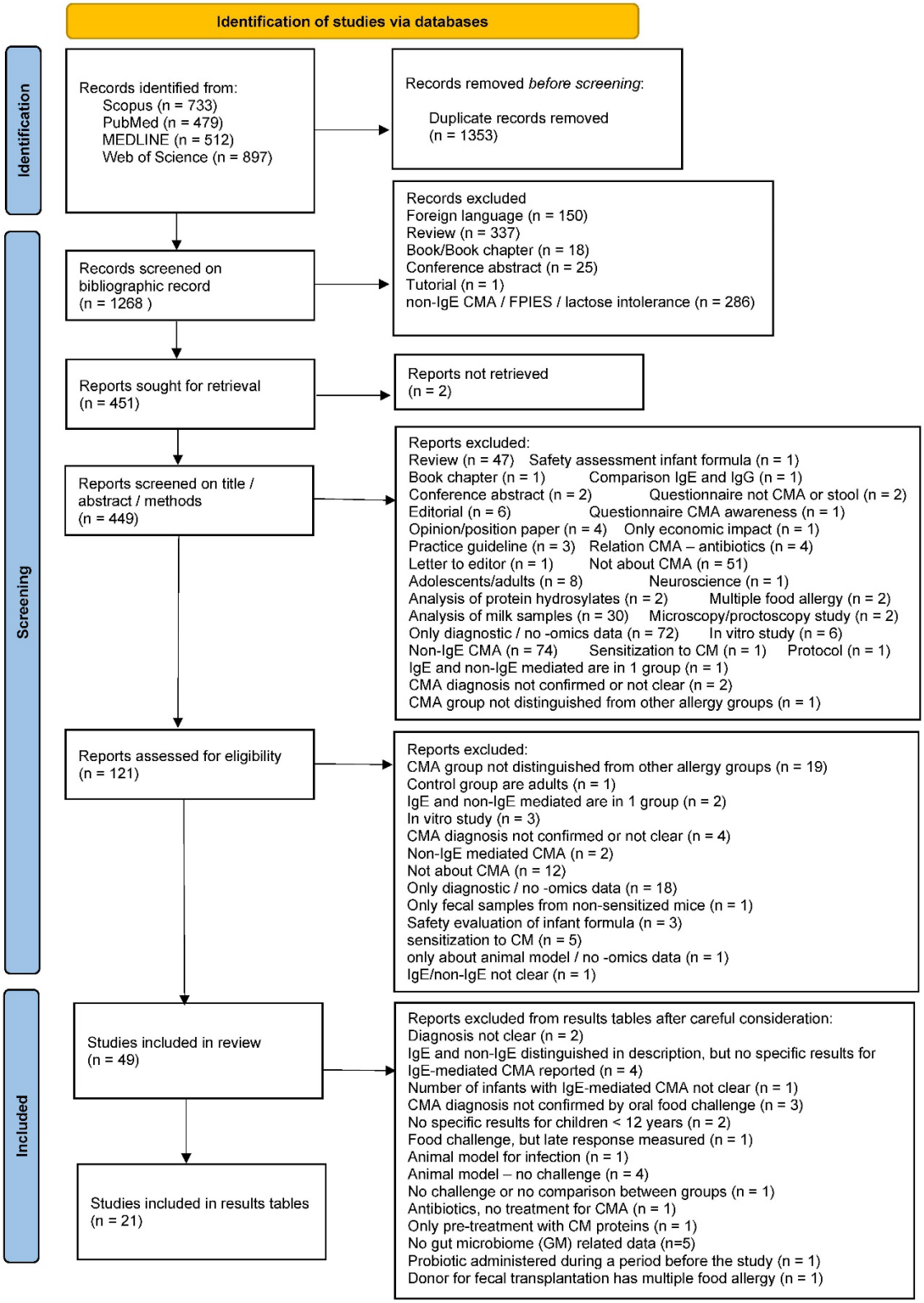


Figure 1. PRISMA flow chart for this systematic review.

3.2 Study findings

3.2.1 Human studies

CMA diagnosis criteria and measured parameters in human studies are summarized in Table S3.

3.2.1.1 Case and case-control studies

Human studies include one case and nine case-control studies (Table 1), among which four examined both the microbiome and metabolome,[14–17] five the microbiome,[18–22] and one the metabolome.[23] For all case-control studies, healthy controls (HC) were used except for one study[23] that considered atopic eczema/dermatitis syndrome infants as controls.

GM modifications

The GM-related studies include four case-control reports,[15,19,17,20] four case-control findings in intervention studies,[14,16,18,21] and one case study.[22] Techniques applied for GM profile identification included bacteria culture[18] and 16S rRNA gene-based approaches (DGGE,[19] FISH[14,15] and gene sequencing[16,17,21,20,22]). Two studies applied specific probes to target certain bacteria groups,[14,15] and six used universal probes or primers to target the V3 region,[19] V4 region[16,22] or both.[17,20,21]

Six studies compared α- and β-diversity between CMA-group and HC, three of them noted increased[16,19] or decreased[20] Shannon α-diversity difference in the CMA-groups, and one reported β-diversity (unweighted UniFrac) difference between CMA-group and HC.[21] A single study reported a higher total bacteria count in the CMA-group.[18]

Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Verrucomicrobia were the primary reported GM phyla. Elevated abundances of the Firmicutes phylum were consistently observed in the CMA-groups.[14–19,21] These included: total Firmicutes;[17,21] the class Clostridia;[17] the families *Lachnospiraceae*[16] and *Ruminococcaceae*[16,17]; the genera *Clostridium*,[14,19] *Faecalibacterium*[16], *Lactobacillus*,[18] *Ruminococcus*[16] and *Subdoligranulum*[19] and the species *Clostridium coccoides*[15] and *Clostridium celerecrescens*.[19] Conversely, certain Firmicutes phylum, including the genus

Granulicatella[21] *and the families Streptococcaceae,*[16] *Enterococcaceae,*[16] and *Acidaminococcaceae,*[20] decreased in the CMA-groups. Additionally, enriched bacteria of the Firmicutes phylum, including the class Clostridia, were also observed in the infants who outgrew CMA.[22]

Bacteroidetes phylum members also showed varying changes in the CMA-groups.[14,17,19–21] These included increased levels of the *Flavobacteriaceae* family,[17] the *Bacteroides*[14,19] and *Prevotella*[21] genera, along with reduced abundance of the *Prevotellaceae* family[20] and the *Parabacteroides* genus.[21] Furthermore, several bacteria from the Proteobacteria phylum, including the *Haemophilus*, *Actinobacillus* and *Klebsiella* genera,[21] and the *Escherichia coli* species,[19] increased in the CMA-groups. In contrast, total Proteobacteria,[17] the *Enterobacteriaceae* family,[16,18] and the *Escherichia* genus[16] decreased. In the Actinobacteria phylum, one study reported increased *Atopobium* cluster (genus) levels,[15] while *Bifidobacteriaceae* family members, including *Bifidobacterium spp.,* consistently exhibited decreased abundance in the CMA-groups.[14,16,18,19] Additionally, the Verrucomicrobia phylum dropped in the CMA-group.[21]

Two studies reported certain bacteria only in the CMA group or the HC. The Clostridium *celerecrescens* species,[19] and the *Burkholderiaceae*, *Nannocystaceae*, *Shewanellaceae*, *Thermomonosporaceae* and *Flavobacteriaceae* families were reported only in the CMA group.[17] In contrast, the *Bifidobacterium bifidum* species[19] and the *Methylophilaceae* and *Dietziaceae* families were exclusively detected in the HC.[17]

Metabolome modifications

Decreased total short chain fatty acid (SCFAs),[14,17] along with increased butyrate and total branched-chain short fatty acids (BCSFAs),[15] were reported in CMA-groups. Besides, lower pyruvate, lactate, threonine and proline, along with higher total esters, ketones, alcohol aldehydes, uridine, histidine, tyrosine, trimethylamine-N-oxide (TMAO) and arginine/histidine,[14] and elevated organic acids were reported in CMA-groups.[23]

Metabolome-microbiome associations

Two studies examined the association between the GM and the metabolome.[15,17] Positive correlations were found between the *Clostridium* genus and butyrate, the *Clostridium coccoides* species and BCSFAs, and the *Bacteroides* genus and propionate.[15] Isocaproate and BCSFAs were negatively related with the *Bifidobacterium* genus.[15] Additionally, lactate was found to be negatively correlated with *Bacteroides* genus[17] *and Clostridium coccoides* species,[15] but positively correlated with *Bifidobacterium* genus.[15]

Table1. Human case and case-control studies in infants/children. Abbreviations: see Supplementary Excel file.

Age years (y); months (m)	**Analytical techniques**	**Analytical data**	**Sample size** (CMA/control)	**Results: modifications in case versus control (case-control study), modifications in allergic versus tolerant (case study)**		**Reference**
				Increase	**Decrease**	
2-12 m	Bacterial culture (CFU)	Microbiome	46/46	**Baseline:** Total bacteria count, Anaerobic bacteria **After 6 months:** Anaerobes count, Lactobacilli count and proportion	**Baseline:** Yeast count **After 6 months:** Bifidobacteria count and proportion, Enterobacteria proportion, Yeast proportion	Thompson-Chagoyan *et al.*[18]
0.55 ± 0.20 y	GC-MS	Metabolomics	16/16	beta-hydroxybutyrate, adipate, isocitrate, homovanillate, suberate, tartarate, 3-indoleacetate, 5-hydroxyindoleacetate	Not reported	#Salmi *et al.*[23]
2-12 m	FISH-FC (16S rRNA gene specific probes); GC-FID	Microbiome, Metabolomics	46/46	*Clostridium coccoides* group, *Atopobium* cluster, butyrate, BCSFA	Not reported	Thompson-Chagoyan *et al.*[15]
6.5-10.4 m	FISH (16S rRNA gene specific probes); GC-MS; NMR;	Microbiome, Metabolomics	18/18	*Bacteroides, Clostridium*, Total esters, ketones, alcohols, aldehydes; Uridine, histidine, tyrosine, TMAO, arginine/histidine	Bifidobacteria, Total SCFAs (major difference: acetate and butyrate), Pyruvate, Lactic acid, threonine, proline	Francavilla *et al.*[14]
5-8 y	PCR-DGGE (V3 regions + 16S rRNA gene- specific primers)	Microbiome	12/12	GM α-diversity (Shannon diversity), *C.coccoides* diversity (Shannon diversity), *Bacteroides, Clostridium, Escherichia coli';* only detected in CMA group: *C. celerecrescens*	*Bifidobacterium* (*B.*) diversity (Shannon diversity), *B. adolescent*, *B. longum*, *B. catenulatum*, *and B. breve* Only detected in control group: *B. bifidum*	Guo *et al.*[19]

Table1. Human case and case-control studies in infants/children. Abbreviations: see Supplementary Excel file.(Continued)

Age years (y), months (m)	**Analytical techniques**	**Analytical data**	**Sample size** (CMA/control)	**Results: modifications in case versus control (case-control study) modifications in allergic versus tolerant (case study)**		**Reference**
				Increase	**Decrease**	
1-12 m	qPCR-16S rRNA (V4 region), GC-FID	Microbiome	19/20	GM α-diversity (Shannon diversity), Gut microbiota evenness (Pielou's evenness), *Ruminococcaceae, Lachnospiraceae*, *Ruminococcus, Faecalibacterium*	*Bifidobacteriaceae, Streptococcaceae, Enterobacteriaceae, Enterococcaceae, Bifidobacterium, Escherichia*	Canani *et al.*[16]
5-8 y	PCR-16s rRNA (V3-V4 regions), HPLC-UV	Microbiome, Metabolomics	6/8	Firmicutes, Clostridia, *Ruminococcaceae, Subdoligranulum* only detected in CMA group: *Burkholderiaceae, Nannocystaceae*, *Shewanellaceae*, *Thermomonosporaceae, Flavobacteriaceae*	Proteobacteria only detected in control group: *Methylophilaceae, Dietziaceae,* Total SCFAs	Dong *et al.*[17]
10-15 m	PCR-16S-rRNA (V3-V4 regions), qRT-PCR	Microbiome	14/14	Firmicutes, *Haemophilus, Actinobacillus, Prevotella, Klebsiella*	Verrucomicrobia, *Parabacteroides, Granulicatella*	Mennini *et al.*[21]
4-6 m	16S-rRNA (V3-V4 regions)	Microbiome	16/34	Not reported	GM α-diversity (Shannon diversity), *Acidaminococcaceae, Prevotellaceae*	Mera-Berriatua *et al.*[20]
3-16 m	16S-rRNA (V4 region)	Microbiome	226/- (3-6m: 29/-)	Fecal microbiome at 3-6 month: *Bacteroidetes*, *Enterobacter* Metagenome functional enrichment of fatty acid metabolism.	Fecal microbiome at 3-6 month: Clostridia, Firmicutes.	Bunyavanich *et al.*[22]

[#]AEDS as basic disease for subjects in both case and control group, and the age is calculated by the pooled mean and sd from the age groups provided in the article

3.2.1.2 Intervention studies

Eight intervention studies for CMA treatment were included (Table2).[14,16,18,21,23–26] Two examined the GM and metabolome,[14,16] one the GM and immune response,[26] four the GM,[18,21,24,25] and one the metabolome.[23] The interventions varied across studies, including synbiotics,[25] prebiotics,[24] probiotics (species of the genus *Bifidobacterium*,[21,26] *Lactobacillus rhamnosus* GG (LGG) species[16,23]) and different formula types.[14,18]

GM modifications

The GM profile was identified with bacteria culture,[18] FISH,[25] 16S rRNA gene sequencing with specific primers/probes[14,24,26] or targeting the V4[16] or V3-V4 regions.[21] Alterations of the phylum Firmicutes in CMA-patients were described in five intervention studies, involving treatment with EHF,[18] lactose-supplemented EHF,[14] LGG,[16] species and strains from the *Bifidobacterium* genus.[21,26] These interventions raised Firmicutes phylum members, including the Turicibacterales order,[48] the *Lactobacillaceae* and *Lachnospiraceae* families[48] and the genera like *Lactobacillus*,[18,48] *Blautia*,[16,21] *Roseburia*,[16] *Coprococcus*,[16] *Anaerofustis*,[16] *Ruminococcus*,[21,26] *Turicibacter*[26] and *Oscillospira*.[26] Conversely, some Firmicutes phylum members, including the Clostridia class,[14] *Christensenellaceae* family[48] and genera like *Enterococcus*, *Streptococcus*,[21] *Anaerovibrio*, *Oscillibacter*, *Bilophila*, *Dorea* and *Roseburia*[26] decreased under treatments.

The interventions also affected the Proteobacteria phylum[21] and its members. The Betaproteobacteria class, the Burkholderiales order, the *Alcalligenaceae* family and the *Sutterella* genus increased in the treated group,[26] while some studies reported decreased levels of the Deltaproteobacteria class,[26] the *Enterobacteriaceae* family[18] and the *Sutterella* genus.[21] In the Bacteroidetes phylum, studies reported the interventions increased levels of the *Porphyromondaceae* family[26] and the *Prevotella* genus,[21,26] and reduced levels of the *Bacteroides* and *Prevotella* genera.[14] Additionally, the Actinobacteria phylum also underwent changes with interventions.[14,18,21,25,26] The use of probiotic *Bifidobacterium* strains consistently elevated the *Bifidobacterium* genus.[21,25,26] Increased *Bifidobacterium* were also noticed after lactose-supplemented EHF diet.[14] In

contrast, the Actinobacteria phylum[21] and its members, the genera *Bifidobacterium*,[18] *Atopobium,*[21] and *Actinomyces,*[21,26] were decreased by the treatments. The Verrucomicrobia phylum and its *Akkermansia* genus were found increased in the treatment group.[21]

In addition to the taxonomy changes, enhanced α-diversity (chao1, observed species),[26] reduced total bacteria[24] and a decreased ratio of the *Eubacterium rectale*/*Clostridium coccoides* species [25] were reported after probiotics, pectin-based thickened AAF and synbiotics treatments, respectively.

Metabolome modifications

After the LGG-supplemented hydrolyzed whey formula (HWF) diet, CMA-patients showed increased kynurenate and decreased 3-indoleacetate.[23] Additionally, butyrate increased in LGG-supplemented extensively hydrolyzed casein (EHC) formula treated CMA-patients.[16] Meanwhile, lactose-supplemented EHF raised SCFAs, lactate, threonine, uridine, histidine, tyrosine, methionine, TMAO, phenylalanine, arginine/histidine and gamma–amino–butyrate/lysine, and lowered the total esters, ketones, alcohols, aldehydes and valine/isoleucine in CMA-patients.[14]

Immune response

The single intervention study reporting findings on the immune response showed that *Bifidobacterium bifidum* reduced allergy symptoms, lowered serum IgE and raised IgG_2 levels in CMA-patients.[26] The IgG_2 and IgE were respectively positively and negatively correlated with GM α-diversity (Chao1 index, observed species, community diversity index, Shannon index). The intervention decreased the pro-inflammatory cytokines TNFα, IL-1β and IL-6 and increased the anti-inflammatory cytokine IL-10 as well.[26]

CMA outcome

Four out of eight intervention studies discussed CMA tolerance or allergic symptoms improvement between treatment and control.[16,24–26] Two studies noted significant improvement in allergic symptoms after treatment,[24,26] and one reported five out of 12 infants in the treated group outgrew CMA after six months, compared to none in the control group.[16]

Table 2. Characteristics of studies that compare CMA infants/children before and after intervention (intervention study). Abbreviations: see Supplementary Excel file.

Age years (y); months (m)	**Analytical techniques**	**Analytical data**	**sample size** (treatment/ control)	**Intervention detail**				**Results: modifications in treatment versus control**		**Reference**
				Duration (months)	**Comparison groups**	**Control diet** (Basic formula (BF))	**Treatment diet** (BF + intervention)	**Increase**	**Decrease**	
0.55 ± 0.20 y	GC-MS	Metabolomics	9/5	1	Treatment vs control	HWF	HWF with LGG	Kynurenate	3-indoleacetate	Salmi *et al.*[23]
2-12 m	Bacteria culture (CFU)	Microbiome	46/46	6	CMA subjects before intervention	-	EHF	*Lactobacilli*	Enterobacteria Bifidobacteria	Thompson-Chagoyan *et al.*[18]
6.5-10.4 m	FISH (16S rRNA-specific probes), GC-MS, NMR;	Microbiome, Metabolomics	16/16	2	CMA subjects before intervention	-	EHF with 3.8% lactose	*Bifidobacteria.* LAB, SCFAs, lactate, threonine, uridine, histidine, tyrosine, methionine, TMAO, Phenylalanine, arginine/histidine, c–amino–butyrate/lysine,	*Atopobium*, *Bacteroides/Prevotella*, clostridia and sulfate-reducing bacteria, Total esters, ketones, alcohols, aldehydes, Valine/isoleucine	Francavilla *et al.*[14]
6.2 ± 4.3 m	qPCR (16S rRNA- specific primers and probes)	Microbiome	23/17	3	Treatment vs control	RAAF	TAAF	Not reported	Total bacteria count	Dupont *et al.*[24]
1-12 m	qPCR-16S rRNA (V4 region), GC-FID	Microbiome; Metabolomics	12/7	6	Treatment vs control, CMA subjects before intervention	EHC formula	EHC formula with LGG	After vs before intervention: *Blautia, Roseburia, Coprococcus*, Compared to control group: *Roseburia,Anaerofustis*. Butyrate	Not observed	Canani *et al.*[16]

Table 2. Characteristics of studies that compare CMA infants/children before and after intervention (intervention study). Abbreviations: see Supplementary Excel file. (Continued)

Age years (y); months (m)	**Analytical techniques**	**Analytical data**	**sample size** (treatment/ control)	**Intervention detail**				**Results: modifications in treatment versus control**		**Reference**
				Duration (months)	**Comparison groups**	**Control diet** (Basic formula(BF))	**Treatment diet** (BF + intervention)	**Increase**	**Decrease**	
0.5-12 m	ELISA qPCR (16S rRNA- specific primers)	Microbiome, Immune response	123/121	6	Treatment vs control	-	*Bifidobacterium bifidum* TMC3115	After 6 months: IL-10, total IgG_2, GM α-diversity (chao1 index, observed species), Bifidobacteriales, *Bifidobacterium*, *Lactobacillaceae Lactobacillus, Turicibacter, Turicibacterales, Betaproteobacteria, Sutterella, Burkholderiales, Alcalligenaceae, Porphyromondaceae, Parabacteroides, Ruminococcus,Oscillospira, Lachnospira*	After 6 months: TNFα, IL-1, IL-6, IL-10, total IgE, *Anaerovibrio, Christensenelaceae, Oscillibacter, Bilophila, Dorea Roseburia*) Desulfovibrionales, Deltaproteobacteria, *Proteobacteria*, *Actinomyces)*	Jing *et al.*[26]
10-15 m	PCR- 16S rRNA (V3-V4 regions), qRT-PCR	Microbiome	14/14	1	CMA subjects before intervention	-	probiotic mix: *Bifidobacterium breve* M-16V, *Bifidobacterium longum subsp. longum* BB536, *Bifidobacterium longum subsp. Infantis* M-63	Verrucomicrobia, Proteobacteria, *Akkermansia, Prevotella, Ruminococcus, Blautia, Bifidobacterium longum subspecies infantis*	Actinobacteria, *Actinomyces, Enterococcus, Streptococcus, Sutterella*	Mennini *et al.*[21]
<13 m	FISH (16S rRNA s-specific probes)	Microbiome	80/89	12	Treatment VS control	AAF	synbiotics: oligosaccharides (oligofructose, inulin), *Bifidobacterium breve* M-16V	After 6 and 12 month: bifidobacteria	After 6 month: ER/CC	Chatchatee *et al.*[25]

3.2.2 Animal studies

The animal studies include two studies on the GM, metabolome and immune response,[27,28] four on the GM and immune response[29–32] and two on the metabolome and immune response[33,34] (Table 3). All animal models were on mice, details are provided in Tables S4 and S5.

GM modifications

Three interventions,[28,31,32] two case-controls[27,30] and one FT[29] study reported GM modifications. Bacteria were identified using 16S rRNA gene-targeted primers, which targeted group/species-specific bacteria[31] or certain hypervariable regions (V3-V4,[27,28,32] V4 [29] and eight other regions[30]).

In two studies comparing GM changes between CMA- and sham mice,[27,30] one observed increased Simpson α-diversity in CMA-male-C57BL/6J mice but decreased Simpson and Shannon α-diversity in CMA-female-BALB/cJ mice.[30] Regardless of the strain and gender, the β-diversity (Bray-Curtis) was significant different between the two groups.[30] Apart from the gender and strain-specific α-diversity difference, CMA-mice showed enrichment in the phyla Bacteroidetes and Patescibacteria (female-C57BL/6J) but reduction in the phyla Verrucomicrobia, Proteobacteria (male-C57BL/6J) and Actinobacteria (female-C57BL/6J).[30] Compared to mice colonized with feces from healthy children (healthy-colonized mice), a FT study reported that mice with feces from CMA children (CMA-colonized mice) had higher abundances of the Clostridiales order and the *Clostridiaceae, Ruminococcaaceae* and *Barnesiellaceae* families, along with lower levels of the *Lachnospiraceae, Erysipelotrichaceae* and *Enterobacteriaceae* families.[29] At the genus level, the CMA-mice exhibited higher *Barnesiella* and *Clostridium*_XIVa,[27] and CMA-colonized mice had enhanced *Enterococcus, Ruminococcus, Coprobacillus, Blautia* and *Parabacteroides.*[29] In contrast, the *Lactobacillus*, *Parvibacter*,[27] *Streptococcus,* and *Salmonella*[29] genera, as well as *Anaerostipes caccae* species[29] decreased in CMA and CMA-colonized mice. Additionally, the *Bosea* genus was absent in CMA-mice.[27]

Species and strains of the *Lactobacillus* and *Bifidobacterium* genera were used as probiotic in CMA-mouse models.[28,31] One study reported that five out of six probiotic strains reduced the total bacteria.[31] Another found significant differences in GM β-diversity (Bray-Curtis, UniFrac) between control and treated groups but only the *Lactobacillus rhamnosus* species increased GM richness.[28] At the family level, it was reported that *Prevotellaceae* and *Marinifilaceae* increased, whereas *Helicobacteraceae*, *Lachnospiraceae*, *Deferribacteraceae*, *Clostridiaceae*, *Peptococcaceae* and *Burkholderiaceae* decreased after taking at least one probiotic.[28] Interestingly, the *Ruminococcaceae* family increased with *Lactobacillus rhamnosus* treatment but decreased with *Bifidobacterium longum subsp. infantis* treatment.[28] Furthermore, one study found that probiotic treatments with *Lactobacillus rhamnosus* and *Bifidobacterium animalis subspecies lactis* increased the *Clostridium* cluster IVa genus and the *Clostridium leptum* species.[31] Conversely, more than three probiotic strains decreased the *Lactobacillus*, *Clostridium* cluster I/II, *Clostridium* cluster XI, *Enterococcus* and *Prevotella* genera, as well as the *Clostridium Coccoides* and *Clostridium Leptum* species.[31] Additionally, it was reported that prebiotic administration with partially hydrolyzed whey reduced the *Lactobacillus* genus and increased the *Prevotella* genus.[32]

Metabolome modifications

Two studies examined fecal SCFAs in CMA-mice with and without synbiotic intervention.[33,34] They reported enhanced acetate[33], butyrate[33] and propionate[34] with synbiotic diet. However, one study only observed reduced kynurenine and N-acetylkunurenine in probiotic-treated mice.[28] Additionally, a FT study compared ileal transcription signatures between CMA and healthy-colonized mice.[29] They found upregulated metabolism of monocarboxylic acid, arachidonic acid, linoleic acid and pyruvate in CMA-colonized mice, while increased carbohydrate metabolic process in healthy-colonized mice.[29]

CMA outcome and immune response

Among all animal studies only Feehley *et al.*[29] and Kostadinova *et al.*[34] correlated the immune response to the GM. Feehley *et al.*[29] reported that growth factor TGF-β receptor

and ROR2 genes in CMA-colonized mice was positively correlated with *Lachnospiraceae* family.[29] Meanwhile, Kostadinova *et al.*[34] showed that propionate was positively correlated with FOXP3+ cell frequency in the colon.[34]

All intervention studies reported immune response data which relates to the treatment outcome.[28,31–34] Unlike post-sensitization,[28] pre-sensitization[31] intake of *Lactobacillus salivarius*, *Lactobacillus rhamnosus* and *Bifidobacterium longum subspecies infantis* successfully lowered the mast cells degranulation marker mucosal mast cell protease-1 (mMCP-1)[35] and BLG-specific IgE.[31] All strains lowered the IL-4 secretion and the BLG-specific $sIgG_1$-to-$sIgG_{2a}$ ratio[31] which indicates the overall T_h2-to-T_h1 response.[36] The rest of the responses were strain-dependent. *Lactobacillus rhamnosus* and *Bifidobacterium longum subspecies infantis* increased T_h1 IFN-γ and T_{reg} IL-10 secretion in stimulated splenocytes, whereas *Lactobacillus salivarius* declined IFN-γ secretion.[31] Post-challenge administration of those probiotic strains predominantly induced regulatory response.[28] All strains significantly increased TGF-β expression, while *Lactobacillus rhamnosus* and *Lactobacillus salivarius* interventions also increased FOXP3 and IL-10 expression. The post-sensitization intake resulted in overall cytokine suppression as well. The reduction in granulocyte-macrophage colony-stimulating factor (GM-CSF), IFN-γ, IL-2, and IL-4 was common among the strains, while IL12p70, IL-10, IL-5 and IL-17A was strain-dependent.[28]

Kostadinova *et al.*[33,34] reported that synbiotic intake alone did not alleviate the acute allergic skin response but its combination with T cell-epitope-containing BLG peptides (PepMix) did.[33,34] Notably, the combined diet reestablished the lost T_h1/T_h2 balance as evidenced by the lymphocyte distribution in the small intestine lamina propria[33] as well as the increased transcription factor (Tbet/GATA3) and cytokine (IFN-γ/IL-13) gene expression in the Peyer's Patches (PP).[34] Right after the intervention the immune response was predominantly regulatory. It was characterized by an increase in the mRNA expression of FOXP3 over the GATA3 and RORγT in the PP, as well as higher FOXP3+ over GATA3+ and T_{reg} over T_h cell frequencies in mesenteric lymph node.[34] Synbiotic addition had a site-dependent effect on IL-22 mRNA expression and also silenced the whey-stimulated splenocyte secretion of cytokines (IL-10, IL-5, IL-13, IL-

17A, IFN-γ) which were induced by the PepMix intake.[33] Kleinjans *et al* showed that the effect of prebiotics on allergic symptoms varied with the composition and treatment duration.[32]

Table 3. CMA intervention studies with animal models. Abbreviations: see Supplementary Excel file.

Groups		Platforms	Results‡		Reference
Case/Intervention	Control		Microbiome/Metabolome	CMA outcome & Immune response	
G1: *L. rhamnosus* **G2:** *B. longum subsp. Infantis* **G3:** *L. salivarius* **G4:** *B. bifidum* **G5:** *L. gasseri* **G6:** *B. animalis subsp. lactis*	**AC**: PBS	**Immunoglobulins** ELISA **Cytokines** IA (ex-BLG) **mRNA expression** q-PCR **Microbiome** qPCR (16s rRNA-specific primers); bacteria culture	**Microbiome** Total bacteria ↓ G1, G2, G3, G4, G5 *Clostridium* cluster IVa ↑ G1, G6 *Staphylococci* abundance ↑ G1 *C. leptum*↑ G1, G6 *Prevotella*↑ G6 *C. leptum* ↓ G2, G3, G4, G5 *Prevotella*↓ G2, G3, G4, Lactobacillus ↓ G2, G3, G4, G5 *Clostridium* cluster I/II↓ G2, G3, G5 Clostridium cluster XI ↓ G2, G3, G4 *C. coccoides*↓ G2, G3, G4, G5 *Enterococcus* ↓ G2, G3, G4, G5 *Enterococcus* ↓ G1	**Allergy markers** mMCP-1 ↓ G1, G2, G3 **Immunoglobulins** BLG-sIgE ↓ G1, G2, G3 BLG-sIgG$_1$/sIgG$_{2a}$↑ G1, G2, G3, G4, G6 **Cytokines** IL-4 ↓ G1, G2, G3, G4 (spleen, MLN) IFN-γ ↑ G1, G2, G6 (spleen) IFN-γ ↓ G3, G4 (spleen) IFN-γ ↑ G6 (MLN) IL-10 ↑ G1, G2, G6 (spleen) IL-10 ↑ G1, G5, G6 (MLN) **mRNA expression** il-4 ↓ G2 IL-10, GATA3, RORγT ↓ G2, G3 FOXP3 ↑ G2, G3 IL-17a ↑ G1, G2, G3	Neau *et al.*[31]
G1: *L. rhamnosus* **G2:** *B. longum subsp. Infantis* **G3:** *L. salivarius*	**AC**: PBS	**Microbiome** PCR - 16S rRNA (V3-V4 regions) **Metabolome** GC-FID, UPLC-MS/MS **Immunoglobulins** ELISA **Cytokines** IA (ex-BLG) **mRNA expression** qPCR	**Metabolome** Kynurenine, N-acetylkunurenine ↓ G1, G2, G3 **Microbiome** Richness (OTU number) ↑ G1 Beta diversity ↑ G1, G2, G3 *Prevotellaceae* ↑ G1, G2, G3 *Marinifilaceae* ↑ G1, G2 *Ruminococcaceae* ↑ G1 *Helicobacteraceae* ↓ G1 *Ruminococcaceae* ↓ G2 *Lachnospiraceae* ↓ G1, G2, G3 *Deferribacteraceae* ↓ G1, G2 *Clostridiaceae* ↓ G1 *Peptococcaceae* ↓ G1, G3 *Burkholderiaceae* ↓ G1 *Anaeroplasmataceae*↓ G2	**Cytokines** GM-CSF, IL-2, IFN-γ, IL-4 ↓ G1, G2, G3 IL12p70 and IL10 ↓ G1 IL-5 ↓ G2, G3 IL17A ↓ G1, G3 **mRNA expression** FOXP3, IL-10 ↑ for G1 and G3 TGFβ ↑ G1, G2, G3	Esber *et al.*[28]
G1 pWH **G2/G3:** pWH + short(G2)/long (G3) scGOS/lcFOS (9:1) **G4/G5:** pWH + short (G4)/long (G5) scGOS/lcFOS (9:1) + pAOS	**TC:** W **AC:** PBS	**Microbiota** PCR (16S rRNA V3-V4 regions) **Immunoglobulins** ELISA	**Microbiome** *Prevotella* ↑ G3, G4, G5 vs G1 *Lactobacillus* ↓ G5 vs G1	**Allergy markers** mMCP-1 ↓ G1, G5 vs AC TSLP ↓ G1 vs AC AASR ↓ TC, G1, G2, G4, G5 vs AC SAS & body-T ↓ TC, G2 vs AC	Kleinjans *et al.*[32]
G1: mix of W peptides (PepMix) **G2:** scFOS and lcFOS (9:1) + *B. breve* M-16V (FF/Bb) **G3:** PepMix + FF/Bb	**TC:** W **AC:** PBS	**Immunoglobulins** ELISA **Metabolites** GC-FID **Lymphocytes** FC **Cytokines** IA (ex-W)	**Metabolites** acetate, butyrate ↑ G2 butyrate ↑ G2 vs G3, TC vs AC	**Allergy markers** AASR ↓ G3, TC vs AC SAS ↓ TC vs AC **Lymphocytes** (SI-LP) T_h1/T_h2 ↑ G3, TC T_{reg}, T_h17 ↑ AC vs TC **Cytokines** (spleen) IFN-γ, IL-17A, IL-13, IL-5, IL-10 ↓ G3 vs G1 & TC vs AC IL-10 ↑ G3	Kostadinova *et al.*[33]

Table 3. CMA intervention studies with animal models. Abbreviations: see Supplementary Excel file.(Continued)

Groups		Platforms	Results‡		Reference
Case/Intervention	**Cotrol**		**Microbiome/Metabolome**	**CMA outcome & Immune response**	
G1: mix of W peptides (**PepMix)** **G2:** scFOS and lcFOS (9:1) + *B. breve* M-16V (**FF/Bb)** **G3:** PepMix + FF/Bb	**TC:** W **AC:** PBS	**Metabolites** GC-FID **Lymphocytes** FC **mRNA expression** qPCR **Immunohisto-chemistry**	Part 1: Post-oral tolerance **Metabolites** butyrate ↑ G3 vs G1 propionate ↑TC, G2, G3 vs AC Positive correlation: propionate and FOXP3+ (colon)	**Allergy markers** AASR ↓ G3, TC vs AC AASR ↑ G1, G2 vs G3 SAS ↓ TC vs AC Part 1: Post-oral tolerance **Lymphocytes** FOXP3+/GATA3+, T_{regs}/T_{effs} ↑ G3 vs AC, G3 vs G2, G3 vs G1 (MLN) T_{regs} ↓ G3 vs AC, G3 vs G2, TC vs AC (spleen) CD25+ ↓ G3 vs G2 **DC** (SI-LP) $CD8\alpha^{-}CD11b^{+}/CD8\alpha^{+}CD11b^{-}$, $CD11b^{+}CD103^{-}$ ↑ G3 $CD8\alpha^{+}CD11b^{-}$↓ G1 **mRNA expression** FOXP3/GATA3 ↑ G3 (PP) FOXP3/RORγT ↑ G3 vs AC, G3 vs G2, G3 vs G1 (PP) TGF-β ↑ G3 vs G2 (proximal SI) TGF-β ↓ G1 (colon) IL-22 ↑ G3 vs AC, G3 vs G1 (PP) IL-22 ↑ for G3 vs G1 (middle SI) IL-22 ↑ G2 vs AC & G2 vs G3 (colon) Galectin 9 ↓ TC Tbet/GATA3 ↓ G1 vs AC, G1 vs G3 (colon) Part 2: Post-challenge **Lymphocytes** (SI-LP) CD25+ Tcells ↑ G3 CD25+ Tcells ↑ G3 vs G2 T_{reg} ↑ G1 **mRNA expression** (PP) Tbet/GATA3 ↑ G3 IFN-γ/IL-13 ↑ G3 vs AC & G3 vs G2	Kostadinova *et al.*[34]

Table 3. CMA intervention studies with animal models. Abbreviations: see Supplementary Excel file.(Continued)

Groups		Platforms	Results‡		Reference
Case/Intervention	**Control**		**Microbiome/Metabolome**	**CMA outcome & Immune response**	
G1: M-C57BL/6J **G2**: M BALB/cJ **G3**: F-C57BL/6J **G4**: F-BALB/cJ	**S:** sham control (sex and strain matched to G1, G2, G3,G4 separately)	**Immunoglobulins** ELISA **Cytokines, chemokines, and acute phase proteins:** IA **Microbiota** 16S rRNA sequencing (8 regions)	**Microbiome** α-diversity ↑ G4 (Simpson and Shannon indices) α-diversity↓ G1(Simpson index) Bacteroidetes↑G3 Patescibacteria↑G3 Verrucomicrobia↓ G1 Proteobacteria↓ G1 Actinobacteria↓G3	**Allergy markers** Body-T ↓ G2 vs S, G4 vs S, G4 vs G3 SAS ↑ G2 vs S, G4 vs S, G4 vs G3 **Immunoglobulins** sIgE ↑ G2 vs S, G1 vs S, G4 vs S, G4 vs G3 $sIgG_1$ ↑ G2 vs S, G2 vs G1, G4 vs S, G4 vs G3 $sIgG_{2a}$ ↑ G2 vs S, G2 vs G1, G4 vs S, G4 vs G3 **Cytokines, chemokines, and acute phase proteins:** G1 vs S: ↑ in CCL1, CSF1, IL-13, CCL17, IL-21, FGF2, CCL12, IL-10, CCL9 G2 vs S: ↓ IL-1β, IL-13, CSF2, TNFRSF1A G4 vs S: ↑ IL-15, TNFRSF1B, ICAM-1	Smith *et al.*[30]
G1: CMA	**S**: Sham control	**Microbiome** PCR-16S rRNA (V3-V4 regions) **Immunoglobulins** ELISA **Cytokines** ELISA **mRNA expression** qPCR **Metabolome** GC-FID, RP, HILIC-MS/MS	**Microbiome** *Barnesiella*↑ *Clostridium*_XIVa↑ *Lactobacillus*↓ *Parvibacter*↓ Only observed in sham mice: *Bosea*	**Allergy markers** Body-T ↓ G1 vs S SAS ↑ G1 vs S Histamine ↑ G1 vs S mMCP-1 ↑ G1 vs S **Immunoglobulins** whey-sIgE, $sIgG_1$, $sIgG_{2a}$ ↑ G1 vs S **Cytokines** IL-6, IL-10 ↑ G1 vs S **mRNA expression** IL-8, IL-33, mTOR mRNA ↑ G1 vs S	Cao *et al.*[27]
G1: CMA-FT **G2:** *Anaerostipes caccae*-FT	**B-HC**: breast-fed HC-FT **F-HC:** formula-fed HC-FT	**Microbiome** PCR - 16S rRNA (V4 region) **Immunoglobulins** ELISA **Transcriptome** RNA-seq, qPCR	**After fecal colonization before sensitization:** **Microbiome** G1 vs F-HC: *Enterococcus*↑ *Barnesiellaceae*↑ *Ruminococcus*↑*Ruminococcaceae*↑ *Coprobacillus* ↑	**Allergy markers** mMCP-1 ↑ G1, G4 vs HC mMCP-1 ↓ G2 vs G1 **Immunoglobulins** BLG-specific IgE, IgG1↑ G1 vs HC **Cytokines** IL-13, IL-4 ↑ G1 vs G2	Feehley *et al.*[29]

			Clostridiaceae ↑ *Clostridiales* ↑ *Blautia* ↑ *Parabacteroides*↑ *Lachnospiraceae*↓ *Erysipelotrichaceae*↓ *Enterobacteriaceae*↓ *Streptococcus*↓*Enterobacteriaceae*↓ *Salmonella*↓*Anaerostipes caccae* ↓ **Transcriptome** **G1 vs F-HC:** (Mroh7, Cntn1, Slc9b2, Letm2, Acot12, Abcc2, Cyp3a59, Cyp2b10, Lrrn1, Me1, Akr1c19, Gstm1, Ces1f) ↑ (Tgfbr3, Acta1, Ror2, Slc22a13, Fbp1, Apcdd1) ↓	**Transcriptome** Tgfbr3 ↓ G1 vs G2, G1 vs HC Ror2 ↓ G1, G2 vs HC Ror2, Tgfbr3 positively correlated to *Lachnospiraceae*	

‡All results are vs AC or C or S unless state otherwise

4 Discussion and Conclusion

In general, no clear conclusion can be drawn about the GM diversity modification in CMA children, because of limited data on β-diversity[21,30] and discordant results regarding to α-diversity in both human[16,19,20] and animal[30] studies.

Taxonomic findings showed that the *Bifidobacteriaceae* family, including *Bifidobacterium spp.,* were consistently reported lower in CMA-children.[14,16,18,19] This result aligns with the consensus on the protective function of *Bifidobacterium spp*. in early life.[37,38] Another noteworthy observation concerning GM in CMA-children is the consistent increase of the Firmicutes phylum,[14–19,21] primarily associated with the Clostridia class. Conversely, decreased levels of bacteria of the Lactobacillales order were observed.[16,21] The trends of Firmicutes alterations align with the findings of an animal study which reported higher *Clostridium* cluster XIVa and lower *Lactobacillus* genus in CMA-mice.[27] However, CMA and healthy-colonized mice were both characterized with bacteria from the Clostridia class, with *Anaerostipes caccae*, a clostridial species, showing protective effects against CMA.[29] Additionally, infants who resolved CMA were reported to have enriched Clostridia class at 3-6 months.[22] Discordant results have also been reported regarding the protective or detrimental effect of the Clostridia class in food allergy.[39,40] Therefore, despite the conflicting findings of the Clostridia class in this review, we lean towards suggesting that GM with enriched Clostridia class, reduced Lactobacillales order and reduced *Bifidobacterium* genus is associated with CMA in early-life.

Various intervention approaches, including probiotics, prebiotics and synbiotics, were applied to restore the balance of GM and the metabolome in CMA-children. Elevated *Bifidobacterium* genus was consistently observed post-treatment with *Bifidobacterium* strains as probiotics[21,25,26] or after lactose-supplemented EHF treatment.[14] However, the impact on the Lactobacillales order in both CMA-children and CMA-mice was less clear. Increased levels of the *Lactobacillaceae* family were reported with *Bifidobacterium*-specific probiotics[26] and EHF in CMA-children,[18] while decreased *Enterococcus* and *Streptococcus* genera were noted in *Bifidobacterium*-treated CMA-

children.[21] Additionally, decreased levels of *Lactobacillus* genus were reported in CMA-mice treated with *Bifidobacterium* and *Lactobacillus*-specific probiotics.[31,32] Similarly, the effect on the Clostridia class varied. Higher levels of its members were reported in CMA-children and mice treated with probiotics.[16,21,26,28,31] Meanwhile, reduced Clostridia class members also noted in CMA-children treated with lactose-supplemented EHF or probiotics,[14,26] and in CMA-mice treated with probiotics.[28,31] Therefore, it is clear that the enhancement of *Bifidobacterium* after *Bifidobacterium*-specific treatment was commonly reported, however the treatment effect on other bacteria remain inconclusive. Despite the uncertainty of most GM profile modifications, there are studies which reported improved allergic symptoms or a high resolution rate in CMA-children treated with probiotics or prebiotics.[16,24,26]

In addition to GM modifications, CMA-children were reported to have decreased total SCFAs[14,16] and altered amino acids and nucleotides levels.[14,23] These findings are consistent with a recent review on the metabolic changes in children with IgE-mediated food allergies,[41] and these metabolome changes appear to be restored with interventions. Increased SCFAs and balanced amino acids were reported after treatment with LGG or lactose-supplemented EHF.[14,23] Enhanced levels of acetate,[33] butyrate,[33,34] and propionate[34] were also reported in synbiotic-treated CMA-mice.

This systematic review provides an overview of the modifications of the GM, metabolome, and immune response in IgE-mediated CMA-children and CMA animal models. Comparing microbiome data between studies is challenging due to methodological variations, diverse intervention approaches, and the reporting of different taxonomic levels. Consequently, only general conclusions can be drawn based on family or higher taxonomic levels. Meanwhile, insights into metabolomics are restricted by limited scope of studied metabolites. Thus, future work should examine broader range of metabolites known to be crucial in the crosstalk between the GM and host's immune system[41,42] and use untargeted metabolomics as hypothesis-generating strategy. Only a single human study reported microbiome and immune response data and their relationship.[26] Similarly, only a single animal study correlated transcriptomics and GM data,[29] including genes related to the immune response. Therefore, there is a

need for both human and animal studies on the correlation of the GM to the immune response. Future animal studies can build on the general treatment outcome findings in the review, namely overall cytokine silencing,[28,33] restoration of the T_h2/T_h1 balance,[31,33,34] and induction of regulatory response.[28,31,34] Moreover, future work can focus on parameters already connected to allergic tolerance acquisition in human, such as induction of T_{reg} response, the production of TGF-β, IgG_4, IgA.[43] No proteomics studies met our inclusion criteria, but a study on the fecal microbiome and metaproteome relationships in CMA-children has been published after our inclusion date.[44] Overall, discussions on multi-omics connections are rare in the reviewed studies, and none of the studies reported shotgun meta-genomics, meta-transcriptomics, or meta-proteomics for microbiome function information. Therefore, there is a clear need for more comprehensive multi-omics studies to gain a better mechanistic understanding of CMA in early life. These efforts would eventually lead to the development of better and effective treatment and preventive strategies.

Acknowledgements

We thank Ria Derkx (Wageningen University library) for her advice on the search strategy. FINANCIAL SUPPORT: This study was part of the EARLYFIT project (Partnership programme NWO Domain AES-Danone Nutricia Research), funded by the Dutch Research Council (NWO) and Danone Nutricia Research (project number: 16490).

Reference:

1. Flom, J. D. & Sicherer, S. H. Epidemiology of Cow's Milk Allergy. *Nutrients* **11**, 1051 (2019).
2. Savage, J. & Johns, C. B. Food Allergy: Epidemiology and Natural History. *Immunol Allergy Clin North Am* **35**, 45–59 (2015).
3. Guidelines for the Diagnosis and Management of Food Allergy in the United States: Report of the NIAID-Sponsored Expert Panel. *Journal of Allergy and Clinical Immunology* **126**, S1–S58 (2010).
4. Høst, A. *et al.* Clinical course of cow's milk protein allergy/intolerance and atopic diseases in childhood. *Pediatric Allergy and Immunology* **13**, 23–28 (2002).
5. Vanto, T. *et al.* Prediction of the development of tolerance to milk in children with cow's milk hypersensitivity. *The Journal of Pediatrics* **144**, 218–222 (2004).
6. Saarinen, K. M., Pelkonen, A. S., Mäkelä, M. J. & Savilahti, E. Clinical course and prognosis of cow's milk allergy are dependent on milk-specific IgE status. *Journal of Allergy and Clinical Immunology* **116**, 869–875 (2005).
7. Schoemaker, A. A. *et al.* Incidence and natural history of challenge-proven cow's milk allergy in European children – EuroPrevall birth cohort. *Allergy* **70**, 963–972 (2015).
8. Hochwallner, H., Schulmeister, U., Swoboda, I., Spitzauer, S. & Valenta, R. Cow's milk allergy: From allergens to new forms of diagnosis, therapy and prevention. *Methods* **66**, 22–33 (2014).
9. Zhao, W., Ho, H. & Bunyavanich, S. The gut microbiome in food allergy. *Annals of Allergy, Asthma & Immunology* **122**, 276–282 (2019).
10. Peroni, D. G., Nuzzi, G., Trambusti, I., Di Cicco, M. E. & Comberiati, P. Microbiome Composition and Its Impact on the Development of Allergic Diseases. *Frontiers in Immunology* **11**, 700 (2020).
11. Shi, J., Wang, Y., Cheng, L., Wang, J. & Raghavan, V. Gut microbiome modulation by probiotics, prebiotics, synbiotics and postbiotics: a novel strategy in food allergy prevention and treatment. *Critical Reviews in Food Science and Nutrition* 1–17 (2022) doi:10.1080/10408398.2022.2160962.
12. Bunyavanich, S. & Berin, M. C. Food allergy and the microbiome: Current understandings and future directions. *Journal of Allergy and Clinical Immunology* **144**, 1468–1477 (2019).
13. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
14. Francavilla, R. *et al.* Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatric Allergy and Immunology* **23**, 420–427 (2012).
15. Thompson-Chagoyan, O. C. *et al.* Faecal microbiota and short-chain fatty acid levels in faeces from infants with cow's milk protein allergy. *International Archives of Allergy and Immunology* **156**, 325–332 (2011).
16. Canani, R. B. *et al.* Lactobacillus rhamnosus GG-supplemented formula expands butyrate-producing bacterial strains in food allergic infants. *The ISME Journal* **10**, 742–750 (2016).
17. Dong, Y., Fei, P., Han, Y. & Guo, L. Characterization of Fecal Microbiota, Short-Chain Fatty Acids and Lactic Acid Concentrations in 5-8-Year-Old Children with Cow Milk Protein Allergy. **28**, 64638 (2018).
18. Thompson-Chagoyan, O. C., Vieites, J. M., Maldonado, J., Edwards, C. & Gil, A. Changes in faecal microbiota of infants with cow's milk protein allergy - A Spanish prospective case-control 6-month follow-up study. *Pediatric Allergy and Immunology* **21**, e394–e400 (2010).
19. Guo, L. *et al.* Comparative Analysis of Fecal Microbiota in 5-8-Year-old Children with and without Cow Milk Protein Allergy. **26**, 6397 (2016).
20. Mera-Berriatua, L. *et al.* Unravelling the Gut Microbiota of Cow's Milk–Allergic Infants, Their Mothers, and Their Grandmothers. *Journal of Investigational Allergology and Clinical Immunology* **32**, 395–398 (2022).

21. Mennini, M. *et al.* Gut Microbiota Profile in Children with IgE-Mediated Cow's Milk Allergy and Cow's Milk Sensitization and Probiotic Intestinal Persistence Evaluation. *International Journal of Molecular Sciences Article* (2021) doi:10.3390/ijms22041649.
22. Bunyavanich, S. *et al.* Early-life gut microbiome composition and milk allergy resolution. *Journal of Allergy and Clinical Immunology* **138**, 1122–1130 (2016).
23. Salmi, H., Kuitunen, M., Viljanen, M. & Lapatto, R. Cow's milk allergy is associated with changes in urinary organic acid concentrations. *Pediatric Allergy and Immunology* **21**, e401–e406 (2010).
24. Dupont, C. *et al.* Safety of a New Amino Acid Formula in Infants Allergic to Cow's Milk and Intolerant to Hydrolysates. *Journal of Pediatric Gastroenterology & Nutrition* **61**, 456–463 (2015).
25. Chatchatee, P. *et al.* Tolerance development in cow's milk–allergic infants receiving amino acid–based formula: A randomized controlled trial. *Journal of Allergy and Clinical Immunology* **149**, 650-658.e5 (2022).
26. Jing, W., Liu, Q. & Wang, W. Bifidobacterium bifidum TMC3115 ameliorates milk protein allergy in by affecting gut microbiota: A randomized double-blind control trial. *Journal of Food Biochemistry* **44**, e13489 (2020).
27. Cao, L.-H. *et al.* Food Allergy-Induced Autism-Like Behavior is Associated with Gut Microbiota and Brain mTOR Signaling. *J Asthma Allergy* **15**, 645–664 (2022).
28. Esber, N. *et al.* Three Candidate Probiotic Strains Impact Gut Microbiota and Induce Anergy in Mice with Cow's Milk Allergy. *Appl Environ Microbiol* **86**, e01203-20 (2020).
29. Feehley, T. *et al.* Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat Med* **25**, 448–453 (2019).
30. Smith, N. A. *et al.* Anxiety-like behavior and intestinal microbiota changes as strain-and sex-dependent sequelae of mild food allergy in mouse models of cow's milk allergy. *Brain Behav Immun* **95**, 122–141 (2021).
31. Neau, E. *et al.* Three Novel Candidate Probiotic Strains with Prophylactic Properties in a Murine Model of Cow's Milk Allergy. *Applied and Environmental Microbiology* **82**, 1722–1733 (2016).
32. Kleinjans, L. *et al.* Mice co-administrated with partially hydrolysed whey proteins and prebiotic fibre mixtures show allergen-specific tolerance and a modulated gut microbiota. *Beneficial Microbes* **10**, 165–178 (2019).
33. Kostadinova, A. I. *et al.* A Specific Mixture of Fructo-Oligosaccharides and Bifidobacterium breve M-16V Facilitates Partial Non-Responsiveness to Whey Protein in Mice Orally Exposed to β-Lactoglobulin-Derived Peptides. *Frontiers in Immunology* **7**, 673 (2017).
34. Kostadinova, A. I. *et al.* Dietary Intervention with β-Lactoglobulin-Derived Peptides and a Specific Mixture of Fructo-Oligosaccharides and Bifidobacterium breve M-16V Facilitates the Prevention of Whey-Induced Allergy in Mice by Supporting a Tolerance-Prone Immune Environment. *Front Immunol* **8**, 1303 (2017).
35. Vaali, K., Puumalainen, T., Wolff, H., Alenius, H. & Palosuo, T. Mucosal mast cell protease-1 (MMCP-1) as a marker of intestinal immunopathology in food allergy model. *Journal of Allergy and Clinical Immunology* **115**, S240 (2005).
36. Stevens, T. L. *et al.* Regulation of antibody isotype secretion by subsets of antigen-specific helper T cells. *Nature* **334**, 255–258 (1988).
37. Di Gioia, D., Aloisio, I., Mazzola, G. & Biavati, B. Bifidobacteria: their impact on gut microbiota composition and their applications as probiotics in infants. *Appl Microbiol Biotechnol* **98**, 563–577 (2014).
38. Saturio, S. *et al.* Role of Bifidobacteria on Infant Health. *Microorganisms* **9**, 2415 (2021).
39. Lee, K. H., Song, Y., Wu, W., Yu, K. & Zhang, G. The gut microbiota, environmental factors, and links to the development of food allergy. *Clin Mol Allergy* **18**, 5 (2020).
40. Huang, Y. J. *et al.* The microbiome in allergic disease: Current understanding and future opportunities—2017 PRACTALL document of the American Academy of Allergy, Asthma & Immunology and the European Academy of Allergy and Clinical Immunology. *Journal of Allergy and Clinical Immunology* **139**, 1099–1110 (2017).
41. De Paepe, E., Van Gijseghem, L., De Spiegeleer, M., Cox, E. & Vanhaecke, L. A Systematic Review of Metabolic Alterations Underlying IgE-Mediated Food Allergy in Children. *Molecular Nutrition & Food Research* **65**, 2100536 (2021).
42. Spertini, F. Metabolomics and allergy: Opening Pandora's box. *Journal of Allergy and Clinical Immunology* **145**, 782–784 (2020).
43. Sampson, H. A. *et al.* Mechanisms of food allergy. *Journal of Allergy and Clinical Immunology* **141**, 11–19 (2018).
44. Hendrickx, D. M. *et al.* Assessment of infant outgrowth of cow's milk allergy in relation to the faecal microbiome and metaproteome. *Sci Rep* **13**, 12029 (2023).

Supplementary Material

Table S1. Search queries

1. MEDLINE	((((cow*.ti. OR cow*.ab. OR cow*.kw. OR cow*.kf.) AND (milk.ti. OR milk.ab. OR milk.kw. OR milk.kf.)) AND ((allerg*.ti. OR allerg*.ab. OR allerg*.kw. OR allerg*.kf.) OR (hypersensitiv*.ti. OR hypersensitiv*.ab. OR hypersensitiv*.kw. OR hypersensitiv*.kf.))) OR milk hypersensitivity.sh.) AND ((microb*.ti. OR microb*.ab. OR microb*.kw. OR microb*.kf.) OR (microflora.ti. OR microflora.ab. OR microflora.kw. OR microflora.kf.) OR (16S*.ti. OR 16S*.ab. OR 16S*.kw. OR 16S*.kf.) OR (bifido*.ti. OR bifido*.ab. OR bifido*.kw. OR bifido*.kf.) OR (bacter*.ti. OR bacter*.ab. OR bacter*.kw. OR bacter*.kf.) OR (lachno*.ti. OR lachno*.ab. OR lachno*.kw. OR lachno*.kf.) OR (rumino*.ti. OR rumino*.ab. OR rumino*.kw. OR rumino*.kf.) OR (veillo*.ti. OR veillo*.ab. OR veillo*.kw. OR veillo*.kf.) OR (entero*.ti. OR entero*.ab. OR entero*.kw. OR entero*.kf.) OR microbiota.sh. OR bifidobacterium.sh. OR bacteroidaceae.sh. OR bacteroides.sh. OR ruminococcus.sh. OR veillonellaceae.sh. OR veillonella.sh. OR enterobacteriaceae.sh.) AND ((child*.ti. OR child*.ab. OR child*.kw. OR child*.kf.) OR (infant*.ti. OR infant*.ab. OR infant*.kw. OR infant*.kf.) OR (baby.ti. OR baby.ab. OR baby.kw. OR baby.kf.) OR (babies.ti. OR babies.ab. OR babies.kw. OR babies.kf.) OR (toddler*.ti. OR toddler*.ab. OR toddler*.kw. OR toddler*.kf.) OR (newborn*.ti. OR newborn*.ab. OR newborn*.kw. OR newborn*.kf.) OR infant.sh. OR child.sh. OR child, preschool.sh. OR infant, newborn.sh.)
2. PubMed	(((((cow[Title/Abstract] OR cow's[Title/Abstract]) AND milk[Title/Abstract]) AND (allerg*[Title/Abstract] OR hypersensitiv*[Title/Abstract])) OR ((milk hypersensitivity[MeSH Terms]) OR (milk hypersensitivities[MeSH Terms]))) AND (((microb*[Title/Abstract]) OR (microflora[Title/Abstract]) OR (16S[Title/Abstract]) OR (bifido*[Title/Abstract]) OR (bacter*[Title/Abstract]) OR (lachno*[Title/Abstract]) OR (rumino*[Title/Abstract]) OR (veillo*[Title/Abstract]) OR (entero*[Title/Abstract])) OR ((microbiota[MeSH Terms]) OR (microbiotas[MeSH Terms]) OR (human microbiome[MeSH Terms]) OR (human microbiomes[MeSH Terms]) OR (microbiome[MeSH Terms]) OR (microbiome, human[MeSH Terms]) OR (microbiomes[MeSH Terms]) OR (16s ribosomal rna[MeSH Terms]) OR (ribosomal rna, 16s[MeSH Terms]) OR (rna, 16s ribosomal[MeSH Terms]) OR (bifidobacterium[MeSH Terms]) OR (bacteroidaceae[MeSH Terms]) OR (bacteroides[MeSH Terms]) OR (ruminococcus[MeSH Terms]) OR (veillonellaceae[MeSH Terms]) OR (veillonella[MeSH Terms]) OR (enterobacteriaceae[MeSH Terms]))))) AND (((child*[Title/Abstract]) OR (infant*[Title/Abstract]) OR (baby[Title/Abstract]) OR (babies[Title/Abstract]) OR (toddler*[Title/Abstract]) OR (newborn*[Title/Abstract])) OR ((infant[MeSH Terms]) OR (child[MeSH Terms]) OR (child, preschool[MeSH Terms]) OR (infant, newborn[MeSH Terms])))
3. Scopus	(TITLE-ABS-KEY (cow* W/6 milk)) AND ((TITLE-ABS-KEY (allergy)) OR (TITLE-ABS-KEY (hypersensitiv*))) AND ((TITLE-ABS-KEY (microb*)) OR (TITLE-ABS-KEY (microflora)) OR (TITLE-ABS-KEY (16s*)) OR (TITLE-ABS-KEY (bifido*)) OR (TITLE-ABS-KEY (bacter*)) OR (TITLE-ABS-KEY (lachno*)) OR (TITLE-ABS-KEY (rumino*)) OR (TITLE-ABS-KEY (veillo*)) OR (TITLE-ABS-KEY (entero*))) AND ((TITLE-ABS-KEY (child)) OR (TITLE-ABS-KEY (infant)) OR (TITLE-ABS-KEY (baby)) OR (TITLE-ABS-KEY (toddler)) OR (TITLE-ABS-KEY (newborn)))
4. Web of Science	(TI=(cow* AND milk) OR AB=(cow* AND milk) OR AK=(cow* AND milk) OR KP=(cow* AND milk)) AND ((TI=(allergy) OR AB=(allergy) OR AK=(allergy) OR KP=(allergy)) OR (TI=(hypersensitiv*) OR AB=(hypersensitiv*) OR AK=(hypersensitiv*) OR KP=(hypersensitiv*))) AND ((TI=(microb*) OR AB=(microb*) OR AK=(microb*) OR KP=(microb*)) OR (TI=(microflora) OR AB=(microflora) OR AK=(microflora) OR KP=(microflora)) OR (TI=(16s*) OR AB=(16s*) OR AK=(16s*) OR KP=(16s*)) OR (TI=(bifido*) OR AB=(bifido*) OR AK=(bifido*) OR KP=(bifido*)) OR (TI=(bacter*) OR AB=(bacter*) OR AK=(bacter*) OR KP=(bacter*)) OR (TI=(lachno*) OR AB=(lachno*) OR AK=(lachno*) OR KP=(lachno*)) OR (TI=(rumino*) OR AB=(rumino*) OR AK=(rumino*) OR KP=(rumino*)) OR (TI=(veillo*) OR AB=(veillo*) OR AK=(veillo*) OR KP=(veillo*)) OR (TI=(entero*) OR AB=(entero*) OR AK=(entero*) OR KP=(entero*))) AND ((TI=(child) OR AB=(child) OR AK=(child) OR KP=(child)) OR (TI=(infant) OR AB=(infant) OR AK=(infant) OR KP=(infant)) OR (TI=(baby) OR AB=(baby) OR AK=(baby) OR KP=(baby)) OR (TI=(toddler) OR AB=(toddler) OR AK=(toddler) OR KP=(toddler)) OR (TI=(newborn) OR AB=(newborn) OR AK=(newborn) OR KP=(newborn)))

Table S2. Information and reasons for the 28 papers excluded after careful consideration

Index	Author and year	Exclusion reason
1	Pohjavuori *et al.*, 2004[1]	Diagnosed IgE-mediated CMA based on a CM challenge and skin prick tests or antigen-specific IgE of any antigen tested (including also egg-white, cat, dog and birch).
2	Viljanen *et al*., 2005a[2]	
3	Barros *et al*., 2017[3]	Distinguished between IgE-mediated and non-IgE mediated CMA in the description of the allergic subjects but did not report any specific results for IgE-mediated CMA.
4	Viljanen *et al*., 2005b[4]	
5	Burks *et al*., 2015[5]	
6	Dong *et al*., 2018[6]	
7	Jarvinen *et al*., 2014[7]	Reported 29 infants with IgE-mediated CMA in their table with clinical characteristics. However, elevated levels of cow's milk specific IgE were reported in only 13 infants. The corresponding author was contacted by email, but was unable to supply additional data because the research was done in a previous institution
8	Mercer *et al*., 2009[8]	CMA was diagnosed based on total and CM specific IgE levels and CMA-related symptoms, but no oral food challenge was used to confirm CMA.
9	Taniuchi *et al*., 2005[9]	Included several subjects whose diagnosis was not confirmed by an oral food challenge, but by a cow's milk elimination diet
10	Kendler *et al*., 2006[10]	Did not confirm CMA by oral food challenge
11	Hol *et al*., 2008[11]	Used a food challenge, but diagnosed children based on their late response, which does not point to IgE-mediated CMA
12	Shek *et al*., 2005[12]	Included both children below 12 years old as well as adolescents and/or adults, but results for children were not reported separately
13	Yamamoto-Hanada *et al*., 2023[13]	
14	Hill *et al*., 1989[14]	Did not include any gut microbiome data or intervention targeting the gut microbiome
15	Hauer *et al*., 1997[15]	
16	Szabó and Eigenmann, 2000[16]	
17	Paparo *et al*., 2016[17]	Studied infants that received probiotics in the past, before entering the study, and therefore could not be compared to the probiotic intervention studies discussed in this review
18	Gotteland *et al*., 1992[18]	Studied CM protein absorption after *E. coli* infection
19	Morin *et al*., 2012[19]	Animals models were sensitized to CM, but did not receive a food challenge, thus focus on CM sensitization rather than CMA
20	Shandilya *et al*., 2016[20]	
21	Wróblewska *et al*., 2020[21]	
22	Maiga *et al*., 2017[22]	
23	Pescuma *et al*., 2019[23]	Two of the three experiments had no challenge, while in the third one there was no comparison between (allergy or treatment) groups
24	Graversen *et al*., 2021[24]	Focused on antibiotics instead of treatment for CMA.
25	Liu *et al*., 2023[25]	Studied the effect of pre-treatment with whey or beta-lactoglobulin (BLG) before sensitization
26	Mauras *et al*., 2019[26]	The CMA donor used for fecal transplantation had multiple food allergy
27	Schouten *et al.*,2009[27]	No GM-related data, do not mention how the treatment changed the GM
28	Adel-Patient *et al.*,2020[28]	

Table S3. CMA diagnosis and measured variables for all human studies. Abbreviations: see Supplementary Excel file.

Author and year	CMA diagnosis	Measured variables		
		Microbiome	Metabolomics	Immune response
Thompson-Chagoyan *et al*., 2010[29]	CM-specific IgE, SPT, DBPCFC	Aerobes, Anaerobes, Enterobacteria, Bifidobacteria, *Lactobacilli*, Clostridia	-	-
Salmi *et al*., 2010[30]	CM-specific IgE, SPT, DBPCFC	-	**Urine**: 37 organic acids, Creatinine	-
Thompson-Chagoyan *et al*., 2011[31]	CM-specific IgE, SPT, DBPCFC	10 targeted probes: *Bifidobacterium*, *Bacteroides,* Enterobacteria*, Streptococcus, Lactobacillus, Atopobium, Clostridium coccoides, Clostridium leptum, Clostridium perfringens sps., Clostridium difficile sps.*	**Feces**: Lactate, SCFA (acetate, propionate, butyrate, isocaproic acid), Branched-chain short fatty acids (BCSFA).	-
Francavilla *et al*., 2012[32]	CM-specific IgE, SPT, DBPCFC	13 targeted probes: Domain bacteria, negative control, *Bifidobacterium*, *Bacteroides/Prevotella*, Eubacterium rectale/Clostridium coccoides, *Lactobacillus/Enterococcus*, *Streptococcus/Lactococcus group*, *Escherichia coli*, Sulfate–reducing bacteria (SRB), *Atopobium* group, *Coriobacterium* group, *Clostridium histolyticum*, *Clostridium lituseburense*	**GC-MS (feces)**: 15 organic metabolites (esters, ketones, Alcohols, sulfur compounds, hydrocarbons, SCFA); **NMR (feces):** pyruvic acid, lactic acid, uridine, histidine, tyrosine, threonine, methionine, proline, TMAO, arginine/histidine, valine / isoleucine, phenylalanine, gamma–amino–butyric acid/lysine	-
Guo *et al*., 2016[33]	Analysis of serum samples, SPT, DBPCFC	Dominant bacteria, *Bifidobacterium, Lactobacillus, C. coccoides*, Microbiota diversity (Shannon-Weaver index, dice similarity coefficient)	-	-
Canani et al., 2016[34]	Clinical history, CM-specific IgE, DBPCFC	Dominant bacteria, Microbiota Alpha diversity (Shannon index) and Evenness (Pielou's evenness index)	**Feces**: butyrate	-
Dong et al., 2018[35]	CM-specific IgE, SPT, DBPCFC	Dominant bacteria, Microbiota Alpha diversity (Chao1, ACE, Simpson, Shannon, and coverage indices)	**Feces**: SCFAs (acetate, butyrate, propionate, isobutyrate), lactate	-
Mennini et al., 2021[36]	CM-specific IgE, SPT, DBPCFC	PCR : Dominant bacteria; qRT-PCR; *B. breve, B. longum subsp. longum, B. longum subsp. infantis* Microbiota Alpha diversity (Observed, Chao1 and Shannon indices) and beta diversity(unweighted UniFrac)	-	-

Mera-Berriatua et al., 2022[37]	Clinical history of IgE-mediated food allergy, SPT	Dominant bacteria Microbiota Alpha diversity (Shannon index) and beta diversity (Bray-Curtis distance)	--	-
Bunyavanich et al., 2016[38]	CM-specific IgE, SPT, CM challenge or AD with CM-specific IgE	Microbiome (feces): Dominant bacteria; Microbiota Alpha diversity (Faith's phylogenetic diversity) and beta diversity (unweighted UniFrac)	-	-
Dupont et al., 2015[39]	CM-specific IgE, SPT, or both positive cutaneous tests and IgE, DBPCFC	Total bacteria, *Clostridium* cluster IV, *Bacteroides/ Prevotella* group, *Bifidobacterium, Lactobacillus/ Leuconostoc/Pediococcus* group, *Clostridium* cluster XIVa, *Clostridium* cluster XI, *Clostridium* cluster I/II, *Staphylococcus, Enterococcus, Escherichia coli*	**Plasma:** Amino acids (cysteine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tyrosine, valine) **Feces:** butyrate	-
Chatchatee et al., 2022[40]	CM-specific IgE, SPT, DBPCFC	bifidobacteria and ER/CC group		
Jing *et al.* 2020[41]	SPT, IgE, DBPCFC	dominant bacteria microbiota Alpha diversity (number of OTUs, Chao1, Shannon, Simpson index) and beta diversity (weighted and unweighted UniFrac)	-	**Immunoglobulins** Total IgE, IgG_2 (serum) **Cytokines** TNFα, IL-1β, IL-6, IL-10 (serum)

Table S4. Model information for all animal studies. Abbreviations: see Supplementary Excel file.

Sensitization					Challenge		Intervention details		n size/group ††	Author and year
					Intradermal	Intragastric				
Animal/Strain (gender)	**Allergen: Dose(mg)**	**Adjuvant: Dose (μg)**	**Period† (wk)**	**Administration**	**Allergen: Dose (ug)**	**Allergen: Dose(mg)**	**Introduction**	**Duration**		
C3H/HeOuJ mice (F)	W:20	CT:10	5	i.g.	W:20	W:50	Pre-S		6-8	Kostadinova *et al.*2017a[42]
C3H/HeOuJ mice (F)	W:20	CT:10	5	i.g.	W:20	W:50	Pre-S	6-9d‡	6-8	Kostadinova *et al* 2017b.[43]
C3H/HeOuJ mice (F)	W:20	CT:10	5	i.g.	W:6	W:50	**Long** WS **Short** Pre-S	**Long** 7.5wk **Short** 5d	7-10	Kleinjans *et al.*2019[44]
BALB/cByJ mice (F)	W:15	CT:10	5	i.g.	-	BLG:60	WS	6wk	30	Neau *et al.*,2016[45]
BALB/cByJ mice (F)	W:15	CT:10	5	i.g.	-	BLG:60	Post-S	20d	10 -12	Esber *et al.*2020[46]
Germ-free C3H/HeN (M and F)	BLG:20	CT:10	5	i.g.	-	BLG:2*100	-	-	6-42	Feehley *et al.*2019[47]
C3H/HeN mice (M)	W/W/W: 10/100/0.5	CT/CT/Alum: 10/10/2	5/2/2	i.g./i.g./i.p.	-	W:50	-	-	3-7	Cao *et al.* 2022[48]
C57BL/6J and BALB/cJ (M and F)	BLG:1	CT:10	5	i.g.	-	W:50	-	-	5-10	Smith *et al.* 2021[49]

†All administrations are performed weekly
†† Intervention group sizes (not control group)
‡ Synbiotic diet for 9 days, peptide mix intake for 6 days

Table S5. Measured variables for all animal studies. Abbreviations: see Supplementary Excel file.

Author and year	Measured variables		
	Microbiome	**Metabolomics**	**Immune response**
Neau *et al.*[45]	11 bacteria primers, all bacteria	-	**Igs:** Total and BLG-s IgE, IgG_1, IgG_{2a} (plasma) **Cytokines:** IFN-γ, IL-12p70, IL-4, IL-5, and IL-10 (spleen, MLN) **mRNA expression:** ifn-g, il-4, il-10, tgf-b, il-17a, t-bet, gata3, rorγt, foxp3 (ileum)
Esber *et al.*[46]	α (Shannon index) and β (Bray-Curtis distance, UniFrac distance) diversity	**Feces**: SCFA, Plasma: other metabolites	**Igs:** BLG- sIgE, $sIgG_1$, $sIgG_2$ (plasma) **Cytokines:** IL-17A, IL-2, GM-CSF, IL-4, IFN-γ, IL-10, IL-5, IL-12p70 (spleen) **mRNA expression:** gata3, tbet, foxp3, rorγt, ifnγ, tnfα, il4, Il10, and tgfβ (ileal)
Kleinjans *et al.*[44]	All bacteria	-	**Igs:** W- sIgE, $sIgG_1$, $sIgG_{2a}$ (serum)
Kostadinova *et al.*[42]	-	**Feces**: acetic acid, propionic acid, butyric acid	**Igs:** W- and BLG- sIgE, $sIgG_1$, $sIgG_{2a}$ (serum) **Lymphocytes:** T cells, DC (spleen, MLN, SI-LP) **Cytokines:** IL-5, IL-13, IL-10, IL-17A, IFN-γ (Spleen, MLN, SILP)
Kostadinova *et al.*[43]		Part 1: Post-oral tolerance **Metabolites** **Feces**: acetic acid, propionic acid, butyric acid, valeric acid	Part 1: Post-oral tolerance **mRNA expression:** Foxp3, Tbet, GATA3, RorγT, IL-10, galectin-9, TGF-β, IL-13, IFN-γ, IL-22 (PP, SI (proximal, middle), colon) **Immunohistochemistry**: Foxp3+ cells (colon) Part 2: Post-challenge **mRNA expression:** Foxp3, Tbet, GATA3, RorγT, IL-10, galectin-9, TGF-β, IL-13, IFN-γ, and IL-22 (PP, spleen) **Lymphocytes:** T_{reg} (LP)
Smith *et al.*[24]	α (Shannon, Simpson indices) and β(Bray-Curtis) diversity	-	**Igs:** BLG-sIgE,s sIgG1; sIgG2a (serum) **Cytokines, chemokines, and acute phase proteins:** e.g. IL-10, IL-13, IL-15, IL-1β, IL-31, IL-21, CCL1, CCL9, CCL12, CCL17, FGF2, CDF1, CSF2, TNFFSF1A, TNFRSF1B, ICAM-1 (plasma)
Cao *et al.*[23]	All bacteria, α and β diversity	-	**Igs:** W- sIgE, $sIgG_1$, $sIgG_{2a}$ (serum) **Cytokines:** IL-6, IL-10 (serum) **mRNA expression:** IL-4, IL-8, IL-33, IL-1β, TGF-β, GAPDH, mTOR mRNA
Feehley *et al.*[47]	α (Shannon index) and β (weighted UniFrac) diversity Pielous's evenness		**Igs:** BLG-specific IgE, IgG_1 (serum) **Cytokines:** IL-13, IL-4 (spleen) ex-W **Transcriptome:** 32 genes (IEC)

Reference:

1. Pohjavuori, E. *et al.* Lactobacillus GG effect in increasing IFN-γ production in infants with cow's milk allergy. *Journal of Allergy and Clinical Immunology* **114**, 131–136 (2004).
2. Viljanen, M. *et al.* Probiotics in the treatment of atopic eczema/dermatitis syndrome in infants: a double-blind placebo-controlled trial. *Allergy* **60**, 494–500 (2005).
3. Barros, K. V. *et al.* Evidence for Involvement of IL-9 and IL-22 in Cows' Milk Allergy in Infants. *Nutrients* **9**, 1048 (2017).
4. Viljanen, M. *et al.* Probiotic effects on faecal inflammatory markers and on faecal IgA in food allergic atopic eczema/dermatitis syndrome infants. *Pediatr Allergy Immunol* **16**, 65–71 (2005).
5. Burks, A. W. *et al.* Synbiotics-supplemented amino acid-based formula supports adequate growth in cow's milk allergic infants. *Pediatric Allergy and Immunology* **26**, 316–322 (2015).
6. Dong, P., jing Feng, J., yong Yan, D., jing Lyu, Y. & Xu, X. Early-life gut microbiome and cow's milk allergy- a prospective case - control 6-month follow-up study. *Saudi Journal of Biological Sciences* **25**, 875–880 (2018).
7. Järvinen, K. M. *et al.* Role of maternal elimination diets and human milk IgA in the development of cow's milk allergy in the infants. *Clinical & Experimental Allergy* **44**, 69–78 (2014).
8. Mercer, N. *et al.* Duodenal Intraepithelial Lymphocytes of Children with Cow Milk Allergy Preferentially Bind the Glycan-Binding Protein Galectin-3. *Int J Immunopathol Pharmacol* **22**, 207–217 (2009).
9. Taniuchi, S. *et al.* Administration of Bifidobacterium to infants with atopic dermatitis: Changes in fecal microflora and clinical symptoms. *Journal of Applied Research* **5**, 387–396 (2005).
10. Kendler, M., Uter, W., Rueffer, A., Shimshoni, R. & Jecht, E. Comparison of fecal microflora in children with atopic eczema/dermatitis syndrome according to IgE sensitization to food. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* **17**, 141–147 (2006).
11. Hol, J. *et al.* The acquisition of tolerance toward cow's milk through probiotic supplementation: A randomized, controlled trial. *Journal of Allergy and Clinical Immunology* **121**, 1448–1454 (2008).
12. Shek, L. P. C., Bardina, L., Castro, R., Sampson, H. A. & Beyer, K. Humoral and cellular responses to cow milk proteins in patients with milk-induced IgE-mediated and non-IgE-mediated disorders. *Allergy* **60**, 912–919 (2005).
13. Yamamoto-Hanada, K. *et al.* Combination of heat-killed Lactiplantibacillus plantarum YIT 0132 (LP0132) and oral immunotherapy in cow's milk allergy: a randomised controlled trial. *Beneficial Microbes* **14**, 17–29 (2023).
14. Hill, D. J., Firer, M. A., Ball, G. & Hosking, C. S. Recovery from milk allergy in early childhood: Antibody studies. *The Journal of Pediatrics* **114**, 761–766 (1989).
15. Hauer, A. C., Breese, E. J., Walker-Smith, J. A. & MacDonald, T. T. The Frequency of Cells Secreting Interferon-γ and Interleukin-4, -5, and -10 in the Blood and Duodenal Mucosa of Children with Cow's Milk Hypersensitivity. *Pediatr Res* **42**, 629–638 (1997).
16. Szabó, I. & Eigenmann, P. a. Allergenicity of major cow's milk and peanut proteins determined by IgE and IgG immunoblotting. *Allergy* **55**, 42–49 (2000).
17. Paparo, L. *et al.* Epigenetic features of FoxP3 in children with cow's milk allergy. *Clin Epigenetics* **8**, 86 (2016).
18. Gotteland, M., Crain-Denoyelle, A. M., Heyman, M. & Desjeux, J. F. Effect of Cow's Milk Protein Absorption on the Anaphylactic and Systemic Immune Responses of Young Rabbits during Bacterial Diarrhoea. *Int Arch Allergy Immunol* **97**, 78–82 (1992).
19. Morin, S. *et al.* Delayed bacterial colonization of the gut alters the host immune response to oral sensitization against cow's milk proteins. *Molecular Nutrition & Food Research* **56**, 1838–1847 (2012).
20. Shandilya, U. K., Sharma, A., Kapila, R. & Kansal, V. K. Probiotic Dahi containing Lactobacillus acidophilus and Bifidobacterium bifidum modulates immunoglobulin levels and cytokines expression in whey proteins sensitised mice. *Journal of the Science of Food and Agriculture* **96**, 3180–3187 (2016).
21. Wróblewska, B. *et al.* Effect of Low-Immunogenic Yogurt Drinks and Probiotic Bacteria on Immunoreactivity of Cow's Milk Proteins and Tolerance Induction-In Vitro and In Vivo Studies. *Nutrients* **12**, 3390 (2020).
22. Maiga, M. A. *et al.* Neonatal mono-colonization of germ-free mice with Lactobacillus casei enhances casein immunogenicity after oral sensitization to cow's milk. *Molecular Nutrition & Food Research* **61**, 1600862 (2017).
23. Pescuma, M., Hébert, E. M., Font, G., Saavedra, L. & Mozzi, F. Hydrolysate of β-lactoglobulin by Lactobacillus delbrueckii subsp. bulgaricus CRL 656 suppresses the immunoreactivity of β-lactoglobulin as revealed by in vivo assays. *International Dairy Journal* **88**, 71–78 (2019).

24. Graversen, K. B. *et al.* Partially Hydrolysed Whey Has Superior Allergy Preventive Capacity Compared to Intact Whey Regardless of Amoxicillin Administration in Brown Norway Rats. *Front Immunol* **12**, 705543 (2021).
25. Liu, M. *et al.* Oral pretreatment with β-lactoglobulin derived peptide and CpG co-encapsulated in PLGA nanoparticles prior to sensitizations attenuates cow's milk allergy development in mice. *Frontiers in Immunology* **13**, 1053107 (2023).
26. Mauras, A. *et al.* Gut microbiota from infant with cow's milk allergy promotes clinical and immune features of atopy in a murine model. *Allergy* **74**, 1790–1793 (2019).
27. Schouten, B. *et al.* Cow Milk Allergy Symptoms Are Reduced in Mice Fed Dietary Synbiotics during Oral Sensitization with Whey. *The Journal of Nutrition* **139**, 1398–1403 (2009).
28. Adel-Patient, K. *et al.* Administration of Extensive Hydrolysates From Caseins and Lactobacillus rhamnosus GG Probiotic Does Not Prevent Cow's Milk Proteins Allergy in a Mouse Model. *Frontiers in Immunology* **11**, 1700 (2020).
29. Thompson-Chagoyan, O. C., Vieites, J. M., Maldonado, J., Edwards, C. & Gil, A. Changes in faecal microbiota of infants with cow's milk protein allergy - A Spanish prospective case-control 6-month follow-up study. *Pediatric Allergy and Immunology* **21**, e394–e400 (2010).
30. Salmi, H., Kuitunen, M., Viljanen, M. & Lapatto, R. Cow's milk allergy is associated with changes in urinary organic acid concentrations. *Pediatric Allergy and Immunology* **21**, e401–e406 (2010).
31. Thompson-Chagoyan, O. C. *et al.* Faecal microbiota and short-chain fatty acid levels in faeces from infants with cow's milk protein allergy. *International Archives of Allergy and Immunology* **156**, 325–332 (2011).
32. Francavilla, R. *et al.* Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatric Allergy and Immunology* **23**, 420–427 (2012).
33. Guo, L. *et al.* Comparative Analysis of Fecal Microbiota in 5-8-Year-old Children with and without Cow Milk Protein Allergy. **26**, 6397 (2016).
34. Canani, R. B. *et al.* Lactobacillus rhamnosus GG-supplemented formula expands butyrate-producing bacterial strains in food allergic infants. *The ISME Journal* **10**, 742–750 (2016).
35. Dong, Y., Fei, P., Han, Y. & Guo, L. Characterization of Fecal Microbiota, Short-Chain Fatty Acids and Lactic Acid Concentrations in 5-8-Year-Old Children with Cow Milk Protein Allergy. **28**, 64638 (2018).
36. Mennini, M. *et al.* Gut Microbiota Profile in Children with IgE-Mediated Cow's Milk Allergy and Cow's Milk Sensitization and Probiotic Intestinal Persistence Evaluation. *International Journal of Molecular Sciences Article* (2021) doi:10.3390/ijms22041649.
37. Mera-Berriatua, L. *et al.* Unravelling the Gut Microbiota of Cow's Milk–Allergic Infants, Their Mothers, and Their Grandmothers. *Journal of Investigational Allergology and Clinical Immunology* **32**, 395–398 (2022).
38. Bunyavanich, S. *et al.* Early-life gut microbiome composition and milk allergy resolution. *Journal of Allergy and Clinical Immunology* **138**, 1122–1130 (2016).
39. Dupont, C. *et al.* Safety of a New Amino Acid Formula in Infants Allergic to Cow's Milk and Intolerant to Hydrolysates. *Journal of Pediatric Gastroenterology & Nutrition* **61**, 456–463 (2015).
40. Chatchatee, P. *et al.* Tolerance development in cow's milk–allergic infants receiving amino acid–based formula: A randomized controlled trial. *Journal of Allergy and Clinical Immunology* **149**, 650-658.e5 (2022).
41. Jing, W., Liu, Q. & Wang, W. Bifidobacterium bifidum TMC3115 ameliorates milk protein allergy in by affecting gut microbiota: A randomized double-blind control trial. *Journal of Food Biochemistry* **44**, e13489 (2020).
42. Kostadinova, A. I. *et al.* A Specific Mixture of Fructo-Oligosaccharides and Bifidobacterium breve M-16V Facilitates Partial Non-Responsiveness to Whey Protein in Mice Orally Exposed to β-Lactoglobulin-Derived Peptides. *Frontiers in Immunology* **7**, 673 (2017).
43. Kostadinova, A. I. *et al.* Dietary Intervention with β-Lactoglobulin-Derived Peptides and a Specific Mixture of Fructo-Oligosaccharides and Bifidobacterium breve M-16V Facilitates the Prevention of Whey-Induced Allergy in Mice by Supporting a Tolerance-Prone Immune Environment. *Front Immunol* **8**, 1303 (2017).
44. Kleinjans, L. *et al.* Mice co-administrated with partially hydrolysed whey proteins and prebiotic fibre mixtures show allergen-specific tolerance and a modulated gut microbiota. *Beneficial Microbes* **10**, 165–178 (2019).
45. Neau, E. *et al.* Three Novel Candidate Probiotic Strains with Prophylactic Properties in a Murine Model of Cow's Milk Allergy. *Applied and Environmental Microbiology* **82**, 1722–1733 (2016).
46. Esber, N. *et al.* Three Candidate Probiotic Strains Impact Gut Microbiota and Induce Anergy in Mice with Cow's Milk Allergy. *Appl Environ Microbiol* **86**, e01203-20 (2020).
47. Feehley, T. *et al.* Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat Med* **25**, 448–453 (2019).

48. Cao, L.-H. *et al.* Food Allergy-Induced Autism-Like Behavior is Associated with Gut Microbiota and Brain mTOR Signaling. *J Asthma Allergy* **15**, 645–664 (2022).
49. Smith, N. A. *et al.* Anxiety-like behavior and intestinal microbiota changes as strain-and sex-dependent sequelae of mild food allergy in mouse models of cow's milk allergy. *Brain Behav Immun* **95**, 122–141 (2021).

Table S6. Abbreviations

Abbreviation	Full name/defination
16S rRNA	16S ribosomal ribonucleic acid
AAF	amino acid formula
AASR	acute allergic skin response (ear swelling)
AC	allergic control
AEDS	atopic eczema/dermatitis syndrome
BCSFAs	branched-chain short fatty acids
BLG	beta-lactoglobulin
body-T	body temperature
CFU	colony-forming unit
CM	cow's milk
CMA	cow's milk allergy
DC	dendritic cells
DGGE	denaturing gradient gel electrophoresis
DBPCFC	Double-blind, placebo-controlled food challenge
EHF	extensively hydrolyzed formula
ELISA	Enzyme-linked immunosorbent assay
ER/CC	Eubacterium rectale/Clostridium coccoides
ex-BLG	ex-vivo res-stimulation with BLG
ex-W	ex-vivo res-stimulation with whey
F	female
FC	flow cytometry
FF/Bb	short and long chain FOS and B. breve M-16V
FISH	fluorescent in situ hybridization
FOS	fructo-oligosaccharides
FOXP3	forkhead box P3
FT	fecal transplantation
G	group
GATA3	GATA Binding Protein 3
GC-FID	GC-flame ionization detector
GC-MS	gas-chromatography-mass spectrometry
GM	gut microbiome
GM-CSF	Granulocyte macrophage colony-stimulating factor
GOS	galacto-oligosaccharides
HC	healthy controls
HILIC	Hydrophilic interaction chromatography
HPLC-UV	high-performance liquid chromatography-ultraviolet detector
HWF	hydrolysed whey formula
IA	immunoassay (other than ELISA)
i.p.	intraperitoneal
i.g.	intragastric
i.d.	intradermally
IEC	Intestinal epithelial cell(s)
IFN-γ	Interferon-gamma
Ig(s)	immunoglobin(s)
IL	interleukin
LAB	lactic acid bacteria
lcFOS	long chain fructo-oligosaccharides
LGG	Lactobacillus rhamnosus GG
LP	lamina propria
M	male
MLN	mesenteric lymph node
mMCP-1	mucosal mast cell protease-1
MS	mass spectrometry
MS/MS	Tandem mass spectrometry
NMR	nuclear magnetic resonance
OTU	operational taxonomic unit
pAOS	pectin-derived acidic oligosaccharide

PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PP	Peyer's Patches
qPCR	quantitative PCR
qRT-PCR	quantitative real-time PCR
RAAF	reference amino acid formula
Ror2	Receptor Tyrosine Kinase Like Orphan Receptor 2
RORγT	retinoid-Related Orphan Receptor gamma t
RP	reverse phase
SAS	systematic anaphylaxis scores
SCFAs	Short-chain fatty acids
scFOS	short chain FOS
scGOS	short chain galacto-oligosaccharides
sd	standard deviation
SI	small intestine
sIg	specific Immunoglobulin
SI-LP	small intestine lamina propria
sp.	single unnamed species (of a certain genus)
spp.	multiple species (of a certain genus)
SPT	skin prick test
TAAF	thickener amino acid formula
Tbet	T-box transcription factor
TC	tolerant control
Tgfbr3	Transforming growth factor beta receptor III
TGF-β	Transforming growth factor beta
T_h	T helper cell
T_{eff}	effector T cells
TMAO	trimethylamine-N-oxide
TNFa	tumor necrosis factor alpha
T_{reg}	T regulatory cell
TSLP	thymic stromal lymphopoietin
UPLC-MS/MS	ultra-performance liquid chromatography with tandem mass spectrometry
W	whey
Pre-S	pre-sensitization
Post-S	post-sensitization
WS	whole study
wk	week(s)

Chapter V

Exploring the fecal metabolome in infants with cow's milk allergy: The distinct impacts of cow's milk protein tolerance acquisition and of synbiotic supplementation

Based on:

Pingping Zhu*, Mariyana V. Savova*, Alida Kindt, the PRESTO study team, Harm Wopereis, Clara Belzer, Amy C. Harms, Thomas Hankemeier

Exploring the fecal metabolome in infants with cow's milk allergy:

The distinct impacts of cow's milk protein tolerance acquisition and of synbiotic supplementation

Molecular Nutrition & Food Research

DOI: 10.1002/mnfr.202400583

*Authors contributed equally

Abstract

Scope: Cow's milk allergy (CMA) is one of the most prevalent food allergies in early childhood, often treated via elimination diets including standard amino acid-based formula or amino acid-based formula supplemented with synbiotics (AAF or AAF-S). This work aimed to assess the effect of cow's milk (CM) tolerance acquisition and synbiotic (inulin, oligofructose, *Bifidobacterium breve* M-16 V) supplementation on the fecal metabolome in infants with IgE-mediated CMA

Methods and results: The CMA-allergic infants received AAF or AAF-S for a year during which fecal samples were collected. The samples were subjected to metabolomics analyses covering gut microbial metabolites including SCFAs, tryptophan metabolites, and bile acids. Longitudinal data analysis suggested amino acids, bile acids, and branched SCFAs alterations in infants who outgrew CMA during the intervention. Synbiotic supplementation significantly modified the fecal metabolome after six months of intervention, including altered purine, bile acid, and unsaturated fatty acid levels, and increased metabolites of infant-type *Bifidobacterium* species: indolelactic acid and 4-hydroxyphenyllactic acid.

Conclusion: This study offers no clear conclusion on the impact of CM-tolerance acquisition on the fecal metabolome. However, our results show that six months of synbiotic supplementation successfully altered fecal metabolome and suggest induced bifidobacteria activity, which subsequently declined after 12 months of intervention.

1. Introduction

Cow's milk allergy (CMA), characterized by an immune-mediated response to cow's milk protein(s), is one of the major food allergies in early life.[1,2] Over the past decades, the estimated CMA prevalence in children of developed countries is approximately 0.5–3%.[3,4] The allergic symptoms typically occur in the first year of life, whereas the resolution age varies and is related to the type of CMA.[5] Based on symptoms and pathophysiology, CMA is categorized into immunoglobin E (IgE)-mediated, non-IgE mediated, and mixed IgE CMA.[6] Subjects with IgE-mediated CMA, constituting approximately 60% of all CMA cases,[3] require longer time for tolerance acquisition to CM than non-IgE mediated CMA subjects.[7,8] In recent decades, the relevance of the gut microbiome (GM) in CMA has been highlighted, and studies show that compared to healthy counterparts, children with IgE-mediated CMA exhibit a reduction in bifidobacteria.[9]

Bifidobacteria, the prototypical health-promoting bacteria, are dominant inhabitants in a breast-fed infants gut[10] and play a pivotal role in GM development in early life.[11,12] As co-evolved bacteria, bifidobacteria possess unique glycosidases to digest complex host-derived glycans, particularly the human milk oligosaccharides (HMOs).[13,14] The oligosaccharide fermentation products not only satisfy the energy and carbon demands of bifidobacteria but also benefit other bacteria through cross-feeding activities, thereby contributing to maintaining the GM homeostasis in early life.[10,11]

Thus, bifidobacteria-related probiotics and HMO-mimicked prebiotics have gained popularity in the management of CMA in early-life, alongside the conventional interventions with extensively hydrolyzed formula or amino acids-based formula (AAF).[15] Indigestible oligosaccharides, such as fructooligosaccharides (FOS) and galactooligosaccharides, are used as prebiotics due to their bifidogenic effect on the GM.[16] *Bifidobacterium* species, including *B. bifidum*,[17] *B. longum*,[18] and particularly *B. breve*.[18–21] are widely used probiotics for IgE-mediated CMA management in infants. These bifidobacteria have key immunomodulatory roles in the cross-talk between GM and host immune system: *B. bifidum*, for example, can induce the expression of FoxP3

in the regulatory T (T_{reg}) cells through cell surface polysaccharides,[22] while *B. longum* in neonatal microbiota can alleviate the risk of allergy by promoting the T_{reg} maturation;[23] *B. breve*, particularly the *B. breve* M-16V, can trigger the anti-allergic process in early infancy by regulating the intestinal microbiota, intestinal epithelial barrier, and immune system.[24] Overall, bifidobacteria with HMO-utilization genes are found to induce intestinal IFN-β and silence Th2 and Th17 cytokines, thereby regulating the systemic immune balance in infants.[25] Additionally, by breaking down HMOs, bifidobacteria can indirectly enhance the production of butyrate[26] which is essential for the interplay between GM and systemic immunity,[27] possibly through epigenetics mechanisms.[28] Bifidobacteria-derived indolelactic acid also actively engages in the immunoregulation during infancy.[25,29] However, despite these findings and the wide application of bifidobacteria-related interventions for IgE-mediated CMA,[17–21] none of the studies have reported comprehensive metabolome exploration.

In this study, we investigated longitudinal fecal metabolome changes of infants with IgE-mediated CMA undergoing dietary management with AAF, with and without synbiotics (*Bifidobacterium breve* M-16V; FOS: oligofructose, inulin). By applying linear mixed models (LMMs) and repeated measures analysis of variance simultaneous component analysis+ (RM-ASCA+), we compared the longitudinal fecal metabolome of infants with persistent CMA to those who developed CM-tolerance, and identified key metabolic changes associated with the synbiotic intervention.

2. Experimental section

2.1 Study design and dosage information

This study arises from a multicenter, randomized, double-blind, controlled clinical study PRESTO (registered as NTR3725 in Netherlands Trial Register). Detailed information on ethics committees, institutional review boards, and regulatory authorities that approved the study was previously published.[30]

PRESTO enrolled infants diagnosed with IgE-mediated CMA who then received either amino acid formula (AAF, produced by Nutricia, Liverpool, United Kingdom) or AAF with synbiotic (AAF-S) to manage their CMA. The synbiotic blend consisted of chicory-

derived neutral FOS: oligofructose and inulin in a 9:1 ratio (total concentration of 0.63g/100 ml formula, BENEO-Orafti SA, Oreye, Belgium) and *Bifidobacterium breve* M-16V (1.47×10^9 cfu/100 ml formula, Morinaga Milk Industry, Tokyo, Japan). Caretakers were instructed to provided subjects with a minimum daily dose of 450mL, 350mL, and 250mL for infants aged 0 to 8 months, 9 to 18 months, and older than 18 months, respectively.[19] After 12 months of intervention, the allergy status was re-evaluated through double-blind, placebo-controlled food challenge (DBPCFC) with CM. Detailed information on the diagnosis and reassessment was previously published.[19] Out of the 169 participants enrolled in PRESTO, 40 subjects (aged 3-13 months) were selected for this study based on sample availability. One subject was excluded due to unclear allergy status after 12 months.[30] Of the 16 AAF and 23 AAF-S participants, 10 and 14 infants, respectively, outgrew CMA within 12 months. Stool samples were available at 0 (baseline, TP0), 6 (TP1), and 12 months (TP2) after the start of the intervention, resulting in a total of 117 samples.

2.2 Sample collection and storage

The sample collection procedure has been described previously.[30] In short, fecal samples were collected at home and immediately stored in freezers, then transferred on ice to the participant hospitals and stored at -80°C until transfer to Danone Research & Innovation (Utrecht, the Netherlands) for wet sample aliquoting and SCFAs and lactic acid analysis. Sample aliquots for LC-MS metabolomics analysis were transferred on dry-ice to Leiden University and stored at -80°C until analysis.

2.3 Metabolomic analysis

2.3.1 SCFAs and lactic acid analysis

Quantitative SCFAs, including branched SCFAs (BSCFAs) analysis was performed using GC coupled to flame ionization detector and lactic acid was measured using lactic acid assay kit (Megazyme, Wicklow, Ireland) as previously described.[31]

2.3.2 LC-MS metabolomics analysis

The wet sample aliquots were lyophilized at 4 mbar and -110°C for 20h (Martin Christ Gefriertrocknungsanlagen GmbH, Germany), weighed (20±0.2mg), and stored at -80°C until extraction. Liquid-liquid extraction was performed as described by Hosseinkhani *et al.*[32] with adjusted sample amount and doubled solvent-to-feces ratio. Detailed information on the chemicals, the sample preparation, and the quality control (QC) is available in supplementary materials.

Polar to semi-polar metabolites, including acetylcarnitines, amines, benzenoids, organic acids, indoles, nucleosides, and nucleotides, were analyzed using reverse phase LC coupled with quadrupole (Q)-TOF-MS operated in full-scan positive and negative ionization modes, as described previously[33] and in the supplementary material. Bile and fatty acids were measured using reverse phase LC separation using Q-TOF-MS operated in full scan negative ionization mode, as described in the supplementary material.

Targeted peak integration was performed using SCIEX OS (version 2.1.6., SCIEX) with a maximum mass error of 10 ppm. The retention times were verified against authentic standards. In case of coelution, the targets were reported using the name or abbreviation of one of the targets followed by a "#". Details on the abbreviations used are listed in Table S2. For the polar to semi-polar metabolites, peak area was used for further data analysis, whereas for the bile and fatty acids, the area ratio of compounds to stable isotopically labelled standards (Table S1) was used. Data quality inspection was performed using an in-house quality assurance software performing between batch correction and removal of metabolites with high technical variance (RSD of QC>30%).

2.3.3 Data analysis

Data handling and statistical analyses were performed in R (version 4.3.2). Metabolites with missingness above 20% and with median signal of the samples less than five times the mean signal of the procedure blanks were removed, leaving 166 metabolites. To identify group bias in missingness, Fisher's exact test was performed for metabolites with missingness above 20% at each time point after grouping the subjects by intervention or CM-tolerance status, and the results are summarized in Table S2. Ratios of secondary to primary and unconjugated to conjugated bile acids (BAs) were added,

resulting in a total of 177 variables. A list of the reported metabolites and their abbreviations can be found in Table S3. The raw data were normalized by dry weight and subsequently log_2-transformed. Missing values were imputed per metabolite using the quantile regression imputation of left-censored (QRILC) method.[34] Available clinical characteristics that potentially associated with CM-tolerance status at TP2 or intervention were analyzed with the two-sided Mann-Whitney U-test for numeric variables and the Fisher's exact test for binary variables as reported previously.[30,35]

To assess the change from TP0 to TP1 and TP2, LMMs were built using the lme4 package in R. Prior to building the model, the data was scaled by the standard deviation of all baseline samples. The metabolites were modelled as response variables with group and time as fixed effects and subject ID as a random effect. After grouping the subject by either their CM-tolerance status at TP2 (CM-allergic versus CM-tolerant) or intervention (AAF versus AAF-S), two models were built, namely tolerance-allergy and intervention. For the tolerance-allergy model (*Metabolite ~ time + CM-tolerance_status + time:CM-tolerance_status + (1|ID))*, TP0 and the CM-allergic group were used as references. Pairwise comparisons between groups at each time point and within a group between the time points were performed using the emmeans package in R. For the intervention model (*Metabolite ~ time + time:intervention + (1|ID)*), TP0 and the AAF group were used as references. The main effect of the intervention was removed from the model but its interaction with time was kept ensuring the groups are equal at baseline. The p-values were calculated to assess a change from baseline with the Satterthwaite's degrees of freedom method using the lmerTest package within the ALASCA package.[36] In this study, the combined CM-tolerance status–intervention model was not performed because CM-tolerance acquisition as investigated in the parent study did not differ between the interventions at TP2 and aligned with natural rates of CMA outgrowth in infants.[19] For most metabolites, the addition of age as a covariate to models led to no improvement of the performance based on akaike information criterion (Tables S4 and S5). Therefore, age was not used as a covariate in the LMMs. Multiple testing correction was performed using the Benjamini-Hochberg method where $Q<0.1$ was considered as statistically significant.

Visualization of the longitudinal metabolomic alterations was achieved using RM-ASCA+ with ALASCA package,[36] as detailed in the supplementary materials. Performances of the analysis was validated using nonparametric bootstrapping, and the 95% confidence intervals (CI) were estimated based on 1000 resampling iterations.

2.4 16S rRNA gene sequencing and pre-processing

Extraction of DNA from stool samples and the subsequent gut microbiota profiling by 16S rRNA gene sequencing was performed as described previously.[30] Correlations between the changes in metabolites and the relative abundance of *Bifidobacterium* were examined using Spearman's rank correlation analysis. Relative abundance comparisons of *Bifidobacterium* between and within the AAF and AAF-S groups were evaluated with two-side unpaired t-tests.

3. Results

3.1 Patient characteristics

The statistical results of important clinical characteristics are summarized in Table S6-S7. When grouping the subjects by the CM-tolerance status at TP2, the father allergy occurrence and the SCORing Atopic Dermatitis (SCORAD) at baseline were significantly higher in the CM-allergic group than in the CM-tolerant group (Table S6). None of the clinical characteristics were significantly different between AAF and AAF-S groups (Table S7).

3.2 More pronounced fecal metabolome changes in the CM-tolerant group

Firstly, RM-ASCA+ was used to examine the longitudinal metabolome alterations within and between infants that remained allergic and those that acquired tolerance to CM by TP2 (CM-allergic vs CM-tolerant). The PC1 score plot (Figure 1A) describes the direction of maximum variance in the modeled data, whereas the loadings plot (Figure 1B) highlights the top metabolites contributing to PC1. Metabolites with positive loadings follow the trend described by the score, whereas the opposite holds for metabolites with negative loadings. Figure 1B shows that almost half of the variation (47%) described by the fixed effects of the tolerance-allergy model was explained by

PC1 (Figure 1A). The scores and loading for PC1 showed that over time ferulic acid, desaminotyrosine, pipecolic acid, 3-hydroxybenzoic acid increased, whereas dodecanoylcarnitine, pregnenolone sulfate, betaine, pyruvate decreased (Figure 1). Few BAs also showed overall change with time. The primary BAs cholic acid (CA), chenodeoxycholic acid (CDCA), and hyocholic acid (HCA) declined over time. In contrast, the secondary BAs deoxycholic acid (DCA) and the ratios of secondary to primary BAs, including DCA/CA, lithocholic acid (LCA)/CDCA, increased. Although with overlapped CIs between the two groups, those changes were more pronounced for the CM-tolerant group where the PC1 score declined more sharply than the CM-allergy group and for which the CI between the time points were separated, suggesting a significant time effect in this group.

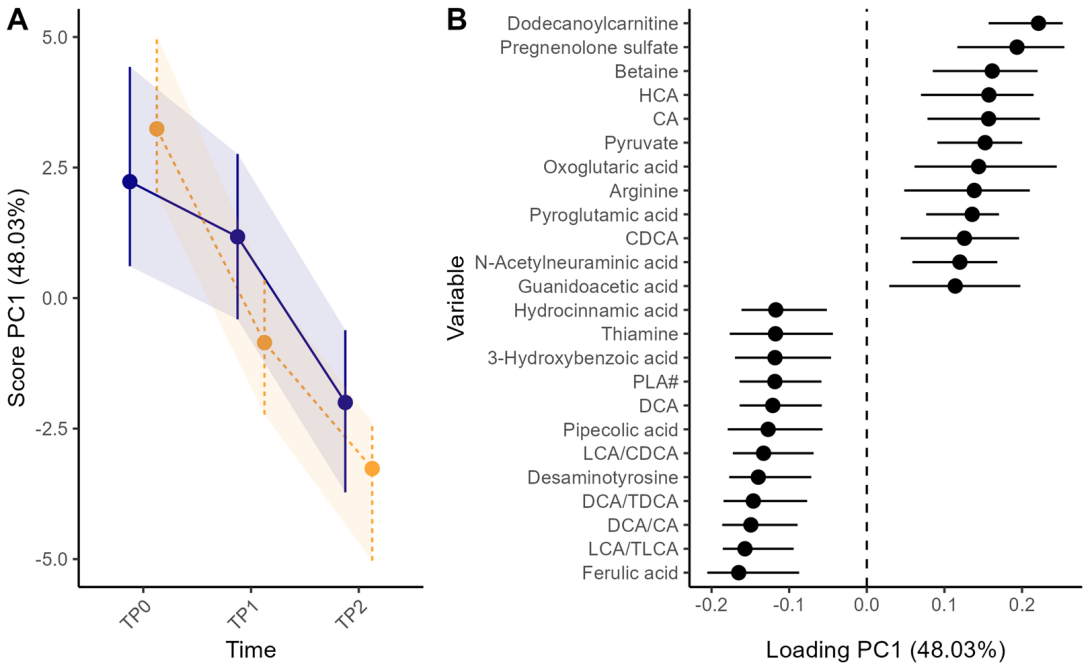


Figure 1. RM-ASCA+ combined effect matrix showing the common metabolome development throughout the study for the CM-allergic (blue solid line, n=15) and CM-tolerant (orange dashed line, n=24) groups as scores (A) and loadings (B). Only the metabolites with 12 highest and 12 lowest loadings are shown in the plot. Error bars representing 95% CI were estimated based nonparametric bootstrapping.

Univariate marginal means comparison showed that around five times more metabolites were significantly altered over time in infants that acquired CM-tolerance versus those

that remained CM-allergic (TP0-TP1: 9 metabolites in CM-tolerant vs 2 metabolites in CM-allergic; TP0-TP2: 30 metabolites in CM-tolerant and 7 in CM-allergic; Figure S1 and Table S8). Pregnenolone sulfate, pyroglutamic acid, pyruvate, oxoglutaric acid, and ferulic acid were significantly affected by time for both groups and follow comparable time-development trends (Figure S1). Similarly, arginine decreased, whereas 3-hydroxybenzoic acid, hydrocinnamic acid, LCA, DCA increased simultaneously in both groups, but significantly only in the CM-tolerant group (Figure S1). Pipecolic acid levels increased over time in both groups, but the rise was steeper and significant only in the CM-tolerant group. Dodecanoylcarnitine followed the trend described by PC1 of the combined effect matrix (Figure 1A) with a decline in time at both TP1 and TP2 significant only in the CM-tolerant group. The rest of the significantly altered metabolites showed dissimilar longitudinal profiles between the groups (Figure S1). Butyric acid, PLA#, desaminotyrosine, and phenylacetic acid were significantly increased, whereas 5-hydroxytryptophan and the primary BAs CA and CDCA showed significant decreases in the CM-tolerant group only. In contrast, threonine#, and tryptophan significantly increased over time only in the CM-allergic group.

Next, the RM-ASCA+ interaction effect matrix was examined to focus on the alterations associated with CM-tolerance acquisition. The PC1 scores and loading of the interaction matrix, Figure 2, suggest that compared to the CM-allergic group, the CM-tolerant group showed overall alterations in amino acid metabolism with an increase in citrulline, lysine, N-acetyltyrosine, phenylacetic acid, gamma-aminobutyric acid (GABA#), glutamate, orotate, ornithine and a decrease in 5-hydroxytryptophan and serotonin. The BAs metabolism was also altered: decline in CDCA, CA, glycochenodeoxycholic acid (GCDCA), tauroursodeoxycholic acid (TUDCA), taurochenodeoxycholic acid (TCDCA) and increase in LCA/CDCA for the CM-tolerant group. The BSCFAs, isobutyrate and isovalerate, also contributed to PC1, showing higher levels in the CM-tolerant group. However, only citrulline and lysine were found significantly different at TP2 between the two groups univariately (Table S6, Figure S2).

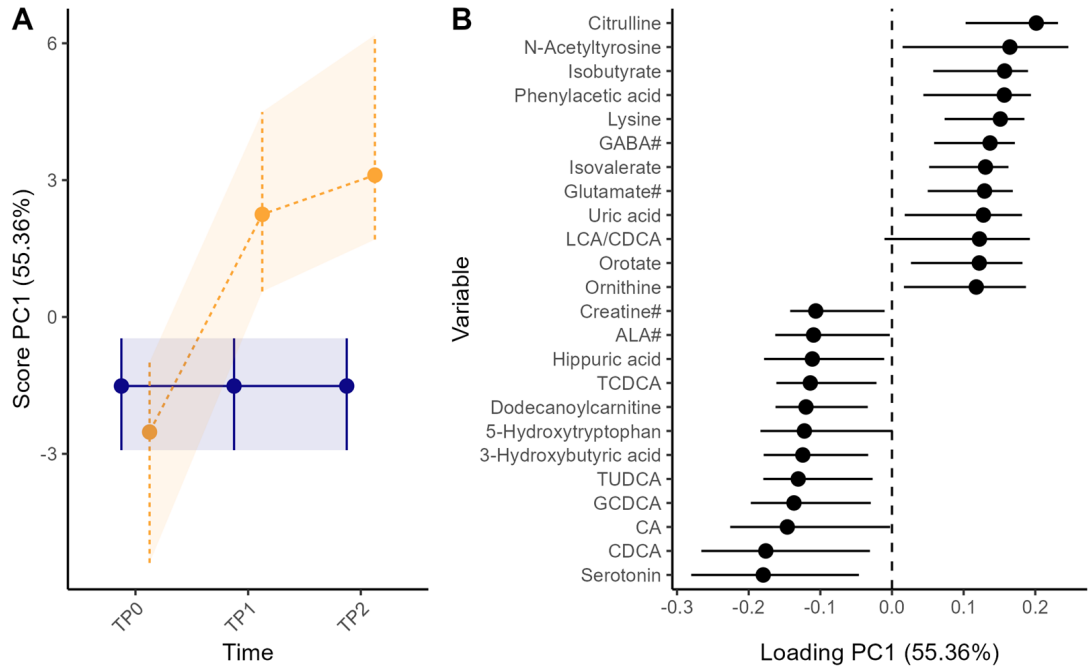


Figure 2. RM-ASCA+ interaction effect matrix showing the metabolome differences between the CM-allergic (blue solid line, n=15) and CM-tolerant group (orange dashed line, n=24) over time as scores (A) and loadings (B). Only the metabolites with 12 highest and 12 lowest loadings are shown in the plot. Error bars representing 95% CI were estimated based nonparametric bootstrapping.

3.3 Synbiotic supplementation altered fecal metabolome after six months of intervention

The longitudinal alterations of the fecal metabolome between the AAF and AAF-S group were studied to understand the effect of the synbiotic supplementation. As shown in Figure 3, clear group separation was observed in PC1 of the RM-ASCA+ interaction effect matrix, especially at TP1.

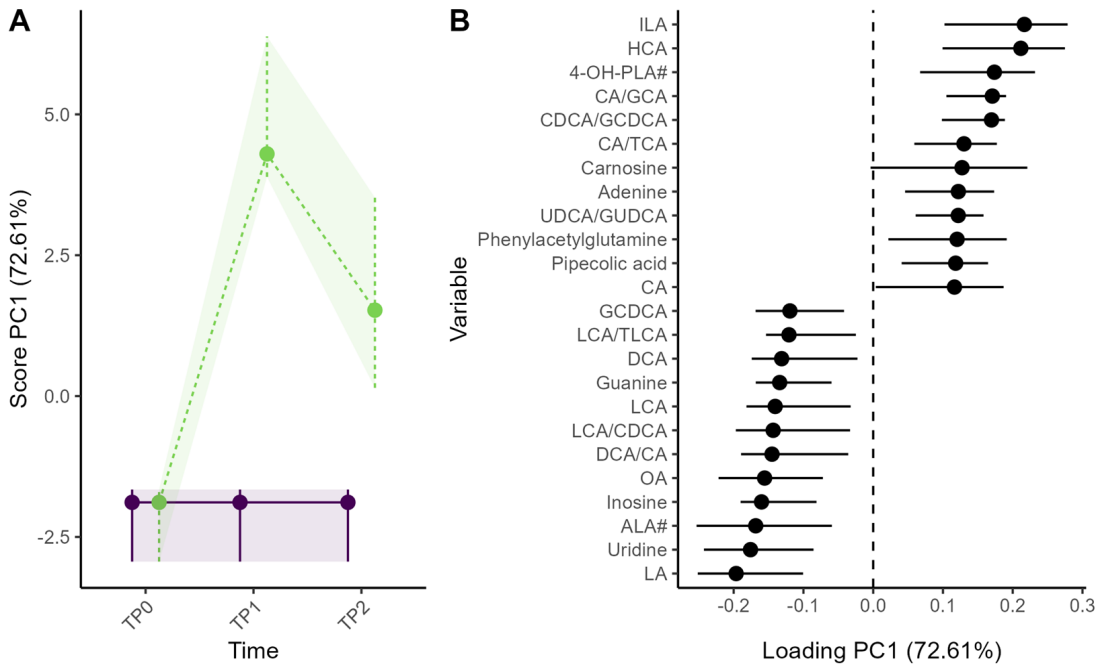


Figure 3. RM-ASCA+ interaction effect matrix showing the metabolome differences between the AAF (purple solid line, n=16) and AAF-S (green dashed line, n=23) group over time as scores (A) and loadings (B). Only the metabolites with 12 highest and 12 lowest loadings are shown in the plot. Error bars representing 95% CI were estimated based nonparametric bootstrapping.

Among all the metabolites, 12 metabolites and three BA ratios were found to be statistically different between the AAF and AAF-S groups at TP1, and only inosine at TP2 (Figure S3, Table S8). The estimated marginal means plot of those analytes can be found in Figure S3. The synbiotic supplementation led to an increase of gut microbial metabolites indolelactic acid (ILA) and 4-hydoxyphenyllactic acid (4-OH-PLA#) and a decline in the fatty acids linoleic acid (LA), alpha-linolenic acid (ALA#), and oleic acid (OA) at TP1 (Figure 4). Amino acid glutamine was also decreased in the AAF-S group at TP1. Three purine metabolites inosine, guanine, and adenine as well as the pyrimidine uridine were also affected by the intervention. While adenine was higher upon the synbiotic addition, the opposite was true for inosine, guanine, and uridine. HCA and CDCA/GCDCA, CA/glycocholic acid (GCA), ursodeoxycholic acid (UDCA)/glycoursodeoxycholic acid (GUDCA) were all significantly higher in the AAF-S than in the AAF group at TP1, whereas GCDCA was significantly lower (Figure

4). A few other BAs were found to be among the main contributors to PC1 of the interaction matrix (Figure 3) or to have significant interaction coefficient at TP1 prior to multiple testing correction (Figure 4), namely, the glyco-conjugated BAs GCA and GUDCA and the secondary BAs and their ratio to primary BAs: LCA, DCA, DCA/CA, and LCA/CDCA.

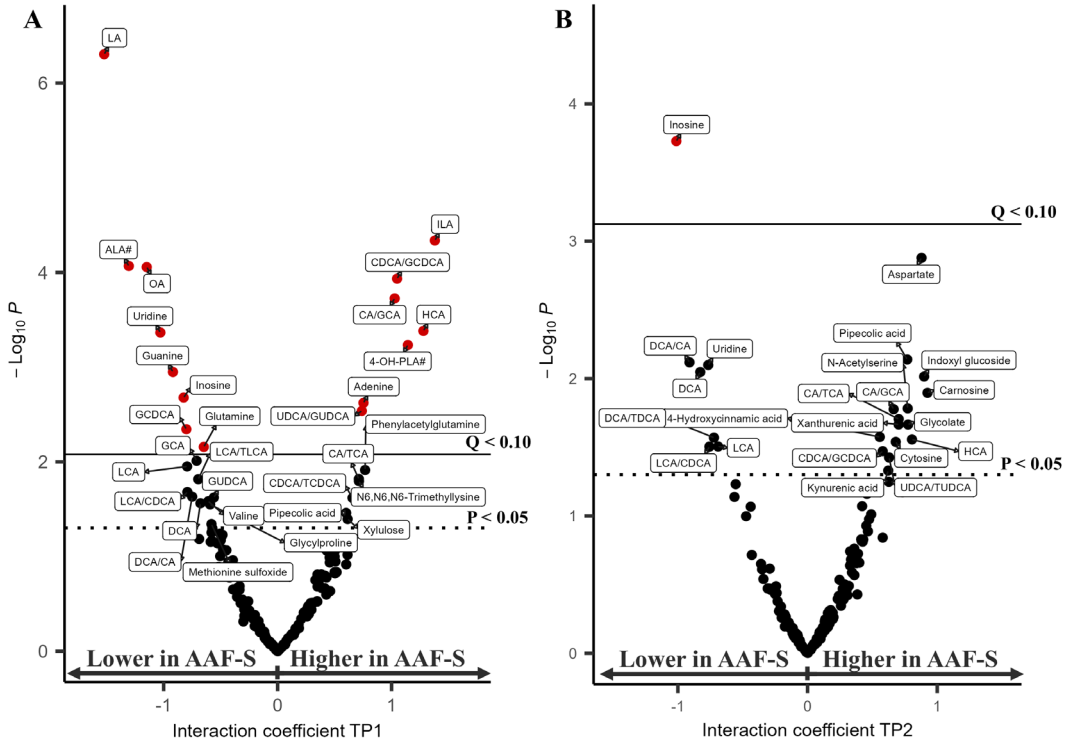


Figure 4. Volcano plot showing the resulting p-value of the interaction coefficient for TP1 (left) and TP2 (right) in intervention LMM, dashed (p = 0.05), solid line (Q = 0.1) for TP1 (A) and TP2 (B). Red symbols indicate metabolites with Q<0.1 after Benjamini-Hochberg procedure.

3.4 Association between changes in *Bifidobacterium* and metabolites significantly altered by the synbiotic

The synbiotic supplementation significantly increased the relative abundance of *Bifidobacterium* in the AAF-S group from baseline to TP1 and TP2 compared to the AAF group (Figure S4).[35] To determine whether these increases were associated with the significantly changed metabolites, Spearman's rank correlation analysis was performed between the changes in metabolite levels and *Bifidobacterium*'s relative

abundance from baseline to TP1 (TP1-TP0) and TP2 (TP2-TP0), respectively (Table S9). In the AAF-S group, changes in ILA and 4-OH-PLA# from TP0 to later time points were positively correlated with those of *Bifidobacterium* ($r > 0.6$, $p < 0.005$), while changes in glutamine were negatively correlated ($r \leq -0.5$, $p < 0.05$) (Figure 5). The changes in *Bifidobacterium* were positively correlated with those of adenine at TP1 and TP2 in both groups ($r > 0.5$, $p < 0.05$), and with CDCA/GCDCA and CA/GCA only at TP1 in the AAF-S group ($r > 0.4$, $p < 0.05$). *Bifidobacterium* also showed negative correlations with GCDCA and inosine in changes from TP0 to TP1 only in the AAF-S group ($r < -0.4$, $p < 0.05$) (Figure S5).

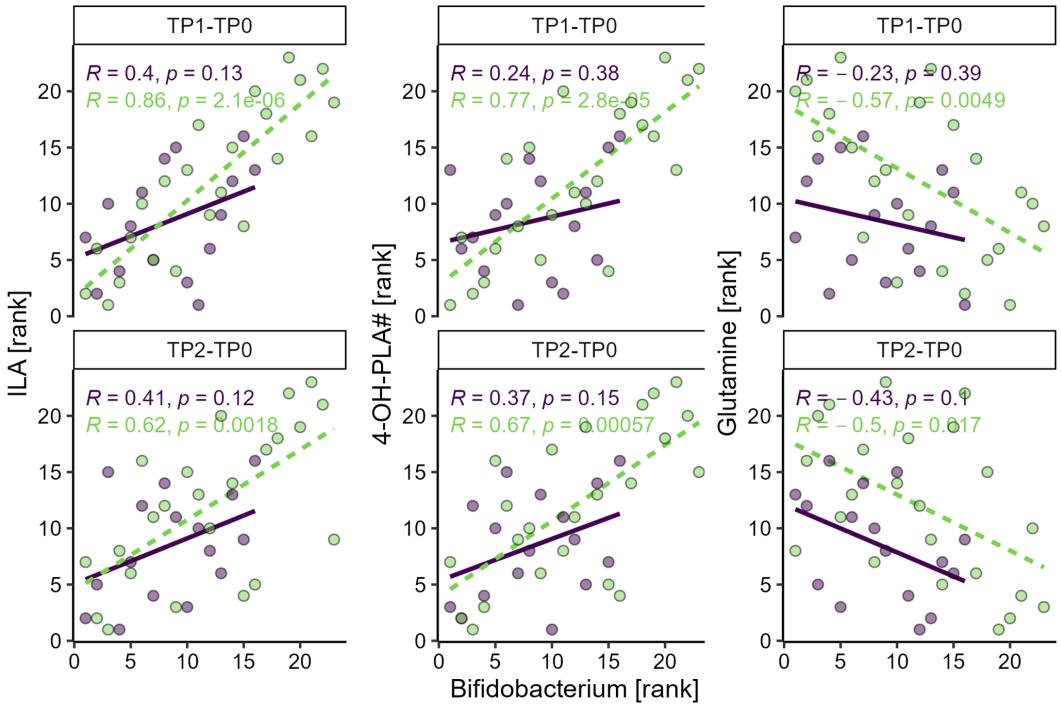


Figure 5. Spearman's rank correlations between the changes in Bifidobacterium and ILA, 4-OH-PLA#, glutamine in AAF (purple solid line, n=16) and AAF-S (green dashed line, n=23) groups from baseline to TP1 (TP1-TP0) and TP2 (TP2-TP0). The rank of the changes in metabolite response and relative abundance of Bifidobacterium within each group were used for plotting. The figure shows p values; the Q values after Benjamini-Hochberg procedure are provided in Table S9.

4. Discussion

In this study we followed the fecal metabolome alterations in infants with IgE-mediated CMA who received AAF with or without synbiotics for a year. Firstly, we examined the effect of CM-tolerance acquisition on the fecal metabolome over time. Time, reflecting growth and diet diversification, had a more pronounced impact on the metabolome than CM-tolerance acquisition (Figure 1, Figure S1). The diet enrichment was evidenced by the overall increase of the phenolic acids which are ubiquitously produced in plants,[37] including ferulic acid, 3-hydroxybenzoic acid, and hydrocinnamic acid. The decrease in the steroid hormone (pregnenolone sulfate), energy metabolites (pyruvate, oxoglutaric acid, dodecanoylcarnitine), and the altered amino acids and derivatives (pyroglutamic acid, arginine, pipecolic acid) suggest metabolome modification associated with somatic growth.[38,39]

The multivariate RM-ASCA+ analysis showed an association of CM-tolerance acquisition status with alterations in amino acids, BAs, and (B)SCFAs (Figure 2). Compared to infants with persistent CMA, citrulline and lysine were significantly higher in the infants who developed CM-tolerance at TP2 (Figure S2). Lower plasma citrulline levels are known marker of increased gut permeability,[40] which can raise the chance of allergen(s) passing the intestinal barrier and triggering the immune system.[41] The increase in fecal citrulline in the CM-tolerant group in this study might suggest improved gut barrier function and gut health. Although not significantly different between the two groups, the amino acids GABA#, glutamate#, threonine#, and ornithine were also higher in the CM-tolerant group compared to the CM-allergic group (Figure S1-S2). Lower fecal threonine levels have previously been reported in infants with IgE-mediated CMA compared to healthy controls.[42] Interestingly, although not significant, 5-hydroxytryptophan and serotonin were higher in the CM-allergic group at TP1 and TP2 (Figure 2), while their precursor tryptophan significantly declined only from TP0 to TP2 in this group (Figure S1). As serotonin is involved in intestinal epithelial proliferation[43] and plays an essential role in regulating intestinal inflammation,[44] the upregulated tryptophan-serotonin metabolism in the CM-allergic group may reflect an inflammatory state of the intestine in the CMA infants.

Children who outgrew CMA showed differences in their BAs profile. The primary BAs (CA, CDCA) significantly decreased, while the secondary BAs (DCA, LCA) and the secondary/primary BAs ratios (DCA/CA, LCA/CDCA) significantly increased from TP0 to TP2 only in the CM-tolerant group (Figure S1). A recent study found that, compared to healthy children, children with IgE-mediated CMA had lower ratios of fecal secondary/primary BAs from the CA pathway, with DCA and other oxidized keto BAs included in the calculation.[45] Secondary BAs from the CDCA pathway, including LCA, were reported lower in children with food allergy compared to healthy controls as well.[46] Although the secondary BAs and secondary/primary BAs ratios were not significantly different between the two groups in our study, the altered BAs profiles in the CMA-tolerant group likely indicate a more mature GM for secondary BAs production. This may contribute to improved intestinal functions in infants outgrowing CMA, as LCA is known to attenuate disruption in the intestinal barrier.[47]

(B)SCFAs were also altered during the CMA tolerance acquisition process. Butyrate significantly increased from TP0 to TP2 only in the CM-tolerant group (Figure S1). Isobutyrate and isovalerate tended to have group separation at TP1, with a continuous elevation in the CM-tolerant group over time, and a decrease at TP1 in the CM-allergic group (Figure S2). Consistent with our finding, those (B)SCFAs, specifically butyrate, are known for their anti-inflammatory effects,[27,48] and are generally observed to be lower in feces of children with IgE-mediated food allergy.[42,48] Additionally, phenylalanine, phenyllactic acid (PLA#), and desaminotyrosine, which are GM metabolites from amino acids and dietary polyphenols,[49–51] were significantly increased from TP0 and TP2 only in the CM-tolerant group (Figure S1). The significant elevations of these metabolites may promote CM-tolerance acquisition, especially considering the recently recognized anti-inflammatory property of desaminotyrosine.[52,53]

The synbiotic (*B. breve* M-16V, FOS: inulin, oligofructose) significantly altered the levels of aromatic lactic acids, purine metabolites as well as fatty acids and BAs, particularly after six months of intervention. The intervention enhanced ILA and 4-OH-

PLA levels (Figure S3), and their increases from baseline to TP1and TP2 were positively correlated with those of bifidobacteria (Figure 5). This finding aligns with reports that ILA and 4-OH-PLA are metabolites of tryptophan[29,54,55] and tyrosine[29] produced by infant-type *Bifidobacterium* species, including *B. breve*. Earlier published microbiome and metaproteomics analysis of stool samples from the same clinical trial revealed that the synbiotic raised the level of bifidobacteria,[19,35] as well as bifidobacterial Carbohydrate-Active enZymes,[35] known to metabolize FOS.[56] Although the proportion of *Bifidobacterium* was significantly higher in the AAF-S group compared to the AAF group at both time points (Figure S4),[19,35] the increases in ILA and 4-OH-PLA# were significantly higher in the AAF-S group only at TP1. These results suggest that the synbiotic promoted the growth and/or the activity of aromatic lactic acids producers, e.g., infant-type *Bifidobacterium* species, especially at TP1. This can be evidenced by stronger positive correlations between changes in the two aromatic lactic acids and bifidobacteria from baseline to TP1 than to TP2 in the AAF-S group (Figure 5). To validate our observations, *Bifidobacterium* species should be quantified. Alternatively, aromatic lactate dehydrogenase reported to convert tryptophan and tyrosine to respectively ILA and 4-OH-PLA in infant-type *Bifidobacterium* species should be analyzed.[29] The possibility that the ILA and 4-OH-PLA# were produced by some lactic acid bacteria should not be ignored neither.[57,58] Overall, the increased ILA and 4-OH-PLA# levels in the AAF-S group suggest enhanced abundance or activity of infant-type bifidobacteria, supporting the successful synbiotic supplementation together with the microbiome and metaproteomics findings.[19,35] Although the parent study found that the CM-tolerance acquisition after 12 (TP2) and 24 months of synbiotic intervention aligned with natural outgrowth,[19] our findings, along with the reported anti-inflammatory effect of ILA,[25,29,55,59] suggest that the synbiotic intervention may pose beneficial effects on infants' immune system. Further metabolomics studies on larger cohorts are required to verify this hypothesis.

In addition to the increase in ILA and 4-OH-PLA, the synbiotic lowered inosine, guanine, and uridine and raised adenine levels. The same purine-pyrimidine trend was observed in conventionally raised and core microbiota-colonized mice in comparison to

germ-free mice,[60] indicating the importance of the GM in purine and pyrimidine metabolism.[60] A decline of inosine and uridine has also been reported in co-culture of *B. breve* with small intestinal-like epithelial cells.[61] *Lactobacillus brevis,* belonging to the *Lactobacillaceae* family, was found to be elevated in the AAF-S group for the same set of samples[35] and was also reported to have inosine degradation capabilities.[62] To link the purine-pyrimidine metabolism to the gut microbiome, and the role of *Bifidobacterium* spp. and *Lactobacillaceae* spp. herein, more research is required.

The AAF-S intervention lowered LA, ALA#, and OA levels, suggesting high consumption of these fatty acids by gut bacteria. This may be a result of hydration by bacteria of the *Lactobacillus* and *Bifidobacterium* genera[63] or production of conjugated fatty acids.[64–68] *Bifidobacterium* strains, especially *B. breve*, are among the best producers of conjugated linoleic acids[66,67] and conjugated linolenic acids.[66,68]

The synbiotic enhanced the deconjugation of BAs, especially at TP1, where significantly decreased GCDCA and increased CDCA/GCDCA, CA/GCA, and UDCA/GUDCA were observed in the AAF-S compared to AAF group (Figure 4). *Bifidobacterium*, in general, are active bile salt hydrolase (BSH) producers,[69] which perform preferred deconjugation activity on glyco-conjugated BAs.[70] This aligns with our results showing that *Bifidobacterium* changes from baseline correlated negatively with those of GCDCA, and positively with those of CA/GCA and CDCA/GCDCA at TP1 in the AAF-S (Figure S5). These correlations in changes disappeared at TP2, possibly due to increased GM diversity. Compared to TP0, families from other phyla, including Bacteroidetes, Firmicutes, and Proteobacteria, were more abundant at later timepoints in both groups, especially at TP2.[35] These bacteria have also been identified as active BSH producers,[71] thus might eliminate the correlation between the activity of BAs deconjugation and *Bifidobacterium*. Unexpectedly, the increased deconjugation activity of BAs failed to promote the production DCA and LCA. In contrast, although not significant, their levels and ratios to precursors (DCA/CA, LCA/CDCA) were lower in the AAF-S than the AAF group (Figure 4). Considering that the conversion of primary BAs to secondary ones is highly conserved in bacteria with the *bai* operon,[72] and that the host liver can further hydroxylate secondary BAs to tertiary BAs after gut-liver

circulation,[73] it is likely that more complex mechanisms underlie the host-gut metabolism of BAs during the intervention.

Our study has several limitations, including the wide age range of the participants at baseline of 3-13 (9.00 ± 2.90) months. Considering the rapid development of the GM in the first two years of life,[39] the wide age range may obscure the observation of fecal metabolome alterations related to CM-tolerance acquisition and the effect of intervention. Another limitation is the lack of information on the CM-tolerance status at TP1. Knowing the status at TP1 could have aided in the interpretation of CM-tolerance acquisition results. The research carried out for this paper is exploratory due to the small samples size (39 subjects). Increasing the sample size is necessary to verify these findings and would also allow to build LMM and RM-ASCA+ models following the intervention and CM-tolerance acquisition simultaneously. In addition, the parent study concluded that the synbiotic supplementation did not significantly affect CMA-resolution. Thus, in this study we cannot draw any conclusions regarding the clinical benefits of the synbiotic supplementation on CM-tolerance acquisition based on fecal metabolome alterations. Despite those limitations, our study revealed several fecal metabolome pathway alterations which may contribute to CMA outgrowth. Most importantly, we found that the AAF-S significantly altered the fecal metabolome after six months of the intervention, not after 12 months, suggesting that early intervention is required to maximize the effect of synbiotics. These findings aid in understanding the link between IgE-mediated CMA-tolerance acquisition, GM, and synbiotics intervention.

List of abbreviations

CMA: cow's milk allergy; CM: cow's milk; GM: gut microbiome; IgE: immunoglobin E; HMO: human milk oligosaccharide; AAF: amino acid-based formula; FOS: fructooligosaccharides; T_{reg}: regulatory T cell; *B.*: *Bifidobacterium*; LMMs: linear mixed models; RM-ASCA+: repeated measures analysis of variance simultaneous component analysis+; AAF-S: amino acid-based formula with synbiotic; DBPCFC: double-blind, placebo-controlled food challenge; BSCFAs: branched short-chain fatty

acids; QC: quality control; CI: confidence intervals; SCORAD: scoring of atopic dermatitis; BSH: bile salt hydrolase

Acknowledgments

This study was part of the EARLYFIT project (Partnership programme NWO Domain AES-Danone Research & Innovation), funded by the Dutch Research Council (NWO) and Danone Research & Innovation (project number: 16490). Pingping Zhu Would like to acknowledge the China Scholarship Council (CSC, No. 201906240049). Diana M Hendrickx (Wageningen University) is gratefully acknowledged for providing the processed 16S rRNA sequencing data. Pascal Mass (Leiden University) is greatly appreciated for his invaluable assistance in metabolomics data pre-processing. We also thank Jolanda Lambert (Danone Research & Innovation) for project management, Guus Roeselers (Danone Research & Innovation) for his input in the study design, and Simone Eussen (Danone Research & Innovation) for her valuable feedback in manuscript review.

Conflict of interest statement

Harm Wopereis is an employee of Danone Research & Innovation. The project is part of a partnership programme between NWO-TTW and Danone Research & Innovation. The other authors declare that they have no known conflicts of interest.

Reference:

1. S. H. Sicherer, H. A. Sampson, *Journal of Allergy and Clinical Immunology* **2010**, *125*, S116–S125.
2. J. Savage, C. B. Johns, *Immunol Allergy Clin North Am* **2015**, *35*, 45–59.
3. J. D. Flom, S. H. Sicherer, *Nutrients* **2019**, *11*, 1051.
4. A. Høst, *Annals of Allergy, Asthma & Immunology* **2002**, *89*, 33–37.
5. A. A. Schoemaker, A. B. Sprikkelman, K. E. Grimshaw, G. Roberts, L. Grabenhenrich, L. Rosenfeld, S. Siegert, R. Dubakiene, O. Rudzeviciene, M. Reche, A. Fiandor, N. G. Papadopoulos, A. Malamitsi-Puchner, A. Fiocchi, L. Dahdah, S. Th. Sigurdardottir, M. Clausen, A. Stańczyk-Przyłuska, K. Zeman, E. N. C. Mills, D. McBride, T. Keil, K. Beyer, *Allergy* **2015**, *70*, 963–972.
6. S. G. O. Johansson, T. Bieber, R. Dahl, P. S. Friedmann, B. Q. Lanier, R. F. Lockey, C. Motala, J. A. Ortega Martell, T. A. E. Platts-Mills, J. Ring, *Journal of Allergy and Clinical Immunology* **2004**, *113*, 832–836.
7. K. M. Saarinen, A. S. Pelkonen, M. J. Mäkelä, E. Savilahti, *Journal of Allergy and Clinical Immunology* **2005**, *116*, 869–875.
8. J. M. Skripak, E. C. Matsui, K. Mudd, R. A. Wood, *Journal of Allergy and Clinical Immunology* **2007**, *120*, 1172–1177.
9. M. V. Savova, P. Zhu, A. C. Harms, R. G. van der Molen, C. Belzer, D. M. Hendrickx, *Pediatric Allergy and Immunology* **2024**, *35*, e14084.
10. F. Turroni, C. Milani, S. Duranti, C. Ferrario, G. A. Lugli, L. Mancabelli, D. van Sinderen, M. Ventura, *Cell. Mol. Life Sci.* **2018**, *75*, 103–118.
11. H. Kumar, M. C. Collado, H. Wopereis, S. Salminen, J. Knol, G. Roeselers, *Microorganisms* **2020**, *8*, 1855.
12. S. Saturio, A. M. Nogacka, G. M. Alvarado-Jasso, N. Salazar, C. G. De Los Reyes-Gavilán, M. Gueimonde, S. Arboleya, *Microorganisms* **2021**, *9*, 2415.
13. J. C. C. Davis, S. M. Totten, J. O. Huang, S. Nagshbandi, N. Kirmiz, D. A. Garrido, Z. T. Lewis, L. D. Wu, J. T. Smilowitz, J. B. German, D. A. Mills, C. B. Lebrilla, *Molecular & Cellular Proteomics* **2016**, *15*, 2987–3002.

14. A. Ioannou, J. Knol, C. Belzer, *Frontiers in Microbiology* **2021**, *12*.
15. Y. Vandenplas, H. A. Brough, A. Fiocchi, M. Miqdady, Z. Munasir, S. Salvatore, N. Thapar, C. Venter, M. C. Vieira, R. Meyer, *JAA* **2021**, *Volume 14*, 1243–1256.
16. Y. Vandenplas, E. D. Greef, G. Veereman, *Gut Microbes* **2014**, *5*, 681–687.
17. W. Jing, Q. Liu, W. Wang, *Journal of Food Biochemistry* **2020**, *44*, e13489.
18. M. Mennini, S. Reddel, F. Del Chierico, S. Gardini, A. Quagliariello, P. Vernocchi, R. Luigi Valluzzi, V. Fierro, C. Riccardi, T. Napolitano, A. Giovanni Fiocchi, L. Putignani, S. Cucchiara, L. Stronati, *International Journal of Molecular Sciences Article* **2021**.
19. P. Chatchatee, A. Nowak-Wegrzyn, L. Lange, S. Benjaponpitak, K. W. Chong, P. Sangsupawanich, M. T. J. van Ampting, M. M. Oude Nijhuis, L. F. Harthoorn, J. E. Langford, J. Knol, K. Knipping, J. Garssen, V. Trendelenburg, R. Pesek, C. M. Davis, A. Muraro, M. Erlewyn-Lajeunesse, A. T. Fox, L. J. Michaelis, K. Beyer, L. Noimark, G. Stiefel, U. Schauer, Hamelman, D. Peroni, Boner, *Journal of Allergy and Clinical Immunology* **2022**, *149*, 650-658.e5.
20. M. Viljanen, M. Kuitunen, T. Haahtela, K. Juntunen-Backman, R. Korpela, E. Savilahti, *Pediatr Allergy Immunol* **2005**, *16*, 65–71.
21. A. W. Burks, L. F. Harthoorn, M. T. J. Van Ampting, M. M. Oude Nijhuis, J. E. Langford, H. Wopereis, S. B. Goldberg, P. Y. Ong, B. J. Essink, R. B. Scott, B. M. Harvey, *Pediatric Allergy and Immunology* **2015**, *26*, 316–322.
22. R. Verma, C. Lee, E.-J. Jeun, J. Yi, K. S. Kim, A. Ghosh, S. Byun, C.-G. Lee, H.-J. Kang, G.-C. Kim, C.-D. Jun, G. Jan, C.-H. Suh, J.-Y. Jung, J. Sprent, D. Rudra, C. De Castro, A. Molinaro, C. D. Surh, S.-H. Im, *Sci. Immunol.* **2018**, *3*, eaat6975.
23. T. Ruohtula, M. C. de Goffau, J. K. Nieminen, J. Honkanen, H. Siljander, A.-M. Hämäläinen, A. Peet, V. Tillmann, J. Ilonen, O. Niemelä, G. W. Welling, M. Knip, H. J. Harmsen, O. Vaarala, *Frontiers in Immunology* **2019**, *10*.
24. B. Cukrowska, J. B. Bierła, M. Zakrzewska, M. Klukowski, E. Maciorkowska, *Nutrients* **2020**, *12*, 946.
25. B. M. Henrick, L. Rodriguez, T. Lakshmikanth, C. Pou, E. Henckel, A. Arzoomand, A. Olin, J. Wang, J. Mikes, Z. Tan, Y. Chen, A. M. Ehrlich, A. K. Bernhardsson, C. H. Mugabo, Y. Ambrosiani, A. Gustafsson, S. Chew, H. K. Brown, J. Prambs, K. Bohlin, R. D. Mitchell, M. A. Underwood, J. T. Smilowitz, J. B. German, S. A. Frese, P. Brodin, *Cell* **2021**, *184*, 3884-3898.e11.
26. A. Belenguer, S. H. Duncan, A. G. Calder, G. Holtrop, P. Louis, G. E. Lobley, H. J. Flint, *Appl Environ Microbiol* **2006**, *72*, 3593–3599.
27. M. T. Siddiqui, G. A. Cresci, *JIR* **2021**, *Volume 14*, 6025–6041.
28. N. Acevedo, B. Alashkar Alhamwe, L. Caraballo, M. Ding, A. Ferrante, H. Garn, J. Garssen, C. S. Hii, J. Irvine, K. Llinás-Caballero, J. F. López, S. Miethe, K. Perveen, E. Pogge von Strandmann, M. Sokolowska, D. P. Potaczek, B. C. A. M. van Esch, *Nutrients* **2021**, *13*, 724.
29. M. F. Laursen, M. Sakanaka, N. von Burg, U. Mörbe, D. Andersen, J. M. Moll, C. T. Pekmez, A. Rivollier, K. F. Michaelsen, C. Mølgaard, M. V. Lind, L. O. Dragsted, T. Katayama, H. L. Frandsen, A. M. Vinggaard, M. I. Bahl, S. Brix, W. Agace, T. R. Licht, H. M. Roager, *Nat Microbiol* **2021**, *6*, 1367–1382.
30. D. M. Hendrickx, R. An, S. Boeren, S. K. Mutte, J. M. Lambert, C. Belzer, *Sci Rep* **2023**, *13*, 12029.
31. H. Wopereis, K. Sim, A. Shaw, J. O. Warner, J. Knol, J. S. Kroll, *Journal of Allergy and Clinical Immunology* **2018**, *141*, 1334-1342.e5.
32. F. Hosseinkhani, A.-C. Dubbelman, N. Karu, A. C. Harms, T. Hankemeier, *Metabolites* **2021**, *11*, 364.
33. P. Zhu, A.-C. Dubbelman, C. Hunter, M. Genangeli, N. Karu, A. Harms, T. Hankemeier, *J. Am. Soc. Mass Spectrom.* **2024**, *35*, 590–602.
34. R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, Y. Ni, *Sci Rep* **2018**, *8*, 663.
35. D. M. Hendrickx, R. An, S. Boeren, S. K. Mutte, the P. study Team, H. Wopereis, C. Belzer, *Beneficial Microbes* **2023**, *14*, 269–280.
36. A. H. Jarmund, T. S. Madssen, G. F. Giskeødegård, *Frontiers in Molecular Biosciences* **2022**, *9*.
37. S. A. Heleno, A. Martins, M. J. R. P. Queiroz, I. C. F. R. Ferreira, *Food Chemistry* **2015**, *173*, 501–513.
38. E. De Peretti, E. Mappus, *The Journal of Clinical Endocrinology & Metabolism* **1983**, *57*, 550–556.
39. E. A. Holzhausen, N. Shen, B. Chalifour, V. Tran, Z. Li, J. A. Sarnat, H. H. Chang, D. P. Jones, M. I. Goran, D. Liang, T. L. Alderete, *Sci Rep* **2023**, *13*, 1886.
40. K. C. Fragkos, A. Forbes, *United European Gastroenterol J* **2018**, *6*, 181–191.
41. M. Niewiem, U. Grzybowska-Chlebowczyk, *Nutrients* **2022**, *14*, 1893.
42. R. Francavilla, M. Calasso, L. Calace, S. Siragusa, M. Ndagijimana, P. Vernocchi, L. Brunetti, G. Mancino, G. Tedeschi, E. Guerzoni, F. Indrio, L. Laghi, V. L. Miniello, M. Gobbetti, M. De Angelis, *Pediatric Allergy and Immunology* **2012**, *23*, 420–427.
43. P. A. Shah, C. J. Park, M. P. Shaughnessy, R. A. Cowles, *Cellular and Molecular Gastroenterology and Hepatology* **2021**, *12*, 1093–1104.

44. S. Haq, J. A. Grondin, W. I. Khan, *The FASEB Journal* **2021**, *35*, e21888.
45. E. De Paepe, V. Plekhova, P. Vangeenderhuysen, N. Baeck, D. Bullens, T. Claeys, M. De Graeve, K. Kamoen, A. Notebaert, T. Van de Wiele, W. Van Den Broeck, K. Vanlede, M. Van Winckel, L. Vereecke, C. Elliott, E. Cox, L. Vanhaecke, *Allergy* **2024**, *79*, 949–963.
46. S.-Y. Lee, Y. M. Park, H. J. Yoo, S.-H. Lee, E. J. Choi, E. Y. Baek, K. B. Song, J. Yoon, S.-J. Hong, *Pediatric Allergy and Immunology* **2023**, *34*, e14003.
47. N. Calzadilla, S. M. Comiskey, P. K. Dudeja, S. Saksena, R. K. Gill, W. A. Alrefai, *Front. Immunol.* **2022**, *13*, 1021924.
48. E. De Paepe, L. Van Gijseghem, M. De Spiegeleer, E. Cox, L. Vanhaecke, *Molecular Nutrition & Food Research* **2021**, *65*, 2100536.
49. N. V. Beloborodov, A. S. Khodakova, I. T. Bairamov, A. Yu. Olenin, *Biochemistry Moscow* **2009**, *74*, 1350–1355.
50. A. Rechner, *Free Radical Biology and Medicine* **2004**, *36*, 212–225.
51. L. Schoefer, R. Mohan, A. Schwiertz, A. Braune, M. Blaut, *Appl Environ Microbiol* **2003**, *69*, 5849–5854.
52. Y. Wei, J. Gao, Y. Kou, M. Liu, L. Meng, X. Zheng, S. Xu, M. Liang, H. Sun, Z. Liu, Y. Wang, *The FASEB Journal* **2020**, *34*, 16117–16128.
53. A. L. Steed, G. P. Christophi, G. E. Kaiko, L. Sun, V. M. Goodwin, U. Jain, E. Esaulova, M. N. Artyomov, D. J. Morales, M. J. Holtzman, A. C. M. Boon, D. J. Lenschow, T. S. Stappenbeck, *Science* **2017**, *357*, 498–502.
54. T. Sakurai, T. Odamaki, J. Xiao, *Microorganisms* **2019**, *7*, 340.
55. A. M. Ehrlich, A. R. Pacheco, B. M. Henrick, D. Taft, G. Xu, M. N. Huda, D. Mishchuk, M. L. Goodson, C. Slupsky, D. Barile, C. B. Lebrilla, C. B. Stephensen, D. A. Mills, H. E. Raybould, *BMC Microbiol* **2020**, *20*, 357.
56. H. Tanno, T. Fujii, K. Hirano, S. Maeno, T. Tonozuka, M. Sakamoto, M. Ohkuma, T. Tochio, A. Endo, *Gut Microbes* **2021**, *13*, 1–20.
57. F. Valerio, P. Lavermicocca, M. Pascale, A. Visconti, *FEMS Microbiology Letters* **2004**, *233*, 289–295.
58. T. Pan, Z. Pei, Z. Fang, H. Wang, J. Zhu, H. Zhang, J. Zhao, W. Chen, W. Lu, *Front. Cell. Infect. Microbiol.* **2023**, *13*.
59. D. Meng, E. Sommella, E. Salviati, P. Campiglia, K. Ganguli, K. Djebali, W. Zhu, W. A. Walker, *Pediatr Res* **2020**, *88*, 209–217.
60. K. Kasahara, R. L. Kerby, Q. Zhang, M. Pradhan, M. Mehrabian, A. J. Lusis, G. Bergström, F. Bäckhed, F. E. Rey, *Cell Host Microbe* **2023**, *31*, 1038-1053.e10.
61. A. Sen, T. Nishimura, S. Yoshimoto, K. Yoshida, A. Gotoh, T. Katoh, Y. Yoneda, T. Hashimoto, J.-Z. Xiao, T. Katayama, T. Odamaki, *Front. Microbiol.* **2023**, *14*, DOI 10.3389/fmicb.2023.1155438.
62. H. Wang, L. Mei, Y. Deng, Y. Liu, X. Wei, M. Liu, J. Zhou, H. Ma, P. Zheng, J. Yuan, M. Li, *Nutrition* **2019**, *62*, 63–73.
63. S. Serra, D. De Simeis, A. Castagna, M. Valentino, *Catalysts* **2020**, *10*, 154.
64. L. Alonso, E. P. Cuesta, S. E. Gilliland, *Journal of Dairy Science* **2003**, *86*, 1941–1946.
65. A. S. Salsinha, L. L. Pimentel, A. L. Fontes, A. M. Gomes, L. M. Rodríguez-Alcalá, *Microbiology and Molecular Biology Reviews* **2018**, *82*, 10–1128.
66. L. Gorissen, K. Raes, S. Weckx, D. Dannenberger, F. Leroy, L. De Vuyst, S. De Smet, *Appl Microbiol Biotechnol* **2010**, *87*, 2257–2266.
67. Y. Mei, H. Chen, B. Yang, J. Zhao, H. Zhang, W. Chen, *International Journal of Food Microbiology* **2022**, *369*, 109593.
68. H. G. Park, H. T. Cho, M.-C. Song, S. B. Kim, E. G. Kwon, N. J. Choi, Y. J. Kim, *J. Agric. Food Chem.* **2012**, *60*, 3204–3210.
69. H. Tanaka, K. Doesburg, T. Iwasaki, I. Mierau, *Journal of Dairy Science* **1999**, *82*, 2530–2535.
70. G.-B. Kim, S.-H. Yi, B. H. Lee, *Journal of Dairy Science* **2004**, *87*, 258–266.
71. Z. Song, Y. Cai, X. Lao, X. Wang, X. Lin, Y. Cui, P. K. Kalavagunta, J. Liao, L. Jin, J. Shang, J. Li, *Microbiome* **2019**, *7*, 9.
72. D. V. Guzior, R. A. Quinn, *Microbiome* **2021**, *9*, 140.
73. J. Zhang, L.-Z. Gao, Y.-J. Chen, P.-P. Zhu, S.-S. Yin, M.-M. Su, Y. Ni, J. Miao, W.-L. Wu, H. Chen, K. L. R. Brouwer, C.-X. Liu, L. Xu, W. Jia, K. Lan, *Drug Metab Dispos* **2019**, *47*, 283–294.

Supplementary Material

Chemicals

Methyl tert-butyl ether (MTBE, ≥99.8%) and ammonium formate (≥99.0%) were purchased from Sigma Aldrich (St. Louis, United States). LC-MS-grade methanol (MeOH), isopropanol and formic acid (FA) were purchased from Biosolve B.V. (Valkenswaard, Netherlands). LC-MS grade acetonitrile was purchased from Actu-all chemicals (Randmeer, The Netherlands) and Biosolve B.V. (Valkenswaard, Netherlacnjugnds). Purified water was obtained from a Milli-Q PF Plus system (Merck Millipore, Burlington, United States). List of the isotopically labelled standards (SILs), including supplier details, can be found in Table S1.

Sample preparation

Briefly, 72 µL of water and 216 µL MeOH, containing stable isotopically labelled standards (SILs) (Table S1), were added to the 20 mg dry-weight fecal sample. After a 3-minute vortex mixing (Marshall Scientific, Cambridge, UK) 120 µL ice-cold MTBE was added, followed by another 3-minute vortex mixing. Following a brief centrifugation (30s, 100g, 4 °C), 200 µL of water and 168 µL of MTBE were added. The samples were vortex mixed for another 3 min, incubated at 4°C for 10 minutes until centrifugation (20 min, 16 000g, 4°C) inducing aqueous and organic layer separation. All solvents used during the LLE were ice-cold and vortex mixing was always at maximum speed. Following layer separation, each layer was transferred to an Eppendorf tube, followed by 5 and 2.5 minutes of centrifugation (16000g, 4°C) for aqueous and organic layers respectively. After extraction, 150 µL of the aqueous layer was aliquoted for polar to semi-polar metabolites analysis, while 48.8 µL of aqueous and 28.8 µL of organic layer was combined for the bile and fatty acids analysis. The aliquots were dried in a Speedvac (Labcono, USA) and stored at -80°C. Prior to LC-MS analysis, the extracts were reconstituted in 50 µL of 0.1% FA in water for polar to semi-polar metabolites analysis, and 200 µL of MeOH for the bile and fatty acids analysis. The reconstitution solvents contained different SILs (Table S1).

Quality Control

Samples were randomized into two batches, with those from the same subject prepared and measured in the same batch. For the preparation of the quality control sample, 30 study samples were weighed and extracted. After the extraction, equal volumes of each layer were taken from each sample and pooled, resulting in pooled QC aqueous and organic layers. Those pooled layers were used to prepare QC samples for each platform. The LLE and aliquoting steps were performed as described in Sample preparation.

LC-MS analysis of polar to semi polar metabolites

Analysis of polar to semi-polar metabolites were performed with a Shimadzu Nexera X2 LC system coupled to a TripleTOF 6600 mass spectrometer (SCIEX, Foster City, CA, USA), as described previously. Briefly, the LC separation was carried out at 40 °C using a Waters Acquity UPLC HSS T3 column (1.8 µm, 2.1 mm × 100 mm) with pre-column in-line stainless steel filter (0.3 µm, Agilent Technologies, Waldbronn, Germany). The mobile phase A was 0.1% FA in water, and the mobile phase B was 0.1% FA in ACN (Actu-all chemicals). With a flow rate of 0.4 mL min^{-1} and 1 µL of injection volume, the gradient starts at 100% A; 0–0.5 min 80% A; 0.5–2.5 min 2% A; 2.5–7.5 min 2% A; 7.5–12 min 2% A; 12 – 15 100% A. The data were acquired under full scan mode over the *m/z* range of 60-800 Da with Analyst TF software 1.7.1 (SCIEX) in negative and positive ionization modes. The preferred ionization mode for metabolites detectable in both polarities was chosen based on lower RSD% and higher signal-to-noise ratio of the QC samples.

LC-MS analysis of bile acid and fatty acids

Analysis of bile and fatty acids was performed on an UPLC-TOF/MS system consisting of ExionLC™ AC UHPLC system and SCIEX ZenoTOF 7600 system (Darmstadt, Germany) equipped with an IonDrive™ Turbo V Source, operated in negative ESI mode. The ion source conditions were as follows: spray voltage of 4.5 kV, capillary temperature of 550°C, ion source gas 1 50 psi, ion source gas 2 50 psi, curtain gas 35 psi, CAD gas 7 psi. The MS data was acquired under full scan mode over the m/z range of 200-900 Da. Accumulation time was set to 0.25 s, delustering potential to -70V and collision energy to -10eV. Chromatographic separation was performed on a Waters Acquity UPLC HSS T3 column (1.8 µm, 2.1 mm × 100 mm) with pre-column in-line stainless

steel filter (0.3 µm, Agilent Technologies, Waldbronn, Germany). The flow rate was set at 0.4 ml min^{-1}, the column was kept at 45 °C, injection volume at 2 µL. Mobile phase A consisted of 10 mM ammonium formate in water/ACN (Biosolve B.V) (95:5, v:v), while mobile phase B was 10 mM ammonium formate in MeOH/water (99:1, v:v). The gradient was as follows: starting at 0% B; 0–0.2 min 70% B; 0.2–7.5 min 100% B; 7.5–11.5 min 100% B; 11.5–11.6 min 0% B; 11.6 – 15 0% B. Isopropanol was used as an external rinsing solution (2 s sip time + rinse port). The flow was directed to waste in the first minute of the run. The autosampler temperature was set at 10 °C. Data acquisition was carried out on SCIEX OS 2.1.6.

Visualization RM-ASCA+

Visualization of the longitudinal metabolomic alterations was achieved using RM-ASCA+, which is an extension of LMMs for multivariate data. In the first step, LMMs are used to decompose the response matrix into effect matrices. The effect matrices are then analyzed using principal component analysis (PCA), and the results are summarized into PCA scores and loadings. The LMMs used for RM-ASCA+ were the LMMs used for the univariate analysis. The visualized effect matrices included the time effect matrix (‘time’) which shows time development of the reference group over time. The interaction matrix (‘time:group’) and the group-interaction matrix (‘group + time:group’) both show the deviations of the study group compared to the reference group over time with the latter also displaying the baseline differences. Lastly, the combined matrix (‘time + time:group’ or ‘time + group + time:group’) shows the time development of both the study and the reference group.

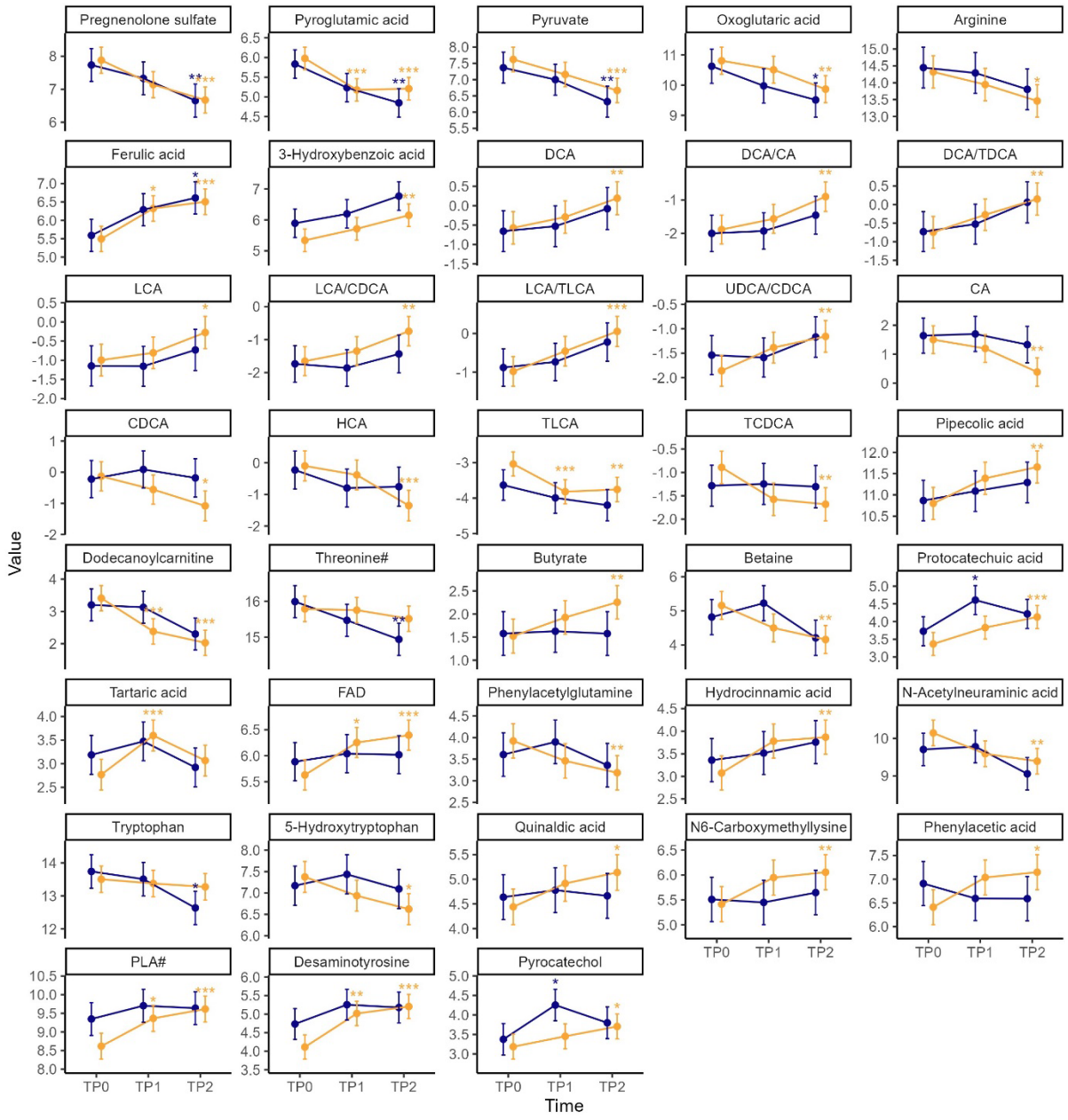


Figure S1. Marginal means estimated from the LMMs for participants who acquired tolerance (CM-tolerant, orange) and those that remained allergic (CM-allergic, blue). Only the metabolites for which pairwise comparison in time was found significant are plotted. The q-values are based on the marginal mean comparison to TP0 for each group, $q < 0.01$ (***), $q < 0.05$ (**), $q < 0.1$ (*).

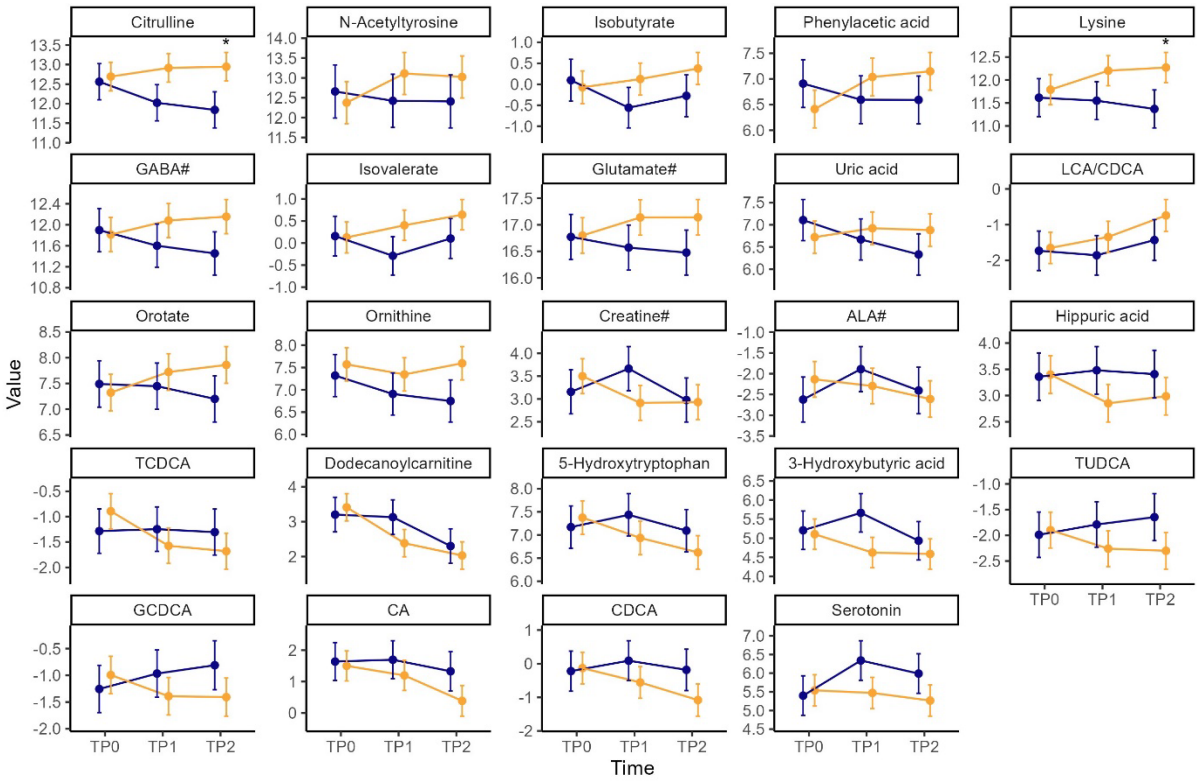


Figure S2. Marginal means estimated from the LMMs for participants who acquired tolerance (CM-tolerant) and those that remained allergic (CM-allergic). The metabolites with top loadings in PC1 of the RM-ASCA+ interaction matrix are plotted. The q-values are based on the marginal mean comparison between the groups at each time point, q <0.1 (*).

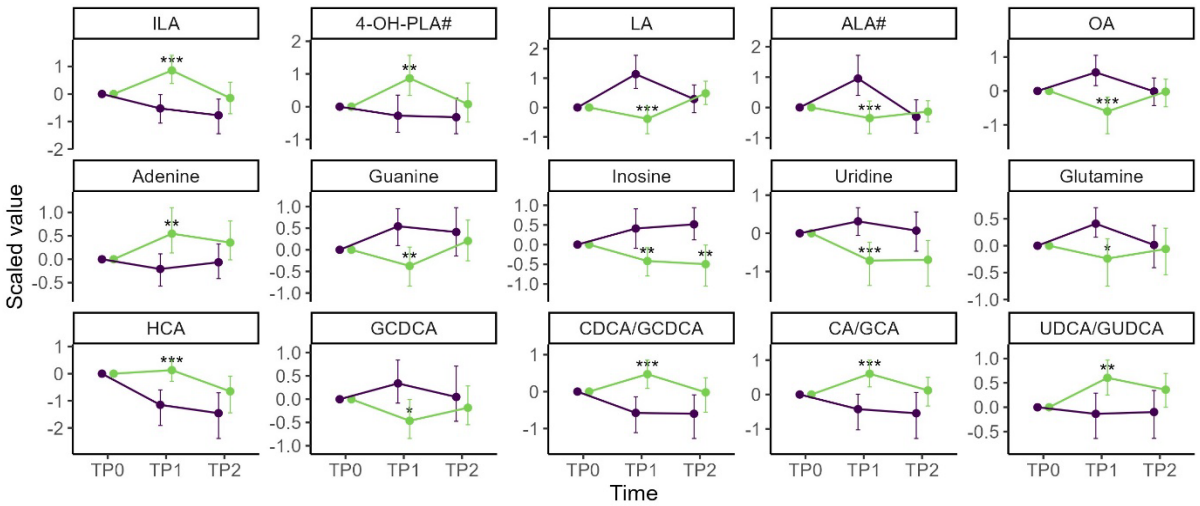


Figure S3. Marginal means estimated from the LMMs for AAF and AAF-S group. Only the metabolites for which an interaction coefficient was found significant are plotted. The response has been scaled. The q-values are based on/denote the significant

between-group change in the within-group change from baseline. q < 0.01 (***), q < 0.05 (**), q <0.1 (*)

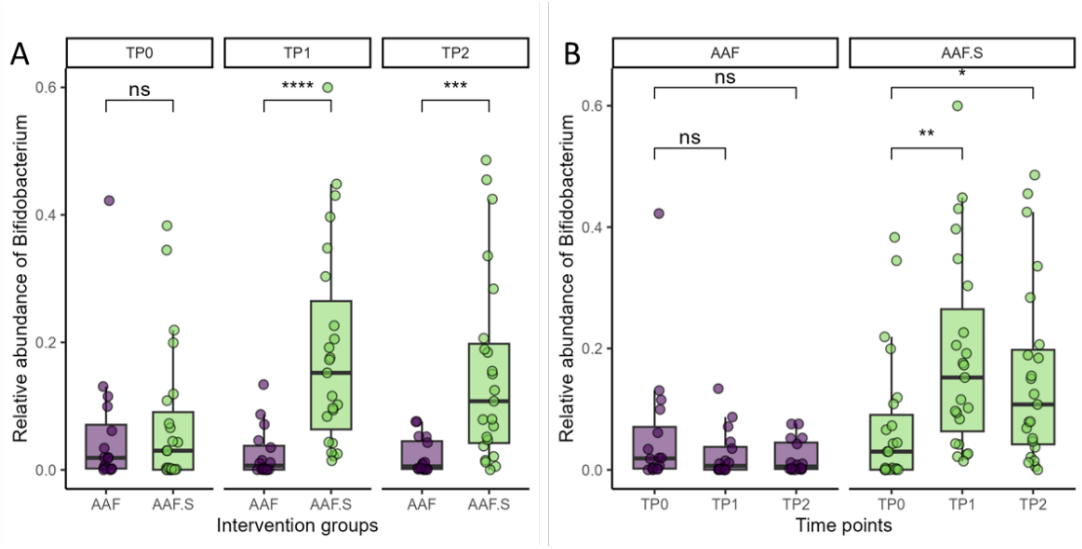


Figure S4. Relative abundance of *Bifidobacterium* comparisons between AAF and AAF-S groups at each time point (A), and between time points in each group (B). Statistical significance was evaluated with two-side unpaired t-tests; p > 0.05 (ns), p ≤ 0.05 (*), p ≤ 0.01 (**), p ≤ 0.001 (***), p ≤ 0.0001 (****).

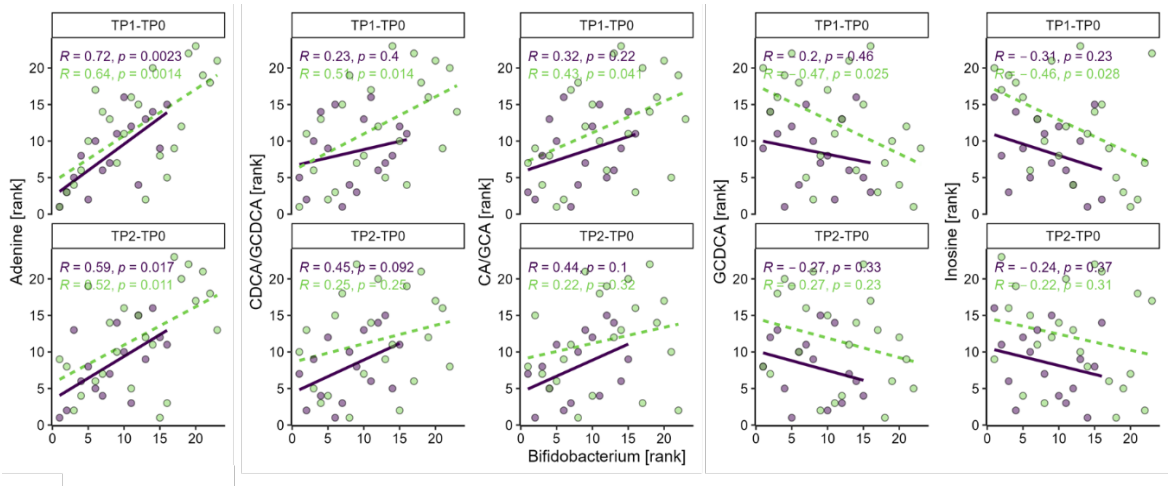


Figure S5. Spearman's rank correlations between the changes in *Bifidobacterium* and adenine, CDCA/GCDCA, CA/GCA. GCDCA, inosine in AAF (purple solid line) and AAF-S (green dashed line) groups from baseline to TP1 (TP1-TP0) and TP2 (TP2-TP0). The rank of the changes in metabolite response and relative abundance of *Bifidobacterium* within each group were used for plotting.

Table S1: General information and solution preparation for all the stable isotopically labeled standards (SILs)

Compound Name	Compound Formula	Supplier	product number	Spiked concentration (µM)	Usage
choline-d_4	C5H9D4NO	CDN	D-2464	16.00	Polar to semi-polar metabolites platform IS
cytidine-$^{15}N_3$	C9H13[15N]3O5	cambridge Isotope laboratories	NLM-3797-50	64.50	Polar to semi-polar metabolites platform IS
DL-leucine-d3	C6H10D3NO2	CDN	D-2400	56.00	Polar to semi-polar metabolites platform IS
DL-proline-d_7	C5H2D7NO2	cambridge Isotope laboratories	DLM-2657-0	58.00	Polar to semi-polar metabolites platform IS
hippuric acid-d_5	C9H4D5NO3	chem Cruz	sc-490158	42.00	Polar to semi-polar metabolites platform IS
hypoxanthine-d_3	C5D3HN4O	cambridge Isotope laboratories	DLM-2923-0.1	12.00	Polar to semi-polar metabolites platform IS
indole-d_5-3-acetic acid	C10H4D5NO2	TRC	I577344	44.00	Polar to semi-polar metabolites platform IS
L-tryptophan-d_3	C11H9D3N2O2	CDN	D-7419	20.00	Polar to semi-polar metabolites platform IS
L-tyrosine-$^{13}C_9$-^{15}N	[13C]9H11[15N]O3	cambridge Isotope laboratories	CNLM-439-H-0.1	26.00	Polar to semi-polar metabolites platform IS
octanoyl-l-carnitine-d_3	C15H26D3NO4	CDN	D-6651	0.40	Polar to semi-polar metabolites platform IS
propionyl-L-carnitine-(n-methyl-d_3)	C10H16D3NO4	CDN	D-6651	4.00	Polar to semi-polar metabolites platform IS
quinaldic acid-d_6	C10HD6NO2	CDN	D-6514	10.00	Polar to semi-polar metabolites platform IS
u-^{15}N-guanosine	C10H13[15N]5O5	Silantes	125303603	114.00	Polar to semi-polar metabolites platform IS
4-hydroxyphenylactic acid-d_6	C8H2D6O3	TRC	H949062	97.78	Polar to semi-polar metabolites platform spiked in reconstitution solution
fludrocortisone-d_5	C21H24D5FO5	TRC	F428102	0.76	Polar to semi-polar metabolites platform spiked in reconstitution solution
caffeine-d_9	C8HD9N4O2	TRC	C080102	2.77	Polar to semi-polar metabolites platform spiked in reconstitution solution
valine-d_8	C5H3D8NO2	cambridge Isotope laboratories	DLM-488	42.12	Polar to semi-polar metabolites platform spiked in reconstitution solution
Lithocholic acid-d_4 LCA-d_4	C24H36D4O3	CDN Isotopes	u501p49	200	Bile and fatty acids platform
cholic acid-d_4 (CA-d_4)	C24H36D4O5	CDN Isotopes	z75p40	65	Bile and fatty acids platform
Deoxycholic acid-d_4 (DCA-d_4)	C24H36D4O4	CDN Isotopes	w133p40	100	Bile and fatty acids platform
Ursodeoxycholic acid-d_4 (UDCA-d_4)	C24H36D4O4	CDN Isotopes	v275p43	100	Bile and fatty acids platform
Glycocholic acid-d_4 (GCA-d_4)	C26H39D4NO6	Cayman Chemical	21889	37.5	Bile and fatty acids platform
Glycoursodeoxycholic Acid-d_4 (GUDCA-d_4)	C26H39D4NO5	Cayman Chemical	21890	37.5	Bile and fatty acids platform
Tauroursodeoxycholic acid-d_5 (TUDCA-d_5)	C26H40D5NO6S	Santa-Cruz Biotechnology	sc-220192	10	Bile and fatty acids platform
Arachidonic Acid-d8 (AA-d_8)	C20H24D8O2	Cayman Chemical	390010	500	Bile and fatty acids platform
Oleic Acid-d_{17} (OA-d_{17})	C18H17D17O2	Cayman Chemical	9000432	1	Bile and fatty acids platform spiked in reconstitution solution
12-[[(cyclohexylamino)carbonyl]amino]-dodecanoic acid (CUDA)	C19H36N2O3	Cayman Chemical	10007923	0.5	Bile and fatty acids platform spiked in reconstitution solution

Table S2: Fisher's exact test results for the metabolites with missingness above 20% in the tolerant-allergy and the treatment groups

Compound name	Time point	CM-Allergic NA(%)	CM-Tolerant NA(%)	P values	model type
Vanillic acid	6	0.0	33.3	0.01457	CM tolerance-allergy
Valerate	6	73.3	33.3	0.02248	CM tolerance-allergy
2-Methylglutaric acid	6	80.0	41.7	0.02441	CM tolerance-allergy
Vanillactic acid	6	13.3	50.0	0.03785	CM tolerance-allergy
Agmatine	6	13.3	50.0	0.03785	CM tolerance-allergy
Creatinine	6	26.7	62.5	0.04837	CM tolerance-allergy
Agmatine	12	20.0	62.5	0.01950	CM tolerance-allergy
Asparagine	12	80.0	41.7	0.02441	CM tolerance-allergy

Compound name	Time point	AAF NA(%)	AAF-S NA(%)	P values	model type
L-Acetylcarnitine	0	62.5	21.7	0.0184	intervention
1,7-Dimethyluric acid	6	62.5	8.7	0.0009	intervention
Xanthosine	6	6.3	52.2	0.0047	intervention
3-Methylindole	6	50.0	8.7	0.0073	intervention
3-Methylhistidine	6	50.0	8.7	0.0073	intervention
GLCA	6	43.8	82.6	0.0172	intervention
2-Ketobutyric acid	6	0.0	30.4	0.0287	intervention
Saccharopine	6	31.3	4.3	0.0332	intervention
Dopamine	6	12.5	47.8	0.0371	intervention
Guanidinosuccinic acid	12	56.3	17.4	0.0172	intervention
Xanthosine	12	6.3	39.1	0.0279	intervention
GLCA	12	20.0	59.1	0.0409	intervention
TLCA-3S	12	26.7	63.6	0.0448	intervention

Table S3: Target list and abbreviations for final data ananlysis

Platform (ionization mode)	Compound_name_reported	abbreviations
Polar to semi polar (negative)	1-Methyluric acid	1-Methyluric acid
Polar to semi polar (negative)	2,5-Furandicarboxylic acid	2,5-Furandicarboxylic acid
Polar to semi polar (negative)	Deoxyinosine	Deoxyinosine
Polar to semi polar (negative)	Deoxyuridine	Deoxyuridine
Polar to semi polar (negative)	ortho-Hydroxyphenylacetic acid	ortho-Hydroxyphenylacetic acid
Polar to semi polar (negative)	Protocatechuic acid	Protocatechuic acid
Polar to semi polar (negative)	Dihydrocaffeic acid/3-hydroxy-3-(3-hydroxyphenyl)propanoic acid/Hydroxyphenyllactic acid	4-OH-PLA#
Polar to semi polar (negative)	3-Hydroxybenzoic acid	3-Hydroxybenzoic acid
Polar to semi polar (negative)	3-Hydroxybutyric acid	3-Hydroxybutyric acid
Polar to semi polar (negative)	3-Methyl-2-oxovaleric acid	3-Methyl-2-oxovaleric acid
Polar to semi polar (negative)	3-Methylxanthine/1-Methylxanthine/ 7-Methylxanthine	3-Methylxanthine/1-Methylxanthine/7-Methylxanthine
Polar to semi polar (negative)	Phenyllactic acid/3-(3-Hydroxyphenyl)propanoic acid	PLA#
Polar to semi polar (negative)	Hydrocinnamic acid	Hydrocinnamic acid
Polar to semi polar (negative)	4-Hydroxybenzoic acid	4-Hydroxybenzoic acid
Polar to semi polar (negative)	4-Hydroxycinnamic acid	4-Hydroxycinnamic acid
Polar to semi polar (negative)	p-Hydroxyphenylacetic acid/Mandelic acid	p-Hydroxyphenylacetic acid#
Polar to semi polar (negative)	Desaminotyrosine	Desaminotyrosine
Polar to semi polar (negative)	4-Pyridoxic acid	4-Pyridoxic acid
Polar to semi polar (negative)	Pyroglutamic acid	Pyroglutamic acid
Polar to semi polar (negative)	alpha-Aminobutyric acid/gamma-Aminobutyric acid/3-Aminoisobutanoic acid/Dimethylglycine	GABA#
Polar to semi polar (negative)	Argininosuccinic acid	Argininosuccinic acid
Polar to semi polar (negative)	Ascorbic acid	Ascorbate
Polar to semi polar (negative)	Carnosine	Carnosine
Polar to semi polar (negative)	Citric acid	Citrate
Polar to semi polar (negative)	Gluconic acid	Gluconate
Polar to semi polar (negative)	Flavin adenine dinucleotide	FAD
Polar to semi polar (negative)	Glutamine	Glutamine
Polar to semi polar (negative)	Glycine	Glycine
Polar to semi polar (negative)	Glycolic acid	Glycolate
Polar to semi polar (negative)	Guanine	Guanine
Polar to semi polar (negative)	Hippuric acid	Hippuric acid
Polar to semi polar (negative)	Histidine	Histidine
Polar to semi polar (negative)	Indolelactic acid	ILA
Polar to semi polar (negative)	Indoxyl glucoside	Indoxyl glucoside
Polar to semi polar (negative)	2-Hydroxyethanesulfonate	2-Hydroxyethanesulfonate
Polar to semi polar (negative)	Isobutyrylglycine	Isobutyrylglycine

Polar to semi polar (negative)	Oxoglutaric acid	Oxoglutaric acid
Polar to semi polar (negative)	Lysine	Lysine
Polar to semi polar (negative)	Malic acid	Malate
Polar to semi polar (negative)	Methionine.sulfoxide	Methionine sulfoxide
Polar to semi polar (negative)	myo-Inositol/ Galactose/ Fructose	Fructose#
Polar to semi polar (negative)	N-alpha-Acetylarginine	N-alpha-Acetylarginine
Polar to semi polar (negative)	N-Acetylglutamine	N-Acetylglutamine
Polar to semi polar (negative)	N-Acetylneuraminic acid	N-Acetylneuraminic acid
Polar to semi polar (negative)	N-Acetylserine	N-Acetylserine
Polar to semi polar (negative)	N-Acetyltryptophan	N-Acetyltryptophan
Polar to semi polar (negative)	N2-gamma-Glutamylglutamine	N2-gamma-Glutamylglutamine
Polar to semi polar (negative)	N6-Carboxymethyllysine	N6-Carboxymethyllysine
Polar to semi polar (negative)	O-Acetylserine/Glutamic acid	Glutamate#
Polar to semi polar (negative)	Orotate	Orotate
Polar to semi polar (negative)	p-Cresol	p-Cresol
Polar to semi polar (negative)	p-Cresol sulfate	p-Cresol sulfate
Polar to semi polar (negative)	Pantothenic acid	Pantothenic acid
Polar to semi polar (negative)	Phenylacetic acid	Phenylacetic acid
Polar to semi polar (negative)	Phenylacetylglutamine	Phenylacetylglutamine
Polar to semi polar (negative)	Phenylpropionylglycine	Phenylpropionylglycine
Polar to semi polar (negative)	Pregnenolone sulfate	Pregnenolone sulfate
Polar to semi polar (negative)	Pseudouridine	Pseudouridine
Polar to semi polar (negative)	Pyrocatechol	Pyrocatechol
Polar to semi polar (negative)	Pyruvate	Pyruvate
Polar to semi polar (negative)	Serine	Serine
Polar to semi polar (negative)	Syringic acid	Syringic acid
Polar to semi polar (negative)	Tartaric acid	Tartaric acid
Polar to semi polar (negative)	Taurine	Taurine
Polar to semi polar (negative)	Thymidine	Thymidine
Polar to semi polar (negative)	trans-Aconitic acid	trans-Aconitic acid
Polar to semi polar (negative)	Ferulic acid	Ferulic acid
Polar to semi polar (negative)	Tryptophan	Tryptophan
Polar to semi polar (negative)	Uric acid	Uric acid
Polar to semi polar (negative)	Uridine	Uridine
Polar to semi polar (negative)	Valine	Valine
Polar to semi polar (negative)	Xanthine	Xanthine
Polar to semi polar (negative)	Xylulose	Xylulose
Polar to semi polar (positive)	1-Methyladenosine/N6-Methyladenosine/2'-O-Methyladenosine	1-Methyladenosine#

Polar to semi polar (positive)	4-Guanidinobutanoic acid	4-Guanidinobutanoic acid
Polar to semi polar (positive)	Dihydrouracil	Dihydrouracil
Polar to semi polar (positive)	5-Aminolevulinic acid/4-Hydroxyproline	5-Aminolevulinic acid#
Polar to semi polar (positive)	5-Aminopentanoic acid	5-Aminopentanoic acid
Polar to semi polar (positive)	5-Hydroxytryptophan	5-Hydroxytryptophan
Polar to semi polar (positive)	Adenine	Adenine
Polar to semi polar (positive)	Adenosine/Deoxyguanosine	Adenosine#
Polar to semi polar (positive)	Alanine/beta-Alanine/Sarcosine	Alanine#
Polar to semi polar (positive)	Aminoadipic acid	Aminoadipic acid
Polar to semi polar (positive)	Arginine	Arginine
Polar to semi polar (positive)	Aspartic acid	Aspartate
Polar to semi polar (positive)	Betaine	Betaine
Polar to semi polar (positive)	Biotin	Biotin
Polar to semi polar (positive)	Cadaverine	Cadaverine
Polar to semi polar (positive)	Carnitine	Carnitine
Polar to semi polar (positive)	Choline	Choline
Polar to semi polar (positive)	Citrulline	Citrulline
Polar to semi polar (positive)	Creatine/Beta-Guanidinopropionic acid	Creatine#
Polar to semi polar (positive)	Cytidine	Cytidine
Polar to semi polar (positive)	Cytosine	Cytosine
Polar to semi polar (positive)	Ethanolamine	Ethanolamine
Polar to semi polar (positive)	Glycerophosphocholine	Glycerophosphocholine
Polar to semi polar (positive)	Glycylproline	Glycylproline
Polar to semi polar (positive)	Guanidoacetic acid	Guanidoacetic acid
Polar to semi polar (positive)	Hypoxanthine	Hypoxanthine
Polar to semi polar (positive)	Indoleacetic acid	Indoleacetic acid
Polar to semi polar (positive)	Inosine	Inosine
Polar to semi polar (positive)	Isoleucine	Isoleucine
Polar to semi polar (positive)	Kynurenic acid	Kynurenic acid
Polar to semi polar (positive)	Feature_mz_130.086	Feature_mz_130.086
Polar to semi polar (positive)	Dodecanoylcarnitine	Dodecanoylcarnitine
Polar to semi polar (positive)	Leucine	Leucine
Polar to semi polar (positive)	Methionine	Methionine
Polar to semi polar (positive)	N-Acetylcadaverine	N-Acetylcadaverine
Polar to semi polar (positive)	N-Acetylputrescine	N-Acetylputrescine
Polar to semi polar (positive)	N-Acetyltyrosine	N-Acetyltyrosine
Polar to semi polar (positive)	Targinine/Homoarginine	Homoarginine#
Polar to semi polar (positive)	N1-Methyl-4-pyridone-3-carboxamide/Nudifloramide	Nudifloramide#

Polar to semi polar (positive)	N2,N2-Dimethylguanosine	N2,N2-Dimethylguanosine
Polar to semi polar (positive)	N6,N6,N6-Trimethyllysine	N6,N6,N6-Trimethyllysine
Polar to semi polar (positive)	Nicotinic acid	Nicotinic acid
Polar to semi polar (positive)	Ornithine	Ornithine
Polar to semi polar (positive)	Phenylalanine	Phenylalanine
Polar to semi polar (positive)	Phenylethylamine	Phenylethylamine
Polar to semi polar (positive)	Picolinic acid	Picolinic acid
Polar to semi polar (positive)	Pipecolic acid	Pipecolic acid
Polar to semi polar (positive)	Proline	Proline
Polar to semi polar (positive)	Pyridoxal	Pyridoxal
Polar to semi polar (positive)	Quinaldic acid	Quinaldic acid
Polar to semi polar (positive)	Riboflavin	Riboflavin
Polar to semi polar (positive)	Serotonin	Serotonin
Polar to semi polar (positive)	Spermidine	Spermidine
Polar to semi polar (positive)	Sphinganine	Sphinganine
Polar to semi polar (positive)	Sphingosine	Sphingosine
Polar to semi polar (positive)	Symmetric dimethylarginine/Asymmetric dimethylarginine	SDMA#
Polar to semi polar (positive)	Thiamine	Thiamine
Polar to semi polar (positive)	Threonine/Homoserine	Threonine#
Polar to semi polar (positive)	Thymine	Thymine
Polar to semi polar (positive)	Trimethylamine	Trimethylamine
Polar to semi polar (positive)	Tryptamine	Tryptamine
Polar to semi polar (positive)	Tyramine	Tyramine
Polar to semi polar (positive)	Tyrosine	Tyrosine
Polar to semi polar (positive)	Uracil	Uracil
Polar to semi polar (positive)	Urocanic acid	Urocanic acid
Polar to semi polar (positive)	Xanthurenic acid	Xanthurenic acid
Bile and fatty acids	Cholic acid	CA
Bile and fatty acids	Chenodeoxycholic acid	CDCA
Bile and fatty acids	Deoxycholic acid	DCA
Bile and fatty acids	Oleic acid	OA
Bile and fatty acids	Linoleic acid	LA
Bile and fatty acids	alpha-Linolenic acid/gamma-Linolenic acid	ALA#
Bile and fatty acids	Dihomo-gamma-linolenic acid/Dihomo-alpha-linolenic acid	DGLA
Bile and fatty acids	Arachidonic acid	AA
Bile and fatty acids	Eicosapentaenoic acid	EPA
Bile and fatty acids	4,8,12,15,19-Docosapentaenoic acid	DPA
Bile and fatty acids	Docosahexaenoic acid	DHA
Bile and fatty acids	Glycocholic acid	GCA
Bile and fatty acids	Glycochenodeoxycholic acid	GCDCA
Bile and fatty acids	Glycoursodeoxycholic acid	GUDCA
Bile and fatty acids	Hyocholic acid	HCA
Bile and fatty acids	Lithocholic acid	LCA
Bile and fatty acids	Taurocholic acid	TCA
Bile and fatty acids	Taurochenodesoxycholic acid	TCDCA
Bile and fatty acids	Taurodeoxycholic acid	TDCA
Bile and fatty acids	Tauroursodeoxycholic acid	TUDCA
Bile and fatty acids	Taurolithocholic acid	TLCA
Bile and fatty acids	Ursodeoxycholic acid	UDCA

SCFA	Acetate	Acetate
SCFA	Butyrate	Butyrate
SCFA	Isobutyrate	Isobutyrate
SCFA	Isovalerate	Isovalerate
SCFA	Propionate	Propionate

Table S4: AIC comparison of model fitting with and without age as a covariate

metabolite	CM tolerance-allergy model		intervention model	
	without age	with age	without age	with age
1-Methyladenosine#	345	351	345	352
4-Guanidinobutanoic acid	345	352	343	350
Dihydrouracil	315	321	314	321
5-Aminolevulinic acid#	328	334	328	334
5-Aminopentanoic acid	328	334	325	332
5-Hydroxytryptophan	324	330	329	335
Adenine	296	302	286	292
Adenosine#	306	313	307	314
Alanine#	319	320	322	323
Aminoadipic acid	279	286	280	287
Arginine	381	386	379	384
Aspartate	314	319	306	310
Betaine	351	356	353	359
Biotin	347	353	345	351
Cadaverine	310	317	313	319
Carnitine	353	358	352	357
Choline	300	306	300	306
Citrulline	322	328	336	340
Creatine#	336	343	341	348
Cytidine	326	333	323	330
Cytosine	340	346	335	341
Ethanolamine	327	325	325	322
Glycerophosphocholine	301	308	301	309
Glycylproline	307	308	305	306
Guanidoacetic acid	364	364	364	366
Hypoxanthine	302	308	301	308
Indoleacetic acid	304	311	304	311
Inosine	320	326	300	307
Isoleucine	309	307	309	306
Kynurenic acid	341	348	340	346
Dodecanoylcarnitine	334	341	338	344
Leucine	306	303	305	300
Methionine	320	319	321	318
N-Acetylcadaverine	316	323	316	323
N-Acetylputrescine	368	375	368	375
N-Acetyltyrosine	407	411	410	413
Homoarginine#	328	334	329	335
Nudifloramide#	301	306	303	308
N2,N2-Dimethylguanosine	330	337	332	339
N6,N6,N6-Trimethyllysine	325	332	325	332
Nicotinic acid	305	312	305	312
Ornithine	329	328	334	332
Phenylalanine	316	315	314	311
Phenylethylamine	325	331	327	333
Picolinic acid	347	348	350	349
Pipecolic acid	325	332	318	325
Proline	333	327	334	327
Pyridoxal	278	285	280	286
Quinaldic acid	322	329	321	328
Riboflavin	357	363	355	361
Serotonin	352	356	359	363
Spermidine	335	341	333	338
Sphinganine	312	313	313	313
Sphingosine	318	324	315	322
SDMA#	350	352	351	354
Thiamine	354	359	353	358
Threonine#	308	309	314	314
Thymine	299	305	294	300
Trimethylamine	319	324	324	329
Tryptamine	308	315	311	318
Tyramine	334	339	333	339
Tyrosine	295	299	291	294
Uracil	332	337	328	333
Urocanic acid	328	332	327	330
Xanthurenic acid	336	343	334	341
1-Methyluric acid	323	329	324	331
2,5-Furandicarboxylic acid	320	327	318	325
Deoxyinosine	315	321	310	317
Deoxyuridine	302	308	296	303
ortho-Hydroxyphenylacetic acid	293	300	295	302
Protocatechuic acid	282	286	289	292
4-OH-PLA#	357	364	348	355
3-Hydroxybenzoic acid	321	325	322	325
3-Hydroxybutyric acid	345	350	354	358
3-Methyl-2-oxovaleric acid	347	347	347	347
Methylxanthine isomers	289	296	289	295
PLA#	312	319	320	326

Hydrocinnamic acid	329	335	329	335
4-Hydroxybenzoic acid	338	344	343	349
4-Hydroxycinnamic acid	296	303	291	297
p-Hydroxyphenylacetic acid#	350	356	351	357
Desaminotyrosine	300	306	304	310
4-Pyridoxic acid	318	320	316	317
Pyroglutamic acid	268	267	270	269
GABA#	296	300	304	307
Argininosuccinic acid	303	309	302	307
Ascorbate	317	323	312	319
Carnosine	383	389	376	382
Citrate	291	296	289	293
Gluconate	333	340	338	345
FAD	272	271	276	276
Glutamine	283	289	275	280
Glycine	321	321	326	324
Glycolate	359	365	355	361
Guanine	320	327	310	317
Hippuric acid	321	328	326	333
Histidine	282	289	284	291
ILA	364	370	350	356
Indoxyl glucoside	366	369	360	364
2-Hydroxyethanesulfonate	353	360	352	358
Isobutyrylglycine	318	325	317	324
Oxoglutaric acid	360	358	362	359
Lysine	299	303	312	315
Malate	355	359	357	361
AGN_mandelic.acid	347	354	348	354
Methionine sulfoxide	324	329	320	325
Fructose#	301	307	311	316
N-alpha-Acetylarginine	296	303	295	302
N-Acetylglutamine	320	327	317	324
N-Acetylneuraminic acid	307	314	310	316
N-Acetylserine	346	348	341	342
N-Acetyltryptophan	324	331	323	330
N2-gamma-Glutamylglutamine	327	333	325	332
N6-Carboxymethyllysine	307	313	309	316
Glutamate#	303	305	310	311
Orotate	318	325	322	329
p-Cresol	315	322	316	323
p-Cresol.sulfate	322	328	322	329
Pantothenic acid	303	309	299	305
Phenylacetic acid	328	335	332	339
Phenylacetylglutamine	336	340	332	337
Phenylpropionylglycine	371	377	374	379
Pregnenolone sulfate	331	328	330	328
Pseudouridine	313	320	309	316
Pyrocatechol	283	289	289	295
Pyruvate	318	319	316	317
Serine	344	344	344	344
Syringic acid	297	304	298	305
Tartaric acid	285	290	288	293
Taurine	379	384	379	385
Thymidine	322	328	320	327
trans-Aconitic acid	304	306	309	312
Ferulic acid	314	321	312	320
Tryptophan	343	346	344	348
Uric acid	321	326	327	331
Uridine	334	341	319	326
Valine	309	312	308	310
Xanthine	285	293	287	295
Xylulose	330	337	326	333
CA	379	385	380	387
CDCA	375	381	379	385
DCA	338	344	329	336
OA	328	332	313	315
LA	341	345	313	315
ALA#	352	357	334	338
DGLA#	311	317	315	321
AA	309	314	315	320
EPA	331	336	335	340
DPA	340	346	343	348
DHA	332	333	339	338
GCA	306	313	303	310
GCDCA	309	316	307	314
GUDCA	281	287	281	287
HCA	371	375	360	364
LCA	339	345	333	339
TCA	304	311	304	311
TCDCA	307	314	311	318
TDCA	333	339	329	334
TUDCA	310	317	316	322
TLCA	287	290	286	288
UDCA	310	317	314	320
Acetate	304	311	305	312
Butyrate	310	314	315	320
Isobutyrate	325	332	333	340
Isovalerate	304	311	312	320
Propionate	304	311	303	310

DCA/CA	352	358	345	351
UDCA/CDCA	289	294	290	296
LCA/CDCA	356	362	352	359
CA/GCA	318	325	299	306
CDCA/GCDCA	314	320	297	303
UDCA/GUDCA	291	298	276	283
CA/TCA	327	334	317	325
CDCA/TCDCA	321	327	316	323
DCA/TDCA	343	348	335	341
UDCA/TUDCA	322	326	320	324
LCA/TLCA	320	323	313	317

Table S5: Clinical characteristics associated with outgrowth of cow's milk allergy

characteristics	**Allergic (n=15)**	**Tolerant (n=24)**	**P_values**
egg_allergy : N	10 (67%)	15 (62%)	1.000
egg_allergy : Y	5 (33%)	9 (38%)	
sibling : N	5 (33%)	6 (25%)	0.718
sibling : Y	10 (67%)	18 (75%)	
allergy_father : N	6 (40%)	18 (75%)	**0.044**
allergy_father : Y	9 (60%)	6 (25%)	
allergy_mother : N	5 (33%)	15 (62%)	0.105
allergy_mother : Y	10 (67%)	9 (38%)	
delivery : Caesarean	8 (53%)	18 (75%)	0.185
delivery : Vaginal	7 (47%)	6 (25%)	
race : Asian	12 (80%)	16 (67%)	0.617
race : Caucasian / White	3 (20%)	6 (25%)	
race : Combination of above / Other	0 (0%)	2 (8%)	
sex : F	3 (20%)	8 (33%)	0.477
sex : M	12 (80%)	16 (67%)	
Daily.Formula.Intake.g : TP1	96.6± 34.38	89.75± 31.45	0.664
Daily.Formula.Intake.g : TP2	85.27± 47.37	85.58± 35.27	
Daily.Formula.Intake.mL : TP1	658± 275.95	601.25± 258.58	0.761
Daily.Formula.Intake.mL : TP2	577.33± 352.65	586.46± 281.46	
SCORAD.index : TP1	8.13± 9.67	5.46± 8.32	0.338
SCORAD.index : TP2	10.37± 8.77	6.77± 8.25	0.266
SCORAD.index : TP0	16.27± 13.24	8.98± 14.41	**0.036**
breastfeding duration until study entry (days)	206.87± 116.53	182.33± 107.6	0.453
age : TP1	15.59± 2.54	14.62± 3.02	0.427
age : TP2	21.88± 3.01	20.84± 3.05	0.411
age : TP0	9.68± 2.63	8.57± 3.04	0.254
AAF	6 (40%)	10 (42%)	1.000
AAF-S	9 (60%)	14 (58%)	

bottle.feeding.type until study entry	**Allergic (n=15)**	**Tolerant (n=24)**
Amino Acid Formula	6 (40%)	4 (18%)
Hydrolysate	0 (0%)	2 (9%)
Hydrolysate;Amino Acid Formula	4 (27%)	1 (5%)
Whole protein (milk / soy)	0 (0%)	1 (5%)
Whole protein (milk / soy);Amino Acid Formula	1 (7%)	2 (9%)
Whole protein (milk / soy);Hydrolysate	1 (7%)	5 (23%)
Whole protein (milk / soy);Hydrolysate;Amino Acid Formula	3 (20%)	7 (32%)
missing	0	2

Numeric variables are presented as mean ± standard deviation; categorical variable are presented as number (%)

Table S6: Clinical characteristics associated with interventions

characteristics	AAF (n=16)	AAF-S (n=23)	P_values
egg_allergy : N	12 (75%)	13 (57%)	0.32
egg_allergy : Y	4 (25%)	10 (43%)	
sibling : N	5 (31%)	6 (26%)	0.73
sibling : Y	11 (69%)	17 (74%)	
allergy_father : N	11 (69%)	13 (57%)	0.52
allergy_father : Y	5 (31%)	10 (43%)	
allergy_mother : N	9 (56%)	11 (48%)	0.75
allergy_mother : Y	7 (44%)	12 (52%)	
delivery : Caesarean	9 (56%)	17 (74%)	0.31
delivery : Vaginal	7 (44%)	6 (26%)	
race : Asian	10 (62%)	18 (78%)	0.62
race : Caucasian / White	5 (31%)	4 (17%)	
race : Combination of above / Other	1 (6%)	1 (4%)	
sex : F	6 (38%)	5 (22%)	0.31
sex : M	10 (62%)	18 (78%)	
Daily.Formula.Intake.g : TP1	91.44± 32.93	93.04± 32.64	0.86
Daily.Formula.Intake.g : TP2	72± 30.23	94.83± 43.38	0.07
Daily.Formula.Intake.mL : TP1	596.25± 285.61	641.74± 251.41	0.68
Daily.Formula.Intake.mL : TP2	475.62± 254.16	657.61± 322.09	0.07
SCORAD.index : TP1	8.03± 10.27	5.41± 7.74	0.24
SCORAD.index : TP2	8.75± 7.9	7.74± 9.08	0.54
SCORAD.index : TP0	13.34± 16.06	10.7± 13.12	0.69
breastfeding duration until study entry (days)	217.25± 105.31	174.04± 112.4	0.20
age : TP1	15.06± 2.88	14.95± 2.89	0.83
age : TP2	21.24± 2.84	21.24± 3.23	0.99
age : TP0	9.09± 2.91	8.93± 2.96	0.91
Allergic: TP2	6 (38%)	9 (39%)	1.00
Tolerant: TP2	10 (62%)	14 (61%)	

bottle.feeding.type until study entry	AAF (n=16)	AAF-S (n=23)
Amino Acid Formula	3 (20%)	7 (32%)
Hydrolysate	0 (0%)	2 (9%)
Hydrolysate;Amino Acid Formula	2 (13%)	3 (14%)
Whole protein (milk / soy)	0 (0%)	1 (5%)
Whole protein (milk / soy);Amino Acid Formula	3 (20%)	0 (0%)
Whole protein (milk / soy);Hydrolysate	3 (20%)	3 (14%)
Whole protein (milk / soy);Hydrolysate;Amino Acid Formula	4 (27%)	6 (27%)
missing	1	1

Numeric variables are presented as mean ± standard deviation; categorical variable are presented as number (%)

Table S7: Significantly altered metabolites in CM-allergic and CM-tolerant groups from marginal means comparison

CM-Allergic				
Metabolite	TP0	TP1	P value	Q value
Protocatechuic acid	3.727 (3.317, 4.137)	4.607 (4.197, 5.017)	0.0006	0.0674
Pyrocatechol	3.374 (2.969, 3.778)	4.252 (3.847, 4.656)	0.0008	0.0674
CM-Allergic				
Metabolite	TP0	TP2	P value	Q value
Pyroglutamic acid	5.833 (5.472, 6.194)	4.849 (4.489, 5.21)	0.0002	0.0328
Threonine#	15.992 (15.544, 16.44)	14.939 (14.491, 15.387)	0.0004	0.0340
Pyruvic acid	7.365 (6.89, 7.841)	6.322 (5.847, 6.798)	0.0006	0.0347
Pregnenolone sulfate	7.735 (7.236, 8.234)	6.658 (6.159, 7.157)	0.0011	0.0460
Tryptophan	13.74 (13.233, 14.247)	12.636 (12.129, 13.142)	0.0030	0.0806
Oxoglutaric acid	10.624 (10.057, 11.191)	9.508 (8.941, 10.076)	0.0033	0.0806
Ferulic acid	5.59 (5.151, 6.029)	6.61 (6.171, 7.049)	0.0034	0.0806
CM-Tolerant				
Metabolite	TP0	TP1	P value	Q value
Tartaric acid	2.769 (2.443, 3.094)	3.597 (3.271, 3.922)	0.0001	0.0082
Pyroglutamic acid	5.977 (5.691, 6.262)	5.178 (4.893, 5.463)	0.0001	0.0082
Dodecanoylcarnitine	3.411 (3.02, 3.802)	2.383 (1.992, 2.774)	0.0002	0.0082
TLCA	-3.038 (-3.378, -2.698)	-3.819 (-4.159, -3.479)	0.0002	0.0082
Desaminotyrosine	4.112 (3.785, 4.44)	5.017 (4.689, 5.344)	0.0003	0.0095
Ferulic acid	5.495 (5.148, 5.842)	6.323 (5.976, 6.67)	0.0026	0.0707
PLA#	8.617 (8.268, 8.965)	9.364 (9.015, 9.712)	0.0042	0.0920
FAD	5.631 (5.342, 5.92)	6.255 (5.966, 6.544)	0.0044	0.0920
Pregnenolone sulfate	7.88 (7.486, 8.275)	7.141 (6.746, 7.535)	0.0051	0.0945
CM-Tolerant				
Metabolite	TP0	TP2	P value	Q value
Dodecanoylcarnitine	3.411 (3.02, 3.802)	2.03 (1.639, 2.421)	0.000001	0.0001
Pregnenolone sulfate	7.88 (7.486, 8.275)	6.675 (6.281, 7.07)	0.000004	0.0003
Desaminotyrosine	4.112 (3.785, 4.44)	5.204 (4.877, 5.532)	0.000012	0.0007
LCA/TLCA	-0.98 (-1.363, -0.598)	0.054 (-0.335, 0.443)	0.0001	0.0025
Pyruvate	7.619 (7.243, 7.995)	6.664 (6.288, 7.04)	0.0001	0.0025
PLA#	8.617 (8.268, 8.965)	9.615 (9.267, 9.963)	0.0001	0.0030
Protocatechuic acid	3.367 (3.043, 3.691)	4.128 (3.804, 4.452)	0.0002	0.0040
HCA	-0.1 (-0.574, 0.373)	-1.354 (-1.836, -0.871)	0.0002	0.0040
Ferulic acid	5.495 (5.148, 5.842)	6.505 (6.158, 6.852)	0.0002	0.0040
Pyroglutamic acid	5.977 (5.691, 6.262)	5.21 (4.924, 5.495)	0.0002	0.0043
FAD	5.631 (5.342, 5.92)	6.396 (6.107, 6.685)	0.0004	0.0064
TLCA	-3.038 (-3.378, -2.698)	-3.756 (-4.101, -3.41)	0.0008	0.0113
Pipecolic acid	10.801 (10.423, 11.178)	11.656 (11.279, 12.034)	0.0010	0.0141
DCA/CA	-1.888 (-2.322, -1.454)	-0.899 (-1.341, -0.456)	0.0015	0.0188
Oxoglutaric acid	10.807 (10.359, 11.256)	9.865 (9.417, 10.314)	0.0016	0.0188
DCA/TDCA	-0.745 (-1.17, -0.321)	0.149 (-0.283, 0.582)	0.0019	0.0209
3-Hydroxybenzoic acid	5.341 (4.976, 5.706)	6.151 (5.786, 6.516)	0.0023	0.0234
Betaine	5.159 (4.751, 5.567)	4.16 (3.752, 4.568)	0.0027	0.0262
TCDCA	-0.893 (-1.24, -0.547)	-1.681 (-2.035, -1.327)	0.0030	0.0278
N-Acetylneuraminic acid	10.142 (9.802, 10.483)	9.392 (9.051, 9.732)	0.0032	0.0279
CA	1.501 (1.022, 1.979)	0.386 (-0.104, 0.875)	0.0040	0.0327
Hydrocinnamic acid	3.076 (2.698, 3.454)	3.869 (3.491, 4.246)	0.0041	0.0327
UDCA/CDCA	-1.858 (-2.175, -1.542)	-1.156 (-1.479, -0.832)	0.0054	0.0392
LCA/CDCA	-1.654 (-2.092, -1.216)	-0.744 (-1.191, -0.297)	0.0054	0.0392
Butyrate	1.522 (1.153, 1.891)	2.255 (1.893, 2.617)	0.0055	0.0392
Phenylacetylglutamine	3.922 (3.524, 4.321)	3.185 (2.786, 3.583)	0.0073	0.0481
DCA	-0.571 (-0.988, -0.155)	0.192 (-0.232, 0.616)	0.0076	0.0481
N6-Carboxymethyllysine	5.415 (5.064, 5.767)	6.055 (5.704, 6.406)	0.0076	0.0481
5-Hydroxytryptophan	7.375 (7.013, 7.737)	6.623 (6.261, 6.985)	0.0131	0.0802
CDCA	-0.131 (-0.601, 0.339)	-1.081 (-1.561, -0.601)	0.0139	0.0809
Quinaldic acid	4.44 (4.08, 4.8)	5.138 (4.777, 5.498)	0.014754102	0.080889787
Arginine	14.321 (13.842, 14.799)	13.461 (12.983, 13.94)	0.015285229	0.080889787
Pyrocatechol	3.184 (2.864, 3.504)	3.703 (3.383, 4.023)	0.015426353	0.080889787
Phenylacetic acid	6.412 (6.044, 6.78)	7.148 (6.78, 7.516)	0.015766331	0.080889787
LCA	-0.995 (-1.408, -0.583)	-0.276 (-0.696, 0.145)	0.015995156	0.080889787
TP2				
Metabolite	Allergic	Tolerant	P value	Q value
Citrulline	11.841 (11.378, 12.303)	12.946 (12.58, 13.311)	0.0003	0.0537
Lysine	11.371 (10.957, 11.785)	12.273 (11.946, 12.601)	0.0010	0.0823

Table S8: Spearman's rank correlation between the changes of *bifidobacterium* and metabolites/ratios which are significantly altered in the AAF-S group

Compound	Rho	P value	time points	Intervention	Q value
ILA	0.858695652	2.13E-06	TP1-TP0	AAF-S	0
4-OH-PLA#	0.768774704	2.81E-05	TP1-TP0	AAF-S	2.00E-04
Adenine	0.637351779	0.001379994	TP1-TP0	AAF-S	0.0069
Glutamine	-0.57312253	0.004939677	TP1-TP0	AAF-S	0.0185
Adenine	0.720588235	0.002305841	TP1-TP0	AAF	0.0346
CDCA/GCDCA	0.507905138	0.014441199	TP1-TP0	AAF-S	0.0433
Inosine	-0.462450593	0.027527897	TP1-TP0	AAF-S	0.059
GCDCA	-0.468379447	0.025413081	TP1-TP0	AAF-S	0.059
CA/GCA	0.43083004	0.041340241	TP1-TP0	AAF-S	0.0775
Guanine	-0.31027668	0.149497361	TP1-TP0	AAF-S	0.2492
Uridine	-0.244071146	0.260522306	TP1-TP0	AAF-S	0.3908
UDCA/GUDCA	-0.185770751	0.394330031	TP1-TP0	AAF-S	0.5377
Inosine	-0.314705882	0.234711639	TP1-TP0	AAF	0.5423
4-OH-PLA#	0.235294118	0.379021393	TP1-TP0	AAF	0.5423
Glutamine	-0.229411765	0.391370422	TP1-TP0	AAF	0.5423
Guanine	-0.267647059	0.315146037	TP1-TP0	AAF	0.5423
ILA	0.397058824	0.128882583	TP1-TP0	AAF	0.5423
OA	0.208823529	0.436322033	TP1-TP0	AAF	0.5423
ALA	0.247058824	0.354996461	TP1-TP0	AAF	0.5423
GCDCA	-0.197058824	0.463185494	TP1-TP0	AAF	0.5423
HCA	-0.194117647	0.470031308	TP1-TP0	AAF	0.5423
CA/GCA	0.323529412	0.221281258	TP1-TP0	AAF	0.5423
CDCA/GCDCA	0.226470588	0.397628203	TP1-TP0	AAF	0.5423
UDCA/GUDCA	-0.294117647	0.268071938	TP1-TP0	AAF	0.5423
Uridine	-0.164705882	0.541223723	TP1-TP0	AAF	0.5799
ALA	0.136363636	0.533413277	TP1-TP0	AAF-S	0.6668
HCA	0.12055336	0.582372852	TP1-TP0	AAF-S	0.672
OA	0.032608696	0.883337839	TP1-TP0	AAF-S	0.9119
LA	0.024703557	0.911925712	TP1-TP0	AAF-S	0.9119
LA	0	1	TP1-TP0	AAF	1
4-OH-PLA#	0.674901186	0.000570581	TP2-TP0	AAF-S	0.0086
ILA	0.624505929	0.001818976	TP2-TP0	AAF-S	0.0136
Adenine	0.523715415	0.011326837	TP2-TP0	AAF-S	0.0566
Glutamine	-0.497035573	0.016967834	TP2-TP0	AAF-S	0.0636
Adenine	0.594117647	0.017246545	TP2-TP0	AAF	0.2147
Uridine	-0.467391304	0.025756077	TP2-TP0	AAF-S	0.0773
LA	-0.571428571	0.028623176	TP2-TP0	AAF	0.2147
OA	-0.521428571	0.048830208	TP2-TP0	AAF	0.2442
Guanine	-0.407114625	0.05494236	TP2-TP0	AAF-S	0.1374
CDCA/GCDCA	0.453571429	0.091529268	TP2-TP0	AAF	0.2572
Glutamine	-0.426470588	0.101056074	TP2-TP0	AAF	0.2572
CA/GCA	0.439285714	0.103199216	TP2-TP0	AAF	0.2572
ILA	0.405882353	0.120020814	TP2-TP0	AAF	0.2572
4-OH-PLA#	0.373529412	0.154767766	TP2-TP0	AAF	0.2902
Guanine	-0.35	0.184066376	TP2-TP0	AAF	0.3068
LA	-0.29079616	0.188679939	TP2-TP0	AAF-S	0.4043
GCDCA	-0.267080745	0.228602156	TP2-TP0	AAF-S	0.423
CDCA/GCDCA	0.25352908	0.253790419	TP2-TP0	AAF-S	0.423
Inosine	-0.219367589	0.313053704	TP2-TP0	AAF-S	0.4424
CA/GCA	0.219649915	0.324461463	TP2-TP0	AAF-S	0.4424
GCDCA	-0.267857143	0.333445517	TP2-TP0	AAF	0.5002
Inosine	-0.241176471	0.366896119	TP2-TP0	AAF	0.5003
Uridine	-0.208823529	0.436322033	TP2-TP0	AAF	0.5454
UDCA/GUDCA	0.185714286	0.506673971	TP2-TP0	AAF	0.5519
HCA	0.182142857	0.515060927	TP2-TP0	AAF	0.5519
UDCA/GUDCA	-0.119141728	0.596160223	TP2-TP0	AAF-S	0.7452
ALA	0.128571429	0.648201799	TP2-TP0	AAF	0.6482
OA	-0.086391869	0.701645989	TP2-TP0	AAF-S	0.7598
ALA	0.084133258	0.709154443	TP2-TP0	AAF-S	0.7598
HCA	-0.049124788	0.828554014	TP2-TP0	AAF-S	0.8286

Chapter VI

Conclusions and Perspectives

The cross-talk between the gut microbiome and the human host has been increasingly recognized as an important factor influencing human health and disease,[1] including cow's milk allergy (CMA), which is the most common type of food allergy in early life.[2] Although advancements in omics techniques have significantly improved our understanding of this interplay, uncovering the complex mechanisms by which the gut microbiome affects the host remains a challenge. In recent decades, growing evidence suggests that the gut microbiome-derived metabolites serve as important mediators in this interaction.[3] This highlights metabolomics as a key technique for elucidating the gut microbiome's role in human health and disease by providing insights at the molecular level. In metabolomics studies, approaches can be broadly categorized into targeted and untargeted metabolomics, based on hypothesis-driven and hypothesis-generating strategies, respectively.[4] Targeted metabolomics focuses on quantifying a limited number of known metabolites, while untargeted metabolomics aims to profile both known and unknown metabolic features.[4] One of the primary challenges for metabolite quantification in targeted and untargeted metabolomics is matrix effect.[6] Matrix effect is primarily caused by co-eluting matrix components, which can impact the accuracy and reliability of signals detected with liquid chromatography-mass spectrometry (LC-MS), particularly when using an electrospray ionization (ESI) source.[5] In this thesis two hypotheses were investigated. The first hypothesis was that the matric effect in untargeted metabolomics can be monitored and corrected by implementing the PCIS technique with LC-MS methods. The second hypothesis was that the fecal metabolome can provide insights into the cross-talk between the gut microbiome and food allergy in infants with the most prevalent type of food allergy in early life: cow's milk allergy (CMA). life.

Matrix effect in untargeted metabolomics

Untargeted metabolomics, a powerful approach for unbiased metabolome profiling, has demonstrated potential for biomarker discovery in diverse fields. However, despite its wide applications, several challenges remain that impact the reliability of untargeted metabolomics. Among these, matrix effect is a major concern, as it can greatly affect the reproducibility, selectivity, and accuracy of metabolome profiling.[6] Stable

isotopically labeled (SIL) standards, the most commonly applied strategy for addressing the matrix effect, are limited to targeted metabolomics due to the requirement of standards spiking. This limitation makes another approach, PCIS, the only applicable method for mitigating matrix effects in untargeted metabolomics, as it is independent of retention time.[6] The effectiveness of PCIS in monitoring and correcting matrix effects has been well demonstrated in targeted metabolomics,[7–10] and it has also been recommended as a quality control tool for matrix effect evaluation in untargeted metabolomics.[11] However, reports on its actual application in untargeted metabolomics remain limited.[12] Therefore, effective strategies to address the matrix effect with PCIS in untargeted metabolomics are still lacking. To tackle this, in **Chapters 2** and **3** of the thesis, we outlined strategies using PCIS to overcome matrix effect in LC-ESI-MS-based untargeted metabolomics, covering matrix effect monitoring and matrix effect compensation.

First, in **Chapter 2**, an untargeted method was developed and applied to evaluated the matrix effect in plasma and fecal samples with PCIS. As part of the method development, the injection amount and reconstitution solvent were first optimized for both plasma and fecal samples. The results showed that optimizing the reconstitution solvent was crucial for balancing the trade-off between peak shape distortion and metabolite solubility, and that proper sample dilution was essential for maximizing metabolites signal intensity while preventing detector saturation in MS. To assess the analytical performance of our untargeted method, the method was validated using a targeted approach with stable isotope-labeled (SIL) standards in plasma and fecal samples. The method exhibited good precision, accuracy, recovery, and repeatability with plasma and fecal samples. By evaluating the matrix effect, it was found that high relative matrix effect (RME) among samples could significantly impact measurement accuracy and reproducibility. However, the SIL standards can only point out the matrix effect at specific retention times. To assess the matrix effect across the entire chromatogram, a PCIS approach was introduced to the developed untargeted metabolomics method. In this approach, xenobiotic compounds were infused post-column during the injection of different plasma and fecal samples, enabling overall

monitoring of absolute matrix effect (AME) and RME by examining the matrix effect profiles of the infused compounds.

The results demonstrated that the PCIS approach effectively identified chromatographic regions exhibiting large AME and RME. Notably, PCIS yielded comparable RME results to those obtained using the traditional post-extraction spiking method, demonstrating its potential as a reliable technique for RME evaluation in untargeted metabolomics. The PCIS approach was applied to predict the RME of over 300 targets covered in our in-house library. The predictions revealed that more targets exhibited RME > 15% in fecal samples compared to plasma. Additionally, for metabolites detectable in both positive and negative ionization modes, most of them experienced larger RME in negative mode than in positive mode. Overall, **Chapter 2** established a comprehensive framework for developing an LC-ESI-MS untargeted metabolomics method using PCIS to monitor the matrix effect in plasma and fecal samples. The findings demonstrated that PCIS is an effective approach for matrix effect monitoring in untargeted metabolomics. This approach has strong potential to improve better data reliability of untargeted metabolomics by identifying regions with severe matrix effect and high matrix effect variation.

The proposed PCIS approach can be further applied to guide the optimization of specific LC parameters, such as the gradient and sample injection amount, to mitigate matrix effects in a reverse-phase (RP) LC-MS untargeted method. A recent study also demonstrated that PCIS contributed to column selection and mobile phase pH optimization for an untargeted hydrophilic interaction liquid chromatography (HILIC)-MS method.[13] Moreover, although both our RPLC-MS and their HILIC-MS methods targeted polar to semi-polar metabolomic features with a mass less than 800 Da, the application of PCIS is not limited by the polarity or mass range of the metabolites. In principle, with careful selection of PCIS candidates, PCIS can serve as a valuable approach for guiding method development to minimize matrix effects in any untargeted metabolomics method.

In addition to matrix effect monitoring, PCIS also has potential for compensating matrix effect in untargeted metabolomics due to its retention time independence. A key challenge in its implementation lies in selecting multiple PCISs for the wide range of metabolic features and determining which one is most effective for correcting the matrix effect specific to each feature. To address this, the application of PCIS from matrix effect monitoring to compensation was investigated in **Chapter 3.**

In this chapter, the workflow for developing a PCIS approach for an LC-ESI-MS-based untargeted metabolomics method was first outlined. Key factors, such as structural diversity, infusion concentration, and room temperature stability, were thoroughly evaluated to select suitable PCIS candidates. The results demonstrated that, at the optimized infusion concentration, the selected PCISs (five standards for positive ionization mode and four for negative ionization mode) exhibited diverse matrix effect profiles, stable infusion signals, and no significant matrix effect interference. Additionally, these compounds remained stable for one week at room temperature, further supporting their long-term usage along with analysis runs. Next, to match a specific feature with its suitable PCIS for matrix effect correction, a novel approach was proposed: post-column infusion of artificial matrices. This matching process was achieved by comparing the ability of a PCIS to compensate for the artificially created matrix effect (ME_{art}).

To ensure that the artificial matrices properly mimicked the biological matrix in inducing matrix effects, multiple artificial matrix compounds were selected based on their relevance to matrix effects mechanisms in an ESI source. L-homoarginine hydrochloride, sodium acetate, and tridodecylmethylammonium chloride were selected as artificial matrix compounds for positive ionization mode, while sodium dodecyl sulphate and sodium acetate were used for negative mode. These compounds can interfere with ESI process of analytes by competing for ionization or increasing the surface tension in droplets, preventing coulombic explosion. Since the presence of ME_{art} was essential for selecting the suitable PCIS, the infused concentrations of these artificial matrix compounds were optimized to obtain 70% artificial absolute matrix effect (AME_{art}) and more than 15% artificial relative matrix effect (RME_{art}). By injecting

samples into the LC-PCIS-MS system with and without artificial matrices infusion, the ME_{art} could be determined across detected features, including both known and unknown metabolites. The selected PCIS could then be used to compensate for biological matrix effects (ME_{bio}).

The effectiveness of ME_{art} was evaluated in selecting PCIS using 19 diverse SIL standards spiked in plasma, urine, and feces. In this evaluation, ME_{bio} and ME_{art} were calculated and used to select the suitable PCIS for each SIL standard in each biological matrix. To incorporate both absolute matrix effect (AME) and relative matrix effect (RME) into the comparison, a matrix effect scoring system was introduced that averaged AME and RME scores as the final ME score. The ME (ME_{art}, ME_{bio}) scores across plasma, urine, and feces were summed to identify the matrix-independent PCIS for each SIL standard. The PCISs selected based on ME_{art} score sums were compared with those identified using ME_{bio} score sums. As a result, 17 out of 19 (89%) SIL standards exhibited consistent PCISs selection based on ME_{art} and ME_{bio} score sums. Considering that ME_{bio} correction is the most commonly applied strategy for PCIS selection in targeted metabolomics,[9,14,15] our results highlight the efficacy of ME_{art} in selecting the suitable PCISs for ME_{bio} compensation.

Subsequently, ME_{art}-selected PCISs were applied to correct for the ME_{bio} in plasma, urine, and feces for the 19 SIL standards. These PCISs improved or maintained the matrix effect scores for 19 (100%) standards in plasma, 16 (84%) in urine, and 18 (95%) in feces. The results demonstrated the efficacy and reliability using ME_{art} to identify suitable PCISs for ME_{bio} correction across various biological matrices. More importantly, since ME_{art} can be determined for any measurable feature by comparing signals acquired with and without artificial matrix infusion, this establishes post-column ME_{art} creation as a feasible approach for selecting PCIS to correct matrix effect in LC-PCIS-MS-based untargeted metabolomics. Ideally, a feature-PCIS-matched library could be constructed using artificial matrix infusion with one or multiple biological matrices, and then applied to compensate for matrix effect in untargeted metabolomics studies.

Following the successful proof-of-concept demonstration of the matrix effect compensation method using artificial matrix-based PCIS selection, further efforts should focus on building the feature-PCIS-matched library to facilitate routine matrix effect correction in untargeted metabolomics. Additionally, comparing significantly altered features before and after PCIS correction in applied studied remains of great interest for further validating this method for matrix effects correction. Although this study included structurally diverse PCISs, over-corrected matrix effects were observed in a few examined SIL standards. This highlights the need to further expand the diversity of PCIS candidates to improve correction those standards and enable more comprehensive matrix effect correction across the metabolome. Furthermore, with well-defined PCIS candidates and a robust ME_{art}-based matching strategy, the LC-PCIS-MS platform can be extended beyond biomedical matrices to applications in food safety, environment science, and other fields where complex matrix effects are commonly encountered.

Fecal metabolome exploration in infants with cow's milk allergy

In **Chapters 4** and **5**, the aim was to deepen our understanding of the interplay between the gut microbiome and CMA in early life through the exploration of the fecal metabolome. To provide a comprehensive overview of current studies on this topic, a systematic review was conducted in **Chapter 4**. This review focused on the modifications and post-treatment alterations in the gut microbiome, metabolome, and immune response in both CMA children (0-12 years) and CMA animal models. By conducting thorough searches in MEDLINE, PubMed, Scopus, and Web of Science, 21 articles published before March 2023 were included, consisting of 13 studies on CMA children and 8 studies on animal models.

In the reviewed studies, no consistent conclusions were drawn regarding the modifications of α- and β-diversity in the gut microbiome in CMA. At the taxonomic level, multiple studies across both CMA children and animal models reported a decrease in the *Bifidobacterium* genus and Lactobacillales order, alongside an increase in the Clostridia class. Regarding CMA management, various intervention approaches,

including different formulas, prebiotics, probiotics, and synbiotics, were applied across several studies. These studies consistently showed increased *Bifidobacterium* levels in both CMA children and animal models following interventions, particularly with *Bifidobacterium* strains-specific treatments. However, the impact of these interventions on other bacterial populations remained inconclusive. In terms of metabolome modifications, decreased short-chain fatty acids (SCFAs), as well as altered amino acid and organic acid profiles, were observed in CMA children. These metabolomic changes appeared to be restored through interventions, with increased SCFAs and balanced amino acid levels. For the immune response, only one study involving CMA children was available, but studies on CMA animal models suggested that interventions could reduce overall cytokine levels, restore the T_h2/T_h1 balance, and induce a regulatory immune response. Additionally, this review highlighted that no study has investigated early-life CMA using multi-omics strategies, such as metagenomics, metatranscriptomics, and metaproteomics. Although several metabolomics studies have been reported, they focused on a limited range of metabolites, emphasizing the need for comprehensive metabolomics studies on CMA in early life.

In **Chapter 5**, a comprehensive investigation of the fecal metabolome in CMA infants undergoing dietary intervention with and without a synbiotic (inulin, oligofructose and *Bifidobacterium breve* M-16 V) was conducted using the untargeted metabolomics method developed in **Chapter 2**, along with an additional in-house platform. Considering the broad metabolite coverage, we primarily focused on known features in this study. By grouping the infants based on CMA status after one year or the type of intervention they received, we explored the distinct impacts of CMA tolerance acquisition and of the synbiotic supplementation on the fecal metabolome of CMA infants. The longitudinal changes in the fecal metabolome across the three time points were analyzed using linear mixed models (LMMs) and repeated measures analysis of variance simultaneous component analysis+ (RM-ASCA+).

By comparing the fecal metabolome of infants with persistent CMA to those who developed CM-tolerance, more pronounced changes in the fecal metabolome related to amino acids, bile acids, and SCFAs were observed in the CM-tolerant group. The CM-

tolerant group exhibited significantly higher levels of lysine and citrulline after one year of intervention compared to the CM-allergic group. Although no significant group differences were found for other metabolites, the metabolome trends along with time indicated a down-regulation of tryptophan-serotonin metabolism, up-regulation of secondary bile acid production, and an increase in butyrate in the CM-tolerant group compared to the CM-allergic group. These alterations might suggest a healthier gut with improved barrier function and a more mature gut microbiome in the CM-tolerant group.

Regarding the impact of the synbiotic, this study demonstrated that the synbiotic significantly altered the fecal levels of aromatic lactic acids, purine metabolites, fatty acids, and bile acids, especially after six months of supplementation. Two aromatic lactic acids (4-hydroxyphenyllactic acid and indolelactic acid), known as infant-type *Bifidobacterium*-derived metabolites, showed a significant increase in the synbiotic group. Moreover, the changes in these metabolites from baseline to later time points were strongly positively correlated with the changed levels of *Bifidobacterium* in the group with synbiotic supplementation. These findings suggested an enhanced abundance and/or activity of infant-type *Bifidobacterium* species, indicating the successful supplementation of the synbiotic. Additionally, the synbiotic supplementation was found to lower the levels of inosine, guanine, and uridine, increase adenine level, and enhance the deconjugation of glycine-conjugated bile acids.

The study in **Chapter 5** contributed to revealing the linkages between early-life CMA, the gut microbiome, and synbiotic intervention. We observed several alterations in fecal metabolomic pathways that may play a role in the outgrowth of CMA in early life. Additionally, Those findings provided evidence for the impact of synbiotic supplementation on modifying the fecal metabolome in CMA infants. This impact was more pronounced after six months of intervention, highlighting the importance of early intervention to maximize the effects of synbiotics. However, no clear conclusions can be drawn regarding the clinical benefits of the synbiotic supplementation on CM-tolerance acquisition, as the tolerance rate observed after one-year synbiotic intervention was consistent with natural outgrowth for infants involved in our study. Despite this, the significant enhancement of metabolites with anti-inflammatory properties, such as

indolelactic acid,[16] suggested a potential beneficial effect of synbiotics in promoting CMA outgrowth. Therefore, it is suggested that further research with larger cohorts is needed to verify our findings and evaluate the therapeutic potential of synbiotics supplementation for CMA in early life.

Although over 300 targets were involved in our study, there are still opportunities for further improvements in metabolomic exploration. First, our study only reported the relative abundance of these targets. Achieving absolute quantification of the targeted metabolites would enhance the accuracy and depth of our interpretation. Second, despite covering a wide range of targets, analyzing the data in an untargeted manner is still necessary to identify other potential metabolomic changes that may not have been captured in our targeted analysis. Moreover, integrating multiple analytical platforms for global profiling, such as HILIC along with RP, could significantly expand the metabolomic coverage and provide a more comprehensive picture. Lastly, this study was conducted with the PCIS setup, providing the opportunity to reanalyze the data and apply matrix effect correction. We believe that matrix effect correction with PCIS holds substantial potential to further enhance the quality of the data presented in this chapter.

Further perspectives

In this thesis aimed to tackle the issue of matrix effect in untargeted metabolomics (**Chapter 2-3**) and expand the understanding of the relationship between the gut microbiome and CMA in early life (**Chapter 4-5**). From a technical perspective, the thesis demonstrated the potential of applying PCIS to address matrix effect in LC-ESI-MS-based untargeted metabolomics methods. The developed PCIS method enabled two primary functions: matrix effect monitoring and matrix effect correction. Matrix effect monitoring is particularly useful during the development phase of an untargeted metabolomics method to help mitigate matrix effect. Matrix effect correction holds great potential to enhance data reliability, advance (semi)quantitative analysis, and ensure more accurate data interpretation in untargeted metabolomics. One direct application of matrix effect correction is solving the problem of matrix dilution when examining the dynamic range using endogenous metabolites with serially diluted quality control (dQC)

samples.[17] This enables the exploration of linearity with endogenous features by correcting matrix effect in a calibration curve constructed using dQC. Additionally, leveraging a dQC series with corrected matrix effect in routine untargeted metabolomics analysis can also improve the data fidelity via advancing the feature filtering with estimated linearity range and response for detected features.[18] Overall, the advances in addressing matrix effect presented in our study will contribute to the broader application of untargeted metabolomics in diverse research fields. However, to expand the implementation of matrix effect correction using PCIS, further efforts are required to develop an automated pipeline that increases throughput for efficiently selecting appropriate PCISs for the hundreds to thousands of features detected in untargeted metabolomics. Meanwhile, incorporating the PCIS pre-processing workflow into existing untargeted data analysis tools could also further promote its application.

With the developed method, **Chapter 5** uncovered several potential metabolomic pathway modifications related to CMA resolution in early life and highlighted significant metabolite changes following the synbiotic intervention. This study contributed to gaining insights into the interplay between the gut microbiome and early-life CMA from a metabolomics perspective. To further reveal the underlying mechanisms regarding the impact of the gut microbiome on early-life CMA, we recommend carrying out more studies with larger-scale cohorts. This will also enable researchers to develop more complex data analysis models to explore how synbiotic interventions can influence CM-tolerance acquisition in early life. Meanwhile, since the fecal metabolome serves as an ideal readout for gut microbial functions,[19] the primary focus of this chapter was on fecal metabolome profiling In the future, combining the fecal metabolome with metabolomes obtained from other biological samples, particularly peripheral blood, could significantly enhance our interpretation of the cross-talk between the gut microbiome and the host. Additionally, comprehensive studies that integrate multi-omics research combining metabolomics, metagenomics, metatranscriptomics, and metaproteomics are still urgently needed to gain a complete understanding of impact of the gut microbiome on the prevention, development, and treatment of CMA in early life. Furthermore, as we enter the era of artificial intelligence,

incorporating techniques like machine learning with integrated multi-omics data holds great promise for advancing our knowledge of the role of the gut microbiome in human health and disease, including CMA in early life.

Final remarks

The rapid expansion of research on human gut microbiome in recent decades has highlighted its role in human metabolism, immune regulation, and behavior.[20] Despite significant progress in deciphering how the gut microbiome affects human health and disease, a long journey lies ahead to fully solve the puzzle. Combining multi-omics analyses has become a trend to unravel the intricate relationship between the gut microbiome and the human host. Among the omics techniques, as a direct readout of phenotypes, metabolomics provides a snapshot reflecting the functional properties of the gut microbiome at the molecular level. This emphasizes the crucial role of metabolomics in revealing this complex relationship and underscores the needs for advances in metabolomics techniques. In this thesis, by proposing strategies to address the matrix effect in LC-ESI-MS-based analytical method, we advanced untargeted metabolomics towards quantitative analysis. The focus then shifted to deepening our understanding of the interactions between the gut microbiome and CMA in early life from a metabolomics perspective. Overall, the research in this thesis suggested that several gut microbiome-involved metabolic pathways may play a role in the acquisition of CM tolerance, and provided evidence that the fecal metabolome can serve as a potential readout to reflect the impact of early synbiotic supplementation in infants. These findings offered valuable insights into the relationship between the gut microbiome and CMA, aiding future research in developing microbiome-targeted strategies for the prevention and management of CMA in early life.

Reference:

1. Chen Y, Zhou J, Wang L. Role and Mechanism of Gut Microbiota in Human Disease. *Front Cell Infect Microbiol*. 2021;11. doi:10.3389/fcimb.2021.625913
2. Sicherer SH, Warren CM, Dant C, Gupta RS, Nadeau KC. Food Allergy from Infancy Through Adulthood. *The Journal of Allergy and Clinical Immunology: In Practice*. 2020;8(6):1854-1864. doi:10.1016/j.jaip.2020.02.010
3. Hosseinkhani F, Heinken A, Thiele I, Lindenburg PW, Harms AC, Hankemeier T. The contribution of gut bacterial metabolites in the human immune signaling pathway of non-communicable diseases. *Gut Microbes*. 2021;13(1):1-22. doi:10.1080/19490976.2021.1882927
4. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J Am Soc Mass Spectrom*. 2016;27(12):1897-1905. doi:10.1007/s13361-016-1469-y
5. Panuwet P, Hunter RE, D'Souza PE, et al. Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring. *Critical Reviews in Analytical Chemistry*. 2016;46(2):93-105. doi:10.1080/10408347.2014.980775
6. Cortese M, Gigliobianco MR, Magnoni F, Censi R, Di Martino P. Compensate for or Minimize Matrix Effects? Strategies for Overcoming Matrix Effects in Liquid Chromatography-Mass Spectrometry Technique: A Tutorial Review. *Molecules*. 2020;25(13):3047. doi:10.3390/molecules25133047
7. Liao HW, Chen GY, Tsai IL, Kuo CH. Using a postcolumn-infused internal standard for correcting the matrix effects of urine specimens in liquid chromatography–electrospray ionization mass spectrometry. *Journal of Chromatography A*. 2014;1327:97-104. doi:10.1016/j.chroma.2013.12.066
8. Liao HW, Chen GY, Wu MS, Liao WC, Tsai IL, Kuo CH. Quantification of endogenous metabolites by the postcolumn infused-internal standard method combined with matrix normalization factor in liquid chromatography–electrospray ionization tandem mass spectrometry. *Journal of Chromatography A*. 2015;1375:62-68. doi:10.1016/j.chroma.2014.11.073
9. Rossmann J, Renner LD, Oertel R, El-Armouche A. Post-column infusion of internal standard quantification for liquid chromatography-electrospray ionization-tandem mass spectrometry analysis – Pharmaceuticals in urine as example approach. *Journal of Chromatography A*. 2018;1535:80-87. doi:10.1016/j.chroma.2018.01.001
10. Huang M, Li HY, Liao HW, et al. Using post-column infused internal standard assisted quantitative metabolomics for establishing prediction models for breast cancer detection. *Rapid Communications in Mass Spectrometry*. 2020;34(S1):e8581. doi:10.1002/rcm.8581
11. González O, Dubbelman AC, Hankemeier T. Postcolumn Infusion as a Quality Control Tool for LC-MS-Based Analysis. *J Am Soc Mass Spectrom*. 2022;33(6):1077-1080. doi:10.1021/jasms.2c00022
12. Tisler S, Pattison DI, Christensen JH. Correction of Matrix Effects for Reliable Non-target Screening LC–ESI–MS Analysis of Wastewater. *Anal Chem*. 2021;93(24):8432-8441. doi:10.1021/acs.analchem.1c00357
13. Zhu M, Lamont L, Maas P, et al. Matrix effect evaluation using multi-component post-column infusion in untargeted hydrophilic interaction liquid chromatography-mass spectrometry plasma metabolomics. *Journal of Chromatography A*. 2025;1740:465580. doi:10.1016/j.chroma.2024.465580
14. Jhang RS, Lin SY, Peng YF, et al. Using the PCI-IS Method to Simultaneously Estimate Blood Volume and Quantify Nonvitamin K Antagonist Oral Anticoagulant Concentrations in Dried Blood Spots. *Anal Chem*. 2020;92(3):2511-2518. doi:10.1021/acs.analchem.9b04063
15. Stahnke H, Reemtsma T, Alder L. Compensation of Matrix Effects by Postcolumn Infusion of a Monitor Substance in Multiresidue Analysis with LC−MS/MS. *Anal Chem*. 2009;81(6):2185-2192. doi:10.1021/ac802362s
16. Henrick BM, Rodriguez L, Lakshmikanth T, et al. Bifidobacteria-mediated immune system imprinting early in life. *Cell*. 2021;184(15):3884-3898.e11. doi:10.1016/j.cell.2021.05.030
17. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. *Journal of Chromatography A*. 2014;1353:99-105. doi:10.1016/j.chroma.2014.04.071
18. Sands CJ, Gómez-Romero M, Correia G, et al. Representing the Metabolome with High Fidelity: Range and Response as Quality Control Factors in LC-MS-Based Global Profiling. *Anal Chem*. 2021;93(4):1924-1933. doi:10.1021/acs.analchem.0c03848
19. Zierer J, Jackson MA, Kastenmüller G, et al. The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet*. 2018;50(6):790-795. doi:10.1038/s41588-018-0135-7
20. Cresci GA, Bawden E. Gut Microbiome. *Nutrition in Clinical Practice*. 2015;30(6):734-746. doi:10.1177/0884533615609899

Appendix

Summary

Nederlandse Samenvatting

Curriculum Vitae

List of Publications

Acknowledgements

Summary

The incidence of food allergy has increased over the last few decades, with cow's milk allergy (CMA) being one of the most prevalent food allergies in early life. In recent years, growing research on the gut microbiome has highlighted its crucial role in human health and disease, including its potential impact on CMA in early life. The gut microbiome is thought to exert a dynamic influence on the immune system, thereby potentially regulating the onset and progression of CMA. During this process, gut-derived metabolites have been increasingly recognized as important mediators of the crosstalk between the gut microbiome and the host. This highlights the essential role of metabolomics in deciphering the influence of the gut microbiome on CMA in early life.

In metabolomics, approaches can be broadly categorized into targeted and untargeted metabolomics, with targeted metabolomics focusing on accurate quantification of known metabolites, whereas untargeted metabolomics aims to discover novel biomarkers by performing comprehensive metabolome profiling. Liquid chromatography coupled with mass spectrometry using an electrospray ionization source (LC-ESI-MS) is one of the most applied techniques in metabolomics due to its high sensitivity and robustness. However, because of the ionization mechanism of the ESI source, the LC-ESI-MS method is susceptible to matrix effect, which is caused by co-eluted matrix components and can significantly impact the accuracy and reproducibility of the analysis. The matrix effect remains a major challenge in LC-ESI-MS-based metabolomics, particularly in untargeted metabolomics, where there is lack of effective compensation strategies.

Therefore, the first aim of this thesis is to address the problem of matrix effect in untargeted metabolomics with the technique of post-column infusion of standards (PCIS). The second aim is to gain insights into the crosstalk of the gut-microbiome and food allergy in early life by exploring the fecal metabolome in CMA infants.

Chapter 1 outlines current approaches used to address matrix effect in metabolomics and presents a summary of the interconnections among metabolomics, gut microbiome,

and food allergy in early life. This chapter also provides an overview of the scope of the thesis.

Chapter 2 describes the development of an untargeted metabolomics method incorporating PCIS for matrix effect monitoring in plasma and feces. To assess the analytical performance of this LC-PCIS-MS-based untargeted method, it was validated using diverse stable-isotope labeled (SIL) standards in a targeted manner. The method exhibited good precision, accuracy, recovery, and repeatability. The validation highlighted the issue of matrix effect in the developed method, as it demonstrated that high variation in matrix effect among samples could significantly impact the accuracy and reproductivity of the measurements. To evaluate the matrix effect across the entire chromatogram, the PCIS approach was implemented. This evaluation was performed by post-column infusion of xenobiotic compounds during the injection of blank or diverse matrix samples, enabling the overall monitoring of both absolute matrix effect (AME) and relative matrix effect (RME). The PCIS approach successfully identified chromatographic regions exhibiting severe matrix effect. Moreover, it yielded comparable RME results to those obtained using the traditional post-extraction spiking method, demonstrating its potential as a reliable technique for evaluating RME in untargeted metabolomics.

Chapter 3 extends the application of PCIS from monitoring to matrix effect correction using a novel artificial matrix infusion strategy. The artificial matrix is composed of compounds that interfere with the ESI process of analytes by competing for ionization or increasing the surface tension in droplets, thereby preventing Coulombic explosion. The matrix effect created by the artificial matrix (ME_{art}) for a specific feature can be determined by injecting a matrix sample both with and without the artificial matrix, and subsequently used to select its ideal PCIS for biological matrix effect (ME_{bio}) correction. This approach enabled the matching of metabolic features, including known and unknown ones, to their appropriate PCISs. The concept of this method was validated using diverse SIL standards spiked into plasma, urine, and feces. To incorporate AME and RME into the validation, a matrix effect scoring system was introduced, which calculates the AME and RME scores separately and averages them as the final matrix

effect score. The PCISs selected based on ME_{art} were compared to those selected by biological matrix effect (ME_{bio}), with 17 out of 19 SIL standards (89%) showing consistent PCIS selection, demonstrating the effectiveness of ME_{art} in identifying suitable PCIS. Applying ME_{art}-selected PCISs to correct the ME_{bio} resulted in improved or maintained matrix effect for 100% of SIL standards in plasma, 84% in urine, and 95% in feces. Since ME_{art} can be assessed at any retention time, the study in this chapter suggests that implementing PCIS with artificial matrix infusion shows great potential in identifying suitable PCISs to correct matrix effect for features detected using untargeted metabolomics.

Chapter 4 presents a systematic review of the modifications and post-treatment alterations in the gut microbiome, metabolome, and immune response in children with CMA aged 0-12 years and in CMA animal models. At the taxonomic level, multiple studies consistently reported decreases in *Bifidobacterium* genus and Lactobacillales order, alongside increases in Clostridia class in CMA children. Various intervention approaches, such as different formulas, prebiotics, probiotics, and synbiotics, were applied to manage CMA across several studies. However, a constant increase in *Bifidobacterium* levels was observed only with *Bifidobacterium* strains-specific treatments. Regarding metabolome modification, altered short-chain fatty acids (SCFAs), amino acids, and organic acids were reported in CMA children. These metabolomic changes appeared to be partially restored through interventions, with increased SCFAs levels and balanced amino acid profiles. Limited data were available regarding immune response. Overall, the review highlights that no study has applied multi-omics approaches to investigate the relationship between gut microbiome and CMA in early life. Although several metabolomics studies have been reported, they focused on a limited range of metabolites, emphasizing the need for more comprehensive metabolomics research on CMA in early life.

To fill the gap identified in **Chapter 4**, **Chapter 5** explores the fecal metabolome in CMA infants undergoing dietary intervention with and without synbiotic supplementation (inulin, oligofructose and *Bifidobacterium breve* M-16 V). By grouping the infants based on their CMA status after one year or the type of intervention

they received, the impacts of CM tolerance acquisition and the influence of the synbiotic supplementation on the fecal metabolome of CMA infants were investigated. More pronounced changes in fecal amino acids, bile acids, and SCFAs were observed the infants who acquired tolerance. The tolerant group showed significantly higher levels of lysine and citrulline, and potential evidence of downregulated tryptophan-serotonin metabolism, upregulated secondary bile acid production, and increased butyrate level. These alterations might suggest a healthier gut with improved barrier function and a more mature gut microbiome in the CM-tolerant group. Regarding the impact of the synbiotic, this study demonstrated that the synbiotic significantly altered the fecal levels of aromatic lactic acids, purine metabolites, fatty acids, and bile acids, especially after six months of supplementation. The significant increase of the two aromatic lactic acids (4-hydroxyphenyllactic acid and indolelactic acid), known as infant-type *Bifidobacterium*-derived metabolites, suggested an enhanced abundance and/or activity of infant-type *Bifidobacterium* species, indicating the successful supplementation of the synbiotic.

Chapter 6 concludes the thesis with a general summary and discussion. It outlines potential improvements in the implementation of PCIS for addressing matrix effects in untargeted metabolomics and provides recommendations and perspectives on applying metabolomics to study the gut microbiome and CMA in early life.

Samenvatting

De incidentie van voedselallergie is de afgelopen decennia toegenomen, waarbij koemelkallergie (KMA) een van de meest voorkomende voedselallergieën in de vroege kindertijd is. In recente jaren heeft groeiend onderzoek naar het darm microbioom de cruciale rol ervan in de menselijke gezondheid en ziekte benadrukt, inclusief de potentiële impact op KMA in de vroege kindertijd. Men denken dat het darm microbioom een dynamische invloed uitoefent op het immuunsysteem, en daarmee mogelijk de aanvang en progressie van KMA reguleert. Tijdens dit proces worden in de darmen geproduceerde metabolieten in toenemende mate erkend als belangrijke mediator van de wisselwerking tussen het darm microbioom en de gastheer. Dit onderstreept de essentiële rol van metabolomics in het ontrafelen van de invloed van het darm microbioom op KMA bij jonge kinderen.

In metabolomics kunnen metabolomics methoden grofweg worden ingedeeld in targeted (doelgericht) en untargeted (niet specifiek gericht): bij targeted metabolomics ligt de nadruk op nauwkeurige kwantificering van bekende metabolieten, terwijl untargeted metabolomics streeft naar de ontdekking van nieuwe biomarkers via uitgebreide profilering van het metaboloom. Vloeistofchromatografie gekoppeld aan massaspectrometrie met electrospray-ionisatie (LC-ESI-MS) is een van de meest toegepaste technieken binnen metabolomics vanwege zijn hoge sensitiviteit en robuustheid. Echter, door het ionisatiemechanisme van de ESI-bron is de LC-ESI-MS-methode gevoelig voor matrixeffecten, die worden veroorzaakt door mee-eluerende matrixcomponenten en de nauwkeurigheid en reproduceerbaarheid van de analyse aanzienlijk kunnen beïnvloeden. Dit blijft een grote uitdaging, met name in untargeted metabolomics, waar effectieve compensatiestrategieën ontbreken.

Daarom is het eerste doel van dit proefschrift om het probleem van matrixeffecten in untargeted metabolomics aan te pakken met de techniek van *post*-kolom infuseren van standaarden (PCIS). Het tweede doel is om inzicht te krijgen in de wisselwerking tussen het darmmicrobioom en voedselallergie in de vroege kindertijd door het fecale metaboloom te onderzoeken bij zuigelingen met KMA.

Hoofdstuk 1 schetst de huidige methodes voor het tegengaan van matrixeffecten in metabolomics en biedt een overzicht van de interconnecties tussen metabolomics, het darmmicrobioom en voedselallergie in de vroege kindertijd. Tevens wordt de reikwijdte van het proefschrift uiteengezet.

Hoofdstuk 2 beschrijft de ontwikkeling van een untargeted metabolomics-methode met PCIS voor het monitoren van matrixeffecten in plasma en feces. De analytische prestatie van deze LC-PCIS-MS-methode werd beoordeeld met targeted analyse van diverse stabiel isotopen-gelabelde (SIL) standaarden. De methode vertoonde goede precisie, nauwkeurigheid, opbrengst en reproduceerbaarheid. Validatie toonde aan dat variatie in matrixeffecten tussen monsters de meetnauwkeurigheid aanzienlijk kan verminderen. Door PCIS te implementeren via post-column infusion van xenobiotische verbindingen tijdens injectie van blanco- en verschillende matrixmonsters, konden zowel absolute (AME) als relatieve (RME) matrixeffecten in kaart worden gebracht. De PCIS-aanpak identificeerde succesvol chromatografische gebieden met ernstige matrixeffecten en leverde resultaten op die vergelijkbaar waren met traditionele *post*-extractie spike-methoden, wat de betrouwbaarheid als techniek voor RME-evaluatie in untargeted metabolomics bevestigt.

Hoofdstuk 3 breidt de toepassing van PCIS uit van monitoren naar matrix effect correctie door een innovatieve strategie van infusie van artificiële matrix. De artificiële matrix bestaat uit verbindingen die het electrospray-ionisatie proces verstoren door te concurreren in de ionisatie of de door de oppervlaktespanning van druppels te verhogen, waardoor Coulomb explosie wordt tegengegaan. Het matrixeffect gecreëerd door de artificiële matrix (ME_{art}) voor een specifieke feature wordt bepaald door monsters zowel met als zonder de artificiële matrix te injecteren; dit wordt vervolgens gebruikt om de ideale PCIS te selecteren voor correctie van biologische matrixeffecten (ME_{bio}). Validatie met diverse SIL-standaarden in plasma, urine en feces liet zien dat PCIS-selectie op basis van ME_{art} in 89% van de gevallen overeenkwam met ME_{bio}-selectie. Correctie met PCIS geselecteerd via ME_{art} resulteerde in vergelijkbare of verbeterde matrix effect scores voor respectievelijk 100% (plasma), 84% (urine) en 95% (feces). Aangezien ME_{art} op elk retentietijdstip kan worden beoordeeld, suggereert dit hoofdstuk

dat PCIS gecombineerd met artificiële matrixinfusie veel potentie heeft om matrix effect correctie te realiseren voor features gevonden met untargeted metabolomics.

Hoofdstuk 4 geeft een systematisch literatuurreview van aanpassingen en veranderingen na behandeling in het darm microbioom, metaboloom en immuunrespons bij kinderen (0-12 jaar) met KMA en diermodellen met KMA. Op het taxonomische niveau meldden meerdere studies afnamen in *Bifidobacterium* genus en Lactobacillales-orde, naast toenames in de klasse Clostridia bij kinderen met KMA. Interventiebenaderingen zoals verschillende voedingen, prebiotica, probiotica en synbiotica werden onderzocht. Echter, een duurzame toename van *Bifidobacterium*-niveaus werd alleen gezien bij behandelingen specifiek met *Bifidobacterium*-strains. Veranderingen in het metaboloom betroffen korte keten vetzuren (KKVZ), aminozuren en organische zuren; interventies hielpen deze gedeeltelijk te herstellen met verhoogde KKVZ-niveaus en genormaliseerde aminozuurprofielen. Er is slechts beperkt bewijs over immuunrespons. De review onderstreept dat er nog geen studies zijn die multi-omics benaderingen hebben toegepast om de relatie tussen darm microbioom en KMA in de vroege kindertijd te onderzoeken. De behoefte aan breder metabolomics onderzoek bij KMA is daarmee evident.

Hoofdstuk 5 sluit aan op de lacune beschreven in **hoofdstuk 4** en onderzoekt het fecale metaboloom van zuigelingen met KMA die een dieet kregen met of zonder synbiotische suppletie (inuline, oligofructose en *Bifidobacterium breve* M-16V). Door de zuigelingen in te delen op basis van KMA-status na één jaar of het type interventie, zijn de effecten van zowel KMA-tolerantie-acquisitie als synbiotische suppletie op het fecale metaboloom onderzocht. Zuigelingen die tolerant werden, vertoonden significant meer veranderingen in fecale aminozuren, galzuren en KKVZ, met hogere niveaus van lysine en citrulline, plus aanwijzingen voor verminderde tryptofaan-serotoninemetabolisme, toegenomen secundaire galzuurproductie en verhoogde butyraatspiegels. Dit wijst mogelijk op een gezonder darmmilieu met verbeterde barrièrefunctie en meer volwassen microbiota bij de tolerante groep. Verder toonde de synbiotische suppletie na zes maanden aanzienlijke veranderingen in fecale aromatische melkzuren, purinemetabolieten, vetzuren en galzuren. Een significante stijging van de aromatische

melkzuren (4-hydroxyphenyllactaat en indolelactaat), typische metabolieten van infant-type *Bifidobacterium*, suggereerde een verhoogde aanwezigheid en activiteit van deze soorten, wat de effectiviteit van de synbiotische suppletie bevestigt.

Hoofdstuk 6 sluit het proefschrift af met een algemene samenvatting en discussie. Het schetst mogelijkheden voor verbeteringen in de implementatie van PCIS voor matrix effect correctie in untargeted metabolomics en biedt aanbevelingen en perspectieven voor de toepassing van metabolomics in het bestuderen van het darm microbioom en KMA in de vroege kindertijd.

Curriculum vitae

Pingping Zhu was born on May 25th, 1992, in Fuling, Chongqing, China. After graduating from No.5 High School of Fuling in 2011, she was admitted to Chongqing Medical University for a five-year bachelor's program majoring in Clinical Pharmacy. In 2016, she obtained her bachelor's degree in clinical pharmacy with a minor in English (Medical English direction).

In the same year, she started her master's study at the West China School of Pharmacy, Sichuan University, China, where she was supervised by Prof. Ke Lan, focusing on studying the tertiary metabolism of deoxycholate in human and pre-clinical animal models. During the three-year study, she developed skills in the bioanalysis of small molecular using liquid chromatography – mass spectrometry and gained a strong interest in exploring the crosstalk between human host and gut microbiome through their mutual metabolism.

In September 2019, after obtaining her master's diploma, she continued her PhD journey at the Metabolomics and Analytical Center at Leiden University, under the supervision of Prof. Dr. Thomas Hankemeier, Dr. Amy Harms, and Dr. Anne-Charlotte Dubbelman. Her doctoral research began with the development and advancement of untargeted metabolomics methods using innovative techniques to address the problem of matrix effects. She then dedicated herself to deepening the understanding of the interplay between the gut microbiome and human health by exploring the fecal metabolome in infants with cow's milk allergy.

After her PhD, Pingping is currently working with Dr. Robert Jansen and Prof. dr. LA.M.P. van Niftrik in the Department of Microbiology at Radboud University. As a postdoctoral researcher, she pursues her interests further in identifying gut-microbial metabolite biomarkers for the early prediction of immunotherapy response in patients with non-small cell lung cancer.

List of Publications

Publications related to this thesis

1. **Zhu, P.**, Dubbelman, A. C., Hunter, C., Genangeli, M., Karu, N., Harms, A., & Hankemeier, T. (2024). Development of an untargeted LC-MS metabolomics method with post-column infusion for matrix effect monitoring in plasma and feces. *Journal of the American Society for Mass Spectrometry*, 35(3), 590-602.
2. **Zhu, P.,** Harms, A. C., Maas, P., Bakas, M., Whien, J. J., Dubbelman, A. C., & Hankemeier, T. (2025). Matrix Effects in Untargeted LC-MS Metabolomics: From Creation to Compensation with Post-Column Infusion of Standards. *Journal of Chromatography A*, 466508.
3. Savova, M. V.*, **Zhu, P.***, Harms, A. C., van Der Molen, R. G., Belzer, C., & Hendrickx, D. M. (2024). Current insights into cow's milk allergy in children: Microbiome, metabolome, and immune response—A systematic review. *Pediatric Allergy and Immunology*, 35(2), e14084.
4. **Zhu, P.***, Savova, M. V.*, Kindt, A., PRESTO Study Team, Wopereis, H., Belzer, C., ... & Hankemeier, T. (2025). Exploring the Fecal Metabolome in Infants with Cow's Milk Allergy: The Distinct Impacts of Cow's Milk Protein Tolerance Acquisition and of Synbiotic Supplementation. *Molecular Nutrition & Food Research*, 69(1), e202400583.

*Authors contributed equally

Other publications

1. Savova, M. V., **Zhu, P.**, Kindt, A., the TEMPO study team, Wopereis, H., Belzer, C., Harms, A. C., & Hankemeier, T.. Fecal metabolome alterations in infants at risk of developing allergies during the first year of life. (under submission)
2. Hendrickx, D. M., Savova, M. V., **Zhu, P.**, An, R., Boeren, S., Klomp, K., ... & Presto Study Team. (2025). A multi-omics machine learning classifier for outgrowth of cow's milk allergy in children. *Molecular Omics*.
3. **Zhu, P.**, Zhang, J., Chen, Y., Yin, S., Su, M., Xie, G., ... & Jia, W. (2018). Analysis of human C24 bile acids metabolome in serum and urine based on enzyme digestion

of conjugated bile acids and LC-MS determination of unconjugated bile acids. *Analytical and bioanalytical chemistry*, 410(21), 5287-5300.

4. Chen, Y. J., Zhang, J., **Zhu, P. P.**, Tan, X. W., Lin, Q. H., Wang, W. X., ... & Lan, K. (2019). Stereoselective oxidation kinetics of deoxycholate in recombinant and microsomal CYP3A enzymes: deoxycholate 19-hydroxylation is an in vitro marker of CYP3A7 activity. *Drug Metabolism and Disposition*, 47(6), 574-581.

5. Zhang, J., Gao, L. Z., Chen, Y. J., **Zhu, P. P.**, Yin, S. S., Su, M. M., ... & Lan, K. (2019). Continuum of host-gut microbial co-metabolism: host CYP3A4/3A7 are responsible for tertiary oxidations of deoxycholate species. *Drug Metabolism and Disposition*, 47(3), 283-294.

6. Gui, L., Wu, Q., Hu, Y., Zeng, W., Tan, X., **Zhu, P.**, ... & Lan, K. (2021). Compensatory transition of bile acid metabolism from fecal disposition of secondary bile acids to urinary excretion of primary bile acids underlies rifampicin-induced cholestasis in beagle dogs. *ACS Pharmacology & Translational Science*, 4(2), 1001-1013.

7. Zeng, W., Gui, L., Tan, X., **Zhu, P**., Hu, Y., Wu, Q., ... & Lan, K. (2021). Tertiary oxidation of deoxycholate is predictive of CYP3A activity in dogs. *Drug Metabolism and Disposition*, 49(5), 369-378.

Acknowledgements

Looking back on this wonderful and challenging PhD journey, I am deeply grateful to have crossed path with so many talented and generous people, each of whom imparted invaluable lessons and sights.

First and foremost, I would like to express my sincere gratitude to my promotor and supervisors. Without your guidance, support, and mentorship, I would not be where I am today. Thomas, I would like to extend my deepest appreciation to you for offering me the opportunity to work in such a state-of-the-art lab and alongside an inspiring team. Your enthusiasm for science and dedication to excellence have deeply inspired me. Amy, you have always been the person I could turn to whenever I encountered a problem or need a thoughtful discussion. I could always rely on your clear thinking and constructive suggestions to help move things forward. Your insightful feedback and careful guidance have been instrumental in shaping my academic growth. Anne-Charlotte, I feel so fortunate to have had you as one of my supervisors. I will always cherish our long weekly meetings filled with idea exchanges, constructive planning, and problem-solving. Thank you for your expert guidance, patience, encouragement, and kindness, all of which have profoundly contributed to the completion of my PhD.

I would also like to thank my collaborators, Diana, Clara, Harm, Rennet, and the EARLYFIT study team, for your valuable insights, discussions, and feedback throughout our collaboration. Especially for Mariyana, I am truly grateful for the time we worked together. I will always remember the laughter, intense discussions, and the two publications that grew from our efforts.

To my dear collogues, Michael, Pascal, and Alida, thanks you for being wonderful teammates in the data analysis. To Gerwin, Tim, and Fiasa, I am deeply appreciative of your timely help with daily instrument maintenance and troubleshooting. To Sabine, Sam, Mich, and Charlotte, thank you for the great experience of working together on the development of the untargeted platform. Ina and Ariadna, thank you for your efficiency and kindness in organizing meetings and handling countless administrative matters. To my lovely office mates, Ischa, Vincent, Isabella and Laurens, thanks for the

being such nice and friendly companions. To Madhu, Kanchana, and Manchu, thank you for the laughter and friendship we shared during this journey.

A very special thank you goes to Bingshu. I feel so lucky that we started our PhD on the same day. Your companionship, help, and encouragement over the past years have meant so much to me. I will always miss our countless coffee chats and heartfelt conversations. To Wei Zhang, Yupeng, Wei Yang, Luojiao, Zhengzheng, Xinyu, Congrou, Fujia, Mengle, Chunyuan, Lu, Yu, Xiaoyue, and Yiwei, thank you for all the wonderful memories we've created together, both at work and beyond. Your companionship has been one of the most cherished parts of my PhD, and I know it will last well beyond this journey.

To my friends back home, Qian, Ying, Jun, Xiaomi, and Yujie, thanks you for always being supportive whenever I need you, and for sharing your life moments with me even when we were far apart. Every reunion with you brings me immense joy and gives me strength to keep going.

To Shunsheng, thank you for your unwavering support, understanding, and encouragement throughout this journey. Your companion and belief in me had been my calm in the storm, helping me persevere through the most challenging moments of my PhD.

To my beloved family, my mom, and my sisters Tingting and Yan, thank you for your endless love and support. The courage you give me to not fear taking a step backward has always been the strength that carries me forward through every chapter of my life.

Finally, I would like to extend my heartfelt thanks to everyone I may have unintentionally forgotten to mention. Your support and kindness are deeply appreciated, even if not mentioned by name.